Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Origins and Dispersal of the Sweet Potato and Bottle Gourd in Oceania: Implications for Prehistoric Human Mobility

Andrew Christopher Clarke

2009

Origins and Dispersal of the Sweet Potato and Bottle Gourd in Oceania: Implications for Prehistoric Human Mobility

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Plant Biology

at Massey University, Palmerston North,
New Zealand.

Andrew Christopher Clarke

2009

Copyright © 2009 by Andrew Christopher Clarke

What manner of men were they who by surpassing the achievements of the Phœnicians in the Mediterranean and the Vikings of the north Atlantic are worthy of being called the supreme navigators of history?

Sir Peter Buck (Te Rangi Hiroa)
Vikings of the Sunrise (1938, p. 13)

ABSTRACT

The origins of the sweet potato (*Ipomoea batatas*) and bottle gourd (*Lagenaria siceraria*), two important commensal species in prehistoric Polynesia, have remained elusive. Most recently, a South American origin has been favoured, which prompts a number of interesting questions surrounding how, when, from where and by whom these species dispersed into the Pacific. For this project, hypotheses were formulated based on existing archaeological, linguistic and maritime evidence, and tested using a molecular approach. For both species, extensive marker development was necessary.

For the bottle gourd, a set of seven molecular markers (two chloroplast and five nuclear) was developed to test the hypothesis of a South American origin for the Polynesian bottle gourd. These were sequenced in 36 accessions of bottle gourd from Asia, the Americas and New Zealand. Analyses of these markers support a dual origin for the Polynesian bottle gourd: the chloroplast markers identify an Asian origin, but the nuclear markers reveal alleles that originate in both the Americas and Asia. By combining information from a number of sources, a model for the domestication(s) and global dispersal of the bottle gourd is proposed.

For the sweet potato, the amplified fragment length polymorphism (AFLP) technique was used. First, using a new procedure that will be applicable to other studies, AFLP scoring parameters were optimised to improve phylogenetic resolution. Second, to elucidate sweet potato dispersal in Oceania, AFLP profiles were generated for 270 unique accessions of sweet potato from Asia, Island Melanesia, Polynesia and the Americas. The putative *kumara* lineage, which represents a prehistoric, Polynesian-mediated introduction from South America, was identified. Sweet potato accessions from Asia to Western Polynesia were found to be genetically diverse, and the relationships between them are more complex than previously recognised. The phylogenetic positions of the Māori varieties 'Hutihuti', 'Rekamaroa' and 'Taputini' are inconsistent with these accessions representing pre-European cultivars; instead it is more likely that they are early European introductions.

To answer questions about the prehistoric dispersal patterns of the bottle gourd, future work could make use of high resolution markers and ancient DNA (aDNA) from archaeological and early historic-period samples. Future work on the sweet potato is needed to narrow down the point of Polynesian contact on the South American coast, and to answer this question more intensive sampling is required. Integration of genetic, linguistic, historical and morphological data will also be important.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Professor David Penny. The opportunity to work with you, and in your lab, has been an exciting, challenging and rewarding experience. Your unyielding enthusiasm for biology, relentless pursuit of interesting questions, and blatant disregard for interdisciplinary boundaries have been particularly appreciated. Thank you also for providing funding to complete research here at Massey and to travel and meet the large number of people who have contributed to this project.

I would also like to thank my team of co-supervisors: Dr John McCallum, Professor Peter Lockhart and Professor Lisa Matisoo-Smith. John, thank you for allowing me to spend time in your lab, and for lots of interesting ideas. Pete, thank you for advice on data analysis and encouragement along the way. Lisa, thank you for your advice in interpreting data, and helping me see my research in an anthropological context.

Trish McLenachan helped with everything from PCR-fixing, cloning-fixing and sequencing-fixing, to experimental design, emotional support ("It'll be fine" when I needed it), editing and proof-reading and even the occasional meal and alcoholic beverage. All have been appreciated.

For this project I have relied a great deal on those who have generously provided samples. For bottle gourd samples I especially thank Charles Heiser and Mike Burtenshaw, as well as Richard Cross, John Palmer and Peter Matthews. For allowing me to sample the Yen Sweet Potato Collection in Japan, I thank Makoto Nakatani and staff at the National Institute of Crop Science, Tsukuba. Steve Lewthwaite also provided many sweet potato samples of New Zealand material, Pam Fletcher provided tissue culture material from the Yen Collection, Mary Taylor provided modern samples from the Pacific, and Richard Walter provided samples from Mangaia, Cook Is.

At the National Museum of National History, Smithsonian Institution, I would like to thank Bruce Smith, Noreen Tuross (now of Harvard University), Dave Erickson, Sarah Zabriskie and members of the Laboratories of Analytical Biology. Your generosity in hosting me for an internship was greatly appreciated, as was the opportunity to write a paper with you afterwards.

Roger Green generously provided me with draft copies of his 2005 *Sweet Potato Transfers in Polynesian Prehistory* paper, which greatly helped develop the background material for the project. Roger was also very generous with his time — lending his vast knowledge of Pacific prehistory to the interpretation of the bottle gourd and sweet potato results in archaeological, historical and linguistic contexts.

Heidi Meudt helped with many aspects of the project, including high-throughput DNA extractions, AFLP fingerprinting, scoring and analysis of AFLP data, and proof-reading of various chapters. Heidi's patience even extended to writing an AFLP review with me, an opportunity for which I am very grateful — even if I sometimes wondered whether writing a review paper concurrently with a thesis was really a good idea!

Barbara Holland assisted with the data analysis for both the bottle gourd and sweet potato, lending her extensive phylogenetic expertise to a daunting task. I have particularly appreciated Barbara's ability to see research potential in even the scariest data matrix. If David's "tame mathematician" exists then she is certainly Barbara.

I would like to thank the many others who helped during the course of this project. Mike Burtenshaw was an excellent collaborator on the bottle gourd project. Peter Matthews assisted with the sweet potato sampling in Japan, and the bottle gourd manuscript. Peter's extensive knowledge of Pacific ethnobotany and tendency to ask really hard questions were both appreciated. Ashley Gould provided excellent guidance on the historical aspects of the sweet potato in New Zealand. Meeghan Pither-Joyce and the other members of the Genome Research Lab at Crop & Food Research (now Plant & Food Research), Lincoln taught me many molecular techniques. Ruth Morrison assisted with germinating bottle gourd seeds to obtain fresh tissue. Sue Gardiner's lab at HortResearch (now Plant & Food Research), particularly Mike Cook, helped grind the sweet potato tissue for DNA extraction. Paul Hocquard, Karen Sinclair and Joy Wood

made sure I had all the consumables I required, often at short notice. Warwick Allen wrote several time-saving Perl scripts. Tim White assisted with the network analysis of the bottle gourd. Lorraine Berry did all the capillary electrophoresis of the DNA sequencing and AFLP products, and was always helpful with troubleshooting when things didn't go quite according to plan.

Māori consultation was an important part of the project, and I would like to acknowledge the guidance of Nick Roskruge, Jonathan Procter and Rangitāne O Manawatu, and the late Dell Wihongi.

For stipend funding I would like to acknowledge a Massey University Doctoral Scholarship and the generosity of Pete Lockhart and Mike Hendy. For travel funding to attend conferences, conduct field work and complete lab work away from Massey, I would like to thank the Institute of Molecular BioSciences, the Allan Wilson Centre for Molecular Ecology and Evolution, the Royal Society of New Zealand, the Smithsonian Institution, John McCallum and the New Zealand Institute for Crop and Food Research.

I would also like to thank all the staff, students and visitors — both past and present — of the Allan Wilson Centre and Institute of Molecular BioSciences who have contributed towards this project, especially the occupants of the Farside and PLEB Labs, the Boffin Lounge and the Fish Bowl. I have benefited enormously from everything from help with lab work and literature searches, to interesting discussions about molecular biology, data analysis and anthropology (amongst *many* other things). You have provided a stimulating, supportive and entertaining environment in which to work. I would especially like to acknowledge the good friends I have made during my time here: Richard Carter, Fred Delsuc, Chrissie Down, Barbara Holland, Anu Idicula, Kim McBreen, Heidi Meudt, Matt Phillips, Melanie Pierson, Renae Pratt, Scott Roy and Alice Storey. Of course there are many others I could name, including outside the Allan Wilson Centre.

I would like to acknowledge and thank Andrea Bryant, whose love, support and tolerance were often far in excess of what I deserved. I would also like to thank Andrea's family for being so supportive and welcoming, especially Ralph and Raewyn, and fellow PhD-ers Jeremy and Michelle.

Finally, I would like to acknowledge the support of my family. Mum and Dad, your support — in all areas — has been unwavering for much longer than the course of this degree, but has been particularly appreciated during the last few years. Thank you also to my sister Sarah, who is always encouraging despite still having no idea what it is I actually *do*.

PREFACE

The sweet potato (*Ipomoea batatas*) and bottle gourd (*Lagenaria siceraria*) were two crop species fundamental to many agricultural systems in Polynesia. A long-standing scientific interest in these species centres on their origins; both are thought to have arrived in the Pacific from South America. If the prehistoric Oceanic lineages of these species are of South American origin, then this prompts a number of interesting questions surrounding how, when, from where and by whom these species were dispersed into the Pacific. Based on existing archaeological, linguistic, botanical and sailing technology evidence we are able to formulate testable hypotheses around the dispersal of these species. These hypotheses are tested in this thesis using a molecular (DNA) approach.

The thesis is divided into five chapters. The first chapter is a general introduction that frames the remaining research by providing an overview of prehistoric human settlement and mobility in the Pacific, with the focus on the contribution of molecular studies of humans and their commensals (animals and plants) to understanding human mobility in the Pacific. The discussion then narrows to the question of contact between Eastern Polynesia and the Americas, specifically the evidence and likelihood of such contact. Chapter One concludes with the aims and hypotheses of the research.

Chapter Two describes the molecular strategies employed for the bottle gourd and sweet potato. The chapter begins by outlining the considerations and approaches for developing appropriate molecular markers for a given taxonomic group and scientific question, with a focus on closely-related taxa where genetic variation is relatively low. In the context of these considerations, the molecular methods of choice are justified for both the bottle gourd and sweet potato. For the bottle gourd, PCR and sequencing markers derived from inter-SSR multilocus genetic fingerprints (so-called ISSR-derived SCAR markers) were chosen as an appropriate marker system. The development of these, which employed a novel combination of existing molecular methods, is described in detail. For the sweet potato, AFLP fingerprinting was chosen as an appropriate

marker system. Although the AFLP technique is firmly established and widely used, there are several aspects of the method which require further development — primarily scoring, i.e., methods for converting raw AFLP profiles into a binary data matrix. During the course of the AFLP work it was found, in collaboration with Drs Barbara Holland and Heidi Meudt, that the widely-used parameters for generating binary data matrices from unedited AFLP profiles are non-optimal, and that by adjusting these parameters the number of characters can be significantly increased and the amount of homoplasy significantly decreased.

Chapter Three describes research on the origins and dispersal of the bottle gourd in Oceania. The introduction brings together disparate and rarely synthesised bottle gourd research, allowing the Oceanic origins of this species to be placed in context: as with humans themselves, the introduction of the bottle gourd into the Pacific represents the most recent event in the global dispersal of this species. Because the bottle gourd occurs in Africa, Asia and the Americas, gourds from all areas are germane to the origins in Oceania. A range of bottle gourd accessions from Asia, the Americas and Oceania were obtained and these are described. This is followed by the methods, which describe how the five SCAR markers developed in Chapter Two were used, along with two additional chloroplast markers, to amplify and sequence polymorphic loci from 36 accessions of bottle gourd. The analyses of these data are presented, followed by a discussion of Oceanic dispersal scenarios compatible with the results. The implications of bottle gourd research using ancient DNA, which was undertaken at the Smithsonian Institution, Washington, D.C. at the same time as my own research (and with which I was involved), are also discussed. The discussion is extended to the global dispersal and domestication(s) of the bottle gourd, as the use of the chloroplast markers in two outgroup species and the reinterpretation of the archaeological data allow some tentative conclusions to be made in these areas.

Chapter Four deals with the origins and dispersal of the sweet potato in Oceania. The introduction reviews the literature in this area. Unlike the bottle gourd, a reasonably clear and consistent picture is emerging for the sweet potato; the field was brought together in 1974 with the publication of Douglas Yen's landmark *The Sweet Potato and Oceania: An Essay in Ethnobotany* (Yen, 1974), and was resynthesised in a multidisciplinary, multi-author volume in 2005 (Ballard *et al.*, 2005). Over 400 sweet

potato accessions from Oceania were sampled for this study, and these are described. This is followed by the methods, which describe the application of the amplified fragment length polymorphism (AFLP) multilocus fingerprinting technique to the analysis of approximately 300 of the sweet potato accessions. The optimised scoring parameters determined in Chapter Two were used to convert the raw AFLP profiles into a binary data matrix AFLP. The matrix was used to construct phylogenetic trees that have been interpreted, with assistance from Emeritus Professor Roger Green, in the context of available linguistic and historical data. The large number of taxa, wide geographic coverage in Oceania, narrow geographic coverage in the Americas, and large amount of historical, linguistic and morphological information throw open a multitude of avenues for further research and interpretation. While some of these avenues were pursued here, more detailed work is necessary in certain areas, e.g., the origins of prehistoric and historic sweet potato in New Zealand.

Chapters Two, Three and Four, which form the core of the thesis, are published, or intended for publication, as scientific papers (full bibliographic information below). Chapter Two includes papers published in *Trends in Plant Science* (Meudt & Clarke, 2007) and *Systematic Biology* (Holland *et al.*, 2008), and a manuscript intended for publication in *Plant Systematics and Evolution* or similar (Clarke & McLenachan, *in prep*). All three of these publications have been, or are being, written in collaboration with other researchers, and therefore contain additional material not presented in the thesis. Chapter Three is already published as two papers — one in *Molecular Biology and Evolution* (Clarke *et al.*, 2006), and one in the *Proceedings of the National Academy of Sciences of the United States of America* (Erickson *et al.*, 2005). The Erickson *et al.* (2005) paper, which was written in collaboration with researchers at the Smithsonian Institution, contains material not presented in the thesis. A third bottle gourd publication (on the origin, domestication and global dispersal of the species) is being prepared for *Horticultural Reviews* (Clarke & Penny, *in prep*). Chapter Four is intended for publication in a multi-disciplinary science journal.

Chapter Five includes a general summary, with the goal of bringing together research described in the other chapters and placing it in the context of human mobility in the Pacific and the extent to which it addresses the question of contact between prehistoric Polynesia and the Americas. Chapter Five also suggests a number of avenues

of future research for both the sweet potato and bottle gourd — to further elucidate the origins and dispersal of these species in Oceania and in other regions in which they are found. The potential of other crop species as markers of prehistoric human contact between Polynesian and the New World is briefly outlined.

Appendices 1–10 are provided as hard copies at the end of the thesis, and appendices 11–21 are on the enclosed CD. Appendices include accession lists for the sweet potato, oligonucleotide sequences, DNA sequence data matrices for bottle gourd, AFLP matrices for sweet potato, and reprints of published papers.

Relevant PhD Publications

	Appendix
Meudt, H. M., and **Clarke, A. C.** 2007. Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends in Plant Science* **12:** 106–117.	7
Holland, B. R.,[†] **Clarke, A. C.**,[†] and Meudt, H. M.[†] 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology* **57:** 347–366. [†] *Equal contributors*	8
Clarke, A. C., Burtenshaw, M. K., McLenachan, P. A., Erickson, D. L., and Penny, D. 2006. Reconstructing the origins and dispersal of the Polynesian bottle gourd (*Lagenaria siceraria*). *Molecular Biology and Evolution* **23:** 893–900.	9
Erickson, D. L., Smith, B. D., **Clarke, A. C.**, Sandweiss, D. H., and Tuross, N. 2005. An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 18315–18320.	10

TABLE OF CONTENTS

Abstract v
Acknowledgements vii
Preface xi
Abbreviations xxiii
List of Figures xxv
List of Tables xxvii

CHAPTER 1: GENERAL INTRODUCTION 1
1.1 Chapter Overview — Research Context 2
1.1.1 A Note on Dates 3
1.2 Human Origins, Migration and Mobility in Oceania 4
1.2.1 Introduction 4
1.2.2 The Settlement of Oceania 4
1.2.2.1 The First Wave (from 60–40,000 yr BP) 4
1.2.2.2 The Austronesian Expansion and the Settlement of Remote Oceania 6
1.2.2.3 Human Genetic Variation in Oceania 10
1.2.2.4 The Lapita Cultural Complex and Commensal Species 12
1.3 Commensal Animals and Plants in Oceania 14
1.3.1 The Commensal Model 14
1.3.2 Application of Commensal Models 15
1.4 Sailing from Polynesia to the New World in Prehistory 17
1.4.1 Evidence of Polynesia–New World Contact 17
1.4.1.1 Crop Species 17
1.4.1.2 The Chicken (*Gallus gallus*) 18
1.4.1.3 Cultural Evidence 19
1.4.1.4 Human Genetic Evidence 20
1.4.2 Sailing to the New World Coast (and Back) 21
1.5 Aims and Hypotheses 23

CHAPTER 2: MOLECULAR MARKERS IN BOTTLE GOURD AND SWEET POTATO 25
2.1 Chapter Overview 26
2.1.1 A Note on Attribution 27
2.2 Introduction 28
2.2.1 Molecular Markers and the Detection of Genetic Variation 28
2.2.1.1 Multilocus DNA Fingerprinting 29
2.2.1.2 Low-Copy Nuclear Markers versus Multilocus Markers 30

2.2.2 Considerations in Choosing Molecular Marker Systems 31
2.2.3 Molecular Marker Systems: Strategies in Crop Plants 32
2.2.4 Molecular Markers in Bottle Gourd: Considerations, Available Markers and Strategies 34
2.2.5 Molecular Markers In Sweet Potato: Considerations, Available Markers and Strategies 36
2.3 SCAR Markers: Introduction and Methodology 38
2.3.1 An Introduction to SCAR Markers 38
2.3.1.1 Development of SCAR Markers 38
2.3.1.2 SCAR Markers Versus Multilocus Fingerprinting 41
2.3.1.3 Sources of SCAR Markers 41
2.3.2 SCAR Markers Derived from Inter-SSR (ISSR) Fingerprinting 42
2.3.2.1 ISSR Fingerprinting: Introduction 42
2.3.2.2 ISSR Fingerprinting: Methodology 43
2.3.3 Problems and Strategies in SCAR Marker Development 44
2.3.3.1 Obtaining Flanking Regions to Improve SCAR Marker Development 46
2.3.4 Using TAIL PCR to Improve SCAR Marker Development 48
2.3.4.1 Thermal Asymmetric Interlaced (TAIL) PCR: Introduction 48
2.3.4.2 TAIL PCR: Methodology 48
2.3.5 Overview of ISSR-Derived SCAR Marker Development 52
2.3.6 Aims 53
2.4 Development of Bottle Gourd SCAR Markers: Methods 54
2.4.1 Preliminary Technical Remarks 54
2.4.2 ISSR Fingerprinting and Isolation of Polymorphic ISSRs 54
2.4.2.1 ISSR Fingerprinting 54
2.4.2.1.1 Materials and DNA Extraction 54
2.4.2.1.2 ISSR PCR 55
2.4.2.1.3 Betaine as an Enhancing Agent for PCR 56
2.4.2.1.4 Agarose Gel Electrophoresis of ISSR PCR Products 57
2.4.2.2 Polyacrylamide Gel Electrophoresis (PAGE) of ISSR PCR Products 58
2.4.2.2.1 Preparation of Polyacrylamide Electrophoretic Glass Plates 58
2.4.2.2.2 Preparation of Acrylamide Gel Solution 59
2.4.2.2.3 Setup of the Electrophoresis Apparatus 59
2.4.2.2.4 Preparation of DNA Size Ladder 60
2.4.2.2.5 Preparation of ISSR PCR Products for PAGE 60
2.4.2.2.6 Loading and Running of ISSR PAGE Gels 60
2.4.2.2.7 Detection of DNA Fragments by Silver Staining of the ISSR Gels 61
2.4.2.3 Isolation and Re-Amplification of Polymorphic ISSRs 62
2.4.2.3.1 Excision of Polymorphic Bands From the ISSR PAGE Gels 62
2.4.2.3.2 Re-Amplification from Excised Band Eluate 62
2.4.2.3.3 Agarose Gel Extraction of Re-Amplification PCR Products 63
2.4.2.3.4 NanoDrop® Spectrophotometry 63
2.4.3 Conversion of ISSRs into SCAR Markers 64
2.4.3.1 Cloning and DNA Sequencing of ISSRs 64
2.4.3.1.1 Overview of the TA Cloning and Blue–White Screening Techniques 64
2.4.3.1.2 Ligation of Re-Amplified ISSRs into pGEM®-T Easy 65

2.4.3.1.3 Transformation of *Escherichia coli* DH5α™ 66
2.4.3.1.4 Blue–White Screening, Colony PCR and Inoculation of LB Broth.... 67
2.4.3.1.5 Plasmid DNA Extraction and Digestion.... 68
2.4.3.1.6 DNA Sequencing of Insert DNA 69
2.4.3.1.7 Purification of DNA Sequencing Products by Ethanol Precipitation 69
2.4.3.1.8 Capillary Electrophoresis of DNA Sequencing Reaction Products 70
2.4.3.1.9 DNA Sequence Data Analysis.... 70
2.4.3.2 TAIL PCR of ISSR Flanking Regions 71
2.4.3.2.1 Design of TAIL PCR Primers.... 71
2.4.3.2.2 1°, 2° and 3° TAIL PCR 71
2.4.3.2.3 Agarose Gel Extraction of TAIL PCR Products 72
2.4.3.2.4 Ethanol Precipitation of TAIL PCR Products 73
2.4.3.2.5 DNA Sequencing of Gel-Extracted TAIL PCR Products 74
2.4.3.2.6 DNA Sequence Data Analysis.... 74
2.4.3.3 Development of ISSR SCAR Markers.... 74
2.4.3.3.1 ISSR SCAR Marker Primer Design 74
2.4.3.3.2 Optimisation PCR for ISSR SCAR Markers.... 75
2.4.3.3.3 Trial PCR of ISSR SCAR Markers on Several Accessions 76
2.4.3.3.4 SAP/Exo I Cleanup of PCR Products.... 76
2.4.3.3.5 DNA Sequencing of Trial ISSR SCAR Markers 76
2.4.3.3.6 DNA Sequence Data Analysis.... 77
2.5 Development of Bottle Gourd SCAR Markers: Results and Discussion.... 78
2.6 Screening of Other Markers for Bottle Gourd 84
2.6.1 Screening of Chloroplast Markers 84
2.6.2 Screening of Nuclear Markers 85
2.6.3 *trn*S–*trn*G Marker from Inter-ccSSR PCR 85
2.7 The AFLP Technique: Introduction and Methodology 87
2.7.1 Introduction 87
2.7.2 Generating an AFLP Fingerprint: Methodology.... 88
2.8 Optimising AFLP Scoring Parameters in Sweet Potato: Introduction.... 93
2.8.1 Aims 98
2.9 Optimising AFLP Scoring Parameters in Sweet Potato: Materials and Methods.... 99
2.9.1 Sampling Strategy.... 99
2.9.2 Generation of Raw AFLP Data.... 100
2.9.3 Generation of Datasets using Different Automated Scoring Parameter Settings.... 101
2.9.4 Comparison of Datasets to Determine Optimal Parameter Settings 104
2.9.4.1 Measures of Accuracy.... 104
2.9.4.2 Optimal Parameter Settings 107
2.9.4.3 Robustness of the Phylogenies to Changes in Parameter Settings.... 107
2.10 Optimising AFLP Scoring Parameters in Sweet Potato: Results and Discussion.... 108
2.10.1 Measures of Accuracy 108
2.10.1.1 Phylogenetic Resolution 108
2.10.1.2 Correct Assignment of Replicates.... 111
2.10.1.3 Error Rates 112

2.10.2 Optimal Parameter Settings 116
2.10.2.1 GeneMapper® 116
2.10.2.2 GeneMarker® 117
2.10.3 Robustness of the Phylogenies to Changes in Parameter Settings 118
2.10.4 Additional Independent Evidence 123
2.11 Optimising AFLP Scoring Parameters in Sweet Potato: Conclusions 124
CHAPTER 3: ORIGINS AND DISPERSAL OF THE BOTTLE GOURD IN OCEANIA 127
3.1 Chapter Overview 128
3.2 Introduction 130
3.2.1 Biology of the Bottle Gourd 130
3.2.1.1 Taxonomy, Systematics and Genetics 131
3.2.1.2 Reproductive Biology 133
3.2.2 Human Uses of the Bottle Gourd, Especially in Polynesia 133
3.2.3 The Bottle Gourd in Africa — Evolution and Domestication 135
3.2.3.1 An African Origin for the Bottle Gourd? 135
3.2.3.2 Discovery of the Wild African Bottle Gourd and its Pollinator? 136
3.2.3.3 Domestication of the Bottle Gourd and the Archaeological Record 137
3.2.4 Global Dispersal of the Bottle Gourd 138
3.2.4.1 The Bottle Gourd in the New World 138
3.2.4.2 The Bottle Gourd in Asia 144
3.2.4.3 A Pre-Austronesian Expansion Pacific Rim Distribution 145
3.2.5 The Bottle Gourd in Island Southeast Asia and Oceania 149
3.2.5.1 The Bottle Gourd in the Western Pacific 149
3.2.5.2 The Wax Gourd in Polynesia 152
3.2.5.3 Distribution of the Polynesian Bottle Gourd 153
3.2.5.4 The Polynesian Bottle Gourd: A New World Origin? 155
3.2.5.5 Natural or Human-Mediated Dispersal from the New World? 158
3.2.5.6 The Bottle Gourd in New Zealand 159
3.2.5.7 A Dual Origin for the Eastern Polynesia Bottle Gourd? 163
3.2.6 Summary 165
3.2.7 Hypothesis and Aims 166
3.2.7.1 Hypothesis 166
3.2.7.2 Aims 167
3.3 Materials and Methods 168
3.3.1 Overview 168
3.3.2 Preliminary Technical Remarks 168
3.3.3 Collection of Bottle Gourd Accessions 168
3.3.4 Genomic DNA Extraction 171
3.3.4.1 Germination of Bottle Gourd Seeds 171
3.3.4.2 CTAB DNA Extraction 172

3.3.4.3 Qiagen DNeasy® DNA Extraction 175
3.3.4.4 Agarose Gel Electrophoresis of Genomic DNA 176
3.3.5 PCR and Sequencing of Chloroplast and ISSR SCAR Markers 176
3.3.5.1 PCR of Chloroplast and SCAR Markers 177
3.3.5.2 DNA Sequencing of Chloroplast and ISSR SCAR Markers 178
3.3.5.2.1 DNA Sequencing of Chloroplast and ISSR SCAR Markers 179
3.3.5.2.2 Purification of DNA Sequencing Products by CleanSEQ® 179
3.3.5.2.3 Capillary Electrophoresis of Chloroplast and ISSR SCAR Markers 179
3.3.5.3 Editing and Alignment of DNA Sequence Data 180
3.3.6 DNA Sequence Data Analysis 180
3.3.6.1 Genotype Frequency Pie Charts 180
3.3.6.2 Network Analysis 181
3.4 Results 182
3.4.1 DNA Sequence Data Analysis 182
3.4.1.1 Genotype Frequency Pie Charts 182
3.4.1.2 Network Analysis 185
3.4.1.3 Outgroup Analysis 186
3.5 Discussion 187
3.5.1 Origins of the Polynesian Bottle Gourd 187
3.5.1.1 Genetic Evidence 187
3.5.1.2 Reassessing the Evidence Against an Asian Origin 187
3.5.1.3 Summary 188
3.5.2 Reconciling Erickson *et al.* (2005) and Clarke *et al.* (2006) 188
3.5.2.1 Background to Collaboration 189
3.5.2.2 Summary of Erickson *et al.* (2005) 189
3.5.2.3 Implications of a Non-African Origin for the New World Bottle Gourd 192
3.5.2.4 Implications for Clarke *et al.* (2006) 193
3.5.2.4.1 Hypothesis 1: Introgression of Modern Māori Samples 194
3.5.2.4.2 Hypothesis 2: A Non-Asian, Non-American Origin for Polynesian ssp. *siceraria* Alleles? 194
3.5.2.4.3 Hypothesis 3: Archaeological Bottle Gourd Seeds in Peru: An African Origin? 195
3.5.2.5 A Reconciliation? 200
3.5.3 Domestication and Dispersal 201
3.5.4 Assumptions, Limitations and Caveats 201
3.6 Conclusions 204
3.6.1 Molecular Markers in Bottle Gourd 204
3.6.2 The Polynesian Bottle Gourd 204
3.6.3 Domestication and Dispersal 206
CHAPTER 4: ORIGINS AND DISPERSAL OF THE SWEET POTATO IN OCEANIA 209
4.1 Chapter Overview 210
4.2 Introduction 211
4.2.1 The Sweet Potato in Oceania — Hypotheses 211

4.2.2 Sweet Potato Evolution, Domestication and Distribution in the Americas 212
4.2.3 The Sweet Potato in Oceania 214
4.2.3.1 The Linguistic Aspect....214
4.2.3.2 Distribution and Dispersal of the Sweet Potato in Polynesia....216
4.2.4 The Possibility of Natural Transfer? 219
4.2.5 Hypotheses and Aims....220
4.3 Materials and Methods....221
4.3.1 Overview....221
4.3.2 Preliminary Technical Remarks 222
4.3.3 Collection of Sweet Potato Samples 222
4.3.4 Sampling Strategy 223
4.3.5 DNA Extraction 225
4.3.5.1 Leaf Tissue Disruption 225
4.3.5.2 Qiagen DNeasy® DNA Extraction....225
4.3.5.3 Quantification of DNA Yield and Assessment of DNA Quality 227
4.3.5.3.1 Agarose Gel Electrophoresis 227
4.3.5.4 Concentration of DNA for AFLP 227
4.3.6 AFLP Fingerprinting....228
4.3.6.1 Restriction Endonuclease Digestion of DNA 228
4.3.6.2 Ligation of Oligonucleotide Linkers to the Restriction Fragments 229
4.3.6.2.1 Preparation of Oligonucleotide Linkers 229
4.3.6.2.2 Ligation of Linkers to the Restriction Fragments 230
4.3.6.3 Pre-Selective Amplification of Ligation Products....230
4.3.6.4 Selective Amplification of Pre-Selective PCR Products 231
4.3.6.4.1 Primer Screening of Selective PCR Primers....231
4.3.6.4.2 Selective Amplification of Pre-Selective PCR Products 233
4.3.6.5 Capillary Electrophoresis of AFLPs....234
4.3.6.5.1 Poolplexing of Selective Amplification PCR Products 234
4.3.6.5.2 Capillary Electrophoresis of Selective Amplification PCR Products....235
4.3.7 AFLP Scoring....235
4.3.8 Phylogenetic Reconstruction....236
4.3.8.1 Preliminary Phylogenetic Analyses....236
4.3.8.2 Choosing Taxa to Omit for Final Analysis....237
4.3.8.3 Final Phylogenetic Analysis 238
4.3.8.4 Path Lengths Between Replicates....238
4.4 Results and Discussion 240
4.4.1 Proportion of “1” Character States....240
4.4.2 Neighbour-Joining Tree of Complete Dataset....242
4.4.3 Path Lengths Between Replicates 245
4.4.4 Origins of the Oceanic Sweet Potato Lineages 246
4.4.5 Origins of the New Zealand Sweet Potato Lineages 250
4.5 Conclusions....254

CHAPTER 5: SUMMARY AND FUTURE WORK **257**
5.1 Summary 258
5.2 Future Work: Bottle Gourd 259
5.2.1 The Oceanic Bottle Gourd 259
5.2.1.1 Ancient DNA 260
5.2.1.2 Additional Sampling and Germplasm Conservation 264
5.2.1.3 High Resolution DNA Sequencing Markers 264
5.2.1.4 Linguistic Analysis 265
5.2.2 Domestication and Dispersal 267
5.2.2.1 Landrace Sampling 267
5.2.2.2 Ancient DNA and Histological Examination 268
5.3 Future Work: Sweet Potato 270
5.3.1 Sweet Potato Phylogenetics 270
5.3.2 Improved Sampling 270
5.3.3 An Integrated Approach 271
5.4 Commensal Plants and New World Contact 272

REFERENCES **273**

APPENDICES **311**
Appendix 1: Markers Screened for Bottle Gourd 312
Appendix 2: Bottle Gourd PCR and Sequencing Primers 314
Appendix 3: Bottle Gourd ISSR Sequences 317
Appendix 4: Yen Sweet Potato Accessions 320
Appendix 5: AFLP Oligonucleotides 340
Appendix 6: AFLP Scoring Parameters 341
Appendix 7: Reprint of Meudt & Clarke (2007) 343
Appendix 8: Reprint of Holland *et al.* (2008) 355
Appendix 9: Reprint of Clarke *et al.* (2006) 375
Appendix 10: Reprint of Erickson *et al.* (2005) 383
Appendix 11: Cloning Application .... CD
Appendix 12: Bottle Gourd NEXUS Files .... CD
Appendix 13: Table of Bottle Gourd Variable Sites .... CD
Appendix 14: Bottle Gourd Seed Length and Width Data .... CD
Appendix 15: Sweet Potato Accessions .... CD
Appendix 16: Sweet Potato AFLP Profiles .... CD
Appendix 17: Sweet Potato NEXUS Files .... CD
Appendix 18: Suppl. Material for Holland *et al.* (2008) .... CD
Appendix 19: Suppl. Material for Clarke *et al.* (2006) .... CD
Appendix 20: Support. Info for Erickson *et al.* (2005) .... CD
Appendix 21: Electronic Copy of Thesis .... CD

ABBREVIATIONS

A	adenine
AD	*anno Domini*
AD	arbitrary degenerate (PCR primer)
aDNA	ancient DNA
AFLP	amplified fragment length polymorphism
AMOVA	Analysis of Molecular Variance
AMS	accelerator mass spectrometry
APS	ammonium persulfate
a.s.l.	above sea level
AWCGS	Allan Wilson Centre Genome Service
ATP	adenosine 5′-triphosphate
BAC	bacterial artificial chromosome
BC	before Christ (prior to AD 1)
bp	base pair
BW	bin width
°C	degrees Celsius
ca	*circa*
C	cytosine
cal	calorie
CAPS	cleaved amplified polymorphic sequence
ccSSR	consensus chloroplast SSR
cDNA	complementary DNA
CE	capillary electrophoresis
cpDNA	chloroplast DNA
CRoPS	Complexity Reduction of Polymorphic Sequences
CTAB	hexadecyltrimethylammonium bromide
cv.	cultivar
d	day
DArT	Diversity Arrays Technology
DMF	*N,N*′-dimethyl-formamide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide 5′-triphosphate
DTE	dithioerythritol
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid · disodium salt
ERMA	Environmental Risk Management Authority
EST	expressed sequence tag
Exo I	exonuclease I
g	gram
$\times g$	times gravity
G	guanine
GMO	genetically modified organism
GTR	GENOTYPER REARRANGER
h	hour
HLA	human leukocyte antigen
Hz	Hertz
indel	insertion–deletion
I-PCR	inverse PCR
IPTG	isopropyl-β-D-thiogalactopyranoside
ISSR	inter-SSR
kb	kilobase pairs (10^3 bp)
L	litre
LB	lysogeny broth/Luria–Bertani
LGDP	local and global detection percentages (GeneMarker®)
LMS	ligation-mediated suppression (PCR)
LSC	long single-copy (cpDNA)
L:W	length:width ratio
m	metre
M	Molar (mol L^{-1})
MAS	marker-assisted selection
Mbp	megabase pairs (10^6 bp)
MDS	multidimensional scaling
MFL	minimum fragment length
min	minute
MLF	multilocus fingerprinting

mtDNA	mitochondrial DNA
nDNA	nuclear DNA
NJ	neighbour-joining
nrITS	nuclear ribosomal internal transcribed spacer
NRY	nonrecombining portion of the Y chromosome
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCA	principal component analysis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PHT	peak height threshold
PNG	Papua New Guinea
POc	proto-Oceanic language
PVP	polyvinyl-pyrrolidone
QTL	quantitative trait locus
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
rfu	relative fluorescent unit
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
s	standard deviation (sample)
SAP	shrimp alkaline phosphatase
SEC	South Equatorial Current
SCAR	sequence-characterised amplified region
SNP	single nucleotide polymorphism
SP	specific (PCR primer)
sp.	species (singular)
SPF	stutter peak filter (GeneMarker®)
spp.	species (plural)
SSC	short single-copy (cpDNA)
SSCP	single-stranded conformation polymorphism
ssp.	subspecies
SSR	simple sequence repeat (microsatellite)
STR	short tandem repeat
syn.	synonym
T	thymine
T_m	melting temperature
TAE	Tris acetate EDTA
TAIL	thermal asymmetric interlaced (PCR)
TBE	Tris borate EDTA
T-DNA	transfer DNA
TEMED	*NNN'N'*-tetramethylethylenediamine
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
U	unit
UBC	University of British Columbia
UPGMA	Unweighted Pair Group Method with Arithmetic mean
V	Volt
v/v	volume per volume
W	Watt
WGA	whole genome amplification
w/v	weight per volume
w/w	weight per weight
$\bar{x}$	mean (sample)
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
yr BP	calendar years before present (non-radiocarbon date, or calibrated ^{14}C radiocarbon date; by convention before AD 1950)
^{14}C yr BP	years before present (uncalibrated ^{14}C radiocarbon date; by convention before AD 1950)

LIST OF FIGURES

Fig. 1.1 Regions of Oceania 5
Fig. 1.2 The Austronesian Language Family 8
Fig. 1.3 The Lapita Expansion into Oceania 9
Fig. 2.1 Traditional Process of SCAR Marker Development 39
Fig. 2.2 Overview of ISSR Fingerprinting 43
Fig. 2.3 Obtaining SCAR Flanking Regions 46
Fig. 2.4 Overview of TAIL PCR Technique 50
Fig. 2.5 Theoretical Results of TAIL PCR Electrophoresis 51
Fig. 2.6 Overview of ISSR-Derived SCAR Marker Development in Bottle Gourd 52
Fig. 2.7 Cucumber Chloroplast Genome Showing Gourd Marker Positions 86
Fig. 2.8 Overview of the AFLP Fingerprinting Technique 89
Fig. 2.9 AFLP Poolplexing and Capillary Electrophoresis 91
Fig. 2.10 Typical Electronic Profiles for AFLP 92
Fig. 2.11 Effects of Changing Bin Width and Peak Height Threshold 97
Fig. 2.12 Flow Chart of Scoring Parameter Optimisation Process 102
Fig. 2.13 Phylogenetic Resolution and Error Rates versus Bin Width 114
Fig. 2.14 Size Difference Between Identical Alleles of Replicate Pairs 115
Fig. 2.15 Consensus Networks of Different AFLP Parameter Settings 119
Fig. 2.16 Consensus Networks of GeneMapper® and GeneMarker® Trees 121
Fig. 2.17 Consensus Networks of Default and Optimised Trees 122
Fig. 3.1 The Bottle Gourd (*Lagenaria siceraria*) 130
Fig. 3.2 Phytoliths from Bottle Gourd and *Cucurbita* 140
Fig. 3.3 Prehistoric Bottle Gourd Floats, Huaca Prieta, Peru 141
Fig. 3.4 Prehistoric Bottle Gourd in Asia, the New World and Oceania 146
Fig. 3.5 Distributions and Dispersal Routes of the Oceanic Bottle Gourd 156
Fig. 3.6 Bottle Gourd Exocarp from Kohika, New Zealand 160
Fig. 3.7 Seeds of Bottle Gourd and Outgroup Accessions 173
Fig. 3.8 Genotype Frequencies for 36 Accessions of Bottle Gourd 183
Fig. 3.9 Spectronet Network of Bottle Gourd Nuclear Data 186
Fig. 3.10 Prehistoric Bottle Gourd Seeds from Five Sites in Peru 197
Fig. 3.11 Hypothesised Global Dispersal and Domestication of Bottle Gourd 207
Fig. 4.1 The Introduction of Sweet Potato into Oceania: The Tripartite Hypothesis 213
Fig. 4.2 The Banks & Solander “Kumala” Sweet Potato (1769) 217
Fig. 4.3 Overview of AFLP Fingerprinting in Sweet Potato 221
Fig. 4.4 Workflow for Generation of AFLP Data 224
Fig. 4.5 Taxa Ranked by Percentage of “1” Character States 241

Fig. 4.6 Sweet Potato Neighbour-Joining Tree..242
Fig. 4.7 Path Lengths Between AFLP Replicates...245
Fig. 4.8 Leaf Shapes of Eastern Polynesian Sweet Potato...247

LIST OF TABLES

Table 2.1 ISSR Primers Used from UBC Primer Set #9 ... 56
Table 2.2 Polymorphic ISSR Bands Isolated for Bottle Gourd ... 78
Table 2.3 TAIL PCR and SCAR Marker Development ... 80
Table 2.4 Polymorphisms Present in Developed SCAR Markers ... 81
Table 2.5 Rates of Polymorphism in TAILed vs. non-TAILed SCAR Markers ... 82
Table 2.6 30-Sample Dataset for Optimising AFLP Scoring Parameters ... 99
Table 2.7 Resolution Scores and Numbers of Parsimony Informative Sites ... 109
Table 2.8 Normalised Resolution Scores ... 110
Table 2.9 Replicate Error Rates for Representative Datasets ... 112
Table 2.10 Binning Error Simulation for Identical Fragments ... 116
Table 3.1 Defining Characters of the *Lagenaria siceraria* subspecies ... 132
Table 3.2 Bottle Gourd Accession Details ... 169
Table 3.3 Chloroplast and Nuclear Bottle Gourd PCR Markers ... 177
Table 3.4 Ancient New World Bottle Gourds from Erickson *et al.* (2005) ... 190
Table 3.5 Seed L:W Ratios in New Zealand, the New World, Africa and Asia ... 198
Table 3.6 Differences in Mean Seed Length:Width Ratios Between Regions ... 198
Table 4.1 Distribution of *Kumara* and Cognate Terms in South America and Polynesia ... 215
Table 4.2 Results of PCR Primer Combination Screening for AFLP ... 232
Table 4.3 Eco+3 and Mse+3 Selective Amplification Primer Combinations ... 233
Table 4.4 Absorbance and Emission Maxima for AFLP Fluorophores ... 233
Table 5.1 Priority Bottle Gourd Samples for aDNA Analysis ... 262
Table 5.2 Indigenous Words for Bottle Gourd in the New World and Polynesia ... 266
Table 5.3 Germplasm Collections of Bottle Gourd Landraces ... 268
Table A1.1 Chloroplast Markers Screened for Use in Bottle Gourd ... 312
Table A1.2 Nuclear Markers Screened for Use in Bottle Gourd ... 313
Table A2.1 pGEM®-T Easy PCR Primers ... 314
Table A2.2 TAIL PCR Primers ... 314
Table A2.3 Chloroplast and Nuclear PCR and Sequencing Primers ... 316
Table A4.1 Yen Sweet Potato Collection (June 2004) ... 320
Table A5.1 Oligonucleotides Used in AFLP ... 340
Table A6.1 AFLP Scoring Parameters in GeneMapper® v. 3.7 ... 341

CHAPTER ONE

GENERAL INTRODUCTION

Māori Women outside a Meeting House
The woman standing is carrying a kete of kumara. New Zealand, ca *1880s*
Coloured wood engraving by Thomas Selby Cousins
Ref. No. B-059-017
Reproduced with permission of the Alexander Turnbull Library, Wellington, New Zealand

1.1 CHAPTER OVERVIEW — RESEARCH CONTEXT

The aim of this project is to use a molecular (DNA) approach to reconstruct the origins and dispersal in Oceania of the sweet potato (*Ipomoea batatas*) and bottle gourd (*Lagenaria siceraria*), and to use the inferences of these crops' patterns of mobility (phylogeographic structure) as a proxy for human mobility. Specifically, the interest is in using both these crop species to test for prehistoric Polynesian contact with South America, because it has been proposed that lineages of both sweet potato and bottle gourd found in Polynesia are the result of a prehistoric human-mediated introduction from South America. Although this thesis is focussed on sweet potato and bottle gourd, it is necessary to give a brief introduction to prehistoric human settlement of the Pacific, and the dispersal of commensal plants and animals.

The chapter begins with an overview of Pacific settlement, from the settlement of Near Oceania 60–40,000 yr BP, to the Austronesian expansion and the settlement of Remote Oceania by Lapita people from ~3,500 yr BP. Finally, the settlement of Eastern Polynesia is covered. The contribution of molecular genetic studies to understanding the origin of Polynesians is outlined. The spread of agricultural species into Remote Oceania is discussed, focussing on the contributions both from Southeast Asia (associated with the Austronesian expansion) and those from Island Melanesia. Next a general commensal model is described, the considerations when applying the model are outlined, and examples of the application of a commensal model to various animal and plant species are provided. The chapter then addresses to the specific question of contact between Eastern Polynesia and the New World, and reviews the biological and cultural evidence for contact, as well as the ability of Polynesian voyagers to reach the Americas (and return). The chapter concludes with the aims and hypotheses of the research.

1.1.1 A NOTE ON DATES

Dates within the last 2,000–3,000 years are generally given as before Christ (BC) or *anno Domini* (AD). Dates older than ~3,000 years are given as either calendar years before present (yr BP) or uncalibrated radiocarbon years before present (^{14}C yr BP). In both cases "present" is, by convention, AD 1950.

For readability, error limits on both calibrated and uncalibrated ^{14}C dates are not stated unless they are particularly large or significant in the context discussed here e.g., for a date reported as 7,000 ± 100 yr BP, the error is not shown in the thesis.

The terms "prehistoric", "prehistory", etc. are frequently used to define the period before European contact with the New World and Pacific, and the terms "historic", etc. for the period that began with European contact. Although "prehistoric" can have negative connotations in some contexts (like "primitive" character states in systematics), its usual definition in Pacific anthropology is well constrained, and it is used simply to define the period before European contact.

1.2 HUMAN ORIGINS, MIGRATION AND MOBILITY IN OCEANIA

1.2.1 INTRODUCTION

The Pacific Ocean is truly immense. It covers approximately 180 million square kilometres — one third of the Earth's surface. There are within it more than 25,000 islands, whose total land area contributes only ~1% of the Pacific's total surface area (and most of this comprises New Guinea, Borneo, Japan, the Philippines and New Zealand). Nevertheless, some of the most remote islands of the Pacific were settled prehistorically. This was accomplished by skilled voyagers for whom the vast expanse of the Pacific was ultimately a highway, not a barrier.

1.2.2 THE SETTLEMENT OF OCEANIA

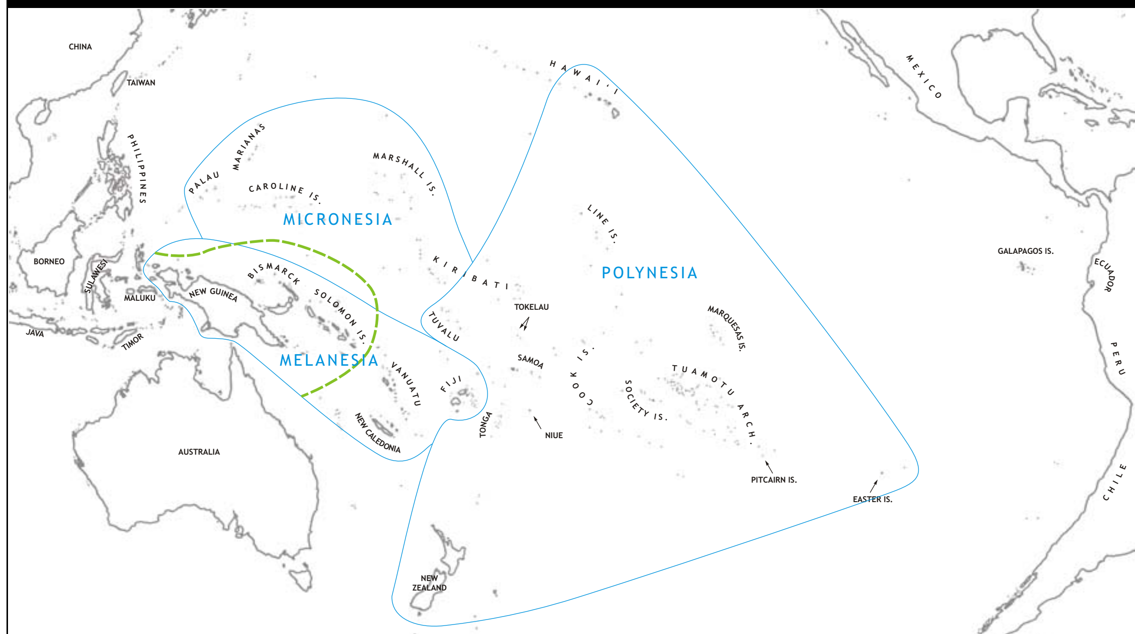
1.2.2.1 The First Wave (from 60-40,000 yr BP)

Anatomically modern humans first moved into the Pacific from Asia (Fig. 1.1) about 40,000 yr BP and possibly as early as 60,000 yr BP (Hurles *et al.*, 2003a; Howe, 2006, p. 17). At this time — during the Pleistocene — sea levels were lower, and the western Pacific was dominated by two massive landmasses: Sunda and Sahul. Sunda comprised mainland Southeast Asia and what are now the islands of Sumatra, Java and Borneo. Sahul comprised the present-day islands of New Guinea, Australia and Tasmania. Sunda and Sahul were separated by Wallacea, an archipelago including the islands of Sulawesi and most of the Philippine Island group except Palawan. Wallacea, and especially the boundary between it and Sunda (termed "Wallace's Line") represented a major barrier to dispersal, especially for mammals (Irwin, 2006, p. 59).

By 40,000 yr BP, human voyagers, probably using bamboo rafts (Irwin, 2006, p. 62), crossed Wallace's Line, spreading into New Guinea and Australia (Hurles *et al.*, 2003a), and reaching Tasmania by 34,000 yr BP (Howe, 2006, p. 19). These early Pacific settlers also began moving farther east, developing deep-sea sailing technologies

which allowed them to progressively colonise the islands of the Bismarck Archipelago and the western Solomon Islands from 40,000 yr BP (Howe, 2006, p. 19). These easternmost islands (approx. 160°E) represented the limits of what these early voyagers' sailing technologies would allow (Howe, 2006, p. 19). The threshold to the distant islands still farther east — in what is now termed Remote Oceania — would not be crossed by humans for another 35,000 years. Even though there was no dispersal beyond Near Oceania during this period, significant cultural changes (beyond those involving voyaging) continued (Kirch, 2000, p. 68–78).

Fig. 1.1
Regions of Oceania



Oceania is typically defined as the region of the Pacific excluding Island Southeast Asia and Australia (Kirch, 2000, p. 5). Oceania itself is divided into the regions of Melanesia, Micronesia and Polynesia. Although these divisions were originally made along perceived cultural boundaries, it is now clear that neither 'Melanesians' nor 'Micronesians' form discrete historical/cultural units (Green, 1991a). Only Polynesia has remained a term with historical significance (Kirch, 2000, p. 5). Although all three terms are used in this thesis to describe geographical regions, it should be clear that in the case of Melanesia and Micronesia, no cultural unity is implied. Two other terms are used in this thesis to divide Oceania: Near Oceania and Remote Oceania (Green, 1991a; see also Kirch & Kahn, 2007). Near Oceania (west of the green line above) includes New Guinea, the Bismarck Archipelago and the Solomon Islands as far east as San Cristobal. Remote Oceania (east of the green line above) includes all islands to the north, east and southeast. The occupation of Near Oceania began in the late Pleistocene (at least 40,000 yr BP), and is currently occupied by both Austronesian- and non-Austronesian-speaking "Papuan" peoples. Remote Oceania has only been occupied from 1500 BC, and is occupied exclusively by Austronesian speakers (Green, 1991a). Thus the boundary between Near and Remote Oceania describes the geographical, chronological and, to some extent, the cultural boundary between two major phases of Oceanic settlement.

Map modified from Kirch (2000, p. 6)

Initially, the settlers of Near Oceania were hunter–gatherers, as were contemporaneous humans in other parts of the world. From 20,000 yr BP however, there is evidence of increasing management of animal resources, in what some would consider a proto-agriculture phase. The islands farther east in the Bismarck Archipelago were increasingly depauperate in terrestrial animal species, and there is limited evidence that settlers in this region compensated for this by translocating several marsupial species (Irwin, 2006, p. 61). For example, the Northern common cuscus (*Phalanger orientalis*) was introduced from New Guinea to New Ireland approximately 20,000 yr BP, and the common bandicoot (*Echimypera kalubu*) from New Guinea or New Britain to Manus (Irwin, 2006, p. 61). In the same way, Yen (1995, 1996) has argued for a proto-horticultural phase of arboricultural beginnings, starting in the late Pleistocene or early Holocene, with the best example of this being *Canarium* (see also Green, 2000a).

Agriculture "proper" was developed in New Guinea at the end of the last ice age (approx. 10,000 yr BP). The domestication of plants and animals in New Guinea appears to have occurred independently of other regions, but, temporally, does coincide with the invention of agriculture in up to nine other regions worldwide (Denham *et al.*, 2003; Diamond & Bellwood, 2003). In New Guinea, as elsewhere, the invention of agriculture dramatically increased the efficiency of food production, and fundamentally altered the way in which human society was structured and developed.

1.2.2.2 The Austronesian Expansion and the Settlement of Remote Oceania

After ~35,000 years of exploration stasis at the limits of Near Oceania, the human settlement of Remote Oceania occurred comparatively explosively, beginning approximately 2000–1500 BC in the case of Western Micronesia, and approximately 1500 BC in Island Melanesia and Western Polynesia (Irwin, 2006, p. 64). Many aspects concerning the settlement of Remote Oceania are debated, including:

- the initial events that culminated in people moving from Near into Remote Oceania,
- the identity and origin of the people(s) involved,
- the extent to which voyaging and settlement were deliberate and planned,

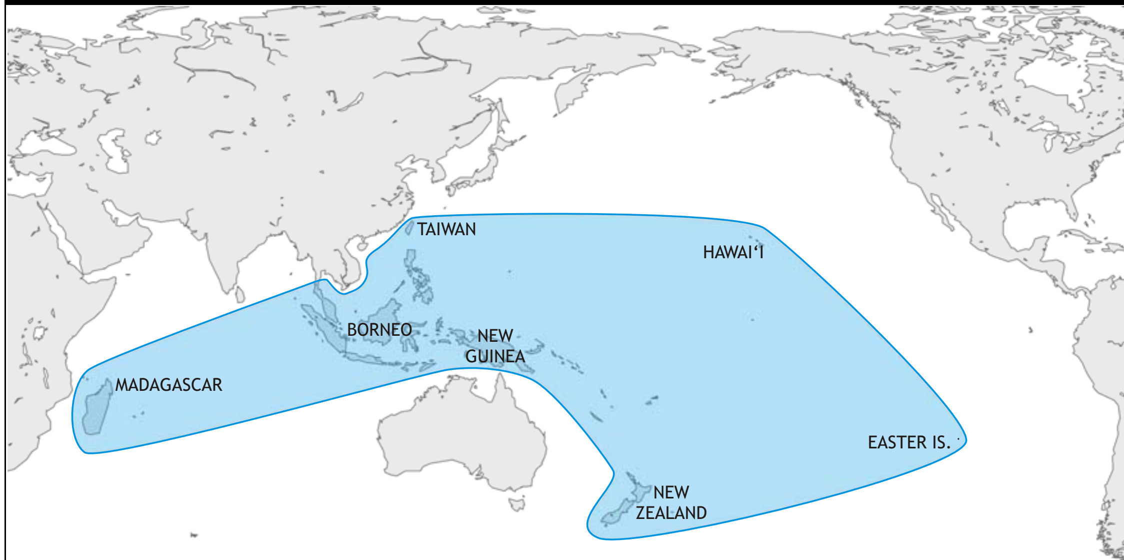
- which species of plants and animals were translocated by voyagers,
- the chronology of settlement, especially in Eastern Polynesia,
- the extent of "back-voyaging" and post-settlement interaction, and
- how long-distance voyaging was achieved.

It is outside the scope of this thesis to cover these issues in any detail (but see Kirch & Kahn, 2007; Matisoo-Smith, 2007a); instead, just a brief overview will be given, focussing on the settlement of Island Melanesia and Polynesia because these are the regions relevant to the origins and dispersal of the sweet potato and bottle gourd in Oceania.

Humans first began moving into Remote Oceania ~3,150 yr BP (Irwin, 2006, p. 64), an event which is generally associated with the Austronesian expansion out of Southeast Asia 6–5,000 yr BP (Hurles *et al.*, 2003a; Fig. 1.2), and particularly with the appearance from ~3,500 yr BP of the Lapita cultural complex in western Island Melanesia (Hurles *et al.*, 2003a; Fig. 1.2), and with southern Island Melanesia at 3,150 yr BP (Green *et al.*, 2008). It should be noted, however, that the extent to which the Austronesian expansion and the appearance of Lapita are connected is still debated (Hurles *et al.*, 2003a; see Kirch & Kahn, 2007; Matisoo-Smith, 2007a). About 3,000 yr BP Lapita peoples and their descendents moved eastwards from the archipelagos of Island Melanesia into Western Polynesia (Samoa and Tonga). Migration farther east then appeared to stop for probably more than a thousand years in what is called by some the "long pause" (Irwin, 2006, p. 76).

Although it is clear that Lapita descendents then moved from Western Polynesia farther east (Kirch & Kahn, 2007), the chronology of settlement in Eastern Polynesia remains controversial (Fig. 1.3). This is largely due to debate about the reliability of previous radiocarbon dates obtained from Eastern Polynesia and, specifically, the degree to which inadequate laboratory pre-treatment and old carbon (carbon from marine organisms or old wood) has made dates for colonisation appear significantly older than they should (Anderson & Sinoto, 2002).

Fig. 1.2
The Austronesian Language Family

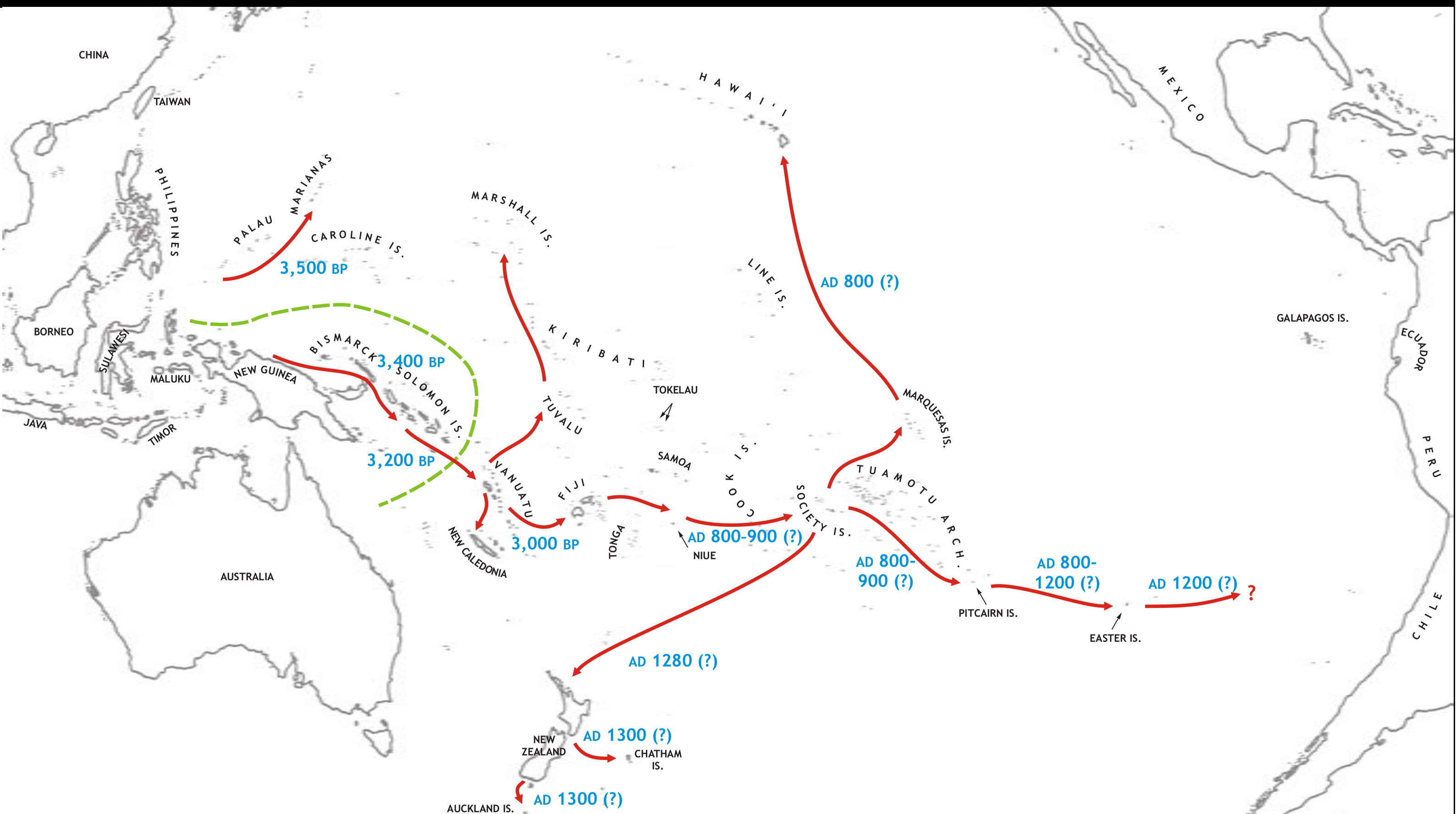


The Austronesian language family is widely distributed — from Island Southeast Asia to parts of New Guinea, Remote Oceania and Madagascar. It is thought that these languages have their origins in Taiwan, where the Austronesian languages are most diverse (Gray & Jordan, 2000; Irwin, 2006, pp. 64–65). Based on linguistic analyses, Austronesian languages, and probably sea-faring voyagers, spread south from Taiwan into the Philippines and Indonesia, and from there into coastal New Guinea and ultimately Remote Oceania.

Modified from Irwin (2006, p. 64)

New radiocarbon dates obtained from key sites using rigorous "chronometric hygiene" (Spriggs & Anderson, 1993; see also Rieth & Hunt, 2008) now place colonisation of Eastern Polynesia at AD 900 or later, favouring the so-called "short chronology" (Anderson & Sinoto, 2002; see also Conte & Anderson, 2003). Two of the sites are in the Society Islands: Motu Paeao cemetery (Maupiti), which has been dated to AD 1200–1400 (Anderson *et al.*, 1999), and Vaito'otia-Fa'ahia (Huahine), which has been dated to AD 1050–1450 (Anderson & Sinoto, 2002). Another two important sites are in the Marquesas Islands: Hane (Ua Huka), which has been dated to AD 1000 at the earliest (Anderson & Sinoto, 2002), and Ha'atuatua (Nuku Hiva), which has been dated to AD 900–1400 (Rolett & Conte, 1995).

Fig. 1.3
The Lapita Expansion into Oceania



Lapita peoples began moving into Remote Oceania and West Polynesia from about 3,500 yr BP, and, favouring the "short chronology", into Eastern Polynesia about AD 900–1200. Although the dates presented here favour the "short chronology" for Eastern Polynesia settlement, it is acknowledged that the chronology of settlement in this region remains highly controversial (reviewed in Kirch & Kahn, 2007).

Compiled from Rolett & Conte (1995), Anderson et al. *(1999), Anderson & Sinoto (2002), Green & Weisler (2002), Hunt & Lipo (2006), Irwin (2006, p. 67), Kirch & Kahn (2007), Wilmshurst* et al. *(2008)*

For Hawai‘i, settlement no later than about AD 800 has been suggested (Kirch & Kahn, 2007 and references therein), and for the chain of islands from the Australs to Mangareva, Pitcairn–Henderson and Easter Island (Rapa Nui), settlement no earlier than about AD 800–900 (Green & Weisler, 2002; Kirch & Kahn, 2007 and references therein). In the case of Easter Island, settlement as late as AD 1200 has been suggested (Hunt & Lipo, 2006), although many authors favour the AD 900–1000 range (see, for example Shepardson *et al.*, 2008; Weisler & Green, 2008).

The most recent radiocarbon dates for New Zealand, based on Pacific rat bones (*R. exulans*) and distinctive rat-gnawed seeds, place settlement at *ca* AD 1280 (Wilmshurst *et al.*, 2008). This date is consistent with the earliest-dated archaeological sites, human-induced faunal extinctions and deforestation (Wilmshurst *et al.*, 2008). Wilmshurst *et al.* (2008) conclude that the earlier radiocarbon date of 2,000 yr BP obtained from *R. exulans* in New Zealand (Holdaway, 1996) is in error and was most likely caused by laboratory pre-treatment procedures at the time (Wilmshurst *et al.*, 2008). If rats were introduced to New Zealand as a food source prior to permanent settlement then it might have been only a few decades earlier, at most.

From about AD 1500, and after the settlement of the farthest reaches of the Pacific, long-distance voyaging began to decline (Irwin, 2006, p. 91), and some small islands such as Palmerston, Norfolk, Raoul (the Kermadecs group), Christmas and Pitcairn Islands were abandoned. These so-called “mystery islands” were unoccupied at the time of European discovery (Anderson, 2001; Irwin, 2006, p. 91).

1.2.2.3 Human Genetic Variation in Oceania

Genetic analyses of human populations have significantly improved our understanding of the origins and dispersal of Polynesians (e.g., Merriwether *et al.*, 2005; Trejaut *et al.*, 2005; Kayser *et al.*, 2006; Pierson *et al.*, 2006; Kayser *et al.*, 2008). The majority of these studies have focussed on the maternally inherited mitochondrial DNA (mtDNA), and the nonrecombining section of the paternally inherited Y chromosome (NRY) (Hurles *et al.*, 2003a). Polynesian and other Oceanic populations have now been the subject of a large number of mtDNA-based studies. Although the detailed findings of

these will not be reviewed here (but see Pierson (2007) and Matisoo-Smith (2007a) for recent syntheses), they all show the general pattern of a recent mainland east Asian or Southeast Asian origin for the dominant Polynesian mtDNA haplotypes. Specifically, the characteristic "Polynesian motif" (part of the B4a1a1 haplogroup) is found amongst speakers of the central/eastern Malayo-Polynesian group of Austronesian languages spoken in Wallacea and the Pacific (Hurles *et al.*, 2003a), and the most closely-related haplotypes are found across the entire range of Austronesian speakers (Hurles *et al.*, 2003a). Mitochondrial DNA variation has also been used to estimate the size of a founding population in Polynesia, with Penny & Meyer (2006) estimating that 70–190 women were amongst the founding population of New Zealand Māori (this is revised slightly from the 50–100 women estimated by Murray-McIntosh *et al.* (1998)). This is consistent with a planned voyage of several waka, and is also in agreement with Māori oral history (Murray-McIntosh *et al.*, 1998).

Subsequent analyses dealing with question of Polynesian origins have combined mtDNA and NRY markers, allowing determination of sex-specific dispersal patterns. Kayser *et al.* (2006) analysed over 1300 Oceanic individuals, including 400 Polynesians, revealing distinct origins for male and female Polynesian ancestors. Whereas Polynesian mtDNA was almost entirely of Asian origin (Asian = 93.8%; Melanesian = 6%), Y chromosomes were largely derived from populations in Melanesia (Asian = 28.3%; Melanesian = 65.8%), consistent with matrilocal residence in proto-Polynesian societies.

Polynesian origins have been further elucidated through the analysis of nuclear DNA. In order to determine the origin of Polynesian autosomes, Kayser *et al.* (2008) screened 377 autosomal short tandem repeat (STR) loci in 47 Polynesians, 44 Han Chinese and 24 Highland New Guineans. Their data indicate that on average about 79% of the Polynesian autosomal gene pool is of East Asian origin, and 21% is of Melanesian origin, which, along with previous mtDNA and NRY studies, is consistent with a dual origin of Polynesian populations.

1.2.2.4 The Lapita Cultural Complex and Commensal Species

The Lapita culture is characterised by several important components, including — perhaps most famously — the distinctive dentate-stamped pottery recovered from Lapita sites in Island Melanesia and Western Polynesia. But more relevant to the research here are the advanced sailing technologies, well-developed agricultural systems, and the deliberate translocation of crop and livestock species (or the so-called "transported landscapes" of the Lapita peoples (Kirch, 2000, p. 109)).

The Lapita settlers and their Polynesian descendents took with them a large number of crop and livestock species into Remote Oceania, including rats, pigs, dogs, chickens, taro, breadfruit and bananas (Matisoo-Smith, 1994; Irwin, 2006, pp. 74–75). The majority of the transported crop species are of Southeast Asian origin, e.g., the greater yam (*Dioscorea alata*), paper mulberry (*Broussonetia papyrifera*) and coconut (*Cocos nucifera*) (Whistler, 1991; Bellwood, 2005). The introduction of these species into Remote Oceania is generally associated with the Austronesian expansion (Kirch, 1997, Table 7.2; 2000, pp. 109–110; see also Matthews, 2006, p. 95). Other crop species transported by Lapita peoples were domesticated in Near Oceania by Papuan-speaking populations, e.g., sugarcane (*Saccharum officinarum*), ti (*Cordyline fruticosa*), breadfruit (*Artocarpus altilis*), the *Canarium* almond, cutnut (*Barringtonia* sp.), and the aroids taro (*Colocasia esculenta*), elephant ear (*Alocasia macrorrhizos*) and swamp taro (*Cyrtosperma chamissonis*) (Kirch, 1997, Table 7.2; 2000, pp. 78–79; Allaby, 2007). In the case of breadfruit, the New Guinea *A. altilis* was transported to Micronesia where it was hybridised with the native *A. mariannensis*. The resulting *A. altilis* × *mariannensis* hybrid was transported throughout the wetter, southern regions of Micronesia (Zerega *et al.*, 2004, 2005; Matthews, 2006, p. 94).

As settlers moved from Near into Remote Oceania the number of cultivated species fell from hundreds to less than one hundred. Approximately 50–75 introduced crop plants were cultivated in Tonga and Samoa, about 30 in the Cook, Society and Marquesas Islands, and then about 30 species in Hawai'i, six to eight in Easter Island and six in New Zealand (Leach, 1984; Whistler, 1991; Bevacqua, 1994; Irwin, 2006, pp. 74–75; Matthews, 2006, p. 96). The successful introductions to New Zealand are generally agreed to be: taro (*Colocasia esculenta*), yam (*Dioscorea* sp.), paper mulberry

(*Broussonetia papyrifera*), ti (*Cordyline fruticosa*), bottle gourd (*Lagenaria siceraria*) and sweet potato (*Ipomoea batatas*) (Leach, 1984, p. 53; Horrocks & Lawlor, 2006; Horrocks *et al.*, 2008). New Zealand would have been too cold for the successful introduction of other Polynesian staples (Whistler, 1991). In terms of animal species, both the dog and Pacific rat were successfully introduced to New Zealand, but there is no evidence of pigs nor chickens (Matisoo-Smith, 2007a).

1.3 COMMENSAL ANIMALS AND PLANTS IN OCEANIA

1.3.1 THE COMMENSAL MODEL

In its broader scope, this project is an application of a commensal model like those increasingly being applied to infer human mobility in Oceania through molecular studies of commensal plants, animals and pathogens. In the sense used here, commensal models include molecular data that are analysed using population genetic, phylogenetic and phylogeographic methods (Storey *et al.*, *in prep.*). These data are complemented by (and necessarily include) relevant biological and anthropological (including linguistic and archaeological) information. The result is inferred patterns of plant and animal dispersal, and indirectly of human mobility. Whereas human genetic research reveals patterns of migration and settlement, commensal models potentially go further by revealing patterns of human mobility that may not necessarily be associated with settlement.

The following is a list of some of the considerations and limitations when employing commensal models. It is based, in part, on Storey *et al.* (*in prep.*). Some of these considerations will be used to develop hypotheses for testing, others are important during the course of the project, and others are important in the interpretation of molecular data. They include:

- the origin of the species, including where, when, and how many times it was domesticated.
- the modern and prehistoric distributions of the wild and domesticated lineages of the species. Knowledge of the prehistoric distribution will also allow determination of whether there are sufficient samples available for aDNA analysis (e.g., Matisoo-Smith, 2002).
- whether there are biotic and/or abiotic factors which may limit/have limited the distribution of the commensal species (e.g., competition, climate, etc.), and whether these vary for different islands.
- the importance of the commensal species to humans and how this might change depending on which islands are settled and by which cultures.

- whether single or multiple introductions occurred and, in the case of multiple introductions, whether the same source populations were involved. The extent of ongoing trade and exchange with source populations can also be important.
- the species' dispersal mechanisms (across land or sea; natural and/or human-mediated (intentional or "stowaway")), and how this might vary between wild and domesticated lineages.
- how well extant populations are likely to represent prehistoric distribution patterns and genetic diversity (e.g., the effects of population extinctions and historical population bottlenecks).
- the species' ability to survive without human intervention (e.g., to form feral populations).
- whether ancestral, wild and/or European-era lineages are present with which the commensal taxa may interbreed.
- which molecular marker systems are most appropriate. This will depend on the level of genetic variability, the genetic resolution required to test the hypotheses, whether processes such as hybridisation and sex-specific dispersal patters are important, whether sexual and/or asexual lineages are present (Zerega *et al.*, 2005; Hinkle, 2007), and ploidy levels (see also Chapter Two). The types of available tissue also need to be considered (e.g., ancient/archaeological, museum/ethnographic, modern).

1.3.2 APPLICATION OF COMMENSAL MODELS

Commensal models have been successfully applied to a number of plant and animal species in the Pacific, and the integration of these studies, along with direct human genetic research, has significantly enriched our understanding of prehistoric human mobility in the Pacific. Commensal animal species studied include Pacific rat (*Rattus exulans*) (Matisoo-Smith *et al.*, 1998; Matisoo-Smith & Robins, 2004), pig (*Sus scrofa*) (Allen *et al.*, 2001; Larson *et al.*, 2007), chicken (*Gallus gallus*) (Storey *et al.*, 2007; see also Section 1.4.1.2), the lizard *Lipunia noctua* (Austin, 1999), and the land snail *Partula hyalina* (Lee *et al.*, 2007). Commensal plant species studied include breadfruit

(*Artocarpus* spp.) (Zerega *et al.*, 2004, 2005), ti (*Cordyline fruticosa*) (Hinkle, 2007) and the bananas (*Musa* spp.) (Kennedy, 2008).

An important application of a commensal model has been to the Pacific rat (*Rattus exulans*) (Matisoo-Smith, 1994, 1996, 2007b). The Pacific rat possesses many attributes that make it a good commensal model: it is widely dispersed throughout Near and Remote Oceania, it is well-represented by modern and archaeological material, it is unable to breed with the European-introduced rat species, its introduction to Oceanic islands was human mediated (it cannot disperse naturally), and it has high mitochondrial (mtDNA) D-loop variability (Matisoo-Smith, 1994; Matisoo-Smith *et al.*, 1998; Matisoo-Smith, 2007b). Analyses of patterns of mtDNA diversity in populations of *R. exulans* have allowed different models for Pacific migration and settlement to be tested. For example, based on analyses of mainly Near Oceanic populations of *R. exulans*, Matisoo-Smith & Robins (2004) found that, of the four sets of models proposed by Green (2003) to explain the origins of Lapita in Near Oceania, there was strongest support for Green's Voyaging Corridor Triple I (VC Triple-I) models (Green, 1991b, 2003). Genetic analyses of *R. exulans* have also been informative in Remote Oceania, where genetic variation in the archipelagos of New Zealand and Hawai'i is more consistent with post-settlement contact with central east Polynesia (the Southern Cook Islands and the Societies, but not the Marquesas), rather than isolation of New Zealand and Hawai'i after colonisation (Matisoo-Smith *et al.*, 1998). This is in contrast to the more remote Easter Island, where limited mtDNA diversity is consistent with a single or at least a limited number of introductions of the Pacific rat, suggesting relative isolation following initial colonisation (Barnes *et al.*, 2006). Similar patterns are observed in the Kermadec and Chatham Islands, where high and low levels of *R. exulans* mtDNA variation are consistent with the hypothesised roles of these islands as "stepping stone" and "end-of-line" islands respectively (Matisoo-Smith *et al.*, 1999).

1.4 SAILING FROM POLYNESIA TO THE NEW WORLD IN PREHISTORY

1.4.1 EVIDENCE OF POLYNESIA–NEW WORLD CONTACT

1.4.1.1 Crop Species

There exist several lines of evidence for prehistoric Polynesian contact with the Americas/New World. Firstly is the sweet potato itself. The sweet potato evidence is covered in more detail in Chapter Four, but briefly: the sweet potato is a South American domesticate and was grown there from at least 9,000 yr BP (Engel, 1970), prehistoric tubers (the earliest of which date to AD 988–1155) have been recovered from Mangaia, Cook Islands (Hather & Kirch, 1991), the sweet potato is almost certainly not capable of natural dispersal nor surviving outside of cultivation, and *kumara*, the South American Quechua word for sweet potato, was used in Polynesia (Yen, 1974). Other crop species that may have been transferred by humans from the New World to Polynesia include bottle gourd (see Chapter Three), the Polynesian tomato and soapberry, whereas the coconut may have been introduced into the Americas from Polynesia (see Chapter Five).

The Polynesian tomato (*Solanum repandum*) was recorded from the Marquesas Islands to Fiji, and always in association with human activities (Whistler, 1991). It has been suggested that *S. repandum* is conspecific with the South American *S. sessiliflorum* (see Whistler, 1991), and although *S. repandum* may be an early Spanish introduction into the Marquesas, Whistler (1991) suggests its wide distribution in Polynesia is more consistent with a Polynesian introduction. Whistler (1991) goes on to suggest that the Marquesan name *koko‘u* may be a cognate of the South American *cocona*. Alternatively, *S. repandum* may be conspecific with (or closely related to) the Southeast Asian *S. lasiocarpum* (Heiser, 1987), a hypothesis offered some support by chloroplast DNA (cpDNA) analyses of section *Lasiocarpa* (to which all three species belong) (Bohs, 2004). Although *S. repandum* is now rare or extinct over much of its

former range (Whistler, 1991), DNA analysis of herbarium specimens using higher resolution markers could be used to test the hypothesis of a South American origin, as well as clarify its relationship to other taxa in section *Lasiocarpa*.

There is some evidence that the soapberry (*Sapindus saponaria*) was grown in all or some of: Easter Island, the Marquesas Islands, Mangareva and Pitcairn (Langdon, 1996). This American plant could have been a prehistoric human-mediated introduction to Polynesia, but as Green (2005) states, "only with suitable emendations to the dubious theory" proposed by Langdon (1996).

The coconut (*Cocos nucifera*) may also provide evidence of long-distance Polynesian voyaging, this time as an introduction *from* Polynesia *to* the New World (Green, 2005). Coastal groves of this Indo-Pacific crop species were recorded on Cocos Island (600 km southwest of Costa Rica) at the time of the first European visit in the 1530s (see Ward & Brookfield, 1992 and references therein). Early 16th century Spanish explorers described other populations of coconut on the western coasts of Costa Rica, Panama and Colombia (Ward & Brookfield, 1992; Zizumbo-Villarreal & Quero, 1998). Modelling of coconut dispersal suggests its ability to drift to the west coasts of Panama and remain viable is extremely low, and that a human-mediated introduction is more likely (Ward & Brookfield, 1992; Green, 2005).

1.4.1.2 The Chicken (*Gallus gallus*)

A Polynesian-mediated introduction of the chicken (*Gallus gallus*) to South America is supported by recent archaeological and molecular research (Storey *et al.*, 2007; Storey *et al.*, 2008b). A chicken bone recovered from El Arenal in coastal south-central Chile was dated to AD 1304–1424 (at two sigma), consistent with a pre-Columbian introduction (Storey *et al.*, 2007). This date is also compatible with the period during which Polynesian sailors were undertaking long-distance voyages (see Section 1.2.2.2), including the proposed voyage to South America (Green, 2005). This pre-European date is consistent with the archaeological context as well as early European historical records.

Ancient DNA analysis of the bone showed that it shares a haplotype with ancient chickens from Tonga and American Samoa (Storey *et al.*, 2007). Subsequent molecular analyses (Gongora *et al.*, 2008a) showed this haplotype to be common among modern domestic breeds, which lead the authors to argue that the haplotype of the El Arenal chicken did not necessarily originate in Polynesia. Storey *et al.* (2008b) argue that inferring patterns of prehistoric genetic diversity based on modern samples is confounded by both the widespread movement of commercial chickens, and the fact that modern European chickens as well as Pacific chickens are ultimately derived from Asia and therefore are expected to share lineages. Gongora *et al.* (2008a, 2008b) also questioned the accuracy of the radiocarbon date, and the potential contribution of marine carbon that would push the date into the Columbian era. Two additional chicken bones from El Arenal have now been dated, and these are also securely pre-Columbian (Storey *et al.*, 2008b), consistent with the first date as well as stratigraphic and artefactual evidence (Storey *et al.*, 2007). Furthermore, stable isotope determinations ($\delta\ ^{13}C$, $\delta\ ^{15}N$, $\delta\ ^{34}S$) confirm a terrestrial diet with no detectable marine contribution (Storey *et al.*, 2008b). Further genetic analyses of ancient chicken remains from Oceania and the New World (see Storey *et al.*, 2008a) and possibly *well-provenanced* modern chickens will further elucidate prehistoric patterns of chicken dispersal, especially the relationships between the El Arenal chickens, the Eastern Polynesian chickens, and chickens associated with the Lapita expansion into Remote Oceania (Storey *et al.*, 2008a).

1.4.1.3 Cultural Evidence

In addition to the biological evidence presented above, there is also possible cultural evidence of Polynesia–New World contact, including:

- early historic records of Polynesian-style sailing rafts in northern Peru and Ecuador (Doran, 1971; Green, 2001; *cf.* Langdon, 2001),
- the presence of Polynesian-style sewn-plank canoes and compound fishhooks in the archaeological record of the Channels islands off southern California (Jones & Klar, 2005; but see Anderson, 2006; Arnold, 2007), and
- various artefacts of the Mapuche area of south-central Chile that are similar to those from Polynesia, including obsidian points similar to Easter Island *mata'a*,

polished stone adzes called *toki* that are similar to Polynesian adzes of the same name, and *clava mere okewa* stone hand clubs which bear a resemblance to their wooden, Māori counterparts (Ramírez, 1991).

Although these cultural linkages may provide support for Polynesia–New World contact, their usefulness is limited by difficulties in distinguishing, especially, quantitatively, between independent invention versus various scenarios of cultural diffusion pre and post European contact.

1.4.1.4 Human Genetic Evidence

Despite mounting biological and cultural evidence for at least one, and probably more, episodes of Polynesian contact with the New World, there remains a lack of conclusive human genetic evidence. In a study of 1,178 Polynesians, two individuals were found to have "native South American" mtDNA haplotypes (Sykes *et al.*, 1995), which the authors suggested may be evidence for prehistoric human contact between Polynesia and South America. Further analysis however, revealed that these haplotypes are also present in continental Asian populations and other basal positions in the human mtDNA phylogeny (Bonatto *et al.*, 1996; Hurles *et al.*, 2003b). Hurles *et al.* (2003b) discovered the American-specific Q3 Y chromosome lineage in two (of 16) men from Rapa (Austral Group), consistent with prehistoric movement of South Americans into the Pacific. But further research, this time historical, revealed that South American crew from a Peruvian slave ship settled on Rapa in 1863, and subsequently produced descendents, meaning a European-era introduction of the Q3 lineage could not be excluded (Hurles *et al.*, 2003b).

More recently, Lie *et al.* (2007) analysed mtDNA and Y chromosome markers and performed high resolution human leukocyte antigen (HLA) genotyping on 48 reputedly non-admixed native Easter Islanders. Several individuals were found to have typically Amerindian HLA alleles, and by combining this information with the mtDNA and Y chromosome data, as well as the known genealogy, the introduction of the alleles to Easter Island could be dated to the early 1800s or earlier (Lie *et al.*, 2007). Thus, these alleles are derived from either a very early (pre-slave trade) historic introduction of Amerindian individuals to Easter Island, or alternatively, from Amerindian

individuals who joined Polynesians on a prehistoric return trip from South America (Lie *et al.*, 2007).

1.4.2 SAILING TO THE NEW WORLD COAST (AND BACK)

Despite vigorous debate, there is growing agreement, based on translocated artefacts from archaeological sites, computer simulations, experimental craft and oral histories recorded by European voyagers, that pre-European Polynesian voyagers were competent sailors, with sophisticated canoe technologies, and that voyaging was largely deliberate and planned (Irwin, 1992; Finney, 1994, 1997; Irwin, 2006, p. 79; Collerson & Weisler, 2007; Finney, 2007). Voyages, at least initially, were mainly upwind because this makes the return easier and safer (Irwin, 2006, p. 82). And in Eastern Polynesia the settlement chronologies are consistent with prioritising safe and strategic dispersal over settling islands that were simply closest in terms of geographical proximity (Irwin, 2006, p. 82). Notions that voyagers sailed away from islands to which they could not return, became lost at sea, or were caught in unexpected winds and happened upon islands by chance are increasingly rejected (Irwin, 1992; Murray-McIntosh *et al.*, 1998).

Analysis of weather patterns, currents and Polynesian sailing technologies, often combined with computer simulations, have shown that return sailing voyages from various departure points in Eastern Polynesia to the New World were possible and often had high probabilities of success. Sailing from Eastern Polynesia to the New World appears difficult because it involves sailing against the direction both of the prevailing trade winds and accompanying currents (Finney, 1994). There are however, regular (annual and El Niño) westerly wind shifts that Polynesian voyagers could have exploited (Finney, 1994). In the case of a voyage to South America, there are also more southerly routes where westerlies prevail (Finney, 1994). Upon approaching South America, southerly winds and the north-flowing Humboldt Current would help carry the voyagers up the coast (Irwin, 2006, p. 85). Similarly, canoes from Hawai‘i could sail north to clear the high pressure system in that region (clockwise rotating in the Northern

Hemisphere), before turning east with the westerly winds. In both hemispheres voyagers could return with the trade winds (Irwin, 2006, p. 85).

Using computer simulations, Irwin (1992) tested the feasibility of a Polynesian double-hulled sailing canoe reaching the coast of South America from Easter Island. Input parameters included weather data, canoe performance, navigational skill and voyaging strategy. It was found that between 50–70% of canoes reached South America, and, by allowing for a return to Easter Island or Sala y Gómez, 95% survival was achieved. Conversely, less than 0.1% of canoes originating from the Marquesas Islands reached South America (Irwin, 1992, pp. 163–164).

More recent computer simulations (Fitzpatrick & Callaghan, 2009) have determined success rates of voyages to the New World from various Polynesian islands (Samoa, Tonga, Niue, the Tuamotus, Easter Island and Hawai‘i). Vessels launched from the Tuamotus were successful in January and December, with 15% and 23% of them making landfall in South America (Fitzpatrick & Callaghan, 2009). Of vessels launched from Easter Island, 100% reached South America and in all months. For Hawai‘i, only 1% of vessels reached South America (January), but in August and September, 67% and 76% of vessels made landfall in the region from southern Mexico to Central America (Fitzpatrick & Callaghan, 2009).

The specific point(s) of suggested contact in South America, and the likely Polynesian island(s) to which the voyagers returned will be dealt with further in Chapter Four.

1.5 AIMS AND HYPOTHESES

The aims of the project are:

1. to reconstruct the origins and dispersal of the sweet potato (*Ipomoea batatas*) and bottle gourd (*Lagenaria siceraria*) in Oceania using high resolution molecular (DNA) data obtained from modern accessions, and to analyse these data using appropriate population genetic and phylogenetic methods,
2. to interpret these data within an integrated framework of a commensal model, and in the context of relevant biological and anthropological (linguistic and archaeological) information, and
3. to infer patterns of crop dispersal and human mobility.

The hypothesis for the bottle gourd is that Polynesian accessions of this species are of South American origin. The alternative hypothesis is that they are of Asian origin (Green, 2000b).

There is now very little doubt that the initial introduction of the sweet potato into Oceania was the result of a prehistoric human-mediated introduction to Polynesia from the New World (Green, 2005). The hypotheses to be tested in this thesis focus on more specific events: the location(s) in South America where Polynesians made landfall, the number of Polynesian (*kumara*) lineages introduced into Polynesia, the dispersal routes within Eastern Polynesia, the dispersal routes of 16th century Iberian (*camote* and *batata*) introductions in the Western Pacific, and the origin(s) of New Zealand sweet potato accessions (Green, 2005)

More specific aims and hypotheses for the bottle gourd and sweet potato are stated in Sections 3.2.7 and 4.2.5 respectively.

CHAPTER TWO

MOLECULAR MARKERS IN BOTTLE GOURD AND SWEET POTATO

A Man Wearing a Bottle Gourd Mask
Kealakekua Bay, Island of Hawai'i, Hawai'i, March 1779
Drawing by John Webber, artist on Captain James Cook's Third Voyage
Reproduced, with permission of the Hawaiian Historical Society, Honolulu, Hawai'i, USA, from Nordyke (1999, p. 126)
Copyright © 1999 by the Hawaiian Historical Society

2.1 CHAPTER OVERVIEW

This chapter begins by outlining the considerations and approaches for developing appropriate molecular markers for a given taxonomic group and scientific question, with a focus on closely-related taxa where genetic variation is low. In the context of these considerations, the molecular methods of choice are justified for the bottle gourd and sweet potato.

For the bottle gourd, PCR and sequencing markers derived from inter-SSR multilocus genetic fingerprints (so-called ISSR-derived SCAR markers) were chosen as an appropriate marker system. To improve the success rate of SCAR marker development the flanking regions of the ISSR products were obtained from genomic DNA using the thermal asymmetric interlaced (TAIL) PCR technique. Designing new PCR primers that will amplify the entire ISSR locus as well additional flanking regions can allow a monomorphic SCAR marker to be rescued, a dominant SCAR marker to be converted into a more informative codominant marker, and/or additional polymorphisms to be discovered in the flanking regions. The development of the SCAR markers using this combination of molecular methods is described in detail.

For the sweet potato, amplified fragment length polymorphism (AFLP) fingerprinting was chosen as an appropriate marker system. The AFLP technique is an increasingly popular component of the phylogenetic toolbox, particularly for plant species. Technological advances in capillary electrophoresis now allow very precise estimates of DNA fragment mobility and amplitude, and current AFLP software allows greater control of data scoring and the production of the binary character matrix. However, for AFLP to become a useful modern tool for large datasets, improvements to automated scoring are required. Therefore, this section describes a procedure that can be used to optimise AFLP scoring parameters to improve phylogenetic resolution. This procedure, which includes measures of resolution, number of characters and error rates, is used to determine the optimum parameter settings for a 30-taxon *Ipomoea* dataset. These optimum settings are then used to analyse the large (300+) sweet potato dataset in Chapter Four.

2.1.1 A NOTE ON ATTRIBUTION

This chapter is mostly my own work, although the sections dealing with the AFLP technique (Sections 2.2, 2.7–2.11) are based on research carried out in collaboration with other researchers. This collaborative research includes:

1. a review of the AFLP technique (applications, analyses and advances), which was carried out in collaboration with Dr Heidi Meudt (Museum of New Zealand Te Papa Tongarewa, Wellington), and was published in *Trends in Plant Science* (Meudt & Clarke, 2007). A reprint of the paper is provided in Appendix 7. Both Dr Meudt and I decided upon the scope of the review, performed the literature search, synthesised the primary literature, and managed the review towards publication. My individual contribution was weighted more towards a review of AFLP protocols, lab implementation and analysis software, whereas Dr Meudt's contribution was weighted more towards AFLP theory, analysis methods and future research.
2. the tuning of automated AFLP scoring parameters to improve phylogenetic resolution, which was carried out in collaboration with Dr Barbara Holland (Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North) and Dr Heidi Meudt, and was published in *Systematic Biology* (Holland *et al.*, 2008). A reprint of the paper is provided in Appendix 8. For this research it is difficult to separate out each author's contribution, although Dr Meudt and I contributed more towards the biological aspects (including generating the raw data), whereas Dr Holland contributed more towards the mathematical aspects (including analyses of the scored data).

Sections of Meudt & Clarke (2007) and Holland *et al.* (2008) relevant to the thesis are summarised below. Some of this content will include contributions from Drs Meudt and Holland (either directly or indirectly), but I have tried to include only my own contributions where possible. For more detailed information, the papers themselves should be consulted.

2.2 INTRODUCTION

2.2.1 MOLECULAR MARKERS AND THE DETECTION OF GENETIC VARIATION

Using molecular markers to reconstruct evolutionary relationships between recently-diverged plant taxa, especially at the intrageneric and intraspecific levels, can be very challenging. The reasons for this difficulty include low genetic diversity (i.e., few polymorphic loci), variations in ploidy levels and complex patterns of evolution (e.g., incomplete lineage sorting and hybridisation) (Hughes *et al.*, 2006). All of these factors constrain the identification of appropriate molecular markers.

Strategies for reconstructing evolutionarily recent relationships are increasingly relying on marker systems which sample multiple, independent loci. These systems comprise both low-copy, codominant nuclear markers and dominant, multilocus fingerprinting marker systems. The low-copy markers include DNA sequencing markers, microsatellite (or simple sequence repeat, SSR) markers (Zane *et al.*, 2002), single nucleotide polymorphisms (SNPs) (Morin *et al.*, 2004) and sequence-characterised amplified region (SCAR) markers (Bailey *et al.*, 2004). Dominant, multilocus marker systems include amplified fragment length polymorphisms (AFLPs), randomly amplified polymorphic DNA (RAPDs) and inter-SSRs (ISSRs). The shift towards these low-copy and/or multilocus marker systems is being driven by the need for markers which can resolve more complex evolutionary histories, coupled with the increasing ease with which these more technically demanding markers can be developed and implemented.

Conventional markers mainly include chloroplast DNA (cpDNA) markers (e.g., Taberlet *et al.*, 1991) and the nuclear ribosomal Internal Transcribed Spacer (nrITS) (White *et al.*, 1990; Baldwin, 1992). Whilst technically straight-forward and often highly-polymorphic, they are increasingly recognised as inappropriate for reconstructing recent evolutionary relationships, especially if they constitute the only source of data (Chiang, 2000; Álvarez & Wendel, 2003; Zhang & Hewitt, 2003; Bailey *et al.*, 2004; Mort & Crawford, 2004; Small *et al.*, 2004). Both the chloroplast genome

and nrITS evolve, and are transmitted, by processes that are different from the rest of the genome; for example, the chloroplast is usually maternally inherited (Provan *et al.*, 2001) and the nrITS evolves by concerted evolution (Álvarez & Wendel, 2003). These and other molecular processes mean, especially over short timescales, that the evolutionary history of the chloroplast and the nrITS may not be representative of the evolutionary history of the remainder of the genome, nor approximate the evolutionary history of the lineage as a whole (Hughes *et al.*, 2006). In contrast, multilocus and low-copy nuclear markers, because of their Mendelian, bi-parental inheritance patterns and broader genome coverage, are increasingly recognised as superior sources of raw genetic data for population genetics and the reconstruction of shallow phylogenies.

2.2.1.1 Multilocus DNA Fingerprinting

The three most common techniques for multilocus genomic fingerprinting are AFLPs (Vos *et al.*, 1995), RAPDs (Welsh & McClelland, 1990; Williams *et al.*, 1990), and ISSRs (Zietkiewicz *et al.*, 1994). They are PCR-based techniques that amplify previously uncharacterised DNA fragments, and can therefore be used on organisms for which there is no *a priori* sequence information (Schlötterer, 2004).

The AFLP, RAPD and ISSR techniques vary with respect to technical ease, the amount of DNA required, reproducibility, data quality, genetic variability and discriminatory power (Savelkoul *et al.*, 1999; McGregor *et al.*, 2000; Saliba-Colombani *et al.*, 2000; Archak *et al.*, 2003; Bussell *et al.*, 2005). In many studies AFLPs outperform both ISSRs and RAPDs in their high reproducibility, robustness, informativeness, and fewer reported reaction artefacts (Jones *et al.*, 1997; Blears *et al.*, 1998; Savelkoul *et al.*, 1999; McGregor *et al.*, 2000; Archak *et al.*, 2003; Belaj *et al.*, 2003; Bussell *et al.*, 2005). AFLP, however, requires more DNA, is more technically difficult, and is generally more expensive, than the ISSR and RAPD techniques (Archibald *et al.*, 2006b).

Multilocus DNA fingerprinting techniques can be ideal in the following situations: when there is no *a priori* sequence information, for intra-specific studies, when genomic heterogeneity is high (i.e., when it is necessary to amplify many loci to

ascertain an accurate measure of genomic diversity, e.g., outcrossing species), when genetic variability is low (i.e., when it is necessary to amplify many loci to locate the few that are polymorphic, e.g., crop species), in polyploids, when hybridisation is occurring, for the rapid generation of data, when high quality DNA is available, and where there are no suitable established markers (Mueller & Wolfenbarger, 1999; Sunnucks, 2000; Schlötterer, 2004; Bensch & Åkesson, 2005; Meudt & Clarke, 2007).

2.2.1.2 Low-Copy Nuclear Markers versus Multilocus Markers

Recent studies have compared dominant marker systems, such as AFLP, RAPD and ISSR fingerprinting, to codominant markers such as microsatellites, SNPs and SCAR markers (e.g., Jones *et al.*, 1997; Gerber *et al.*, 2000; Belaj *et al.*, 2003; Campbell *et al.*, 2003; Squirrell *et al.*, 2003; Schlötterer, 2004; Bensch & Åkesson, 2005). Codominant microsatellite and sequencing marker systems typically comprise a few (5–20) highly informative multiallelic loci with high discriminating capacity. Codominant markers are very powerful because they allow the number of copies of a given allele at a specific locus to be determined, in turn allowing heterozygous (+ −) and homozygous (+ +) individuals to be distinguished. This additional information makes codominant markers much more powerful for linkage mapping, parentage analysis, identifying hybrids and population genetics. The use of codominant markers is largely restricted to diploids however; as ploidy increases, the ability to accurately determine allele dosage is limited. For example, in sweet potato — a hexaploid — it would be extremely difficult to use sequencing markers to obtain data for phylogenetic analysis, especially for hundreds of individuals.

In contrast, dominant markers reveal only whether an allele is present or absent, thus heterozygous (+ −) and homozygous (+ +) individuals cannot be distinguished, with both genotypes being rendered as identical (+) phenotypes. On a per locus basis, this makes dominant marker systems, such as AFLP, less informative than codominant marker systems. However, dominant marker systems derive their statistical power from their sheer number, with a typical study comprising many hundreds of markers distributed genome-wide (Sunnucks, 2000; Mariette *et al.*, 2002; Belaj *et al.*, 2003;

Campbell *et al.*, 2003). In addition, multilocus marker systems are much more easily implemented in polyploids than most codominant marker systems.

Both codominant and dominant marker systems are commonly employed for measuring population genetic structure and diversity (Nybom, 2004; Woodhead *et al.*, 2005). The few explicit comparisons of the relative utility of microsatellites and multilocus fingerprinting markers in plants (e.g., Nybom, 2004; Woodhead *et al.*, 2005; Garoia *et al.*, 2007; Jump & Peñuelas, 2007) show that the two systems give congruent results when a sufficiently large number of microsatellite loci are analysed. For example, in a study of the fern *Athyrium distentifolium*, 18 microsatellite loci gave results comparable to 265 AFLP markers (Woodhead *et al.*, 2005). In some studies, multilocus fingerprinting markers, because of the very large number of markers generated, out-perform microsatellites for discriminating taxa and populations (Campbell *et al.*, 2003; Woodhead *et al.*, 2005), although comparisons are difficult when only a few microsatellite loci have been used (Jones *et al.*, 1997; Campbell *et al.*, 2003; Woodhead *et al.*, 2005; Garoia *et al.*, 2007). Other variables, such as homoplasy, genomic heterogeneity, and population heterogeneity can shift the boundary at which a given number of dominant, multilocus fingerprinting-derived markers becomes more informative than a given number of microsatellite markers (Mariette *et al.*, 2002; Kremer *et al.*, 2005).

2.2.2 CONSIDERATIONS IN CHOOSING MOLECULAR MARKER SYSTEMS

Many molecular marker systems are available for reconstructing evolutionary relationships in plants, and the choice of the most appropriate marker system should be based on considerations under each of three broad categories:

1. ***Biological***

   Including the amount of genetic variability (if known), taxonomic breadth (e.g., inter-specific, intra-specific), genome size, occurrence of hybridisation, and ploidy (Woodhead *et al.*, 2005).

2. ***Research Questions***

   Including the application of the technique (e.g., population genetics, phylogenetics) (Sunnucks, 2000; Belaj *et al.*, 2003; Schlötterer, 2004), whether codominant markers are required, the need for cross-study comparisons, whether ancient DNA (aDNA) will be used, and the life-expectancy of the research project (i.e., cost–benefit analysis of different marker systems through time; for recent divergences, multilocus fingerprinting can be better in the short-term, microsatellites better in the long-term) (Campbell *et al.*, 2003; Squirrell *et al.*, 2003; Woodhead *et al.*, 2005).

3. ***Available Resources***

   Including the quality and quantity of available tissue, previously established genetic resources (e.g., linkage maps, markers established in the same or closely related taxa, sequence data, expressed sequence tag (EST) libraries), and logistics (e.g., available funding and time, technical difficulty, access to training and laboratory facilities) (Jones *et al.*, 1997; Belaj *et al.*, 2003; Bensch & Åkesson, 2005; Woodhead *et al.*, 2005). In New Zealand, the time and financial cost required to obtain permission to develop genetically modified organisms (GMOs) (e.g., for cloning of PCR products) is a significant resource constraint.

2.2.3 MOLECULAR MARKER SYSTEMS: STRATEGIES IN CROP PLANTS

The problem of isolating sufficient polymorphic loci can be compounded in crop species, where the bottlenecks of domestication and ongoing selection may have further reduced genetic variation. However, because of their commercial/economic importance, crop plants often have the significant advantage over wild species of well-developed genetic resources. These resources, although usually developed for agronomic purposes, can be applied to ethnobotanical, evolutionary and anthropological questions such as crop domestication and dispersal, and human mobility.

The strategies for obtaining molecular markers for evolutionary studies in non-model crop plants fit into five groups. These groups are listed below in approximately increasing order of technical difficulty and time required. However, the ease with which markers can be developed should be balanced against their appropriateness based on the biological, scientific and resource considerations listed above.

1. ***Employ established species-specific markers***

   Depending on the species, resources can range from a few codominant markers (e.g., microsatellites, SNPs, sequencing markers), to EST sequences, genetic maps, and, in a few cases, whole genome sequences (e.g., rice (*Oryza sativa*)).

2. ***Transfer species-specific markers from closely-related taxa***

   It may be possible to transfer markers to the species of interest from a closely-related species with well-developed genetic resources (e.g., to bottle gourd from watermelon), although this requires screening and usually PCR re-optimisation.

3. ***Screen universal organellar and nuclear markers***

   Universal markers (e.g., chloroplast markers, nrITS, introns of conserved nuclear genes) can be screened (Chung *et al.*, 2003; Chung & Staub, 2003; Mort & Crawford, 2004). Although universal markers may not have sufficient variation at the intra-specific level (Bailey *et al.*, 2004; Hughes *et al.*, 2006) nor be appropriate as the only source of data (Álvarez & Wendel, 2003), they are usually very easy to amplify and can be easily screened before more complicated molecular techniques are employed (Mort & Crawford, 2004).

4. ***Use multi-locus fingerprinting***

   Multilocus fingerprinting techniques such as AFLP, ISSR and RAPD are easily implemented in most species, and, because no prior sequence information is required, data can be generated quickly (see above).

5. ***Develop species-specific markers* de novo**

   Species-specific markers developed *de novo* include microsatellite and SCAR markers. Microsatellite markers are usually developed by screening microsatellite-enriched genomic libraries. SCAR markers are derived from multilocus fingerprinting techniques such as AFLP, ISSR and RAPD (and will be discussed in more detail below).

It should be noted that the above strategies are already being improved and superseded by new technologies, especially by next-generation sequencing technologies (e.g., Margulies *et al.*, 2005). Although the potential applications (and limitations) of next-generation sequencing technologies are still being realised, it is already clear that they will allow rapid re-sequencing and assembly of simple genomes as well as polymorphism discovery via derived techniques such as Complexity Reduction of Polymorphic Sequences (CRoPS), which allows SNP and microsatellite markers to be obtained from pyrosequenced AFLP fragments (van Eijk, 2006). Both next-generation sequencing technologies and increasingly efficient *de novo* sequencing using conventional (i.e., Sanger sequencing) methods will allow polymorphic loci to be isolated more efficiently, and in greater numbers, allowing new and more complex evolutionary questions to be answered. The next generation sequencing technologies were not available when this study began.

2.2.4 MOLECULAR MARKERS IN BOTTLE GOURD: CONSIDERATIONS, AVAILABLE MARKERS AND STRATEGIES

No appropriate marker systems were available for the bottle gourd at the start of this project. As a crop, bottle gourd is of little commercial importance; it is primarily grown as a utilitarian crop by small land holders in tropical, developing countries or, in the developed world, as a novelty/ornamental/culturally important crop (especially the United States) (Decker-Walters *et al.*, 2001). Commercially, the bottle gourd is used as a rootstock for various cucurbit species (mainly watermelon (*Citrullus lanatus*)) in an effort to improve the yield and quality of fruit produced by the scion species, and the little scientific research undertaken on the bottle gourd has focused in this area (e.g., Yetisir & Sari, 2003; Han *et al.*, 2005).

There are, however, extensive genetic resources (e.g., EST libraries, microsatellite and sequencing markers) developed for cucurbit species that are of commercial importance — including watermelon (e.g., Jarret *et al.*, 1997), rockmelon/cantaloupe (*Cucumis melo*) (e.g., Staub *et al.*, 2000; Chiba *et al.*, 2003;

Gonzalo *et al.*, 2005), cucumber (*Cucumis sativus*) (e.g., Danin-Poleg *et al.*, 2001), and the pumpkins and squashes (mainly *Cucurbita pepo* and *Cucurbita maxima*).

Other considerations when choosing an appropriate marker system (see Section 2.2.2) for the bottle gourd are: it is diploid (Singh, 1990, p. 20) (i.e., it is possible to use sequencing markers, *cf.* polyploids), the markers only need to operate at the intraspecific level (although some sequence data from outgroup species may be useful), and that the available tissue is of variable quality (ranging from fresh seeds and leaf tissue, to 9,000 yr BP archaeological material).

Taking into account the considerations detailed above, three strategies were decided on for developing molecular markers in bottle gourd:

1. ***Develop a set of bottle gourd-specific, ISSR-derived SCAR markers (i.e., nuclear, codominant, sequencing markers)***
   The development of SCAR markers is comparatively expensive and is time-consuming. However, SCAR markers can be tremendously powerful, providing high resolution, co-dominant sequence data from multiple, un-linked loci. Also, once developed, the markers are easily implemented by, and the resulting sequence data shared with, other researchers (*cf.* multi-locus marker systems, for which methods and data are difficult to transfer between research groups).
2. ***Attempt to transfer microsatellite markers from closely-related cucurbit species***
   It is a quick, simple and relatively inexpensive process to screen existing cucurbit microsatellite markers in bottle gourd. The first step is to amplify the marker in bottle gourd (this may require re-optimising the PCR conditions), and, if this is successful, the second step is to determine whether the marker is polymorphic by amplifying it in a small number of diverse bottle gourd accessions and characterising the products using DNA sequencing or capillary electrophoresis.
3. ***Screen established chloroplast markers***
   It is also a quick, simple and relative inexpensive process to screen established, universal cpDNA markers, including the *trn*L–F and *trn*C–D intergenic spacers and Chung's 23 consensus chloroplast SSR (ccSSR) markers (Chung *et al.*, 2003; Chung & Staub, 2003). Although there are limitations of using

uniparentally-inherited chloroplast markers for reconstructing recent divergences (see above), these markers still possess significant advantages. Firstly, the high copy number means they are more likely to be amplified from poor quality and ancient samples, and, secondly, the chloroplast markers will be unaffected by paternal (i.e., pollen-mediated) genetic "contamination" from other cultivars — a concern when living collections of diverse, sexually reproducing accessions are maintained together.

2.2.5 MOLECULAR MARKERS IN SWEET POTATO: CONSIDERATIONS, AVAILABLE MARKERS AND STRATEGIES

Sweet potato is a commercially important crop on a global scale. World production of sweet potato in 2007 was 126 million tonnes, the twelfth largest crop by total production (FAO, 2008). In Melanesia, Micronesia and Polynesia, sweet potato is the fifth, sixth and ninth largest crop respectively (FAO, 2008). Sweet potato is also implicated in important questions about early crop domestication and dispersal in the New World (e.g., Zhang *et al.*, 2004; Srisuwan *et al.*, 2006). For both these reasons, there has been a relatively large amount of genetic research carried out on sweet potato. A majority of this research has employed AFLP markers, which have been used both for assessing germplasm diversity (e.g., Zhang *et al.*, 2004; He *et al.*, 2006), and for constructing linkage maps (e.g., Kriegner *et al.*, 2003). In addition, microsatellite markers have been developed for sweet potato (Hu *et al.*, 2004a; Hu *et al.*, 2004b).

Sweet potato is hexaploid (Fajardo *et al.*, 2002), which constrains the type of molecular markers that can be used for phylogenetic applications. Single-locus DNA sequencing markers are impracticable, especially for large-scale projects, because of the multiple (up to six) alleles that might be present at each locus. Multilocus fingerprinting techniques are commonly used in polyploid species, although relative to diploids, polyploids generally produce higher numbers of AFLP fragments (Kardolus *et al.*, 1998; Fay *et al.*, 2005; Guo *et al.*, 2006) with highly complex polymorphisms involving multiple loci and alleles that can make determination of allele dosages problematic (Alonso-Blanco *et al.*, 1998; McGregor *et al.*, 2000).

The sweet potato genome, at 2205 Mbp (Bennett & Leitch, 2004), is relatively large. Organisms with large genomes, which can contain large amounts of repetitive DNA and retrotransposons, tend to have more fragments in their AFLP profiles than those with smaller genomes, but also frequently give rise to profiles with many low-intensity peaks that are difficult to score (Kardolus *et al.*, 1998; Fay *et al.*, 2005).

Other considerations when choosing an appropriate marker system (see Section 2.2.2) for the sweet potato are that the markers only need to operate at the intraspecific level (although some sequence data from outgroup species may be useful), and that the available tissue is of high quality (fresh tissue collected from living collections).

Sweet potato is generally vegetatively propagated (Fajardo *et al.*, 2002), either from tubers or stem cuttings. Vegetative propagation is currently used both in South America and Polynesia, and even in prehistory there is no evidence in either region of sweet potato being grown from true seed (see Yen, 1974). Modern commercial propagation is also vegetative, and is achieved by excising young stems from sprouted tubers. These vegetative cuttings (called slips) are the "seedlings" often available from plant nurseries. In Polynesia, as elsewhere, new varieties of sweet potato were selected from mutants, or sports, of existing varieties (see Yen, 1974). This is true of new commercial varieties also (see Lewthwaite, 2004). A known exception to vegetative propagation is the sexual reproduction carried out as a part of some modern breeding programmes. Because of the asexual reproduction of sweet potato, the species can be expected to evolve in a relatively tree-like way. Therefore phylogenetic methods for reconstructing relationships between taxa are appropriate, and these should be unaffected by the problems of reticulation, incomplete lineage sorting and 'gene tree' versus 'species tree' incongruence that can affect the reconstruction of shallow divergence events (see also Holland *et al.*, 2008; Appendix 8).

Taking into account the considerations detailed above, it was decided the AFLP technique would be the best strategy for developing multiple, genome-wide markers for sweet potato. AFLPs are more reproducible and robust than other multilocus fingerprinting techniques (see Section 2.2.1.1 above). They are also commonly used for in polyploid taxa (see above), and when genetic variability is low, e.g., in crop species (Kilian *et al.*, 2007). In addition, AFLP data are increasingly being used to estimate phylogenies, including for crops and their wild progenitors (e.g., Spooner *et al.*, 2005; Kilian *et al.*, 2007), and have been shown to result in well-resolved trees that are consistent with independent data (Bussell *et al.*, 2005; Koopman, 2005; see Section 2.8).

2.3 SCAR MARKERS: INTRODUCTION AND METHODOLOGY

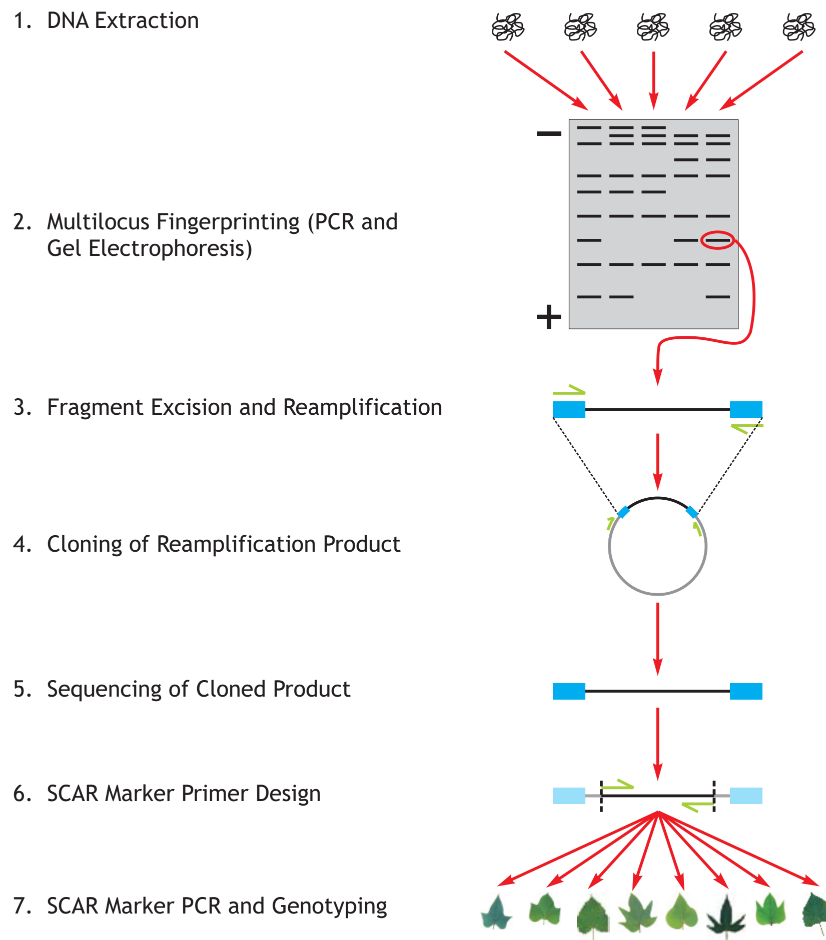
2.3.1 AN INTRODUCTION TO SCAR MARKERS

Sequence-characterised amplified region (SCAR) markers are (usually) nuclear, codominant, single-locus PCR markers that are derived using multilocus fingerprinting techniques. SCAR markers were originally, and are still primarily, developed for plant breeding applications, especially for marker assisted selection (MAS) where the SCAR is linked to an agronomically-important trait (e.g., Negi *et al.*, 2000; Flachowsky *et al.*, 2001; Xu *et al.*, 2001; Boukar *et al.*, 2004; Noguera *et al.*, 2005; Hong *et al.*, 2006). SCAR markers are, however, increasingly important in plant evolutionary research — especially as a source of low-copy nuclear markers for phylogenetic and population genetic applications (e.g., Lockhart & McLenachan, 1997; McLenachan *et al.*, 2000; Scotti *et al.*, 2000; Hagen *et al.*, 2001; Bailey *et al.*, 2004; Mort & Crawford, 2004). They have also been used to elucidate the molecular basis of heteromorphy, e.g., the famous pin and thrum flower development in *Primula vulgaris* (Manfield *et al.*, 2005).

2.3.1.1 Development of SCAR Markers

SCAR markers are usually derived from the multilocus DNA fingerprinting techniques such as AFLP, ISSR and RAPD. Polymorphic loci — and therefore potential SCAR markers — are identified by subjecting a small number of core taxa (~5–20) to multilocus fingerprinting and polyacrylamide gel electrophoresis (PAGE). Polymorphic bands are excised from the gel, re-amplified, and either sequenced directly or cloned and sequenced. The sequence information is used to design locus-specific PCR primers that can be used to amplify the polymorphic marker directly from genomic DNA of any taxon in the group under study (Fig. 2.1).

Fig. 2.1
Traditional Process of SCAR Marker Development



1. DNA Extraction
DNA is extracted from a small number of individuals (often 5–20).

2. Multilocus Fingerprinting
A multilocus fingerprint is generated using techniques such as amplified fragment length polymorphism (AFLP), inter-simple sequence repeat (ISSR) or randomly amplified polymorphic DNA (RAPD). The resulting PCR products are separated on a denaturing polyacrylamide gel and visualised (e.g., by silver-staining).

3. Fragment Excision and Re-amplification
Bands present in one individual, and not in another, represent polymorphic loci. These bands are excised from the gel and re-amplified using PCR.

4.–5. Cloning and Sequencing of Re-amplification Products
Each re-amplified fragment is cloned into a vector and sequenced using the vector's universal primers.

6. SCAR Marker Primer Design
Sequence-characterised amplified region (SCAR) marker primers are designed from the ends of the sequenced product. The primers allow the corresponding locus to be amplified from genomic DNA.

7. SCAR Marker PCR and Genotyping/Sequencing
The SCAR markers are amplified in multiple accessions. SCAR markers are usually characterised by DNA sequencing, but can also be developed into fluorescently-labelled microsatellite markers, cleaved amplified polymorphic sequence (CAPS) markers or single nucleotide polymorphism (SNP) markers.

The composition of the core set for the initial multilocus fingerprinting depends on the application. For plant breeding, the core set represents the phenotypes which need to be distinguished. For evolutionary applications, the core set should represent the range of genetic diversity (estimated from previous genetic work, morphology, geography, etc.) to which the resulting SCAR markers will be applied. This ensures the most closely-related taxa can be distinguished, and the most distant taxa can be amplified.

Electrophoresis and isolation of multilocus fingerprinting products for SCAR marker development can be performed in various ways. Electrophoresis systems include agarose, Separide™ (Lockhart & McLenachan, 1997), and polyacrylamide. Of these, polyacrylamide provides the best resolution. A new technique to isolate fragments using capillary electrophoresis systems (Polanco *et al.*, 2005) may offer an alternative method, but the required destruction of the capillary array (!) would make it prohibitively expensive in most cases.

Multilocus fingerprint products separated by polyacrylamide electrophoresis are usually visualised either by silver-staining or radio-labelling. For isolating potential SCAR markers, silver-staining is a superior method for two reasons: first, because the bands are visualised directly and are easily excised (the radio-labelled method requires the awkward overlay of the developed film onto the gel), and, second, because all bands are visualised, it is easier to avoid inadvertently excising unwanted fragments (*cf.* radio-labelled products (see Xu *et al.*, 2001)).

SCARs are often developed into DNA sequencing markers because this ensures that the maximum amount of information is obtained. In cases where DNA sequencing markers cannot be used (e.g., for microsatellites and polyploids) or a simpler system is adequate, SCARs can be developed into microsatellite, SNP markers or cleaved amplified polymorphic sequence (CAPS) markers (e.g., Bensch *et al.*, 2002; Nicod & Largiadèr, 2003).

2.3.1.2 SCAR Markers Versus Multilocus Fingerprinting

A set of SCAR markers can possess several advantages over multilocus fingerprinting techniques (Paran & Michelmore, 1993; McLenachan *et al.*, 2000; Negi *et al.*, 2000; Zhang & Stommel, 2001; Brugmans *et al.*, 2003):

1. they are often codominant (homozygotes and heterozygotes can be distinguished — important for hybridisation questions),
2. the presence of DNA sequence data allows more powerful analysis methods,
3. results can be obtained more quickly, at reduced financial cost, and from many more individuals,
4. the quantity and quality of DNA required is lower (e.g., herbarium specimens and crude DNA extraction techniques can be used),
5. amplification is more robust and technically easier,
6. reproducibility (e.g., in different labs, or through time) is higher, and
7. cross-study comparisons and the addition of data to existing datasets are easier.

2.3.1.3 Sources of SCAR Markers

The most frequently-used fingerprinting techniques — RAPD, AFLP and ISSR — differ significantly in terms of technical ease, variability, reproducibility, and several other factors (see Section 2.2.1.1). These variables appear most important for determining which technique is superior for SCAR marker development. However, other variables, such as the distribution of polymorphisms, may make one fingerprinting technique superior to another for SCAR marker development. For example, SNPs that occur primarily in MLF primer-binding sites (which is thought to be the case with AFLP (Abbo *et al.*, 2001; Dussle *et al.*, 2002; Brugmans *et al.*, 2003)) will not be captured by traditional SCAR markers.

2.3.2 SCAR MARKERS DERIVED FROM INTER-SSR (ISSR) FINGERPRINTING

2.3.2.1 ISSR Fingerprinting: Introduction

Inter-SSR (ISSR) DNA fingerprinting (Zietkiewicz *et al.*, 1994) is a firmly-established marker technique for reconstructing plant evolutionary relationships, especially for recently-radiated taxa where DNA sequence data fail to show variability (Archibald *et al.*, 2006b). ISSRs have been particularly valuable for reconstructing recent radiations (Archibald *et al.*, 2006a) and in understanding the evolutionary importance of hybridisation (e.g., in *Penstemon* (Wolfe *et al.*, 1998a, 1998b; Datwyler & Wolfe, 2004)) and polyploidisation (e.g., in *Brassica* (Liu & Wang, 2006)). ISSRs have also been used to genetically characterise germplasm resources for crop species such as lemon (*Citrus limon*) (Capparelli *et al.*, 2004), oca (*Oxalis tuberosa*) (Pissard *et al.*, 2006), and peanut (*Arachis hypogaea*) (Raina *et al.*, 2001) and for wild species such as the Easter Is. endemic toromiro tree (*Sophora toromiro*) (Maunder *et al.*, 1999).

The performance of ISSR relative to the AFLP and RAPD techniques is covered in Section 2.2.1.1. AFLP is generally the superior of the three fingerprinting techniques, and was the first technique attempted for SCAR marker development in bottle gourd. However, AFLP proved to be insufficiently variable in bottle gourd, with only ~1% of loci identified as putatively polymorphic (results not shown). Because this is close to the random error rate of AFLP (see Bonin *et al.*, 2004, and references therein), the technique was not pursued. ISSR, on the other hand, proved to be much more variable, with ~5–10% putatively polymorphic loci, and was selected as the best method for SCAR marker development. Compared to RAPDs, ISSRs are more reproducible and often more variable (Esselman *et al.*, 1999; Crawford *et al.*, 2001; Archibald *et al.*, 2006b).

ISSR-derived SCAR markers, although apparently not yet used for evolutionary research, have recently been applied in plant breeding to isolate markers linked to phenotypes of interest: for example, somaclonal variation in maize (*Zea mays*) (Osipova *et al.*, 2003), seasonal flowering in the wild diploid strawberry (*Fragaria vesca*) (Albani *et al.*, 2004), and canola quality in oriental mustard (*Brassica juncea*) (Ripley & Roslinsky, 2005).

2.3.2.2 ISSR Fingerprinting: Methodology

ISSR PCR products are generated using a PCR primer comprising a ~15 nt repeat region with a 1–3 nt anchor at either the 5′ or 3′ end. The repeat region contains a 2–4 nt repeat motif (Zietkiewicz *et al.*, 1994). The anchor helps prevent the primer binding at multiple locations within a single repeat region (Wolfe *et al.*, 1998b). But even with the anchor, some slippage may occur, producing a characteristic stutter/ladder pattern of bands that represents products of different length but derived from the same locus (Yang *et al.*, 2001; Archibald *et al.*, 2006b). Slippage can be mitigated by using high-stringency PCR conditions (e.g., touchdown PCR) and a 3′ anchor. Although 5′ anchored-primers have been shown to bind non-specifically (Blair *et al.*, 1999; Pharmawati *et al.*, 2005; Archibald *et al.*, 2006b), there has been recent success using 5′-anchored ISSR primers to develop microsatellite markers in *Sphagnum* (Provan & Wilson, 2007). The process of generating ISSRs is outlined in Fig. 2.2.

Fig. 2.2
Overview of ISSR Fingerprinting

Genomic DNA
5′ NNNNNNNNGTGTGTGTGTGTGTGTGTGTANNNNNNNNNNNNNTACACACACACACACACACACNNNNNNNN 3′
3′ NNNNNNNNCACACACACACACACACACACATNNNNNNNNNNNNNATGTGTGTGTGTGTGTGTGTGNNNNNNNN 5′

5′ NNNNNNNNGTGTGTGTGTGTGTGTGTGTANNNNNNNNNNNNNTACACACACACACACACACACNNNNNNNN 3′
3′ A(TG)$_8$ 5′
5′ (GT)$_8$A 3′
3′ NNNNNNNNCACACACACACACACACACACATNNNNNNNNNNNNNATGTGTGTGTGTGTGTGTGTGNNNNNNNN 5′

PCR

PCR Product
5′ GTGTGTGTGTGTGTGTANNNNNNNNNNNNNTACACACACACACACAC 3′
3′ CACACACACACACACACATNNNNNNNNNNNNNATGTGTGTGTGTGTGTG 5′

The $(GT)_8A$ primer (green) binds to two inversely orientated $(GT)_{\geq 8}A$ SSRs (blue) in the genomic DNA, anchoring at the 3′ ends of the repeat region. The primers will only bind when the anchoring base at the 3′ end of the primer (in this example an A) is complementary to the base flanking the SSR in the genomic sequence. PCR is used to amplify the region between the two SSRs. The primers bind to and amplify multiple loci, generating a multilocus fingerprint.

Modified from Fig. 1, Zietkiewicz et al. (1994).

ISSR polymorphisms can be caused by a change in the microsatellite sequence (e.g., loss of repeat units), a change in the anchoring base (e.g., SNP), or a length polymorphism between the two SSRs (e.g., indel).

2.3.3 PROBLEMS AND STRATEGIES IN SCAR MARKER DEVELOPMENT

Ideally, all developed SCARs would retain the polymorphism observed on the fingerprinting gel and yield one of the following outcomes:

1. ***A codominant SCAR marker containing a length polymorphism***

   This occurs when the SCAR marker amplifies alleles of different length, and may indicate that the excised band was part of an allelic band pair (homologous alleles of different mobility on the fingerprinting gel). If only the character state of the length polymorphism needs to be determined, then standard electrophoresis techniques can be used. If the marker contains other useful polymorphisms then more advanced techniques can be used (e.g., DNA sequencing, SSCP and CAPS). Although these methods are suitable when all alleles are known, DNA sequencing often gives superior results — providing the maximum amount of information and ensuring that previously unknown alleles are identified.

2. ***A dominant presence–absence SCAR marker***

   This occurs when the SCAR marker only amplifies a product in some taxa (the plus-allele) and not in others (the null-allele). Although these SCAR markers are easily characterised by agarose electrophoresis, their dominant nature means they are of limited use. In addition, false negatives are difficult to detect without multiplexing the SCAR marker with a marker that always amplifies a product (e.g., a conserved gene).

In practice, however, a third outcome of SCAR marker development is common:

3. ***Amplification of equal-length (monomorphic) products in all taxa***

   This occurs if the polymorphism responsible for the original presence–absence phenotype on the fingerprinting gel is lost because it is located in the regions flanking the SCAR primer-binding sites (Yang *et al.*, 2001; Zhang & Stommel, 2001; Dussle *et al.*, 2002; Brugmans *et al.*, 2003; Albani *et al.*, 2004). For

ISSRs, this includes polymorphisms in the ISSR primer-binding site (e.g., SSR, SNP) or in the region between the ISSR primer and the SCAR primer. For AFLPs, this includes polymorphisms caused by the gain or loss of a restriction site, changes in the selective primer regions, or length polymorphisms between the selective region and the SCAR primer will not be amplified. There are two further explanations for a monomorphic SCAR: first, the 'polymorphism' observed on the gel was due to inconsistent amplification. This can be avoided by fingerprinting replicates of all taxa, and only isolating bands which are present in both replicates. Second, a monomorphic band close to the band of interest was inadvertently excised and re-amplified. This can be avoided by ensuring maximum separation between bands (e.g., by using polyacrylamide), sequencing multiples clones of the re-amplification product, and comparing the size of the clone insert to the size of the original fragment (De Jong *et al.*, 1997; Xu *et al.*, 2001).

An apparently monomorphic SCAR (outcome "3" above) can sometimes be 'rescued' despite the loss of the original polymorphism. Approaches to recover polymorphisms include optimisation of PCR (to favour amplification of one allele over another — producing a dominant marker), sequencing the product to locate SNPs that can be used to develop codominant CAPS or sequencing markers, and primer redesign (e.g., Paran & Michelmore, 1993; Deng *et al.*, 1997; Zhang & Stommel, 2001; Brugmans *et al.*, 2003; Albani *et al.*, 2004).

Even employing the rescue techniques outlined, standard SCAR marker development is plagued by a low success rate, which is compounded by the technical difficulty of development (Negi *et al.*, 2000; Xu *et al.*, 2001; Yang *et al.*, 2001; Zhang & Stommel, 2001; Brugmans *et al.*, 2003). For example, McLenachan *et al.* (2000) found that of 46 AFLP-derived SCAR markers, only ~50% were polymorphic when electrophoresed on polyacrylamide or sequenced. Paran & Michelmore (1993) found that for nine RAPD polymorphisms, six were caused by SNPs in the binding sites of the RAPD primers, and therefore would be lost if traditional SCAR markers were developed from the internal regions. These low success rates are consistent with other studies (e.g., Bradeen & Simon, 1998; Dussle *et al.*, 2002; but see Brugmans *et al.*, 2003).

2.3.3.1 Obtaining Flanking Regions to Improve SCAR Marker Development

The success rate of SCAR marker development is improved by obtaining the regions flanking the MLF fragment from genomic DNA. The flanking regions are used to design *new* PCR primers that amplify a larger region which includes the entire MLF locus, allowing polymorphisms occurring in the flanking regions to be amplified (Fig. 2.3). Designing new primers in the flanking regions can also allow dominant SCAR markers to be converted into more informative codominant markers (see Deng *et al.*, 1997).

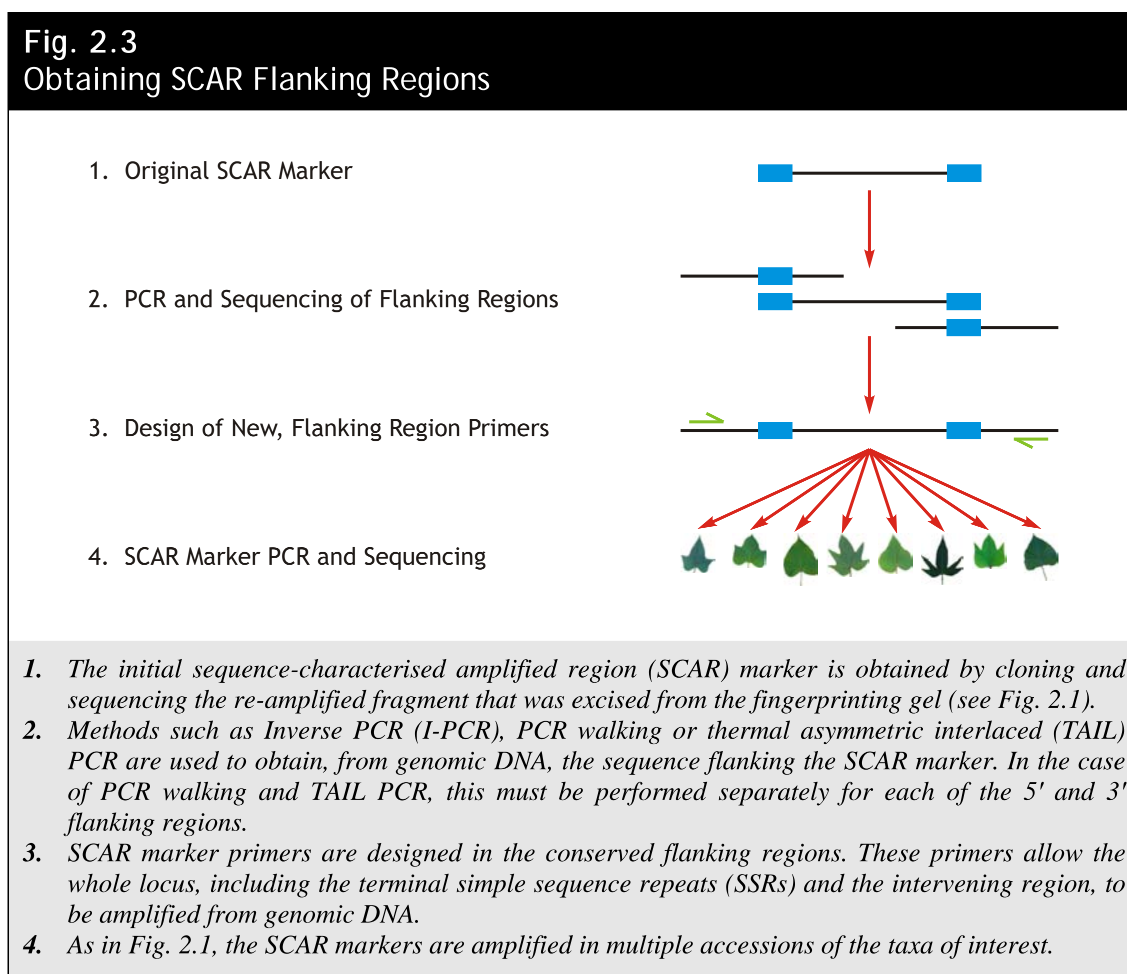


Fig. 2.3
Obtaining SCAR Flanking Regions

1. *The initial sequence-characterised amplified region (SCAR) marker is obtained by cloning and sequencing the re-amplified fragment that was excised from the fingerprinting gel (see Fig. 2.1).*
2. *Methods such as Inverse PCR (I-PCR), PCR walking or thermal asymmetric interlaced (TAIL) PCR are used to obtain, from genomic DNA, the sequence flanking the SCAR marker. In the case of PCR walking and TAIL PCR, this must be performed separately for each of the 5′ and 3′ flanking regions.*
3. *SCAR marker primers are designed in the conserved flanking regions. These primers allow the whole locus, including the terminal simple sequence repeats (SSRs) and the intervening region, to be amplified from genomic DNA.*
4. *As in Fig. 2.1, the SCAR markers are amplified in multiple accessions of the taxa of interest.*

It was hypothesised that, for this project, using PCR to obtain the flanking regions would increase the efficiency of SCAR marker development by ensuring that polymorphisms causing the presence–absence phenotype (especially those in the MLF primer-binding regions) would be captured in the SCAR markers. It was also hoped

that, by obtaining the flanking regions, additional polymorphisms in the regions flanking the ISSR locus might also be 'captured' (microsatellites have been found to cluster together (van der Nest *et al.*, 2000; Yang *et al.*, 2001; see also Mogg *et al.*, 2002)).

Techniques for obtaining uncharacterised flanking regions are increasingly being used to improve the success of SCAR marker development, allowing informative, polymorphic markers to be obtained from previously uninformative, monomorphic SCARs. For example, Brugmans *et al.* (2003) developed SCAR markers for 10 polymorphisms (based on the presence–absence phenotype) from AFLP fragments. All 10 markers produced monomorphic (by length) PCR products in all taxa. However, seven of these markers were "rescued" by the discovery of internal SNPs. The remaining three markers were rescued by using PCR-walking to obtain the flanking regions. This revealed SNPs responsible for the original presence–absence phenotype, and these were used to develop CAPS markers. Although techniques for obtaining flanking regions are increasingly being used to improve the success of SCAR marker development for plant breeding applications (e.g., De Jong *et al.*, 1997; Bradeen & Simon, 1998; Negi *et al.*, 2000; Li & Garvin, 2003; Hong *et al.*, 2006), their application in evolutionary biology is limited (although see Bensch *et al.*, 2002).

Several techniques are available for amplifying flanking regions, including Inverse PCR (I-PCR) (Ochman *et al.*, 1988), targeted gene walking (Parker *et al.*, 1991), PCR walking (Siebert *et al.*, 1995), ligation-mediated suppression (LMS) PCR (Schupp *et al.*, 1999) and TAIL PCR (Liu & Whittier, 1995). Although these techniques have all been used widely, many have serious drawbacks when applied to plant genomes which are typically large and repetitive. These drawbacks include the need for a Southern hybridisation step (I-PCR), genomic digestion with potentially finicky restriction enzymes (I-PCR), chimeric artefacts (I-PCR) and extensive non-specific amplification (targeted gene walking) (Liu & Whittier, 1995; Bradeen & Simon, 1998).

2.3.4 USING TAIL PCR TO IMPROVE SCAR MARKER DEVELOPMENT

2.3.4.1 Thermal Asymmetric Interlaced (TAIL) PCR: Introduction

TAIL PCR (Liu & Whittier, 1995) is a flexible, robust molecular technique for amplifying, often from genomic DNA, a region of unknown DNA sequence flanking a region of known sequence. Although its use as a method for obtaining flanking sequences in SCAR marker development appears limited (but see Xu *et al.*, 2001; Sakata *et al.*, 2006), it may be a more robust alternative — especially in plants — to the alternative techniques listed above. TAIL PCR has been used primarily for isolating promoter sequences (e.g., Terauchi & Kahl, 2000) and sequences flanking T-DNA and transposon insertion sites (e.g., Liu *et al.*, 1995; Tsugeki *et al.*, 1996).

2.3.4.2 TAIL PCR: Methodology

The primers used for TAIL PCR are the specific (SP) primer and the arbitrary degenerate (AD) primer. The SP primer is designed from the region of known sequence, with the 3′ end of the primer nearest the flanking region to be amplified. The AD primer is a short, degenerate, universal primer that binds at an unknown position in the flanking region. PCR is then used to amplify the region between the SP and AD primers (Fig. 2.4).

The non-specific AD primer will bind at many sites in the genome in addition to the region immediately flanking the region of known sequence. The background amplification that occurs due to the AD primer binding at many loci is limited in two ways:

1. ***TAIL PCR is carried out asymmetrically***

   The main part of the PCR program comprises 15 ‘supercycles’, with each supercycle itself comprising two high-stringency cycles (annealing temp = 60°C) and one low-stringency cycle (annealing temperature = 44°C). The high-stringency cycles favour annealing and extension of the SP primer only (a ~20-mer with a T_m of 60–62°C. The high-stringency cycles allow the formation of only single-stranded products extending from the region of known sequence into

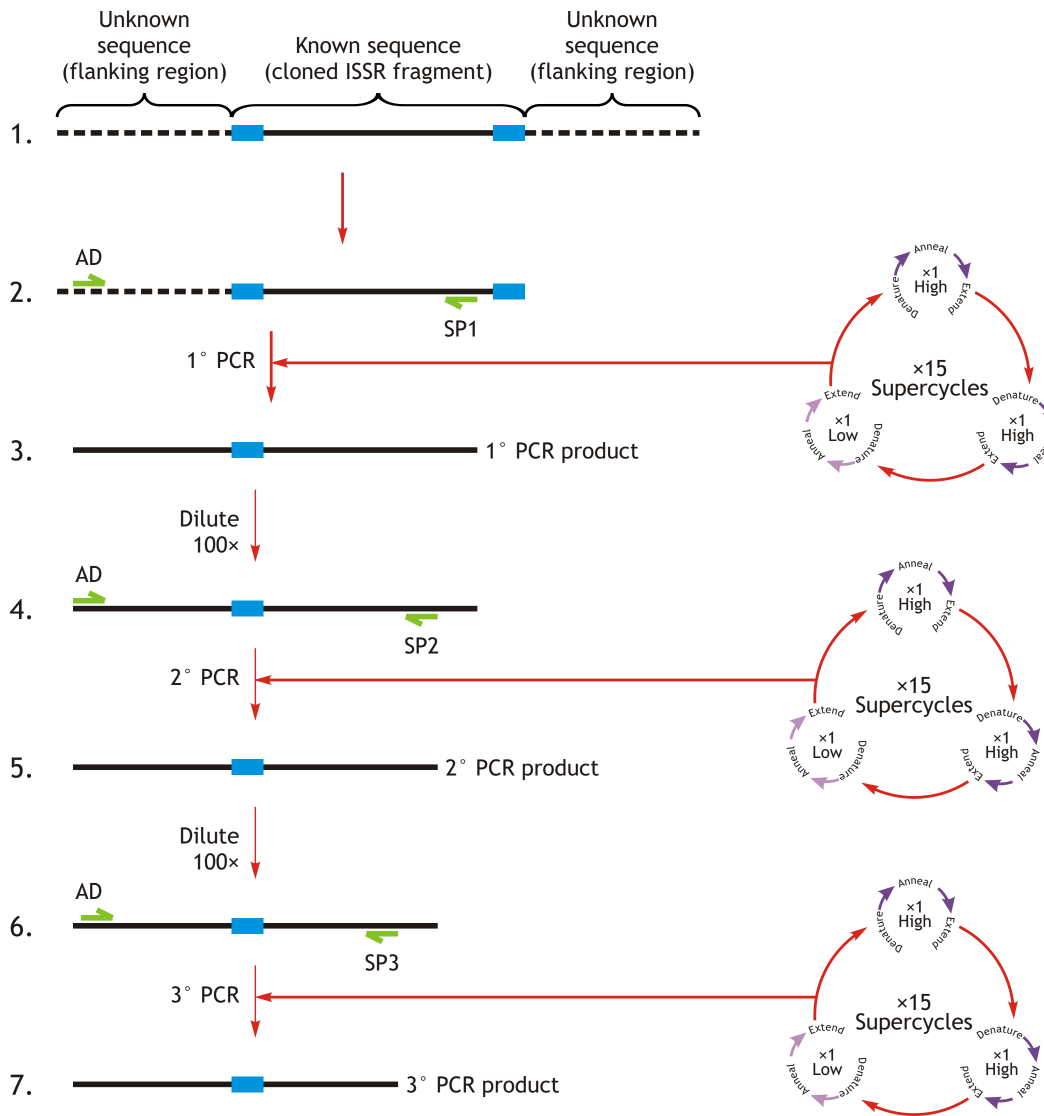
the flanking region. The low-stringency cycle allows annealing of the AD primer (a 16-mer with a T_m of ~45°C) to synthesise the strand complementary to that synthesised by the SP primer. This asymmetric amplification helps overcome the bias towards non-specific products that would result from the AD primer binding at multiple loci.

2. ***Three consecutive rounds of nested PCR are carried out***

   Three SP primers are used for TAIL PCR: SP1, SP2 and SP3. All are designed from the known sequence. SP1 is farthest from the flanking region, SP2 is closer to the flanking region and SP3 is closer still, allowing SP2 and SP3 to be used for nested PCR. The first round of PCR (the primary (1°) PCR) is carried out using genomic DNA with the SP1 and AD primer. Even employing the asymmetric technique described above, the 1° PCR produces a large number of non-specific products in addition to the SP-primed product. The second round of PCR (the secondary (2°) PCR) is carried out using product from the 1° PCR as template, the nested SP2 primer, and the AD primer. The 2° PCR reduces the number of unwanted products further. The third round of PCR (the tertiary (3°) PCR) is carried out using product from the 2° PCR as template, the nested SP3 primer, and the AD primer. Ideally, the 3° PCR will produce only one product — generated by the SP3 and AD primers and straddling the known and unknown (flanking) region.

1°, 2° and 3° rounds of PCR are electrophoresed in adjacent lanes of an agarose gel, producing the typical step-down pattern of bands shown in Fig. 2.5.

Fig. 2.4
Overview of TAIL PCR Technique



1. *The cloned and sequenced inter-SSR (ISSR) region is flanked by unknown sequence in the genomic DNA.*

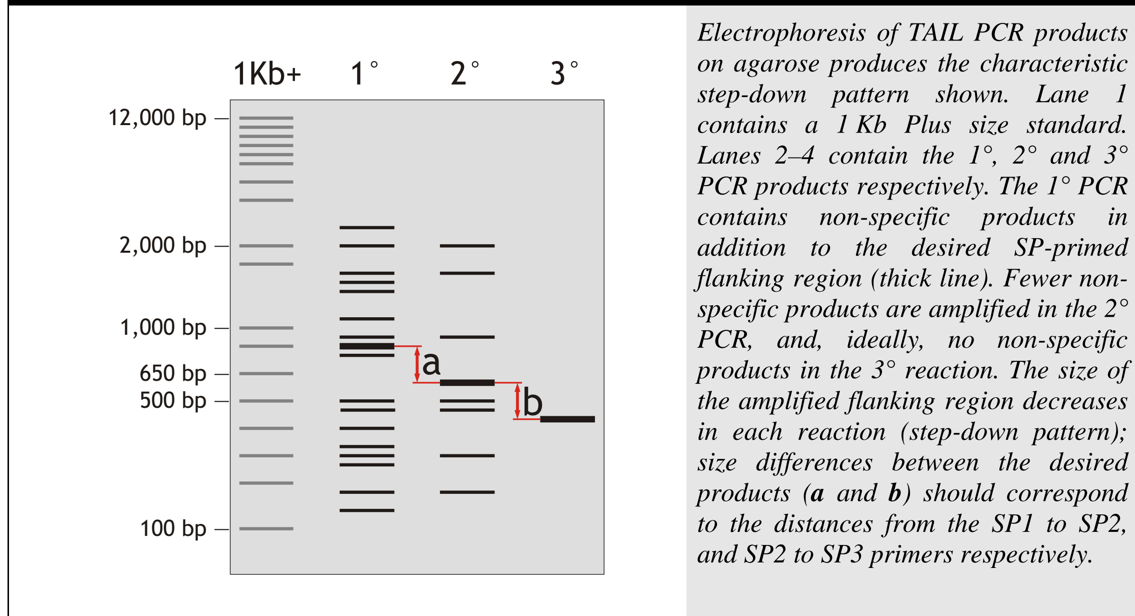
2. *An arbitrary degenerate (AD) primer, which binds at an unknown location in the flanking region, and a specific (SP1) primer, which is designed from the sequenced ISSR, are used to amplify the flanking region. This is achieved using thermal asymmetric interlaced (TAIL) PCR, which alternates two high-stringency cycles and one low-stringency cycle in 15 supercycles (see main text).*

3. *The SP1-AD primed product (1° PCR product) is produced, in addition to a number of non-specific AD-AD products (not shown; see Fig. 2.5).*

4.–5. *The 1° PCR product is diluted and nested TAIL PCR is performed using the SP2 and AD primers.*

6.–7. *This process is repeated, using the SP3 primer and the diluted 2° PCR product as template in the 3° TAIL PCR. The 3° TAIL PCR should produce a single 3° PCR product covering the flanking region, which can be sequenced using the SP3 primer. To amplify the right-hand (3') flanking region, this process is repeated using a new set of SP primers.*

Fig. 2.5
Theoretical Results of TAIL PCR Electrophoresis



*Electrophoresis of TAIL PCR products on agarose produces the characteristic step-down pattern shown. Lane 1 contains a 1 Kb Plus size standard. Lanes 2–4 contain the 1°, 2° and 3° PCR products respectively. The 1° PCR contains non-specific products in addition to the desired SP-primed flanking region (thick line). Fewer non-specific products are amplified in the 2° PCR, and, ideally, no non-specific products in the 3° reaction. The size of the amplified flanking region decreases in each reaction (step-down pattern); size differences between the desired products (**a** and **b**) should correspond to the distances from the SP1 to SP2, and SP2 to SP3 primers respectively.*

For the development of SCAR markers from ISSRs, both 5′ and 3′ flanking regions are amplified by performing TAIL PCR at each end of the fragment.

2.3.5 OVERVIEW OF ISSR-DERIVED SCAR MARKER DEVELOPMENT

ISSR-derived SCAR markers were developed using the procedure outlined in Fig. 2.6.

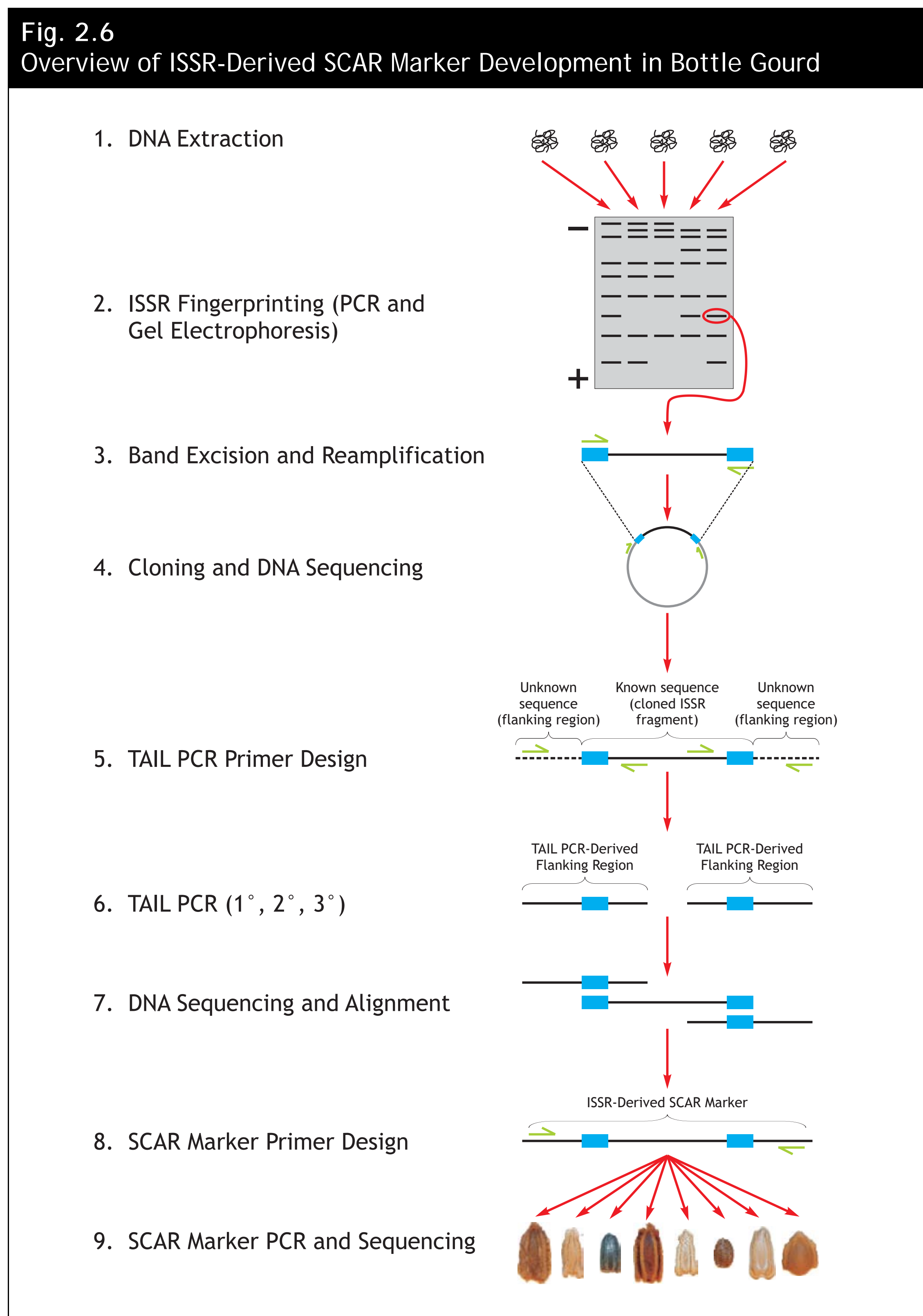


Fig. 2.6
Overview of ISSR-Derived SCAR Marker Development in Bottle Gourd

1. ***DNA Extraction***
   Genomic DNA was extracted from a small number of accessions (5) that represented the most diverse taxa available. Replicate extractions were performed from each individual.

2. ***ISSR Fingerprinting***
   PCR was used to amplify inter-SSRs (ISSRs) from genomic DNA (see Fig. 2.2). Multi-locus ISSR PCRs were separated on a denaturing polyacrylamide gel and visualised by silver-staining.

3. ***Band Excision and Re-Amplification***
   ISSR bands present in one accession, and not in another, represent polymorphic loci and were excised from the acrylamide gel. The DNA from these bands was re-amplified by PCR using the ISSR primer used to generate the band in the first instance.

4. ***Cloning and DNA Sequencing***
   Each re-amplified ISSR fragment was cloned into a vector and sequenced.

5. ***TAIL PCR Primer Design***
   To allow amplification of the regions flanking the ISSR, thermal asymmetric interlaced (TAIL) PCR primers were designed from the sequenced ISSR locus.

6. ***TAIL PCR (1•, 2•, 3•)***
   The 5' and 3' flanking regions of the ISSR locus were amplified from genomic DNA using TAIL PCR.

7. ***DNA Sequencing and Alignment***
   The amplified flanking regions were sequenced and aligned with the original cloned ISSR sequence.

8. ***SCAR Marker Primer Design***
   SCAR marker primers were designed in the conserved flanking regions. These primers allow the whole locus, including the terminal SSRs and the intervening region, to be amplified from genomic DNA

9. ***SCAR Marker PCR and DNA Sequencing***
   The ISSR-derived SCAR markers were amplified and sequenced in all accessions of bottle gourd.

2.3.6 AIMS

The aim of the research is to develop a set of codominant ISSR-derived SCAR markers in bottle gourd that are of sufficient resolution to allow the origins of the Polynesian bottle gourd to be determined.

2.4 DEVELOPMENT OF BOTTLE GOURD SCAR MARKERS: METHODS

2.4.1 PRELIMINARY TECHNICAL REMARKS

1. For obscure items, the supplier is given in parentheses the first time the item is mentioned. Suppliers are not given for common chemicals (e.g., NaCl).
2. The type of thermal cycler used for each different PCR is not listed, as PCR was performed in a range of thermal cyclers (i.e., T1 (Biometra), TGradient (Biometra), PCR Express (Hybaid) and PTC-200 DNA Engine (MJ Research)) and the results should be replicable in any of these instruments (and others) if the parameters described in the methods section are met (e.g., ramping speed).
3. PCR primers and other oligonucleotides were supplied desalted and lyophilised (from Invitrogen or Sigma–Aldrich) and were diluted to a stock concentration of 1 mM (1 nmol μL^{-1}) with Milli-Q H_2O and were stored at −80°C. Stock solutions were typically diluted 100× to a working solution of 10 µM (10 pmol μL^{-1}) with Milli-Q H_2O and were stored at −20°C.
4. All solutions were made with Milli-Q H_2O unless otherwise stated. If available, all reagents were molecular biology grade, otherwise they were an appropriate alternative grade.

2.4.2 ISSR FINGERPRINTING AND ISOLATION OF POLYMORPHIC ISSRS

2.4.2.1 ISSR Fingerprinting

2.4.2.1.1 Materials and DNA Extraction

For the ISSR PCR, a core set of five accessions was assembled, composed of one individual each of the following accessions: ‘183’, ‘195’, ‘GD’, ‘MR’ and ‘BR’ (see Table 3.2 for an explanation of the accession codes). These five accessions were chosen

for ISSR fingerprinting before the Heiser accessions (Table 3.2) were available, otherwise some of the Heiser material would have been included in the core set.

DNA was extracted, quantified and diluted as described in Section 3.3.4. To ensure reproducibility of the ISSRs, replicate DNA extractions were performed for each individual, giving a core set of 10 DNA samples for ISSR. Only DNA which appeared on an agarose gel as a single, intact band of high molecular weight (> 12 kb) was used for ISSR; degraded DNA reduced the reproducibility of the ISSR fingerprint profiles.

2.4.2.1.2 ISSR PCR

ISSR PCR reactions were performed in 96-well semi-skirted PCR plates (ABgene). The 12 templates (10 DNA samples and two negative controls) were coloured with ~1 mM cresol red · sodium salt (Sigma–Aldrich). Cresol red does not interfere with PCR and is extremely useful as a tracking dye for high-throughput applications (Hoppe *et al.*, 1992). In total, the 12 templates were amplified with 16 different ISSR primers (Table 2.1). Each ISSR PCR reaction consisted of 1× PCR buffer (Roche; 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 1.25 mM additional $MgCl_2$ (for a total $MgCl_2$ concentration of 2.75 mM), 250 μM of each dNTP, 1 M betaine (Sigma–Aldrich; see below), 0.75 μM ISSR primer (UBC; see Table 2.1 for primer sequences), 1 U *Taq* DNA polymerase (Roche) and ~1 ng of genomic DNA in a total volume of 20 μL. ISSR PCR was carried out in a thermal cycler using the following touchdown program: 94°C for 1 min; 7 cycles of 94°C for 1 min, 55°C for 1 min (thereafter decreasing 1°C/cycle), 72°C for 1 min; 35 cycles of 94°C for 1 min, 48°C for 1 min, 72°C for 1 min; 72°C for 5 min; hold at 10°C. Ramping speed was set at 1°C s^{-1}. The touchdown protocol improved reproducibility of the results by creating stringent primer-binding conditions. It also allowed reactions containing ISSR primers with different annealing temperatures to be amplified in the same 96-well plate. ISSR products were stored at 4°C overnight (and never longer), before being subjected to polyacrylamide gel electrophoresis (PAGE).

Table 2.1
ISSR Primers Used from UBC Primer Set #9

96-Well Plate	Row	UBC Primer #[a]	Sequence (5′–3′)[b]	T_m (°C)[c]
One	**A**	808	$(AG)_8C$	52
	B	811	$(GA)_8C$	52
	C	812	$(GA)_8A$	50
	D	813	$(CT)_8T$	50
	E	814	$(CT)_8A$	50
	F	816	$(CA)_8T$	50
	G	817	$(CA)_8A$	50
	H	822	$(TC)_8A$	50
Two	**A**	823	$(TC)_8C$	52
	B	843	$(CT)_8RA$	52–54
	C	846	$(CA)_8RT$	52–54
	D	852	$(TC)_8RA$	52–54
	E	853	$(TC)_8RT$	52–54
	F	881	$(GGGTG)_3$	54
	G	884	$HBH(AG)_7$	48–54
	H	891	$HVH(TG)_7$	48–54

[a] All ISSR primers were from the UBC Primer Set #9 (University of British Columbia, BC, Canada).
[b] H = A, C and T; B = C, G and T; V = A, C and G; R = A and G.
[c] The T_m (melting temperature) is defined as the temperature at which 50% of the helical structure of DNA is lost (Stryer, 1995, pp. 84, 86) and is calculated (approximately) using the following formula:
$T_m = 4°C(N_G + N_C) + 2°C(N_A + N_T)$

2.4.2.1.3 Betaine as an Enhancing Agent for PCR

Betaine (*N,N,N*-trimethylglycine monohydrate), at a final concentration of 1 M, was routinely added to PCR reactions as an enhancing agent. Its use increased product yield and improved amplification of difficult templates. Betaine acts by reducing the formation of secondary structures caused by GC-rich regions (Rees *et al.*, 1993; Henke *et al.*, 1997; Frackman *et al.*, 1998), increasing the optimal range of $MgCl_2$ concentration in the PCR (Weissensteiner & Lanchbury, 1996), and possibly increasing the thermal stability of *Taq* DNA polymerase (Weissensteiner & Lanchbury, 1996;

Hengen, 1997). The Q-Solution available from Qiagen has been shown to contain betaine (Frackman *et al.*, 1998).

2.4.2.1.4 Agarose Gel Electrophoresis of ISSR PCR Products

To verify successful amplification of the ISSRs in bottle gourd, an aliquot of each ISSR PCR reaction was electrophoresed on an agarose gel prior to PAGE. A 3 µL aliquot of each ISSR PCR reaction was combined with 1 µL 10× loading buffer (25% (w/v) Ficoll 70 (Pharmacia), 0.16% (w/v) bromophenol blue · sodium salt (3′-3″-5′-5″-tetrabromophenol sulfonphthalein; Serva) and 0.16% (w/v) xylene cyanol FF (Sigma–Aldrich)) in a total volume of 10 µL. Each 10 µL aliquot was loaded onto a 2.0% (w/v) agarose/1× Tris acetate EDTA (TAE) buffer gel. The 1× TAE buffer was 40 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) (pH 8.0), 20 mM (0.115% (v/v)) acetic acid and 1 mM ethylenediaminetetraacetic acid · disodium salt (EDTA) (pH 8.0). A 1 Kb Plus DNA Ladder™ (Invitrogen) was loaded as a size standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. For agarose gels run using the large, multi-channel pipette compatible, 102-well Sub-Cell® Model 96 Cell (Bio-Rad) electrophoresis apparatus, gels contained 0.3 µg mL^{-1} ethidium bromide and the running buffer contained 1 µg mL^{-1} ethidium bromide. All other agarose gels, which were much smaller and easier to handle, were stained post-electrophoresis in a 1 µg mL^{-1} ethidium bromide solution for 15 min and then rinsed with Milli-Q H_2O for 15 s. Gels were visualised by ethidium bromide fluorescence on a Gel Doc 2000™ UV transilluminator (λ = 302 nm; Bio-Rad) and photographed using the transilluminator's built-in CCD camera and Quantity One v. 4.4.0 software (Bio-Rad). Successful amplifications, which appeared on the agarose gel as a smear of DNA < 1.5 kb with some distinct bands, were subjected to PAGE.

2.4.2.2 Polyacrylamide Gel Electrophoresis (PAGE) of ISSR PCR Products

ISSR polyacrylamide gels were electrophoresed and silver-stained using a protocol modified from Caetano-Anollés & Gresshoff (1994) and Promega (1998).

2.4.2.2.1 Preparation of Polyacrylamide Electrophoretic Glass Plates

In preparation for pouring the polyacrylamide gel, one side of each of the long and the short plates was cleaned thoroughly. The cleaning process consisted of scrubbing the plates with detergent under hot running water, followed by rinsing the plates to remove any trace of detergent. The plates were then dried with a paper towel and cleaned with 70% (v/v) ethanol, which was also wiped off using a paper towel. This process was repeated twice more for each plate so that each had been cleaned three times with detergent and three times with ethanol. After the final ethanol wash the plates were again cleaned with ethanol but this time were wiped dry with a Kimwipe® (Kimberly–Clark) to prevent paper towel fibres being incorporated into the gel.

The cleaned face of the short plate was then treated with a Bind–Silane solution to promote adherence of the gel to this plate following electrophoresis. The Bind–Silane solution consisted of 2 mL 100% ethanol, 0.5% (v/v) (87 mM) acetic acid and 0.05% (v/v) PlusOne™ Bind–Silane (γ-methacryloxy-propyl-trimethoxysilane; Pharmacia) prepared in a 2 mL microtube. The solution was applied to the plate using a Kimwipe and left to dry for ~3 min. Excess Bind–Silane was removed from the plate by cleaning the plate three times with 3 mL 100% ethanol and wiping with a Kimwipe after each application.

Approximately 2 mL Windshield Rain Repellent (Prestone) was applied to the cleaned face of the long plate (according to the manufacturer's instructions) using a Kimwipe. This was to prevent the adherence of the gel to this plate following electrophoresis. The plate was then cleaned with 3 mL 100% ethanol, again using a Kimwipe.

The plates were immediately assembled, treated faces facing, using 0.4 mm spacers (Gibco BRL) to form a glass sandwich. It was imperative at this stage that the faces of the long and short plates did not touch, as the resulting interaction inevitably caused the gel to adhere to the long plate (in addition to the short plate). The glass sandwich was then placed in a S2 casting clamp (Gibco BRL) to hold the plates securely.

2.4.2.2.2 Preparation of Acrylamide Gel Solution

A 5% (v/v) acrylamide gel solution was prepared (7 M urea, 5% (v/v) 19:1 acrylamide/bis-acrylamide (Bio-Rad) and 1× Tris borate EDTA (TBE) buffer in a total volume of 70 mL). The 1× TBE buffer was 89 mM Tris-HCl (pH 8.0), 89 mM boric acid and 2 mM EDTA (pH 8.0). The solution was filtered through two No. 1, ∅ 125 mm filter circles (Whatman). Immediately prior to pouring the gel, 350 µL 10% (v/v) ammonium persulfate (APS; Bio-Rad) and 35 µL *NNN'N'*-tetramethylethylenediamine (TEMED; BDH) were added to the filtrate. The acrylamide solution was mixed by gentle agitation and immediately drawn into a 50 mL syringe (from which the needle had been removed).

The acrylamide solution was injected into the glass sandwich via the hole in the casting clamp. A pair of 5.7 mm point-to-point sharkstooth combs (Gibco BRL) were inserted, in reverse orientation, to form the gel wells. The combs were clamped in place with four large bulldog clips. The top of the gel was sealed with cling film to prevent dehydration, and the gel laid at an angle of 5° (using the feet of the casting clamp). The gel was left overnight to polymerise.

2.4.2.2.3 Setup of the Electrophoresis Apparatus

Once polymerised, the gel was prepared for electrophoresis. The cling film, bulldog clips, casting clamp, and combs were removed from the gel sandwich. The outside of the glass plates were cleaned with 70% (v/v) ethanol to remove any excess buffer and polyacrylamide. The sandwich was loaded into a model S2 electrophoresis apparatus

(Gibco BRL) and the upper and lower buffer tanks filled with 1× TBE buffer. A plastic Pasteur pipette was used to flush excess urea, polyacrylamide fragments and any small air bubbles from the sample wells. The gel was then pre-run at 70 W for 1 h.

2.4.2.2.4 Preparation of DNA Size Ladder

A 100 bp DNA Ladder (Invitrogen) was used as a size standard for the ISSR PAGE gels. This was prepared in a 200 µL aliquot by combining 198µL of formamide loading dye and 2 µL 100 bp DNA Ladder (1 µg μL^{-1}). The formamide loading dye was 98% (v/v) formamide, 10 mM EDTA (pH 8.0), 0.05% (w/v) bromophenol blue and 0.05% (w/v) xylene cyanol. Using a thermal cycler, the solution was denatured at 95°C for 4 min before being rapidly cooled to 4°C. The ladder was stored at −20°C until required.

2.4.2.2.5 Preparation of ISSR PCR Products for PAGE

A 3 µL aliquot of each ISSR PCR product was combined with 4 µL formamide loading dye. Using a thermal cycler, the samples were then irreversibly denatured at 95°C for 4 min before being rapidly cooled to 4°C.

2.4.2.2.6 Loading and Running of ISSR PAGE Gels

After pre-running the gel, the wells were flushed again (as described above). The combs were inserted into the wells so the teeth rested on the gel surface and the wells flushed a final time. A 6 µL aliquot of each prepared PCR product was loaded onto the gel. A 6 µL aliquot of the 100 bp DNA Ladder solution was loaded approximately every 10–12 lanes as a size standard. Replicate samples from the same individual were run in adjacent lanes to allow reproducibility to be checked easily. Negative control ISSR PCR reactions were also included. ISSR PAGE gels were run at 70 W until the xylene cyanol dye band was approximately 5 cm from the base of the gel (usually about 2 h).

2.4.2.2.7 Detection of DNA Fragments by Silver Staining of the ISSR Gels

At the commencement of electrophoresis, the developing solution (4 L of 280 mM anhydrous sodium carbonate (Na_2CO_3; Merck) in Milli-Q H_2O) was made and placed at −20°C to chill. 2L of Milli-Q H_2O was also placed at −20°C.

At the completion of electrophoresis, the gel sandwich was removed from the apparatus and the spacers removed from between the plates. A scalpel blade was then inserted between the plates to separate them. The short plate, to which the gel had adhered, was transferred (gel side uppermost) to a plastic tray containing 4 L of 10% (v/v) (1.74 M) acetic acid (made using Milli-Q H_2O) and agitated on an orbital shaker at 30–40 revolutions per minute (rpm) for 1 h to remove urea from the gel. After this wash, 2 L of the acetic acid solution was retained and chilled at −20°C for later use. The gel was then placed in a plastic tray containing 2 L of Milli-Q H_2O and shaken for approximately 3 min. This rinsing step was repeated twice more, each time transferring the gel to a tray containing 2 L of fresh Milli-Q H_2O. After the three rinse steps, which had removed acetic acid from the gel, the gel was placed in a plastic tray containing 3 L of stain solution (6 mM silver nitrate ($AgNO_3$, Merck) and 0.15% (v/v) formalin (37% (w/w) formaldehyde (H_2CO); BDH) in Milli-Q H_2O) and placed on the orbital shaker for 45 min.

Formalin and sodium thiosulfate ($Na_2S_2O_3$; Sigma–Aldrich) were then added to the 4 L of pre-chilled developing solution to a final concentration of 0.16% (v/v) and 12.65 µM respectively.

The gel was removed from the stain solution and placed in the 2 L of pre-chilled Milli-Q H_2O for 20 s to cool the gel and remove unbound silver. The Milli-Q H_2O was retained and kept chilled, but the gel was immediately transferred to a developing tray containing 2 L of the developing solution. The tray was agitated vigorously by hand to disperse the brown precipitate that formed, before being placed on the orbital shaker. As soon as DNA bands became visible, the gel was transferred to the remaining 2 L of developing solution and placed back on the shaker. The gel was then developed until the bands reached the desired intensity but before the background became overly dark. The developing reaction was stopped, and the gel fixed, by the addition of 2 L of the chilled

acetic acid. This caused the developing solution/acetic acid mixture to effervesce and when this ceased, the gel was deemed to have been properly fixed (*ca* 10 min). The gel was transferred back to the 2 L of chilled Milli-Q H_2O and placed on the orbital shaker for about 5 min. This final step removed excess acetic acid from the gel.

2.4.2.3 Isolation and Re-Amplification of Polymorphic ISSRs

2.4.2.3.1 Excision of Polymorphic Bands From the ISSR PAGE Gels

The polyacrylamide gel was transferred to a fluorescent light box. Bands that were present in one or more (but not all) of the five accessions in each primer were hypothesised to represent a polymorphism, but unless a band was present in both replicate lanes of an accession it was ignored. Polymorphic bands that were between 200 bp and 1 kb in size were excised from the gel using a sterile scalpel blade. Bands were excised across both replicate lanes of a single accession. Each gel fragment was numbered and the primer from which it came, along with the accession and its estimated size (from the size standard) were recorded. The excised gel fragments were incubated in 30 μL H_2O at 4°C overnight (~16 h) to allow diffusion of the DNA from the polyacrylamide gel. Preliminary work had shown diffusion periods shorter than 16 h were insufficient for adequate diffusion; periods longer than 16 h caused inhibitors (e.g., formamide) to diffuse from the fragment and prevent successful re-amplification (results not shown).

2.4.2.3.2 Re-Amplification from Excised Band Eluate

DNA eluted from the polyacrylamide gel cuts was re-amplified using PCR. The same PCR primer used for the original ISSR PCR (i.e., from the UBC primer set) was used for the re-amplification PCR. Each re-amplification PCR reaction consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 250 μM of each dNTP, 1 M betaine, 0.75 μM ISSR primer (UBC; see Table 2.1 for primer sequences), 1 U *Taq* DNA polymerase and 1 μL of gel cut eluate in a total volume of 20 μL. PCR was carried out in a thermal cycler as

follows: 94°C for 2 min; 35 cycles of 94°C for 1 min, 48°C for 1 min, 72°C for 1 min; 72°C for 5 min; hold at 10°C. Ramping speed was set at 1°C s^{-1}.

2.4.2.3.3 Agarose Gel Extraction of Re-Amplification PCR Products

Success of the re-amplification PCR reaction was determined by electrophoresis. The entire 20 µL of each ISSR PCR reaction was combined with 2 µL 10× loading buffer, and loaded onto a 1.5% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A 1 Kb Plus DNA Ladder was loaded as a size standard, and a Low DNA Mass™ Ladder (Invitrogen) was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4. Successful re-amplifications appeared on the agarose gel as a discrete band of the expected size of the fragment as estimated from the ISSR PAGE gel (see Section 2.4.2.3.1).

PCR products that were successfully re-amplified were excised from the agarose gels using a sterile scalpel blade. DNA was extracted from the gel slice using the QIAquick® Gel Extraction Kit (Qiagen), according to the manufacturer's protocol. DNA was eluted in 30 µL of the provided EB (elution) buffer (10 mM Tris-HCl (pH 8.5)).

2.4.2.3.4 NanoDrop® Spectrophotometry

The yield of the agarose gel extractions was determined using a NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies). A 2 µL aliquot of gel extraction eluate was placed on the NanoDrop pedestal and analysed according to the manufacturer's protocol. Clean DNA samples gave a 260/280 nm absorbance ratio of ~1.8. Gel-extracted DNA was stored at −20°C until required.

2.4.3 CONVERSION OF ISSRS INTO SCAR MARKERS

2.4.3.1 Cloning and DNA Sequencing of ISSRs

PCR products can normally be sequenced directly with either of the primers used to generate the PCR product, or alternatively, with internal primers if there is prior sequence knowledge. Because ISSR PCR products are generated with only one primer, they cannot be sequenced directly; if the ISSR primer were used as the sequencing primer the sequencing reaction would proceed from both ends of the PCR product, resulting in a mixed signal. Therefore, ISSR PCR products need to be cloned, and then sequenced with a vector primer.

2.4.3.1.1 Overview of the TA Cloning and Blue-White Screening Techniques

Taq DNA polymerase, due to a lack of 3′–5′ exonuclease activity, adds a single deoxyadenosine to the 3′ end of amplified fragments, resulting in PCR products possessing an A overhang at the 3′ ends of both strands (Clark, 1988). This feature is exploited in TA cloning. The vector, in this case pGEM®-T Easy, is supplied as a linear plasmid with a 3′ terminal thymidine added to both ends (Promega Corp., 2003). These 3′-T overhangs are compatible with the 3′-A overhangs on the PCR product and can be employed for sticky-end ligation of the insert into the vector.

The ligated vector is used to transform *Escherichia coli* strain DH5α™ cells. Transformation reactions are plated onto media containing ampicillin, X-Gal and IPTG. pGEM-T Easy contains the ampicillin resistance (β-lactamase) gene so that only cells transformed with the vector will be able to grow on the ampicillin-containing media. pGEM-T Easy also possesses the *lacZ* gene (induced by IPTG), which encodes the enzyme β-galactosidase. β-galactosidase is able to cleave the chromogenic substrate X-Gal, producing a blue dye and turning cells visibly blue. Because the *lacZ* gene straddles the multiple cloning region of pGEM-T Easy, cells with a vector containing PCR product inserts will possess an insertionally inactivated *lacZ* gene, produce no β-

galactosidase, and will be coloured white. It is the insert-containing white colonies which are picked (for downstream analyses).

2.4.3.1.2 Ligation of Re-Amplified ISSRs into pGEM®-T Easy

All cloning and bacterial transformation was carried out with the appropriate approval from the Environmental Risk Management Authority (ERMA), New Zealand (ERMA Approval Code GMD002681). See Appendix 11 for a copy of the application.

Re-amplified ISSR PCR products were ligated into the TA cloning vector pGEM-T Easy using a $mol_{insert} : mol_{vector}$ ratio of 3:1. The amount of insert was calculated using the formula:

$$ng_{insert} = \frac{ng_{vector} \times bp_{insert}}{bp_{vector}} \times (\text{insert : vector molar ratio})$$

where:

ng_{vector} = 50 ng,
bp_{vector} = 3015 bp (pGEM-T Easy),
insert : vector molar ratio = 3,

which can be simplified to:

$$ng_{insert} = \frac{150 \times bp_{insert}}{3015}.$$

Ligation reactions consisted of 1× Rapid Ligation Buffer (Promega; 30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 1 mM adenosine 5′-triphosphate (ATP), 5% (v/v) polyethylene glycol (PEG)), 50 ng pGEM-T Easy Vector (Promega), 3 Weiss U T4 DNA ligase (Promega) and gel-extracted PCR product (amount calculated as above) in a total volume of 10 μL. A positive control ligation reaction was prepared, consisting of the above reagents but including 2 μL of Control Insert DNA (542 bp) (Promega) in place of PCR product. A negative control ligation reaction was prepared, consisting of the above reagents but no insert DNA. Ligation reactions were incubated either at room temperature (20–25°C) for 1 h, or alternatively, at 4°C overnight (~16 h).

2.4.3.1.3 Transformation of Escherichia coli *DH5α™*

Ligation reactions were diluted 5× with 40 μL $T_{10}E_1$ (10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)).

Ligated vectors were transformed into MAX Efficiency® DH5α™ Chemically Competent Cells (*Escherichia coli* strain DH5α, genotype F^- ϕ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk^-, mk^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1) (Invitrogen). Vials of *E. coli* DH5α were thawed on wet ice. A 30 μL aliquot of cells was aliquoted into a cold 10 mL Kimax® culture tube (Kimble). This was done using a pipette with a tip from which the lower 5 mm had been removed using a sterile scalpel blade (the wider bore tip prevents mechanical lysing of the cells). A 10 μL aliquot of the diluted ligation reaction was added to the tube of cells and stirred gently using a pipette tip. A positive transformation control reaction was prepared with 1.5 μL (1.5 pg) pUC19 plasmid (Invitrogen). Transformation tubes were incubated on ice for 15 min, before being heat-shocked at 42°C for 45 s. Tubes were incubated on ice for 2 min, before the addition of 270 μL SOC medium (Invitrogen; 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 20 mM glucose, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 2.5 mM KCl). Cells were then incubated at 37°C at 225 rpm for 1 h.

After incubation with SOC medium, a sterilised cell spreader was used to plate 150 μL of each transformation (20 μL for the pUC19 positive transformation control) onto lysogeny broth/Luria–Bertani (LB) agar plates (LB broth contained 85 mM NaCl, 1% (w/v) tryptone (Merck) and 0.5% (w/v) yeast extract (Merck), with 1.5% (w/v) bacteriological agar (Oxoid) added for LB agar). The LB agar plates contained 100 μg mL^{-1} ampicillin, 400 μg mL^{-1} isopropyl-β-D-thiogalactopyranoside (IPTG) and 40 μg mL^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal). The stock solutions of ampicillin, IPTG and X-Gal were 100 mg mL^{-1} in Milli-Q H_2O, 200 mg mL^{-1} in Milli-Q H_2O and 20 mg mL^{-1} in *N,N′*-dimethyl-formamide (DMF) respectively. All stock solutions of ampicillin, IPTG and X-Gal were stored at −20°C, with the X-Gal in a light-proof container. Plates were left to dry for 10 min at room temperature before being inverted and incubated at 37°C overnight.

2.4.3.1.4 Blue-White Screening, Colony PCR and Inoculation of LB Broth

Plates typically had ~100 colonies, 10–20% of which were white and putatively contained a vector with a PCR product insert. Approximately 6 white colonies were picked from each plate for colony PCR and LB broth culture. The colony PCR was performed on white colonies to verify they contain insert DNA, and that the insert was of the expected size. The primers for the colony PCR (referred to as universal primers) are located either side of the multiple cloning site. The 5′ end of the forward primer (M13F) binds at position #2961 of the vector (115 bp upstream of the insert site) and the 5′ end of the reverse primer (M13R) binds at position #211 (151 bp downstream of the insert site).

Each colony PCR reaction consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 250 μM of each dNTP, 0.5 μM M13F primer, 0.5 μM M13R primer (see Appendix 2 for primer sequences) and 1 U *Taq* DNA polymerase in a total volume of 20 μL. Template DNA was added by lightly touching the white colony with a sterile pipette tip and then briefly dipping the pipette tip into the PCR reaction, ensuring there was no carry-over of LB agar. The pipette tip was then ejected into a tube of LB broth as described below. Colony PCR was carried out in a thermal cycler as follows: 94°C for 2 min; 35 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 45 s; 72°C for 5 min; hold at 10°C.

A 5 μL aliquot of each colony PCR reaction was combined with 1 μL 10× loading buffer in a total volume of 10 μL, and loaded onto a 1.5% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A 1 Kb Plus DNA Ladder was loaded as a size standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4. Successful amplifications appeared on the agarose gel as a discrete band, the size of which corresponded to the size of the insert plus 266 bp (the distance of the universal primers from the insert site of the vector — see above).

Sterile 5 mL aliquots of LB broth containing 100 μg mL^{-1} ampicillin were inoculated with the pipette tip that had been used to provide template for the colony PCR (see above). Broth tubes were then incubated at 37°C at 225 rpm for 16 h.

2.4.3.1.5 Plasmid DNA Extraction and Digestion

The LB broths corresponding to successful colony PCR were used for plasmid DNA extraction. A 1.5 mL aliquot of each 5 mL broth was transferred to a 1.7 mL microtube and centrifuged at 4,000 × *g* for 5 min to pellet the cells. The supernatant was discarded. Plasmid DNA was extracted using the GenElute™ Plasmid Miniprep Kit (Sigma–Aldrich), according to the manufacturer's protocol. Step 7, which is not necessary for *end*A1 strains such as DH5α, was omitted. DNA was eluted in 100 µL of the provided elution buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and stored at 4°C until required.

To verify the plasmid contained the insert of the correct size, an aliquot of each plasmid extraction was subjected to restriction endonuclease digestion with *Eco*R I. pGEM-T Easy possesses two *Eco*R I sites — both in the multiple cloning region and either side of the insert site. Each restriction digest consisted of 1× SuRE/Cut Buffer H (Roche; 5 mM Tris-HCl, 10 mM NaCl, 1 mM $MgCl_2$, 100 µM dithioerythritol (DTE), pH 7.5), 10 U *Eco*R I (Roche) and 5 µL plasmid DNA in a total volume of 20 µL. Restriction digests were incubated at 37°C for 30 min.

A 15 µL aliquot of each digestion reaction and a 5 µL aliquot of one sample of undigested plasmid DNA was combined with 2 µL and 1 µL respectively of 10× loading buffer, and loaded onto a 1.5% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). The undigested sample was included as a negative restriction digest control. A 1 Kb Plus DNA Ladder was loaded as a size standard, and a Low DNA Mass™ Ladder was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4.

Recombinant plasmids appeared on the agarose gel as two discrete bands — one at 3 kb (which corresponds to the linear vector), and the other the size of the insert DNA. (Although not observed, if the insert DNA itself also contained one or more *Eco*R I sites, the combined size of all of the resulting fragments should equal the size of the insert DNA.) The undigested plasmid appeared on the agarose gel as an intense band at ~2.5 kb (the supercoiled form of the plasmid), with fainter bands at 3 kb (the relaxed

and linear forms of the plasmid). The concentration of the insert DNA was estimated by comparison with the Low DNA Mass Ladder.

2.4.3.1.6 DNA Sequencing of Insert DNA

DNA sequencing reactions were performed using the Sanger dideoxy chain termination method (Sanger *et al.*, 1977) with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). At least three clones of each PCR product were sequenced. Plasmid inserts were sequenced with the T7F primer, the 5′ end of which binds at position #2998 of the vector (78 bp upstream of the insert site). The amount of plasmid DNA template required for the sequencing reaction was calculated using the following formula (as recommended by Applied Biosystems):

$$\text{ng required} = \frac{\text{size of insert (bp)}}{40}$$

The relatively short inserts (all less than 1 kb) and the high-quality of the sequence data obtained from plasmids meant that it was necessary to sequence in one direction only. Each DNA sequencing reaction consisted of 0.75× BigDye Terminator v3.1 Sequencing Buffer (Applied Biosystems), 2.0 μL Ready Reaction Mix (BigDye v3.1) (Applied Biosystems), 160 nM T7F primer (see Appendix 2 for primer sequence) and plasmid DNA template (amount calculated as above) in a total volume of 20 μL. DNA sequencing reactions were carried out in a thermal cycler as follows: 27 cycles of 96°C for 10 s, 50°C for 5 s, 60°C for 4 min; hold at 10°C. Ramping speed was set at 1°C s^{-1}.

2.4.3.1.7 Purification of DNA Sequencing Products by Ethanol Precipitation

DNA sequencing products were purified by ethanol precipitation. The 20 μL sequencing reaction was transferred to a 1.7 mL microtube. To this was added 2 μL 125 mM EDTA (pH 8.0), 2 μL 3 M sodium acetate (pH 5.2) and 50 μL 100% ethanol, which was mixed by pipetting. The tube was then incubated at room temperature for approximately 15 min before being centrifuged at 20,000 × *g* at 4°C for 30 min. After centrifugation,

the supernatant was immediately removed using a pipette, taking care not to disturb the invisible pellet. A 150 µL aliquot of 70% (v/v) ethanol was then added. The tube was then centrifuged at 20,000 × *g* at 4°C for 10 min. After centrifugation the supernatant was immediately removed using a pipette, again taking care not to disturb the pellet. The pellet was dried either at 95°C for 5 min, or alternatively, at room temperature overnight (in the dark).

2.4.3.1.8 Capillary Electrophoresis of DNA Sequencing Reaction Products

DNA sequencing products were subjected to capillary electrophoresis (CE) at the Allan Wilson Centre Genome Service (AWCGS), Massey University, Palmerston North. The dried products were resuspended in 10 µL Hi-Di™ formamide (Applied Biosystems), and then irreversibly denatured by incubating at 95°C for 5 min before being rapidly cooled on wet ice. Samples were briefly centrifuged to remove any micro-bubbles. Samples were subjected to CE on a 3730 Genetic Analyzer (Applied Biosystems) using a 50 cm array, and the data collected using Run 3730 Data Collection v. 3.0 software (Applied Biosystems). Samples were analysed with Sequencing Analysis v. 5.2 software (Applied Biosystems) with the Z-BigDyeV3 dye set and LongSeq50_POP7_1 run module. The profiles were generated as electronic files (in *.ab1 format) using the Flat Profile method (i.e., peak height was normalised).

2.4.3.1.9 DNA Sequence Data Analysis

DNA sequence files (electropherograms in ABI format) were imported into Sequencher™ v. 4.2 (Gene Codes) software. The different clone sequences (at least 3) of each single-locus ISSR PCR product were aligned. Vector sequence was trimmed from both ends and uncalled or ambiguous bases edited where possible. A consensus sequence was created from the multiple clones of each ISSR PCR product.

2.4.3.2 TAIL PCR of ISSR Flanking Regions

2.4.3.2.1 Design of TAIL PCR Primers

Each single-locus ISSR consensus sequence was imported into Oligo v. 4.03 (National Biosciences) primer design software. All primers were designed within the following parameters where possible:

1. No primer dimers, or, if primer dimers are unavoidable, they have $\Delta G >$ (i.e., less negative than) −3.6 kcal mol^{-1}
2. No hairpin structures, or, if hairpin structures are unavoidable, they have a melting temperature lower than the melting temperature (T_m) of the primer
3. Length = 19–24-mer
4. T_m = 58–62°C

Where the above parameters allowed, 1°, 2° and 3° TAIL PCR primers were designed more than 20 bp apart. The separation of 20 bp between adjacent SP TAIL primers allowed the difference in size of the consecutive products to be resolved on an agarose gel — giving the distinctive step-down pattern in PCR product size shown in Fig. 2.5. The 3° TAIL PCR primer was designed at least 30 bp from the end of the sequence from which flanking DNA sequence was required; this 'buffer' provided sequence overlap between the 3° TAIL PCR product and the cloned ISSR sequence to ensure the correct TAIL sequence had been obtained.

2.4.3.2.2 1°, 2° and 3° TAIL PCR

Each 1° PCR reaction consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 1.25 mM additional $MgCl_2$ (for a total $MgCl_2$ concentration of 2.75 mM), 250 μM of each dNTP, 1 M betaine, 4 μM AD primer, 0.5 μM SP1 primer (see Appendix 2 for primer sequences), 1 U *Taq* DNA polymerase and ~1 ng of genomic DNA in a total volume of 20 μL. The genomic DNA used was the same as that used to produce the ISSR fingerprint from which the band was originally excised. 1° PCR was carried out in a thermal cycler as follows: 94°C for 2 min; 5 cycles of 94°C for 30 s, 60°C for 1 min, 72°C for 2 min; 94°C for 30 s; ramp from 25°C to 72°C over 3 min (Δ47°C =

+0.25°C s^{-1}); 72°C for 2 min; 10 cycles of 94°C for 30 s, 44°C for 1 min, 72°C for 2 min; 15 supercycles of 94°C for 30 s, 60°C for 1 min, 72°C for 2 min, 94°C for 30 s, 60°C for 1 min, 72°C for 2 min, 94°C for 30 s, 44°C for 1 min, 72°C for 2 min; 72°C for 5 min; hold at 10°C. Ramping speed was set at 1°C s^{-1} except where stated otherwise. Each 1° PCR product was diluted 100× by combining 2 µL PCR product with 198 µL Milli-Q H_2O.

Each 2° PCR reaction consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 1.25 mM additional $MgCl_2$ (for a total $MgCl_2$ concentration of 2.75 mM), 250 µM of each dNTP, 1 M betaine, 3 µM AD primer, 0.5 µM SP2 primer (see Appendix 2 for primer sequences), 1 U *Taq* DNA polymerase and 1 µL of 1° PCR product (diluted 100×) in a total volume of 20 µL. The AD primer used was the same as that used in the 1° PCR. 2° PCR was carried out in a thermal cycler as follows: 94°C for 2 min; 15 supercycles of 94°C for 30 s, 60°C for 1 min, 72°C for 2 min, 94°C for 30 s, 60°C for 1 min, 72°C for 2 min, 94°C for 30 s, 44°C for 1 min, 72°C for 2 min; 72°C for 5 min; hold at 10°C. Ramping speed was set at 1°C s^{-1}. Each 2° PCR product was diluted 100× by combining 2 µL PCR product with 198 µL Milli-Q H_2O.

Each 3° PCR reaction consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 1.25 mM additional $MgCl_2$ (for a total $MgCl_2$ concentration of 2.75 mM), 250 µM of each dNTP, 1 M betaine, 2 µM AD primer, 0.5 µM SP3 primer (see Appendix 2 for primer sequences), 1 U *Taq* DNA polymerase and 1 µL of 2° PCR product (diluted 100×) in a total volume of 20 µL. The AD primer used was the same as that used in the 1° and 2° PCRs. 3° PCR was carried out in a thermal cycler with the same cycling program as for the 2° PCR (see above).

2.4.3.2.3 Agarose Gel Extraction of TAIL PCR Products

Success of the TAIL PCR reactions was determined by electrophoresis of the 1°, 2° and 3° PCR products. A 5 µL aliquot of the 1° PCR reaction, a 5 µL aliquot of the 2° PCR reaction, and the entire 20 µL 3° PCR reaction were combined with 1 µL, 1 µL and 2 µL respectively of 10× loading buffer. The 1°, 2° and 3° PCR reactions were then loaded into adjacent lanes of a 1.5% (w/v) agarose/1× TAE gel (as described in Section

2.4.2.1.4). A 1 Kb Plus DNA Ladder was loaded as a size standard, and a Low DNA Mass Ladder was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4.

Successful amplification of the flanking region appeared on the agarose gel as a discrete band (or bands) in the 2° or 3° PCR reactions. The difference in size between bands in the 2° and 3° PCR reactions usually matched the distance between the 2° and 3° SP primers as calculated from the single-locus ISSR consensus sequence. This was not always the case; presumably some strong bands in the 2° PCR reaction were the result of non-specific amplification, the re-amplification of which was prevented in the more specific 3° PCR reaction.

2° and/or 3° PCR that appeared on the agarose gel as discrete bands were excised from the agarose gels using a sterile scalpel blade. DNA was extracted from the gel slice using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's protocol. DNA was eluted in 30 µL of the provided EB (elution) buffer (10 mM Tris-HCl (pH 8.5)).

2.4.3.2.4 Ethanol Precipitation of TAIL PCR Products

The quality of direct DNA sequencing of gel-extracted PCR products was found to be greatly improved by ethanol precipitation of the eluted PCR product. Presumably, this removes contaminants which otherwise interfere with the sequencing reaction. PCR products were precipitated as described in Section 2.4.3.1.7, dissolved in 10 µL Milli-Q H_2O, and the yield determined using a NanoDrop ND-1000 Spectrophotometer as described in Section 2.4.2.3.4. DNA was stored at −20°C until required.

2.4.3.2.5 DNA Sequencing of Gel-Extracted TAIL PCR Products

TAIL PCR products were sequenced with the SP primer used to generate the product. This was found to be more successful than sequencing with the AD primer, probably due to higher specificity of the SP primer. Each DNA sequencing reaction consisted of 0.75× BigDye Terminator v3.1 Sequencing Buffer, 2.0 µL Ready Reaction Mix (BigDye v3.1), 160 nM SP primer (see Appendix 2 for primer sequences) and TAIL PCR product template (amount calculated as in Section 2.4.3.1.6) in a total volume of 20 µL. DNA sequencing reactions were carried out in a thermal cycler as described in Section 2.4.3.1.6. Sequencing products were purified by ethanol precipitation as described in Section 2.4.3.1.7, and subjected to capillary electrophoresis as described in Section 2.4.3.1.8.

2.4.3.2.6 DNA Sequence Data Analysis

DNA sequence files (electropherograms in ABI format) were imported into Sequencher v. 4.2 software. The regions of overlapping DNA sequence between the TAIL PCR products and the 5′ and 3′ ends of the ISSR consensus sequence were aligned. Ambiguous bases were edited where possible. A new consensus sequence was created, consisting of the left flanking region, cloned ISSR, and right flanking region (Fig. 2.6, No. 7–8).

2.4.3.3 Development of ISSR SCAR Markers

2.4.3.3.1 ISSR SCAR Marker Primer Design

SCAR marker PCR primers were designed in the 5′ and 3′ flanking regions of the ISSR, allowing the entire locus, including the terminal SSR regions and the intervening region, to be amplified and sequenced. Primers were designed within the parameters described in Section 2.4.3.2.1. Each ISSR SCAR marker primer was given a four-part name consisting of: the code for the accession and the number of the individual from which the marker was derived, the number of the excised ISSR band, and L (left) or R

(right) for the end of the marker to which the primer binds e.g., BR01_19_L denotes cv. Bottle Ruku, individual No. 1, ISSR band No. 19, left primer.

2.4.3.3.2 Optimisation PCR for ISSR SCAR Markers

The new ISSR-derived SCAR markers were optimised and tested in separate steps. First, to verify that amplification of the marker was possible, each marker was amplified in the individual from which it was derived. This step was combined with a gradient PCR to determine the optimal annealing temperature of the primers. Next, each ISSR SCAR marker was amplified in a small number of accessions to verify that it was polymorphic between accessions. Polymorphic SCAR markers that could be consistently amplified were used for high-throughput PCR and sequencing.

Each ISSR SCAR marker gradient PCR consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 250 µM of each dNTP, 1 M betaine, 0.5 µM L primer, 0.5 µM R primer (see Appendix 2 for primer sequences), 1 U *Taq* DNA polymerase and ~1 ng of genomic DNA (from the individual from which the SCAR marker was derived) in a total volume of 20 µL. Ten reactions were prepared for each SCAR marker and subjected to a gradient PCR with annealing temperatures at ~1.3°C increments from 50–62°C. SCAR marker gradient PCR was carried out in a TGradient (Biometra) thermal cycler as follows: 94°C for 2 min; 35 cycles of 94°C for 30 s, 50–62°C for 30 s, 72°C for 1 min; 72°C for 5 min; hold at 10°C.

Success of the ISSR SCAR marker optimisation PCR reaction was determined by electrophoresis of the PCR product. A 5 µL aliquot of each PCR reaction was combined with 1 µL 10× loading buffer in a total volume of 10 µL, and loaded onto a 1.5% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A 1 Kb Plus DNA Ladder was loaded as a size standard, and a Low DNA Mass Ladder was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4. Successful amplifications appeared on the agarose gel as a discrete band that matched the expected size of the ISSR SCAR marker. For each marker the annealing temperature that gave a single, discrete product of sufficient yield was chosen as the optimal annealing

temperature for that marker. Markers that failed to amplify, or did not produce a single product of the expected size, were discarded.

2.4.3.3.3 Trial PCR of ISSR SCAR Markers on Several Accessions

To verify that the ISSR SCAR markers were polymorphic, each marker was amplified and sequenced in a small number of accessions (~10), including the accession from which the marker was derived (as a positive control) and those accessions from the ISSR fingerprinting gel in which the band originally appeared as polymorphic. PCR was carried out as described above, and using the optimum annealing temperature from the gradient PCR. PCR reactions were electrophoresed as described above. Successful amplifications appeared on the agarose gel as a discrete band for each of the accessions amplified.

2.4.3.3.4 SAP/Exo I Cleanup of PCR Products

Prior to DNA sequencing, trial ISSR SCAR marker PCR products were treated with shrimp alkaline phosphatase (SAP) and exonuclease I (Exo I) to remove unincorporated dNTPs and primers respectively. SAP dephosphorylates dNTPs, and Exo I digests the single-stranded primers. Both enzymes are irreversibly denatured at 80°C, so there is no carry-over of functional enzyme to the sequencing reaction. Each cleanup reaction consisted of the remaining ISSR SCAR marker PCR product, 2 U SAP (USB) and 10 U Exo I (USB). The reactions were incubated at 37°C for 30 min and then at 80°C for 15 min. The reactions were stored at −20°C until required.

2.4.3.3.5 DNA Sequencing of Trial ISSR SCAR Markers

ISSR SCAR marker PCR products were sequenced in both directions using the L (left) and R (right) PCR primers. Each DNA sequencing reaction consisted of 0.75× BigDye Terminator v3.1 Sequencing Buffer, 2.0 µL Ready Reaction Mix (BigDye v3.1), 160 nM SP primer (see Appendix 2 for primer sequences) and ISSR SCAR marker PCR

product template (amount calculated as in Section 2.4.3.1.6) in a total volume of 20 µL. DNA sequencing reactions were carried out in a thermal cycler as described in Section 2.4.3.1.6. Sequencing products were purified by ethanol precipitation as described in Section 2.4.3.1.7, and subjected to capillary electrophoresis as described in Section 2.4.3.1.8.

2.4.3.3.6 DNA Sequence Data Analysis

DNA sequence files (electropherograms in ABI format) were imported into Sequencher v. 4.2 software. For each marker, sequences (both forward and reverse) for all taxa were aligned. Ambiguous bases were edited where possible. Markers that were polymorphic were used for sequencing the bottle gourd accessions described in Chapter Three.

2.5 DEVELOPMENT OF BOTTLE GOURD SCAR MARKERS: RESULTS AND DISCUSSION

A total of 31 polymorphic ISSR bands were excised from polyacrylamide fingerprinting gels. These are shown in Table 2.2. The estimated size of the bands ranged from 230–900 bp. All but five of the 31 bands could be re-amplified using PCR.

Table 2.2
Polymorphic ISSR Bands Isolated for Bottle Gourd

Band Number	ISSR Primer[a]	Accession Code[b]	PCR Number	Estimated Size (bp)	Successful Re-amplification?
17	#811	BOP	A1-2	230	Yes
18	#808	BR	B9-10	320	Yes
19	#808	BR	B9-10	390	Yes
20	#808	MR	B7-8	250	Yes
21	#808	BR	B9-10	310	Yes
22	#823	MR	C7-8	470	Yes
23	#823	MR	C7-8	800	Yes
24	#823	MR	C7-8	550	Yes
25	#823	BOP	C1-2	460	Yes
26	#823	BOP	C1-2	390	Yes
27	#823	BOP	C1-2	380	Yes
28	#846	BOP	E1-2	900	Yes
29	#846	195	E3-4	830	Yes
30	#846	BOP	E1-2	800	Yes
31	#846	BOP	E1-2	550	Yes
32	#846	MR	E7-8	900	Yes
33	#823	MR	C7-8	380	Yes
34	#823	195	C3-4	360	Yes
35	#812	BOP	A1-2	650	Yes
36	#812	195	A3-4	645	No
37	#813	BOP	B1-2	650	No
38	#813	MR	B7-8	490	No
39	#813	MR	B7-8	430	No

Table continued next page...

Table continued…

Band Number	ISSR Primer[a]	Accession Code[b]	PCR Number	Estimated Size (bp)	Successful Re-amplification?
40	#813	MR	B7-8	310	Yes
41	#813	BR	B9-10	300	Yes
42	#813	MR	B7-8	295	Yes
43	#813	BR	B9-10	290	Yes
44	#813	BOP	B1-2	260	Yes
45	#813	BOP	B1-2	250	Yes
46	#813	BR	B9-10	260	No
47	#814	195	C3-4	780	Yes

[a] See Table 2.1 for ISSR primer sequences.
[b] See Table 3.2 for accession code details.

Of the 26 re-amplified products, a subset of 10 were cloned, sequenced and used to design SP primers for TAIL PCR (Table 2.3). Using TAIL PCR, left (5′) and right (3′) flanking regions were obtained for eight products. For ISSR_26, TAIL products could not be successfully amplified for either flanking region. For ISSR_32, the left flanking region only was obtained. Both these markers were discarded at this point, and the eight markers for which flanking regions had been obtained were taken to the next step.

Table 2.3
TAIL PCR and SCAR Marker Development

Cloned Band[a]	Size (bp)	TAIL Primers Designed?	Flanking Regions				Marker Name
			Left		Right		
			Obtained?	Size (bp)	Obtained?	Size (bp)	
ISSR_19	357	Yes	Yes	108	Yes	182	**BR01_19**
ISSR_21	310	Yes	Yes	167	Yes	122	**BR01_21**
ISSR_23	811	Yes	Yes	122	Yes	126	**MR06_23**
ISSR_24	548	Yes	Yes	197	Yes	43	**MR06_24**
ISSR_26	383	Yes	No		No		
ISSR_27	367	Yes	Yes	295	Yes	79	**BOP19_27**
ISSR_29	842	Yes	Yes	138	Yes	225	**JAP06_29**
ISSR_31	520	Yes	Yes	124	Yes	219	**BOP19_31**
ISSR_32	896	Yes	Yes	222	No		
ISSR_35	656	Yes	Yes	51	Yes	330	**BOP19_35**

[a] The ISSR sequence number corresponds to the band number in Table 2.2.

Of the eight SCAR markers developed, six could be successfully amplified in bottle gourd, producing single PCR products of the expected size (Table 2.4). Sequencing of PCR products for all markers produced the expected sequence (comprising the original ISSR sequence flanked by the two TAIL sequences). It is unclear why the remaining two markers (MR06_23 and JAP06_29) could not be amplified. If the ISSR sequence exists as multiple copies in the genome then it is possible that the contig created for the ISSR and TAIL sequences does not correspond to a single locus. If this were the case then PCR would not be successful.

Five of the markers that were successfully amplified proved to be polymorphic and useful for testing the hypotheses presented in Chapter Three. BR01_19 possesses three concordant SNPs. That is, individuals have alleles that are either G–C–G or A–T–A (although some individuals were found to be heterozygous, there were never any observed cases of recombination between the alleles). MR06_24, BOP19_31 and BOP19_35 each possess a SNP. BOP19_27 possesses a 14 bp indel and a $(GA)_n$

microsatellite. Four alleles were observed for BOP19_27: In–6, In–7, Del–5 and Del–6 (where 5, 7 or 7 = *n*). Although BR01_21 is listed in Table 2.4 as not being polymorphic, there was one accession (#195 Akita, Japan) that possesses a SNP relative to other accessions sequenced. Although not useful for testing the hypotheses in the present study, this marker may be useful in the future for reconstructing fine-scale dispersal patterns within Asia. The five polymorphic markers obtained here are used in Chapter Three to test the origins of the Polynesian bottle gourd.

Table 2.4
Polymorphisms Present in Developed SCAR Markers

Marker Name	**Size (bp)**	**Amplification of SCAR Marker?**	Polymorphic?	Polymorphisms Present
BR01_19	641	Yes	Yes	3 concordant SNPs
BR01_21	582	Yes	No	
MR06_23	1059 (predicted)	No		
MR06_24	738	Yes	Yes	1 SNP
BOP19_27	740	Yes	Yes	1 indel; variable SSR
JAP06_29	1163	No		
BOP19_31	759 (predicted)	Yes	Yes	SNP
BOP19_35	1014	Yes	Yes	SNP

One aim of this study was to determine whether the success rate of SCAR marker development could be improved by using the TAIL PCR technique to obtain flanking regions. It was hoped that designing SCAR marker primers in the flanking regions would allow polymorphisms that caused the original presence–absence phenotype on the fingerprinting gel to be captured. It was also hoped that additional polymorphisms in the flanking regions themselves might be discovered. To determine the extent to which TAIL PCR improved SCAR marker development, the distributions of polymorphisms along the length of the TAILed loci were examined (Table 2.5). These distributions can be compared with the number of polymorphisms that would have been discovered had SCAR markers been developed using the traditional method (i.e., without obtaining flanking regions).

Table 2.5
Rates of Polymorphism in TAILed vs. non-TAILed SCAR Markers

Marker Name	Zone						
	1	2	3	4	5	6	7
BR01_19	—	—	—	3	—	—	—
MR06_24	1	—	—	—	—	—	—
BOP19_27	—	—	—	—	1	1	—
BOP19_31	—	—	—	—	1	—	—
BOP19_35	—	—	—	1	—	—	—

In Table 2.5, zones 1 and 7 correspond to the flanking regions, zones 2 and 6 to the SSRs, zones 3 and 5 to the locations where the SCAR marker primers would be located if traditional SCAR techniques had been employed, and zone 4 corresponds to the central region that would be amplified regardless of whether TAILed or traditional methods were used. Table 2.5 shows the distribution of polymorphisms across the seven zones for each of the TAILed SCAR markers. To determine the location of zones 3 and 5 for each marker the sequence of that marker was examined for suitable primer locations. These hypothetical primers were designed as close to the SSR repeats as possible and using the design parameters listed in Section 2.4.3.2.1.

Table 2.5 shows that of the eight polymorphisms, only four would have been captured if the flanking regions had not been available. Or, in other terms, only two of the five markers would have been polymorphic. It is possible that BOP19_27 may also have been polymorphic, but probably only as a dominant presence–absence marker because the large indel is located in the hypothetical primer-binding site.

Interestingly, only BOP19_27 contains a polymorphism that could explain the original presence–absence phenotype on the ISSR fingerprinting gel (a variable length SSR and an indel). Polymorphisms in the other markers would not have produced visible phenotypes on the original gel (they are all SNPs). This prompts the question of what *did* cause the original phenotypes.

In summary, using TAIL PCR to obtain flanking regions appears to significantly improve the success rate of SCAR marker development. In addition to capturing polymorphisms that caused the original presence–absence phenotype on the fingerprinting gel, success is also improved by discovering more polymorphism by chance (i.e., just from having more sequence data). The potential benefits of using TAIL PCR (and other techniques) to obtain flanking regions in SCAR marker development may need to be assessed on a per study basis, but it does appear that this approach can significantly improve the success rate of SCAR marker development.

2.6 SCREENING OF OTHER MARKERS FOR BOTTLE GOURD

In an attempt to discover other useful markers for bottle gourd, a number of PCR and sequencing markers developed for other taxa were screened (Appendix 1). This was achieved by first attempting to amplify the markers in bottle gourd (using a range of annealing temperatures and $MgCl_2$ from 1.5–3.0 mM). If amplification was successful, the markers were sequenced in a small number of accessions (5–10) from a broad range of geographic locations (i.e., inter-continental) to search for useful polymorphisms.

2.6.1 SCREENING OF CHLOROPLAST MARKERS

A total of 23 established and widely-used (both in terms of number of studies and taxonomic breadth) chloroplast markers were screened, including *rbc*L (Olmstead *et al.*, 1992), the *trn*L–F intergenic spacer (Taberlet *et al.*, 1991), the *trn*C–*trn*D intergenic spacer (Lee & Wen, 2004) and 20 consensus chloroplast SSR (ccSSR) markers (Chung *et al.*, 2003; Chung & Staub, 2003). The results of the chloroplast marker screening are summarised in Appendix 1. Neither ccSSR09 nor ccSSR10 could be amplified, even though these markers were successfully amplified by Chung *et al.* (2003) and showed a 1 bp difference in length between three bottle gourd accessions (Table 3 in Chung *et al.*, 2003). Although *rbc*L, *trn*L–F, *trn*C–*trn*D and 14 of the ccSSR markers were successfully amplified, only *trn*C–*trn*D and ccSSR20 proved to be polymorphic. Although ccSSR20 may be useful for future work, it was not used in this study because the single SNP present is congruent with polymorphisms in *trn*C–*trn*D and *trn*S–*trn*G (see below). *trn*C–*trn*D contains an insertion–deletion (indel) and a SNP, both of which separate Asian and American gourds, so is useful for testing the hypothesis of an American origin for Polynesian gourds. The position of the *trn*C–*trn*D marker on the cucumber chloroplast genome is shown in Fig. 2.7.

2.6.2 SCREENING OF NUCLEAR MARKERS

The screened nuclear markers comprised the ribosomal DNA intergenic spacer (nrITS) (Baldwin, 1992; Hershkovitz & Zimmer, 1996) and 22 microsatellite markers developed from commercially important Cucurbitaceae species: 13 from rockmelon/cantaloupe (*Cucumis melo*) (Katzir *et al.*, 1996; Chiba *et al.*, 2003; Ritschel *et al.*, 2004), 7 from watermelon (*Citrullus lanatus*) (Jarret *et al.*, 1997), and 2 from cucumber (*Cucumis sativus*) (Katzir *et al.*, 1996). The only markers which could be amplified were ITS, 2 of the watermelon markers, 1 rockmelon marker, and 1 cucumber marker[1]. Of these, the rockmelon marker CMMS30-3 (Chiba *et al.*, 2003) proved to be polymorphic, and although promising if further developed (i.e., new primers designed in the flanking regions), it could not be amplified consistently enough to be used for high-throughput PCR and sequencing in bottle gourd. The cucumber marker CSHPRAG (Katzir *et al.*, 1996), although successfully amplified, has not been sequenced; this may be another useful marker. The results of the nuclear marker screening are summarised in Appendix 1.

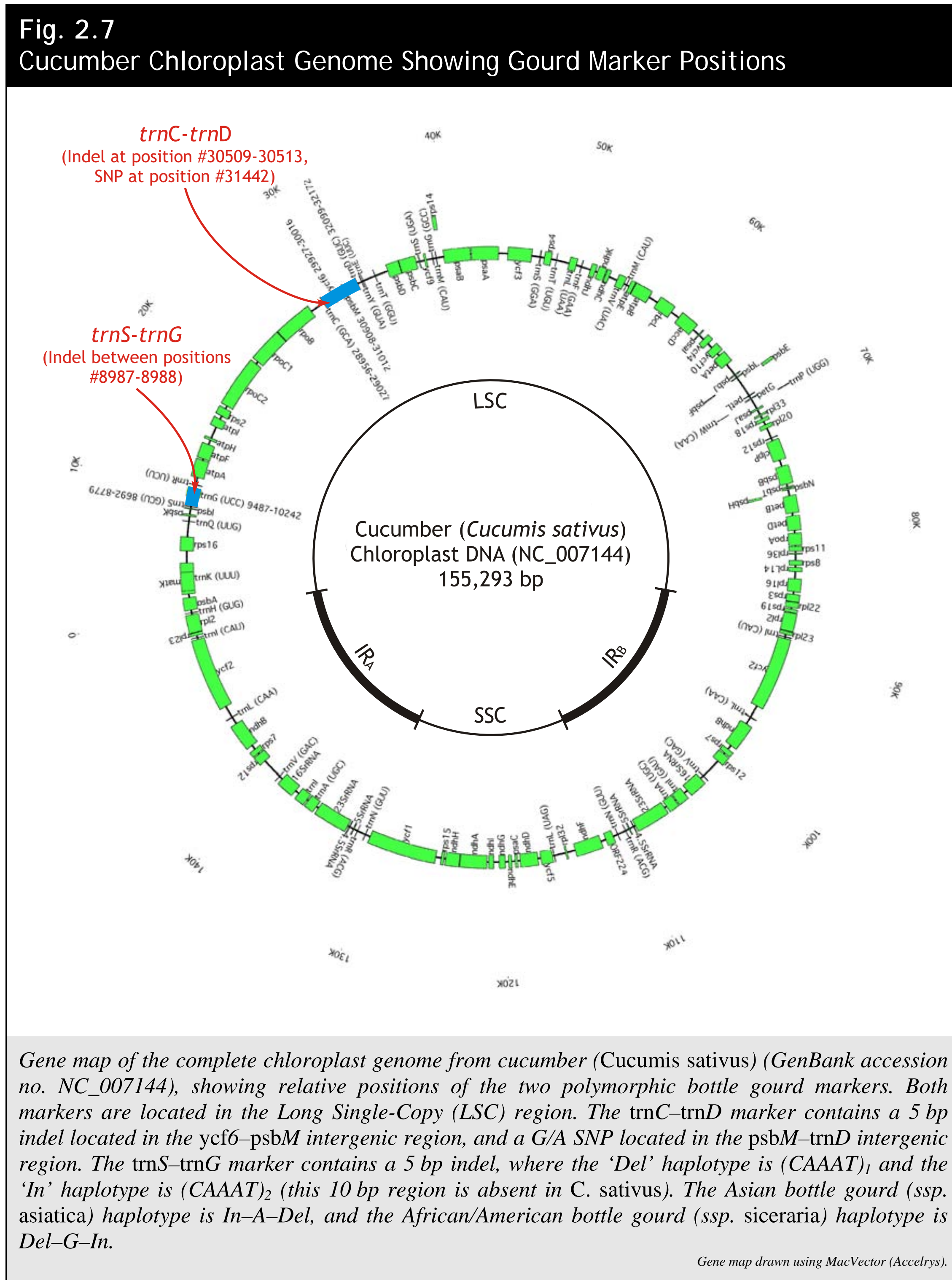
2.6.3 *TRN*S–*TRN*G MARKER FROM INTER-CCSSR PCR

The ~150 bp ccSSR2 marker (Chung *et al.*, 2003; Chung & Staub, 2003) was thought to be variable in bottle gourd (David Erickson, pers. comm.) but I was unable to amplify it, presumably because of primer–template mis-matching. To obtain sequence data for ccSSR2, I used the primers ccSSR2_F with ccSSR3_R (Chung *et al.*, 2003; Chung & Staub, 2003), which amplified the entire ~1,650 bp ccSSR2–ccSSR3 region, with the ccSSR2 marker at one end, and the ~100 bp ccSSR3 marker at the other. Although the ccSSR2 marker itself does not contain any polymorphisms, the intervening region (which corresponds to the *trn*S–*trn*G intergenic spacer) was sequenced, and was found to contain an indel which separates Asian and New World gourds, so is useful for testing the hypothesis. A new primer (trnG_R) was designed to use with the ccSSR2_F

[1] In most cases these markers could not be amplified at all, even in the taxa from which they were developed; in the case of the markers from Chiba *et al.* (2003), and maybe others, the poorly designed PCR primers were probably responsible.

to amplify *trn*S–*trn*G. The position of the *trn*S–*trn*G marker on the cucumber chloroplast genome is shown in Fig. 2.7.

Fig. 2.7
Cucumber Chloroplast Genome Showing Gourd Marker Positions



*Gene map of the complete chloroplast genome from cucumber (*Cucumis sativus*) (GenBank accession no. NC_007144), showing relative positions of the two polymorphic bottle gourd markers. Both markers are located in the Long Single-Copy (LSC) region. The* trnC–trnD *marker contains a 5 bp indel located in the* ycf6–psbM *intergenic region, and a G/A SNP located in the* psbM–trnD *intergenic region. The* trnS–trnG *marker contains a 5 bp indel, where the 'Del' haplotype is (CAAAT)$_1$ and the 'In' haplotype is (CAAAT)$_2$ (this 10 bp region is absent in* C. sativus*). The Asian bottle gourd (ssp.* asiatica*) haplotype is In–A–Del, and the African/American bottle gourd (ssp.* siceraria*) haplotype is Del–G–In.*

Gene map drawn using MacVector (Accelrys).

2.7 THE AFLP TECHNIQUE: INTRODUCTION AND METHODOLOGY

2.7.1 INTRODUCTION

Amplified fragment length polymorphism (AFLP®[2]) DNA fingerprinting (Vos *et al.*, 1995) is a firmly established molecular marker technique (Mueller & Wolfenbarger, 1999; Meudt & Clarke, 2007; Appendix 7). It has broad applications in linkage mapping (Saliba-Colombani *et al.*, 2000), parentage analysis (Gerber *et al.*, 2000), measuring genetic diversity (Mariette *et al.*, 2002; Nybom, 2004), identifying hybrids (Goldman *et al.*, 2004) and cultivars (McGregor *et al.*, 2000), population genetics (Woodhead *et al.*, 2005; Barluenga *et al.*, 2006), reconstruction of shallow phylogenies (Després *et al.*, 2003; Perrie *et al.*, 2003), population assignment (Campbell *et al.*, 2003), and developing single-locus sequence-characterised amplified region (SCAR) markers (McLenachan *et al.*, 2000; Shirasawa *et al.*, 2004; Bussell *et al.*, 2005). AFLP has become the method of choice for many studies on plants and, more recently, for animals, fungi and bacteria (Bensch & Åkesson, 2005), spanning numerous disciplines in genetics, evolution and ecology.

The markers that make up an AFLP fingerprint, although often concentrated in centromeric regions (Alonso-Blanco *et al.*, 1998; Saliba-Colombani *et al.*, 2000), are widely distributed throughout the genome, allowing an assessment of genome-wide variation. These anonymous markers consist largely of non-coding DNA (Wong *et al.*, 2001; Shirasawa *et al.*, 2004). As outlined in Section 2.2.1.1, AFLP is ideal when there is no *a priori* sequence information, for intra-specific studies, when genomic heterogeneity is high, when genetic variability is low, in polyploids, when hybridisation is occurring, for the rapid generation of data, when high quality DNA is available, and where there are no suitable established markers (Mueller & Wolfenbarger, 1999; Bensch & Åkesson, 2005; Meudt & Clarke, 2007).

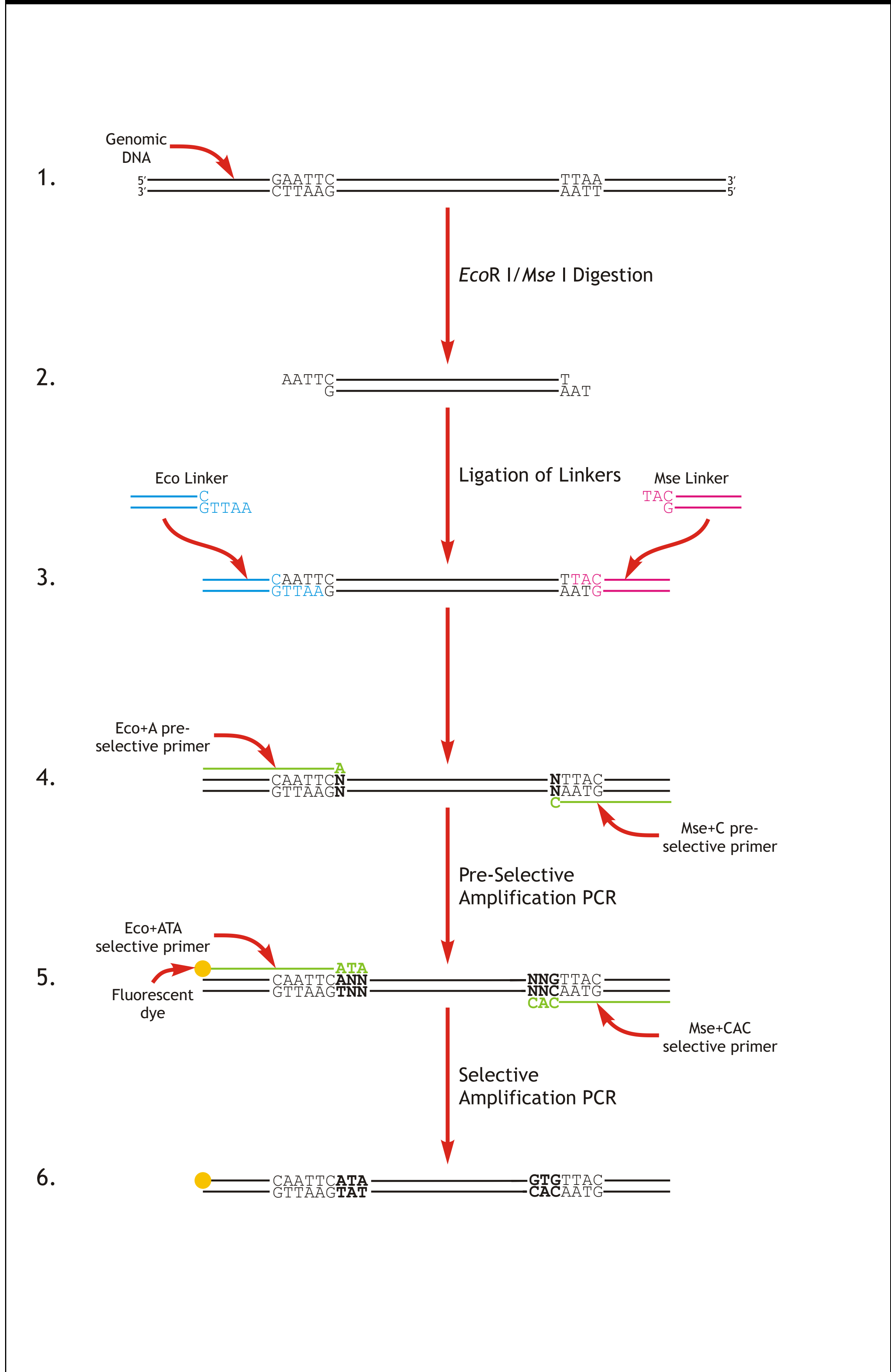
---

[2] Although "AFLP" was not originally intended as an acronym (Vos *et al.*, 1995), I treat it as one here, consistent with widespread usage.

2.7.2 GENERATING AN AFLP FINGERPRINT: METHODOLOGY

AFLPs are generated by complete restriction endonuclease digestion of total genomic DNA, followed by selective PCR amplification and electrophoresis of a subset of the fragments (Fig. 2.8), resulting in a unique, reproducible fingerprint (or profile) for each individual (Mueller & Wolfenbarger, 1999).

Fig. 2.8
Overview of the AFLP Fingerprinting Technique



1.–2. Restriction Endonuclease Digestion of DNA

Genomic DNA is digested with a pair of restriction endonucleases: a rare cutter, typically the six-base cutter EcoR I, and a frequent cutter, typically the four-base cutter Mse I. This results in DNA fragments up to ~1 kb in size. Three species of fragments are produced: EcoR I sticky end at both ends (Eco–Eco), Mse I sticky end at both ends (Mse–Mse), and, as shown, EcoR I sticky end at one end and an Mse I sticky end at the other (Eco–Mse).

3. Ligation of Linkers to the Restriction Fragments

Double-stranded EcoR I (Eco) and Mse I (Mse) linkers (synthetic oligonucleotides of known sequence), which have ends complementary to those produced in the restriction digestion, are ligated to the restriction fragments. Steps 1 and 2 can be performed in the same reaction. The linker serves as template for PCR amplification of the restriction fragment.

4. Pre-Selective Amplification of Ligation Products

A subset of all the fragments is amplified by PCR, using primers which are complementary to the linker sequence with the addition of one nucleotide (A, G, C or T) at the 3' end of the primer (usually Eco+A and Mse+C). These 'pre-amp' primers will only prime DNA synthesis of fragments with bases flanking the restriction sites which are complementary to the selective nucleotides of the primers, thus reducing the number of fragments to about $\frac{1}{16}$ of the initial amount.

5.–6. Selective Amplification of Pre-Selective Amplification PCR Products

The number of fragments is further reduced — to a suitable number to be visualised by electrophoresis — by a second round of PCR (selective amplification), in which PCR product from the pre-selective amplification reaction ('pre-amp product') is used as template. In the second round of PCR, the primers have an additional two selective bases (e.g., Eco+ATA and Mse+CAC). The Eco+3 primer is labelled with a fluorescent dye (a 'fluorophore'), so that all strands synthesised from this primer are fluorescently-labelled (i.e., one strand from Eco–Mse fragments, both strands from Eco–Eco fragments). Alternative subsets of loci can be amplified by using combinations of primers with different selective bases.

Modified from Vos (1995) and Gibco BRL (n.d.)

AFLP PCR products are poolplexed and subject to capillary electrophoresis (Fig. 2.9), producing an electronic profile of fluorescent intensity versus fragment size (Fig. 2.10).

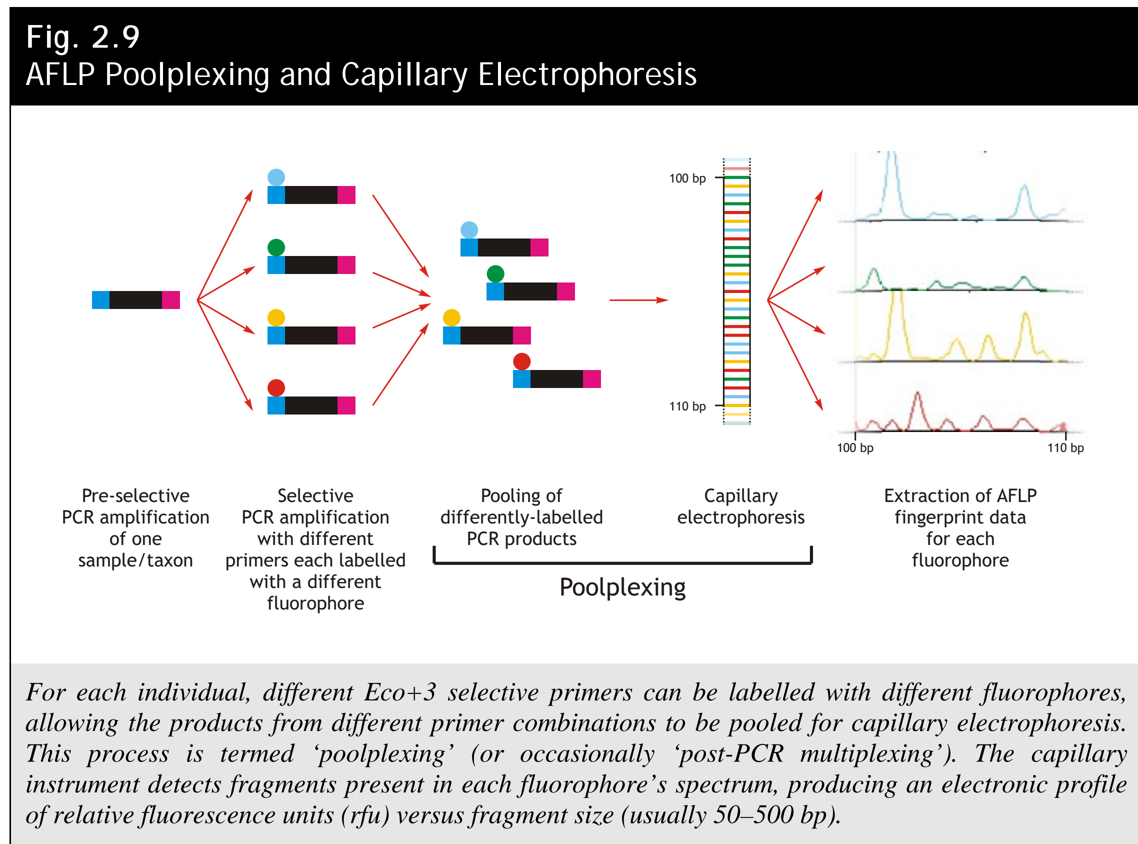
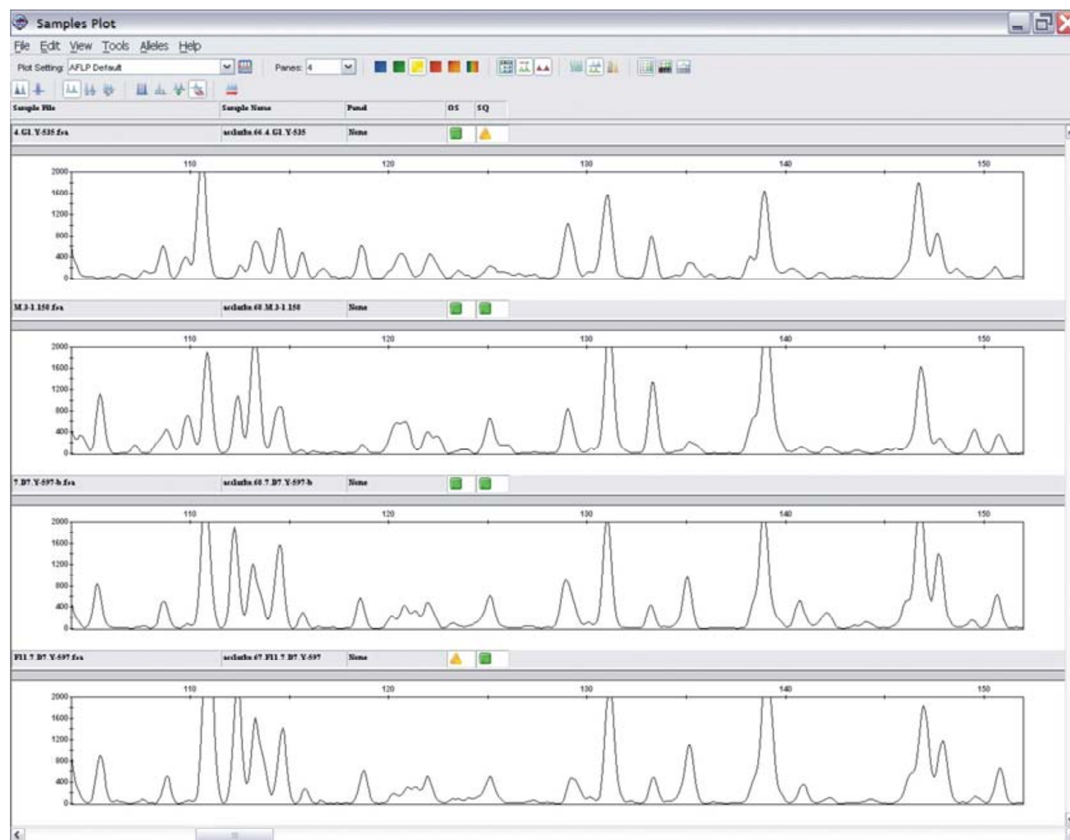


Fig. 2.9
AFLP Poolplexing and Capillary Electrophoresis

For each individual, different Eco+3 selective primers can be labelled with different fluorophores, allowing the products from different primer combinations to be pooled for capillary electrophoresis. This process is termed 'poolplexing' (or occasionally 'post-PCR multiplexing'). The capillary instrument detects fragments present in each fluorophore's spectrum, producing an electronic profile of relative fluorescence units (rfu) versus fragment size (usually 50–500 bp).

Polymorphisms, which are observed as peaks present in some samples and absent in others, are caused by the gain or loss of a restriction site, a change (e.g., SNP) in the selective primer binding site, or a length polymorphism (e.g., indel or variable microsatellite) between the restriction sites (Vos *et al.*, 1995; Mueller & Wolfenbarger, 1999; Bonin *et al.*, 2005). Profiles from multiple individuals are aligned and scored based on the presence (1) or absence (0) of a peak, producing a binary data matrix. Data scoring is covered in more detail in Sections 2.8–2.11 below.

Fig. 2.10
Typical Electronic Profiles for AFLP



This screenshot from GeneMapper® shows AFLP profiles for four sweet potato samples, from top to bottom: Y-535 Society Islands, 158 cv. Mary Anne, Y-597 Peru B, and Y-597 Peru A. Thus, the first three rows show profiles from three different individuals, and the last two samples are replicate AFLP profiles obtained from a single individual (Y-597 Peru). Each profile consists of a plot of fluorescence (relative fluorescent units; rfu) versus fragment mobility/size (base pairs; bp) for one fluorescently-tagged primer pair combination — in this case from approx. 105–150 bp using a NED-labelled primer. These raw fluorescence data are converted to binary data by first binning the data (grouping the similar-sized peaks from different accessions into a single character) and then scoring the peaks in that character as 1 (present; plus-allele) or 0 (absent; null-allele). The resulting character matrix of 0's and 1's is then exported for phylogenetic analysis.

2.8 OPTIMISING AFLP SCORING PARAMETERS IN SWEET POTATO: INTRODUCTION

Of interest in this study is the use of AFLP to generate data for reconstructing phylogenetic relationships between sweet potato accessions. Although some researchers have suggested that AFLP data are inappropriate for phylogenetic applications (Hollingsworth & Ennos, 2004; Kosman & Leonard, 2005), several empirical studies have revealed tree-like properties in AFLP datasets, and AFLP data are increasingly being used to estimate phylogenies, including for recent species radiations (e.g., Barluenga *et al.*, 2006; Ellis *et al.*, 2006; Albach, 2007), and crops and their wild progenitors (e.g., Spooner *et al.*, 2005; Kilian *et al.*, 2007). Because AFLP markers are sampled throughout the genome, they likely uncover rare genetic differences in groups with low sequence variation (Mendelson & Shaw, 2005), and have been shown to result in well-resolved trees that are consistent with independent data (Bussell *et al.*, 2005; Koopman, 2005).

A significant criticism of using AFLP data for phylogenetics centres around the effects of incomplete lineage sorting, a phenomenon which results in the phylogeny of a single locus not matching the species phylogeny (Degnan & Salter, 2005). The alternative view is that phylogenetic analysis of AFLP data can give a robust estimate of the overall species phylogeny (Sullivan *et al.*, 2004; Koblmüller *et al.*, 2007) because the hundreds or thousands of concatenated AFLP loci from a typical AFLP study are more likely, on average, to approximate the species phylogeny, and the individual effects of loci that have evolutionary histories different to that of the species are more likely to be diminished. In any case, the problems of 'gene tree' vs. 'species tree' incongruence, which might hamper AFLP-based phylogenetic reconstruction of sexually reproducing taxa, are not expected to occur in analyses of sweet potato, which is an almost exclusively asexually reproducing species (see Section 2.2.5).

The focus in this section is on optimising AFLP scoring parameters to maximise the phylogenetic signal obtained from the raw data. In capillary electrophoresis of fluorescently labelled AFLP fragments the end result is the production of a profile like

the ones shown in Fig. 2.10. To convert the data for numerous profiles into a binary character matrix of 0's (peak absent; null-allele) and 1's (peak present; plus-allele), two types of decisions have to be made. First, when should a fragment be called as present (character state 1) vs. absent (character state 0)? This is mostly determined by the height of the peak and, depending on the scoring software used, sometimes by other attributes as well (see Applied Biosystems, 2004; SoftGenetics, 2006). Second, when should two fragments be designated as having the same length (number of nucleotides) and therefore be treated as identical plus-alleles of the same locus (Koopman & Gort, 2004)? The ideal is to have all truly identical fragments recognised, scored as present, and assigned to the same column of the character matrix — and to have no non-identical fragments assigned to the same column of the character matrix. In practice it is likely that not all characters in an AFLP data matrix contain identical plus-alleles and identical null-alleles because, firstly, some non-identical fragments will have similar mobility by chance, secondly, identical fragments will have slightly different mobilities and peak heights due to random error (Vekemans *et al.*, 2002), and thirdly, shared absences (null-alleles) may be derived in multiple, independent ways.

New capillary-based technologies allow more precise estimates of AFLP DNA fragment mobility (fragment length) and fluorescence intensity (amplitude) than traditional gel-based systems. Furthermore, analysis of capillary profiles with currently available automated scoring software (see Table 1 in Meudt & Clarke, 2007; Appendix 7) allows the user to control several parameters that influence the resulting data matrix. In contrast to manual scoring, automated scoring is objective, repeatable, and far less time-consuming (Pompanon *et al.*, 2005). In fact, with increasingly large datasets, automated scoring is often the only feasible option, yet to my knowledge no experimental or theoretical studies have explored different automated scoring parameter settings and their effects on downstream analyses. Given that AFLP has many potential applications and that the automated scoring packages have many adjustable parameters it is natural to ask: how can we measure the quality of the AFLP character matrix, and what is the best way to go about optimising AFLP scoring parameters for phylogenetic studies? More specifically, when scoring a particular AFLP dataset, which parameter settings will give the most accurate phylogenetic estimate?

The aim is to find parameter settings for automated scoring software that lead to data matrices whose analysis allows us to accurately recover the true tree (i.e., the species phylogeny). However, there are difficulties in directly measuring the accuracy of phylogenetic estimates. This is because, (1) in general the true tree is not known, and (2) the question is not amenable to study with a simulation based approach, such as AFLP *in silico* (Qin *et al.*, 2001; Bikandi *et al.*, 2004), because the factors which influence bin width and peak height are not currently understood well enough to be simulated accurately. Because we cannot measure accuracy directly and simulation studies are not applicable, we use the resolution of the phylogenetic tree resulting from bootstrap analysis of the data matrices constructed with different parameter settings as a proxy for accuracy. The higher the resolution the more information there is about phylogenetic relationships, and unless there is some systematic bias, high resolution should be correlated with accuracy (Hillis & Bull, 1993). For example, Taylor & Piel (2004) showed empirically that high bootstrap support was strongly correlated with accuracy in their study using a genome-scale yeast dataset. In addition, it is possible to assess the topological congruence of the resulting AFLP phylogeny with independent sources of evidence (e.g., anthropological, historical) as an additional measure of accuracy.

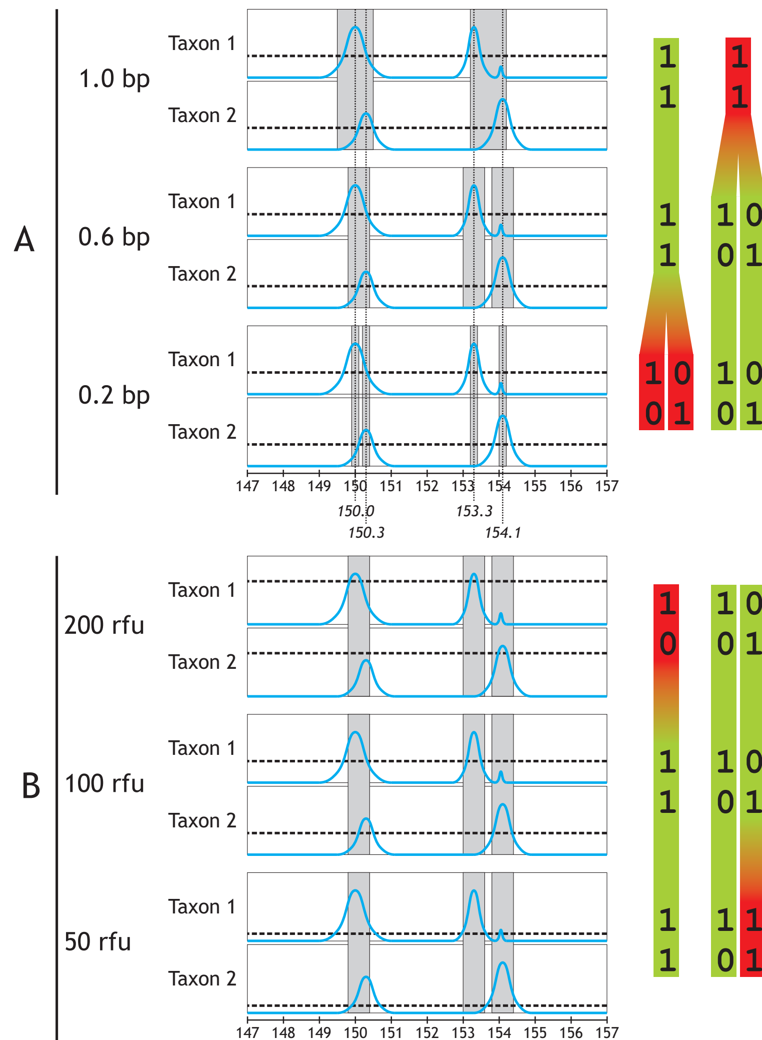
Error in data matrices can be measured through the study of replicate AFLP profiles, where two AFLP fingerprints are independently generated from one individual, either from the same or from independent DNA extractions. When a replicate pair of samples is analysed we expect to see, for a given character, either both samples with a peak called as present (i.e., character states 1,1) or both samples lacking a peak (0,0); situations where a peak is called as present in one replicate sample and absent in the other (0,1 or 1,0) are clearly errors.

The effect of parameter choice is explored using two commonly available software platforms designed for automated AFLP scoring: GeneMapper® v. 3.7 by Applied Biosystems and GeneMarker® v. 1.51 by SoftGenetics LLC. The parameters studied include: the minimum peak height threshold required for a peak to be called as present, the minimum fragment length at which a marker is scored and included as a character in the matrix, and the width of the marker bins in base pairs (bp). Each of these parameters influences the number of characters available for phylogenetic analysis

and whether or not these characters represent homologous fragments (Fig. 2.11). This in turn affects both the resolution and replicate error rate. Introducing more homologous characters should lead to higher resolution, but in practice by including more characters we also risk introducing errors. For each of the main parameters studied there is expected to be a trade-off between getting more characters of lower quality and fewer characters of high quality.

Specifically, these trade-offs include the following. First, lowering the minimum fragment length will increase the number of characters, but these characters may be of lower quality, as previous studies (Vekemans *et al.*, 2002) indicate that smaller fragments are more likely to be homoplasious than larger fragments. Second, reducing the bin width will increase the number of characters, but as bin width is reduced, single characters may split into pairs of characters (Fig. 2.11A) which at the very least can reduce branch support in the resulting tree but could also potentially introduce error into the dataset. For example, if there are single peaks in each bin, reducing the bin width cannot create pairs of characters that are in conflict with each other but it will reduce support for internal edges of the tree, i.e., if bin width is reduced indefinitely all characters will become parsimony uninformative singletons (e.g., one "1" and $n-1$ "0's"). In the case where two peaks have been assigned to the same bin, splitting the bin by reducing the bin width could create two incongruent characters. Conversely, bin width is increased, separate characters will be amalgamated. If these characters are not really identical this could create character conflict in the resulting data matrix (Fig. 2.11A) which may reduce resolution. Third, lowering the peak height threshold will increase the number of characters. If peak height threshold is set too low we will, by scoring background noise, call peaks present when they are really absent, and if it is set too high we will call peaks as absent when they are really present (Fig. 2.11B).

Fig. 2.11
Effects of Changing Bin Width and Peak Height Threshold



The binary matrix on the right shows the effects of changing the AFLP scoring parameters in the profiles on the left; correctly-scored peaks are represented in the binary matrix in green and incorrectly scored peaks in red.

A. *Bin widths (BW) are shown as grey rectangles underneath the profiles; peaks that fall within a given bin width are scored as 1 (present) and outside as 0 (absent). Using a bin width that is too wide (1.0 bp) makes it more likely that identical alleles (whose mobilities differ only slightly because of random error) are treated as a single character, but it can also cause non-identical alleles to be incorrectly treated as one character. Although this problem is corrected by narrowing the bin width (0.6 bp), if the bin width is too narrow (0.2 bp) then even the identical alleles will be wrongly split into separate characters.*

B. *Peak height threshold (PHT) is shown using black dashed lines; a peak above this line is scored as 1 (present) and below as 0 (absent). If the PHT is set too high (200 rfu), then peaks that are present will be scored as absent (Taxon 2, left peak). Although this is corrected by lowering the PHT (100 rfu), if the PHT is too low (50 rfu) then background noise or stutter peaks will be incorrectly scored as present (Taxon 1, right peak). This simplified example shows two hypothetical taxa and three characters, but real datasets may contain hundreds of taxa and hundreds of characters; determining the optimum parameter settings over all taxa and all characters is much more complex.*

It may seem counter-intuitive to consider bin widths less than one base pair. However, both gel- and capillary-based AFLP systems measure mobility and only *estimate* length. Mobility values, and thus the estimated length values, effectively vary continuously due to differences in strand composition (sequence and secondary structure). In fact, DNA fragments with the same number of nucleotides that have different sequences can differ in mobility by as much as 10 “bp” when measured against a size standard (Clarke, 2001). However, it is when the mobility difference of non-identical DNA fragments differs by 1 bp or less that there can be serious problems of identity assessment. In such cases, bin width settings < 1 bp have the potential to separate non-identical fragments with similar mobility.

To study the effect of different parameter choices on automated scoring of sweet potato AFLP data, a 30-profile dataset comprising *Ipomoea* accessions was used. This dataset contained a small number of known replicates so that the replicate error rate could be calculated. In addition, a much larger dataset of 25 replicate pairs was analysed ($n = 6110$ individual pair-wise comparisons) to determine the average sizing error between truly identical fragments.

2.8.1 AIMS

The aims of the research were to: (1) to determine the optimum automated AFLP scoring parameter settings for sweet potato, (2) determine if there are global optimum parameter settings for automatic scoring of AFLP data to maximise phylogenetic resolution, or if these parameters are data dependent[3], (3) determine how robust the resulting phylogeny is to changes in automated scoring parameters, and (4) stimulate more studies of automated scoring of AFLP data and encourage improvements to available software.

---

[3] In collaboration with Dr Heidi Meudt (Museum of New Zealand Te Papa Tongarewa) and Dr Barbara Holland (Massey University), the procedure described below for optimising automated AFLP scoring parameters was also performed on a 30-profile dataset comprising multiple species from the alpine plant genus *Ourisia*. The inclusion of a second dataset allowed comparisons of datasets of the same size, and allowed us to test whether there are global optimum parameter setting for all AFLP datasets. Only the *Ipomoea* results are presented here unless a comparison of the two genera is directly relevant. See Holland *et al.* (2008) (Appendix 8) for complete results.

2.9 OPTIMISING AFLP SCORING PARAMETERS IN SWEET POTATO: MATERIALS AND METHODS

2.9.1 SAMPLING STRATEGY

The 30-profile dataset used to optimise scoring parameters for sweet potato is a subset of the 300+ dataset used in Chapter Four. The dataset was reduced to 30 AFLP profiles to give a representative sample of the larger dataset, and to decrease the time involved in preparing each of the character matrices and running the resulting analyses (see below). The *Ipomoea* dataset included accessions from the extremes of the species' range (from South America to New Caledonia) and accessions from regions where we would expect high genetic diversity (such as northern South America; Zhang *et al.*, 2000a) and low genetic diversity (such as Eastern Polynesia; Green, 2005). This sampling strategy resulted in an *Ipomoea* dataset containing 25 unique accessions *Ipomoea* (24 accessions of *I. batatas* and one accession of the outgroup *I. tiliacea*) and 5 replicates (Table 2.6). AFLP profiles for all replicate pairs were obtained from replicate DNA extractions of the same leaf tissue (see Chapter Four).

Table 2.6
30-Sample Dataset for Optimising AFLP Scoring Parameters

Species	Locality	Accession[a]	Name used in figures[b]
I. batatas	Cook Islands: Mangaia	135	135 Cook Islands
I. batatas	Cook Islands: Mangaia	136	136 Cook Islands
I. batatas	New Zealand Commercial ('Toka Toka Gold')	157-1	157 cv. Toka Toka Gold A
I. batatas	New Zealand Commercial ('Toka Toka Gold')	157-2	157 cv. Toka Toka Gold B
I. batatas	New Zealand Commercial ('Mary Anne')	158	158 cv. Mary Anne
I. batatas	New Caledonia: Balade	Y-411	Y-411 New Caledonia
I. batatas	Fiji	Y-427	Y-427 Fiji
I. batatas	Cook Islands: Aorangi	Y-484	Y-484 Cook Islands
I. batatas	Cook Islands: Rarotonga	Y-485	Y-485 Cook Islands
I. batatas	Cook Islands: Mangaia	Y-491	Y-491 Cook Islands

Table continued next page…

Table continued...

Species	Locality	Accession[a]	Name used in figures[b]
I. batatas	French Polynesia: Society Islands: Moorea	Y-528	Y-528 Society Is.
I. batatas	French Polynesia: Society Islands: Borabora	Y-532	Y-532 Society Is.
I. batatas	French Polynesia: Society Islands: Borabora	Y-535	Y-535 Society Is.
I. batatas	French Polynesia: Society Islands: Raiatea	Y-537	Y-537 Society Is.
I. batatas	French Polynesia: Society Islands: Raiatea	Y-539	Y-539 Society Is.
I. batatas	French Polynesia: Society Islands: Moorea	Y-545	Y-545 Society Is.
I. batatas	French Polynesia: Marquesas Islands: Nukuhiva	Y-556	Y-556 Marquesas Is.
I. batatas	Peru: Piura	Y-594	Y-594 Peru
I. batatas	Peru: Piura	Y-595	Y-595 Peru
I. batatas	Peru: Piura	Y-597-1	Y-597 Peru A
I. batatas	Peru: Piura	Y-597-2	Y-597 Peru B
I. batatas	Peru: Chipillico	Y-610	Y-610 Peru
I. batatas	Peru: Chira	Y-622-1	Y-622 Peru A
I. batatas	Peru: Chira	Y-622-2	Y-622 Peru B
I. batatas	Peru: Recuay: Ancash	Y-662	Y-662 Peru
I. batatas	Colombia: Palmira	Y-680-1	Y-680 Colombia A
I. batatas	Colombia: Palmira	Y-680-2	Y-680 Colombia B
I. batatas	Ecuador: San Horca	Y-695	Y-695 Ecuador
I. tiliacea	–	K233-1-1	K233-1 I. tiliacea A
I. tiliacea	–	K233-1-2	K233-1 I. tiliacea B

[a] All *Ipomoea* accessions are live collections at the Sweet Potato Breeding Lab, National Institute of Crop Science, Tsukuba, Japan except 157 and 158, which are at Plant & Food Research, Pukekohe, New Zealand.
[b] "A" and "B" after a name denote the two replicates AFLP profiles for the same specimen.

2.9.2 GENERATION OF RAW AFLP DATA

AFLPs were generated as described in Chapter Four. Briefly, DNA was digested with the restriction enzymes *Eco*R I and *Mse* I. Eco and Mse linkers were ligated to the restriction fragments and a subset of these were amplified using Eco+A and Mse+C preselective PCR primers. Selective amplifications were performed with four Eco+3/Mse+3 PCR primer combinations. Eco+3 primers were labelled with 6FAM™ (Sigma–Aldrich), VIC®, NED™ or PET™ (Applied Biosystems) fluorescent dyes. The fluorescently-labelled selective amplification products were poolplexed, along with a

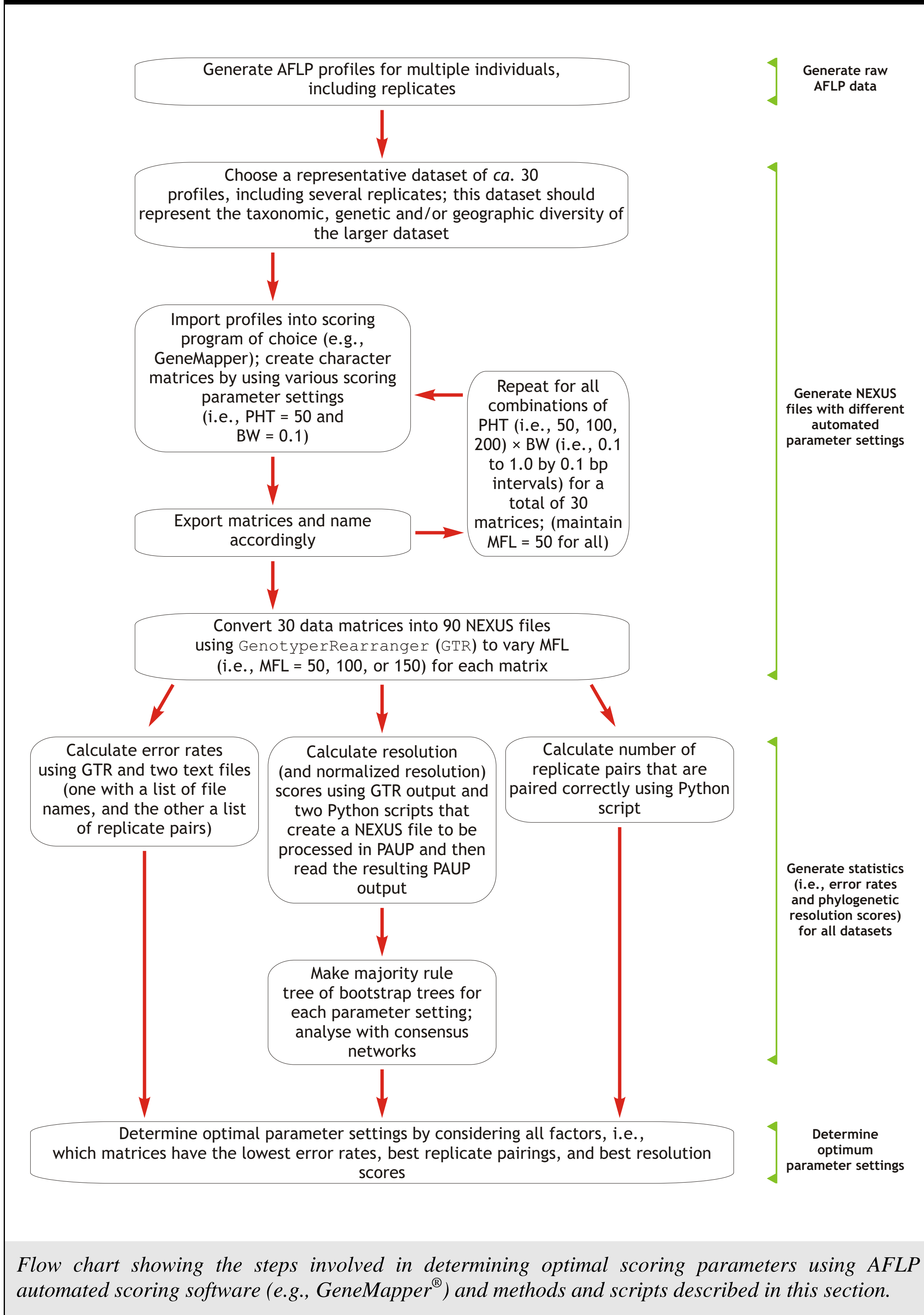
GeneScan™-500 LIZ™ size standard, on a 3730 Genetic Analyzer (Applied Biosystems). Capillary electrophoresis was carried out at the Allan Wilson Centre Genome Service, Massey University.

2.9.3 GENERATION OF DATASETS USING DIFFERENT AUTOMATED SCORING PARAMETER SETTINGS

We designed, and describe below, a procedure to optimise numerous automated scoring parameters. A flow chart showing the different methods used to investigate and optimise parameter settings is shown in Fig. 2.12. From the 30-profile dataset, 90 different character matrices were created from GeneMapper and 36 different character matrices were created from GeneMarker. For computational reasons, it was not feasible to do a complete test of all parameter settings available in each of the automatic scoring software packages. Preliminary testing was performed to determine which parameters were most important. The most important parameters to be subsequently tested here were: for GeneMapper, peak height threshold (PHT), minimum fragment length (MFL), and bin width (BW), and for GeneMarker, PHT, MFL, stutter peak filter (SPF), and local and global detection percentages (LGDP).

A number of other parameters did not have a large effect on scoring (e.g., for GeneMapper, “baseline window”, “polynomial degree”, “peak window size”, and “slope threshold”; and for GeneMarker, “AFLP unconfidence at right-hand side”). These parameters, and all others, were subsequently left at their default values for all analyses. Other parameter settings had a detrimental effect. For example, with respect to smoothing, “heavy” (GeneMapper), “enhanced” (GeneMarker), or “no smoothing” (both programs) all performed worse than the middle option of “light smoothing” (GeneMapper) or “smoothing” (GeneMarker) which we used here. In addition, in GeneMarker, the minimum peak score default of “fail < 1 check < 7 pass” performed worse than other settings with the second value less than 7. Therefore, the minimum peak score was set to “fail < 1 check < 1 pass” in which peaks below a score of 1 were discarded, and those above the score of 1 were automatically accepted, thus fully automating the scoring process.

Fig. 2.12
Flow Chart of Scoring Parameter Optimisation Process



Flow chart showing the steps involved in determining optimal scoring parameters using AFLP automated scoring software (e.g., GeneMapper®) and methods and scripts described in this section.

The 90 GeneMapper matrices (3 × 3 × 10) were obtained by setting PHT to 50, 100 or 200 relative fluorescence units (rfu), setting MFL to 50, 100 or 150 bp, and adjusting BW from 0.1 to 1.0 bp in increments of 0.1. (In GeneMapper, the allele calling threshold was set to the same value as the peak height threshold.) These settings were chosen to test a wide range of possible values for bin width, and to test commonly used values for peak height threshold and minimum fragment length. The 36 GeneMarker matrices (3 × 3 × 2 × 2) were obtained by setting the PHT to either 50, 100 or 200 rfu, the MFL to either 50, 100 or 150 bp, SPF to either its default of 5% or turned off, and LGDP to its default of 1% (both local and global) or turned off. These settings were chosen to test commonly used values for PHT and MFL, and to test the effect of SPF and LGDP. Note that the PHT and MFL parameters are common to both GeneMapper and GeneMarker. The algorithms used by GeneMarker automatically allocate different BW to different characters in the matrix, as opposed to setting one BW for all characters in the matrix in GeneMapper. Although BW in GeneMarker can be subsequently changed so that all characters have identical values, doing so does not appear to greatly alter the number of characters in the matrix or resulting character states, and thus we did not consider BW further in GeneMarker.

Data matrices were exported from both programs and converted into NEXUS format files using the program GENOTYPER REARRANGER (GTR) by Warwick Allen (which is available, along with a detailed protocol, from http://awcmee.massey.ac.nz/aflp/aflp.html). All NEXUS files are in Appendix 18. Creating 126 matrices for the 30-sample dataset via this streamlined process took approximately four hours.

2.9.4 COMPARISON OF DATASETS TO DETERMINE OPTIMAL PARAMETER SETTINGS

2.9.4.1 Measures of Accuracy

For each character matrix, the resolution score and the number of parsimony informative characters were recorded. All phylogenetic analyses were carried out in PAUP* version 4.0b10 (Swofford, 2003) using both neighbour-joining (NJ) on uncorrected distances and with heuristic search using the parsimony optimality criterion (retaining all default settings in PAUP*). Both methods gave congruent trees with some local differences, but NJ gave higher resolution scores for all but 6 of the 252 parameter setting combinations tested (resolution was an average of 11.5% higher). For simplicity, only the NJ resolution scores are reported. To calculate the resolution score for each character matrix, 100 repetitions of 100 bootstrap replicates were performed. For each replicate, all the bootstrap scores over 50% were summed and this number was divided by 27 (each dataset had 30 samples so there were a maximum of 27 internal edges in each tree) to give a value between 0 and 100%. The mean and standard deviation of the resolution score over the 100 repetitions were then calculated.

It is expected that both the quality of characters and the number of characters will have an effect on resolution and accuracy. If two datasets contain characters of the same quality, the dataset with the most characters should give a more accurate phylogenetic estimate; if two datasets contain the same number of characters, the dataset with the highest quality characters should give a more accurate phylogenetic estimate. To try and disentangle these two effects and to get a measure of character quality independent of the total number of characters, a normalised resolution score was used. The number of characters c_{min} in the smallest character matrix was recorded (the datasets with PHT 200, MFL 150 bp, SPF 5%, LGDP 1% from GeneMarker). For each combination of parameter settings, 100 new character matrices were created by sampling c_{min} columns of the original character matrix without replacement, thus creating many datasets of the same length. For each of these character matrices the resolution score was calculated as outlined above. The mean and standard deviation of this normalised resolution score over the 100 resampled alignments was also calculated.

As another measure of accuracy, for each character matrix the number of replicate pairs that were correctly assigned (i.e., as sister to one another) were calculated. For each character matrix the replicate error rate (Bonin *et al.*, 2004; Pompanon *et al.*, 2005) was calculated as

$$\frac{N_{(0,1)} + N_{(1,0)}}{N_{(0,0)} + N_{(1,0)} + N_{(0,1)} + N_{(1,1)}},$$

where $N_{(0,0)}$ and $N_{(1,1)}$ represent the number of correct calls where a replicate pair both have no peak or both have a peak respectively; and, $N_{(0,1)}$ and $N_{(1,0)}$ represent the number of incorrect calls where one half of the replicate pair has a peak and the other half does not. Each category is summed over all the replicate pairs in the data. This error rate is effectively the average Hamming[4] distance between replicate pairs. In the handful of animal and plant AFLP studies that have published them (see Bonin *et al.*, 2004 and references therein), error rates have been calculated to be 2–5% using the above equation. However, the way in which this error rate is calculated makes inter-study comparisons very difficult. First, the formula includes a term for (0,0) calls in the denominator, which means that if the dataset contains many plus-alleles that are not present in a given replicate pair (i.e., scored as (0,0) in the replicates), then the error rate will appear to be lower. Thus, as more data are added to the dataset or very divergent taxa are included, new, unique characters will be introduced and the apparent error rate will decrease. Second, the error rate includes both errors in the raw AFLP profiles themselves (e.g., PCR errors) as well as scoring errors. The number of scoring errors will vary widely between AFLP studies depending on whether a manual, semi-automated or automated scoring procedure is employed, and which scoring software is used.

To check whether the size of the dataset changed the error rate, the error rates for the *Ipomoea* dataset containing 313 samples was calculated, using optimised parameter settings. To check if an increased number of (0,0) calls was masking an increased rate of (1,0) or (0,1), an alternative error rate was defined (effectively the average Jaccard distance between replicate pairs),

$$\frac{N_{(0,1)} + N_{(1,0)}}{N_{(1,0)} + N_{(0,1)} + N_{(1,1)}}.$$

[4] The Hamming distance is incorrectly referred to as the Euclidean distance in Holland *et al.* (2008).

To determine the relative contributions to the error rate of (1) errors in the raw AFLP profiles, and (2) errors introduced during the automated scoring process, the program REPLICATEERROR by Warwick Allen was used (which is available, along with a detailed protocol, from http://awcmee.massey.ac.nz/aflp/aflp.html). REPLICATEERROR approximates the manual editing process by locating errors in a replicate pair (i.e., (0,1) or (1,0)) and testing to see if they can be corrected to (1,1) or (0,0) according to a set of predefined rules. REPLICATEERROR detects three common types of (0,1) and (1,0) scoring errors: first, if a peak is detected but, because it falls below the PHT, it is scored as 0; second, if a peak is detected and is above the PHT, but because it does not meet all required quality criteria (e.g., peak shape), it is scored as 0; and third, if (0,1) and (1,0) errors are adjacent characters that are less than 0.5 bp apart. This third error type is caused by identical fragments that are only slightly different in length (due to random error) being binned as separate characters. See below for justification of why 0.5 bp is an appropriate range over which to amalgamate characters.

A number of scripts were written in Python to streamline the process of analysing the PAUP* output and producing resolution scores and normalised resolution scores. (The Python scripts were written by Dr Barbara Holland, and are available, along with detailed instructions for running them, from http://awcmee.massey.ac.nz/aflp/aflp.html.) The script to calculate error rates requires the set of NEXUS files created by GTR, a text file with a list of the filenames, and a text file with the list of replicate pairs. There are two scripts to calculate the resolution scores. The first uses the files that result from GTR and creates a NEXUS file to be processed by PAUP*. After PAUP* has been run, the second script reads in the resulting bootstrap trees and calculates the resolution scores. A similar process is used to calculate the normalised resolution scores. For the 30-sample dataset, the NEXUS file used to compute the bootstrap trees for calculating the normalised resolution scores for the 126 character matrices took approximately three hours to execute in PAUP* and the Python script took a further 30 minutes (on a Pentium 4). The resolution scores took less than an hour to compute.

2.9.4.2 Optimal Parameter Settings

For each program, optimal parameter settings based on the above analyses of our two datasets were determined. For GeneMapper, to visualise how the error rate, resolution, and normalised resolution change with increasing bin width, over the 9 possible parameter settings for MFL and PHT were averaged and plotted them for each bin width. For both programs, the results from the larger replicate study (see below) were also considered and trends in each of the measures of accuracy to find the optimal settings for each of PHT, MFL, and BW (GeneMapper) and PHT, MFL, SPF, and LGDP (GeneMarker).

In addition to the phylogenetic-based methods described above, the optimum bin width was independently investigated by analysing a set of 25 pairs of replicate AFLP profiles comprising 6110 pairs of identical fragments from *Ipomoea* and *Ourisia* (Heidi Meudt, unpubl. data; see Holland *et al.*, 2008; Appendix 8), and measuring the average size difference (random error) between the peaks of identical fragments. These peaks are known to represent identical fragments because they are from the same or replicate DNA extractions of the same tissue sample from the same individual.

2.9.4.3 Robustness of the Phylogenies to Changes in Parameter Settings

To assess the robustness of the resulting tree to different parameter settings the majority-rule consensus tree of the bootstrap trees for each parameter setting was constructed. The sets of 90 trees from the GeneMapper character matrices and the sets of 36 trees generated from the GeneMarker character matrices were analysed using consensus networks (Holland *et al.*, 2005) as implemented in SplitsTree 4 (Huson & Bryant, 2006). For the GeneMapper datasets, consensus networks of the 63 majority-rule bootstrap trees corresponding to bin width settings of 0.4 and above were also made. Consensus networks also facilitated topological comparison of trees constructed using datasets scored with the software's default vs. optimised parameter settings, and comparison of trees constructed using optimal parameter settings in GeneMarker vs. GeneMapper.

2.10 OPTIMISING AFLP SCORING PARAMETERS IN SWEET POTATO: RESULTS AND DISCUSSION

2.10.1 MEASURES OF ACCURACY

2.10.1.1 Phylogenetic Resolution

There is a wide variation in phylogenetic resolution depending on choice of parameters, with resolution scores ranging from 37–83% (see Appendix 18 for complete results; Table 2.7 compares the resolution scores for selected parameter settings). Importantly, in both programs, default settings are not optimal with respect to phylogenetic resolution. The highest (and the lowest) resolution scores were found by tuning scoring parameters away from the default settings. Default settings in GeneMapper (PHT 100, MFL 100, BW 1.0; shown in Table 2.7) gives a resolution score of 45%. The highest resolution score for GeneMapper is 83%, with parameter settings PHT 50, MFL 50, BW 0.4. Default settings in GeneMarker (MFL 100, PHT 100, LGDP 1%, SPF 5%; shown in Table 2.7) give a resolution score of 49%. The highest resolution score for GeneMarker is 62%, which occur with parameter settings PHT 50, MFL 50, LGDP off, SPF off.

The normalised resolution scores did not vary as widely as the non-normalised resolution scores (Table 2.7; see Appendix 18 for complete results; Table 2.8 compares scores from selected parameter settings). This result indicates that most of the differences in resolution can be explained by a difference in the number of characters, i.e., as expected the presence of more characters leads to higher resolution. However, this result is not simply a question of “more characters equals more resolution”; the point here is that more phylogenetically informative characters and higher resolution can be obtained in both programs from the *same* raw AFLP data by simply optimising scoring parameters (this is quite different from DNA sequence data where new characters can only be added by obtaining more sequence data). Assuming that higher resolution is correlated with higher accuracy, this means it is not always best to strive

for error free datasets at the expense of throwing away many characters. The approach of Althoff *et al.* (2007) that advocates eliminating all error may actually be counter-productive for phylogenetic applications of AFLP. On the other hand, it must be remembered that resolution is only a proxy for accuracy, and if the error introduced by accepting many lower quality characters is biased in some way then higher support for erroneous clades may occur. This is where additional independent evidence (from morphological or other characters, for example) can play an important role (see below).

Table 2.7
Resolution Scores and Numbers of Parsimony Informative Sites[a,b,c]

		GeneMarker®		GeneMapper®	
MFL[d]	**PHT[d]**	**1%,5%[d]**	**off,off[d]**	**BW[d] = 0.5**	**BW[d] = 1.0**
50	**50**	54% (301)	62% (284)	77% (803)	64% (572)
	100	56% (302)	58% (289)	66% (576)	56% (430)
	200	48% (275)	55% (267)	63% (365)	59% (293)
100	**50**	50% (237)	57% (220)	62% (634)	60% (463)
	100	49% (238)	47% (226)	53% (436)	45% (336)
	200	46% (212)	49% (204)	51% (264)	44% (216)
150	**50**	56% (160)	55% (150)	67% (459)	61% (349)
	100	49% (160)	40% (153)	49% (291)	45% (240)
	200	38% (136)	37% (132)	41% (163)	44% (141)

[a] Representative data from two different bin width settings in GeneMapper and two different detection percentage settings in GeneMarker are shown for all three minimum fragment length and peak height threshold settings (see Appendix 18 for complete results).
[b] Values for the software default settings are underlined; optimal settings are double-underlined.
[c] The maximum standard deviation for any of the resolution scores was 2.91 corresponding to a standard error of 0.29.
[d] MFL = minimum fragment length; PHT = peak height threshold; 1%,5% = default values for local and global detection percentages (LGDP, 1%), and stutter peak filter (SPF, 5%); off,off = LGDP and SPF turned off; BW, bin width.

Table 2.8
Normalised Resolution Scores[a,b,c]

		GeneMarker®		GeneMapper®	
MFL[d]	**PHT[d]**	**1%,5%[d]**	**off,off[d]**	**BW[d] = 0.5**	**BW[d] = 1.0**
50	**50**	46%	49%	46%	43%
	100	46%	46%	44%	42%
	200	40%	41%	44%	45%
100	**50**	47%	50%	44%	42%
	100	43%	43%	41%	39%
	200	40%	41%	40%	39%
150	**50**	55%	53%	45%	43%
	100	46%	38%	39%	41%
	200	39%	37%	35%	41%

[a] Representative data from two different bin width settings in GeneMapper and two different detection percentage settings in GeneMarker are shown for all three minimum fragment length and peak height threshold settings (see Appendix 18 for complete results).
[b] Values for the software default settings are underlined; optimal settings are double-underlined.
[c] The maximum standard deviation for any of these values was 6.05 corresponding to a standard error of 0.61.
[d] For abbreviations, see Table 2.7.

One interesting aside is that, in general, GeneMapper gives better non-normalised phylogenetic resolution than GeneMarker, but worse normalised resolution. Thus, comparing data matrices from the two programs with identical MFL and PHT settings (and using the default settings for SPF and LGDP in GeneMarker, and BW 0.5 in GeneMapper) shows that GeneMapper data matrices contain from 1.2 to 2.9 times as many parsimony informative characters as the equivalent GeneMarker dataset. This implies that GeneMarker creates character matrices with higher quality characters than GeneMapper, but because the GeneMapper datasets contain more characters they give more highly-resolved trees.

2.10.1.2 Correct Assignment of Replicates

The number of replicate pairs that were correctly grouped as sister taxa in both *Ipomoea* and *Ourisia* (Appendix 18) provide evidence that BW values below 0.4 are not optimal. In addition, in GeneMapper using a PHT of 100 rfu always gave more correctly assigned replicate pairs than PHT 50 or 200 rfu, although PHT 50 rfu was almost as good (for both datasets). For the *Ipomoea* dataset, using GeneMapper, bin widths of 0.4 or 0.5 gave the most correct replicate pairs — 3.33 (out of 5) on average. Using GeneMarker, all settings with PHT 100 or 200 gave only two correct replicates out of five, settings with PHT 50 gave three correct replicate pairs (9 times) or two correct replicate pairs (3 times). Many settings incorrectly group the three replicate pairs 157 cv. Toka Toka Gold A/B, Y-622 Peru A/B, and Y-680 Colombia A/B. In fact the replicate pair 157 cv. Toka Toka Gold A/B was never correctly recovered by either program for any parameter settings — one or other of the pair always grouped more closely with '158 cv. Mary Anne'. The cultivar 'Mary Anne' is a recently-derived vegetative mutant of 'Toka Toka Gold' (Steve Lewthwaite, pers. comm.), so it is perhaps not surprising that these two cultivars are indistinguishable based on AFLP. Analysis of the distance matrices (data not shown) shows that for each of the three replicate pairs that sometimes group incorrectly there is a third taxon that is also genetically very close. In contrast, in the *Ourisia* data, where the replicates are usually all correctly assigned, the distance between replicates is in the same range as in the *Ipomoea* data, but there aren't any other taxa that are genetically very close to the replicates. This is also seen in the *Ipomoea* dataset in replicates of the outgroup, *I. tiliacea* K233-1 A and B, which are always grouped correctly in all analyses.

The *Ipomoea* dataset comprises cultivars of a single species plus the outgroup, and it appears that there is insufficient signal in the AFLP data to distinguish some ingroup accessions. In *Ipomoea* (for the optimal GeneMarker dataset) the distances between replicates were: 0.07, 0.08, 0.09, 0.10, and 0.20; the distances between a replicate and its closest non-replicate ranged from 0.08 to 0.33, with a median of 0.11. In *Ourisia* the distances between replicates were: 0.07, 0.07, 0.09, 0.11, 0.12, and 0.23; but in contrast to *Ipomoea*, the distances between a replicate and its closest non-replicate ranged from 0.13 to 0.34, with a median of 0.22.

2.10.1.3 Error Rates

The error rates range from 9–18% (see Appendix 18 for complete results; Table 2.9 shows the replicate pair error rates for both programs at selected parameter settings). The observed error rates are higher than those previously reported for AFLP datasets of 2–5% (see Bonin *et al.*, 2004 and references therein). However, as discussed above (see above), we should be cautious regarding inter-study comparisons of error rates because the error rates may be affected by the level of divergence among the individuals included in the study, the number of individuals in the dataset, and both errors resulting from the raw AFLP profiles themselves (e.g., PCR errors) and those resulting from the scoring process (and the type of procedure and software employed).

Table 2.9
Replicate Error Rates for Representative Datasets[a,b]

		GeneMarker®		**GeneMapper®**	
MFL[c]	**PHT[c]**	**1%,5%[c]**	**off,off[c]**	**BW[c] = 0.5**	**BW[c] = 1.0**
50	**50**	12%	11%	15%	14%
	100	13%	13%	14%	13%
	200	14%	13%	15%	14%
100	**50**	11%	10%	15%	13%
	100	13%	13%	14%	12%
	200	14%	14%	16%	15%
150	**50**	11%	9%	15%	13%
	100	13%	13%	16%	13%
	200	15%	15%	18%	16%

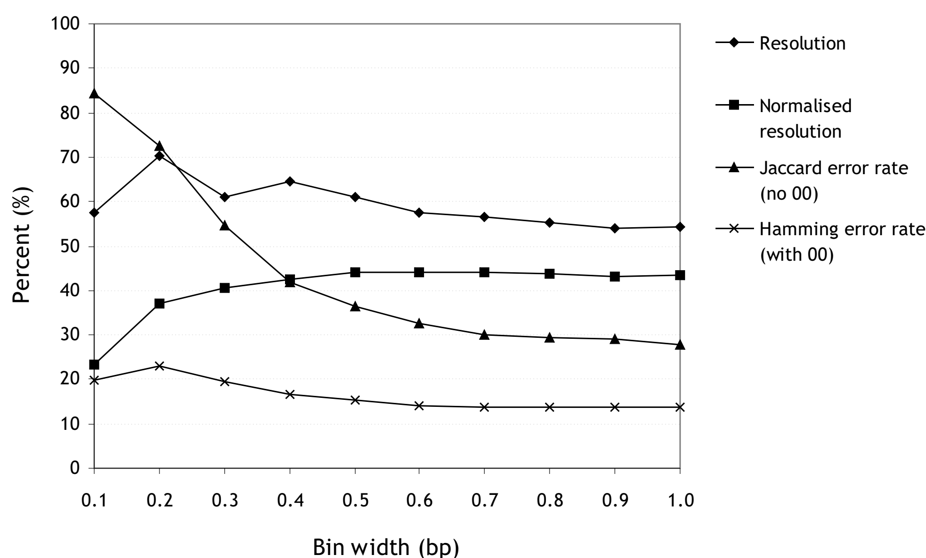
[a] Representative data from two different bin width settings in GeneMapper and two different detection percentage settings in GeneMarker are shown for all three minimum fragment length and peak height threshold settings (see Appendix 18 for complete results).
[b] Values for the software default settings are underlined; optimal settings are double-underlined.
[c] For abbreviations, see Table 2.7.

It appears that all of these factors have affected the error rates in the *Ipomoea* and *Ourisia* datasets, some to a greater degree than others. First, the largely intraspecific *Ipomoea* datasets give higher error rates overall than the interspecific *Ourisia* datasets, which suggests that lower divergence among samples results in higher error rates. This

was tested further by comparing the standard (Hamming) and alternative (Jaccard) error rates. For the GeneMapper datasets, standard error rates in the *Ipomoea* dataset are higher for smaller BW than larger BW (Fig. 2.13), as would be predicted, whereas error rates are almost constant in the *Ourisia* datasets. To check if the difference in error rates and the difference in this trend in error rates in the *Ipomoea* and *Ourisia* datasets were due to an increased number of (0,0) calls was masking an increased rate of erroneous (1,0) or (0,1) calls, the alternative (Jaccard) error rate was calculated. As shown in Fig. 2.13 this alternative error rate decreases sharply from a bin width of 0.1 to 0.5 after which it flattens off at around 30% (the same pattern was observed in *Ourisia*). The *Ipomoea* dataset has an average standard error rate of 15% compared to 10% for the *Ourisia* dataset. However, when the alternative error rate is used, the average error rate for *Ipomoea* is 38% compared to 40% for *Ourisia*, suggesting that the apparently higher standard error rate in *Ipomoea* may not be 'real' but is instead a result of fewer (0,0) calls in the denominator.

Secondly, the higher error rates found in this study are also partly due to the small nature (30 individuals) of the dataset. The recalculated the error rates for the 313-sample dataset with optimal parameter settings were indeed lower compared to the 30-taxon dataset; the error rate dropped from 15% to 9% (GeneMapper) and from 10% to 9% (GeneMarker). Nevertheless, even though the error rates are lower when many more individuals are included, they are still not within the range reported in Bonin *et al.* (2004), which suggests that the nature of the errors is also an important factor.

Fig. 2.13
Phylogenetic Resolution and Error Rates versus Bin Width



Resolution, normalised resolution, and error rate versus bin width for the GeneMapper® analysis. All measures have been averaged over the 9 possible settings for minimum fragment length (MFL) and peak height threshold (PHT).

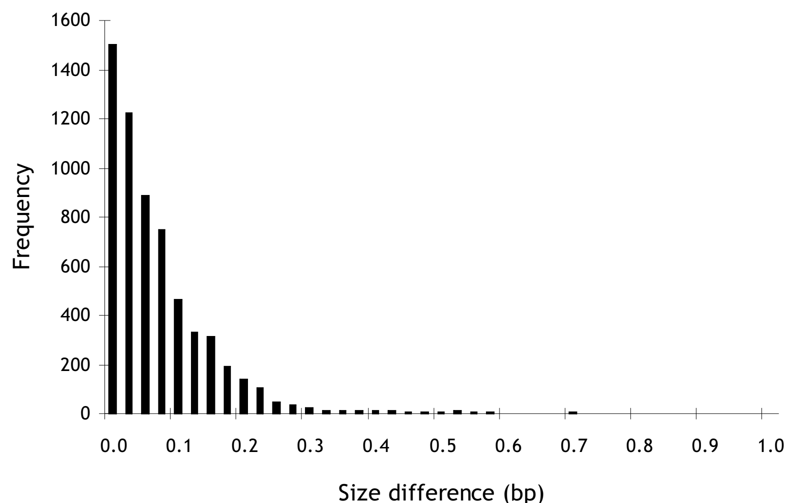
Finally, using the program REPLICATEERROR, we were able to lower the error rate in the optimised GeneMapper dataset from 15% (uncorrected) to 5% (corrected). This suggests that the majority of errors (the difference between the corrected and uncorrected rates) in the dataset are scoring errors, whereas the resulting corrected error rates approximate the number of PCR errors. Because REPLICATEERROR only locates errors between replicates, it is not possible to use this program to reduce errors in the dataset as a whole, but it does indicate that there is significant potential to improve the accuracy of automated scoring.

In summary, it is likely that the higher error rates are due in large part to the combined effects of smaller datasets and a fully automated scoring procedure. Including the (0,0) term in the denominator has a significant effect on error rates, and makes comparison of error rates between datasets unreliable, especially if they contain different numbers of taxa and/or taxa with varying amounts of genetic diversity. The analyses using REPLICATEERROR revealed that the error rate was significantly increased

by errors introduced during the automated scoring process. Although it could be argued that this result supports manual scoring (or at least manual editing of automatically scored data), we still think that automated scoring is preferable because it is more time efficient, makes it easier to maintain consistency in large datasets, and removes both subjectivity and the potential to introduce bias into datasets.

Finally, Fig. 2.14 shows the results of the larger replicate study, which demonstrates that almost all identical fragments fall within 0.4 bp of each other. This suggests that lowering the bin width setting below 1 bp could be beneficial as it would introduce only a small number of extra errors in the character matrix, and may help distinguish between non-identical fragments that differ in mobility by less than 1 bp. There is some measurable distance even between identical fragments, which means it is always possible for truly homologous fragments to fall into different bins.

Fig. 2.14
Size Difference Between Identical Alleles of Replicate Pairs



The size difference between 6110 pairs of identical alleles (fragments) from 25 replicate AFLP profiles was measured to determine the average sizing error between truly identical fragments. Together with error rates and resolution measures, these results help to determine the optimum bin width. The mean size difference between identical alleles was 0.08 bp, with 99% of values falling within 0.42 bp, and 99.9% falling within 0.66 bp. These results suggest that bin width can be set to less than 1.0 bp, without greatly increasing the risk of splitting identical alleles into separate characters. These values are for the experimental set-up we used (see methods), and may need to be determined empirically for other capillary instruments.

Using the empirical distribution of observed distances between identical fragments (Fig. 2.14), a simulation was performed to estimate how many errors of this kind we would expect for different bin width settings. For each simulated pair of identical fragments it was assumed that the position of the leftmost fragment of the pair was uniformly distributed within the bin, the distance to the rightmost fragment of the pair was then sampled from the empirical distribution (Fig. 2.14). It was then recorded whether the rightmost fragment was still in the same bin. This was repeated for 1,000,000 simulated fragments for each bin width between 0.1 and 1.0 (in steps of 0.1). The proportion of expected errors of this kind is shown in Table 2.10. Table 2.10 shows that if the bin width drops below 0.5 bp the error rate starts to rise steeply.

Table 2.10
Binning Error Simulation for Identical Fragments

Bin Width	Errors[a]
0.1	0.59
0.2	0.37
0.3	0.26
0.4	0.20
0.5	0.16
0.6	0.14
0.7	0.12
0.8	0.10
0.9	0.09
1.0	0.08

[a] Each error value is a proportion based on 1,000,000 random fragment pairs.

2.10.2 OPTIMAL PARAMETER SETTINGS

2.10.2.1 GeneMapper®

In general, optimal parameter settings in GeneMapper were PHT 50, MFL 50, and BW 0.5. For the datasets generated using GeneMapper, we can see how the error rate, resolution, and normalised resolution change with increasing bin width (Fig. 2.13). The

values in these figures have been averaged over the 9 possible parameter settings for MFL and PHT. Resolution peaks at a bin width of 0.2 bp with a smaller peak at 0.4 bp. In spite of this, I do not recommend setting the bin width so low, as the high resolution scores may be an artefact of having many unreliable characters.

Instead, I propose that a BW of 0.5 bp is optimal for both datasets, for the following reasons. The normalised resolution scores show that the phylogenetic quality of the characters is increasing up to a bin width of 0.5 bp after which it is fairly stable. Taken together, the results from the larger replicate study (Fig. 2.14), the assignment of replicate pairs, and the trend in the normalised resolution and alternative error rate (Fig. 2.13) suggest that when scoring ABI 3730-derived AFLP data using GeneMapper, lowering the BW setting to 0.5 bp is beneficial. Although some errors are introduced, this is outweighed by the positive effect of having more characters. Decreasing the bin width below 0.5 bp results in even more characters, but splits apart too many characters that are identical plus-alleles of the same locus.

A PHT setting of 50 appears to be optimal in both datasets. Although PHT 50 gives slightly higher replicate error rates in GeneMapper than the higher settings (Table 2.9), in most cases the resolution scores (Table 2.7) and normalised resolution scores (Table 2.8) are better for both datasets at this setting — especially at a bin width setting of 0.5. In contrast, the optimal setting for MFL appears to differ between the two datasets. The highest resolution (Table 2.7) and normalised resolution (Table 2.8) scores are found at MFL 50. The error rate (Table 2.9) is not greatly affected by the choice of MFL.

2.10.2.2 GeneMarker®

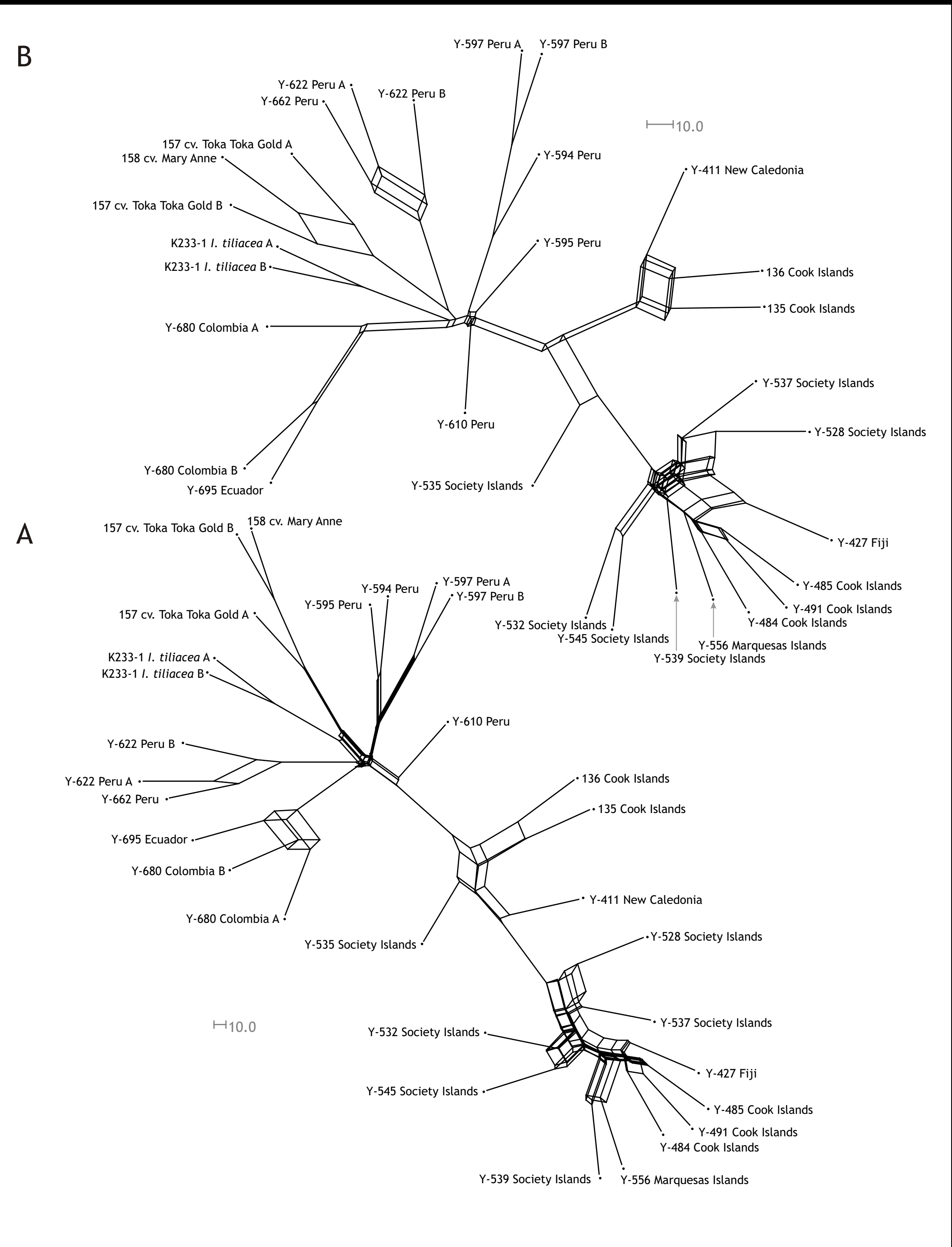
Optimal parameter settings in GeneMarker were PHT 50, MFL 50, and SPF off. Resolution, normalised resolution and number of parsimony informative sites mostly increase with decreasing PHT (Table 2.7 and Table 2.8). In most cases, for both datasets, setting the PHT at 50 gives better resolution than setting it at 100 or 200 (Table 2.7). Error rates decrease with decreasing PHT (Table 2.9). For MFL, the highest resolution was found at 50 bp. Error rates were not affected by MFL (Table 2.9).

Turning off the stutter peak filter (SPF) caused a marked increase in resolution, but this could be artefactual. For instance, by including all the stutter peaks, support for some splits could be inflated. In contrast, LGDP had only a small effect on the resolution of the resulting trees.

2.10.3 ROBUSTNESS OF THE PHYLOGENIES TO CHANGES IN PARAMETER SETTINGS

Consensus networks are a very useful way to visualise the robustness of the phylogenies to AFLP scoring parameter settings and to different software (in this case, GeneMapper and GeneMarker), and specifically examine the difference in phylogeny reconstruction of default vs. optimised settings. Fig. 2.15A shows the consensus network of the 63 GeneMapper majority-rule bootstrap consensus trees corresponding to bin width settings of 0.4 and above, and Fig. 2.15B shows the consensus network of all 36 GeneMarker majority-rule bootstrap consensus trees. Significantly, many parts of the phylogenies in both Fig. 2.15 are robust to parameter choice, and where the trees do differ the boxes in the consensus networks indicate specific areas of conflict due to local rearrangements rather than taxa shifting their position in the tree dramatically. One exception to this is the datasets from GeneMapper with low bin width settings. As indicated by Fig. 2.13 and discussed above, setting the bin width lower than 0.4 probably creates many errors in the character matrix and may lead to the reconstruction of inaccurate trees. Indeed, this appears to be the case, as the consensus network constructed for the majority-rule bootstrap trees for all 90 GeneMapper datasets including those with BW below 0.4 (Appendix 18) show much more conflict than Fig. 2.15. In spite of this, the consensus networks encouragingly show that regardless of parameter settings, the datasets are converging on very similar topologies whose accuracy is corroborated by independent sources of data (see below).

Fig. 2.15
Consensus Networks of Different AFLP Parameter Settings



A. *Consensus network of the 63 GeneMapper® majority-rule bootstrap consensus trees that correspond to parameter settings with bin width > 0.3 (i.e. 0.4 to 1.0), showing all splits (edges) that appear in any of the 63 trees.*

B. *Consensus network of the 36 GeneMarker® majority-rule bootstrap consensus trees showing all splits (edges) that appear in any of the 36 trees.*

Parallel edges represent the same split, and edge length is proportional to the number of trees in which that split appears. Boxes represent areas of conflict among the trees generated using different parameter settings.

To investigate whether the two programs are converging on similar topologies, consensus networks comparing the majority rule consensus trees using default parameter settings and using optimised parameter settings for GeneMapper vs. GeneMarker were constructed (Fig. 2.16). It should be emphasise that the intention of this exercise is not to compare the performance of GeneMapper vs. GeneMarker *per se*, but to show the degree of topological congruence. For default settings (Fig. 2.16A) there are two conflicting edges, and for optimised settings (Fig. 2.16B) there are six areas of conflict in the *Ipomoea* dataset (five affecting just single edges and one more complex area of conflict). In all cases, the conflict is confined to a few, local areas of the tree and does not represent large differences between the topologies recovered using the two programs. Therefore, in addition to differing parameter settings, topologies are also robust to software choice. Because very similar topologies were recovered using different software packages that use diverse algorithms and methods, this result provides further evidence that automated scoring of AFLP profiles results in accurate and robust phylogenies.

Finally, consensus networks also show that optimised parameter settings consistently show an increase in the number of edges with > 50% bootstrap support relative to default settings. Fig. 2.17A illustrates the difference in the majority-rule consensus tree between default parameter settings in GeneMapper (PHT 100, MFL 100, BW 1.0) vs. optimised settings (PHT 50, MFL 50, BW 0.5). Edges shown with dashed lines appear only in the trees built with optimised parameter settings. Optimisation of scoring increased the number of internal edges with > 50% bootstrap increased from 14 to 25 (out of a possible maximum of 27) by optimising scoring parameters. Along with these, 11 new edges — that importantly do not conflict with the default setting tree — four additional edges are changed. The same plot was done for GeneMarker default (MFL 100, PHT 100, LGDP 1%, SPF 5%) vs. optimised (MFL 50, PHT 50, LGDP off, SPF off) parameter settings (Fig. 2.17B). Using optimised settings gives five new edges with bootstrap support above 50% that do not conflict with the default setting tree; it also changes four edges. Thus, even though the topologies are largely robust to scoring parameter settings and choice of software, these consensus networks show that optimised parameter settings can improve resolution and increase support for the resulting phylogenetic trees relative to default parameter settings.

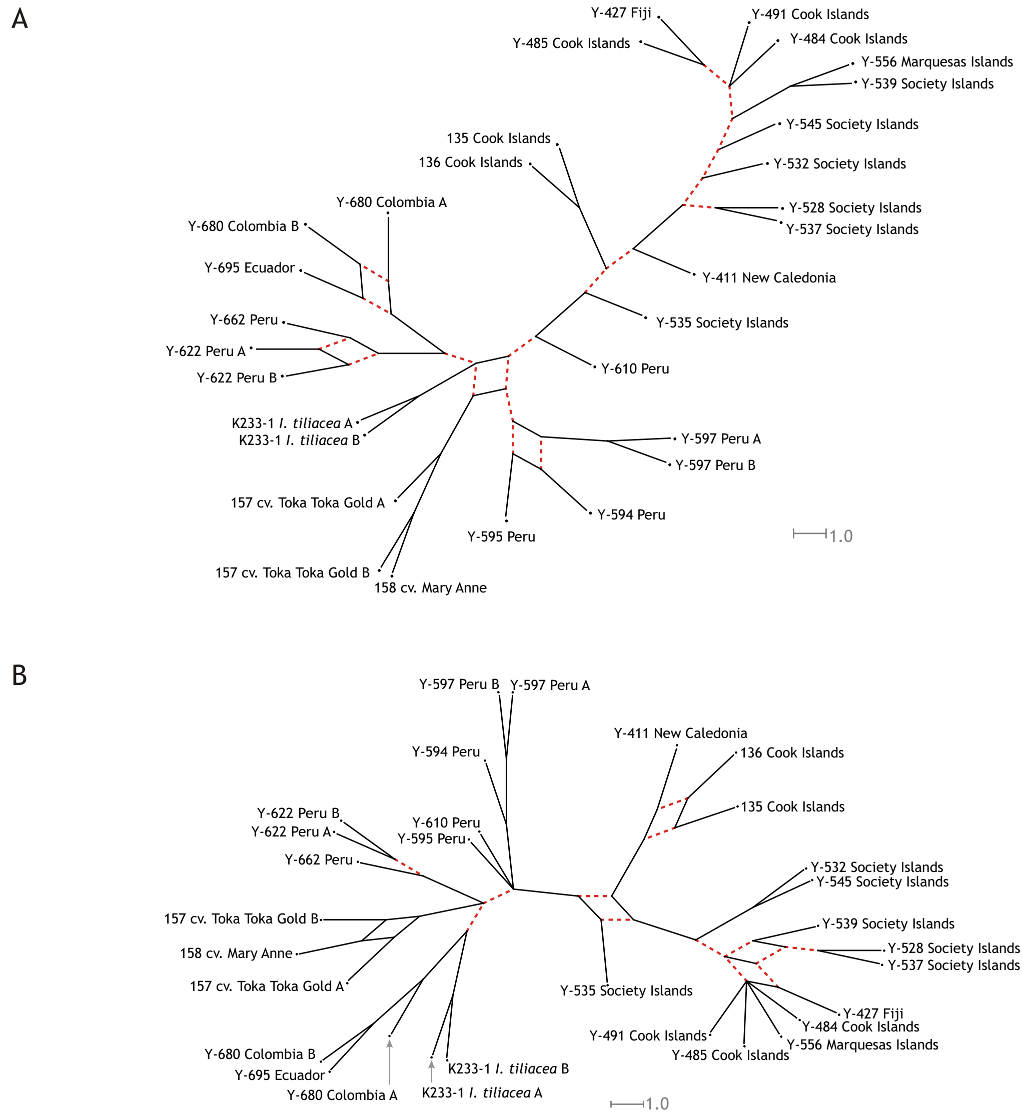
Fig. 2.16
Consensus Networks of GeneMapper® and GeneMarker® Trees

A

B

A. *Consensus network displaying the default GeneMapper and GeneMarker trees.*
B. *Consensus network displaying the optimized GeneMapper and GeneMarker trees.*
Parallel edges represent the same split, and edge length is proportional to the number of trees that split appears in. Boxes represent areas of conflict between the trees generated using the two different programs.

Fig. 2.17
Consensus Networks of Default and Optimised Trees



A. *Majority-rule consensus network of the default and optimised GeneMapper® trees.*
B. *Majority-rule consensus network of the default and optimised GeneMarker® trees.*
Parallel edges represent the same split, and edge length is proportional to the number of trees in which that split appears. The dashed edges indicate splits that only appear in the optimised trees. Boxes represent areas of conflict between the trees generated using default vs. optimised parameter settings.

2.10.4 ADDITIONAL INDEPENDENT EVIDENCE

In addition to resolution scores, another way to investigate the accuracy of trees constructed using the optimum parameter settings is to compare the tree topologies to other sources of data, such as morphological, ecological, or other characters. In the case of sweet potato, the AFLP-derived phylogenies are congruent with genetic, horticultural and anthropological evidence. Firstly, there is a major split (Fig. 2.15 Fig. 2.16 Fig. 2.17) between accessions from Eastern Polynesia/Island Melanesia and accessions from South America, consistent with anthropological and historical evidence (Green, 2005). The South American accessions are then further divided into those from Colombia/Ecuador and those from Peru, consistent with previous genetic work which has separated northern South American germplasm from germplasm farther south in Peru/Chile (Zhang *et al.*, 2000a). It is thought that sweet potato from Eastern Polynesia is derived from a pre-European introduction into the Pacific, whereas the commercial material from New Zealand (incl. 'Toka Toka Gold' and 'Mary Anne') is almost certainly a European-era introduction (Green, 2005). Accordingly, these two groups are separated in all networks (Fig. 2.15 Fig. 2.16 Fig. 2.17). 'Mary Anne' is a purple-fleshed mutant derived from 'Toka Toka Gold' (Steve Lewthwaite, pers. comm.), and indeed these two accessions always group together (with 'Mary Anne' placed between the two 'Toka Toka Gold' replicates; see above). Although the single accession from Fiji is always placed with accessions from Eastern Polynesia, historical evidence reveals that the Fiji accession is in fact a relatively recent introduction from the Polynesian island of Niue (Appendix 4).

2.11 OPTIMISING AFLP SCORING PARAMETERS IN SWEET POTATO: CONCLUSIONS

In any phylogenetic study using AFLP data, the main aim is to recover an accurate species phylogeny. There has been much discussion in the literature regarding appropriate techniques for phylogenetic analysis of AFLP data (Albach, 2007; Bonin *et al.*, 2007; Meudt & Clarke, 2007 and references therein; Appendix 7); by contrast very little has been done with respect to the scoring of AFLP data and the effect this has on downstream analyses. To my knowledge, this study is the first quantitative, objective and thorough investigation on the effect of different automated scoring parameter settings on phylogenetic resolution.

Optimising the parameter settings for automated AFLP scoring significantly increased phylogenetic resolution in this study, allowing relationships to be resolved that were obscured when using default scoring parameters. It is likely that similar improvements in resolution can be obtained in other phylogenetic studies, and automated scoring parameters should be optimised wherever possible. Interestingly, the results show that it is not always best to choose the character matrices with the lowest error rates, as the benefits of increasing the number of characters can outweigh some reduction in character quality.

The optimal settings differed for the *Ourisia* and *Ipomoea* datasets suggesting that, for at least some parameters, there are not universal optimal settings. A peak height threshold of 50 gave good resolution for both datasets and for both programs, but more replicate pairs were correctly assigned with a threshold of 100. The best settings for minimum fragment length varied for the two datasets in terms of both resolution and error rates. In contrast, there is a case for a universally optimal setting for bin width in GeneMapper; several lines of evidence suggest that 0.5 bp is a good choice. This is supported for both datasets by the high resolution scores, the greater number of replicate pairs appearing as sister taxa, and the fact that almost all peaks of truly identical fragments in the replicate study fell within 0.5 bp. However, it should be cautioned that these results are based on analysis of data from two datasets that were run on the same

capillary instrument. Therefore, rather than suggesting particular parameter settings, it is recommended that users of AFLP use a similar procedure to that described here (Fig. 2.12) to investigate the effect of changing AFLP scoring parameters settings on phylogenetic resolution, assignment of replicate pairs, and error rates. As more results are compiled for different datasets it will be possible to determine if there are some universally good settings. Nevertheless, it is clear that reducing bin widths from the default 1.0 bp setting is potentially beneficial.

One potential shortcoming of using the same dataset to optimise parameter settings via bootstrap resolution and to make phylogenetic estimates is that this could upwardly bias support values in the phylogenetic analysis. A way to get around this would be to use only a subset of the taxa of interest to optimise the parameters (as has been done here).

Encouragingly, many parts of the phylogenies were robust to changes in the parameter settings, although local rearrangements did occur. Consensus networks provided a useful tool to visualise which parts of the phylogeny were robust to changes in the parameters. The phylogenetic trees recovered were corroborated by additional evidence, i.e., genetic, horticultural and anthropological evidence. This congruence of AFLP data with various other lines of evidence supports the view that AFLP data can be useful for phylogenetic studies.

Error rates found in this study were higher than those previously reported for datasets generated using semi-automated scoring (2–5%; Bonin *et al.*, 2004). There is evidence that the calculation used to quantify genotyping error rate depends on the number of taxa in the dataset and the genetic distance between them, but these effects are not sufficient to explain the difference in error rates between semi-automated and automated scoring. The results from REPLICATEERROR suggest that the majority of errors were scoring errors (rather than PCR errors). This suggests that the incorporation of improved scoring algorithms into current software packages such as GeneMapper and GeneMarker would further increase their power and usefulness.

Despite error rates that are higher than ideal, automated scoring still produces character matrices that are phylogenetically informative and where most or all replicates are correctly assigned. When parameter settings are chosen carefully, character matrices can be produced using automatic scoring that result in well resolved trees. As well as its application to phylogenetics, it is likely that optimising automated AFLP scoring parameters will provide increased resolution in other important applications of the technique such as linkage mapping and population genetics. In these cases, different measures of accuracy and resolution will be required, although in all applications of the AFLP technique measures of error rate from replicate samples are critical. Future studies should focus on calculating and publishing error rates, optimising parameter settings prior to analysis, improving automated scoring algorithms, as well as taking a step back and thoroughly assessing the appropriateness of AFLP for phylogenetic reconstruction.

CHAPTER THREE

ORIGINS AND DISPERSAL OF THE BOTTLE GOURD IN OCEANIA

Māori Woman Beating Bark, with Two Large Bottle Gourds beside Her

Koriniti, Whanganui River, New Zealand, 1921

Photograph taken during a Dominion Museum ethnographic expedition

Ref. No. PA1-q-257-38-1

Reproduced with permission of the Alexander Turnbull Library, Wellington, New Zealand

3.1 CHAPTER OVERVIEW

The bottle gourd (*Lagenaria siceraria*) was one of the first, most widespread, and most versatile plant species domesticated by humankind, and by at least 2000 BC was grown by peoples in Africa, Asia and the New World (Heiser, 1979a). The bottle gourd was grown by Polynesians by at least AD 1000, although how, when, and from where this species entered the Pacific is largely unknown; this is the focus of the current work.

Literature relevant to the dispersal of the bottle gourd is reasonably extensive but very disparate, and the aim of the introduction is to bring this research together and apply it to questions surrounding the origin of the Polynesian bottle gourd. The introduction begins with an overview of the biology of the bottle gourd, including its taxonomy and reproductive biology. Prehistoric uses of the bottle gourd fruits are described. The evidence for an African origin of this species is presented, and the probable modes of dispersal of this species to Asia and the New World discussed.

In the current research I aim to determine whether the Polynesian bottle gourd is from Asia or the New World. Several authors (e.g., Whistler (1990), Burtenshaw (1999) and Green (2000b, 2005)) suggest a New World, specifically a South American, origin is more likely. A New World origin for the Polynesian bottle gourd has implications for human mobility in the Pacific; human-mediated transfer from South America could have been effected by Polynesian voyagers who reached the west coast of South America, collected the bottle gourd (and sweet potato) and returned to Polynesia with these crop species. Recent evidence supporting prehistoric contact between Hawai‘i and the Channel Islands, California between AD 400 and 800 (Jones & Klar, 2005) is also compatible with human-mediated dispersal of the bottle gourd from North America. A number of alternative hypotheses exist, including natural dispersal from the New World, natural or human-mediated dispersal from Asia, and combinations of these scenarios. The introduction concludes with the hypothesis and aims of the current research.

The materials and methods describe the bottle gourd accessions obtained, and the genetic markers developed to test the hypothesis. The markers comprise the five SCAR markers described in Chapter Two, and two additional cpDNA markers. All seven markers were used to amplify and sequence polymorphic loci from 38 accessions of bottle gourd from Africa, Asia, the New World and Polynesia.

The discussion offers an interpretation of the sequence data, and the contribution of this research to understanding the origins of the bottle gourd in Polynesia. Both the domestication and global dispersal of the bottle gourd have significant implications for the origin of the Polynesian bottle gourd, and the contribution of the current research to these areas is discussed. Particular attention is paid to contemporaneous research carried out by Dr Bruce Smith's group at the Smithsonian Institution, Washington, D.C., the results of which apparently differ from the results described here. Resolving this incompatibility will represent significant progress.

The final section of this chapter brings together the molecular, archaeological, linguistic and morphological research to draw conclusions on the origins of the Polynesian bottle gourd and the global dispersal and domestication of the species. Possible future work is presented in Chapter Five.

3.2 INTRODUCTION

3.2.1 BIOLOGY OF THE BOTTLE GOURD

Fig. 3.1
The Bottle Gourd (*Lagenaria siceraria*)

Māori variety of bottle gourd (hue) from the Auckland region, growing in Otaki, Kapiti Coast during late summer, 1999.

Reproduced with permission of Mike Burtenshaw
Copyright © 1999 by Mike Burtenshaw

The bottle gourd (Fig. 3.1), also known as the white-flowered gourd or calabash, *Lagenaria siceraria* (Molina) Standley[5] is a herbaceous annual vine in the Cucurbitaceae. Except for a single wild accession collected from Zimbabwe in 2004 (Decker-Walters *et al.*, 2004b), the bottle gourd is now known only in cultivation. The wild ancestor of the domesticated bottle gourd is probably of African origin. Even

[5] *Cucurbita lagenaria* L., *Cucurbita siceraria* Molina, *Lagenaria vulgaris* Seringe and *Lagenaria leucantha* (Duchesne) Rusby are all invalid synonyms of *Lagenaria siceraria* (Molina) Standley (Towle, 1961, p. 92; Jeffrey, 1967, pp. 51–52; Heiser, 1973b; Robinson & Decker-Walters, 1997, p. 88).

prehistorically, the bottle gourd was widely cultivated, found throughout the tropics, subtropics, and even some temperate zones, in both the Northern and Southern Hemispheres (Heiser, 1979a, p. 71).

3.2.1.1 Taxonomy, Systematics and Genetics

The Cucurbitaceae contains two subfamilies, eight tribes, 118 genera and approximately 825 species (Jeffrey, 1990, pp. 449–463). The four important cucurbit crops are watermelon (*Citrullus lanatus*), rockmelon/cantaloupe (*Cucumis melo*), cucumber (*Cucumis sativus*), and the pumpkins and squashes (mainly *Cucurbita pepo* and *Cucurbita maxima*). Minor crops include bitter melon (*Momordica charantia*), loofah (mainly *Luffa cylindrica*), wax gourd (*Benincasa hispida*) and bottle gourd itself (see Robinson & Decker-Walters, 1997). The bottle gourd is in the subfamily Cucurbitoideae, tribe Benincaseae, subtribe Benincasinae. Other members of subtribe Benincasinae include the wax gourd and watermelon (Jeffrey, 1990, pp. 457–459). Analyses of cpDNA (Chung *et al.*, 2003; Decker-Walters *et al.*, 2004a) and nrITS variation (Jobst *et al.*, 1998) confirm *Lagenaria*, *Citrullus* and *Benincasa* are closely related.

There are six recognised species in the genus *Lagenaria*: the annual, monoecious cultivated bottle gourd (*L. siceraria*), and five wild, perennial, dioecious species (*L. abyssinica*, *L. breviflora*, *L. guineensis*, *L. rufa* and *L. sphaerica*) (Robinson & Decker-Walters, 1997, p. 88). All five wild species are native to tropical Africa, with *L. sphaerica* extending into the Comoros and Madagascar (Jeffrey, 1967, p. 52). A seventh species, *L. bicornuta* from Ghana (Chakravarty, 1968), has been subsumed into *L. siceraria* (Heiser, 1973b).

The bottle gourd is diploid, possessing 11 pairs of chromosomes ($2n = 2x = 22$) (Singh, 1990, p. 20), and a relatively small genome of 686 Mbp (*Arabidopis thaliana* is 157 Mbp; *Citrullus lanatus* is 441 Mbp; *Ipomoea batatas* is 2205 Mbp) (Bennett & Leitch, 2004). Ploidy levels, chromosome counts and genome sizes for the other species of *Lagenaria* appear not to have been determined. Hybrid plants, however, have been

formed between the bottle gourd and both *L. breviflora* and *L. sphaerica*, although in both cases the offspring have greatly reduced fertility (Heiser, 1979a, pp. 85–87).

Based on morphology, Kobiakova (1930) recognised two subspecies of *L. siceraria*. This classification is supported by Heiser (1973b; 1979a, pp. 90–92), who obtained and grew 180 accessions of bottle gourd from various parts of the world. The subspecies are *L. siceraria* ssp. *siceraria* (the African and American gourds; hereafter abbreviated ssp. *siceraria*) and *L. siceraria* ssp. *asiatica* (the Asian gourds; hereafter abbreviated ssp. *asiatica*). The subspecies taxonomy is also supported by principal component analysis (PCA) of RAPD fingerprinting data (Decker-Walters *et al.*, 2001). The characters that define the subspecies are presented in Table 3.1. Fruit shape was found to be highly variable and particular fruit morphologies did not segregate with the different subspecies (Heiser, 1973b). Decker-Walters *et al.* (2001) also found that some fruit shapes (e.g., bilobal, round and pyriform) have multiple geographic origins; it is likely that common fruit shapes (such as those for water carriers) have been selected independently in both subspecies.

Table 3.1
Defining Characters of the *Lagenaria siceraria* subspecies[a]

Organ	Character	Subspecies	
		***L. siceraria* ssp. *siceraria* (African and American)**	***L. siceraria* ssp. *asiatica* (Asian)**
Leaf	**Margin**	Entire to crenate	Serrate
	Shape	Unlobed or rounded-lobed	Sharply three- to five-lobed
Flower	**Overall size**	Small to medium	Large
	Calyx lobes	Short (2–10 mm) and broad	Long (6–20 mm) and slender
Seed	**Colour**	Usually dark	Light (grey-brown to white)
	Length to width ratio	Usually < 2:1	Usually > 2:1
	Corky wings	Present or absent	Absent
	Longitudinal lines	Present or absent	Present
	Line morphology	Usually glabrous	Sometimes pubescent
	Distal 'ears'	Present or absent	Usually present

[a] Compiled from Heiser (1973a, 1973b; 1979a, p. 92).

3.2.1.2 Reproductive Biology

The bottle gourd is monoecious, with solitary, diclinous (separate staminate (male) and pistillate (female)) flowers. The flowers are white, opening at dusk and closing early the following morning, except on cloudy or overcast days (Heiser, 1979a, p. 72). Plants are self-compatible — that is, pollen from a male flower can pollinate a female flower of the same plant (Heiser, 1979a, p. 72). Heiser suspected the gourds he grew (in North America) were pollinated by striped and spotted cucumber beetles (Heiser, 1979a, pp. 72–74), but was unsure of what pollinated gourds in Africa — the probable centre of origin of this species — and suggested that determining the pollinator in Africa may assist in determining the natural range of the species (Heiser, 1979a, p. 74).

The ovary of the female bottle gourd flower is inferior, so that as the fruit (a pepo) develops the flower is abscised, leaving a flower scar at the distal end of the fruit. Once mature, the exocarp hardens and becomes woody, but the remaining layers of the pericarp remain as soft, watery pith. Over time, water evaporates from the fruit, leaving the hard exocarp (also called a rind or shell) surrounding a hollow interior which contains the loose seeds. The fruits of different bottle gourd cultivars are amazingly diverse, both in size and shape; the length can range from just a few cm to over 2.5 m, and circumference from a few cm (e.g., in the snake gourds) to 1.8 m for the large ovoid types (Heiser, 1979a, p. 75). Fruit shapes include oblate, spheroid, pyriform, dipper-shaped (i.e., an elongated neck at the proximal end), bilobal (i.e., two spheres joined by a constricted neck), and cylindrical (Heiser, 1979a, p. 74–76). The exocarp is typically ~4 mm thick (Heiser, 1979a, p. 82), although the unusual *ipu nui* gourds from Hawai‘i have an average exocarp thickness of 22 mm (Eames & St. John, 1943).

3.2.2 Human Uses of the Bottle Gourd, Especially in Polynesia

The importance of the bottle gourd in human society has gradually diminished as modern, alternative materials have emerged (e.g., plastic and glass) (Dodge, 1943, p. 86; Eames & S$^{t.}$ John, 1943; Prendergast & Decker-Walters, 2000). This section of the thesis is primarily concerned with prehistoric uses, with a focus on Polynesia. The

innumerable uses of the bottle gourd fruit fit into six broad categories: the mature shells (exocarps) are used for containers, apparel, floats and musical instruments, and immature fruits are used for food and medicine.

The predominant and most important use of the dried exocarps of the bottle gourd was hollowed-out as containers. It has been noted that "most probably anything that would fit inside has been kept therein at one time or another" (Maingay, 1985, p. 38). However, the bottle gourd was especially important as a vessel for liquids, being used pantropically as water containers but also as vessels for oil, milk and wine (Maingay, 1985, p. 36). In New Zealand, the smallest bottle gourds were used as vessels for perfumed oils, while the medium-sized ones were used as dishes and water containers (*tahā wai*). The largest gourds (up to 40–50 cm in diameter) were used to preserve birds and rats (*tahā huahua*) (Best, 1976 [1925], p. 251; Colenso, 2001 [1880], pp. 14–15). The large vessels were highly valued, and were sometimes individually named and passed from one generation to the next (Colenso, 2001 [1880], pp. 14–15). The now-extinct *ipu nui* (literally, large gourd) bottle gourds of Hawai'i were thick-shelled and large (up to 48 cm in diameter), and were used as containers for food, water and articles such as clothes (Eames & St. John, 1943).

In New Guinea, bottle gourds are employed as penis sheaths (or phallocrypts) by Highland men. Penis gourds are also used in northern South America and West Africa but such use is thought to be due to independent invention rather than cultural diffusion (Heiser, 1979a, pp. 159–160). Bottle gourds were also used as floats for fishnets, a use documented in south coastal Peru (Whitaker & Bird, 1949) and New Zealand (Best, 1976 [1925], p. 245). Because bottle gourd floats are buoyant and intact (i.e., contain seeds), this use is germane to the discussion of the species' dispersal (see below). Variously shaped bottle gourds have also been used as musical instruments (percussion, string and wind) (Maingay, 1985, p. 42–44). In Hawai'i, *ipu hula* drums were constructed from two huge bottle gourd shells (Dodge, 1943, pp. 37–38), and *uliuli* rattles from a gourd shell with a handle attached (Heiser, 1979a, pp. 183–184). In both Hawai'i (Heiser, 1979a, p. 192) and New Zealand, small gourds were manufactured into nose flutes (Dodge, 1943, pp. 40–44).

The use of the bottle gourd as a food was apparently restricted to China, Japan, Southeast Asia, New Guinea and Eastern Polynesia, where the immature fruits were either boiled and eaten fresh, pickled or dried (see Maingay, 1985, pp. 29–34; as referenced in Walters, 1989). In pre-European New Zealand, the fruit were also eaten but only when young, and were always cooked. In summer, "prodigious numbers" were consumed (Colenso, 2001 [1880], p. 14).

The main medicinal use of the bottle gourd is as a laxative, with particularly bitter cultivars having the most cathartic effect (see Maingay, 1985, pp. 34–35). In China the fruit pulp is used as a diuretic, the cortex of the vine and the flowers are used as poison antidotes, and the seeds are used to treat toothache (as referenced in Walters, 1989).

3.2.3 The Bottle Gourd in Africa — Evolution and Domestication

3.2.3.1 An African Origin for the Bottle Gourd?

The argument for an African origin for the bottle gourd employs two widely-cited criteria of Vavilov's seven-part "differential phytogeographic method" (1951, p. 18) for determining the centre of origin of a cultivated species. These two criteria state that the centre of origin should:

1. Possess more species (both wild and cultivated) closely related to the cultivated plant (i.e., in the same genus) than other locations.
2. Exhibit the widest range of variation of a plant character for the cultivated species.

For contemporary studies the second criterion is extended from morphological and physiological variation to include genetic variation.

Africa meets both of these criteria, and is considered the geographical origin of the bottle gourd by both 20th century bottle gourd authorities — Heiser (1969; 1979a, p. 85) and Whitaker (1971) — and most subsequent authors. Specifically, the five other (wild) species in genus *Lagenaria* (listed above) are all native to Africa (Jeffrey, 1967,

p. 52). The only properly described wild specimen of *L. siceraria* is also from Africa (Decker-Walters *et al.*, 2004b; see below). The bottle gourd is also most variable in Africa, even though each landrace itself is highly inbred (Heiser, 1973b; 1979a, p. 81; Morimoto *et al.*, 2005; Morimoto *et al.*, 2006). Although genetic analysis of the bottle gourd *within* Africa is limited, Decker-Walters *et al.* (2001) show, based on PCA of RAPD fingerprinting, that landraces from southern Africa (Namibia, South Africa, Zimbabwe and Zambia) are derived and divergent, and rule out this region as the origin of *L. siceraria*. Southern Africa does, however, appear to be the source of many modern cultivars (Decker-Walters *et al.*, 2001).

The main evidence *against* an African origin is that the bottle gourd does not appear in the archaeological record of Africa until 6,000 years *after* it appears in the New World. This paradox will be discussed further below, but for now it appears that the Vavilovian evidence for an African origin is strong enough to support this area as the origin of the species.

3.2.3.2 Discovery of the Wild African Bottle Gourd and its Pollinator?

Whilst there have been several discoveries of allegedly wild bottle gourd in Africa, Asia and the New World (see references in Heiser (1969) and Decker-Walters *et al.* (2004b)), none of these can be verified; it is difficult to distinguish plants that are truly wild from those that are escapes, or from those that are remains of earlier gardening. This problem is exacerbated in tropical regions where shifting cultivation is common practice and cultivated plants may remain at an abandoned horticultural site (Maingay, 1985, p. 50).

The first apparently wild bottle gourd subject to detailed morphological and genetic analysis is an accession from southeast Zimbabwe (Wilkins-Ellert, 2003; Decker-Walters *et al.*, 2004b). Although the seed and fruits of the Zimbabwean gourd shared characteristics of both *L. sphaerica* and *L. siceraria*, the foliage morphology, night-blooming flowers and monoecism are more like *L. siceraria* (*L. sphaerica* is a dioecious herbaceous perennial) (Wilkins-Ellert, 2003; Decker-Walters *et al.*, 2004b). Analyses of both RAPD fingerprint and cpDNA sequence data also suggest the Zimbabwean gourd is closer to *L. siceraria* than *L. sphaerica*. However, the presence of

unique bands (36%) in the RAPD profiles and a unique cpDNA SNP[6] in the Zimbabwean accession suggests it could represent a lineage distinct from *L. siceraria* (Decker-Walters *et al.*, 2004b). Although both the morphological and genetic data are consistent with the Zimbabwe accession being a wild bottle gourd, Decker-Walters *et al.* (2004b) suggest the presence of unique alleles indicates it is not part of the wild population from which the bottle gourd was domesticated. The Zimbabwe gourd is an exciting discovery, but further inferences on the origin and domestication of the bottle gourd can only be made with broader sampling of wild bottle gourds from Zimbabwe and other parts of Africa (if such populations still exist).

Because the pollen of the bottle gourd is not windborne (Morimoto *et al.*, 2004), pollinators are required to move pollen from the anther of a male flower to the stigma of a female flower. Determining the origin of the bottle gourd could be aided by locating a wild animal species capable of pollinating this species. To determine the pollinator of the domesticated bottle gourd in Kenya, Morimoto *et al.* (2004) recorded insect visitors to the flowers over a 10 month period. Twenty-two species of insects (from 10 families in 4 orders) visited the flowers. Of these, two species of hawkmoths (*Agrius convolvuli* and *Hippotion celerio*) were suspected to be the major pollinators of bottle gourd, visiting the flowers in the late evening and early morning respectively (Morimoto *et al.*, 2004). The availability of pollinators augments the case for an East African origin for the bottle gourd[7].

3.2.3.3 Domestication of the Bottle Gourd and the Archaeological Record

An African origin for the bottle gourd — buttressed by the recent discovery of an apparently wild bottle gourd in Africa — may suggest domestication occurred here also. However, when domestication occurred, how many times, and whether some domestication events occurred outside of Africa (see Burkill, 1935, pp. 1296–1298;

[6] Any paternal contribution of *L. sphaerica* to the genome of the Zimbabwean accession will not be evident in analysis of the maternally inherited chloroplast DNA. Such a hybridisation event could be resolved by the use of nuclear DNA markers.

[7] Shrivastava (1991) noted that bottle gourd flowers emit an odour similar to that emitted by the now-cosmopolitan tomato mirid (*Nesidiocoris tenuis*, Miridae), which also pollinates the flowers (see Free, 1993, pp. 207–208).

Pickersgill & Bunting, 1969; Harlan, 1976) all remain open questions. These questions will be further addressed in the discussion, but for now at least one domestication event in Africa is assumed.

The oldest bottle gourd remains in Africa come from a 12th Dynasty (2400–2200 BC) Egyptian tomb at Thebes (Luxor) (Schweinfurth, 1884b; 1884a; translation from German kindly provided by Klaus Schliep)[8]. This exocarp tissue is almost certainly from a domesticated bottle gourd. Nearly contemporaneous remains have been recovered from the Gwisho Hot Springs, central Zambia at 2000 BC (Richardson, 1972), and at burial sites at Shongweni Cave, KwaZulu–Natal, South Africa at 2100 BC and 1900 BC (Davies, 1975). Remains also appear in burials excavated by the Leakeys in the Njoro River Cave, Kenya at 1000 BC (see Richardson, 1972). Despite extensive archaeological research in Egypt and east Africa, these are all fairly late appearances for the bottle gourd compared to the age of remains recovered in the New World (see below).

3.2.4 GLOBAL DISPERSAL OF THE BOTTLE GOURD

3.2.4.1 The Bottle Gourd in the New World

Charles Pickering, in the 1840s, was the first European to demonstrate a prehistoric presence of the bottle gourd in the New World, when he recovered exocarp fragments from an Inca cemetery at Pachicamac in Peru (Pickering, 1849, p. 21). Since then, bottle gourd remains, in the form of exocarp, seeds and microfossils, have been found at some of the oldest archaeological sites associated with human settlement in the New World (see Fig. 3.4 for locations).

---

[8] The older, 5th Dynasty date (3500–3300 BC) that is often cited (e.g., Cutler & Whitaker, 1961; Whitaker & Davis, 1962, p. 7; Heiser, 1969; Pickersgill & Bunting, 1969; Pickersgill, 1972; Heiser, 1973b) is incorrect — due to a translation error from German of the Schweinfurth (1884a) article.

Particularly informative in determining the temporal and spatial distribution of the bottle gourd in the New World have been a number of caves in Mexico. Bottle gourd appears in these caves, which were first excavated in the 1950s and 1960s, in association with some of the oldest domesticated plants in the New World: maize (*Zea mays*), the common bean (*Phaseolus vulgaris*) and squash (*Cucurbita pepo*) (Pickersgill & Heiser, 1977, pp. 815–816; Smith, 1997). At the Guilá Naquitz Cave (Oaxaca), bottle gourd remains have been directly dated to 9,920 yr BP with accelerator mass spectrometry (AMS) radiocarbon dating (Smith, 2005; see also Whitaker & Cutler, 1971). About 150 km north of Guilá Naquitz, at the Coxcatlan Cave in the Tehuacán Valley (Puebla), there are new AMS dates for the bottle gourd at 7,200 yr BP, (Smith, 2005; see also Cutler & Whitaker, 1967). Another 500 km farther north, at the Ocampo caves (Tamaulipas), traditional associated radiocarbon dates place bottle gourd exocarp fragments at 9,000 to 7,500 yr BP and the first seeds at 3,400 to 2,400 yr BP (Whitaker *et al.*, 1957; Whitaker, 1971). The earliest AMS date for Ocampo is 6500 yr BP (Smith, 2005), which is compatible with Smith's (2001) hypothesis that the bottle gourd, with the pepo squash, reached Tamaulipas via the Tehuacán Valley from southern Mexico.

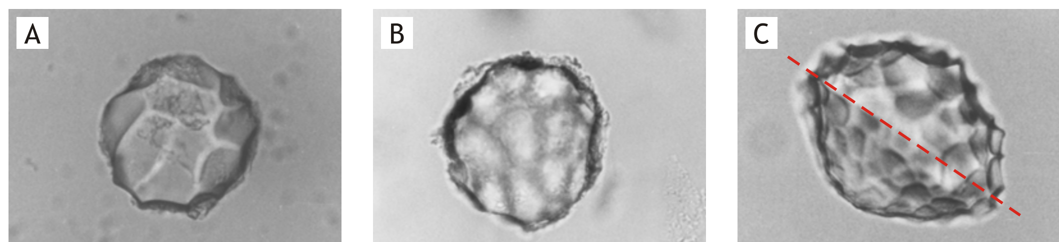
The earliest evidence of the bottle gourd in North America is from the waterlogged Windover site near Titusville in east-central Florida, where a large piece of exocarp dated to 7,290 ^{14}C yr BP was recovered from a subadult human burial (Doran *et al.*, 1990; this date is corroborated by many other dates from the same site). However, this date in Florida is much earlier than the rest of North America, with bottle gourd not appearing in New Mexico until 300 BC, farther north in New Mexico, Colorado, Utah, South Dakota, Illinois and Kentucky by AD 1000 and as far east as Pennsylvania by the mid-16th century (Whitaker, 1948; Cutler & Whitaker, 1961; Richardson, 1972). By the time of European contact, the bottle gourd had reached as far north as Lake Ontario on the present Canadian–American border (Carrier, 1923, p. 218).

Since the 1990s, plant microfossils (primarily pollen, starch granules and phytoliths) recovered from archaeological sites have become increasingly important indicators of prehistoric cultivation and domestication in the New World and elsewhere (e.g., Piperno & Holst, 1998; Piperno *et al.*, 2000b; Piperno *et al.*, 2002; Piperno & Stothert, 2003; Piperno *et al.*, 2004). Unlike the macro evidence of plant remains, which

is rarely preserved — especially in wet, tropical soils — microfossils can persist in the soil thousands of years after plant death (Bryant, 2003; Fig. 3.2).

The Vegas early Holocene coastal site on the Santa Elena Peninsula[9], southwest Ecuador has yielded large, diagnostic bottle gourd phytoliths that date to 9,060 yr BP (Piperno *et al.*, 2000a). Piperno *et al.* (2000a) suggest that these bottle gourds were domesticated, since the phytoliths are large, and this is generally an indicator of domestication (Piperno *et al.*, 2002). Bottle gourd phytoliths at 9,300 to 8,000 yr BP have also been recovered from the Peña Roja site in the Colombian Amazon (Piperno & Pearsall, 1998, pp. 203–204), and at 8,000 to 7,000 yr BP from the Aguadulce rock shelter on the Pacific coastal plain of Panama (Piperno *et al.*, 2000a).

Fig. 3.2
Phytoliths from Bottle Gourd and *Cucurbita*



Phytoliths are microscopic, siliceous (silica-based) bodies deposited during growth, and are thought to be deposited in the soils immediately beneath the plant after death (Horrocks et al.*, 2000a). In species of Cucurbitaceae they are located, if present, at the interface between the epidermis and parenchyma of the fruit rind/exocarp (Piperno* et al.*, 2000a). Phytoliths have a distinctive size and morphology and, depending on the taxonomic group, can be diagnostic to species level. Phytoliths from* Lagenaria siceraria *are diagnostic to at least genus level, and are large (mean length of 91 µm), hemispherical, and have elongated scallops (Piperno* et al.*, 2000a). Bottle gourd, however, produces relatively few phytoliths, leading to a limited distribution in archaeological soils (Piperno & Pearsall, 1998).*

A. *A hemispherical phytolith from bottle gourd, with the scalloped, rounded side face-up.*
B. *A hemispherical phytolith from bottle gourd, with the flat, undecorated flat side face-up.*
C. *A typical* Cucurbita *(squash) phytolith, which is spheroid with deeply scalloped surfaces. The dashed line indicates the axis between the two hemispheres.*

Photographs reproduced, with permission of Elsevier, from Piperno et al. *(2000a, Figs 2, 9, 10)*
Copyright © 2000 by Elsevier

[9] The Santa Elena Peninsula adjoins the territory of the Cañari speakers, where the Polynesian word *kumara* (for sweet potato) may originate (Scaglion, 2005; see Section 4.2.3.1).

In Peru, bottle gourds appear in the highlands at Ayacucho by 8,000–6,000 yr BP[10] and on the coast in Ancón–Chillón region (near Lima) by 8,000–7,000 yr BP (Richardson, 1972; Cohen, 1977, pp. 157, 173). Bottle gourds have also been recovered from the midden of Huaca Prieta in northern coastal Peru. Remains appear here from 5,000 yr BP, and include seeds and immaculate, intact bottle gourd floats attached to a fishnet (Whitaker & Bird, 1949; Cutler & Whitaker, 1961; Towle, 1961, pp. 92–95; Whitaker, 1983; Hudson, 2004; Fig. 3.3). By AD 500–1100 the bottle gourd had reached as far south as Pampa Grande, northern Argentina (Whitaker, 1983).

Fig. 3.3
Prehistoric Bottle Gourd Floats, Huaca Prieta, Peru

Dated to 5,000–4,000 yr BP, eight intact bottle gourd floats attached to a large fishnet have been recovered from Huaca Prieta, coastal Peru. See jar at left for scale.

Reproduced, without objection from HarperCollins, UK, from Heyerdahl (1952, Pl. LXIV)
Copyright © 1952 by Allen & Unwin

[10] An earlier (associated) radiocarbon date of 13,000 yr BP has been assigned to macrobotanical remains from Pikimachay Cave in Ayacucho, Peru, but this has generally been disregarded as the remains may be intrusive from a younger layer (see Flannery, 1973). However, direct AMS dating of the Pikimachay material should now be possible.

It is now clear that the oldest known archaeological remains in the New World predate those in Africa by about 6,000 years (10,000 yr BP in Mexico versus 4,000 yr BP in Egypt). This problem of chronology is yet to be resolved (although it has been noted). Despite the chronology, the New World bottle gourd is generally considered to be of African, rather than Asian, origin. For the purposes of this project an African origin for the New World bottle gourd was assumed because this has been the consensus opinion (see Heiser, 1990), and is based on the studies which show that the modern gourds of the New World belong to ssp. *siceraria* (the African subspecies) rather than ssp. *asiatica* (the Asian subspecies) (Heiser, 1973b; 1979a, p. 97). PCA of RAPD fingerprinting data for seven African, nine New World and 15 Asian landraces confirms that New World germplasm is primarily — but not solely — of African origin (Decker-Walters *et al.*, 2001).

It is worth discussing briefly the modes by which the bottle gourd could have dispersed to the New World because realistic scenarios may also explain how the bottle gourd dispersed to Polynesia, a fact perhaps too lightly addressed by some recent authors (e.g., Green, 2000b, 2005). Several theories have been proposed to explain dispersal of the bottle gourd from Africa to the New World, based both on natural dispersal (fruits floating from Africa to the New World) or human-mediated dispersal. Natural dispersal is biologically possible; Whitaker & Carter's (1961) elegant experiment found that gourds remain afloat in seawater for up to 347 days (even when host to barnacles and mussels) and that there is no significant loss in viability of the seeds due to this treatment. Whitaker & Carter (1954) calculated a bottle gourd travelling 6500 km from the Gulf of Guinea to Brazil with the South Equatorial Current (SEC) at a velocity of 67.5 cm s^{-1} would have to remain afloat for at least 145 days. Recent estimates suggest the velocity of the SEC may only be 11.3–30 cm s^{-1} (Bonhoure *et al.*, 2004). This increases the time required for a bottle gourd to cross the Atlantic to 248–659 days (8 to 22 months) — still within the timeframe for it to remain afloat and retain viable seed

Because the bottle gourd is not a littoral plant (Camp, 1954; Whitaker, 1971), a fruit washed up on a South American beach would have to be transferred inland to a site suitable for growth. This could have been effected naturally (e.g., by a hurricane (Heiser, 1979a, p. 115) or tsunami) or by a human who found the bottle gourd and took

it back to his or her village. Whitaker & Carter (1961) found that the seeds remain viable for a least six years after the fruit has been floated in seawater for a year, providing ample time for a gourd to reach a suitable site.

Although dispersal from Africa to South America by floatation is possible, this event most likely involved a domesticated bottle gourd, and not a wild bottle gourd as suggested by some authors (e.g., Decker-Walters *et al.*, 2004b). In the New World, the adoption and cultivation of the bottle gourd has often been linked to its domestication, i.e., because the bottle gourd was cultivated in the New World, it either floated there as a wild plant and was domesticated and then cultivated (Decker-Walters *et al.*, 2004b), *or* was introduced by humans in a domesticated state and then cultivated (Whitaker & Carter, 1961). A third possibility, that comes from dissociating cultivation and domestication, is that the bottle gourd was domesticated in Africa, dispersed naturally (i.e., floated) to the New World, and was independently adopted as a crop plant by indigenous Americans (a combination of diffusion and independent invention models). Intact fruits (i.e., containing seeds) of domesticated bottle gourds grown by coastal communities in western Africa could easily have found their way out to sea, especially those used as floats for fish nets. Furthermore, the exocarp of the *wild* bottle gourd discovered in Zimbabwe probably has limited ability to float for long periods; in plants grown, the exocarp was "not durable as is typical of *L. siceraria*… [and] became very thin, was easily cracked, and ultimately disintegrated after several years" (Decker-Walters *et al.*, 2004b). On the surface, this would favour a domesticated plant floating to the New World.

Although several authors (e.g., Heiser, 1979a, p. 99–117; Maingay, 1985, p. 63) have speculated on human-mediated transfer of the bottle gourd from Africa to South America, only one (Lathrap, 1977, p. 727) favours this hypothesis over natural dispersal. Lathrap (1977, p. 727) proposes that African fishermen crossed the Atlantic to South America, bringing with them the bottle gourd, cotton (*Gossypium* sp.) and leguminous plants used as fish poisons[11]. However, the New World cottons are a

[11] It is paradoxical to have African fisherman crossing the 2,500 km-wide Atlantic Ocean to South America 10,000 yr BP, but not also crossing the 400 km-wide Mozambique Channel to Madagascar (assuming the shortest, direct routes in both cases). African settlers only began arriving in Madagascar after the initial settlement by Austronesian voyagers between AD 300–800 (see Section 1.2.2.2).

different species to those used in Africa and the taxonomic status of the leguminous fish poisons makes African–New World comparisons difficult (Heiser, 1979b). Also, there is no direct evidence (e.g., genetic) of any contact between Africa and South America before 10,000 yr BP. Human-mediated dispersal across the Atlantic must be considered very unlikely.

Although dispersal *within* the New World would have been largely human-mediated, the ability of the fruits to survive sea water allows for a natural dispersal component. This could explain the presence of the bottle gourd very early at Windover in Florida when it does not appear at other North American locations until much later — it could have floated directly across the Caribbean from South America (Heiser, 1985, p. 20; 1990). A separate introduction from Africa to North America may also have occurred; there is some morphological evidence to support repeated introductions from Africa to the New World (Heiser, 1973b). Heiser (1990) also suggests overland dispersal of wild bottle gourd to account for the distribution in the New World.

3.2.4.2 The Bottle Gourd in Asia

What was thought to be the earliest dates for the bottle gourd in Asia were reported from northwest Thailand, where fragments of shell (exocarp) were recovered from Spirit Cave and Banyan Valley Cave and dated to 11,000–9,000 yr BP and 9,000–8,000 yr BP respectively (Gorman, 1969; Yen, 1977, Table 1; Fig. 3.4). Yen (1977, Table 1, p. 575) was cautious in his identification of these remains as *L. siceraria*, (as acknowledged by both Green (2000b) and Golson (2002)), and Heiser (1979a, pp. 82–83) subsequently concluded, based on examination of shell thickness and cell morphology, that none of these remains were of bottle gourd (although he could not positively identify them).

Later remains (seeds and gourd-shaped pots) have been recovered from Hemudu (syn. Ho-mu-tu), Zhejiang Province, China (130 km south of Shanghai) at 7,200 and 6,900 yr BP (Chang, 1986, pp. 208–210; Bellwood, 1997, p. 208; Fig. 3.4). Although both Green (2000b) and Golson (2002) appear hesitant in accepting the identification of these remains as *L. siceraria*, the seeds should allow accurate identification. The bottle

gourd was certainly in China by 1000 to 500 BC, when it appears in Chinese literature (Walters, 1989). Gourd remains have also been recovered from Early to Middle Jomon sites in Japan (Crawford, 1992, p. 18), including the coastal Torihama shell midden (5,500 yr BP) at Mikata, Fukui Prefecture in western Honshu (Imamura, 1996, p.108), and the Sannai Maruyama site (5,900 to 4,300 yr BP) at Aomori, Aomori Prefecture in northern Honshu (Habu *et al.*, 2001, pp. 9, 13; Fig. 3.4).

It is also likely that bottle gourd was grown by the Proto-Austronesian Ta-p‘en-k‘eng culture in western Taiwan from 6,300 to 4,500 yr BP (Bellwood, 1997, pp. 212, 217). This allows for the bottle gourd to have been taken out of Taiwan with colonising humans during the Austronesian expansion, which began about 5,000 yr BP with the settlement of the Philippine island of Luzon (Bellwood, 1997, p. 241).

The bottle gourd could easily have been dispersed by people to Asia from Africa via the Middle East, or alternatively across the Indian Ocean by either natural or human-mediated means. Interestingly, accessions from Ethiopia and Syria were both, based on PCA of RAPD fingerprinting data, intermediate between African and Asian landraces (Decker-Walters *et al.*, 2001), consistent with the bottle gourd dispersing from the Horn of Africa, through the Middle East and into Asia.

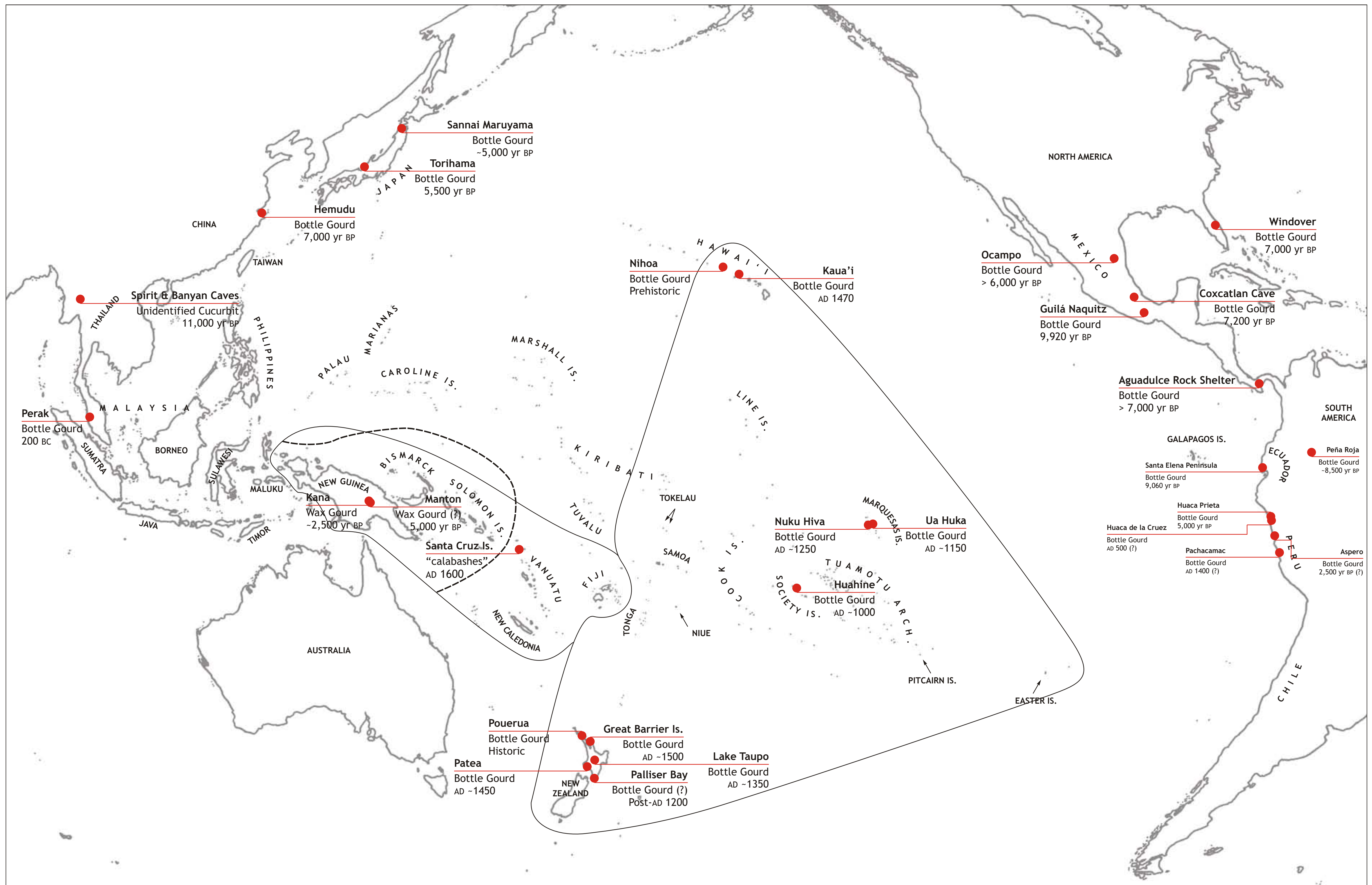
3.2.4.3 A Pre-Austronesian Expansion Pacific Rim Distribution

A range of evidence is now providing a tentative reconstruction of the domestication and dispersal of the bottle gourd prior to the entrance of this species into Oceania. The bottle gourd is almost certainly native to Africa (possibly to east Africa) where it may have been domesticated 10,000 yr BP or earlier. Bottle gourd, probably in a domesticated state, dispersed from Africa to South America by at least 10,000 yr BP — it appears in archaeological deposits in Mexico about this time. Dispersal from Africa was most likely effected by intact gourds (such as seed-containing fishnet floats) floating across the Atlantic Ocean — with current patterns suggesting departure from West Africa and landfall in Brazil. Seeds from such transatlantic bottle gourds could have germinated if they were carried far enough inland (by high tides or tsunami) or, perhaps more likely, could have been collected from the beach by South Americans and

taken inland. Under either scenario, the gourd was quickly adopted as a crop plant by South Americans, and was rapidly dispersed throughout the New World, including west to Peru, south to Argentina and north as far as the Great Lakes. In a separate dispersal event, the bottle gourd spread from Africa eastwards into Asia (where it was or became ssp. *asiatica*). This could have been effected either by humans moving into Arabia or by floatation across the Indian Ocean to India. The Asian gourd spread east, reaching Japan and China by at least 7,200 yr BP. Thus, the bottle gourd was poised for entry into the Pacific from either direction — from Asia in the west, or from the New World in the east.

Fig. 3.4
Prehistoric Bottle Gourd in Asia, the New World and Oceania

Bottle gourd remains, usually exocarp fragments but sometimes seeds, have been recovered from a number of archaeological sites in the New World, Asia and Oceania. The bottle gourd first appears in the New World at Guilá Naquitz in Mexico at 9,920 yr BP, and by 5,000 yr BP at a number of sites from Peru to Florida. In Asia, the bottle gourd first appears in China and Japan by 7,200 and 5,500 yr BP respectively. Gourd remains from New Guinea dated to 5,000 yr BP were thought to be bottle gourd, but are probably from a different species — wax gourd. The bottle gourd first appears in Southeast Asia at 200 BC, and may have reached the Santa Cruz Islands by late in prehistory. In Polynesia, the bottle gourd first appears at AD 1050 in Huahine, Society Islands, and within 300–400 years in New Zealand, Hawai'i and the Marquesas Islands.



Sannai Maruyama
Bottle Gourd
~5,000 yr BP
Torihama
Bottle Gourd
5,500 yr BP
JAPAN
Hemudu
Bottle Gourd
7,000 yr BP
CHINA
TAIWAN
Spirit & Banyan Caves
Unidentified Cucurbit
11,000 yr BP
THAILAND
PHILIPPINES
MARIANAS
PALAU
CAROLINE IS.
MARSHALL IS.
MALAYSIA
Perak
Bottle Gourd
200 BC
SUMATRA
BORNEO
SULAWESI
MALUKU
JAVA
TIMOR
NEW GUINEA
BISMARCK
SOLOMON IS.
Kana
Wax Gourd
~2,500 yr BP
Manton
Wax Gourd (?)
5,000 yr BP
Santa Cruz Is.
"calabashes"
AD 1600
VANUATU
NEW CALEDONIA
FIJI
TUVALU
KIRIBATI
AUSTRALIA
TONGA
SAMOA
TOKELAU
NIUE
COOK IS.
LINE IS.
HAWAI'I
Nihoa
Bottle Gourd
Prehistoric
Kaua'i
Bottle Gourd
AD 1470
MARQUESAS IS.
Nuku Hiva
Bottle Gourd
AD ~1250
Ua Huka
Bottle Gourd
AD ~1150
TUAMOTU ARCH.
SOCIETY IS.
Huahine
Bottle Gourd
AD ~1000
PITCAIRN IS.
EASTER IS.
Pouerua
Bottle Gourd
Historic
Great Barrier Is.
Bottle Gourd
AD ~1500
Lake Taupo
Bottle Gourd
AD ~1350
Patea
Bottle Gourd
AD ~1450
Palliser Bay
Bottle Gourd (?)
Post-AD 1200
NEW ZEALAND
NORTH AMERICA
MEXICO
Windover
Bottle Gourd
7,000 yr BP
Ocampo
Bottle Gourd
> 6,000 yr BP
Coxcatlan Cave
Bottle Gourd
7,200 yr BP
Guilá Naquitz
Bottle Gourd
9,920 yr BP
Aguadulce Rock Shelter
Bottle Gourd
> 7,000 yr BP
SOUTH AMERICA
GALAPAGOS IS.
ECUADOR
Peña Roja
Bottle Gourd
~8,500 yr BP
Santa Elena Peninsula
Bottle Gourd
9,060 yr BP
Huaca Prieta
Bottle Gourd
5,000 yr BP
Huaca de la Cruez
Bottle Gourd
AD 500 (?)
Pachacamac
Bottle Gourd
AD 1400 (?)
PERU
Aspero
Bottle Gourd
2,500 yr BP (?)
CHILE

3.2.5 THE BOTTLE GOURD IN ISLAND SOUTHEAST ASIA AND OCEANIA

3.2.5.1 The Bottle Gourd in the Western Pacific

From linguistic studies, Green (2000b) concludes that, despite the very early presence of the bottle gourd in China and Japan, the bottle gourd may not have appeared in Southern Asia and Island Southeast Asia until much later — after 2000 BC and maybe as late as 200 BC. In the Sanskrit of India spoken from 1500 to 800 BC the word for bottle gourd is *alābū* (and cognates) (see Green, 2000b). The Sanskrit *alābū*, is probably the origin of **labu*[12], the word used for the bottle gourd in the Austronesian languages of Sumatra, Java, Sulawesi, central Maluku (the islands between Sulawesi and Western New Guinea) and the South Philippines (original sources referenced in Green, 2000b). The direction of this linguistic borrowing, combined with archaeological evidence for trade between India and western Indonesia about 200 BC, persuades Green (2000b) of the spread of the bottle gourd from India (but also possibly Taiwan) farther east into Island Southeast Asia and ultimately into the New Guinea Highlands. Green's hypothesis is apparently supported by archaeological evidence of bottle gourd found in mounds in the state of Perak in Peninsular Malaysia and dated to approximately 200 BC (Nik Hassan Shuhaimi, 1991, p. 150; Golson, 2002). An earlier introduction of the bottle gourd from India into Taiwan–Indonesia (along with the term **tabu* which appears in Austronesian languages from Maluku to Sumatra and Madagascar) had been supported by remains of bottle gourd in East Timor at 2000 BC (as reported in Glover, 1977; Bellwood, 1997, p. 231; Green, 2000b). However, this early Timor date was in error and has now been revised to no older than AD 1000 (Golson, 2002)), bringing the oldest well-supported age of the bottle gourd in Southeast Asia forward to the 200 BC **labu* introduction.

In 1966, crushed gourd was recovered from the Manton site[13], Waghi Valley, near Mt Hagen in the Western Highlands Province of Papua New Guinea (1,600 m a.s.l.). The remains were identified as *Lagenaria siceraria* and dated to

---

[12] In linguistics, an asterisk immediately preceding a word indicates a reconstructed form.

[13] The Manton Site is approximately 6 km south of the famous Kuk Swamp site, which has been very important in reconstructing the origins of New Guinea agriculture (e.g., Denham *et al.*, 2003).

4,600 ± 140 yr BP (Golson *et al.*, 1967) (although the lack of detail in the paper meant that the date was subsequently cited as only pre-2,300 ± 120 yr BP (Golson, 2002)). The identification as *L. siceraria* was based on information from workmen at the site that the gourd was like one grown in the villages, was widespread in the rest of New Guinea, and was used to carry water (Golson, 2002).

In 1994, gourd remains were recovered from Kana, another agricultural site in the Waghi Valley of the Papua New Guinea central highlands (1,500 m a.s.l.) (Muke & Mandui, 2003). The remains, which consisted of exocarp and seeds, were dated to 2,950–2,000 yr BP. However, the presence of seeds allowed the identification of the material not as bottle gourd but as wax gourd (*Benincasa hispida*) (Matthews, 2003); based on seed and fruit morphology, *L. siceraria* and other species of Cucurbitaceae recorded from the Western Highlands could be excluded.

In light of the identification of the gourd remains at Kana as wax gourd, Golson (2002) suggests the Manton gourd is also wax gourd, and not bottle gourd as first reported. (Although the identification is questioned, the date is confirmed, with a second date from the same layer at 4,880 ± 90 ^{14}C yr BP (Golson, 2002)). There are extensive records of wax gourd being cultivated in New Guinea in the late 19th century (see references in Golson, 2002). Although the wax gourd is most often used as a food plant (Heiser, 1979a, p. 60–63; Walters & Decker-Walters, 1989), there are records (see Golson, 2002) of the orange-sized, hard exocarp variety found in Papua New Guinea being used as a container to hold lime for betel nut (*Areca catechu*) chewing. With the identification if the New Guinea Highland material as wax gourd, there is now no evidence for the bottle gourd in Island Southeast Asia until the 200 BC date for Peninsular Malaysia.

It is unclear how far east into Island Southeast Asia and Island Melanesia the bottle gourd spread prehistorically although "calabashes" were observed during the Spanish voyages to the eastern Melanesian islands of Santa Cruz, Solomon Islands (1595) and Espiritu Santo, Vanuatu (1606) (Yen, 1973). These were assumed to be true bottle gourds by Yen, but were described by the Spanish as "very small melons" so may actually have been *Benincasa hispida*; Yen suggested bottle gourd because it is "the only indigenous cultivated cucurbit in the region," but with the identification of wax

gourd in New Guinea by at least 2,000 yr BP (Matthews, 2003), it is now clear that at least one other cucurbit was present in Island Melanesia.

Morphological studies of six bottle gourd cultivars from three locations in New Guinea (one in Western New Guinea, two in Papua New Guinea) to determine the origin of New Guinea bottle gourd (see Section 3.2.2) produced conflicting results (Heiser, 1973a; 1979a, pp. 157–160). For five of eight characters, the plants were more like the African–American subspecies (ssp. *siceraria*) than the Asian subspecies (ssp. *asiatica*), for one character the gourds were more like the Asian subspecies, and two characters were intermediate between the two subspecies (see Table 3.1) (Heiser, 1973a; 1979a, pp. 93, 158). Although Heiser conceded his sample was too small to make any firm conclusions (1979a, pp. 158–159), he was able to advance several theories to explain why the New Guinea gourds appear more like the African and New World ssp. *siceraria*, including:

1. Human-mediated or natural dispersal from the New World or Africa to New Guinea followed by hybridisation with Asian cultivars.
2. Introduction of an early form of ssp. *asiatica* possessing African morphology into New Guinea followed by replacement of that early type in mainland Asia by the modern ssp. *asiatica*.
3. Convergent evolution of the New Guinean cultivars so that they now more closely resemble ssp. *siceraria* (1973a).

A connection between New Guinea and American cultivars is not supported by PCA of RAPD fingerprint data for two New Guinean phallocrypt bottle gourds (one from Western New Guinea and the other from Papua New Guinea); both cultivars group with other cultivars from Asia (Decker-Walters *et al.*, 2001), consistent with only the last two of Heiser's hypotheses. DNA analysis of the six Heiser accessions should be performed to confirm the result of Decker-Walters *et al.* (2001).

In summary, the remains from China and Japan confirm that the bottle gourd was cultivated in the Far East by at least 7,000 yr BP but there is no evidence for it being present in Southeast Asia and Island Melanesia until much later. The remains from Thailand at 9,000 to 11,000 yr BP are apparently not from domesticated bottle gourd (as for New Guinea they may prove to be wax gourd), and the remains from Kana, and probably Manton, are from the wax gourd. The linguistic evidence suggests the bottle

gourd reached Southeast Asia, specifically Peninsular Malaysia, *ca* 200 BC from India, via the extensive trading networks operating in the Bay of Bengal during that time (Green, 2000b). From Southeast Asia the bottle gourd may have spread as far as the Solomon Islands and Vanuatu in prehistory (Yen, 1973) but this remains unsubstantiated.

3.2.5.2 The Wax Gourd in Polynesia

As in New Guinea, accurate temporal and spatial distributions for the bottle gourd in Polynesia have been obscured by widespread misidentifications of the wax gourd (*B. hispida*) as bottle gourd (*L. siceraria*). These misidentifications occurred for two reasons:

1. The wax gourd in Melanesia and Polynesia is not typical of the species elsewhere, and instead its superficial morphology is closer to bottle gourd. In Asia, the wax gourd is known only as a food crop (Heiser, 1979a, pp. 60–63; Walters & Decker-Walters, 1989), and is large, ovoid and does not form a hard pericarp. The variety found in Polynesia and eastern Melanesia however, has a utilitarian purpose (like the bottle gourd), and is small (5–12 cm), spheroid and forms a hard pericarp, often leading it to be misidentified as *L. siceraria* (Whistler, 1990). The wax gourds from the Kana (and Manton) site in New Guinea also appear to be of the Polynesian/Island Melanesian type.
2. When anthropologists and naturalists found the wax gourd in Eastern Polynesia (it was known from Asia to the Marquesas Islands (Whistler, 1990)), it was often assumed to be bottle gourd because the latter was so abundant and widespread in this region (and in tropical regions worldwide). Even when the wax gourd was discovered in Fiji and Western Polynesia, where there is no evidence of the bottle gourd ever having a prehistoric presence (Whistler, 1990), it was identified as bottle gourd.

To distinguish the hard pericarp variety of *B. hispida* found in the Pacific from varieties found elsewhere, Whistler (1990) proposes the name *B. hispida* var. *pruriens* (Parkinson) Whistler. In Fiji, Tonga, Samoa, ‘Uvea (Wallis Is.), Futuna and Tahiti, *B.*

hispida var. *pruriens* is used hollowed-out as a container for scented coconut oil (Whistler, 1990).

It is feasible that the wax gourd, in addition to human-mediated transfer, also dispersed naturally in Polynesia by floating in the sea. Henry Guppy (1906, p. 570) observed what were probably wax gourds[14] floating in the open sea and in Fijian estuaries, and also stranded on beaches. Guppy collected one of these gourds from the sea, floated it in sea water for a further two months (after which time it was still buoyant) and planted the seeds, some of which grew. Decker-Walters & Walters (2000) found a wax gourd contained viable seeds after floatation in sea water for six months.

3.2.5.3 Distribution of the Polynesian Bottle Gourd

Gourd remains first appear in the archaeological record of Eastern Polynesia between AD 850 and 1200 in Huahine, Society Islands (Emory, 1979, pp. 202–204; Leach, 1984, p. 23), and are assumed to represent bottle gourd, although the possibility that they may be wax gourd has not be examined.

In the Marquesas Islands, bottle gourd fragments have been recovered from both Ua Huka and Nuku Hiva. On Ua Huka, five bottle gourd fragments (identified as *L. siceraria* exocarp) from the Vaipikoau cave site have an associated radiocarbon date of *ca* AD 1150 (Kirch, 1973; Leach, 1984, p. 25). On Nuku Hiva, remains have been recovered from the Akipou rock shelter (NBM 1) with (very rough) associated radiocarbon dates of AD 1046–1596 (Suggs, 1961, pp. 20, 22–24, 99). Undated remains have also been recovered from the Nahotoa Cave (NH 4) on Nuku Hiva (Suggs, 1961, pp. 29–30, 99)[15].

---

[14] Guppy (1906, p. 570) said the fruits were “calabashes or bottle-gourds” but that they were “more or less globular, 3 or 4 inches across” and the seeds are “not those figured ...as belonging to Lagenaria vulgaris [s*iceraria*], and more resemble those of Cucurbita.” This description suggests Guppy was examining *B. hispida* and not *L. siceraria* (Matthews, 2003).

[15] Richardson (1972) incorrectly reports the Suggs (1961) date for the Akipou remains as AD 760 ± 150, and incorrectly ascribes a date of AD 1100 ± 1790 to the Nahotoa remains when these are actually undated.

In Hawai‘i, prehistoric bottle gourd remains have been recovered from the Māhā‘ulepū Caves on the island of Kaua‘i. These remains, the identification of which is aided by the presence of many seeds, exocarp fragments and stems, have a direct AMS date of AD 1425–1520 (Burney *et al.*, 2001). Later historic-period remains appear at the same site at AD 1645–1950 (Burney *et al.*, 2001). The Māhā‘ulepū Caves are unusual for the volcanic Hawaiian Islands because they have formed from Pleistocene eolianite (a limestone) (Burney *et al.*, 2001); these alkaline conditions make preservation of DNA in the bottle gourd remains much more likely (*cf.* Bramanti *et al.*, 2003; Zink & Nerlich, 2003).

Fragments of bottle gourd have been recovered from two sites (57 and 58) in a rock shelter on remote Nihoa Island in the Hawaiian Archipelago (Emory, 1971 [1928], pp. 34–36, 47). The fragments were found in association with numerous other archaeological items. Although some of the fragments were from a small gourd, fragments of large gourds were found at both sites, suggesting they belong to *Lagenaria siceraria* and not *Benincasa hispida*. The bottle gourd remains are certainly prehistoric; Nihoa was uninhabited at the time of its European discovery in 1769, and was only known to Hawaiians in their oral histories (Emory, 1971 [1928], pp. 3, 8). Less reliable remains have also been recovered from AD 1300–1700 shelters at Lapakahi on Hawai‘i (Big Island), Hawai‘i, and although were only identified to the level of “cucurbit” by Yen (Griffin *et al.*, 1971), are considered as probably bottle gourd by Leach (1984, p. 31).

For Easter Island (Rapa Nui), I have been unable to find any reports of dated prehistoric bottle gourd remains. This lack of evidence, despite extensive archaeological work on the island, leads some to suggest the bottle gourd was only introduced historically (Roger Green, pers. comm.). However, perhaps it was just rare on Easter Island or there are taphonomic issues — bottle gourd is recorded in the oral history of the island (as *ipu*) as being introduced with the first settlers (along with dogs, chickens, paper mulberry, ti pore, yams, etc.) (Métraux, 1971 [1940], p. 60), it was recorded by both Spanish voyagers and Captain Cook in the early 1770s (Métraux, 1971 [1940], p. 157), and Métraux (1971 [1940], p. 157) noted that they were “abundant” and “grew wild [?] over the island” in the late 19th century. Métraux also found some bottle gourd fragments in Easter Island graves, although there is no date ascribed to these.

3.2.5.4 The Polynesian Bottle Gourd: A New World Origin?

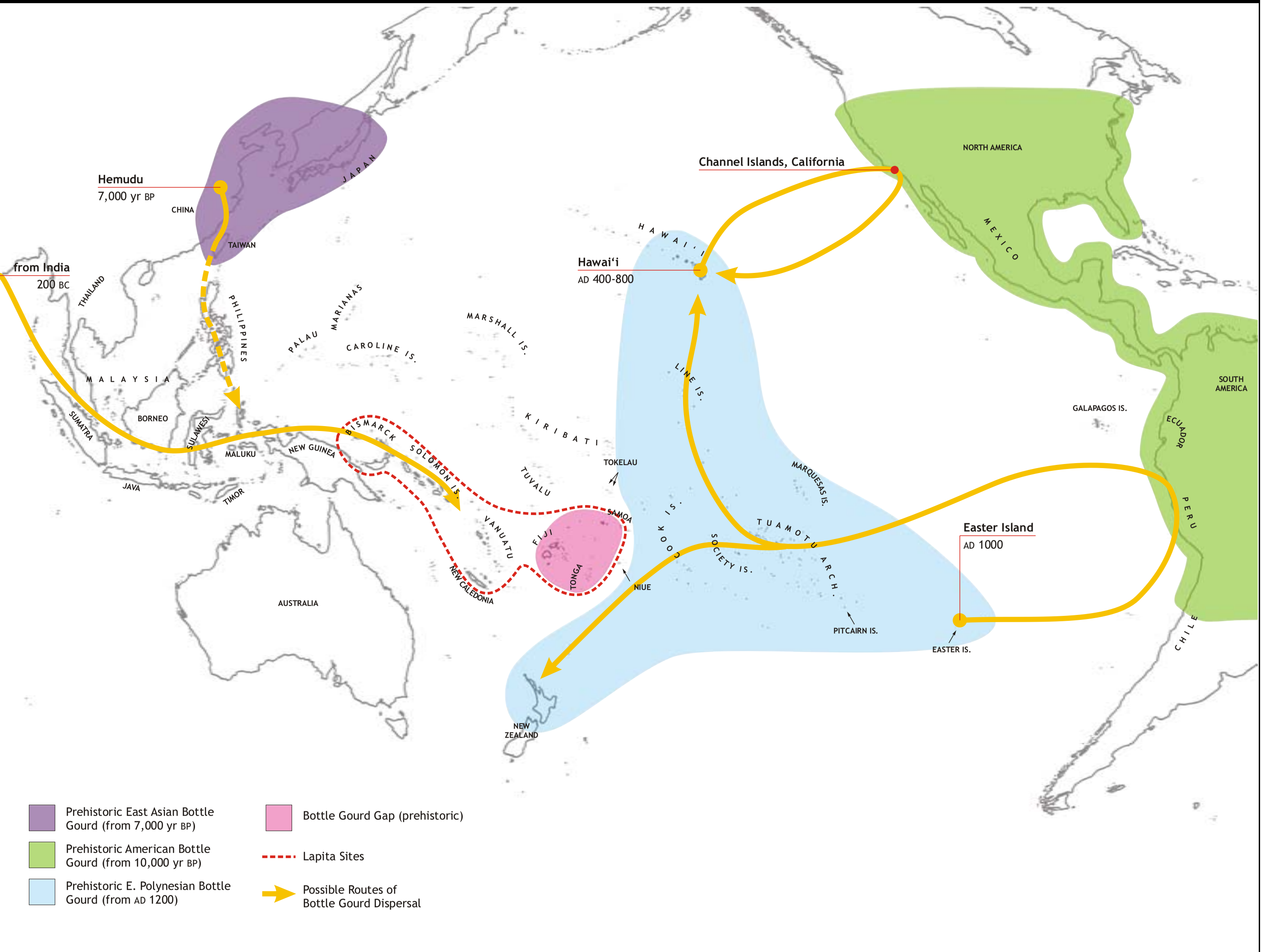
Several lines of evidence have lead recent authors to suggest a New World, specifically South American, origin for the Polynesian bottle gourd (e.g., Whistler, 1990; Burtenshaw, 1999; Green, 2000b, 2005). The evidence, detailed below, includes the unusual distribution of the bottle gourd in Oceania, the lack of archaeological evidence for the bottle gourd in Island Melanesia, the origin of Polynesian vernacular terms used for the bottle gourd (linguistic evidence), and morphological evidence from prehistoric and modern Polynesian bottle gourds. The human-mediated introduction of the sweet potato has also been used as evidence (the validity of this will be discussed in Section 3.5.4).

Teasing apart the distributions of the bottle gourd and the wax gourd reveals an area straddling eastern Melanesia (Fiji) and Western Polynesia (Samoa, Tonga and Niue) where there is no evidence for a prehistoric presence of the bottle gourd[16] (Whistler, 1990). This area is flanked by two regions with a prehistoric presence of the bottle gourd (Fig. 3.5). The absence of the bottle gourd in the "Bottle Gourd Gap" suggests it was not transferred through this area from the west (unless it has subsequently become extinct). It is also unlikely that the bottle gourd came from the northwest, through Micronesia, because as Whistler (1990) points out "there is no evidence of any significant contact between that region and Eastern Polynesia. Also, the bottle gourd is rarely found on atolls, the major type of island in Micronesia." The Bottle Gourd Gap leads Whistler (1990) to favour a South American origin for the Polynesian bottle gourd.

[16] The only records of the bottle gourd in Western Polynesia are from Tonga in 1959 and Samoa in 1972 and these are thought to be recent introductions (see Whistler, 1990).

Fig. 3.5
Distributions and Dispersal Routes of the Oceanic Bottle Gourd

The bottle gourd has been present in the Americas and East Asia since 9,920 and 7,200 yr BP respectively. It is unclear how far south the East Asian bottle gourd spread in prehistory (indicated by dashed line). The Southeast Asian bottle gourd may be a very recent arrival from India 200 BC, spreading only as far as Vanuatu in prehistory. The bottle gourd was apparently not present in Western Polynesia (the Bottle Gourd Gap), suggesting that it was not introduced from Asia into Polynesia via human-mediated dispersal. However, the bottle gourd may not have been required in the Gap region as Lapita pottery was widely available as an alternative for containers (distribution of Lapita sites from Kirch (2000)). The bottle gourd was present in Eastern Polynesia from AD 1200 and may have been introduced from the Americas by human-mediated dispersal. A human-mediated introduction from South America could have been effected by Polynesian voyagers who sailed from Easter Island around AD 1000 to the Peruvian Coast, and returned (also with the sweet potato), or by Polynesian voyagers who sailed from Hawai'i around AD 400–800 to the Californian Channel Islands, and returned.



Linguistic evidence offers no support for an Asian origin for the Polynesian bottle gourd[17]. For many of the commensal Polynesian plants of Asian origin it is relatively easy to reconstruct at least one appropriate proto-form attributable to the proto-Oceanic (POc) language subgroup of *ca* 3,000–3,500 yr BP (Ross, 1996; see Green, 2000b), but this is not the case for the bottle gourd, which according to Ross (1996, p. 166) suggests the bottle gourd was "*not* known to POc speakers…. The gourd may well have reached Oceania from two directions, arriving in Melanesia from the Indo–Malaysian region and in Eastern Polynesia from South America…. It is thus possible that the bottle gourd reached the Bismarck Archipelago *after* the break-up of POc" (emphases in original). This is compatible with the evidence presented in Section 3.2.5.1 above that the bottle gourd did not reach Southeast Asia until 200 BC and New Guinea and the Bismarck Archipelago some time later — by which time Central Polynesia was already settled (see Section 1.2.2.2). The bottle gourd literally missed the boat. Green (2000b) finds support for this in the archaeological record; no bottle gourd macroremains have been recovered from the rich plant assemblages of three waterlogged sites in Near Oceania: at Dongan on the north coast of New Guinea at 5,800 yr BP, at Kumbun Island off New Britain at 4,400 yr BP, or at the Talepakemalai Lapita site in the Mussau group at 3,000 yr BP (see also Golson, 2002). Such negative evidence, however, is fraught with problems (see for example Matthews & Gosden, 1997).

In Eastern Polynesia (Easter Island, Hawai‘i, New Zealand, Mangareva, Rarotonga, Tahiti, and the Tuamotu Islands) cognates of the term **fue* (literally "vine" (Whistler, 1988)) are used to describe the bottle gourd (Best, 1976 [1925], p. 244; Green, 1998, p. 98). This is not the case in Western Polynesian and eastern Oceanic languages where **fue* applies to a different vine — *Ipomoea macrantha*, a prostrate plant of littoral regions (Whistler, 1988; Green, 2000b)[18]. This indicates that in Eastern Polynesia a semantic shift for **fue* took place, which may have coincided with the introduction of the bottle gourd. With the widespread use of the new meaning in Eastern Polynesia, the shift probably occurred before AD 1200 (Green, 1998, pp. 98, 100;

---

[17] Importantly, the linguistic evidence also offers no direct support for a South American origin (*cf.* the sweet potato where it does).
[18] In Hawai'i, *pōhuehue* is used for beach morning glory (*I. pes-caprae*), a similar (or perhaps conspecific?) species (Kepler, 1998, pp. 135–138).

2000b). If the bottle gourd was introduced into Polynesia from South America with the sweet potato[19] (Green, 2000b; 2005; see Chapter Four) this must have occurred prior to *ca* AD 1000 — fragments of sweet potato tubers appear in Mangaia Is. at AD 988–1155 (Hather & Kirch, 1991), and fragments of probable gourd exocarp in Huahine, Society Is. between AD 850 and 1200 (Emory, 1979, pp. 202–204; Leach, 1984, p. 23).

In addition to the kumara voyage to South America, it is possible that there was contact between Hawai'i and California in AD 400–800 (Jones & Klar, 2005). This voyage could also be the source of the Polynesian bottle gourd — the distributional, linguistic and archaeological evidence presented above is equally compatible with a North American origin.

3.2.5.5 Natural or Human-Mediated Dispersal from the New World?

Recent publications (e.g., Green, 2000b, 2005) have focussed on human-mediated dispersal of the bottle gourd from the New World. Given that the gourd probably floated from Africa to the New World, this dispersal mechanism should also be considered for the Polynesian bottle gourd. As Maingay (1985, p. 104) states, in favour of natural dispersal is the "prevailing winds and currents and prehistoric Peruvian net floats made of gourds" and in favour of human-mediated dispersal is the "kumara and Pacific seamanship."

If the bottle gourd dispersed naturally from the New World, this would require its independent adoption as a crop plant in Polynesia, but not independent domestication. Because Polynesians were probably already familiar with the wax gourd (*B. hispida*) (Whistler, 1990), it seems plausible for a prehistoric beachcombing Polynesian to collect a bottle gourd, and for it to enter cultivation in the same way as that proposed for the New World (see Section 3.2.4.1). A strong case for human-

[19] Although Thor Heyerdahl discusses the bottle gourd extensively (1952, pp. 439–446), he does not explicitly advance the theory that it was introduced into the Pacific by South American voyagers. He simply states that its dispersal was certainly human-mediated — at that time it was thought the bottle gourd could not tolerate saline conditions, and therefore could not disperse naturally by floating across the sea (Heyerdahl, 1952, p. 440). Subsequent experiments (Whitaker & Carter, 1954) show that the bottle gourd can in fact survive in seawater for nearly 12 months (see Section 3.2.4.1 above).

mediated dispersal and cultural contact could be made if there were linguistic evidence connecting New World and Polynesian gourds. This is currently lacking.

3.2.5.6 The Bottle Gourd in New Zealand

An introduction of the bottle gourd into the Pacific from North or South America by AD 1000 would allow it to have been taken to New Zealand with the first Māori colonisers in AD 1100 to 1300 (Green, 2000b). The introduction of the bottle gourd to New Zealand would certainly have been human-mediated. This is supported by the linguistic evidence and the deliberate introduction of at least five other crop species (see Section 1.2.2.4), as well as Pacific food rat and dog). Both the marginal climatic conditions in New Zealand, and the apparent inability of the bottle gourd to grow in the wild preclude any possibility of natural dispersal from Central Polynesia to New Zealand.

In Māori myth the bottle gourd (*hue*), as with the sweet potato, is honoured by a personified form. Among the Ngati Awa of the Bay of Plenty the bottle gourd is believed to have originated with Pū-tē-hue, who was one of the offspring of Tane (Best, 1976 [1925], p. 245). Such a myth clearly indicates the bottle gourd's status as Māori taonga. According to the oral history of Bay of Plenty Māori, the bottle gourd was introduced to New Zealand long before the sweet potato and taro (*Colocasia esculenta*); perhaps bottle gourd seeds would more easily survive long ocean voyages than sweet potato tubers or taro corms (Best, 1976 [1925], p. 245).

In New Zealand, prehistoric bottle gourd exocarp remains have been recovered from Whakamoenga Cave on the north-eastern shore of Lake Taupo, New Zealand. The earliest of these remains have an associated radiocarbon date of AD 1345 ± 56, and were found in association with remains of the fibre plant New Zealand flax (*Phormium tenax*) and the food and bedding plant bracken fern (*Pteridium esculentum*) (Leahy, 1976; Williams & Walton, 2003). Bottle gourd seeds were recovered from a more recent layer in Whakamoenga Cave (Leahy, 1976), and whilst they cannot be definitively ascribed as prehistoric (the associated radiocarbon dates are AD 1616–1760), they are early enough that they are unlikely to represent European introductions. Pieces of exocarp

have also been recovered further south in the North Island — from the Waitore Site (near Patea), Taranaki, dated to AD 1380–1500 (Cassels, 1979).

A large number of bottle gourd exocarp fragments have been recovered from Kohika, a waterlogged archaeological site in the coastal Bay of Plenty. Kohika was a Māori village that was abandoned in the latter half of the 17th century due to flooding, leaving gourd fragments immaculately preserved in peat (Irwin, 2004b, pp. 58, 69, 74, Plate 4.29; Fig. 3.6). A single gourd seed was found at Kohika in February 2005 (Geoff Irwin, pers. comm.).

Fig. 3.6
Bottle Gourd Exocarp from Kohika, New Zealand

Well-preserved bottle gourd exocarp has been recovered from excavations of the pre-European Māori village of Kohika, Bay of Plenty, New Zealand. See trowel for scale.

Reproduced, with permission, from Irwin (2004a, Pl. 4.29)
Copyright © 2004 by G. J. Irwin

In New Zealand, as in the New World, microfossil analysis provides direct evidence for the cultivation of bottle gourd, with diagnostic bottle gourd pollen identified at several sites. Pollen from insect-pollinated plants such as bottle gourd is produced in small quantities and is thought to be deposited on the ground within a few

metres of the source plant, providing a level of spatial and temporal discrimination similar to that of phytoliths (Horrocks *et al.*, 2000; see Fig. 3.2). Bottle gourd pollen has been recovered from prehistoric coprolites (fossil/sub-fossil faecal material) discovered in sand dunes at Harataonga Bay, Great Barrier Island. The coprolites are of either human or canine origin — Horrocks *et al.* (2002) suggests human is more likely — and have an associated age of 467 ± 60 ^{14}C yr BP (Horrocks *et al.*, 2002). If the coprolites are of human origin, the pollen may have been ingested as residue on the young fruit when they were eaten (Colenso, 2001 [1880], p. 14), or, since the bottle gourd was hand-pollinated by Māori (Best, 1976 [1925], p. 250), inadvertently transferred from gardeners' hands to their mouths (Horrocks *et al.*, 2002).

Bottle gourd pollen has also been recovered from a Māori stone mound located at the base of Pouerua, an extinct volcano located 15 km inland from the Bay of Islands, Northland (Horrocks *et al.*, 2000). (Sweet potato was also cultivated at Pouerua, after Māori cleared the forest in *ca* AD 1400 (Yen & Head, 1993; Horrocks *et al.*, 2000)). However, the bottle gourd pollen was found at the same levels as pollen from European-introduced pine (*Pinus* sp.), so although the gourds grown were probably prehistoric cultivars, they may have been cultivated in the historic period (Horrocks *et al.*, 2000).

Early European explorers also recorded the bottle gourd in cultivation in New Zealand. In late October 1769, Sir Joseph Banks, aboard Captain Cook's *Endeavour*, observed seedlings of the bottle gourd at Anaura Bay on the East Coast of the North Island[20] (Beaglehole, 1962, p. 417; Leach, 1984, p. 65). In December of the same year, Monneron and l'Horne, aboard De Surville's *Saint Jean Baptiste*, saw bottle gourds at Doubtless Bay in Northland (McNab, 1914, pp. 287, 341; Leach, 1984, p. 66).

[20] Banks recorded seeing in the Māori gardens "some one of the cucumber kind, as we judgd [*sic*] from the seed leaves which just appeard [*sic*] above ground... the Cucumbers were set in small hollows or dishes much as we do in England" (Beaglehole, 1962, p. 417). It is likely that Beaglehole (1962, p. 417) is incorrect in his conclusion that the plants in "small hollows or dishes" were taro; based on the cotyledons, Banks clearly identified these plants as cucurbits (which were therefore almost certainly bottle gourd). Taro, as a monocot, has a completely different seedling morphology and, in any case, was not propagated from seed.

Both William Colenso (2001 [1880]) and Elsdon Best (1976 [1925]) provide valuable accounts of bottle gourd cultivation in early European New Zealand. Although Colenso noted “only one species and no varieties” of Māori bottle gourd, he did observe fruits ranging in size from “that of a cricket ball up to that of a globular, pear-shaped, or spheroidal figure, capable of holding several gallons” (Colenso, 2001 [1880], pp. 14–15). To me, this observation suggests many, rather than one, genetically distinct varieties. The presence of many varieties is supported by Best (1976 [1925], p. 247), who recorded 13 Māori names for different varieties of bottle gourd — although some names may be synonyms. At least one early European introduction of the bottle gourd was given a Māori name (*kōki*) (Best, 1976 [1925], p. 250), suggesting that for genetic and morphological analyses, there should be caution in accepting modern ‘Māori’ cultivars as being derived from true prehistoric Māori lineages.

The southern limit of bottle gourd cultivation in New Zealand is unknown, although Maingay (1985, pp. 77–78), summarises evidence that it could have been grown in the lower North Island and warmer parts of the South Island such as Marlborough and Banks Peninsula, but probably not farther south. Burtenshaw (2003) found that the seed of New Zealand bottle gourd cultivars appears to germinate at a lower temperature than other cultivars, suggesting selection for traits beneficial in the relatively cool New Zealand conditions. Although conditions would have been marginal, the bottle gourd was probably grown at Palliser Bay, Wairarapa, New Zealand. The land in this area was cleared in AD 1256 ± 72, and some time later mounds were constructed, in which a central post was placed. This post is thought to have served as a support for a bottle gourd vine (Leach, 1981; Leach, 1984, p. 42). Gourd fragments recovered from a cave in Fiordland at the southern tip of the South Island are almost certainly from gourds cultivated farther north and transported south (Maingay, 1985, p. 77).

3.2.5.7 A Dual Origin for the Eastern Polynesia Bottle Gourd?

The linguistic (*fue*) and distributional (Bottle Gourd Gap) evidence presented thus far is more consistent with a New World origin for the Eastern Polynesian bottle gourd, however several morphological studies (Heiser, 1979a, p. 93; Maingay, 1985, p. 102–105; Burtenshaw, 1999) all suggest an Asian contribution to the variation found in Eastern Polynesia.

Heiser grew nearly 400 bottle gourd accessions from various parts of the world, including Oceanic material from the Philippines, the Caroline Islands, New Guinea, Niue and New Zealand (Heiser, 1973b, 1973a; 1979a, p. 93). Based on morphological characters (refer to Table 3.1), the accessions from the Philippines, Caroline Islands and Niue were classified as the Asian ssp. *asiatica* (Heiser, 1973b; 1979a, p. 93), consistent with an Asian origin. The five New Zealand accessions were also classified as ssp. *asiatica*, although they did show some features more typical of ssp. *siceraria* (Heiser, 1973a). If the Niuean and New Zealand accessions were derived from New World cultivars we would expect them to be ssp. *siceraria*. The Niuean accession may be a historic introduction from Asia, because, as with the rest of Western Polynesia, it was not known in Niue prehistorically (Whistler, 1990). This leaves unexplained the five ssp. *asiatica* accessions from New Zealand (Heiser, 1973a).

Joan Maingay, in her comprehensive thesis *Te Hue: People and a Plant* (1985), similarly found that the gourds she grew from New Zealand and Eastern Polynesia possessed some characters which were atypical of the Asian subspecies (Maingay, 1985, pp. 102–105) — results which lend support to a New World origin. Maingay (1985, p. 102) found the lobing of the leaves of the New Zealand plants was intermediate between the New World ssp. *siceraria* and the Asian ssp. *asiatica*, although the leaf surfaces and margins in all varieties were closer to ssp. *asiatica*. With the seeds also, some cultivars displayed both ssp. *asiatica* and ssp. *siceraria* features (Maingay, 1985, p. 102). Whilst Asian features observed in the living plants could be attributed to introgression with historic-period introductions, this does not explain the typically ssp. *siceraria* features of two prehistoric archaeological seeds examined (Maingay, 1985, pp. 101–102). Of these seeds, Maingay says (1985, p. 103) that the "distinctive characteristics… strongly suggests that early New Zealand gourds had more

than one geographic origin" (Maingay, 1985, p. 103). Both seeds can be ascribed to ssp. *siceraria*, the New Guinean and American type, and although nonconclusive, the dimensions of one seed indicate an American rather than New Guinean origin.

Burtenshaw (1999) grew six bottle gourd cultivars, all thought to be of pre-European Māori origin, and found the leaf lobes more closely resembled ssp. *siceraria* than ssp. *asiatica* (but with slightly serrated margins). The seed colour and dimensions for these six cultivars conflicts with these results, with five cultivars having the morphology of ssp. *asiatica* and only one with the morphology of ssp. *siceraria*. So, like Maingay (1985), Burtenshaw's Māori cultivars exhibited both ssp. *siceraria* and ssp. *asiatica* features. This leads Burtenshaw to suggest that New Zealand (and Eastern Polynesian) bottle gourds have a dual origin — arriving from both Asia and the New World. It should be noted that Burtenshaw's definition of "dual" is quite different to that of Green (2000b), who proposes that *all* Eastern Polynesian bottle gourds are of South American origin, while only those from Island Southeast Asia and Island Melanesia are of Asian origin.

In summary, both Maingay and Burtenshaw found evidence for Eastern Polynesian gourds having a New World origin, but that these gourds also possessed some Asian characteristics. If the New World hypothesis is correct then the Asian characteristics could easily be explained by the modern introduction of ssp. *asiatica* cultivars to Eastern Polynesia. Of course the opposite could be the true: if Eastern Polynesian gourds have an Asian origin and the presence of the ssp. *siceraria* characteristics is due to modern introductions from Africa or the New World. Maingay cites several ways in which very early European introductions could have occurred (Maingay, 1985, p. 103).

3.2.6 SUMMARY

The bottle gourd was an important utilitarian crop species of prehistoric societies in tropical and temperate regions worldwide. It is probably of African origin, and archaeological evidence shows it was grown by 9,920 yr BP in the New World, and by 7,200 yr BP in East Asia. Morphological variation in the bottle gourd is sufficient to recognise two subspecies: ssp. *siceraria* from Africa and the New World, and ssp. *asiatica* from Asia.

Archaeological and linguistic evidence are consistent with the bottle gourd spreading into Island Southeast Asia fairly late — perhaps only about 200 BC. It may have reached as far east as the Solomon Islands by the time of the first European voyagers into the Pacific in the late 16th and early 17th centuries. Morphological analysis of cultivars from this region, specifically the phallocrypts of New Guinea, suggests they belong to ssp. *siceraria*, but genetic evidence places them in ssp. *asiatica*. In Polynesia, the bottle gourd first appears at AD 850–1200 in Huahine, Society Is. It has also been recovered from other prehistoric sites in Eastern Polynesia, including the Marquesas Is., Hawai‘i, New Zealand and possibly Easter Is.

Accurate temporal and spatial distributions of the bottle gourd have been obscured by the presence of the morphologically similar wax gourd (*Benincasa hispida* var. *pruriens*). Careful morphological and historical analyses of these two species have teased apart their distributions, revealing an area — the Bottle Gourd Gap — straddling Fiji and Western Polynesia where it appears that the bottle gourd was absent in prehistory. The Bottle Gourd Gap, together with linguistic evidence leads several authors to suggest a South American origin for the Polynesian bottle gourd. The bottle gourd could have been introduced from South America with the sweet potato. Alternatively, it could have been introduced from North America; recent research suggests Polynesian contact between Hawai‘i and the Channel Islands, California between AD 400 and 800.

A New World origin for the Polynesian bottle gourd is supported by morphological analysis of modern Māori gourds supposedly derived from prehistoric introductions, as well as morphological analysis of prehistoric seeds recovered from archaeological sites. However, these plants also possess some typically ssp. *asiatica* features, consistent with a partly Asian origin. The potential for further research, including DNA analysis, to test the hypothesis of a New World origin for the Polynesian bottle gourd is widely recognised (Heiser, 1979a, p. 97; Maingay, 1985, pp. 266–270; Burtenshaw, 1999; Green, 2000b).

3.2.7 HYPOTHESIS AND AIMS

3.2.7.1 Hypothesis

The aim is to test the hypothesis that the prehistoric Polynesian bottle gourd has a New World origin. The morphological evidence suggests at least some contribution from Asia; consequently the extent to which the Polynesian accessions represent a polyphyletic assemblage which has arisen from multiple introductions of the bottle gourd from both continents combined with post-establishment gene flow will be examined.

To test the New World hypothesis, a set of nuclear and chloroplast sequencing markers have been developed. The nuclear markers will be of sufficient resolution to differentiate Asian and New World bottle gourd cultivars, and to determine from which of these continents a collection of Polynesian (all Māori) bottle gourd cultivars are derived. The use of codominant nuclear markers will also allow Burtenshaw's hypothesis of a dual origin for the Polynesian bottle gourd to be tested.

All of the Polynesian bottle gourd accessions that could be collected with good provenance are from New Zealand. If the Polynesian bottle gourd has a single origin (from Asia *or* the New World), and the aim is to distinguish between these two basic scenarios, then the accessions from New Zealand will be a good proxy for the remainder of Eastern Polynesia. However, if the origins are more complicated (for example a dual

origin), and the bottle gourds sampled from New Zealand represent only a subset of the broader Polynesian diversity, then it is unlikely that the origins of all Polynesian lineages will be fully revealed by analysis of a limited number of New Zealand bottle gourd accessions. This caveat needs to be considered in the interpretation of the data presented here.

3.2.7.2 Aims

1. Collect seed of indigenous cultivars (landraces) of bottle gourd from Asia, the New World and Polynesia that will allow proper testing of the hypothesis.
2. Distinguish closely related cultivars of bottle gourd by developing high resolution nuclear and chloroplast markers.
3. Employ appropriate analysis techniques to analyse the DNA sequence data.
4. Determine whether the New Zealand accessions are derived from the New World, consistent with the hypothesis, or whether alternative hypotheses such as an Asian or dual Asian–New Wold origin are supported.
5. If the markers are of sufficient resolution, establish from which regions *within* the New World or Asia the Polynesian (New Zealand) bottle gourd is derived (e.g., South America, North America).
6. Establish whether the DNA data are consistent with the bottle gourd being collected from Peru–Ecuador with the sweet potato.
7. Use patterns of bottle gourd dispersal to infer patterns of prehistoric human mobility in the Pacific, and the extent to which these patterns support Polynesian voyagers reaching the New World, collecting the bottle gourd, and returning.

3.3 MATERIALS AND METHODS

3.3.1 OVERVIEW

PCR and sequencing markers were obtained for the bottle gourd using two methods. First, sequence-characterised amplified region (SCAR) markers were derived from inter-SSR (ISSR) genomic fingerprints. Second, existing markers in use in other taxa were screened for their use in bottle gourd. The existing markers included universal (amongst angiosperms) chloroplast markers, and microsatellite markers derived from commercially important Cucurbitaceae species.

3.3.2 PRELIMINARY TECHNICAL REMARKS

The preliminary technical remarks in Section 2.4.1 apply here also.

3.3.3 COLLECTION OF BOTTLE GOURD ACCESSIONS

A total of 38 accessions of bottle gourd were obtained: 13 from Asia, 15 from the Americas/New World, 8 from New Zealand (Polynesia), and 2 from Africa. The origin, cultivar name (if available), and source of each accession is provided in Table 3.2. All accessions were obtained as seeds — usually 5–20 seeds (individuals) per accession.

Table 3.2
Bottle Gourd Accession Details

Region	Accession Code	Location (Cultivar)	Source[a]	Representative Individual
Asia	020(1)Ind	Indonesia	Charles Heiser	020(1)Ind-03
	020(2)Ind	Indonesia	Charles Heiser	020(2)Ind-02
	061Ind	India	Charles Heiser	061Ind-01
	101Ind	India	Charles Heiser	101Ind-01
	111Mal	Malaysia	Charles Heiser	111Mal-01
	149Phi	Philippines	Charles Heiser	149Phi-01
	157Phi	Philippines	Charles Heiser	157Phi-01
	159Ind	India	Charles Heiser	159Ind-02
	161Mal	Malaysia	Charles Heiser	161Mal-02
	174Phi	Philippines	Charles Heiser	174Phi-01
	188Tha	Thailand	Charles Heiser	188Tha-02
	195	Yuwa-machi, Japan	Mike Burtenshaw	195-08C
	AK	Akita, Japan	Peter Matthews	AK-01
Polynesia	183(BOP)	Poverty Bay, New Zealand[b]	Mike Burtenshaw	183-23C
	BR	New Zealand (Bottle Ruku)[c]	Mike Burtenshaw	BR-06
	GA	New Zealand (Gourd 'A')	Richard Cross	GA-02C
	GD	New Zealand (Gourd 'D')	Richard Cross	GD-03C
	MA	New Zealand (Māori Gourd 1973–74)	Steve Lewthwaite	MA-01
	MG	New Zealand (Māori Gourd)	Richard Cross	MG-01
	NB	New Zealand (New Zealand Bottle)	Mike Burtenshaw	NB-04
	NP	New Zealand (Nga Puhi)	Mike Burtenshaw	NP-06
New World	006Cos	Costa Rica	Charles Heiser	006Cos-02
	027Ecu	Ecuador	Charles Heiser	027Ecu-03
	035BArg	Argentina	Charles Heiser	035BArg-03
	036(1)Per	Peru	Charles Heiser	036(1)Per-03
	036(2)Per	Peru	Charles Heiser	036(2)Per-02
	037Per	Peru	Charles Heiser	037Per-03
	051Bra	Brazil	Charles Heiser	051Bra-01
	059Cos	Costa Rica	Charles Heiser	059Cos-01
	079Mex	Mexico	Charles Heiser	079Mex-01
	093Hop	Hopi, Arizona, USA	Charles Heiser	093Hop-01

Table continued next page…

Table continued...

Region	Accession Code	Location (Cultivar)	Source[a]	Representative Individual
New World	152Mex	Mexico	Charles Heiser	152Mex-01
	153Mex	Mexico	Charles Heiser	153Mex-01
	195Per	Peru	Charles Heiser	195Per-01
	315Bra	Brazil	Charles Heiser	315Bra-01
	407Bra	Brazil	Charles Heiser	407Bra-01
Africa	291Mad	Madagascar	Charles Heiser	291Mad-02
	MR	Africa (Maranka)	Mike Burtenshaw	MR-05C

[a] Charles Heiser, Indiana University, Bloomington, Indiana, USA.
Mike Burtenshaw, The Open Polytechnic of New Zealand, Lower Hutt, New Zealand.
Peter Matthews, National Museum of Ethnology, Osaka, Japan.
Richard Cross, New Zealand Institute for Plant and Food Research, Lincoln, New Zealand.
Steve Lewthwaite, New Zealand Institute for Plant and Food Research, Pukekohe, New Zealand.
[b] The location of '183' is incorrectly given as Bay of Plenty in Clarke *et al.* (2006).
[c] Also called 'Bottle Ruka'.

The majority of the Asian and New World accessions were obtained from Charles Heiser (Indiana University, Bloomington), and most of these are described in Heiser (1973b). None of the Heiser accessions are directly derived from commercial seed companies, and most should represent indigenous varieties (landraces) of the region from which they were obtained (Heiser, 1973b).

The historic period has seen the nearly complete loss of pre-European gourds in Polynesia (Dodge, 1943, p. 86), making it difficult to obtain samples to test hypotheses about the dispersal of the species. The 8 New Zealand Māori accessions are believed to be genuine pre-European introductions, with provenance based both on discussions with the people from whom seed was obtained and the areas (isolated Māori communities) from which accessions were collected. The accessions from Plant & Food Research ('Gourd "A"', 'Gourd "D"' and 'Māori Gourd' from Richard Cross; 'Māori Gourd 1973–74' from Steve Lewthwaite) were collected from the East Cape region, North Island, New Zealand by Doug Yen *ca* 1960s (John Palmer, pers. comm.). Accession '183' was collected by Bill Sykes from Poverty Bay, North Island, New Zealand *ca* 1960s (it was subsequently sent to Charles Heiser, who then sent it to Mike

Burtenshaw). The remaining three accessions from Mike Burtenshaw ('Bottle Ruku', 'New Zealand Bottle' and 'Nga Puhi') were obtained from Koanga Gardens, Maungaturoto, New Zealand, and are all believed to represent pre-European Māori cultivars from Northland, New Zealand.

Two other species in subtribe Benincasinae (to which bottle gourd belongs) were obtained as outgroups: wax gourd (*Benincasa hispida*) and watermelon (*Citrullus lanatus*) cv. Candy Red. The wax gourd seeds were from Peter Matthews (see Table 3.2) and the watermelon seeds from Yates New Zealand.

Seeds of the 38 accessions of bottle gourd (as well as the outgroup taxa) are shown in Fig. 3.7. All remaining seeds have been archived at 4°C at the Allan Wilson Centre, Palmerston North.

3.3.4 GENOMIC DNA EXTRACTION

DNA was extracted from one randomly chosen individual of each accession — the "representative individual" (see Table 3.2).

All seeds that were known to be viable or were < 5 years old were germinated, and DNA was extracted from fresh cotyledon leaf tissue using an inexpensive CTAB protocol. Alternatively, for seeds which were non-viable or > 5 years old, DNA was extracted directly from the seed embryo using the DNeasy® Plant Mini Kit (see Section 3.3.4.3). CTAB DNA extractions from embryo tissue were unsuccessful — probably due to the high amount of starchy carbohydrate in the seeds.

3.3.4.1 Germination of Bottle Gourd Seeds

Bottle gourd seeds for germination were dusted with the fungicide Thiram (Yates New Zealand) and spread between sheets of moist paper, which were in turn placed between moist towels. The paper–towel sandwiches were placed on stainless steel trays and each

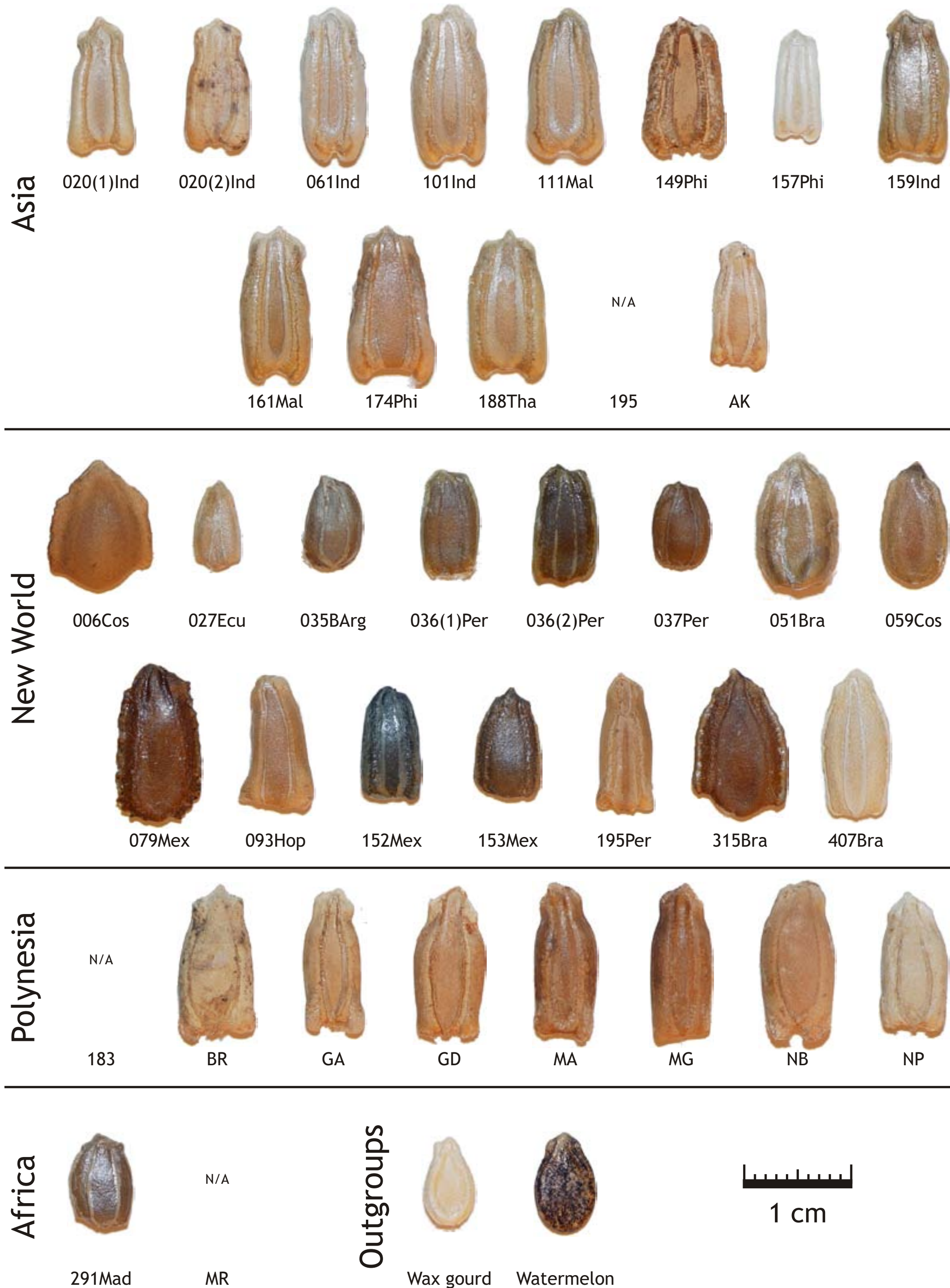
tray sealed inside a plastic bag. Trays were placed in a germinating incubator (Warren Sherer) under the following conditions: 30°C for 8 h with lights on, 20°C for 16 h with lights off (optimal conditions established by Mrs Ruth Morrison, Centre for Plant Reproduction and Seed Technology, Massey University). The seeds were incubated until they had germinated and the seedling hypocotyls had reached ~5 cm in length (10–20 d from planting).

3.3.4.2 CTAB DNA Extraction

Prior to the extraction of DNA from the samples it was necessary to manufacture a disposable pestle for grinding the tissue in a 1.7 mL disposable tube. A glass Pasteur pipette was heated in a Bunsen flame, close to the point where the pipette narrows, until the two sections separated. The wider section was then held in the flame until a small bulb formed at the separation point, forming the pestle. A new glass pestle was used for each DNA extraction to avoid cross-contamination of samples.

Total genomic DNA was extracted using a method based on that of Doyle & Doyle (1990). Approximately one half of the distal end of each cotyledon leaf (~50 mg) was used for each extraction. Tissue was transferred, using a pair of forceps wiped with 70% (v/v) ethanol, to a 1.7 mL microtube. The tube was then suspended in liquid nitrogen (N_2) for approximately 30 s. The tube was removed from the liquid nitrogen and the glass pestle used to grind the tissue until a fine powder was obtained. The tube was occasionally resuspended in liquid nitrogen during the grinding process to keep the tissue frozen and brittle.

Fig. 3.7
Seeds of Bottle Gourd and Outgroup Accessions



13 Asian, 15 American, 8 New Zealand (Polynesian) and 2 African accessions of bottle gourd, as well as the wax gourd and watermelon outgroup taxa were used. The Asian accessions possess the typical Lagenaria siceraria *ssp.* asiatica *morphology — seeds generally more than twice as long as broad, and a light colour. The New World accessions are much more variable (in shape, size and colour) but still possess the typical* L. s. *ssp.* siceraria *morphology — seeds generally less than twice as long as broad, and a dark colour. In terms of seed morphology, the Polynesian accessions are clearly closer to ssp.* asiatica*, although leaf characters place them closer to ssp.* siceraria *(Burtenshaw, 1999). See also Table 3.1.*

A 600 μL aliquot of hexadecyltrimethylammonium bromide (CTAB) isolation buffer (2% (w/v) CTAB (Sigma–Aldrich), 1% (w/v) polyvinyl-pyrrolidone (PVP; Fluka), 1.4 M sodium chloride (NaCl), 100 mM Tris-HCl (pH 8.0) and 20 mM EDTA (pH 8.0)) was added to the tube and mixed with the ground tissue using the glass pestle. The tube was then placed in a heating block at 60°C and incubated for approximately 45 min with occasional inversion of the tube.

Following incubation, a 700 μL aliquot of chloroform was added to the isolation mixture. The tube was then inverted several times and centrifuged at $5{,}000 \times g$ for approximately 10 s. This resulted in the separation of the mixture into organic and aqueous phases and compaction of cell debris at the interface. The upper aqueous phase, containing the DNA, was then transferred to a 1.7 mL microtube. This was done using a pipette with a tip from which the lower 5 mm had been removed using a sterile scalpel blade. The resulting wide-bore tip prevents mechanical shearing of the DNA during the pipetting process. Wide-bore tips were used for all subsequent steps in the extraction process that involved pipetting the DNA.

A 700 μL aliquot of ice-cold propan-2-ol was then added to the tube containing the aqueous phase solution. The tube was inverted several times and incubated on ice for 20–30 min. During this incubation period the DNA forms a cream-coloured, cotton wool-like precipitate in the propanol solution.

The DNA precipitate was then carefully removed from the solution (using a pipette) and transferred to a 1.7 mL microtube containing a 600 μL aliquot of ice-cold 80% (v/v) ethanol. The tube was gently inverted several times to remove contaminants from the DNA. This step required some care as too vigorous agitation resulted in the DNA either collapsing in on itself, trapping the contaminants and making subsequent cleaning difficult or, alternatively, breaking apart and subsequently lowering the yield. This process of transferring the DNA precipitate to a fresh 600 μL aliquot of ice-cold 80% ethanol was repeated a further two times for a total of three ethanol washes. With each wash it was possible to invert the tube with increasing vigour as progressively more contaminants were removed from the DNA and the DNA aggregated further. After the third wash the DNA appeared as a small (usually white) fluffy pellet suspended in the ethanol solution.

The DNA pellet was transferred to a 1.7 mL microtube, with as little ethanol carry-over as possible. The DNA pellet was dried for 4 h at 37°C. 50 μL $T_{10}E_{0.1}$ (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA (pH 8.0)) was added to the dried pellet and the DNA left overnight at 4°C to dissolve. The slow, overnight solubilisation of the DNA was preferable to rapid dissolution by pipetting the DNA up and down; the latter causes mechanical shearing of the DNA.

3.3.4.3 Qiagen DNeasy® DNA Extraction

As outlined in Section 3.3.4, for seeds which were no longer viable or > 5 years old, total genomic DNA was extracted directly from the seed embryo using the DNeasy Plant Mini Kit (Qiagen). The seed testa was removed using a sterile scalpel blade and the whole embryo used for the extraction. The testa was removed prior to extraction, as it is derived from the parent plant and would otherwise contaminate the DNA with the maternal genotype. Perhaps the maternal genotype could be determined from DNA extracted from the testa (useful for examining the hybridisation), but this is yet to be tested.

Tissue was ground as described in Section 3.3.4.2. DNA extractions were performed according to the manufacturer's protocol. At Step 3 the optional centrifugation step at 20,000 × *g* for 5 min was always performed. This greatly reduces shearing of the DNA and increases yield. DNA was eluted in 200 μL AE (elution) buffer (Qiagen; 10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0).

While in use, DNA (from both CTAB and DNeasy extractions) was stored at 4°C. Remaining DNA and leaf tissue (from the germinated samples) is archived at −80°C at the Allan Wilson Centre, Palmerston North.

3.3.4.4 Agarose Gel Electrophoresis of Genomic DNA

To quantify the yield and assess the quality of DNA from the extraction process, aliquots of each DNA solution were electrophoresed on an agarose gel.

A 2 μL aliquot of each CTAB DNA extraction or a 5 μL aliquot of each DNeasy DNA extraction sample reaction was combined with 1 μL 10× loading buffer in a total volume of 10 μL, and loaded onto a 1.0% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A High DNA Mass™ Ladder (Invitrogen) was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4.

Successful extractions appeared on the agarose gel as a single, intact band of high molecular weight DNA (> 12 kb). A smear of low molecular weight RNA was visible in the CTAB DNA extractions (the DNeasy DNA extractions were treated with RNase so no RNA was visible). Whilst only high molecular weight DNA could be used for generating ISSR fingerprints, slightly degraded DNA — which appeared as a smear rather than a single, intact band — was satisfactory for the more robust single-locus chloroplast and SCAR PCRs. DNA concentration was estimated by comparison with the High DNA Mass Ladder standard. CTAB DNA extractions typically yielded DNA concentrations of ~50 ng μL^{-1}. DNeasy DNA extractions typically yielded 10 ng μL^{-1}. For PCR, DNA extractions were diluted to ~1 ng μL^{-1} in $T_{10}E_{0.1}$ (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA (pH 8.0)).

3.3.5 PCR AND SEQUENCING OF CHLOROPLAST AND ISSR SCAR MARKERS

A total of 7 markers were used to sequence a representative individual from each of the 36 accessions of bottle gourd and 2 outgroup taxa (Table 3.2, Fig. 3.7). The 2 chloroplast and 5 nuclear, ISSR-derived SCAR markers are shown in Table 3.3.

Table 3.3
Chloroplast and Nuclear Bottle Gourd PCR Markers

Location	Marker	Expected Size of PCR Product (bp)	PCR Annealing Temperature	Primer Name[a]	Primer Use[b]
Chloroplast	*trn*C–trnD[c]	2800	56°C	trnC_F[d]	P
				trnD_R[d]	P
				psbM_2R[d]	S
				psbM_3F	S
	*trn*S–*trn*G	1000	50°C	ccSSR02F[e]	P
				trnG_R	P, S
Nuclear[c]	BOP19_27	740	52°C	BOP19_27_L	P
				BOP19_27_R	P, S
	BOP19_31	759	52°C	BOP19_31_L	P, S
				BOP19_31_R	P
	BOP19_35	1014	52°C	BOP19_35_L	P, S
				BOP19_35_R	P
	BR01_19	641	52°C	BR01_19_L	P
				BR01_19_R	P, S
	MR06_24	738	52°C	MR06_24_L	P, S
				MR06_24_R	P

[a] See Appendix 2 for primer sequences.
[b] P = PCR primer; S = sequencing primer.
[c] The nuclear markers are all ISSR-derived SCAR markers.
[d] From Lee & Wen (2004).
[e] From Chung & Staub (2003).

3.3.5.1 PCR of Chloroplast and SCAR Markers

Each chloroplast or SCAR marker PCR consisted of 1× PCR buffer (Roche; 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 250 µM of each dNTP, 1 M betaine (see Section 2.4.2.1.3 for a description of the benefits of betaine), 0.5 µM L primer, 0.5 µM R primer (see Appendix 2 for primer sequences), 1 U *Taq* DNA polymerase (Roche) and ~1 ng of genomic DNA in a total volume of 20 µL. The annealing temperature for each marker is shown in Table 3.3. PCR of the ISSR SCAR markers was carried out in a thermal cycler as follows: 94°C for 2 min; 35 cycles of 94°C for

30 s, annealing for 30°s, 72°C for 1 min; 72°C for 5 min; hold at 10°C. PCR of the chloroplast markers was identical except that the extension step was 68°C for 3 min for *trn*C–*trn*D, and 68°C for 1 min 30 s for *trn*S–*trn*G; the final extension step was 68°C for 5 min for both chloroplast markers. The lower extension temperature of 68°C (compared to 72°C) was found to dramatically improve yield for all products > 1 kb (and was *essential* for *trn*C–*trn*D).

The only markers that could be amplified in the outgroup taxa (wax gourd and watermelon) were the chloroplast markers *trn*C–*trn*D and *trn*S–*trn*G, and the SCAR marker BOP19_27. The chloroplast markers were amplified using the conditions described above. For BOP19_27, the annealing temperature was decreased to 50°C and the number of cycles increased to 40. The remaining 4 SCAR markers could not be amplified in either of the outgroup taxa.

Success of the ISSR SCAR marker and chloroplast marker PCR reaction was determined by electrophoresis of the PCR product. A 5 μL aliquot of each PCR reaction was combined with 1 μL 10× loading buffer in a total volume of 10 μL, and loaded onto a 1.5% (w/v) agarose/1× TAE buffer gel (as described in Section 2.4.2.1.4). A 1 Kb Plus DNA Ladder™ was loaded as a size standard, and a Low DNA Mass™ Ladder was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4. Successful amplifications appeared on the agarose gel as a single, discrete band that matched the expected size of the marker.

3.3.5.2 DNA Sequencing of Chloroplast and ISSR SCAR Markers

Prior to DNA sequencing, trial ISSR SCAR marker and chloroplast marker PCR products were treated with SAP and Exo I as described in Section 2.4.3.3.4. Due to the low level of variation, known positions of useful polymorphisms and high-quality sequence data, PCR products were sequenced in one direction only. For 6 of the 7 markers, one of the PCR primers was used as the sequencing primer (Appendix 2). *trn*C–*trn*D was sequenced with the internal sequencing primers psbM_2R (Lee & Wen,

2004) and the bottle gourd-specific psbM_3F. These sequencing primers generate ~1.5 kb overlapping sequence data in the centre of the 2.8 kb *trn*C–*trn*D marker.

3.3.5.2.1 DNA Sequencing of Chloroplast and ISSR SCAR Markers

For high-throughput DNA sequencing, the amount of BigDye® was reduced from 2.0 µL (¼ reaction) to 1.0 µL (⅛ reaction). This produced no noticeable loss in signal strength or read length.

Each DNA sequencing reaction consisted of 0.875× BigDye Terminator v3.1 Sequencing Buffer, 1.0 µL Ready Reaction Mix (BigDye v3.1), 160 nM SP primer (see Appendix 2 for primer sequences) and PCR product template (amount calculated as in Section 2.4.3.1.6) in a total volume of 20 µL. DNA sequencing reactions were carried out in a thermal cycler as described in Section 2.4.3.1.6.

3.3.5.2.2 Purification of DNA Sequencing Products by CleanSEQ®

DNA sequencing products were purified using the CleanSEQ® Sequencing Reaction Clean-Up system (Agencourt), according to the manufacturer's protocol. Each sequencing product was eluted in 40 µL 0.1 mM EDTA (pH 8.0). A 20 µL aliquot of each eluate was removed (avoiding carry-over of beads) for capillary electrophoresis.

3.3.5.2.3 Capillary Electrophoresis of Chloroplast and ISSR SCAR Markers

Each 20 µL aliquot of CleanSEQ-purified sequencing product was submitted to the AWCGS and subjected to capillary electrophoresis as described in Section 2.4.3.1.8. Because products had been eluted in EDTA, it was not necessary to add formamide (nor heat-denature the samples). Samples were briefly centrifuged to remove any micro-bubbles.

3.3.5.3 Editing and Alignment of DNA Sequence Data

DNA sequence files (electropherograms in ABI format) were imported into Sequencher™ v. 4.2 software. For each marker, sequences were aligned and uncalled or ambiguous bases edited where possible, and a NEXUS format file (Maddison *et al.*, 1997) was exported.

PCR products from individuals which were heterozygous (either for single nucleotide polymorphisms (SNPs) or length polymorphisms) were cloned and sequenced. Cloning and sequencing were carried out as described in Section 2.4.3.1. Cloning of heterozygotes was necessary to obtain clean sequence data, and also to check for potential recombination between homologous alleles.

3.3.6 DNA SEQUENCE DATA ANALYSIS

3.3.6.1 Genotype Frequency Pie Charts

Pie charts were constructed to show haplotype frequencies (cpDNA) or genotype frequencies (nDNA) for each marker in each of the three geographical regions (Asia, Polynesia and the New World). Only characters informative for testing the hypothesis were included in the analysis. Other polymorphic characters, many of which are singleton sites (a character state present in only one individual), were not included in the analysis. Although some of these singletons may be due to PCR error, others may be rare alleles that will be useful in future work when more accessions are available.

Data for the *trn*C–D and *trn*S–G chloroplast markers were concatenated because the polymorphisms are congruent between markers, and the chloroplast represents a single 'locus'. For the nDNA markers, the method of displaying genotype frequencies, rather than haplotype (allele) frequencies, allows information about each individual (i.e., whether it is homozygous or heterozygous at a given site) to be retained.

3.3.6.2 Network Analysis

Relationships between bottle gourd cultivars and their regions of origin were investigated through construction of a Spectronet network (Huber *et al.*, 2002). Each informative character in the aligned data was recoded as two genotype characters (e.g., a heterozygous SNP 'Y' was recoded as 'CT'). Gap data were recoded as nucleotide data (e.g., a homozygous deletion '00' was recoded as 'CC', a homozygous insertion '11' as 'TT', and a heterozygous indel '01' as 'CT').

3.4 RESULTS

3.4.1 DNA SEQUENCE DATA ANALYSIS

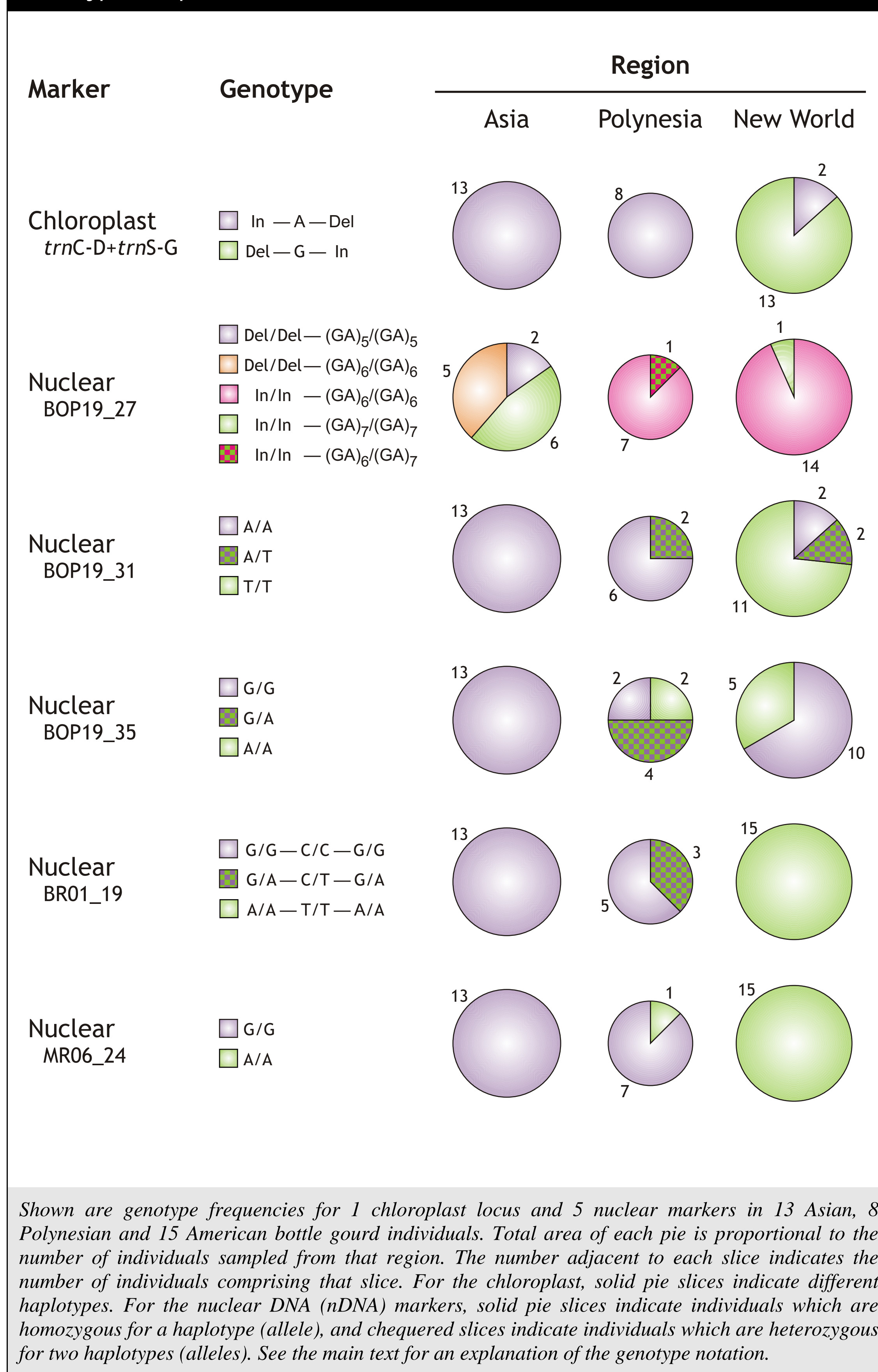
The five nuclear SCAR markers and two chloroplast markers were successfully amplified and sequenced in all 38 accessions of bottle gourd, resulting in 5.7 kb of sequence data for each accession. All sequence data have been deposited in GenBank (accession numbers DQ281822–DQ282115). The NEXUS format alignments are provided in Appendix 12 and a table of all variable sites, of which there are 49, is provided in Appendix 13. Because the indels probably arose from a single mutation event they are counted as one ‘site’ (and one character).

A wide variety of analysis methods were explored in a comprehensive search to obtain the best method for analyzing the data (e.g., various tree building methods, network building, statistical analysis) and it was decided the results were most clearly displayed as genotype frequencies presented as simple pie charts, supplemented with a Spectronet network diagram constructed from all informative sites.

3.4.1.1 Genotype Frequency Pie Charts

Genotype frequencies are shown in Fig. 3.8.

Fig. 3.8
Genotype Frequencies for 36 Accessions of Bottle Gourd



Shown are genotype frequencies for 1 chloroplast locus and 5 nuclear markers in 13 Asian, 8 Polynesian and 15 American bottle gourd individuals. Total area of each pie is proportional to the number of individuals sampled from that region. The number adjacent to each slice indicates the number of individuals comprising that slice. For the chloroplast, solid pie slices indicate different haplotypes. For the nuclear DNA (nDNA) markers, solid pie slices indicate individuals which are homozygous for a haplotype (allele), and chequered slices indicate individuals which are heterozygous for two haplotypes (alleles). See the main text for an explanation of the genotype notation.

cpDNA (trnC–D and trnS–G)

Three variable sites are observed in the chloroplast data: two indels and a G/A SNP. The polymorphisms are always concordant so that an individual is always of the In–A–Del haplotype or the Del–G–In haplotype. These In–A–Del (Asian) and Del–G–In (American) haplotypes are fixed on either side of the Pacific except for two South American individuals (one each from Brazil and Argentina, see Appendix 13) that possess the Asian haplotype. These gourds may represent a prehistoric introduction from Asia (or even Africa) but a modern introduction is also quite likely, especially given the ~500 year presence of the Spanish and Portuguese in both Southeast Asia (where gourds could have been collected) and South America (where gourds could have been dispersed). In any case, these anomalous Asian haplotypes in South America were not observed on the west coast (from where the Polynesian bottle gourd would most likely originate), and therefore are probably not implicated in the origin of the Polynesian bottle gourd. All 8 Polynesian individuals possess the Asian chloroplast haplotype.

BOP19_27

BOP19_27 possesses two informative sites (a 14 bp indel and a $(GA)_n$ microsatellite), present as four alleles: In–6, In–7, Del–5, Del–6 (where 5, 6 or 7 = n). The outgroup species wax gourd and watermelon are both In–5 (see below), indicating that the ‘In’ character state is ancestral, and the ‘Del’ state derived. Assuming that the deletion event occurred only once, and that there have been only gains and no losses of repeat units (empirical studies show microsatellite repeat unit gains occur at 2–4 times the rate of losses (see Estoup & Cornuet, 1999, pp. 61–63)), then the deletion event occurred when $n \leq 5$ and this has been followed by repeat unit gains in both lineages independently. Therefore, repeat length is homoplasious and is not a useful character. The indel, however, may be useful for further studies in Asia; although samples sizes are small, the ‘Del’ allele may represent a northern lineage (it is the only one in Japan) and the ‘In’ allele a southern lineage (it is the only one in India).

BOP19_31 and BOP19_35

Both the BOP19_31 and BOP19_35 markers possess an informative SNP where one allele is present Asia, Polynesia and the New World and the other allele in Polynesia and the New World only.

BR01_19

Three informative SNPs are observed and are concordant so that an individual is always of the G–C–G haplotype and/or the A–T–A haplotype. Asia is fixed for the G–C–G haplotype so that all individuals possess two copies of this allele (denoted as the G/G–C/C–G/G genotype). The New World, however, are fixed for the A–T–A haplotype so that all individuals are of the A/A–T/T–A/A genotype. The Polynesian individuals possess either the Asian genotype or are heterozygous for both the Asian G–C–G haplotype and the American A–T–A haplotype; this heterozygous genotype is denoted as G/A–C/T–G/A. Cloning of heterozygous individuals showed there was no recombination between these two haplotypes. BR01_19 was the only marker that had a significant match to any sequence in GenBank (a putative serine/threonine protein kinase in *Arabidopsis thaliana*; E value = 5×10^{-4}).

MR06_24

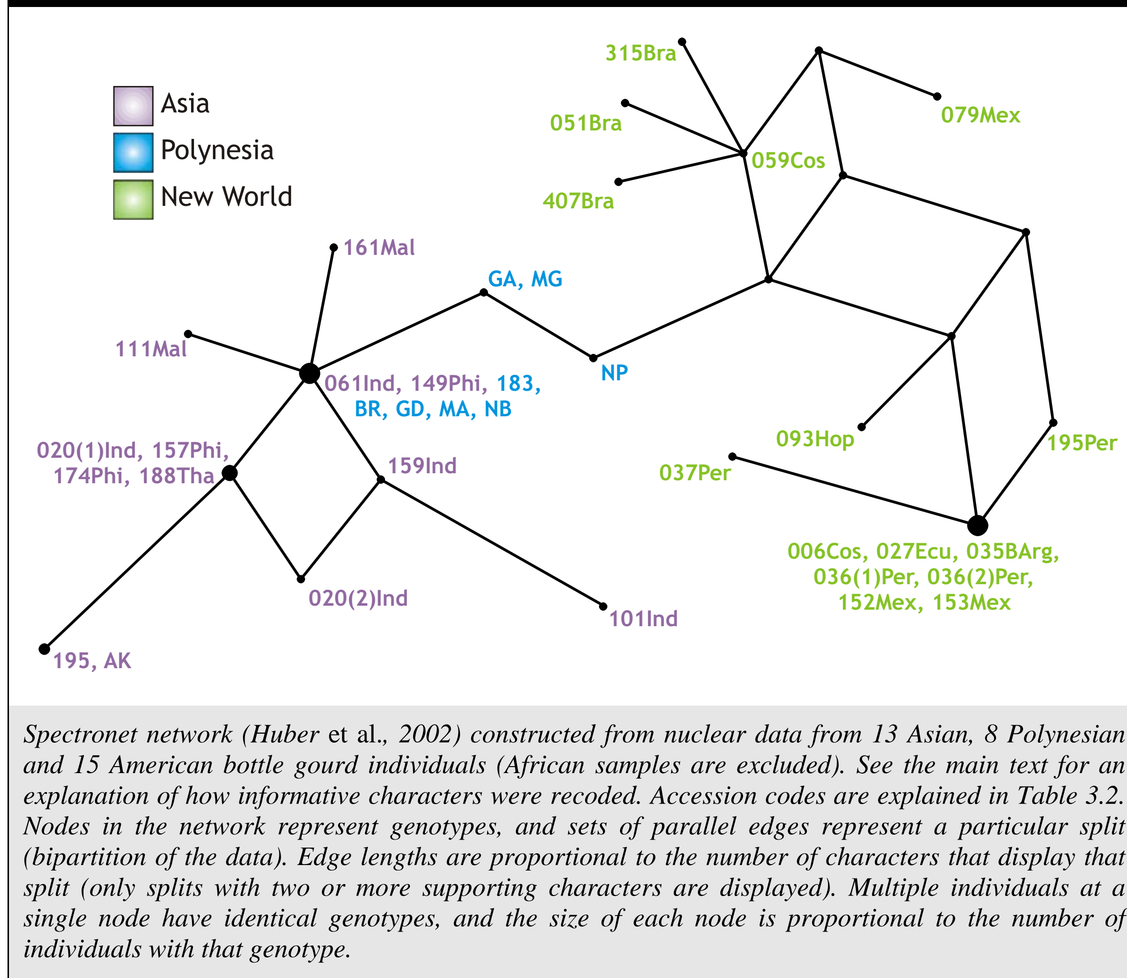
One informative SNP is observed that again is fixed in Asia (G/G genotype) and in the New World (A/A genotype) but has both the G and A alleles present in Polynesia.

The BR01_19 and MR06_24 nuclear markers are fixed for different alleles on either side of the Pacific, clearly separating the Asian and American bottle gourds (congruent with the subspecies taxonomy), and allowing us to infer from which side of the Pacific the Polynesian bottle gourds originate.

3.4.1.2 Network Analysis

The Spectronet network (Fig. 3.9) is consistent with the genotype frequencies shown in Fig. 3.8. The Asian and American bottle gourd accession form two distinct groups, with five of the New Zealand accession grouping with the Asian accessions and the remaining three New Zealand accessions possessing genotypes intermediate between those of the Asian and American accessions. Although the New World accessions form two groups with a number of inferred haplotypes, this might just be a sampling effect; there is no obvious geographic pattern that corresponds with these the New World accessions forming two groups.

Fig. 3.9
Spectronet Network of Bottle Gourd Nuclear Data



Spectronet network (Huber et al.*, 2002) constructed from nuclear data from 13 Asian, 8 Polynesian and 15 American bottle gourd individuals (African samples are excluded). See the main text for an explanation of how informative characters were recoded. Accession codes are explained in Table 3.2. Nodes in the network represent genotypes, and sets of parallel edges represent a particular split (bipartition of the data). Edge lengths are proportional to the number of characters that display that split (only splits with two or more supporting characters are displayed). Multiple individuals at a single node have identical genotypes, and the size of each node is proportional to the number of individuals with that genotype.*

3.4.1.3 Outgroup Analysis

The *trn*C–D and *trn*S–G markers were also sequenced in two outgroups, wax gourd (*Benincasa hispida*) and watermelon (*Citrullus lanatus*), which both possess the 'Asian' haplotype, consistent with the Asian haplotype being ancestral in bottle gourd and the African derived (Appendix 13).

3.5 DISCUSSION

3.5.1 ORIGINS OF THE POLYNESIAN BOTTLE GOURD

3.5.1.1 Genetic Evidence

An Asian origin for the Polynesian bottle gourd is supported by the presence of the Asian chloroplast haplotype in New Zealand bottle gourds. Nuclear markers, however, indicate there is also a significant genetic contribution from the New World and that the New Zealand bottle gourds are the result of hybridisation(s) between cultivars from both continents (also supported by network analysis, see Fig. 3.9). New Zealand individuals which possess American alleles do not possess them at all markers (Appendix 13), suggesting that although some of the New Zealand accessions are hybrids between Asian and American cultivars, these individuals are not the first generation (F_1) of such a cross.

3.5.1.2 Reassessing the Evidence Against an Asian Origin

Both the chloroplast and nuclear data from the New Zealand accessions strongly support a partly Asian origin for the Polynesian bottle gourd. Although human-mediated dispersal is not concordant with both the apparently late introduction of the bottle gourd into Southeast Asia (~200 BC), and the Bottle Gourd Gap, it is premature to exclude human-mediated dispersal of the Polynesian bottle gourd from Asia (Fig. 3.5). Firstly, the bottle gourd was present in Taiwan during the Austronesian expansion out of this area ~5,000 yr BP (Bellwood, 1997, pp. 212, 217) so it was available to be taken farther south as new islands were settled. Secondly, the 'negative evidence' of the lack of bottle gourd in archaeological sites from Melanesia does not prove its absence as there is no body of wetland archaeobotanical research from this area (the excavations from highland New Guinea are exceptional; Peter Matthews, pers. comm.). Thirdly, the apparent Bottle Gourd Gap region is nested within the distribution of known Lapita sites

(Fig. 3.5), where pottery was abundant and the bottle gourd may not have been required as a container (Burtenshaw, 2003)[21].

3.5.1.3 Summary

The nuclear data from the New Zealand accessions are consistent with a partly American origin for the Polynesian bottle gourd although, as for the Asian contribution, it cannot be resolved whether this dispersal was natural or human-mediated. The latter is certainly possible if it is established that Polynesians sailed to South America (and collected the sweet potato (Green, 2005)) or the Channel Islands (Jones & Klar, 2005), as the bottle gourd could have been collected on such a voyage.

3.5.2 RECONCILING ERICKSON *ET AL.* (2005) AND CLARKE *ET AL.* (2006)

Consistent with models for the dispersal of the Polynesian bottle gourd, the ssp. *siceraria* alleles present in the Māori bottle gourds are hypothesised to originate in the New World. Modern American bottle gourds belong to ssp. *siceraria*, and this was thought to be true of prehistoric American bottle gourds also; there was no evidence to suggest that modern American gourds are not derived from the prehistoric diversity. However, recent ancient DNA analysis of 12 prehistoric gourd exocarp fragments from sites across the New World (Erickson *et al.*, 2005) provides convincing evidence that prehistoric American bottle gourds in fact belonged to ssp. *asiatica*! This finding removes the only realistic source of the ssp. *siceraria* alleles found in Polynesia, and creates a number of immediate problems which need to be solved before progress can be made in understanding the origins of the Polynesian bottle gourd using molecular techniques. This section will summarise the ancient DNA research performed, and its implications for understanding the origins of the New World and Polynesian bottle gourds. Three testable hypotheses which may explain the incompatibility in the data will be presented.

---

[21] The replacement of gourds with pottery containers has been noted in China (Whitaker & Bird, 1949; Walters, 1989).

3.5.2.1 Background to Collaboration

In 2004 we formed a collaboration with Bruce Smith's group at the National Museum of Natural History (Smithsonian Institution, Washington, D.C.). Dr Smith's group is also interested in using molecular techniques to answer question about the origins of the bottle gourd, and although we are each interested in dispersal events in different geographical regions — them in the New World, us in Polynesia — it made sense to collaborate on developing markers, which for both groups was proving to be a very difficult process (they had sequenced > 10 kb mtDNA without locating any polymorphisms). From July–September 2004 I undertook an internship in the Smithsonian's Laboratories of Analytical Biology, which resulted in the development of the *trn*C–D and *trn*S–G chloroplast markers, as well as a number of other potential markers which are yet to be pursued.

Back in New Zealand, I continued working on the Polynesian bottle gourd project using the cpDNA markers developed at the Smithsonian as well as five nuclear markers developed at the Allan Wilson Centre. This work was published in 2006 (Clarke *et al.*, 2006; Appendix 9). Meanwhile, the Smithsonian group completed the New World bottle gourd project, and although those results were produced after the results for the Polynesian project, the New World paper was published about six months earlier, appearing in late 2005 (Erickson *et al.*, 2005; Appendix 10).

3.5.2.2 Summary of Erickson *et al.* (2005)

Erickson *et al.* (2005) integrated genetic and archaeological approaches to address a range of long-standing questions regarding the introduction of the bottle gourd to the New World. Did it reach the New World directly from Africa or through Asia? Was it transported by humans or ocean currents? Was it wild or domesticated upon arrival?

Twelve bottle gourd exocarp fragments were collected from eight sites in the New World (Table 3.4). AMS radiocarbon dating confirmed eleven fragments were prehistoric, ranging from 9,920 yr BP (Guilá Naquitz) to AD 1040–1190 (Ancon), and one was historic (AD 1660, Coxcatlan Cave). Rind thickness values for undamaged exocarp ranged from 3.0–7.0 mm (excluding the Guilá Naquitz sample; see Table 3.4), consistent with a domesticated status; *L. abyssinica* and *Cucurbita* spp. rind thickness values were all < 2.0 mm (Erickson *et al.*, 2005).

Table 3.4
Ancient New World Bottle Gourds from Erickson *et al.* (2005)[a]

Archaeological Site	AMS Age (calendar yr BP)	Rind Thickness (mm)	Marker Genotype		
			LS_InDel1	LS_SNP	LS_InDel2
Coxcatlan Cave, Mexico	210 ± 40, 220 ± 20	3.8	African	African	African
Ancon, Peru	790 ± 40	4.8	Asian	Asian	Asian
Ancon, Peru	900 ± 40	4.3	Asian	Asian	Asian
El Coyote Cave, Mexico	1,000 ± 30, 1,090 ± 35	4.9	Asian	Asian	N/A
Cloudsplitter Cave, Kentucky	2,735 ± 35	4.2	Asian	N/A	Asian
Mammoth Cave, Kentucky	2,750 ± 40, 2,760 ± 30	7.0	Asian	Asian	Asian
Coxcatlan Cave, Mexico	7,200 ± 50, 7,230 ± 50	3.2	Asian	Asian	Asian
Windover, Florida	8,105 ± 120	3.0	No PCR amplification		
Quebrada Jaguay, Peru	8,410 ± 50	3.4	Asian	Asian	Asian
Quebrada Jaguay, Peru	8,415 ± 50	3.6	Asian	Asian	Asian
Guilá Naquitz, Mexico	8,685 ± 60, 9,030 ± 35	4.6	Asian	Asian	Asian
Guilá Naquitz, Mexico	9,920 ± 50	2.2[b]	No PCR amplification		

[a] Reproduced from Table 1, Erickson *et al.* (2005).
[b] An unknown amount of the interior wall segment had spalled off the 2.2 mm-thick fragment from Guilá Naquitz, yielding only a partial rind thickness value.

Because the DNA in the specimens was so degraded, the three polymorphic sites in the *trn*C–D and *trn*S–G markers were amplified in separate fragments, ranging from 95–125 bp. Using appropriate aDNA practices, at least one (and in most cases all three) sites were amplified in 10 of the 12 samples — only the oldest sample from Guilá Naquitz (which was also burned) and the waterlogged sample from Windover did not yield amplifiable DNA. Comparison of the aDNA haplotypes with modern reference sets of well-provenanced bottle gourds from Africa and Asia, revealed all prehistoric American samples possess the Asian haplotype, consistent with an Asian origin.

If the exocarp of wild bottle gourd is as thin and fragile as that documented for other cucurbit taxa (including the wild Zimbabwean bottle gourd (Decker-Walters *et al.*, 2004b)), the probability that wild bottle gourd fruits were able to float from Asia to the New World is considerably reduced. Therefore, if dispersal was by floating, it was probably of a thick-shelled, domesticated bottle gourd, which could have been carried eastward from Asia along the north Pacific current rapidly enough to reach the New World with still-viable seeds. This conclusion is based on recent drift and diffusion analyses of container ship spills of buoyant cargo (e.g., plastic bath toys) in the North Pacific (Ebbesmeyer & Ingraham, 1994).

But rather than natural dispersal, Erickson *et al.* (2005) favour human-mediated dispersal from Asia — by Paleoindian groups who carried the bottle gourd and still-viable seeds along the coast of Beringia, either on foot or in near-shore watercraft, rapidly enough to have introduced domesticated bottle gourd to the New World. The introduction of the bottle gourd may have been associated with the introduction of another early 'utilitarian domesticate', the dog (*Canis familiaris*) (Wayne *et al.*, 2006).

Whilst I agree that the bottle gourd introduced to the New World was already domesticated, I think that — in addition to a Paleoindian-mediated introduction via Beringia — natural dispersal by floating also needs to be considered. The Beringian climate was probably too cold for the tropical–sub-tropical bottle gourd. For example, at eight sites in Alaska, mean July (summer) temperatures in the Late Wisconsinan/Early Holocene (20,000–10,000 yr BP) are estimated from 9.3–13.5°C (Elias *et al.*, 2000; Elias, 2001). (As the bottle gourd was present in the interior southern highlands of Mexico by 10,000 yr BP, this is the upper limit for the date of introduction to the New

World.) Based on comparisons with modern cultivars, these Late Holocene temperatures are significantly below those required to grow bottle gourd. Dunedin, New Zealand, with a mean January (summer) temperature of 15.2°C (NIWA Science, 2001) — whilst warmer than early Holocene Beringia — is still too cold for bottle gourd (Maingay, 1985). They will however, grow in Christchurch, New Zealand (John Palmer, pers. comm.), which, with a mean January temperature of 17.4°C (NIWA Science, 2001), is probably near the climatic limit of bottle gourd cultivation. Thus if bottle gourds weren't *grown* in Beringia, the Paleoindian hypothesis requires that seeds were carried, on foot or in near-shore watercraft, from the northern cultivation zone of Asia to at least the northern limit of cultivation in the New World. Based on archaeological and ethnographic data (see Section 3.2.4), this cultivation limit may have been as far south as 40–45°N in both continents (northern Honshu, Japan and southern Great Lakes, USA), requiring a Beringian-era journey of at least 8,000 km.

3.5.2.3 Implications of a Non-African Origin for the New World Bottle Gourd

An Asian origin for the New World bottle gourd resolves a long-standing chronological problem — that the bottle gourd appears in the archaeological record of Africa nearly 6,000 years after it appears in Mexico, despite the New World bottle gourd supposedly being derived from a domesticated African bottle gourd. An Asian origin for the New World bottle gourd instead allows for an independent domestication of this species in Africa much later than in Asia — maybe only as recently as ~5,000 yr BP

Based on morphological and molecular evidence, modern American gourds are of African origin (Heiser, 1973b; Clarke *et al.*, 2006). But if pre-Columbian American bottle gourds are of Asian origin then, during the last 500 years, these have been replaced by post-European contact gourds from Africa — a process which began very early (African bottle gourd are in the Tehuacán Valley, Mexico by the 1660s), was widespread and has gone close, or completely, to completion. Clarke *et al.* (2006) show that of 15 modern American gourds sampled, 13 have the typically African chloroplast haplotype (this haplotype and that of the African reference set in Erickson *et al.* (2005) are identical). In addition, all 15 individuals possess what is a putatively African

haplotype for the BR01_19 and MR06_24 nuclear markers (although no African bottle gourds have yet been sequenced with these markers, the American haplotypes *never* appear in Asia). Such widespread and complete replacement of prehistoric Asian bottle gourds in both North and South America suggests, as well as multiple introductions, that there has been deliberate and sustained selection for African bottle gourds; for some characteristics, which are as yet unknown, African gourds are superior to their Asian counterparts that had been grown in the New World for 10,000 years.

3.5.2.4 Implications for Clarke *et al.* (2006)

In determining the origin of the Polynesian bottle gourd it was assumed that the bottle gourd in prehistoric New World originated from Africa, not Asia, and therefore belonged to ssp. *siceraria*. This was the position of the overwhelming majority of authors (e.g., Kobiakova, 1930; Whitaker & Carter, 1954; Organ, 1963, p. 34; Heiser, 1973b; Lathrap, 1977; Heiser, 1979a, 1989, 1990; Cooke, 2005)[22]. It was also supported by morphological and genetic studies of modern American accessions (Heiser, 1973b; Decker-Walters *et al.*, 2001). Under this assumption, the 'African' spp. *siceraria* alleles present in bottle gourds from New Zealand were hypothesised to have come from the New World (Clarke *et al.*, 2006). The surprising result from Erickson *et al.* (2005) was that the prehistoric American bottle gourds in fact possess only ssp. *asiatica* alleles, consistent with dispersal from Asia, not Africa. Therefore, if only ssp. *asiatica* alleles were present in both Asia and the New World prehistorically, the origin of the New Zealand ssp. *siceraria* alleles becomes an immediate problem, forcing a number of new, but testable, hypotheses to account for their presence in Māori bottle gourds:

1. The Māori bottle gourds sampled have been introgressed with historic-period introductions of African or modern American gourds possessing ssp. *siceraria* alleles.
2. The ssp. *siceraria* alleles were present in Polynesia prehistorically, but were introduced from somewhere other than Asia or Africa.

---

[22] Very few authors have suggested an Asian origin for the New World bottle gourd. Exceptions include Camp (1954) and Carter (1977, pp. 118–120).

3. There were ssp. *siceraria* alleles present in the New World prehistorically (perhaps from multiple independent introductions from different regions), but:
    a) the sample size of Erickson *et al.* (2005) was too small for them to be detected, or
    b) the aDNA results of Erickson *et al.* (2005) are contaminated with Asian bottle gourd DNA.

Each of these hypotheses will now be explored in greater detail because they provide a framework for resolving the incompatibility between Clarke *et al.* (2006) and Erickson *et al.* (2005).

3.5.2.4.1 Hypothesis 1: Introgression of Modern Māori Samples

As described in Section 3.3.3, all of the Māori bottle gourds are thought to be derived from prehistoric New Zealand cultivars, with provenance based on information from seed suppliers (heritage seed companies, Māori families) and the locations from which seed was obtained (isolated East Cape Māori communities in the 1950s and 1960s). Nevertheless, these gourd accessions have had to withstand up to ~200 years of potential introgression, especially pollen-mediated, from historically-introduced gourds. In North America, preservation of true cultivars is hindered by failure to isolate cultivars from each other and recent cross-pollination (Decker-Walters *et al.*, 2001). aDNA analysis of prehistoric and protohistoric bottle gourd material from New Zealand and other Polynesian island will be the best method for establishing prehistoric genotypes in Eastern Polynesia. The potential of ancient DNA research is described in more detail in Chapter Five.

3.5.2.4.2 Hypothesis 2: A Non-Asian, Non-American Origin for Polynesian ssp. siceraria *Alleles?*

If aDNA research of Polynesian material reveals only ssp. *asiatica* alleles in prehistoric Polynesia, then, consistent with Erickson *et al.* (2005), they could have come from either Asia or the New World. With the current markers, which can probably only differentiate accessions to the subspecies level, it will be difficult to discriminate

between these two alternatives. New markers, which can separate Asian ssp. *asiatica* gourds from American ssp. *asiatica* gourds, may be required.

If aDNA research of Polynesian material confirms ssp. *siceraria* alleles in prehistoric Polynesia, then, consistent with Erickson *et al.* (2005), these alleles have most likely come from somewhere other than Asia or the New World. The origin of the morphologically African penis gourds in New Guinea is unresolved — the analysis of two penis gourds by RAPD fingerprinting which places them closer to ssp. *asiatica* (Decker-Walters *et al.*, 2001) is insufficient to make any firm conclusions. More wide-scale testing of the large number of penis gourds in collections worldwide (e.g., at the Smithsonian Institution) is required. An unexplored explanation, but mentioned by Heiser (1973a), to account for the morphologically African gourds in Oceania (New Guinea and Eastern Polynesia) is prehistoric contact with Africa, specifically Madagascar. Austronesian voyagers reached Madagascar about between AD 300 and 800 (Diamond, 1997, pp. 392–393), probably from Borneo (Hurles *et al.*, 2005). Perhaps return voyagers to Island Southeast Asia could have introduced African bottle gourds into this region, where they subsequently spread into the New Guinea highlands. This hypothesis is compatible with the apparently late introduction of the bottle gourd into New Guinea, although the linguistic evidence supports an introduction from India via Island Southeast Asia, rather than Africa. However, an introduction at this time is probably far too late to account for African gourds in Eastern Polynesia — AD 300 postdates the break-up of the Proto-Oceanic language subgroup by at least 1,500 years, and the bottle gourd would have to have spread quickly through the Bottle Gourd Gap region to be deposited in the Societies by AD 1050 (Emory, 1979, pp. 202–204).

3.5.2.4.3 Hypothesis 3: Archaeological Bottle Gourd Seeds in Peru: An African Origin?

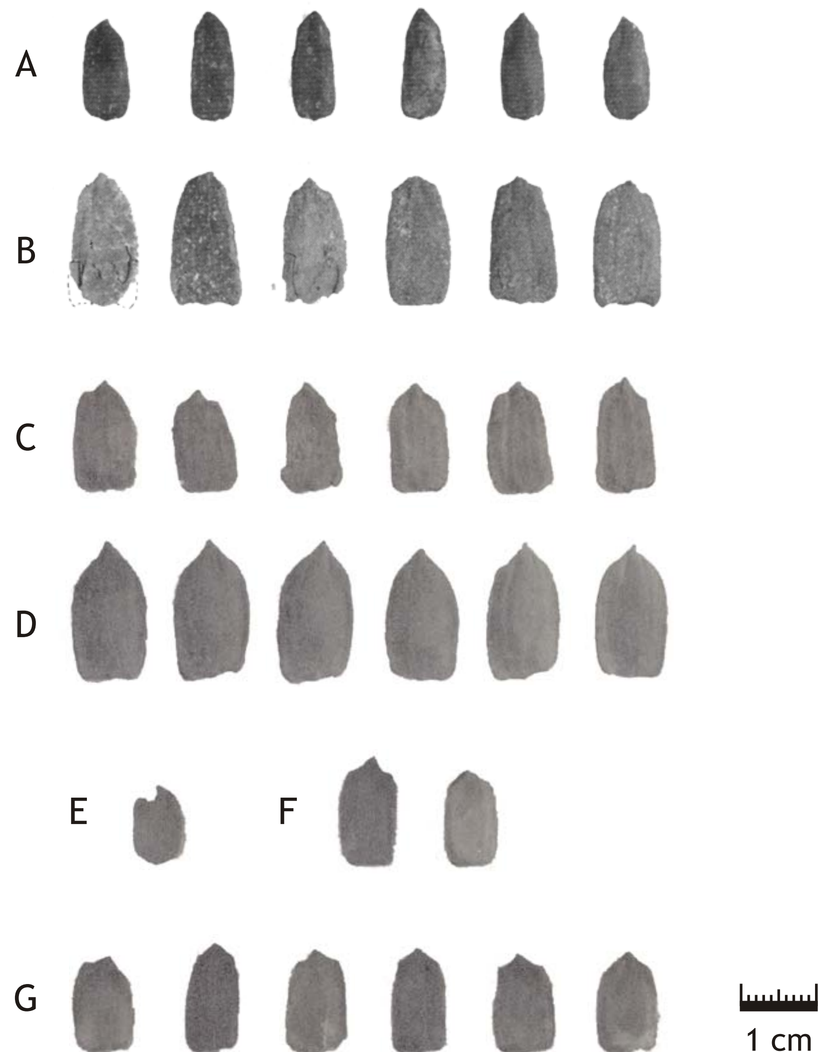
Despite the results of Erickson *et al.* (2005), it is premature to rule out the possibility that ssp. *siceraria* alleles existed in the New World prehistorically. Although Erickson *et al.* (2005, Fig. 2, Fig. 8) show two 1,000-year-old, typically ssp. *asiatica*, seeds from Cold Oak Rock Shelter, Kentucky, other sites in the New World, namely five sites in Peru, have yielded numerous prehistoric seeds that are morphologically closer to ssp.

siceraria (Whitaker & Bird, 1949; Towle, 1952). Seeds from these five sites — Huaca Prieta, Aspero, Huaca de la Cruez, Castillo de Tomaval, and Pachacamac — are shown in Fig. 3.10. The locations of the sites are shown in Fig. 3.4.

Modern African and American (ssp. *siceraria*) seeds typically have a length:width ratio (L:W) < 2, while modern Asian (ssp. *asiatica*) seeds typically have a L:W ratio > 2 (Heiser, 1973b; Table 3.1). To allow a subspecies designation of the prehistoric Peruvian material, the seeds shown in Fig. 3.10 were measured, and a mean L:W value calculated. There are, however, significant taphonomic issues with calculating L:W from archaeological seeds. From examination of surface features, it is clear that archaeological bottle gourd seeds erode through time (Towle, 1952). This occurs preferentially along the sides of the seeds (i.e., along the longitudinal axis), causing the seeds to become narrower, increasing L:W, and effecting a more 'Asian' appearance. For example, Towle (1952) found, that for seeds recovered from a single gourd fruit, L:W increased from 1.75 for well preserved seeds, to 2.0 for badly eroded seeds.

The seeds in Fig. 3.10B–D, F–G were measured, providing a mean L:W of 1.89. The seeds from inside the fishnet float (Fig. 3.10A) were excluded from the analysis because Towle (1952) provides convincing evidence that these are eroded. The single, damaged seed from Castillo de Tomaval (Fig. 3.10E) was also excluded. To establish a modern baseline for comparison, mean L:W values were calculated for 37 New World seeds, 37 African seeds, and 28 Asian seeds (photos from Heiser, 1973b). In addition, the mean L:W for 25 New Zealand seeds (photos from Maingay, 1985, pp. 85–87) was calculated. Although some L:W data were already available for the material, all seeds were re-measured to remove measurer-associated bias. Raw measurement data are provided in Appendix 14. Mean L:W values are shown in Table 3.5.

Fig. 3.10
Prehistoric Bottle Gourd Seeds from Five Sites in Peru



*Most of these prehistoric bottle gourd (*Lagenaria siceraria*) seeds recovered from five sites in coastal Peru possess the typically African ssp.* siceraria *morphology, with length:width ratios of < 2:1 (see Table 3.1 and Fig. 3.7). The only seeds with a length:width ratio > 2:1 are shown in* **A.**, *but these seeds may be badly eroded, causing them to become narrower (Towle, 1952; see main text). The seeds from Huaca Prieta date to 5,000 yr BP (Whitaker & Bird, 1949; Whitaker, 1983), and the others range from 2,500 yr BP to AD 1400 (Towle, 1952). See Fig. 3.4 for the locations of the sites.*

A. *From Huaca Prieta;* $\bar{x}_{L:W} = 2.52$. *These seeds are from a fishnet float shown in Fig. 3.3.*

B. *From Huaca Prieta;* $\bar{x}_{L:W} = 1.99$.

C. *From Aspero (specimen 747/41A);* $\bar{x}_{L:W} = 1.90$.

D. *From Huaca de la Cruez (specimen 3/V-162);* $\bar{x}_{L:W} = 1.84$.

E. *From Castillo de Tomaval (specimen 21/V-51); seed not measured (damaged).*

F. *From Pachacamac (specimen 135/41A);* $\bar{x}_{L:W} = 1.96$.

G. *From Pachacamac (specimen G81);* $\bar{x}_{L:W} = 1.78$.

A.–B. reproduced, with permission of the American Museum of Natural History, New York, NY, USA, from Whitaker & Bird (1949, Fig. 3)
Copyright © 1949 by the American Museum of Natural History
C.–G. reproduced, with permission of the Economic Botany Library of Oakes Ames, Harvard University, Cambridge, MA, USA, from Towle (1952, Pl. LVII–LIX)
Copyright © 1952 by Harvard University

Table 3.5
Seed L:W Ratios in New Zealand, the New World, Africa and Asia[a]

Context	Region	Seed Length:Width Ratio			Reference
		n	Mean ($\bar{x}$)	Std Dev. (s)	
Modern	**New Zealand**	25	2.28	0.310	Maingay (1985, pp. 85–87)
	New World	37	1.98	0.439	Heiser (1973b)
	Africa	37	2.03	0.450	Heiser (1973b)
	Asia	28	2.31	0.299	Heiser (1973b)
Prehistoric	**Peru**	26	1.89	0.146	Whitaker & Bird (1949); Towle (1952)

[a] Raw measurement data are provided in Appendix 14.

Pair-wise differences between L:W values for the five regions are shown in Table 3.6. To establish whether any significant differences between mean L:W values for the five regions exists, pair-wise t-tests were performed ($\alpha = 0.05$). The resulting P-values are shown in Table 3.6.

Table 3.6
Differences in Mean Seed Length:Width Ratios Between Regions[a,b]

	New Zealand	New World	Africa	Asia
New World	−0.30 (**0.002**)			
Africa	−0.26 (**0.010**)	0.05 (0.660)		
Asia	0.02 (0.781)	0.33 (**0.001**)	0.28 (**0.004**)	
Peru	−0.40 (< **0.001**)	−0.09 (0.234)	−0.14 (0.085)	−0.42 (< **0.001**)

[a] Differences shown = $\bar{x}_{row} - \bar{x}_{column}$. Differences are calculated from the mean values in Table 3.5.
[b] P-values for pair-wise t-tests between seed:length ratios of each region are shown in parentheses. H_0: $\mu_2 - \mu_1 = 0$; H_1: $\mu_2 - \mu_1 \neq 0$; $\alpha = 0.05$. Significant P-values are shown in bold.

No significant difference in mean L:W values between modern African and American seeds is observed ($P = 0.660$), consistent with the current taxonomy (Heiser, 1973b) and genetics (Erickson *et al.*, 2005; Clarke *et al.*, 2006) placing gourds from these regions in ssp. *siceraria*. Bottle gourds from Asia were significantly different from gourds from both Africa ($P = 0.004$) and the New World ($P = 0.001$), consistent with the taxonomy placing Asian gourds in ssp. *asiatica*.

Based on Erickson *et al.* (2005), who conclude an Asian origin for the American bottle gourds, we would predict no significant difference between L:W values for modern Asian gourds and prehistoric Peruvian gourds. However, the difference is highly significant ($P < 0.001$), and therefore inconsistent with Peruvian gourds originating from Asia. Furthermore, the Peruvian gourds are not significantly different from modern American ($P = 0.234$) or African ($P = 0.085$) gourds. Therefore, L:W analysis of the prehistoric Peruvian seeds is more consistent with an African, rather than an Asian, origin, and apparently incompatible with the aDNA results of Erickson *et al.* (2005).

If the aDNA results prove to be incorrect or incomplete, then there are three significant implications of prehistoric American bottle gourds actually belonging to ssp. *siceraria*:

1. The paradox of an early presence in the New World (10,000 yr BP) but late presence in Africa (2000 BC) is re-established.
2. It is not necessary to evoke widespread and total replacement of indigenous American bottle gourds with modern African gourds; the modern diversity in the New World could derive from prehistoric American gourds.
3. The New World is re-established as a potential source of ssp. *siceraria* gourds, i.e., ssp. *siceraria* alleles found in Polynesia — such as those described in Clarke *et al.* (2006) — could be derived from the New World. In addition, the presence of morphologically ssp. *siceraria* seeds in Peru matches the predicted landing point for Polynesian voyagers (Green, 2005; Scaglion, 2005).

In summary, the morphological analysis of the prehistoric Peruvian gourds suggests that the origins of the American gourds may not be completely resolved by Erickson *et al.* (2005). The morphological analysis does however have limitations:

1. Seed morphology may have changed through time, leading to convergence.
2. The Peruvian seeds may not be representative of prehistoric material from the wider New World (see Whitaker & Bird (1949) and Towle (1952) for further discussion).

The only accurate L:W value I have obtained for North America is from a photograph of a single ~3,000 yr BP seed from the Ocampo caves, Mexico (Whitaker *et al.*, 1957, Fig. 3). This seed has an L:W value of 1.87, comparable to the values obtained for prehistoric Peru.

3.5.2.5 A Reconciliation?

A resolution to the American origins question will most likely come from further aDNA research that is carried out on greater numbers of bottle gourd remains that more fully represent the temporal and spatial diversity of the species in the New World. It will be particularly interesting to perform aDNA analysis of the prehistoric Peruvian material shown in Fig. 3.10. Genetically, do the seeds belong to ssp. *siceraria*, consistent with the morphology, or to ssp. *asiatica*, consistent with the results and conclusions of Erickson *et al.* (2005)? Morphological analyses of prehistoric bottle gourd seeds from other New World locations may also be useful, but, as is the case with Towle (1952), reports of seeds are often published in obscure publications that are difficult to find (but see Towle, 1961, p. 94). Without a better understanding of the origins of the American bottle gourd, it will be impossible to attack the same question in Oceania, and further research on the origins of the New World bottle gourd should be a priority.

Also included in the morphological analyses were the 27 New Zealand seeds from Maingay (1985, pp. 85–87). L:W values for these seeds show they are significantly different from modern African, modern New World, and prehistoric Peruvian seeds, but not from Asian seeds. This result suggests an Asian, not South American, origin for the New Zealand bottle gourds. But it is already evident from the cpDNA and nDNA analysis of the eight New Zealand accessions that a relatively complex picture is emerging for Polynesia, with hybridisation between ssp. *siceraria* and ssp. *asiatica* gourds occurring. As for the New World, the questions will be answered with aDNA analysis of a large sample of prehistoric Polynesian bottle gourds.

3.5.3 DOMESTICATION AND DISPERSAL

Data from the chloroplast and two of the nuclear markers (BR01_19 and MR06_24) support the Asian subspecies and African/American subspecies each comprising a monophyletic group, consistent with the current taxonomy (Heiser, 1973b). The large number of genetic differences between the subspecies are inconsistent with their recent (i.e., post-domestication) divergence, and the cpDNA data from the outgroup species are inconsistent with the prevailing hypothesis that Asian cultivars are derived from African cultivars (instead the Asian chloroplast haplotype is basal). Decker-Walters *et al.*, (2001) also concluded that, based on PCA of RAPD fingerprinting data, southern African landraces are derived. Although a rapid 'subspecies' divergence may post-date a single domestication event (followed by a change in allele frequencies due to migration, genetic drift and selection events), a perhaps simpler explanation is divergence of the subspecies predating independent domestication of the bottle gourd in Asia and Africa.

3.5.4 ASSUMPTIONS, LIMITATIONS AND CAVEATS

Finally, I have made several assumptions when interpreting the data and inferring dispersal routes for the bottle gourd. These include:

1. That there have not been multiple, possibly contemporaneous, introductions of the bottle gourd to the New World (i.e., from Africa *and* Asia).
2. That, prehistorically, alleles that differentiate Asia and the New World were fixed in each continent.
3. That, for a single region, allele frequencies have remained fairly constant through time.

It may be that these assumptions are not borne out in future work, but deviating from them with the present data makes interpretation of the data too complex and too speculative.

The most significant limitation of the current work is that the only samples obtained from Polynesia are modern, and only from New Zealand. Because these

accessions may not be representative of the prehistoric diversity in Eastern Polynesia, it will be important for future work to include archaeological material from as many Polynesian islands as possible. This is outlined further in Chapter Five.

For each accession, only one individual was sequenced. In the case of Asia and the New World, intra-continental diversity is so low that it is unlikely that new genotypes would have been discovered by sampling additional individuals of each accession. In the New Zealand accessions, however, where hybridisation between the subspecies is clearly occurring, sequencing more individuals per accessions may reveal different combinations of alleles. For future work, a small pilot study should be carried out to establish levels of genetic diversity within accessions, and the effect this has, if any, on the analysis.

In the literature, the introduction of the bottle gourd into Polynesia has been tightly linked with the introduction of the sweet potato from South America into Polynesia, the latter being almost-certainly human-mediated. Whilst the dispersal of the Polynesian bottle gourd should be considered in the *context* of the sweet potato introduction, there are several reasons why the dispersal of the bottle gourd should be assessed independently:

1. The bottle gourd has the ability to disperse naturally; the sweet potato probably does not.
2. For the Polynesian bottle gourd, the morphological (and now genetic) evidence suggests a partial origin in Asia and a partial origin in tropical or temperate region of the New World; the sweet potato could *only* have come from South or Central America.
3. The linguistic evidence suggests the bottle gourd was not known to proto-Oceanic speakers but it does not support an American origin either; the term *kumara* was clearly available in South America.

A more recent linkage of two independent ideas is the introduction of the bottle gourd to the New World (Erickson *et al.*, 2005) and the introduction of another important, early utilitarian domesticate — the dog (Zeder, 2006; Zeder *et al.*, 2006). Ancient DNA analysis of New World dog remains points to an Old World origin of the domestic dog, and suggests at least five founding lineages accompanied late Pleistocene

humans across the Bering land bridge (Leonard *et al.*, 2002; Wayne *et al.*, 2006). Pairing the dog findings with those for the bottle gourd has allowed the formation of a new hypothesis: "Together, these studies [Erickson *et al.*, 2005; Wayne *et al.*, 2006] indicate that Paleoindians entered the New World with the world's two earliest domestic species, dogs and bottle gourds, and that initial domesticates served utilitarian functions, but not as sources of food." (Zeder, 2006). To test the gourd component of this hypothesis it will be necessary to:

1. Establish whether the bottle gourd could have grown in the Beringian climate, or whether it is realistic that seeds were transported from the northern limit of cultivation in Asia, through Beringia, to the northern limit of cultivation in the New World (see Section 3.5.2.2).
2. Explain the apparently African morphology of prehistoric bottle gourd seeds from Peru (see Section 3.5.2.4.3).

3.6 CONCLUSIONS

3.6.1 MOLECULAR MARKERS IN BOTTLE GOURD

This project includes the development of the first reported set of codominant, nuclear markers specifically for bottle gourd — an acknowledged requirement for understanding this species' diversity (Decker-Walters *et al.*, 2001). The cpDNA markers are also a significant resource for studying the bottle gourd. Because the chloroplast is maternally inherited, the cpDNA markers can help distinguish female and male (e.g., pollen contamination) contributions to the genome. For aDNA research, the cpDNA markers are easier to amplify than the low-copy nuclear markers, and because they amplify length polymorphisms, are immune to many of the post-depositional mutations that plague interpretation of ancient material based on SNP analysis (see Erickson *et al.*, 2005).

3.6.2 THE POLYNESIAN BOTTLE GOURD

The Polynesian bottle gourd may have a dual origin, with the chloroplast data from the New Zealand accessions indicating a partly Asian origin and the nuclear data supporting genetic contributions from both Asia and the New World (consistent with the morphological data).

The Asian alleles present in the Māori bottle gourd may be derived from an old *ca* 5,000 yr BP Austronesian (Chinese–Taiwanese) lineage or the more recent 200 BC Southeast Asian (Indian–Indonesian) lineage. In either case, it is unclear how bottle gourds could have moved through the Bottle Gourd Gap into Eastern Polynesia. Natural dispersal may play a role, or the striking correlation between the presence of Lapita pottery and the absence of bottle gourd may offer an alternative explanation.

The ssp. *siceraria* alleles present in the Māori bottle gourds support a partly American origin for the Polynesian bottle gourd. There is very strong evidence for the prehistoric Polynesian-mediated transfer of the sweet potato from Peru–Ecuador into Polynesia, and the bottle gourd may have been introduced from this region also; it was certainly available in Peru. The new evidence for contact between Hawai‘i and California from AD 400–800 also allows for the introduction of the bottle gourd from California. In fact, the distribution of the bottle gourd along the entire Pacific coast from Peru to California allows for an introduction from anywhere in this region.

In addition to human-mediated dispersal, natural dispersal could explain the introduction of the bottle gourd from the New World to Polynesia. Perhaps this hypothesis has not been given enough attention recently in the literature. Certainly, within Polynesia, the linguistic and other anthropological evidence suggests dispersal of the bottle gourd was mostly human-mediated.

The results of Erickson *et al.* (2005) have confounded the situation in Polynesia and added new complexities — their data are apparently incompatible with the conclusions I have made about the origins of the bottle gourd in Polynesia. Under the Erickson *et al.* (2005) model, the ssp. *siceraria* alleles found in the Māori bottle gourds cannot have come from the New World and must be the result of either introgression with historic-period gourd introductions, or a more complex dispersal event which is, as yet, unknown (e.g., an introduction from Africa via Island Southeast Asia). There are, however, questions remaining about the origins of the New World bottle gourd and these need to be solved before more molecular research is carried out on the Polynesian bottle gourd. Temporally, the origins of the Polynesian bottle gourd are the last pieces of the jigsaw in the global dispersal of this species, and rely on many assumptions about bottle gourd dispersal in the rest of the world.

3.6.3 DOMESTICATION AND DISPERSAL

A parsimonious interpretation of the data for this project, as well as data from other recent research (Decker-Walters *et al.*, 2001; Decker-Walters *et al.*, 2004b; Erickson *et al.*, 2005) and important earlier work (Heiser, 1973b, 1979a), allows the extension of models for the domestication and global dispersal of the bottle gourd (Fig. 3.11).

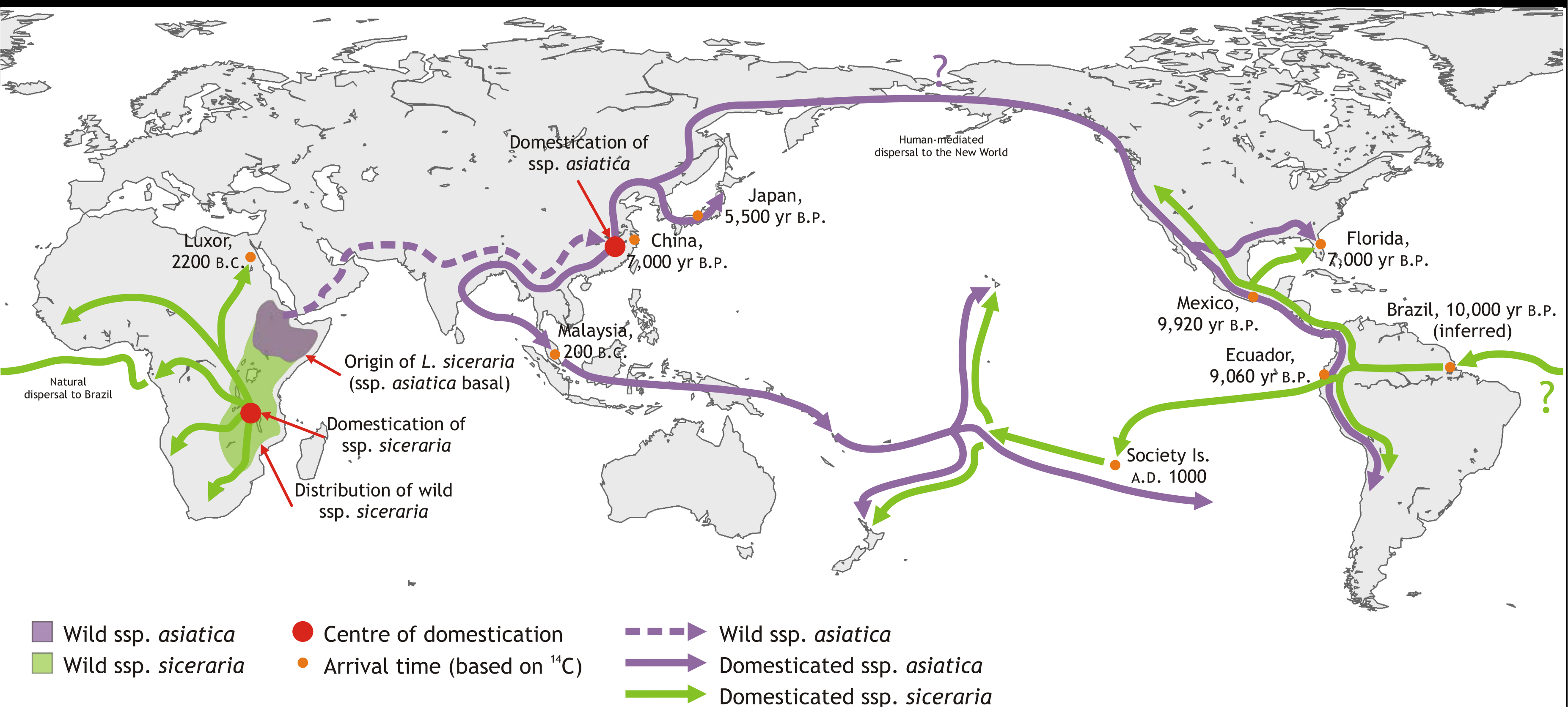
Based on the distribution of other *Lagenaria* species, the great morphological diversity of bottle gourds in Africa, and the discovery of a wild Zimbabwean gourd, tropical Africa is still the most likely centre of origin for the species. The exciting finding that the ssp. *asiatica* chloroplast haplotype is basal, and the ssp. *siceraria* haplotype derived, suggests either:

1. domestication of ssp. *asiatica* followed by divergence of the subspecies, or, perhaps more likely,
2. divergence of the subspecies followed by independent domestication

It is now clear that a number of crop and livestock species have been domesticated more than once (Diamond, 2002), and this may also be the case for bottle gourd.

Based on the wild gourd from Zimbabwe, and augmented by the availability of pollinators in Kenya, the wild ssp. *siceraria* gourd may have been distributed throughout tropical east Africa. African, ssp. *siceraria*, bottle gourds most likely floated (i.e., without human assistance) from the Gulf of Guinea in Africa, with the South Equatorial Current, to the coast of Brazil; there is no evidence for human voyaging across the tropical Atlantic Ocean 10,000 yr BP. If this dispersal event was of a domesticated gourd (the rind of the wild bottle gourd may be too fragile for long ocean voyages (see Decker-Walters *et al.*, 2004b)), then ssp. *siceraria* must have domesticated in Africa by at least 10,000 yr BP because bottle gourd appears in a domesticated state (based on large phytoliths (Piperno *et al.*, 2000a) and thick rinds (Erickson *et al.*, 2005)) at a number of sites in the New World and by this time. If the bottle gourd that floated to the New World was wild (and was domesticated upon arrival), then this allows for the bottle gourd in Africa to be domesticated much later, consistent with its appearance in African archaeological sites from only 2000 BC, as well as the apparent divergence of

Fig. 3.11
Hypothesised Global Dispersal and Domestication of Bottle Gourd



The wild bottle gourd almost certainly originates in tropical Africa. It may have been domesticated in this region also. A separate domestication event may have occurred in East Asia, following dispersal of a wild ssp. asiatica lineage from Africa. The New World bottle gourd was originally believed to have dispersed from Africa by floating, but recent research supports an Asian origin, and human-mediated dispersal through late Pleistocene Beringia is inferred. Both models are shown here. The Polynesian bottle gourd could have dispersed from either side of the Pacific (see Fig. 3.5).

southern African landraces from those found in the New World (Decker-Walters *et al.*, 2001).

The morphological variation in Africa is not consistent with domestication of ssp. *asiatica* on this continent; perhaps with the exception of gourds from the Horn of Africa, African gourds belong to ssp. *siceraria*. Instead, wild ssp. *asiatica* may have dispersed naturally from Africa — either overland or by floating — as far as East Asia. Although no wild bottle gourds have yet been discovered in Asia, the rind fragments from Spirit and Banyan Valley Caves in Thailand (Yen, 1977, Table 1, p. 575) could represent wild *L. siceraria*; the Thai material is certainly not from *domesticated* bottle gourd (Heiser, 1979a, pp. 82–83). There have also been unconfirmed reports of wild bottle gourd in Myanmar (Peter Matthews, pers. comm.). Although more effort should be made to locate wild bottle gourd plants in Mainland Southeast Asia, the recent political situation means that botanical and archaeological information for much of Mainland Southeast Asia — especially Myanmar, Laos and Cambodia — is difficult to obtain (Bellwood, 2005, p. 131).

Domestication of ssp. *asiatica* in East Asia by at least 10,000 yr BP would explain the plant's early presence in China by 7,000 yr BP and Japan by 5,500 yr BP. Perhaps the bottle gourd was domesticated in coastal China with the earliest cereals; whole-grain millet and rice may have been domesticated as early as 14,000–11,000 yr BP (Bellwood, 2005, p. 114). An early domestication of the bottle gourd in East Asia would also allow human-mediated dispersal through Beringia and south to Mexico by 9,920 yr BP (based on the Erickson *et al.* (2005) model of an Asian origin for the New World bottle gourd). An Asian origin for the New World bottle gourd allows for domestication in Africa 2000 BC, consistent with the late appearance of the bottle gourd in the archaeological record of Africa.

CHAPTER FOUR

ORIGINS AND DISPERSAL OF THE SWEET POTATO IN OCEANIA

Māori Man Sorting Kumara

The Far North, New Zealand, ca *1910s*

Photograph taken by Arthur James Northwood

Ref. No. 1/1-6227-G

Reproduced with permission of the Alexander Turnbull Library, Wellington, New Zealand

4.1 CHAPTER OVERVIEW

Various lines of botanical, archaeological and genetic evidence are consistent with a South American origin of the sweet potato (*Ipomoea batatas*) and its prehistoric transfer to Oceania. The evidence suggests transfer effected by Polynesian voyagers between AD 1000 and 1100, who reached the remote western coasts of South America, collected the sweet potato and returned to Polynesia.

This chapter begins with an overview of the hypotheses accounting for the presence of the sweet potato in Oceania, with a focus on the Tripartite Hypothesis, for which there is the best support. The linguistic, archaeological and historical support for the Tripartite Hypothesis is reviewed, as are the proposed dispersal routes for the sweet potato in Polynesia. A recent paper (Zhang *et al.*, 2004) favours natural dispersal of the sweet potato into Oceania, and the implications of this are discussed. The introduction concludes with the hypotheses and aims of the current research.

The materials and methods describe the sweet potato accessions obtained, the application of the AFLP technique and the use of the optimised scoring parameters developed in Chapter Two.

The results and discussion include interpretation of AFLP phylogenetic analysis in the context of linguistic and historical data. The focus first is on the broad-scale dispersal patterns in Oceania, and then on questions surrounding the origin of commercial and Māori varieties in New Zealand. Possible future work is presented in Chapter Five.

4.2 INTRODUCTION

4.2.1 THE SWEET POTATO IN OCEANIA — HYPOTHESES

There has been a long history of debate regarding when, from where, and how, the sweet potato arrived in Oceania. Douglas Yen, in his seminal monograph *The Sweet Potato and Oceania: An Essay in Ethnobotany* (Yen, 1974), was the first to assimilate and critically evaluate, in one volume, various hypotheses (both prehistoric and historic) to account for the presence of the sweet potato in Oceania (Yen, 1974, pp. 1–5). Of the eight hypotheses reviewed by Yen, five invoke prehistoric explanations, including: independent evolution of the sweet potato in multiple regions, natural dispersal from South America by floatation, an African or Asian origin of the sweet potato followed by a human-mediated introduction into Oceania, or a South American origin of the sweet potato followed by a human-mediated (either Polynesian or indigenous South American) introduction into Oceania. A further two hypotheses invoke historic-period explanations: that the sweet potato was introduced into Oceania by late 16th or early 17th century explorers and that further dispersal was effected by Polynesian voyagers, or that early European explorers introduced the sweet potato to India and that Persian, Hindu or other traders introduced it into Indonesia from where it spread into New Guinea and Polynesia (Yen, 1974, pp. 1–5).

The final hypothesis, for which there is now very strong support, is known as the Tripartite Hypothesis, and allows for both prehistoric and historic dispersal events of the sweet potato from South America to explain the distribution of the species in Oceania. The Tripartite Hypothesis was first proposed by Barrau (1957), before being developed extensively by Yen (1974), and modified and updated by Green (2005). The elements of the Tripartite Hypothesis, including the modifications of Green (2005), are:

1. the Polynesian sweet potato (*kumara* lineage(s)) was introduced by Polynesian voyagers, who collected it from the west coast of South America between AD 1000 and 1100 (brought forward from Yen's earlier dates of AD 400–700),

2. the sweet potato of the Philippines and Marianas (*camote* lineage(s)) was introduced by Spanish traders sailing the Manila galleons between Acapulco, Mexico and Manila in the 16th century (Yen, 1974, p. 3), and
3. the Indonesian sweet potato (*batata* lineage(s)) was introduced indirectly, from Brazil via Africa and India, along the Portuguese trade routes of the 16th century.

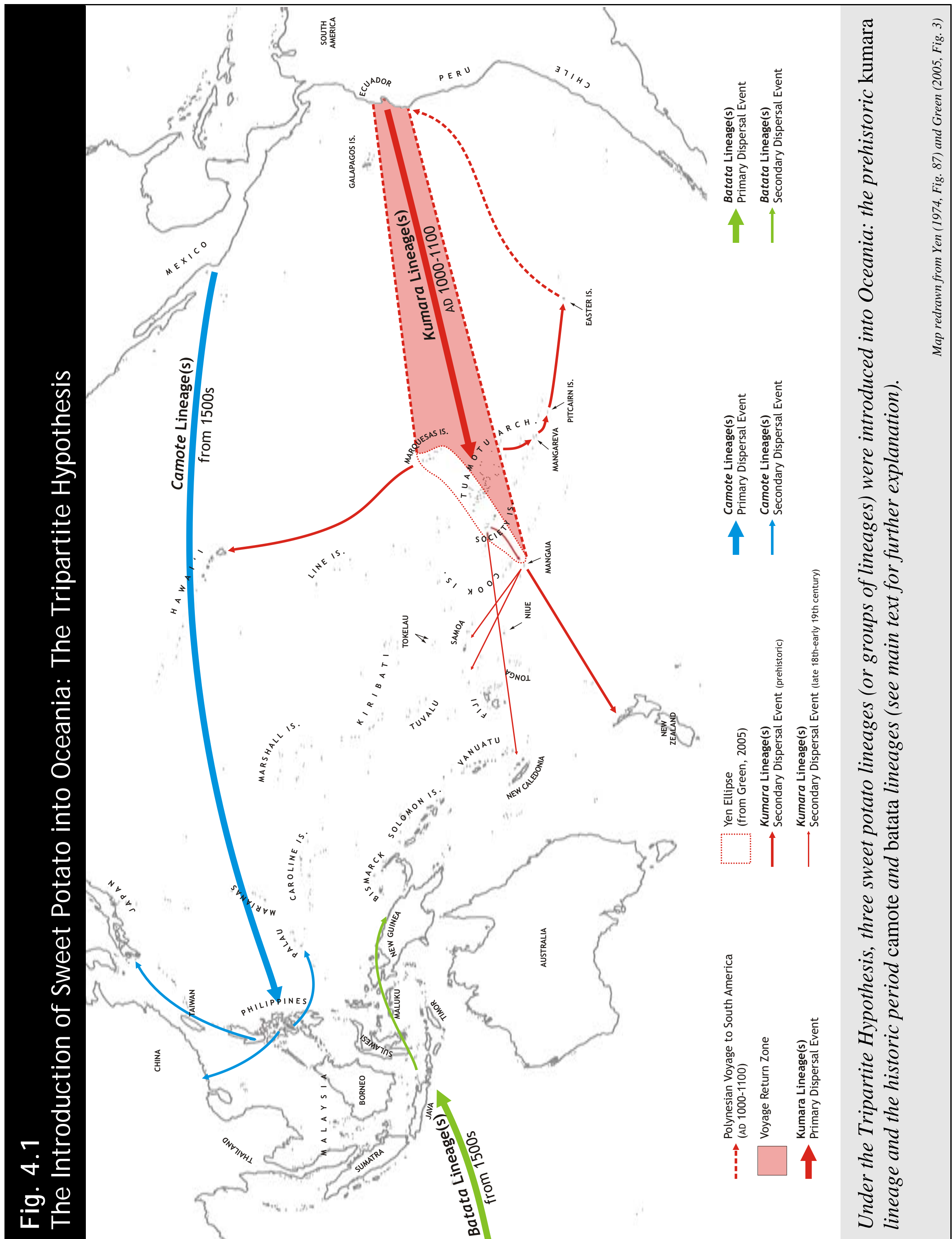
Expanding on 1. above, South America was most likely reached by a Polynesian double-hulled canoe sailing from Easter Island (Rapa Nui). Landfall was made somewhere along the Peru–Ecuador coast, and returning Polynesian voyagers arrived in the islands of the Tuamotu Archipelago or Society Islands (the Yen Ellipse) between AD 1000 and 1100. The sweet potato was then introduced by Polynesians to Hawai‘i and Easter Island, which were already settled. Transfer was also effected to New Zealand, at about the same time as permanent settlement. Transfer westwards was effected later, with the sweet potato arriving in the Western Polynesia and eastern Island Melanesia during the early historic period.

4.2.2 SWEET POTATO EVOLUTION, DOMESTICATION AND DISTRIBUTION IN THE AMERICAS

The centre of origin of the sweet potato is thought to be Central America, which, based on AFLP (Zhang *et al.*, 1998; Zhang *et al.*, 2000a) and allele specific-SSR analyses (Zhang *et al.*, 2000b), is more diverse than Colombia–Venezuela, the Caribbean and Peru–Ecuador. This is consistent with the hypothesis of Austin (1988), who predicted that the centre of origin of sweet potato was between the Yucatán Peninsula, Mexico and the mouth of the Orinoco River, Venezuela, and that the wild progenitor of sweet potato formed as a hybrid between *I. trifida* and *I. triloba*. Restriction fragment length polymorphism (RFLP) and ISSR analyses have confirmed *I. trifida* is a close relative and possible ancestor of *I. batatas* (Jarret *et al.*, 1992; Huang & Sun, 2000).

Peru–Ecuador is a secondary centre of diversity (Zhang *et al.*, 1998; Zhang *et al.*, 2000b; Zhang *et al.*, 2000a) and, compared to the other American regions, is the most diverged from the Central American accessions (Zhang *et al.*, 1998). This is consistent with the long presence (at least 10,000 years) of sweet potato in this area (see

below), and the observations of Yen (1974), who found that the broadest range of morphological variation was found in accessions from western South America.



Under the Tripartite Hypothesis, three sweet potato lineages (or groups of lineages) were introduced into Oceania: the prehistoric kumara lineage and the historic period camote and batata lineages (see main text for further explanation).

The earliest evidence for sweet potato cultivation in western South America is from dried *I. batatas* remains recovered from a cave in Chilca Canyon, Peru (3,900m a.s.l.), dated from 10,000 to 8,000 yr BP (Engel, 1970). These remains are likely to be a product of early agriculture rather than of plant gathering because the sweet potato has not been recorded outside of cultivation (Yen, 1974). Sweet potato remains (including both macro remains and starch grains) have been recorded from numerous other archaeological sites in Central and South America (e.g., Piperno & Pearsall, 1998, pp. 293–294, 297). Sweet potato tubers are also depicted in ceramics, including in vessels attributed to the Mochica, a pre-Incan civilisation that inhabited coastal Peru 2,100–1,400 yr BP (Yen, 1974).

4.2.3 THE SWEET POTATO IN OCEANIA

4.2.3.1 The Linguistic Aspect

The most important evidence in the argument for Polynesian contact with South America, and moreover with the indigenous people of South America, is the apparent lexical borrowing, by Polynesians, of the western South American word for sweet potato. Seeman (1865–1873, p. 170) was the first to note the similarity of the word *cumar* used by the Quechuans of the Ecuadorian highlands to the Polynesian word *kumara*. Yen (1974, Tables A1–F1) compiled lists of many vernacular names for the sweet potato in South America and Oceania, both from historical accounts and his own research. Some of these (*kumara* cognates only) are shown in Table 4.1.

The incidence of the word *cumar* and cognates in South America and the Pacific has been used as evidence for Polynesians making landfall in South America, acquiring the sweet potato and the name *cumar* and returning to Polynesia. One significant problem however, was that *cumar* was only ever recorded in the Ecuadorian and Peruvian highlands and not in coastal regions, leading Brand (1971) to argue that the "Polynesians could not have obtained the sweet potato with such a name on the coasts of South America". A solution to his problem has only been proposed recently, with Scaglion (2005) noting, based on a report by a late 16th century Spanish priest, that the term *comal* and/or *cumal* was used by the Cañari people of Ecuador, whose territory

extended to the Ecuadorian coast at the Gulf of Guayaquil. The incidence of *kumara*-like words across Polynesia and into the eastern regions of Melanesia suggests that voyagers transporting the sweet potato and the name *kumara* (or cognates of this word) between islands are at least partly responsible for the intra-Polynesian distribution of sweet potato (see Green, 2005).

Table 4.1
Distribution of *Kumara* and Cognate Terms in South America and Polynesia[a]

Region	Term	Area
South America	*cumara*	Urubamba, Peru; Cuzco, Peru
	cumar	Ecuador highlands
	cumar’	Peru; Ecuador
	umar’	Peru; Ecuador
	kumar	Lima region, Peru; Aymara, Peru; N. Quechua, Peru
	kumal	Lima region, Peru
	kumara	Peru; Ecuador; Colombia
	ckumara	Peru; Ecuador
	umala	Colombia
	kuala	Colombia
	cjumara	Sierras, Peru
Polynesia	*kumara*	New Zealand; Easter Island; Tuamotu; Mangareva, Rarotonga, Niue, Rotuma
	kuma‘a	Nuku Hiva, Marquesas; Hiva Oa, Marquesas
	ku‘a‘ra	Mangaia
	ku‘ara	Mangaia; Aitutaki
	kuara	Mangaia; Aitutaki; Rarotonga
	umara	Tahiti; Tubuai
	umaa	Tahiti
	uara	Mangaia
	oomara	New Zealand
	kumala	Tonga; Futuna; ‘Uvea; New Zealand
	umala	Samoa
	uala	Hawai‘i
	uwala	Hawai‘i

[a] Compiled from Yen (1974, Tables A2, B1).

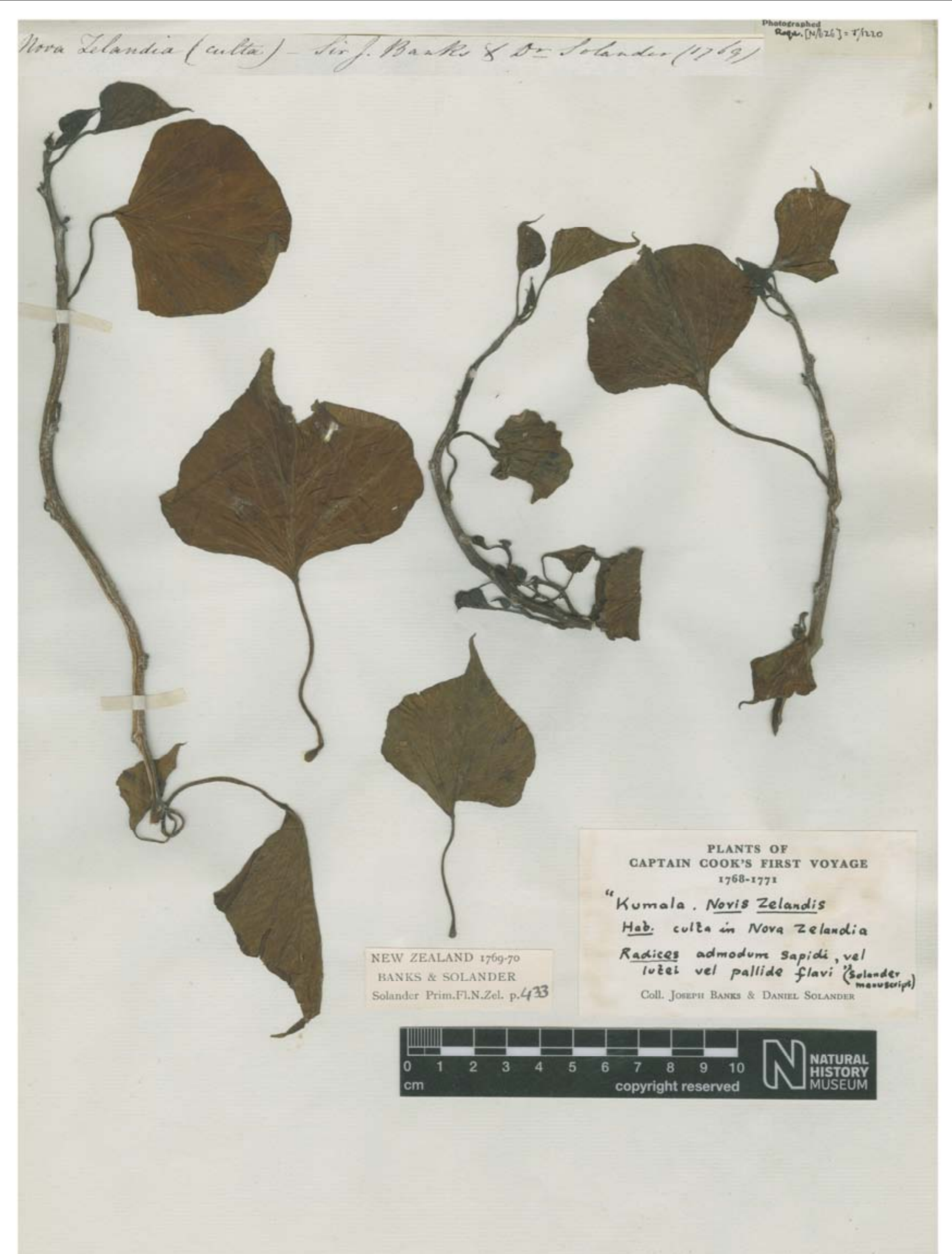
Of the many names used for the sweet potato in Oceania, four are derived from foreign borrowings. Three of these (*kumara*, *camote* and *batata*) are Amerindian terms. The use of *camote* is restricted to the Philippines and Micronesia, whereas *batata* is used in Ambon, Timor and some of the northern Maluku Islands (Yen, 1974, p. 16). The terms *camote* and *batata* are both thought to be Iberian introductions to Oceania (Yen, 1974, p. 16). The fourth foreign term is *putete* (including the cognate *butete*), which is derived from the English *potato*, and is used in central Melanesia (Yen, 1974). Other words used for the sweet potato in Melanesia and are of unknown origin, although a few are obvious derivations from names of other underground crop species (e.g., yam (*Dioscorea* spp.) and taro (*Colocasia esculenta*)) (Yen, 1974).

4.2.3.2 Distribution and Dispersal of the Sweet Potato in Polynesia

A prehistoric presence of the sweet potato in Polynesia was confirmed by Hather & Kirch (1991), who recovered sweet potato remains from Tangatatau, a large, well-stratified rock shelter on Mangaia (Cook Islands). Three fragments of sweet potato tuber were radiocarbon dated at AD 988–1155, AD 1162–1280 and AD 1327–1438. For New Zealand, archaeological studies suggest the sweet potato may have been present in New Zealand by AD 1150–1250 (Davidson, 2000), or about the same time as initial settlement (Wilmshurst *et al.*, 2008).

Early historic accounts have also been useful in determining the prehistoric distribution of the sweet potato in Polynesia. For example, sweet potato was recorded by Roggeveen in Easter Island in 1722 and by James Cook in Tahiti and New Zealand in 1769 and in Hawai‘i in 1778 (Yen, 1974, pp. 8–9). During Cook’s voyage to New Zealand in 1769 a herbarium specimen of sweet potato was collected by Sir Joseph Banks and Dr Daniel Solander (Fig. 4.2).

Fig. 4.2
The Banks & Solander "Kumala" Sweet Potato (1769)



This sweet potato herbarium specimen was collected in New Zealand by Sir Joseph Banks and Dr Daniel Solander during Cook's first voyage, 1768–1771. Steve Lewthwaite (Plant & Food Research, Pukekohe, New Zealand) has not observed this leaf morphology among New Zealand sweet potato accessions, including among the "Māori" accessions (Steve Lewthwaite, pers. comm.).

Ref. No. Photographed Regn. [N/6z6] = J/1220
Reproduced with permission of the Natural History Museum, London, UK
Copyright © Natural History Museum

By integrating archaeological and linguistic evidence, historical records and information about sweet potato morphology, Green (2005) has been able to propose a model (extending Yen's (1974) Tripartite Hypothesis) for the introduction of sweet potato to Polynesia and its dispersal within Polynesia during the prehistoric and early historic periods. Some of the key points of this model include:

1. The transfer of the sweet potato was effected about AD 1000–1100 by Polynesian voyagers who reached South America from the Mangareva–Pitcairn–Easter Island region.
2. The sweet potato was first introduced and established within a modified central ellipse region (see Fig. 4.1), including the Marquesas Islands, the Societies, the Tuamotu Archipelago and Mangaia.
3. Subsequent transfers in about AD 1100–1300 were to Hawai'i (see also Kirch, 1985, pp. 225, 231; Kirch, 1998; Coil & Kirch, 2005) and Easter Island (which were both already settled), and to New Zealand (at about the same time as settlement). In the case of Easter Island, the sweet potato was introduced from the Marquesas Islands (via the eastern Tuamotu atolls, Mangareva and Pitcairn). In the case of Hawai'i, the sweet potato may have been introduced from the southeast Marquesas (and certainly not as a direct introduction from Colombia (*cf.* Rensch, 1991)).
4. A large tuber variety of sweet potato from Hawai'i was transferred to the Society Islands by European traders in the late-18th to mid-19th centuries, and subsequently to Mangaia, Mangareva and New Zealand. This Hawaiian type may have been developed in Hawai'i or may have been introduced from the Marquesas (and be derived from an independent introduction from South America).
5. Transfers to Tonga and a few nearby Fijian islands, Samoa, Futuna, East 'Uvea and the Australs occurred during the historic period (with subsequent transfers from Tonga to the northeastern tip of New Caledonia).
6. In the mid-19th century, sweet potato from the Society Islands was imported into Rarotonga (where there was apparently no prehistoric cultivation of the species).

4.2.4 The Possibility of Natural Transfer?

Despite most recent authors (e.g., Yen, 1991, 1998; Green, 2005) no longer considering natural dispersal of the sweet potato into Oceania a realistic proposition, Zhang *et al.* (2004) favour this mode of transfer over human-mediated dispersal. Zhang *et al.* (2004) generated AFLP profiles for 76 cultivars of sweet potato: 31 from Oceania, 12 from the Philippines and 11 each from Ecuador, Peru and Mexico. Based on multidimensional scaling (MDS) of Euclidean distances, all of the Peru-Ecuador cultivars group together, with 3 of the Oceanic cultivars. The remaining 28 Oceanic cultivars grouped with those from Mexico and the Philippines. From this the authors conclude there has been significant gene flow from Mexico to Oceania, contrasted against a lack of association between the Oceanic and Peru-Ecuador samples. This leads them to suggest that Peru-Ecuador may not be the source of the Oceanic germplasm and that natural dispersal from Mesoamerica, possibly by birds carrying seeds, could account for the relationships of the cultivars, and therefore the origin of the Oceanic sweet potato (Zhang *et al.*, 2004).

The fundamental problem with Zhang *et al.* (2004), and dealt with by Green (2005), lies in the choice of samples. Of the 31 Oceanic samples, 10 are from the Solomon Islands, nine from Papua New Guinea, two from Fiji and one from New Caledonia. These 22 Melanesian samples lie in the probable area of confluence of the Iberian and Polynesian introductions and therefore cannot be assigned, with any confidence, as being prehistoric. The remaining nine Oceanic samples are from Tonga (7) and the Cook Islands (2). No samples are included from Hawai‘i, New Zealand, Mangaia or from eastern Polynesia (e.g., Marquesas Is. or Easter Is.). Therefore, as Green (2005) states, Zhang *et al.* (2004) “are in no position to make claims about how the sweet potato entered that region prehistorically”.

There are two significant implications of natural dispersal of the sweet potato into Oceania. First, cognates of *kumara* must have arisen independently in South American and Polynesia. Second, the sweet potato must have organs that could survive a journey from the Americas (e.g., floating seed pods, seeds or seed pods carried by birds, tubers atop drifting wood or earth rafts). Following on from this second point, are questions about how the sweet potato could have become established in Polynesia. If

the sweet potato floated or rafted to Polynesia, then it would have to have been transferred away from the littoral zone to a location suitable for growth. This would probably require human intervention, because the sweet potato is non-invasive and has poor survival when competing with wild species (Yen, 1974).

4.2.5 HYPOTHESES AND AIMS

The aim is to test separate elements of the Tripartite Hypothesis. The potential to do this using a molecular approach is widely acknowledged (e.g., Hurles *et al.*, 2003a; Green, 2005; Yen, 2005). The specific aims of this research are to:

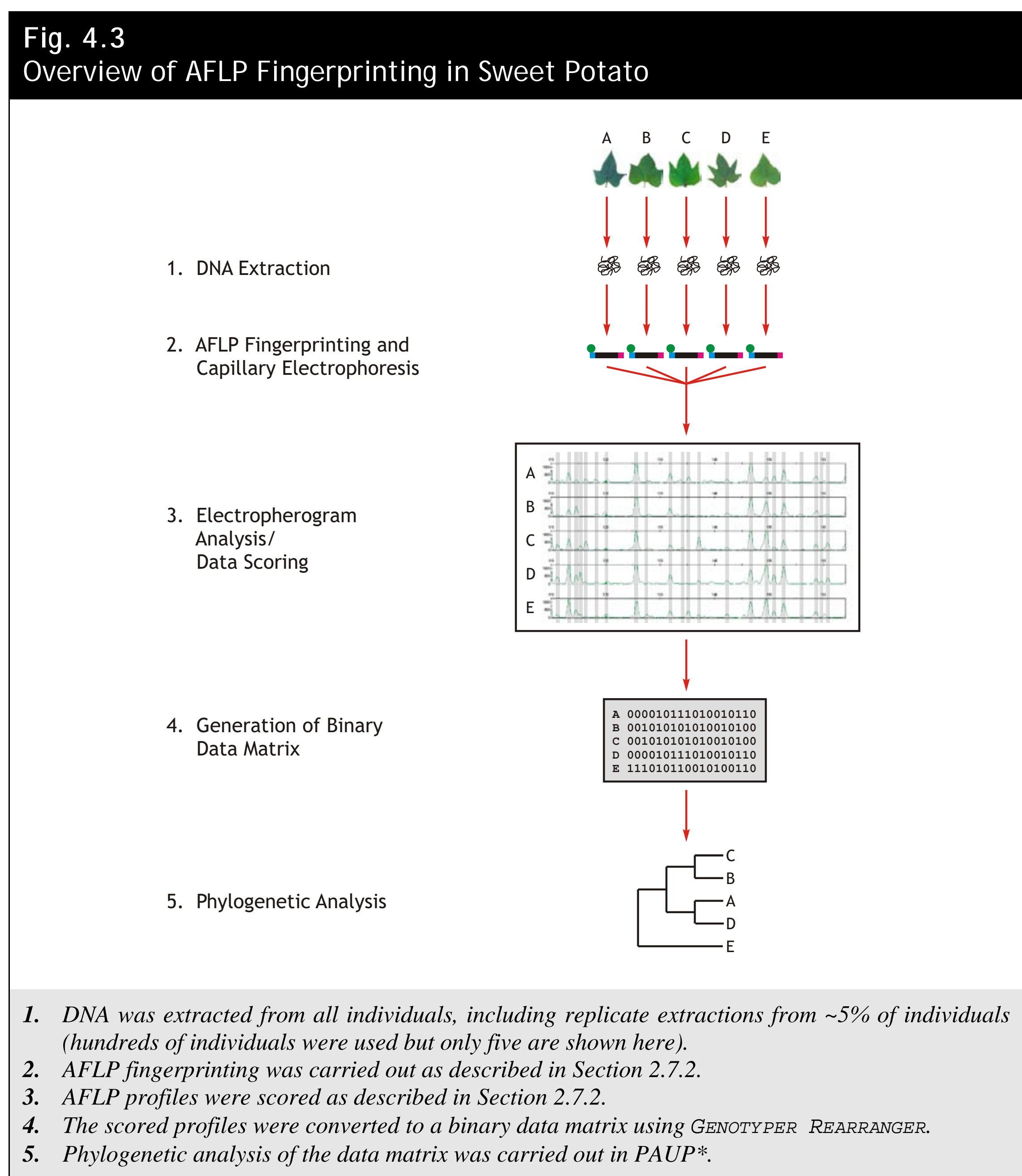
1. Sample varieties of sweet potato from Asia, Melanesia, Polynesia and the Americas that will allow proper testing of the hypotheses.
2. Distinguish closely related varieties of sweet potato using the AFLP fingerprinting technique (including using the optimised scoring parameters to improve phylogenetic resolution).
3. Employ appropriate analysis techniques to analyse the AFLP data.
4. Establish:
    a. The point on the South American coast where Polynesians made landfall,
    b. The number of Polynesian (*kumara*) lineages introduced into Polynesia,
    c. The entry point of the sweet potato into the Pacific (within the Yen ellipse),
    d. Dispersal routes of the *kumara* lineages within Polynesia, including how far west the lineages spread,
    e. The origin of the New Zealand sweet potato accessions, including the Waina group and the Māori varieties ‘Hutihuti’, ‘Rekamaroa’ and ‘Taputini’, and
    f. The dispersal routes of the *camote*, *batata* and other lineages in the Western Pacific, including the origin of the sweet potato in New Guinea.
5. Use patterns of sweet potato dispersal to infer patterns of prehistoric human mobility in Oceania.

4.3 MATERIALS AND METHODS

4.3.1 OVERVIEW

The methodology for generating AFLP data and a general overview of the technique is provided in Section 2.7. The specific process by which AFLP data were generated for this project is described in Fig. 4.3.

Fig. 4.3
Overview of AFLP Fingerprinting in Sweet Potato



1. *DNA was extracted from all individuals, including replicate extractions from ~5% of individuals (hundreds of individuals were used but only five are shown here).*
2. *AFLP fingerprinting was carried out as described in Section 2.7.2.*
3. *AFLP profiles were scored as described in Section 2.7.2.*
4. *The scored profiles were converted to a binary data matrix using GENOTYPER REARRANGER.*
5. *Phylogenetic analysis of the data matrix was carried out in PAUP*.*

4.3.2 PRELIMINARY TECHNICAL REMARKS

The preliminary technical remarks made in Section 2.4.1 apply here also.

4.3.3 COLLECTION OF SWEET POTATO SAMPLES

The majority of samples used in this research were collected in June, 2004 from the Yen Sweet Potato Collection at the National Institute of Crop Science in Tsukuba, Japan. Collecting was carried out by Dr Peter Matthews (National Museum of Ethnology, Osaka) and myself. Leaves were collected on silica gel.

The Yen Sweet Potato Collection contains approximately 300 accessions of sweet potato from East Asia, Island Southeast Asia, Island Melanesia, Polynesia and South America. The Yen Collection was assembled by Doug Yen in the 1950s and was initially housed at the DSIR Crop Research Unit in Otara, New Zealand. A copy of the collection was sent to Japan in 1969, after which the New Zealand copy of the collection died out (in the 1980s). A comprehensive history of the Yen Sweet Potato Collection can be found in Gould (2007). Complete accession lists for the Yen Sweet Potato Collection made during the 2004 sampling trip are provided in Appendices 4 and 15. Photos of the leaves of all accessions in the Yen Collection were also made in 2004, and these are shown in Appendix 15.

In addition to the Yen Collection, a number of sweet potato accessions were obtained from other sources, including New Zealand material from Plant & Food Research (Pukekohe, New Zealand and Lincoln, New Zealand), modern Pacific material from the Secretariat of the Pacific Community (SPC, Fiji) and two samples of a supposedly prehistoric sweet potato *kumara toru* from Mangaia (courtesy of Richard Walter, University of Otago). Complete accession lists for all of these accessions are provided in Appendix 15.

Some of the material obtained from Plant & Food Research ultimately derives from the Yen Collection and was part of the material "repatriated" to New Zealand from Japan in 1988 (see Gould, 2007). The material from the SPC was included to establish the variation in the SPC's germplasm collection (for agronomic purposes), to identify duplicate (redundant) accessions and to examine the extent to which the varieties of sweet potato grown in the Pacific have changed between the 1950s (when the Yen Collection was made) and the present day.

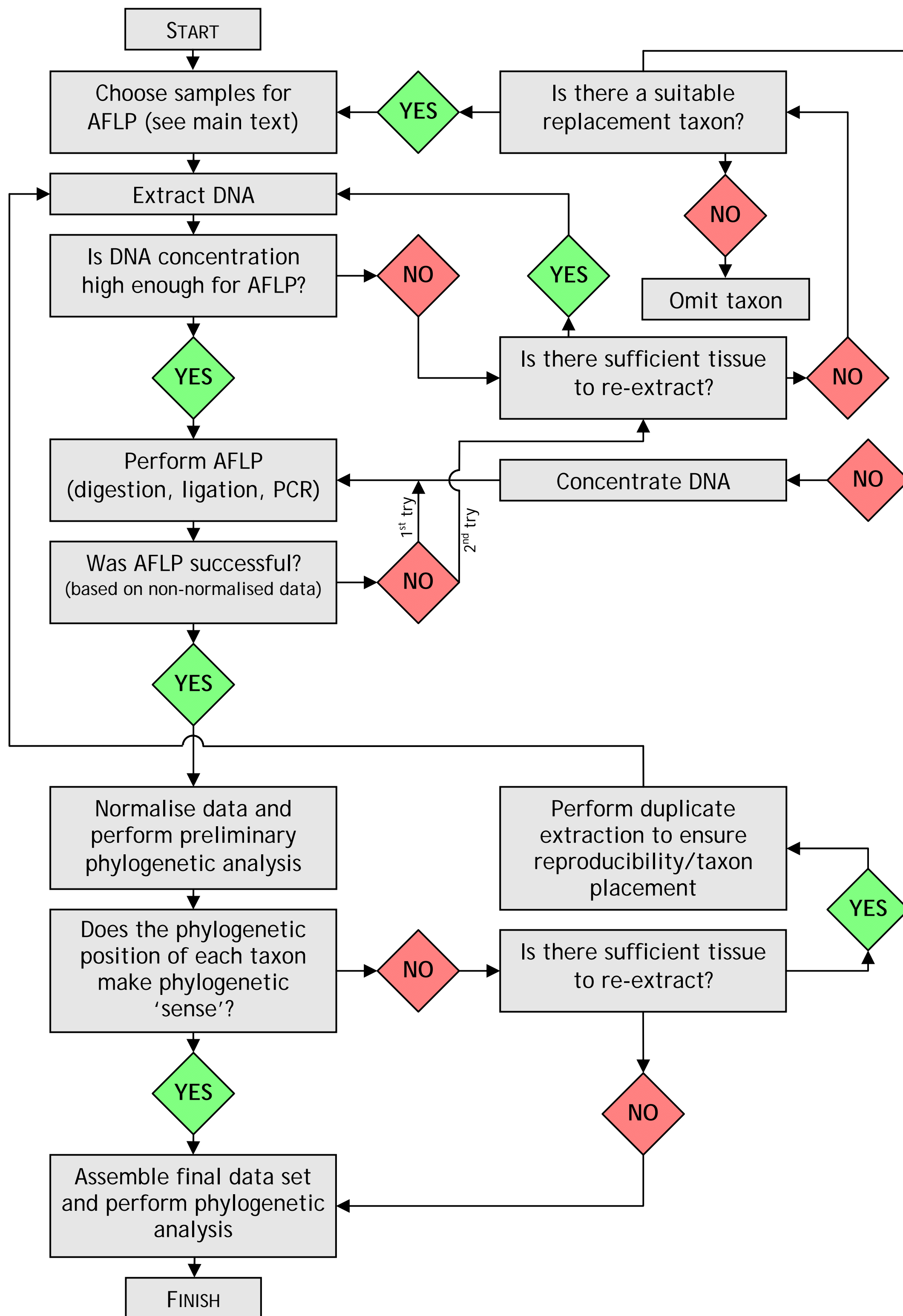
All Yen and non-Yen sweet potato samples collected for this research are stored in silica gel and have been deposited in the Allan Wilson Centre Herbarium (Science Tower D, Massey University).

4.3.4 SAMPLING STRATEGY

Although all 296 accessions present in the Yen Sweet Potato Collection in 2004 were sampled, it was only possible (due to logistical and financial constraints) to generate AFLP profiles for 246 of these. Accessions were excluded if there was already good geographical representation from that region, or the location was not thought to be critical for the hypotheses being tested here. The workflow for generating the AFLP data is shown in Fig. 4.4.

Replicate extractions were performed on 5–10% of accessions. The accessions for which replicates were generated were chosen randomly.

Fig. 4.4
Workflow for Generation of AFLP Data



There were multiple points in the process of generating AFLP profiles at which a sample could fail. Thus, it was necessary to design a workflow to manage samples efficiently.

4.3.5 DNA EXTRACTION

4.3.5.1 Leaf Tissue Disruption

Approximately 25 mg of silica-dried leaf tissue from each accession was loaded into the wells of a 1.2 mL 96 deep-well plate (ABgene). Each 96-well plate contained approximately 6 replicate DNA extractions (where two DNA extractions were performed from a single accession). These replicates were processed through to phylogenetic analysis to ensure the technique produced consistent results and to ensure any large-scale loading or labelling errors could be detected.

After each deep-well plate had been loaded with tissue, two stainless steel beads (∅ 4 mm) were added to each well, and the plate freeze-dried overnight to ensure the tissue was as dry and brittle as possible before grinding. After freeze-drying, each deep-well plate was sealed with Thermo-Seal heat sealing foil (ABgene) using an electric Thermo-Sealer (ABgene). The tissue was reduced to a fine powder by placing each deep-well plate in the grinding apparatus at HortResearch (now Plant & Food Research, Palmerston North, New Zealand), and shaking at approximately 7–8 Hz for two periods of 1 min each. Plates containing the ground tissue were stored at −20°C until required.

4.3.5.2 Qiagen DNeasy® DNA Extraction

DNA was extracted directly from dried leaf tissue using the DNeasy® 96 Plant Kit (Qiagen) with a modified version of the manufacturer's protocol. The following description is for one 96-well plate.

A working lysis solution was prepared by preheating 38.8 mL Buffer AP1 to 65°C and adding 97 μL RNase A (100 mg mL^{-1}) and 97 μL of Reagent DX. Using a multi-channel pipette, 133 μL of working lysis solution was added to each well in the first column of the deep-well plate by piercing the foil seal. The solution was pipetted up and down vigorously to resuspend the ground plant tissue. The suspended plant tissue was transferred to a new rack of collection microtubes. This process was repeated

a further two times for the first column, so that a total of 400 μL of working lysis solution had been used to resuspend each sample. This three-aliquot process was repeated for columns 2–12.

130 μL Buffer AP2 was added to each of the collection microtubes, the microtubes closed with caps, and the clear cover placed over the rack. The rack was shaken vigorously by hand for 15 s and then the solution collected in the bottom of the microtubes by centrifuging the rack briefly at 1,300 × *g* in a Labofuge 4000 centrifuge (Heraeus). The rack was incubated at −20°C for 10 min and then centrifuged at 3,000 × *g* for 10 min.

The supernatant from each well was then transferred to a new rack of collection microtubes, ensuring that no pelleted material was transferred. Typically 300–360 μL of supernatant was recovered. 1.5 volumes of Buffer AP3/E was added to each well in the new rack, the microtubes closed with caps, and the clear cover placed over the rack. The rack was shaken vigorously by hand for 15 s and then the solution collected in the bottom of the microtubes by centrifuging the rack briefly at 1,300 × *g*.

The DNeasy 96 Plate was placed on top of an S-Block, and the samples from the rack of collection microtubes transferred into the plate. The plate was sealed with an AirPore™ Tape Sheet and centrifuged at 1,000 × *g* for 10 min and 2,000 × *g* for 10 min (brake off). The plate was then centrifuged for additional 10 min periods (at 2,000 × *g*, brake off) until all liquid had flowed through the DNeasy membranes.

The flow-through from the S-block was discarded, the AirPore Tape Sheet removed, and 800 μL Buffer AW added to each sample. The plate was sealed with a new AirPore Tape Sheet and centrifuged at 1,000 × *g* for 10 min and 2,000 × *g* for 15 min (brake off). The plate was then spun for additional 10 min periods (at 2,000 × *g*, brake off) until all liquid had flowed through the DNeasy membranes. The plate was then centrifuged at 2,000 × *g* for a final 10 min to properly dry the DNeasy membranes.

The AirPore Tape Sheet was removed, and the DNeasy 96 Plate placed on a rack of elution microtubes RS. The DNA was eluted by adding 100 μL Buffer AE (10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0) to each sample and incubating the plate at room

temperature for 5 min. The plate was sealed with a new AirPore Tape Sheet and centrifuged at 3,000 × *g* for 5 min. The AirPore Tape Sheet was removed and a second 100 μL aliquot of Buffer AE was added to each sample, and the plate incubated and centrifuged as above. The DNeasy 96 Plate was discarded and the DNA stored in the elution microtubes while being used. Remaining DNA has been transferred to 1.7 mL tubes and archived at −80°C at the Allan Wilson Centre, Palmerston North.

Additional DNA extractions were performed individually using the DNeasy Plant Mini Kit (Qiagen) as described in Section 3.3.4.3.

4.3.5.3 Quantification of DNA Yield and Assessment of DNA Quality

4.3.5.3.1 Agarose Gel Electrophoresis

To quantify the yield of DNA from the extraction process aliquots of each DNA solution were electrophoresed on an agarose gel. A 5 μL aliquot of each DNA extraction was combined with 1 μL 10× loading buffer in a total volume of 10 μL, and loaded onto a 1.0% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A High DNA Mass™ Ladder was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4.

Successful extractions appeared on the agarose gel as a single, intact band of high molecular weight DNA (> 12 kb). DNA concentration was estimated by comparison with the High DNA Mass Ladder standard. Only those DNA samples which contained high molecular weight DNA, and had a DNA concentration of over 10 ng μL^{-1} were used for AFLP.

4.3.5.4 Concentration of DNA for AFLP

It was necessary to concentrate some DNA samples that yielded an inadequate concentration of DNA for AFLP (< 10 ng μL^{-1}). As a control experiment,

unconcentrated and concentrated DNA from a high-concentration extraction was subjected to AFLP. Identical AFLP profiles were obtained, demonstrating the concentration protocol itself has no effect on the quality of the AFLP profiles.

The volume of unconcentrated DNA was measured using a pipette. One-tenth of this volume of 3 M sodium acetate (NaOAc (pH 5.2)) and 2.5 volumes of 100% ethanol were added, and the solution mixed by pipetting. The tube was then incubated at −20°C for 1 h before being centrifuged at 20,000 × *g* at −10°C for 30 min. After centrifugation, the supernatant was immediately removed using a pipette, taking care not to disturb the (sometimes invisible) pellet. A 150 µL aliquot of 70% (v/v) ethanol was then added. The tube was centrifuged at 20,000 × *g* at −10°C for 10 min. After centrifugation the supernatant was immediately removed using a pipette, again taking care not to disturb the pellet. The pellet was dried at 70°C for 10 min before being dissolved in 35 µL Buffer AE (Qiagen). A 2 µL aliquot of each concentrated DNA sample was combined with 1 µL 10× loading buffer in a total volume of 10 µL, and loaded onto a 1.0% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A High DNA Mass Ladder was loaded as a mass standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4. Concentrated samples with > 10 ng μL^{-1} were used for AFLP.

4.3.6 AFLP FINGERPRINTING

4.3.6.1 Restriction Endonuclease Digestion of DNA

Preliminary experiments showed that the AFLP technique is very robust to varying amounts of DNA in the digestion reaction (from 100 ng–1 µg). For this reason it was possible to add 18 µL (50–250 ng) DNA to every reaction — the ability to add a fixed volume of DNA greatly increases the ease and speed at which high-throughput digestion reactions can be performed.

Genomic DNA was digested with the restriction endonucleases *Eco*R I and *Mse* I. Each digest reactions contained 1× restriction buffer (50 mM potassium acetate,

10 mM magnesium acetate and 10 mM Tris-HCl (pH 7.5)), 10 U *Eco*R I (Roche), 10 U *Mse* I (New England Biolabs) and 18 μL (50–250 ng) genomic DNA in a total volume of 25 μL. The reactions were incubated at 37°C for 2 h (to digest the DNA), followed by 70°C for 15 min (to irreversibly denature the enzymes).

To determine whether complete digestion had occurred, a 5 μL aliquot of each digestion reaction was combined with 1 μL 10× loading buffer in a total volume of 10 μL, and loaded onto a 1.0% (w/v) agarose/1× TAE gel (as described in Section 2.4.2.1.4). A 1 Kb Plus DNA Ladder™ was loaded as a size standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4.Complete digestion appeared on the gel as a smear of low molecular weight DNA up to ~1.0 kb. The next step usually proceeded immediately, but if not, the digestion reactions were stored at −20°C until required.

4.3.6.2 Ligation of Oligonucleotide Linkers to the Restriction Fragments

Double-stranded oligonucleotide linkers were ligated to the restriction fragments produced from the digestion reaction. The linkers provide targets for the pre-selective amplification and selective amplification primers.

4.3.6.2.1 Preparation of Oligonucleotide Linkers

The double-stranded oligonucleotide linkers (also called adapters) were prepared by annealing two single-stranded oligonucleotides. The *Eco*R I (Eco) and *Mse* I (Mse) linkers were prepared separately. The Eco linker reaction mixture contained 45% (v/v) $T_{10}E_{0.1}$ buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)), 1 nmol Eco Linker I and 1 nmol Eco Linker II in a total volume of 200 μL. The Mse linker reaction mixture contained 45% (v/v) $T_{10}E_{0.1}$ buffer, 10 nmol Mse Linker I and 10 nmol Mse Linker II in a total volume of 200 μL. Using a thermal cycler, both reaction mixtures were heated to 95°C for 5 min and then cooled to 5°C at a rate of 0.05°C s^{-1} (Δtemp = 90°C = 30 min).

Slow cooling of the linkers favours correct annealing. The linkers were stored at −20°C until required. See Appendix 5 for linker sequences.

4.3.6.2.2 Ligation of Linkers to the Restriction Fragments

Each linker ligation reaction contained 1× ligation buffer (Roche; 66 mM Tris-HCl, 5 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 1 mM adenosine 5′-triphosphate (ATP), pH 7.5), 5 pmol (1 µL) Eco Linker, 50 pmol (1 µL) Mse Linker, 1 U T4 DNA Ligase (Roche) and 5 µL of digested genomic DNA in a total volume of 20 µL. The reactions were incubated at 37°C for 3 h. The next step usually proceeded immediately, but if not, the ligation reactions were stored at −80°C until required (they were found to degrade at −20°C).

4.3.6.3 Pre-Selective Amplification of Ligation Products

Pre-selective amplifications were performed using primers with one selective base — an A in the case of the Eco pre-selective primer and a C in the case of the Mse pre-selective primer (as described in Section 2.7).

Each PCR reaction consisted of 1× PCR buffer (Roche; 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 250 µM of each dNTP, 1 M betaine (see Section 2.4.2.1.3 for a description of the benefits of betaine), 0.5 µM Eco+A pre-selective primer, 0.5 µM Mse+C pre-selective primer, 1 U *Taq* DNA polymerase (Roche) and 1 µL of ligation product in a total volume of 20 µL. PCR was carried out in a thermal cycler as follows: 20 cycles of 94°C for 30 s, 56°C for 1 min, 72°C for 1 min; hold at 10°C. Ramping speed was set at 1°C s^{-1}. The PCR products were stored at 4°C until required. See Appendix 5 for primer sequences.

Success of the pre-selective amplification PCR reactions was determined by electrophoresis of an aliquot of a subset (~5) of the total PCR reactions in each batch. A 5 µL aliquot of each reaction was combined with 1 µL 10× loading buffer in a total volume of 10 µL, and loaded onto a 1.0% (w/v) agarose/1× TAE buffer (as described in

Section 2.4.2.1.4) gel. A 1 Kb Plus DNA Ladder was loaded as a size standard. Samples were electrophoresed at 5 V cm^{-1} in 1× TAE buffer. Gels were stained, visualised and photographed as described in Section 2.4.2.1.4. Successful amplification appeared on the gel as a faint smear of 50–500 bp.

4.3.6.4 Selective Amplification of Pre-Selective PCR Products

4.3.6.4.1 Primer Screening of Selective PCR Primers

A brief search of the literature was undertaken to determine the specific Eco+3 and Mse+3 primers most often used for AFLPs in plants, and particularly for sweet potato. Based on this research, the four Eco+3 primers were chosen, and each labelled with a different fluorophore (because of the fluorescent labels, it was prohibitively expensive to screen different Eco+3 primers).

Eight commonly-used Mse+3 primers were identified from the literature search, and these were used with each of the four Eco+3 primers (i.e., eight primer combinations per fluorophore). Each primer combination was tested on six accessions to assess levels of polymorphism and test that the primers amplified consistently. Selective amplification, poolplexing and capillary electrophoresis were carried out as described below (Sections 4.3.6.4.2–4.3.6.5.2). The results of the primer screening are shown in Table 4.2. For each primer combination, the profiles were qualitatively assessed for simplicity (peaks should be well-defined and well-separated), signal:noise ratio, polymorphism (at least some polymorphic peaks should be visible), and length of profile (there should be a range of peaks from 100–500 bp). Primer combinations were chosen that possessed all of these qualities (Table 4.2). The large variation in quality between different primer combinations affirmed the importance of primer screening before embarking on a large-scale AFLP project.

Table 4.2
Results of PCR Primer Combination Screening for AFLP

Eco+3	Mse+3	Simplicity[a]	Signal:Noise[b]	Length of Profile[c]	Polymorphism[d]	Overall[e]
BLUE 6FAM-Eco+ACT	Mse+CAA	✓✓	✓	< 325 bp	✓	–
	Mse+CAT	✓✓	✓✓	< 300 bp (low 200-300 bp)	✓	–
	Mse+CAC	✓✓✓	✓✓✓	< 320 bp (low 250-320 bp)	✓✓	2nd
	Mse+CAG	✓✓	✓	< 340 bp	✓✓	–
	Mse+CTT	✓	✓✓	< 350 bp (low 240-350 bp)	✓✓	–
	Mse+CTG	✓✓	✓✓	< 360 bp	✓✓✓	–
	Mse+CCC	✓✓✓	✓✓✓	< 360 bp (low 280-360 bp)	✓✓✓	1st
	Mse+CGG	✓✓	✓	< 380 bp (low 220-380 bp)	✓✓✓	–
GREEN VIC-Eco+AGC	Mse+CAA	✓✓	✓✓✓	< 325 bp	✓✓	–
	Mse+CAT	✓✓	✓✓✓	< 325 bp	✓✓✓	2nd=
	Mse+CAC	✓✓✓	✓✓✓	< 325 bp	✓✓	–
	Mse+CAG	✓✓	✓✓✓	< 325 bp	✓✓	1st
	Mse+CTT	✓✓	✓✓	< 350 bp	✓✓	–
	Mse+CTG	✓✓	✓✓✓	< 325 bp	✓✓✓	–
	Mse+CCC	✓	✓✓✓	< 325 bp	✓	2nd=
	Mse+CGG	✓✓	✓✓	< 340 bp	✓✓	–
YELLOW NED-Eco+ATA	Mse+CAA	✓✓	✓	< 260 bp	✓✓	–
	Mse+CAT	✓✓✓	✓✓	< 320 bp	✓✓	–
	Mse+CAC	✓✓✓	✓✓	< 330 bp (low 220-320 bp)	✓✓	1st
	Mse+CAG	✓✓	✓✓	< 220 bp	✓✓✓	2nd
	Mse+CTT	✓✓	✓✓	< 245 bp	✓	–
	Mse+CTG	✓✓	✓✓	< 380 bp (low 250-380 bp)	✓✓✓	–
	Mse+CCC	✓	✓	< 320 bp (low 180-320 bp)	✓✓	–
	Mse+CGG	✓✓✓	✓✓	< 250 bp (low 170-250 bp)	✓✓	–
RED PET-Eco+AAG	Mse+CAA	✓✓	✓✓	< 410 bp (low 230-410 bp)	✓	–
	Mse+CAT	✓✓	✓✓	< 320 bp (low 230-320 bp)	✓✓	–
	Mse+CAC	✓✓✓	✓✓	< 410 bp (low 240-410 bp)	✓✓	2nd
	Mse+CAG	✓✓✓	✓✓	< 300 bp	✓✓✓	1st
	Mse+CTT	✓✓	✓	< 190 bp	✓✓	–
	Mse+CTG	✓✓	✓✓✓	< 270 bp	✓✓✓	–
	Mse+CCC	✓✓	✓✓	< 230 bp	✓✓	–
	Mse+CGG	✓✓	✓	< 220 bp	✓✓	–

[a] Rated from ✓ (low simplicity i.e., peaks poorly separated, many shoulder peaks) to ✓✓✓ (high simplicity i.e., well-separated peaks with few shoulder peaks).

[b] Rated from ✓ (noisy data i.e., low signal:noise ratio) to ✓✓✓ (clean data i.e., high signal:noise ratio).

[c] The range in which peaks are found. "Low" corresponds to the range in which only very low intensity peaks are found.

[d] Rated from ✓ (few polymorphisms between the six taxa tested) to ✓✓✓ (many polymorphisms between the six taxa tested).

[e] The best two Mse+3 primers for each Eco+3 primer are denoted "1st" and "2nd". Primers rated 1st were used (see Table 4.3). Primers denoted 2nd and "–" were not used, although the 2nd primers are recommended for future work.

4.3.6.4.2 Selective Amplification of Pre-Selective PCR Products

Selective amplification reactions were performed using Eco and Mse selective primers with three selective bases each (Table 4.3). The Eco+3 primers were fluorescently-labelled using the fluorophores described in Table 4.4.

Table 4.3
Eco+3 and Mse+3 Selective Amplification Primer Combinations

Colour	Primer Combination	
	Eco+3 Primer	**Mse+3 Primer**
BLUE	6FAM-Eco+ACT	Mse+CCC
GREEN	VIC-Eco+AGC	Mse+CAG
YELLOW	NED-Eco+ATA	Mse+CAC
RED	PET-Eco+AAG	Mse+CAG

Table 4.4
Absorbance and Emission Maxima for AFLP Fluorophores

Colour	Fluorophore	Absorbance maximum (nm)[a]	Emission maximum (nm)[a]	Supplier
BLUE	**6FAM**™ (6-carboxy-fluorescein)	494	522	Sigma–Aldrich
GREEN	**VIC**®[b,c]	533	554	Applied Biosystems
YELLOW	**NED**™[b]	553	575	Applied Biosystems
RED	**PET**™[b]	558	595	Applied Biosystems
ORANGE	**LIZ**™[b,d]	638	655	Applied Biosystems

[a] Absorbance and emission wavelengths were obtained from Table 1–2, Schoske (2003).
[b] Applied Biosystems proprietary fluorophores.
[c] HEX is sometimes used as a less expensive alternative to VIC, however we found that its emission spectrum significantly overlaps with that of NED, leading to bleed-through (spectral overlap) of HEX signal into the NED channel.
[d] Size standard (GeneScan™-500).

Each selective amplification PCR consisted of 1× PCR buffer (containing 1.5 mM $MgCl_2$), 3.125 mM additional $MgCl_2$ (for a total $MgCl_2$ concentration of 4.625 mM), 250 µM of each dNTP, 0.5 µM •XXX-Eco+ANN selective primer (where •XXX denotes the fluorophore — see Table 4.3), 0.5 µM Mse+CNN selective primer,

1 U *Taq* DNA Polymerase and 1 μL of undiluted pre-selective amplification PCR product in a total volume of 20 μL. PCR was carried out in a thermal cycler using the following touchdown program: 94°C for 2min; 10 cycles of 94°C for 30 s, 65°C for 30 s (thereafter decreasing 1°C/cycle), 72°C for 1 min; 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min; 72°C for 30 min; hold at 10°C. Ramping speed was set at 1°C s^{-1}. See Appendix 5 for primer sequences.

The prolonged incubation at 72°C for 30 min ensures the non-templated adenylation of the 3′ ends of the synthesised strands (production of an adenosine (A) overhang) is driven to completion (Clark, 1988; Smith *et al.*, 1995; Brownstein *et al.*, 1996). The one-base variation in PCR product length caused by a mixture of adenylated and non-adenylated products could otherwise lead to 'stutter' or 'split peaks' in the AFLP profiles.

Usually the PCR products were poolplexed and subjected to capillary electrophoresis immediately, but if this was not possible they were wrapped in tin foil and stored at −80°C until required.

4.3.6.5 Capillary Electrophoresis of AFLPs

4.3.6.5.1 Poolplexing of Selective Amplification PCR Products

For each accession, 6FAM-, VIC-, NED- and PET-labelled selective amplification PCR products were pooled together, and an aliquot of the pooled mixture electrophoresed on the 3730 Genetic Analyzer. The optimal pooling ratio for 6FAM:VIC:NED:PET was found to be 1:1:1:2 (PET fluoresces less brightly than the other fluorophores so is required in higher concentration).

4.3.6.5.2 Capillary Electrophoresis of Selective Amplification PCR Products

Poolplexed AFLP reactions were subjected to capillary electrophoresis (CE) at the Allan Wilson Centre Genome Service (AWCGS), Massey University, Palmerston North. A 1 μL aliquot of each poolplexed sample was mixed with 0.3 μL GeneScan™-500 LIZ™ size standard (Applied Biosystems) and 8.7 μL Hi-Di™ formamide (Applied Biosystems). Samples were then irreversibly denatured by incubating at 95°C for 5 min before being rapidly cooled on wet ice. Samples were briefly centrifuged to remove any micro-bubbles.

Samples were subjected to CE on a 3730 Genetic Analyzer (Applied Biosystems) using a 50 cm array. Samples were analysed using the G5 Reduced Cross Talk (G5-RCT) dye set and GeneMapper50_POP7_1 run module (both Applied Biosystems). The profiles were provided as electronic files in the default Applied Biosystems format (*.fsa).

4.3.7 AFLP SCORING

In total, 313 AFLP profiles were successfully generated for 261 unique accessions of sweet potato (the replicate taxa included 44 accessions with two profiles each, and four accessions with three profiles each).

All 313 AFLP profiles for the preliminary dataset were imported into GeneMapper® v. 3.7 (Applied Biosystems) and scored using the optimised parameter settings from Chapter Two, i.e., minimum fragment length (MFL) = 50 bp, peak height threshold (PHT) = 50 rfu and bin width (BW) = 0.5 bp. Smoothing was set to "light smoothing", and the allele-calling threshold was set to the same value as the peak height threshold. All other parameters, preliminary testing of which showed negligible effects on scoring, were left at their default values. A full description of the scoring parameters used is provided in Appendix 6. Peaks above the detection and labelling thresholds (50 rfu) were labelled as "1" and peaks below the threshold or absent as "0". The binary data matrix was exported from GeneMapper and converted into a NEXUS format file

(Maddison *et al.*, 1997) using the program GENOTYPER REARRANGER (GTR) by Warwick Allen (which is available, along with a detailed protocol, from http://awcmee.massey.ac.nz/aflp/aflp.html). The character matrix contained 2271 characters. The NEXUS file is in Appendix 17.

4.3.8 PHYLOGENETIC RECONSTRUCTION

4.3.8.1 Preliminary Phylogenetic Analyses

Preliminary phylogenetic analyses were carried out in PAUP* version 4.0b10 (Swofford, 2003) using both neighbour-joining (NJ) on uncorrected distances and with heuristic search using the parsimony optimality criterion (retaining all default settings in PAUP*). Both Hamming (shared presences (1's) and absences (0's)) and Jaccard (shared presences only) distance measures were used. See Section 2.2.5 for an explanation of why tree-building methods are appropriate.

The number of known replicates that were correctly placed as sister taxa were counted for each of the three trees. Using this measure, the NJ analysis of Jaccard distances was found to perform best. The better performance of NJ over parsimony is consistent with analyses on the 30-taxon data sets (see Chapter Two; Holland *et al.*, 2008). The finding that Jaccard distances preserved more replicate relationships than Hamming distances is also consistent with previous AFLP studies (e.g., Wong *et al.*, 2001; Simmons *et al.*, 2007). These studies show that shared presence distance measures are more appropriate for AFLP data than those that also incorporate shared absences because shared absences (null-alleles) are particularly susceptible to homoplasy (because of the multiple, independent, ways in which a fragment can be lost) (Bussell *et al.*, 2005; Kosman & Leonard, 2005; Simmons *et al.*, 2007).

4.3.8.2 Choosing Taxa to Omit for Final Analysis

Before constructing the final tree, taxa for which there was evidence that they might bias the analysis were removed.

The first group of taxa that were removed were all of the outgroup species. These are genetically distant from the ingroup species, and most may be too genetically distant to give a reliable phylogenetic position. Correct placement of outgroup taxa is a known problem in phylogenetic analysis, and can disrupt ingroup relationships that would otherwise have been found (Holland *et al.*, 2003). The excluded outgroup species were: *I. indica* (#137), *I. minuta* (syn. *I. plummerae*; #011), *I. nil* (#126), *I. pes-caprae* (#129), *I. tiliacea* (K233) and *I. trifida* (K270, K500, T10).

The second group of taxa that were removed included 24 unique sweet potato accessions obtained from the Secretariat of the Pacific Community (SPC), Fiji. There were several aims associated with including these accessions (all with accession numbers beginning "IB") in the larger dataset (see Section 4.3.3). Unfortunately, the quality and quantity of leaf tissue provided for these accessions was very low, as was the quality of the resulting AFLP profiles (they were characterised by very low or very high proportions of "1" character states; see below). Also, replicate SPC accessions did not group together as predicted. Compared to the Yen accessions, the SPC accessions had little information about provenance (they sometimes had country of origin information, but might still have been a recent commercial introduction into that area), which makes it difficult to have confidence in their placement on the tree. For these reasons, the SPC accessions were excluded from the final analysis. Future work (see Chapter Five) will focus on obtaining better quality leaf tissue for the SPC accessions, and re-genotyping these samples.

The third group of accessions that were removed were those with an unusually high or low proportion of "1" character states. Preliminary work had shown that the number of "1" character states as a proportion of total characters can vary widely. Although some of this variation may reflect true differences in the number of AFLP fragments generated for each accession, other variation may be the result of poorly amplified accessions (low proportion of 1's) or accessions in which a lot of background

noise has incorrectly been scored as true peaks. The negative side effect of this variation in the proportion "1" character states might be that taxa incorrectly group together because they share many 1's or many 0's (rather than grouping based on true genetic relatedness). The proportion of "1" character states was calculated for all taxa, and then taxa were ranked in order of increasing proportion.

4.3.8.3 Final Phylogenetic Analysis

A NJ tree of the Jaccard distances of the remaining 270 taxa was created in PAUP* and exported in Newick tree format. This tree was imported into the program Dendroscope (Huson *et al.*, 2007), which is specially designed for viewing large phylogenies, and then edited in CorelDRAW 12 (Corel Corporation).

4.3.8.4 Path Lengths Between Replicates

Preliminary work had shown that replicates do not always group together as sister taxa. This may occur for a number of reasons. Firstly, there may be other taxa that are so closely related to the replicate pair that they cannot be distinguished from the replicates using the AFLP technique (either because there are no true polymorphisms that separate them, or the number of polymorphisms is significantly less than the random error of the technique). The neighbour-joining technique will always result in a fully resolved tree, regardless of whether this is an accurate reflection of the resolving power of the data; it may be the case that some very closely related taxa cannot be meaningfully distinguished from each other. Replicates may also not appear as sister taxa if they vary significantly in their proportion of "1" character states (see above). In this case, replicates may pair with taxa that share a similar proportion of "1" character states, rather than with the other half of the replicate pair. Another reason replicates may not group as sister taxa is human error leading to a labelling mistake at one of the many stages between DNA extraction and production of the final tree (this can be corrected by generating another AFLP fingerprint for the taxon). To quantify the frequency at which replicates were grouped as sister taxa, the path lengths (number of edges) that separate replicates was measured. Path length was defined as follows: in the tree

((1A,1B),(2A,(3,2B))), replicates 1A and 1B appear as sister taxa, and have a path length of 2. Replicates 2A and 2B are separated by taxon 3, and therefore have a path length of 3.

4.4 RESULTS AND DISCUSSION

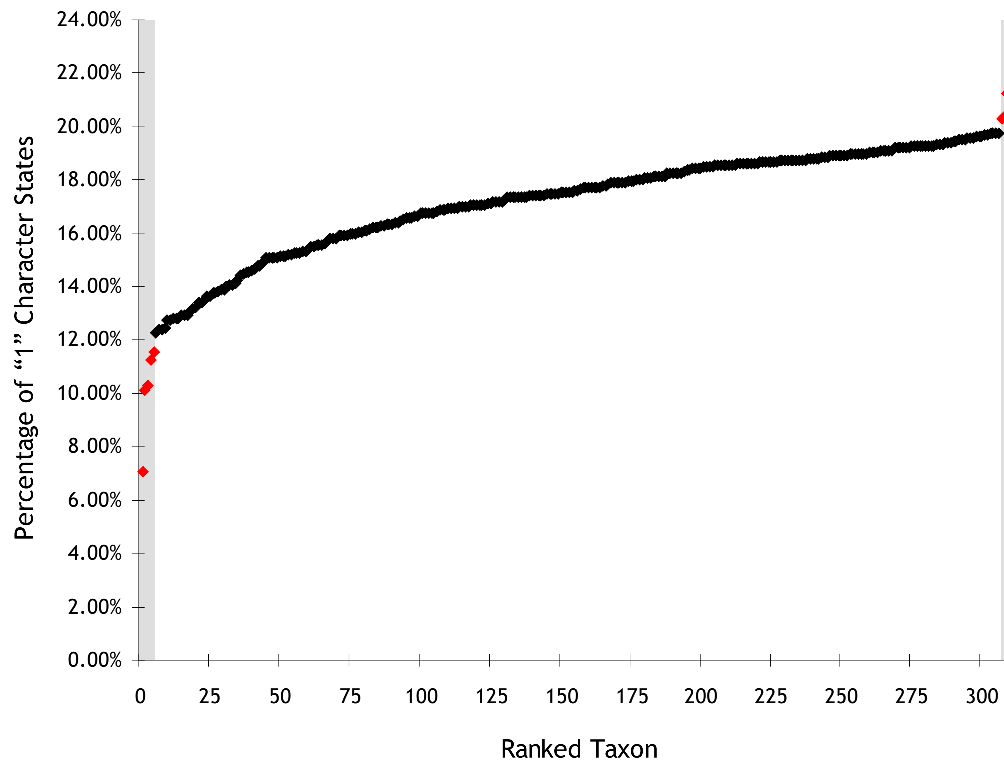
4.4.1 PROPORTION OF "1" CHARACTER STATES

The number of "1" character states as a proportion of the total number of characters per taxon varied from 7.09 to 22.46%. To test the effect of this variation, and to ensure that taxa were not incorrectly grouping together only because they shared many 1's or many 0's, the proportion of 1's was compared with tree topology. Reassuringly, there was found to be a very weak (if any) correlation between proportion of "1" characters and tree topology.

Replicates that varied significantly in their proportion of 1's were also useful for examining the effects of proportion of 1's on topology. Again reassuringly, replicate accessions often still grouped as sister taxa even though they varied significantly in proportion of 1's. For example the replicate samples for accession Y-630 had proportion of 1's of 12.42% and 17.35%, yet still grouped together as sister taxa.

All 313 taxa were ranked in order of increasing proportion of 1's (Fig. 4.5). This shows a smooth curve, but with two distinct tails representing taxa with very few or lots of "1" character states (five taxa were below 12%; seven taxa were above 20%). These taxa were well outside the normal range of proportions, and re-examination of the corresponding AFLP profiles showed problems with the PCR amplification. These 12 taxa were excluded from the final analysis.

Fig. 4.5
Taxa Ranked by Percentage of "1" Character States



All 313 taxa in the preliminary dataset were ranked in increasing order of their proportion of "1" character states. A smooth distribution was observed but with two distinct tails. The taxa representing these tails (red; shaded area) were excluded from further analysis.

In total, 43 taxa were excluded from the final analysis: 13 outgroup taxa (one outgroup taxon also had proportion 1's < 12%), 23 SPC taxa (4 SPC taxa also had proportion 1's < 12% or > 20%), and the remaining 7 taxa with proportion 1's < 12% or > 20%.

4.4.2 NEIGHBOUR-JOINING TREE OF COMPLETE DATASET

The sweet potato neighbour-joining tree is show in Fig. 4.6.

Fig. 4.6
Sweet Potato Neighbour-Joining Tree

The colours of the taxon labels indicate broad-scale geographic origins (the extent to which these correlate with cultural groupings and dispersal events is dealt with in the main text). Taxa are organised into phylogenetic units (Groups 1–33) for ease of reference. Three large groups of particular significance are highlighted. These are the prehistoric Ia and Ib kumara *lineages (blue), the large Melanesian cluster (purple) and the Waina group (red).*

Taxon labels comprise: the accession number, the geographic origin of the accession (from least to most specific), the local or cultivar name (where available), and any other relevant information.

Accession numbers preceded by a "Y" are from the Yen Collection (Appendix 4); accessions not preceded by a "Y" are from other sources (in the main text these accession numbers are preceded by a "#" to distinguish them from the Yen accessions).

AFLP replicates are denoted by "(A)", "(B)" or "(C)" at the end of the taxon label. Note that four replicates (in all cases the "(B)" half of the replicate pair) were amongst the taxa omitted from the final analysis (see Section 4.3.8.2) and therefore do not appear in this figure. In each case the other half of the replicate pair is still denoted with an "(A)" even though the "(B)" sample is absent. These four taxa are #148 (Group 11), Y-570-2 (Group 23), #160 (Group 27) and Y-309 (Group 28).

Abbreviations: Jap. = Japan; Linc. = Lincoln; N. Cal. = New Caledonia; NZ = New Zealand; Puk. = Pukekohe; TTG = Toka Toka Gold

Melanesia

Waina

Kumara Ib

Kumara Ia

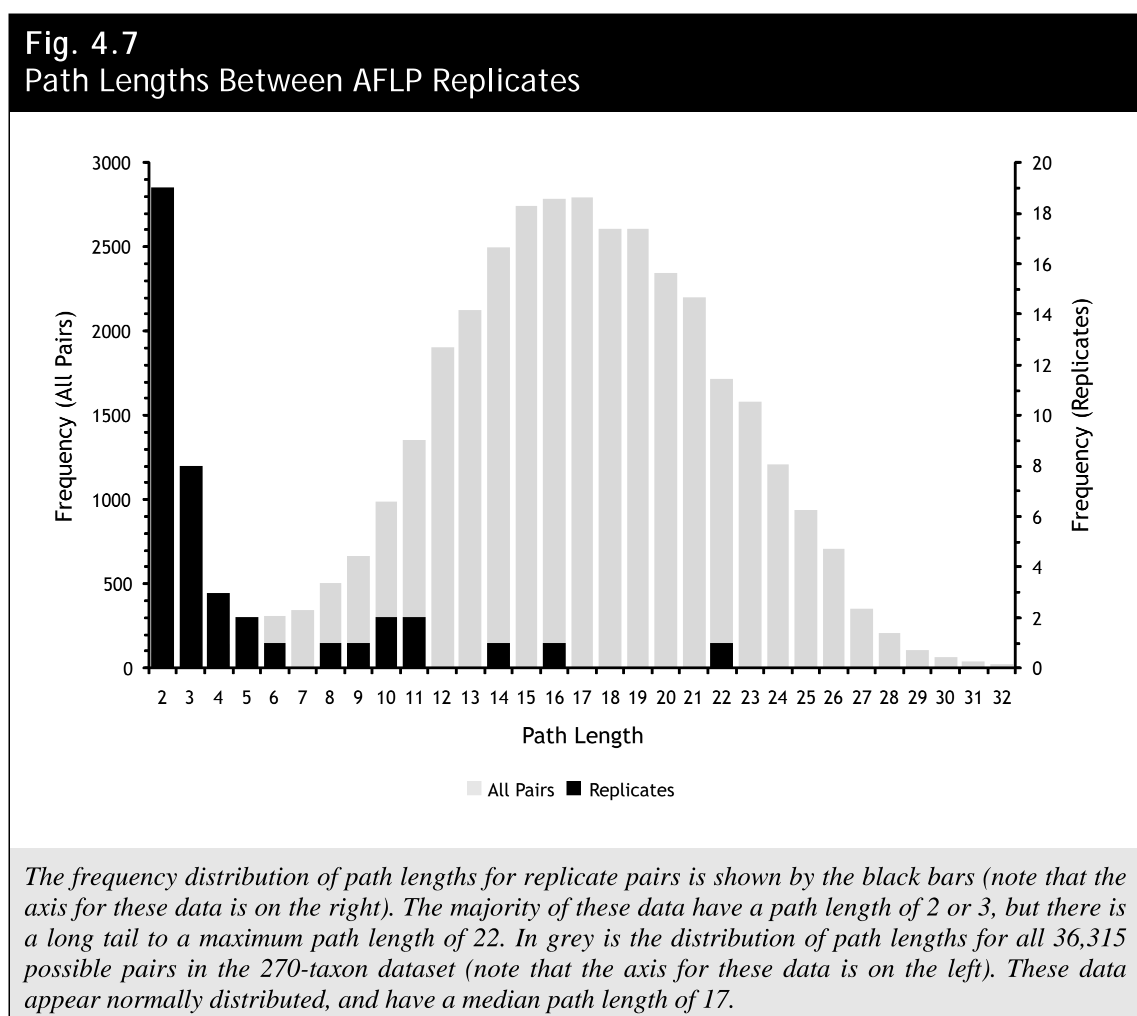
0.1

- Eastern Polynesia
- New Zealand "Māori" varieties (Hutihuti, Rekamaroa, Taputini)
- Western Polynesia
- Island Melanesia (excl. New Guinea)
- New Guinea
- Philippines, Thailand, Japan
- South America
- Waina Group
- Other (incl. commercial, North America)

4.4.3 PATH LENGTHS BETWEEN REPLICATES

The distribution of path lengths between replicates is shown in Fig. 4.7. Of the 42 pairs of replicates, 19 group together as sister taxa (path length of 2). A further 8 pairs have a path length of 3. There are four taxa (Y-106, Y-240, Y-385 and Y-609) that have three replicates each, rather than the usual two. These taxa account for three of path length = 3, and two of path length = 4. Overall, the grouping of replicates as sister taxa, or as near-sister taxa (path length of 3 or 4), is reassuring. The median path length for all replicates is 3.

Fig. 4.7
Path Lengths Between AFLP Replicates



The frequency distribution of path lengths for replicate pairs is shown by the black bars (note that the axis for these data is on the right). The majority of these data have a path length of 2 or 3, but there is a long tail to a maximum path length of 22. In grey is the distribution of path lengths for all 36,315 possible pairs in the 270-taxon dataset (note that the axis for these data is on the left). These data appear normally distributed, and have a median path length of 17.

Fig. 4.7 also shows the distribution of paths lengths for all 36,315 possible pairs in the tree. In this case, the median path length is 17. Taken together, these data suggest the replicates are performing relatively well, and certainly much better than random. There

are, however, some replicates that do not group closely at all (path lengths of 8, 9, 10, 11, 14, 16, 22), and this is concerning. Some of these larger path lengths may reflect groups of taxa that are very closely related and cannot be confidently separated, but it is likely that most reflect mis-placed taxa. Future work should include generating new AFLP profiles for these taxa so that their position on the tree can be accurately determined.

4.4.4 ORIGINS OF THE OCEANIC SWEET POTATO LINEAGES

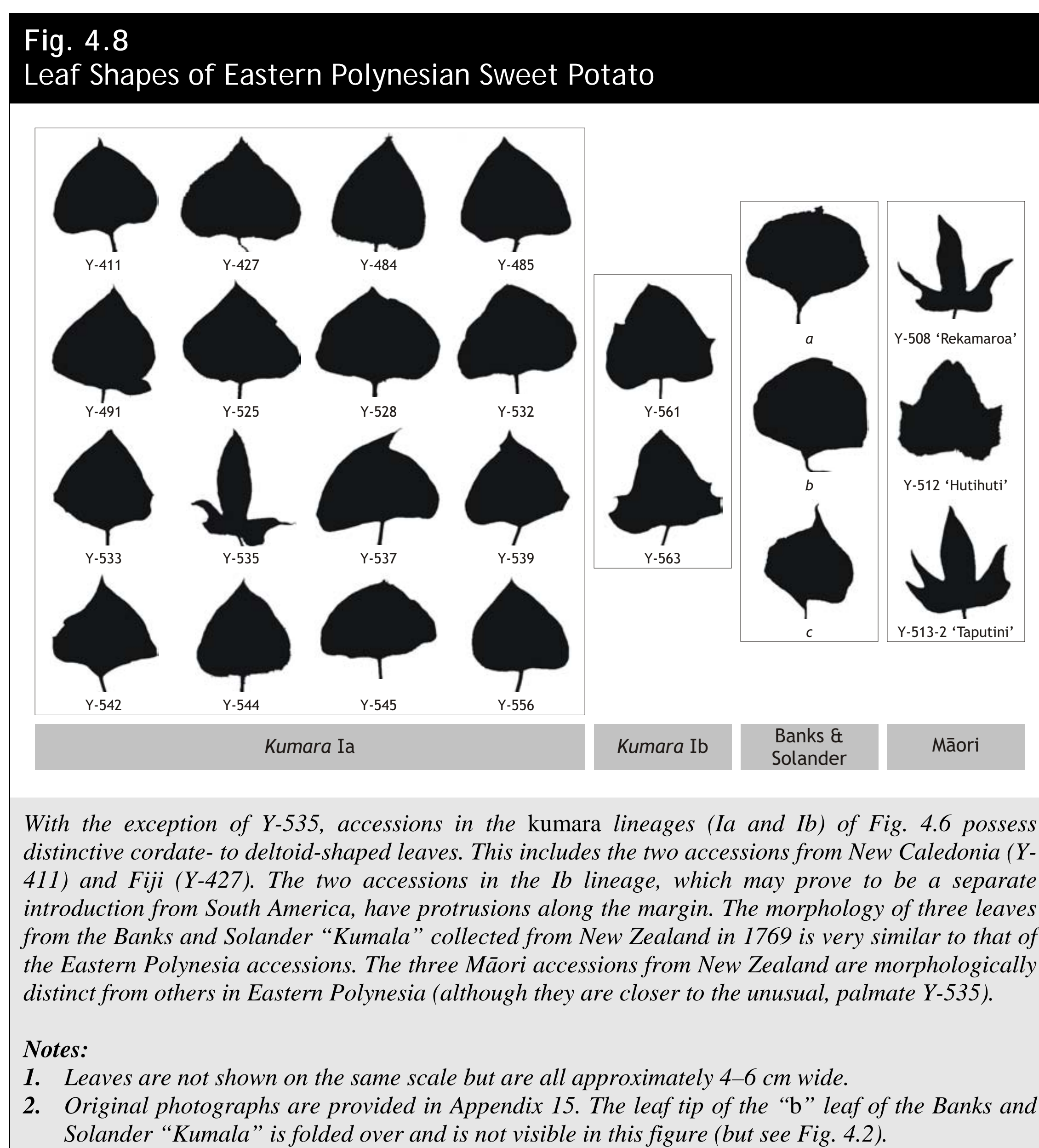
It is clear from Fig. 4.6 that there is a high degree of phylogeographic structure amongst the sweet potato accessions sampled. These patterns are broadly consistent with the main *kumara*, *camote* and *batata* introductions of sweet potato predicted under the Tripartite Hypothesis, followed by more recent transfers of specific lineages to diverse locations. There are also more complex patterns, with sweet potato accessions collected from a single region having multiple origins.

The majority of accessions from Eastern Polynesia (Groups 14–15 in Fig. 4.6) form a group which is tentatively identified as representing the Polynesian-mediated prehistoric introduction of sweet potato from South America to Polynesia. Included in this group are accessions from the Marquesas Islands, the Societies and the Cook Islands (Rarotonga and Mangaia). The identification of this group as the *kumara* lineage of the Tripartite Hypothesis is supported by linguistic and historical analyses (Roger Green, pers. comm.).

In addition to the accessions from the Yen Collection, this *kumara* group includes two accessions collected from Mangaia in 2003 by Richard Walter (University of Otago). These samples (both with the varietal name *kumara toru marama*) were claimed by the local people to be “original” plants (Richard Walter, pers. comm.). They also possess the small, multiple, yellow flesh tubers that are thought to be typical of prehistoric sweet potato in Polynesia (Roger Green, pers. comm.). Accessions in the putative *kumara* lineage possess distinctive cordate- to deltoid-shaped leaves, as does the Banks and Solander “Kumala” collected from New Zealand in 1769 (Fig. 4.2; see

Fig. 4.8 for a comparison of the leaf shapes). This offers further support for the Eastern Polynesian accessions in Groups 14–15 belonging to the prehistoric *kumara* lineage.

Fig. 4.8
Leaf Shapes of Eastern Polynesian Sweet Potato



With the exception of Y-535, accessions in the kumara *lineages (Ia and Ib) of Fig. 4.6 possess distinctive cordate- to deltoid-shaped leaves. This includes the two accessions from New Caledonia (Y-411) and Fiji (Y-427). The two accessions in the Ib lineage, which may prove to be a separate introduction from South America, have protrusions along the margin. The morphology of three leaves from the Banks and Solander "Kumala" collected from New Zealand in 1769 is very similar to that of the Eastern Polynesia accessions. The three Māori accessions from New Zealand are morphologically distinct from others in Eastern Polynesia (although they are closer to the unusual, palmate Y-535).*

Notes:
1. *Leaves are not shown on the same scale but are all approximately 4–6 cm wide.*
2. *Original photographs are provided in Appendix 15. The leaf tip of the "*b*" leaf of the Banks and Solander "Kumala" is folded over and is not visible in this figure (but see Fig. 4.2).*

A further two accessions in the *kumara* group are from Island Melanesia (Fiji and New Caledonia). Historical evidence is consistent with both of these accessions being part of the westward intrusion of the *kumara* lineage into the western Pacific during the historic period. First, is the accession Y-427 from Fiji, which has the varietal name "Samoa", and is recorded by Yen as being an introduction from Western Samoa (now just called Samoa; Appendix 4). The second Island Melanesian accession is Y-411 from the

Balade/Balaide region of the northwestern tip of Grande Terre, New Caledonia (Appendix 4). Yen records this as being grown by migrants originally from the Loyalty Islands (northeast of Grande Terre; Appendix 4), which is consistent with the movement of sweet potato from Samoa and West and East ‘Uvea into the Loyalties in the late 18th and early 19th centuries (Green, 2005; see also Fig. 4.1).

The accessions labelled as representing the *kumara* introductions in Fig. 4.6 form two separate lineages (labelled “Ia” and “Ib”). Although these accessions may represent a single introduction from South America (with subsequent diversification), it is also possible that they are the result of separate introductions (either two varieties collected during a single voyage, or from separate voyages). This hypothesis receives support from examination of leaf morphology: accessions in the “Ia” lineage typically have cordate to deltoid leaves with an entire leaf margin, whereas those in the “Ib” lineage have protrusions along the margin which give the leaves an appearance between deltoid and palmate (Fig. 4.8).

If there were multiple, prehistoric, Polynesian-mediated introductions of the sweet potato into Oceania, then we would not necessarily expect all accessions derived from these introductions to form a monophyletic group when accessions from other regions are also included. For example, if diverse cultivars of different origins were grown at the same location in South America at which Polynesians made landfall, and these cultivars were subsequently introduced into Polynesia then they may be located in quite distant positions in the tree. Following from this, it is possible that other Eastern Polynesian accessions located elsewhere in the tree (Groups 6, 9, 16, 19, 23 and 32) may also be derived from prehistoric, Polynesian introductions. This could be tested with further sampling and historical research. But in the meantime the labelled *kumara* group is a distinct lineage in Fig. 4.6.

Based on the current analysis, it is not possible to confidently distinguish the *camote* and *batata* lineages that are the result of the Iberian-mediated sweet potato introductions to the Western Pacific. Instead there appears to be very weak geographical structuring in the general region of Island Southeast Asia (Groups 1–13), consistent with widespread exchange of diverse sweet potato lineages in this region. Identification of the *camote* and *batata* lineages may be possible with more intensive sampling from

the Americas. Under the Tripartite Hypothesis, the *camote* sweet potato are the result of Spanish introductions to the Philippines and would be expected to be nested within the diversity of sweet potato accessions from Mexico (indeed, the single accession from Mexico (Y-710) is sister to a Philippines accession (Y-106)). Similarly, the *batata* sweet potato are the result of Portuguese introductions to Indonesia and would be expected to be nested within the diversity of sweet potato accessions from Brazil and the Caribbean. Identification of the *batata* lineage would be further aided by more samples from southern Indonesia. Currently, the only samples from Indonesia (excluding Western New Guinea) are from East Timor (Dili and Liquiçá districts) but these always group closely with accessions from the Philippines (Groups 1, 5, 7 and 12), further supporting multiple cases of long-distance exchange in this region. Identification of the *camote* and *batata* lineages is potentially made more challenging by multiple introductions of diverse cultivars from Mexico, Brazil and the Caribbean.

One interesting and unexpected feature of Fig. 4.6 is the large group of 57 unique accessions from Papua New Guinea, Island Melanesia (Vanuatu, New Caledonia, Fiji) and Western Polynesia (Tonga, Samoa). This large group (named "Melanesia" in Fig. 4.6) contains no accessions from other regions, with the exception of a single Japanese commercial variety (#163 'Beniazuma'), Y-482 ("tauranga"/"torumarama" from Aitutaki) and Y-513-2 ('Taputini' from New Zealand). Although the Melanesia cluster contains many taxa and is geographically widespread (from New Guinea to Western Polynesia), it appears not to have spread west into Island Southeast Asia or east into Eastern Polynesia (with the possible exception of the accession from Aitutaki). The origin of the sweet potatoes in this group remains unclear, but this might also be resolved with further sampling from the Americas (there are currently no American accessions in this group). The most parsimonious explanation is that the Melanesian cluster is the result of a European-era introduction(s) to the general area between Papua New Guinea (PNG) and Fiji. The area of Western New Guinea (corresponding to the Indonesian provinces of Papua and West Papua) can probably be excluded as the origin of the Melanesian cluster; all New Guinea accessions in the cluster are from Papua New Guinea (the PNG Highlands, the coastal city of Lae and New Britain). More complex explanations for the Melanesia cluster include an origin farther west in Island Southeast Asia, followed by dispersal into New Guinea and replacement of original varieties in the west, although these sorts of explanations seem

less likely based on the current data. In any case, it will be important to determine the origins of this large group.

From Fig. 4.6, it is clear that the sweet potatoes in New Guinea have multiple, diverse origins. In addition to many New Guinea accessions belonging to the Melanesia cluster just discussed, there are also accessions from New Guinea amongst East Asian accessions in Groups 1–13, as well as the unique Group 16. Thus, there does appear to be some geographic structuring *within* the region of New Guinea. As discussed above, New Guinea accessions in the Melanesia cluster are only from the PNG Highlands (18 accessions), Lae (2) and New Britain (2). In contrast, New Guinea accessions in Groups 1–13 are typically from Western New Guinea, New Britain and the PNG coastal cities of Port Moresby and Lae (only four new Guinea accessions in Groups 1–13 are from the PNG Highlands).

As expected, accessions from South America are found in multiple positions in the tree, reflecting the high genetic diversity of this region. It is also clear from including the South American accessions that the Oceanic lineages have multiple, diverse origins. Perhaps most notable is the lack of South American accessions within the large Melanesia cluster. Conversely, Groups 17–24 include mainly South American accessions, suggesting this genetic diversity is not well-represented in Oceania. As already alluded to above, more intensive sampling from the Americas will be critical for resolving the origins of the Oceanic lineages (see also Chapter Five).

4.4.5 ORIGINS OF THE NEW ZEALAND SWEET POTATO LINEAGES

The majority of commercial sweet potato cultivars in New Zealand belong to the Waina group, including the most commercially important cultivar in New Zealand 'Owairaka Red' (Lewthwaite, 2004). Varieties in this group are all thought to be derived from a single sweet potato variety introduced by whalers who visited New Zealand aboard the *Rainbow* in the 1860s (Gould, 2007). This original variety was given the name "Waina" by Māori (from the English "vine") (Gould, 2007). The majority of the Waina accessions cluster together in Fig. 4.6 (Groups 26–27), consistent with historical

records. Also included in this group are accessions from Vanuatu (Y-386) and Peru (Y-607), the Māori accession "Kotepo" and two accessions of the Māori "Taputini II"). There is no historical evidence that the Vanuatu and Peruvian accessions should be within the variation observed in the Waina group, and instead the placement of these taxa is more consistent with the current data having insufficient resolution to resolve relationships within this group. The suggestion that these accessions should be considered an unresolved conglomerate (there are no problems with doing that for the current work) is further supported by replicates not grouping as sister taxa for two accessions (Y-503 and #153).

Based on the current analysis, neither 'Tauranga Red' (Group 19) nor 'Owairaka Pink' (Group 13) appear to be closely related to accessions in the Waina group (Groups 26–27). In the case of 'Tauranga Red' this is consistent with recent historical research which suggests 'Tauranga Red' was derived from 'Red Bermuda' imported into the Tauranga Horticulture Station between 1910 and 1913 (Gould, 2007). 'Owairaka Pink' also appears to be unrelated to the Waina group, and further historical research may reveal the route into New Zealand (in the tree it groups with accessions from Ecuador, China and southern Japan). The remaining New Zealand accession of commercial importance is the yellow-fleshed 'Toka Toka Gold'. The position of 'Toka Toka Gold' in Fig. 4.6 is not stable, with one replicate located in Group 18 and another in Group 20 (along with the purple-fleshed 'Mary Anne', a known mutant of 'Toka Toka Gold'). More replicates are required to confirm the position of 'Toka Toka Gold', but Group 20 is more likely since 'Mary Anne' is located there also.

The origin of the New Zealand sweet potato varieties 'Hutihuti', 'Rekamaroa' and 'Taputini' is of long-standing interest because all three varieties have been classified as probable prehistoric Māori varieties (Yen, 1963, 1974; Harvey *et al.*, 1997; Green, 2005). Interestingly, none of these accessions group with the putative *kumara* lineage identified in Fig. 4.6, nor with any other Eastern Polynesian accessions. Two accessions of 'Hutihuti' are included in Fig. 4.6, but they do not group together as sister taxa as expected (one is in Group 11, the other in Group 19). The Y-512 accession of 'Hutihuti' (Group 19) is from material collected from the Yen Collection in 2004 by Peter Matthews and myself. The #145 'Hutihuti' is from material provided by Pam Fletcher (Plant & Food Research, Lincoln) in 2006, and is derived from living material

that was obtained from the Yen Collection in 1988 as part of the "repatriation" process (Gould, 2007). The incongruence of these two accessions of 'Hutihuti' can probably be explained by examining the positions of the 'Rekamaroa' samples. Both the sample of 'Rekamaroa' collected from the Yen Collection in 2004 and the sample collected in 2006 from the repatriated material group together in Group 11 (although a replicate of the 2004 sample is located in Group 12, indicating there is some instability in the placement of this variety). A parsimonious interpretation of the placement of 'Hutihuti' and 'Rekamaroa' in the tree is that the identities of the samples collected from Japan in 2004 and the 'Rekamaroa' sample collected from Lincoln in 2006 are correct, but that the plant maintained as 'Hutihuti' at Plant & Food Research, Lincoln is actually mislabelled 'Rekamaroa'. The implication is that 'Hutihuti' has been lost from tissue culture in New Zealand, and perhaps from New Zealand generally. If this duplication of one variety and loss of another has occurred, then it must have been between 1988 and 2006. It is possible that 'Hutihuti' material derived from tissue culture may have been released to New Zealand growers before the mix-up occurred, but this cannot be determined without broader sampling of material in New Zealand grown under the name of 'Hutihuti'. This proposed mix-up is consistent with morphological observations made by Steve Lewthwaite of field-grown plants at Plant & Food Research, Pukekohe, where 'Rekamaroa' and 'Hutihuti' are indistinguishable, and the morphology of 'Hutihuti' does not match Yen's (1963) description (Steve Lewthwaite, pers. comm.). The third possible prehistoric Māori variety is 'Taputini'. The 2004 sample of this variety (from the Yen Collection) is located in Group 30, but the 2006 sample (from Lincoln, New Zealand) is located in the Waina group (Group 27). This incongruence could be indicative of another mix-up, although this cannot be confirmed without resampling these accessions. In any case, the samples of 'Taputini' do not group with the putatively prehistoric *kumara* lineage or any other Eastern Polynesian samples.

Based on Fig. 4.6, the tentative conclusion is that neither 'Hutihuti', 'Rekamaroa' nor 'Taputini' are derived from true, prehistoric Māori varieties, and instead all are derived from 19th or 20th century European introductions. Although this contradicts Yen (1974) and Green (2005), it is compatible with more recent historical research that favours European-era introductions for at least two of these varieties. Gould (2007) provides evidence that 'Rekamaroa' was introduced from California in about 1900 (it was given the Māori name *Reke Mau Roa*, long lingering sweetness).

'Hutihuti' also appears to be a European introduction, although in this case the name was borrowed from a true Māori variety that has since died out (Gould, 2007). Independent historical analysis recently completed by Green supports the European-era introductions for 'Hutihuti' and 'Rekamaroa'. The origin of 'Taputini' remains unresolved, but historical research is continuing (Roger Green, pers. comm.).

4.5 CONCLUSIONS

The large number of samples, broad geographic coverage and high resolution of the sweet potato neighbour-joining tree allows testing of a number of specific elements of the Tripartite Hypothesis.

A number of Eastern Polynesian accession group together in what has been identified, based on linguistic and morphological evidence, as the *kumara* lineage that Polynesians introduced from South America in prehistory. During the historic period, this lineage spread west, reaching Samoa and Fiji and also the Loyalty Islands. There is some evidence of two separate introductions into Polynesia (the Ia and Ib lineages) although this requires further testing with additional samples.

The sweet potatoes of Island Southeast Asia are highly admixed, with evidence for extensive exchange of diverse cultivars over a large area. Because of this it is not possible to distinguish separate *camote* and *batata* lineages in the current analysis.

Unexpectedly, a very large number of accessions from Papua New Guinea, Island Melanesia and Western Polynesia group together in what I have termed the Melanesia cluster. There are effectively no accessions from Island Southeast Asia or Eastern Polynesia in this cluster. This cluster is not predicted under the Tripartite Hypothesis, and some modification of the model for sweet potato dispersal in Oceania may be required.

An aim not yet achieved is to determine the South American region whose sweet potato varieties are most closely related to sweet potato accessions in the *kumara* lineage, and use this to infer the specific location in South America where Polynesians made landfall. To do this successfully would require a large number of sweet potato samples from coastal regions in western South America. The potential for further sampling is outlined in Chapter Five.

Commercial New Zealand varieties of sweet potato fall into four distinct lineages: the Waina group (which contains 'Waina', 'Owairaka Red' and 'Gisborne Red'), the 'Tauranga Red' lineage, the 'Owairaka Pink' lineage and the 'Toka Toka Gold' lineage (although 'Toka Toka Gold' is quite unstable in the tree, it is still clear that it is separate from all other commercial New Zealand varieties).

The placement of 'Hutihuti', 'Rekamaroa' and 'Taputini' is inconsistent with their status as prehistoric Māori sweet potato. Although this contradicts Yen (1963, 1974) and Green (2005), it is in agreement with more recent historical research which suggests 'Hutihuti' and 'Rekamaroa' and possibly 'Taputini' too are European-era introductions (Gould, 2007; Roger Green, pers. comm.). Even if all three of these varieties prove to be historic introductions, they are clearly still of importance to Māori, which makes the discovery that 'Hutihuti' grown in New Zealand is probably just mislabelled 'Rekamaroa' highly relevant. This may result in a desire to re-introduce to New Zealand living material of Y-512 'Hutihuti' from the Yen Collection in Tsukuba, Japan.

CHAPTER FIVE

SUMMARY AND FUTURE WORK

Pai Kanohi, with Tahā Huahua Bottle Gourd Containers for Preserving Kererū (Wood Pigeons)

Ruatahuna, Huiarau Range (near Lake Waikaremoana), New Zealand, 1903

Ref. No. G-17332-1/1

Reproduced with permission of and acknowledgement to Archives New Zealand (National Publicity Studios Collection), and permission of the Alexander Turnbull Library, Wellington, New Zealand

5.1 SUMMARY

Using the bottle gourd and sweet potato to test hypotheses of contact between Polynesia and the New World required significantly different approaches. Reasons for this include:

- differences in biology (ploidy and reproductive systems) and levels of genetic variability,
- numbers of available samples,
- prehistoric and historic distribution patterns,
- available archaeological, linguistic and historical information, and
- the types of questions being asked.

These differences affected the choice of molecular marker systems, the types of analyses performed and the interpretation of the data.

For the bottle gourd, a set of seven markers (two chloroplast and five nuclear) was sequenced in 36 accessions of bottle gourd from Asia, the Americas and Polynesia. Genetically distinguishing American and Asian gourds allowed the origins of the Polynesian gourd to be inferred; the current data are consistent with a dual origin. By combining information from a number of sources, a model for the domestication(s) and global dispersal of the bottle gourd is proposed.

For the sweet potato, AFLP profiles were generated for 270 unique accessions of from Asia, Island Melanesia, Polynesia and the Americas. The putative *kumara* lineage was identified (corresponding to the prehistoric, Polynesian introduction from South America). Sweet potato accessions from Asia to Western Polynesia proved more diverse, and the relationships between them are more complex than previously realised. The phylogenetic positions of the Māori varieties 'Hutihuti', 'Rekamaroa' and 'Taputini' are inconsistent with these accessions representing pre-European cultivars.

The remainder of this chapter suggests avenues for future research in the bottle gourd and sweet potato, and other commensal plant species that may be implicated in contact between Polynesia and the New World.

5.2 FUTURE WORK: BOTTLE GOURD

5.2.1 THE OCEANIC BOTTLE GOURD

Many interesting questions remain about the origins of the Polynesian bottle gourd:

- When was the bottle gourd introduced to Polynesia?
- Was it ever present in the Bottle Gourd Gap region?
- Were the ssp. *siceraria* alleles that are present in modern Māori accessions also present in prehistoric Eastern Polynesia?
- Does the Asian origin of the Polynesian bottle gourd derive from an old 5,500 yr BP Austronesian lineage, or a recent 200 BC Indian–Indonesia lineage?
- What is the origin of the morphologically African phallocrypt gourds from New Guinea — are they from Asia, consistent with preliminary molecular research (Decker-Walters *et al.*, 2001), or is a late AD 300 introduction from Africa (perhaps Madagascar) supported?
- From where in the New World (e.g., Peru–Ecuador or California) does the New World lineage derive?
- Is there evidence (e.g., linguistic) for human-mediated dispersal, or was dispersal effected naturally?

Many of these questions are solvable through a multidisciplinary approach including: additional sampling, ancient DNA, AMS radiocarbon dating, isotope analysis, the use of high resolution genetic markers, linguistic research, archaeological analysis of macro-remains and microfossils, and histological examination.

5.2.1.1 Ancient DNA

> *"Many gourds are preserved in museums, of course, but the empty fruits are of little aid in attempting to determine their relationships."*
>
> Heiser, C. B., Jr (1973a)

The techniques of ancient DNA have become a powerful tool in reconstructing evolutionary relationships (e.g., Matisoo-Smith & Robins, 2004; Shapiro *et al.*, 2004; Larson *et al.*, 2005), and there is massive potential for ancient DNA analysis of prehistoric and protohistoric bottle gourd material to further elucidate the domestication and dispersal of this species in Polynesia and globally. It is possible that the New Zealand bottle gourds used for this study have been contaminated by gene flow (especially pollen-mediated) from post-European contact introductions. Therefore, the tentative conclusion of a dual origin should be further tested by aDNA analysis of protohistoric herbarium and anthropological bottle gourd material, as well as prehistoric archaeological material from sites throughout Polynesia, Southeast Asia and the New World. The high-copy number *trn*C–D and *trn*S–G chloroplast markers will be most amenable to this, but these may not fully represent the diversity within Polynesia, and the biparentally inherited, codominant nuclear markers will also be required.

Despite the loss of bottle gourds worldwide, and especially in Polynesia (Dodge, 1943; Eames & St. John, 1943; Métraux, 1971 [1940], p. 157), a large amount of bottle gourd material exists in research institutions around the world, representing archaeological and ethnographic collections from Asia, Africa, the New World and Oceania. For the extinct Hawaiian *ipu nui* gourd, Eames & St. John (1943) report "scores in the Bishop Museum." The good representation of bottle gourd material is partly due to the species' unusual (amongst plants) property of producing woody exocarp tissue which can survive *post mortem*. For example, in the Smithsonian Institution, there are over 1,000 gourd artefacts in the archaeological and ethnographic collections (Sarah Zabriskie, pers. comm.) — probably more than for any other plant species. Current research (i.e., Erickson *et al.*, 2005) has already shown the potential of this material for aDNA analysis.

For questions surrounding the origin of the bottle gourd in Polynesia, it will be necessary to perform aDNA analysis of gourds from Asia, the New World and Oceania. Analysis of New World bottle gourd will be particularly important to determine if ssp. *siceraria* alleles were present there prehistorically (see below). Table 5.1 lists priority gourd samples for testing the hypothesis of a New World origin for the Polynesian bottle gourd. For Polynesia, I have selected prehistoric archaeological material and some early historic samples. For Asia, I have listed the 7,000 yr BP seeds from Hemudu, China. For the New World, due to the large amount of material available, I have selected mainly seeds for two reasons:

1. Based on morphology, seeds can be ascribed to one of the subspecies, and it will be interesting to examine the concordance between morphology and genetics.
2. Seeds probably contain DNA in higher amounts, and of better quality, than exocarp fragments (after all, seeds have evolved to successfully transmit functional DNA to the next generation). Both the yield and quality of DNA extracted from the 40-year-old Māori gourds is very high. Obtaining high quality DNA will also increase the probability that nuclear DNA can be amplified from archaeological material. In Polynesia, where hybridisation between Asian and American gourds may have occurred, it will be critical to amplify and analyse nuclear DNA.

For all three regions, I have focussed on material from the Smithsonian Institution because this is readily available. Samples displayed in red have already been obtained.

From the small amounts of sample available for aDNA analysis, it will also be possible to obtain AMS radiocarbon dates and isotope (e.g., ^{13}C, ^{15}N) data. These additional techniques will add important temporal and spatial (e.g., latitude, altitude, distance from Ocean) dimensions (e.g., Hawke & Holdaway, 2005; Noreen Tuross, pers. comm.).

Table 5.1
Priority Bottle Gourd Samples for aDNA Analysis

Region	Location	Tissue	Age	Sample Held[b]	Ref. No	Reference
Polynesia	Huahine, Society Islands	Exocarp	AD 850–1200	?	?	Emory (1979, pp. 202–204)
	Akipou, Nuku Hiva, Marquesas Islands	Exocarp	AD 1046–1596	? AMNH, NY, USA	?	Suggs (1961)
	Vaipikoau, Ua Huka, Marquesas Islands	Exocarp	*ca* AD 1150	?	?	Kirch (1973)
	Whakamoenga Cave, Lake Taupo, New Zealand	Exocarp	AD 1289–1401	? AWMM, New Zealand	?	Leahy (1976)
	Waitore, Patea, Taranaki	Exocarp	AD 1380–1500	? PM, New Zealand	?	Cassels (1979)
	Māhā‘ulepū Caves, Kaua‘i, Hawai‘i	Seeds	AD 1425–1520	?	?	Burney *et al.* (2001)
	Whakamoenga Cave, Lake Taupo, New Zealand	Seeds	AD 1616–1760	? AWMM, New Zealand	?	Leahy (1976)
	Kohika, Bay of Plenty, New Zealand	Exocarp & Seeds	AD 1650	UA, New Zealand	?	Irwin (2004a)
	Nihoa, Hawai‘i	Exocarp	Prehistoric?	? BP, HI, USA	?	Emory (1971 [1928], pp. 34–36, 47)
	Hawai‘i	Exocarp	1858	SI, D.C., USA	E003555-0	—
	Easter Island	Exocarp	1888	SI, D.C., USA	E129756-0	—
	Easter Island	Exocarp	1888	SI, D.C., USA	E129757-0	—
	Hawai‘i	Exocarp	1909	SI, D.C., USA	E257965-0	—
	Hawai‘i	Exocarp	1909	SI, D.C., USA	E257976-0	—
Asia	Hemudu, Zhejiang Province, China	Seeds	7,200 and 6,900 yr BP	?	?	Chang (1986, pp. 208–210)
	Nias Island, Indonesia	Exocarp	1903	SI, D.C., USA	E221756-0	—
	Mindoro Island, Philippines	Exocarp	1909	SI, D.C., USA	E253263-0	—
	Thailand	Exocarp	1969	SI, D.C., USA	E410875-0	—

Table continued next page...

Table continued...

Region	Location	Tissue	Age	Sample Held[a,b]	Ref. No	Reference
New World	Huaca Prieta, northern Peru	Seeds	5,000 yr BP	?	?	Whitaker & Bird (1949)
	Ocampo Caves, Mexico	Seeds	3,400–2,400 yr BP	NMC, Canada	?	Whitaker (1971)
	Huaynuma, Casma Valley, Peru	Seeds	4,000 yr BP	?	?	Whitaker (1983)
	Pampa Llamas, Casma Valley, Peru	Seeds	3.500 yr BP	?	?	Whitaker (1983)
	Tortugas, Casma Valley, Peru	Seeds	3,500 yr BP	?	?	Whitaker (1983)
	Las Haldas, Casma Valley, Peru	Seeds	3,500 yr BP	?	?	Whitaker (1983)
	Las Haldas, Casma Valley, Peru	Seeds	3,500 yr BP	?	?	Whitaker (1983)
	San Diego, Casma Valley, Peru	Seeds	3,400 yr BP	?	?	Whitaker (1983)
	Pampa Rosario, Casma Valley, Peru	Seeds	500 BC	?	?	Whitaker (1983)
	Aspero, Puerto de Supe, Peru	Seeds (288)	500 BC ?	CU, NY, USA	747/41A	Towle (1952)
	Huaca de la Cruez, Virú Valley, Peru	Seeds (88)	AD 500 ?	CU, NY, USA	3/V-162	Towle (1952)
	Ica Valley, Peru	Seeds	AD 1250	MAUCB, CA, USA	4-5680	Whitaker (1948)
	Chincha, Peru	Seeds	AD 1250	MAUCB, CA, USA	4-3965s	Whitaker (1948)
	Pachacamac, Lurin Valley, Peru	Seeds (2)	AD 1400 ?	CU, NY, USA	135/41A	Towle (1952)
	Pachacamac, Lurin Valley, Peru	Seeds (10)	AD 1400 ?	CU, NY, USA	G81	Towle (1952)
	Chincha, Peru	Seeds	*ca* AD 1500	MAUCB, CA, USA	4-3889b	Whitaker (1948)
	Peru	Exocarp	Prehistoric? (coll. 1887)	SI, D.C., USA	A132611-1	—
	Peru	Exocarp	Prehistoric? (coll. 1887)	SI, D.C., USA	A132616-0	—
	Peru	Exocarp	1919	SI, D.C., USA	E307843-0	—
	Colombia	Exocarp	1931	SI, D.C., USA	E362513-0	—

[a] Samples in red have already been obtained and are stored at the Ancient DNA Laboratory, Allan Wilson Centre, Massey University, Palmerston North, New Zealand

[b] AMNH = American Museum of Natural History, New York; AWMM = Auckland War Memorial Museum, Auckland; PM = Patea Museum, Patea; UA = University of Auckland, Auckland; BP = Bishop Museum, Honolulu; SI = Smithsonian Institution, Washington; NMC = National Museum of Canada, Ottawa; CU = Columbia University, New York; MAUCB = Museum of Anthropology, University of California — Berkeley, Berkeley

5.2.1.2 Additional Sampling and Germplasm Conservation

Although molecular research on prehistoric and protohistoric material is most promising, it may be possible to sample additional modern material that is supposedly derived from pre-European Polynesian introductions. The Māori gourds described by Maingay (1985, pp. 85–87), if available, are ideal. This research, combined with aDNA, would allow us to determine changes in allele frequency through time.

Effort should be made to preserve existing Māori cultivars. The Māori cultivars ‘GA’, ‘GD’ and ‘MG’ (Table 3.2) probably only exist as a few seeds of decreasing viability in the Plant & Food Research collection. Attempted germination of these seeds (Section 3.3.4.1) resulted in ~50% germination, and, of the seeds that did germinate, all produced weak, abnormal seedlings. Mike Burtenshaw attempted to grow four of these seedlings to maturity; three died before flowering, and one produced fruit but the seed did not mature (Mike Burtenshaw, pers. comm.). The efficient tissue culture method developed to vegetatively propagate bottle gourd plants from cotyledon explants (Han *et al.*, 2004) may be a useful technique for preserving Māori gourd cultivars from old seed with reduced viability.

5.2.1.3 High Resolution DNA Sequencing Markers

Based on their ability to distinguish modern Asian and New World cultivars of bottle gourd, the SCAR markers developed for this project will allow us to accurately determine from which of these regions the Polynesian bottle gourd is derived. They will also be useful for addressing some of the other outstanding questions regarding the dispersal of the bottle gourd.

The SCAR markers, however, are unlikely to provide the resolution required to resolve intra-continental dispersal events, i.e., the specific regions within Asia or the New World from which the Polynesian bottle gourd is derived. The effectiveness of the markers will also be reduced if the aDNA results of Erickson *et al.* (2005) are supported, i.e., prehistoric bottle gourds from the New World are ssp. *asiatica*. In these cases, it will be necessary to develop additional markers that provide intra-continental

resolution. These could be SCAR-derived — a number of the potential markers described in Chapter Two are at various stages of development. Microsatellite markers could also be developed, although the success of this approach is likely to be limited in bottle gourd (see Chapter Two). The cpDNA markers ccSSR09 and ccSSR10, which I was unable to amplify, but were polymorphic for Chung *et al.* (2003), may be useful. Marker development based on new pyrosequencing technologies is another option, e.g., the Complexity Reduction of Polymorphic Sequences (CRoPS) technique may allow rapid development of SNP and microsatellite markers from pyrosequenced AFLP fragments (van Eijk, 2006).

5.2.1.4 Linguistic Analysis

There appears to have been little, or no, research to determine if the New World is a source of prehistoric words for the bottle gourd found in Eastern Polynesia. This is despite the comparatively large amount of research on the origin of the Polynesian word *kumara* (and cognates), and the importance of this evidence for supporting a prehistoric, human-mediated transfer of the sweet potato into Polynesia. The sweet potato research has even led to a relatively small area in the Ecuadorian Gulf of Guayaquil being suggested as the point of Polynesian–New World contact (Scaglion, 2005).

Comparisons of words for the bottle gourds in the New World (especially near the hypothesised contact points of Peru–Ecuador and southern California) may prove similarly fruitful in supporting human-mediated dispersal and locating a contact point. I have begun some very preliminary research in collating indigenous words for the bottle gourd in Polynesia and the New World (Table 5.2). I have attempted to restrict Table 5.2 to vernacular terms for the plant only; terms for parts of the plant, different shapes and colours of gourds, articles made from the exocarp, and metaphorical extensions of gourd terms have been excluded. Some of these additional terms, however, may be useful when linguistic studies are pursued.

Table 5.2
Indigenous Words for Bottle Gourd in the New World and Polynesia

Region	Word	Location	Reference(s)
New World	*puru, poro*	Peru (Quechua)	Yacovleff & Herrera (1934–1935, pp. 312–313)[a]; Whitaker (1948); Towle (1961, p. 93)
	matti, maté	Peru (Quechua)	Yacovleff & Herrera (1934–1935, pp. 312–313); Whitaker (1948); Towle (1961, p. 93)
	ancana	Peru	Whitaker (1948)
Polynesia	*hue, fue*	Eastern Polynesia	Green (2000b); see refs in Maingay (1985, pp. 291–299)
	hue mao'i	Marquesas Is.	see refs in Maingay (1985, pp. 291–299)
	hue kava, 'ue	Cook Is.	see refs in Maingay (1985, pp. 291–299)
	uhe	Tuamotu Arch.	see refs in Maingay (1985, pp. 291–299)
	ipu	Hawai'i, Easter Is.	Eames & St. John (1943); Métraux (1971 [1940])
	hua ipu	Hawai'i	see refs in Maingay (1985, pp. 291–299)
	kaha	Easter Is.	Métraux (1971 [1940], p. 157)
	kotawa	New Zealand	Bulmer (1989)
	kowenewene	New Zealand	Best (1976 [1925], pp. 244–245)
	wenewene	New Zealand	Best (1976 [1925], pp. 244–245)
	fagu	Samoa	see refs in Maingay (1985, pp. 291–299)
	fangu	Tonga	see refs in Maingay (1985, pp. 291–299)

[a] Translation from Spanish kindly provided by Carlos Lehnebach

For North America, words for the bottle gourd in indigenous languages of Mexico and the American Southwest (New Mexico, Arizona, Texas, California and Nevada) should be collated to determine if there is evidence for linguistic borrowing between this region and Polynesia. This should include the Chumashan and Gabrielino languages of California's Channel Islands where there is evidence for Hawaiian voyagers making contact between AD 400–800 (Jones & Klar, 2005).

5.2.2 DOMESTICATION AND DISPERSAL

Many questions also remain about the domestication and early worldwide dispersal of the bottle gourd. Although these events preceded dispersal into Polynesia, current research shows they have important implications for understanding the distribution of the species in Oceania. Immediate questions include: Are the prehistoric Peruvian gourds of Asian origin, consistent with the aDNA results of Erickson *et al.* (2005), or are they from Africa, consistent with their morphology? What is the relationship between domesticated bottle gourds and the wild Zimbabwean accession? What is the distribution of ssp. *asiatica* and ssp. *siceraria* alleles in Africa? Specifically, is there support for Ethiopia containing both subspecies? Was the bottle gourd domesticated in Africa and, if so, where and when? What were the dispersal routes within Africa — are gourds of southern and western Africa derived from east African gourds? Is there evidence for a separate domestication event in East Asia — are gourds from this region more diverse than elsewhere in Asia? What were the dispersal routes from Africa to India, China, Japan, and Island Southeast Asia? What were the dispersal routes within the New World — from North to South America (or vice-versa), across the Caribbean, or overland through Central America? How do allele frequencies change through time — especially in the New World, where Erickson *et al.* (2005) conclude that pre-Columbian Asian gourds were rapidly replaced with African gourds early in the European settlement of the New World? What are the implications of dispersal patterns of the bottle gourd for prehistoric human mobility?

These questions will be answered using the same strategies as in Polynesia — but primarily aDNA analysis of prehistoric and protohistoric material, additional sampling of modern landraces, and the use of high resolution genetic markers.

5.2.2.1 Landrace Sampling

Despite the recent, rapid decrease in gourd cultivation and the loss of landraces from which ancestry is known (Prendergast & Decker-Walters, 2000; Heiser, 2001), recent research (e.g., Decker-Walters *et al.*, 2001; Erickson *et al.*, 2005; Clarke *et al.*, 2006),

indicates there is a lot of potential in sampling modern bottle gourd landraces from Africa, Asia and possibly the New World. Unlike for New Zealand, these continental regions appear less affected by modern gene flow and thus better represent the prehistoric genetic diversity. A large number of bottle gourd landraces have already been collected, and are listed in Table 5.3.

Table 5.3
Germplasm Collections of Bottle Gourd Landraces

Collection Held By	Accessions (*N*)	Regions Represented	Reference
International Plant Genetic Resources Inst., Nairobi, Kenya	269	Kenya (52 locations)	Morimoto *et al.* (2005)
Vavilov Research Institute, St. Petersburg, Russia	255	Unknown (but ≥ 45 "landraces")	Piskunova (2002)
Indiana University, Bloomington, IN, USA	102	Africa (37); Asia (28); New World (37)	Heiser (1973b)
The Cucurbit Network, Miami, FL, USA	31	Africa (7); Asia (15); New World (9)	Decker-Walters *et al.* (2001)
Smithsonian Institution, Washington, D.C., USA	30	Africa (18); Asia (12)	Erickson *et al.* (2005)

5.2.2.2 Ancient DNA and Histological Examination

Ancient DNA analysis and histological examination (including phytoliths) of exocarp tissue from the wild Zimbabwean gourd (Decker-Walters *et al.*, 2004b) and the 11–8,000 yr BP thin-shelled "gourd" from Thailand (Gorman, 1969; Yen, 1977) may be useful for understanding domestication of the species. The Thai gourd, although excluded as domesticated by Heiser (1979a, pp. 82–83), could be wild *Lagenaria*. Ancient DNA analysis could be performed on the unidentified cucurbit seeds which were recovered from the Thai site (Yen, 1977). The oldest (and widely-cited) bottle gourd remains from Africa — from the Egyptian tomb at Luxor — should also be re-

examined (they are held at Kew) and the identification confirmed (Schweinfurth, 1884b). These remains would also be interesting for ancient DNA analysis.

The same New World samples listed in Table 5.3 will also be useful for elucidating the origin and diversity of the American gourds. Remains from Coxcatlan Cave in Mexico would be ideal for determining changes in allele frequency through time as rind samples appear at 7,200 yr BP, AD 1340 and six samples in the 17th century (Smith, 2005).

5.3 FUTURE WORK: SWEET POTATO

5.3.1 SWEET POTATO PHYLOGENETICS

In many ways the analysis presented here is still preliminary and there are several avenues for future research. As discussed in Holland *et al.* (2008; Appendix 8), there is still scope to get more phylogenetic information out of AFLP data. In particular there could be gains from moving away from a single peak threshold applied to all characters. For instance, an examination of individual characters on the estimated tree reveals that some characters change state ($0 \leftrightarrow 1$) many times across the tree. Rather than reflecting true mutations, this might also mean that the peak height threshold has been chosen poorly with respect to this character. Also, the two distance measures considered here (Hamming and Jaccard) represent two extremes of how shared absences can be treated: either identically to shared presence (Hamming) or giving them weight 0 (Jaccard). It would be interesting to explore schemes that gave intermediate weights to shared absences.

Another important issue is assessing statistical support for the different groups found within the tree. Traditional bootstrap methods (Felsenstein, 1985) are known to perform poorly for large intra-population datasets (Matisoo-Smith *et al.*, 1998). Future work will include more relaxed methods of assessing support such as those suggested in Matisoo-Smith *et al.* (1998).

Another goal is to see if the outgroup taxa can be brought back into the analysis. This may be a simple matter of constraining the outgroup taxa to form a group and checking to see how many of the ingroup relationships are affected.

5.3.2 IMPROVED SAMPLING

To establish the location in South America at which Polynesians made contact, it will be necessary to obtain more samples from the coastal regions of western South America. It

may be possible to access well-provenanced accessions from the International Potato Center (CIP) in Lima. More samples from Mexico, Brazil and the Caribbean may also allow the *camote* and *batata* lineages to be identified in the Western Pacific, as well as the origins on the large Melanesia cluster. In terms of samples from within Oceania, the priority should be additional samples from Polynesia that are thought to be part of the *kumara* lineage, especially for New Zealand and Hawai‘i, which are currently not represented in the *kumara* group identified in Fig. 4.6. There is also a *kumara toia ua moa*, a probable prehistoric variety, believed to be available on Mangaia (Richard Walter, pers. comm.). Additional samples from Island Southeast Asia may also help resolve the complicated relationships in this region, especially samples from Borneo, Sulawesi and the Maluku Islands which lie in the probably area of confluence of the *camote* and *batata* lineages.

5.3.3 AN INTEGRATED APPROACH

The most informative reconstructions of sweet potato dispersals in Oceania will come from fully integrated studies that include genetics, linguistics, morphological data and historical evidence. Much more detailed research in all of these areas should prove fruitful, especially in determining the origins for the sweet potato in New Zealand. The lack of actual sweet potato remains (for example, from archaeological and ethnographic sources) and the complicated genetics of the sweet potato probably mean that the potential for ancient DNA (aDNA) research is limited (especially compared to the bottle gourd) but it may still be worthwhile exploring whole genome amplification techniques (WGA) to obtain usable quantities of DNA from herbarium specimens. For example, it would be useful to know the genetic relationships between the Banks and Solander “Kumala” and the *kumara* Ia and Ib lineages.

5.4 COMMENSAL PLANTS AND NEW WORLD CONTACT

Other plant species that may be prehistoric Polynesian introductions to the Pacific are the Polynesian tomato (*Solanum repandum*) and the soapberry (*Sapindus saponaria*). The coconut (*Cocos nucifera*) may have gone in the other direction (see Section 1.4.1.1). The relationship between the Chilean wine palm (*Jubaea chilensis*) and the extinct Easter Island palm (*Paschalococos disperta*?) also remains unresolved. The origins of all these species could be tested using molecular approaches, especially if herbarium and archaeological material can be included.

Marker development in minor crop species for which there are no available genetic resources remains a significant time, technical and financial constraint. Probably the most promising developments are coming from next generation sequencing technologies such as 454 (Roche) and Solexa (Illumina). The potential of these technologies is still to be fully realised, and it is not yet clear whether the data from these instruments will be used directly for population genetic and phylogenetic analyses, or whether they will be useful as an intermediate step in marker development (e.g., van Eijk, 2006; Cronn *et al.*, 2008). But in any case, future research on commensal species in Oceania will be aided by these new sequencing technologies.

REFERENCES

Abbo, S., Lev-Yadun, S., and Ladizinsky, G. 2001. Tracing the wild genetic stocks of crop plants. *Genome* **44:** 309–310.

Albach, D. C. 2007. Amplified fragment length polymorphisms and sequence data in the phylogenetic analysis of polyploids: multiple origins of *Veronica cymbalaria* (Plantaginaceae). *New Phytologist* **176:** 481–498.

Albani, M. C., Battey, N. H., and Wilkinson, M. J. 2004. The development of ISSR-derived SCAR markers around the *SEASONAL FLOWERING LOCUS* (*SFL*) in *Fragaria vesca*. *Theoretical and Applied Genetics* **109:** 571–579.

Allaby, R. 2007. Origins of plant exploitation in Near Oceania: a review. *In* Friedlaender, J. S. (Ed.), *Genes, language, and culture history in the Southwest Pacific*, Human Evolution Series, pp. 181–198. Oxford University Press: New York, New York, United States.

Allen, M. S., Matisoo-Smith, E., and Horsburgh, A. 2001. Pacific 'Babes': issues in the origins and dispersal of Pacific pigs and the potential of mitochondrial DNA analysis. *International Journal of Osteoarchaeology* **11:** 4–13.

Alonso-Blanco, C., Peeters, A. J. M., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J., and Kuiper, M. T. R. 1998. Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. *The Plant Journal* **14:** 259–271.

Althoff, D. M., Gitzendanner, M. A., and Segraves, K. A. 2007. The utility of amplified fragment length polymorphisms in phylogenetics: a comparison of homology within and between genomes. *Systematic Biology* **56:** 477–484.

Álvarez, I., and Wendel, J. F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29:** 417–434.

Anderson, A. 2001. No meat on that beautiful shore: the prehistoric abandonment of subtropical Polynesian Islands. *International Journal of Osteoarchaeology* **11:** 14–23.

———. 2006. Polynesian seafaring and American horizons: a response to Jones and Klar. *American Antiquity* **71:** 759–763.

Anderson, A., and Sinoto, Y. 2002. New radiocarbon ages of colonization sites in East Polynesia. *Asian Perspectives* **41:** 242–257.

Anderson, A. J., Conte, E., Clark, G. R., Sinoto, Y., and Petchey, F. J. 1999. Renewed excavations at Motu Paeao, Maupiti Island, French Polynesia: preliminary results. *New Zealand Journal of Archaeology* **21:** 47–66.

Applied Biosystems. 2004. *GeneMapper® Software Version 3.7 User Guide*. Applied Biosystems: Foster City, California, United States.

Archak, S., Gaikwad, A. B., Gautam, D., Rao, E. V. V. B., Swamy, K. R. M., and Karihaloo, J. L. 2003. Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. *Genome* **46:** 362–369.

Archibald, J. K., Crawford, D. J., Santos-Guerra, A., and Mort, M. E. 2006a. The utility of automated analysis of inter-simple sequence repeat (ISSR) loci for resolving relationships in the Canary Island species of *Tolpis* (Asteraceae). *American Journal of Botany* **93:** 1154–1162.

Archibald, J. K., Mort, M. E., Crawford, D. J., and Santos-Guerra, A. 2006b. Evolutionary relationships within recently radiated taxa: comments on methodology and analysis of inter-simple sequence repeat data and other hypervariable, dominant markers. *Taxon* **55:** 747–756.

Arnold, J. E. 2007. Credit where credit is due: the history of the Chumash oceangoing plank canoe. *American Antiquity* **72:** 196–209.

Austin, C. C. 1999. Lizards took express train to Polynesia. *Nature* **397:** 113–114.

Austin, D. F. 1988. The taxonomy, evolution and genetic diversity of sweetpotatoes and related wild species. *In* Gregory, P. (Ed.), *Exploration, maintenance, and utilization of sweetpotato genetic resources*, pp. 27–60. International Potato Center: Lima, Peru.

Bailey, C. D., Hughes, C. E., and Harris, S. A. 2004. Using RAPDs to identify DNA sequence loci for species level phylogeny reconstruction: an example from *Leucaena* (Fabaceae). *Systematic Botany* **29:** 4–14.

Baldwin, B. G. 1992. Phylogenetic utility of the Internal Transcribed Spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* **1:** 3–16.

Ballard, C., Brown, P., Bourke, R. M., and Harwood, T. (Eds). 2005. *The Sweet Potato in Oceania: A Reappraisal*. Ethnology Monographs 19/Oceania Monograph 56. Ethnology, University of Pittsburgh: Pittsburgh, Pennsylvania, United States/Oceania Publications, University of Sydney: Sydney, New South Wales, Australia.

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., and Meyer, A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439:** 719–723.

Barnes, S. S., Matisoo-Smith, E., and Hunt, T. L. 2006. Ancient DNA of the Pacific rat (*Rattus exulans*) from Rapa Nui (Easter Island). *Journal of Archaeological Science* **33:** 1536–1540.

Barrau, J. 1957. L'Enigme de la patate douce en Océanie. *Etudes d'Outre-Mer* **40:** 83–87.

Beaglehole, J. C. (Ed.). 1962. *The Endeavour Journal of Joseph Banks 1768–1771*. Vol. 1 of 2. Angus & Robertson: Sydney, New South Wales, Australia.

Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L., and Trujillo, I. 2003. Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theoretical and Applied Genetics* **107:** 736–744.

Bellwood, P. 1997. *Prehistory of the Indo-Malaysian Archipelago*. Revised edition. University of Hawai‘i Press: Honolulu, Hawai‘i, United States.

———. 2005. *First Farmers: The Origins of Agricultural Societies*. Blackwell: Malden, Massachusetts, United States.

Bennett, M. D., and Leitch, I. J. 2004. *Plant DNA C-values Database: Angiosperm DNA C-values database* [electronic source]. http://www.rbgkew.org.uk/cval/homepage.html. Last updated: 12/2004. Accessed: 21/06/2006. Royal Botanic Gardens, Kew: Richmond, Surrey, United Kingdom.

Bensch, S., and Åkesson, M. 2005. Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14:** 2899–2914.

Bensch, S., Åkesson, S., and Irwin, D. E. 2002. The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology* **11:** 2359–2366.

Best, E. 1976 [1925]. *Maori Agriculture*. Dominion Museum Bulletin 9. Board of Maori Ethnological Research for the Dominion Museum: Wellington, New Zealand.

Bevacqua, R. F. 1994. Origin of horticulture in Southeast Asia and the dispersal of domesticated plants to the Pacific islands by Polynesian voyagers — the Hawaiian Islands case study. *HortScience* **29:** 1226–1229.

Bikandi, J., San Millán, R., Rementeria, A., and Garaizar, J. 2004. *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* **20:** 798–799.

Blair, M. W., Panaud, O., and McCouch, S. R. 1999. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **98:** 780–792.

Blears, M. J., De Grandis, S. A., Lee, H., and Trevors, J. T. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology & Biotechnology* **21:** 99–114.

Bohs, L. 2004. A chloroplast DNA phylogeny of *Solanum* section *Lasiocarpa*. *Systematic Botany* **29:** 177–187.

Bonatto, S. L., Redd, A. J., Salzano, F. M., and Stoneking, M. 1996. Lack of ancient Polynesian–Amerindian contact. *American Journal of Human Genetics* **59:** 253–256.

Bonhoure, D., Rowe, E., Mariano, A. J., and Ryan, E. H. 2004. *Ocean Surface Currents: The South Equatorial Current System* [electronic source]. http://oceancurrents.rsmas.miami.edu/atlantic/south-equatorial.html. Last updated: 2004. Accessed: 13/01/2005. HYCOM Consortium for Data Assimilative Modeling, University of Miami: Miami, Florida, United States.

Bonin, A., Pompanon, F., and Taberlet, P. 2005. Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. *Methods in Enzymology* **395:** 145–161.

Bonin, A., Ehrich, D., and Manel, S. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology* **16:** 3737–3758.

Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13:** 3261–3273.

Boukar, O., Kong, L., Singh, B. B., Murdock, L., and Ohm, H. W. 2004. AFLP and AFLP-derived SCAR markers associated with *Striga gesnerioides* resistance in cowpea. *Crop Science* **44:** 1259–1264.

Bradeen, J. M., and Simon, P. W. 1998. Conversion of an AFLP fragment linked to the carrot Y_2 locus to a simple, codominant, PCR-based marker form. *Theoretical and Applied Genetics* **97:** 960–967.

Bramanti, B., Hummel, S., Chiarelli, B., and Herrmann, B. 2003. Ancient DNA analysis of the delta F508 mutation. *Human Biology* **75:** 105–115.

Brand, D. D. 1971. The sweet potato: an exercise in methodology. *In* Riley, C. L., Kelley, J. C., Pennington, C. W., and Rands, R. L. (Eds), *Man across the Sea: Problems of Pre-Columbian Contacts*, pp. 343–365. University of Texas Press: Austin, Texas, United States.

Brownstein, M. J., Carpten, J. D., and Smith, J. R. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20:** 1004–1010.

Brugmans, B., van der Hulst, R. G. M., Visser, R. G. F., Lindhout, P., and van Eck, H. J. 2003. A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Research* **31:** e55.

Bryant, V. M. 2003. Invisible clues to New World plant domestication. *Science* **299:** 1029–1030.

Buck, P. 1938. *Vikings of the Sunrise*: New York, New York, United States.

Bulmer, S. 1989. Gardens in the south: diversity and change in prehistoric Maaori agriculture. *In* Harris, D. R. and Hillman, G. C. (Eds), *Foraging and Farming: The Evolution of Plant Exploitation*, One World Archaeology 13, pp. 688–705. Unwin Hyman: London, United Kingdom.

Burkill, I. H. 1935. *A Dictionary of the Economic Products of the Malay Peninsula*. Vol. 2 of 2. Crown Agents for the Colonies: London, United Kingdom.

Burney, D. A., James, H. F., Pigott Burney, L., Olson, S. L., Kikuchi, W., Wagner, W. L., Burney, M., McCloskey, D., Kikuchi, D., Grady, F. V., Gage, R., II, and Nishek, R. 2001. Fossil evidence for a diverse biota from Kaua'i and its transformation since human arrival. *Ecological Monographs* **71:** 615–641.

Burtenshaw, M. K. 1999. Māori gourds: an American connection? *Journal of the Polynesian Society* **108:** 427–433.

———. 2003. The first horticultural plant propagated from seed in New Zealand: *Lagenaria siceraria*. *New Zealand Garden Journal* **6:** 10–16.

Bussell, J. D., Waycott, M., and Chappill, J. A. 2005. Arbitrarily amplified DNA markers as characters for phylogenetic inference. *Perspectives in Plant Ecology, Evolution and Systematics* **7:** 3–26.

Caetano-Anollés, G., and Gresshoff, P. M. 1994. Staining nucleic acids with silver: and alternative to radioisotopic and fluorescent labeling. *Promega Notes* **45:** 13.

Camp, W. H. 1954. A possible source for American pre-Columbian gourds. *American Journal of Botany* **41:** 700–701.

Campbell, D., Duchesne, P., and Bernatchez, L. 2003. AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Molecular Ecology* **12:** 1979–1991.

Capparelli, R., Viscardi, M., Amoroso, M. G., Blaiotta, G., and Bianco, M. 2004. Inter-simple sequence repeat markers and flow cytometry for the characterization of closely related *Citrus limon* germplasms. *Biotechnology Letters* **26:** 1295–1299.

Carrier, L. 1923. *The Beginnings of Agriculture in America*. McGraw-Hill: New York, New York, United States.

Carter, G. F. 1977. A hypothesis suggesting a single origin of agriculture. *In* Reed, C. A. (Ed.), *Origins of Agriculture*, World Anthropology, pp. 89–133. Mouton: The Hague, The Netherlands.

Cassels, R. 1979. Early prehistoric wooden artefacts from the Waitore Site (N136/16), near Patea, Taranaki. *New Zealand Journal of Archaeology* **1:** 85–108.

Chakravarty, H. L. 1968. A new species of African *Lagenaria* (Cucurbitaceae). *Annals of the Missouri Botanical Garden* **55:** 69–72.

Chang, K.-C. 1986. *The Archaeology of Ancient China*. 4th edition. Yale University Press: New Haven, Connecticut, United States.

Chiang, T.-Y. 2000. Lineage sorting accounting for the disassociation between chloroplast and mitochondrial lineages in oaks of southern France. *Genome* **43:** 1090–1094.

Chiba, N., Suwabe, K., Nunome, T., and Hirai, M. 2003. Development of microsatellite markers in melon (*Cucumis melo* L.) and their application to major cucurbit crops. *Breeding Science* **53:** 21–27.

Chung, S.-M., and Staub, J. E. 2003. The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theoretical and Applied Genetics* **107:** 757–767.

Chung, S.-M., Decker-Walters, D. S., and Staub, J. E. 2003. Genetic relationships within the Cucurbitaceae as assessed by consensus chloroplast simple sequence repeats (ccSSR) marker and sequence analyses. *Canadian Journal of Botany* **81:** 814–832.

Clark, J. M. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research* **16:** 9677–9686.

Clarke, A. C. 2001. *Did Polynesian voyagers introduce the sweet potato (*Ipomoea batatas*) into Oceania from South America? A start on the molecular evidence*. Unpublished BSc(Hons) Report. Massey University, Palmerston North, New Zealand.

Clarke, A. C., Burtenshaw, M. K., McLenachan, P. A., Erickson, D. L., and Penny, D. 2006. Reconstructing the origins and dispersal of the Polynesian bottle gourd (*Lagenaria siceraria*). *Molecular Biology and Evolution* **23:** 893–900.

Cohen, M. N. 1977. Population pressure and the origins of agriculture: an archaeological example from the coast of Peru. *In* Reed, C. A. (Ed.), *Origins of Agriculture*, World Anthropology, pp. 135–177. Mouton: The Hague, The Netherlands.

Coil, J., and Kirch, P. V. 2005. An Ipomoean landscape: archaeology and the sweet potato in Kahikinui, Maui, Hawaiian Islands. *In* Ballard, C., Brown, P., Bourke, R. M., and Harwood, T. (Eds), *The Sweet Potato in Oceania: A Reappraisal*, Ethnology Monographs 19/Oceania Monograph 56, pp. 71–84. Ethnology, University of Pittsburgh: Pittsburgh, Pennsylvania, United States/Oceania Publications, University of Sydney: Sydney, New South Wales, Australia.

Colenso, W. 2001 [1880]. *On the Vegetable Food of the Ancient New Zealanders before Cook's Visit*. Transactions of the New Zealand Institute. Kiwi Publishers: Christchurch, New Zealand.

Collerson, K. D., and Weisler, M. I. 2007. Stone adze compositions and the extent of ancient Polynesian voyaging and trade. *Science* **317:** 1907–1911.

Conte, E., and Anderson, A. 2003. Radiocarbon ages for two sites on Ua Huka, Marquesas. *Asian Perspectives* **42:** 155–160.

Cooke, R. 2005. Prehistory of Native Americans on the Central American land bridge: colonization, dispersal, and divergence. *Journal of Archaeological Research* **13:** 129–187.

Crawford, D., Tago-Nakazawa, M., Stuessy, T. F., Anderson, G. J., Bernardello, G., Ruiz, E., Jensen, R. J., Baeza, C. M., Wolfe, A. D., and Silva, M. 2001. Intersimple sequence repeat (ISSR) variation in *Lactoris fernandeziana* (Lactoridaceae), a rare endemic of the Juan Fernández Archipelago, Chile. *Plant Species Biology* **16:** 185–192.

Crawford, G. W. 1992. Prehistoric plant domestication in East Asia. *In* Cowan, C. W. and Watson, P. J. (Eds), *The Origins of Agriculture: An International Perspective*, Smithsonian Series in Archaeological Enquiry, pp. 7–38. Smithsonian Institution Press: Washington, D.C., United States.

Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., and Mockler, T. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* **36:** e122.

Cutler, H. C., and Whitaker, T. W. 1961. History and distribution of the cultivated cucurbits in the Americas. *American Antiquity* **26:** 469–485.

———. 1967. Cucurbits from the Tehuacán Caves. *In* Byers, D. S. (Ed.), *The Prehistory of the Tehuacán Valley: Environment and Subsistence*, vol. 1, pp. 212–219. University of Texas Press: Austin, Texas, United States.

Danin-Poleg, Y., Reis, N., Tzuri, G., and Katzir, N. 2001. Development and characterization of microsatellite markers in *Cucumis*. *Theoretical and Applied Genetics* **102:** 61–72.

Datwyler, S. L., and Wolfe, A. D. 2004. Phylogenetic relationships and morphological evolution in *Penstemon* subg. *Dasanthera* (Veronicaceae). *Systematic Botany* **29:** 165–176.

Davidson, J. M. 2000. The Polynesian settlement of New Zealand: the last frontier of the Oceanic world. *People and Culture in Oceania* **16:** 19–39.

Davies, O. 1975. The dispersal of the bottle gourd. *Current Anthropology* **16:** 665.

De Jong, W., Forsyth, A., Leister, D., Gebhardt, C., and Baulcombe, D. C. 1997. A potato hypersensitive resistance gene against potato virus X maps to a resistance gene cluster on chromosome 5. *Theoretical and Applied Genetics* **95:** 246–252.

Decker-Walters, D. S., and Walters, T. 2000. The bobbing *Benincasa*. *The Cucurbit Network News* **7:** 1–3.

Decker-Walters, D. S., Chung, S.-M., and Staub, J. E. 2004a. Plastid sequence evolution: a new pattern of nucleotide substitutions in the Cucurbitaceae. *Journal of Molecular Evolution* **58:** 606–614.

Decker-Walters, D. S., Staub, J., López-Sesé, A., and Nakata, E. 2001. Diversity in landraces and cultivars of bottle gourd (*Lagenaria siceraria*; Cucurbitaceae) as assessed by random amplified polymorphic DNA. *Genetic Resources and Crop Evolution* **48:** 369–380.

Decker-Walters, D. S., Wilkins-Ellert, M., Chung, S.-M., and Staub, J. E. 2004b. Discovery and genetic assessment of wild bottle gourd [*Lagenaria siceraria* (Mol.) Standley; Cucurbitaceae] from Zimbabwe. *Economic Botany* **58:** 501–508.

Degnan, J. H., and Salter, L. A. 2005. Gene tree distributions under the coalescent process. *Evolution* **59:** 24–37.

Deng, Z., Huang, S., Xiao, S., and Gmitter, F. G., Jr. 1997. Development and characterization of SCAR markers linked to the citrus tristeza virus resistance gene from *Poncirus trifoliata*. *Genome* **40:** 697–704.

Denham, T. P., Haberle, S. G., Lentfer, C., Fullagar, R., Field, J., Therin, M., Porch, N., and Winsborough, B. 2003. Origins of agriculture at Kuk Swamp in the Highlands of New Guinea. *Science* **301:** 189–193.

Després, L., Gielly, L., Redoutet, B., and Taberlet, P. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution* **27:** 185–196.

Diamond, J. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. W. W. Norton & Company: New York, New York, United States.

———. 2002. Evolution, consequences and future of plant and animal domestication. *Nature* **418:** 700–707.

Diamond, J., and Bellwood, P. 2003. Farmers and their languages: the first expansions. *Science* **300:** 597–603.

Dodge, E. S. 1943. *Gourd Growers of the South Seas: An Introduction to the Study of the Lagenaria Gourd in the Culture of the Polynesians*. The Gourd Society of America: Boston, Massachusetts, United States.

Doran, E., Jr. 1971. The sailing raft as a great tradition. *In* Riley, C. L., Kelley, J. C., Pennington, C. W., and Rands, R. L. (Eds), *Man across the Sea: Problems of Pre-Columbian Contacts*, pp. 115–138. University of Texas Press: Austin, Texas, United States.

Doran, G. H., Dickel, D. N., and Newsom, L. A. 1990. A 7,290-year-old bottle gourd from the Windover site, Florida. *American Antiquity* **55:** 354–360.

Doyle, J. J., and Doyle, J. L. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12:** 13–15.

Dussle, C. M., Quint, M., Xu, M. L., Melchinger, A. E., and Lübberstedt, T. 2002. Conversion of AFLP fragments tightly linked to SCMV resistance genes *Scmv1* and *Scmv2* into simple PCR-based markers. *Theoretical and Applied Genetics* **105:** 1190–1195.

Eames, A. J., and St. John, H. 1943. The botanical identity of the Hawaiian ipu nui or large gourd. *American Journal of Botany* **30:** 255–259.

Ebbesmeyer, C. C., and Ingraham, W. J., Jr. 1994. Pacific toy spill fuels ocean current pathways research. *Eos — Transactions, American Geophysical Union* **75:** 425, 427, 430.

Elias, S. A. 2001. Mutual climatic range reconstructions of seasonal temperatures based on Late Pleistocene fossil beetle assemblages in Eastern Beringia. *Quaternary Science Reviews* **20:** 77–91.

Elias, S. A., Berman, D., and Alfimov, A. 2000. Late Pleistocene beetle faunas of Beringia: where east met west. *Journal of Biogeography* **27:** 1349–1363.

Ellis, A. G., Weis, A. E., and Gaut, B. S. 2006. Evolutionary radiation of “stone plants” in the genus *Argyroderma* (Aizoaceae): unraveling the effects of landscape, habitat, and flowering. *Evolution* **60:** 39–55.

Emory, K. P. 1971 [1928]. *Archaeology of Nihoa and Necker Islands*. Bernice P. Bishop Museum Bulletin 53. Bishop Museum Press: Honolulu, Hawai‘i, United States.

———. 1979. The Societies. *In* Jennings, J. D. (Ed.), *The Prehistory of Polynesia*, pp. 200–221. Australian National University Press: Canberra, A.C.T., Australia.

Engel, F. 1970. Exploration of the Chilca Canyon, Peru. *Current Anthropology* **11:** 55–58.

Erickson, D. L., Smith, B. D., Clarke, A. C., Sandweiss, D. H., and Tuross, N. 2005. An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 18315–18320.

Esselman, E. J., Jianqiang, L., Crawford, D. J., Windus, J. L., and Wolfe, A. D. 1999. Clonal diversity in the rare *Calamagrostis porteri* ssp. *insperata* (Poaceae): comparative results for allozymes and

random amplified polymorphic DNA (RAPD) and intersimple sequence repeat (ISSR) markers. *Molecular Ecology* **8:** 443–451.

Estoup, A., and Cornuet, J.-M. 1999. Microsatellite evolution: inferences from population data. *In* Goldstein, D. B. and Schlötterer, C. (Eds), *Microsatellites: Evolution and Applications*, pp. 49–65. Oxford University Press: New York, New York, United States.

Fajardo, S., La Bonte, D. R., and Jarret, R. L. 2002. Identifying and selecting for genetic diversity in Papua New Guinea sweetpotato *Ipomoea batatas* (L.) Lam. germplasm collected as botanical seed. *Genetic Resources and Crop Evolution* **49:** 463–470.

FAO. 2008. *FAOSTAT: ProdSTAT: Crops* [electronic source]. http://faostat.fao.org/site/567/default.aspx#ancor. Last updated: 2007. Accessed: 4/12/2008. Food and Agriculture Organization of the United Nations: Rome, Italy.

Fay, M. F., Conwan, R. S., and Leitch, I. J. 2005. The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Annals of Botany* **95:** 237–246.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39:** 783–791.

Finney, B. 1994. Polynesian voyagers to the New World. *Man and Culture in Oceania* **10:** 1–13.

———. 1997. Experimental voyaging, oral traditions and long-distance interaction in Polynesia. *In* Weisler, M. I. (Ed.), *Prehistoric Long-Distance Interaction in Oceania: An Interdisciplinary Approach*, New Zealand Archaeological Association Monograph No. 21, pp. 38–52. New Zealand Archaeological Association: Auckland, New Zealand.

———. 2007. Tracking Polynesian seafarers. *Science* **317:** 1873–1874.

Fitzpatrick, S. M., and Callaghan, R. 2009. Examining dispersal mechanisms for the translocation of chicken (*Gallus gallus*) from Polynesia to South America. *Journal of Archaeological Science* **36:** 214-223.

Flachowsky, H., Schumann, E., Weber, W. E., and Peil, A. 2001. Application of AFLP for the detection of sex-specific markers in hemp. *Plant Breeding* **120:** 305–309.

Flannery, K. V. 1973. The origins of agriculture. *Annual Review of Anthropology* **2:** 271–310.

Frackman, S., Kobs, G., Simpson, D., and Storts, D. 1998. Betaine and DMSO: enhancing agents for PCR. *Promega Notes* **65:** 27–30.

Free, J. B. 1993. *Insect Pollination of Crops*. 2nd edition. Academic Press: London, United Kingdom.

Garoia, F., Guarniero, I., Grifoni, D., Marzola, S., and Tinti, F. 2007. Comparative analysis of AFLPs and SSRs efficiency in resolving population genetic structure of Mediterranean *Solea vulgaris*. *Molecular Ecology* **16:** 1377–1387.

Gerber, S., Mariette, S., Streiff, R., Bodénès, C., and Kremer, A. 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology* **9:** 1037–1048.

Gibco BRL. *n.d. AFLP® Analysis System I; AFLP Starter Kit.* Instruction Manual. Life Technologies, Inc.: Rockville, Maryland, United States.

Glover, I. C. 1977. The Late Stone Age in eastern Indonesia. *World Archaeology* **9:** 42–61.

Goldman, D. H., Jansen, R. K., van den Berg, C., Leitch, I. J., Fay, M. F., and Chase, M. W. 2004. Molecular and cytological examination of *Calopogon* (Orchidaceae, Epidendroideae): circumscription, phylogeny, polyploidy, and possible hybrid speciation. *American Journal of Botany* **91:** 707–723.

Golson, J. 2002. Gourds in New Guinea, Asia and the Pacific. *In* Bedford, S., Sand, C., and Burley, D. (Eds), *Fifty years in the field. Essays in honour and celebration of Richard Shutler Jr's archaeological career*, New Zealand Archaeological Association Monograph 25, pp. 69–78.

Golson, J., Lampert, R. J., Wheeler, J. M., and Ambrose, W. R. 1967. A note on carbon dates for horticulture in the New Guinea Highlands. *Journal of the Polynesian Society* **76:** 369–371.

Gongora, J., Rawlence, N. J., Mobegi, V. A., Jianlin, H., Alcalde, J. A., Matus, J. T., Hanotte, O., Moran, C., Austin, J. J., Ulm, S., Anderson, A. J., Larson, G., and Cooper, A. 2008a. Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 10308–10313.

———. 2008b. Reply to Storey *et al.*: More DNA and dating studies needed for ancient El Arenal-1 chickens. *Proceedings of the National Academy of Sciences of the United States of America* **105:** E100.

Gonzalo, M. J., Oliver, M., Garcia-Mas, J., Monfort, A., Dolcet-Sanjuan, R., Katzir, N., Arús, P., and Monforte, A. 2005. Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). *Theoretical and Applied Genetics* **110:** 802–811.

Gorman, C. F. 1969. Hoabinhian: a pebble-tool complex with early plant associations in Southeast Asia. *Science* **163:** 671–673.

Gould, A. 2007. *Kumara in New Zealand: Some Issues and Observations* [report]. Report No. Wai 262 # R35, 16 January 2007. The Waitangi Tribunal, Wellington, New Zealand.

Gray, R. D., and Jordan, F. M. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* **405:** 1052–1055.

Green, R. C. 1991a. Near and Remote Oceania — disestablishing "Melanesia" in culture history. *In* Pawley, A. (Ed.), *Man and a Half: Essays in Pacific Anthropology and Ethnobiology in Honour of Ralph Bulmer*, Memoir No. 48, pp. 491–502. The Polynesian Society, University of Auckland: Auckland, New Zealand.

———. 1991b. The Lapita cultural complex: current evidence and proposed models. *In* Bellwood, P. (Ed.), *Indo-Pacific Prehistory 1990*, vol. 2, Bulletin of the Indo-Pacific Prehistory Association 11, pp. 295–305. Australian National University: Canberra, A.C.T., Australia.

———. 1998. Rapanui origins prior to European contact: the view from Eastern Polynesia. *In* Vargas Casanova, P. (Ed.), *Easter Island and East Polynesian Prehistory*, Second International Congress on Easter Island and East Polynesian Archaeology, pp. 87–110. Instituto de Estudios Isla de Pascua, Facultad de Arquitectura y Urbanismo, Universidad de Chile: Santiago, Chile.

———. 2000a. Lapita and the cultural model for intrusion, integration and innovation. *In* Anderson, A. and Murray, T. (Eds), *Australian Archaeologist: Collected Papers in Honour of Jim Allen*, pp. 372–392. Coombs Academic Publishing, The Australian National University: Canberra, A.C.T., Australia.

———. 2000b. A range of disciplines support a dual origin for the bottle gourd in the Pacific. *Journal of the Polynesian Society* **109:** 191–197.

———. 2001. Commentary on the sailing raft, the sweet potato and the South American connection. *Rapa Nui Journal* **15:** 69–77.

———. 2003. The Lapita horizon and traditions — signature for one set of oceanic migrations. *In* Sand, C. (Ed.), *Pacific Archaeology: Assessments and Prospects: Proceedings of the International Conference for the 50th Anniversary of the First Lapita Excavation*, Koné, Nouméa, 2002, pp. 95–120. Département Archéologie, Service des Musées et du Patrimoine: Nouméa, Nouvelle-Calédonie.

———. 2005. Sweet potato transfers in Polynesian prehistory. *In* Ballard, C., Brown, P., Bourke, R. M., and Harwood, T. (Eds), *The Sweet Potato in Oceania: A Reappraisal*, Ethnology Monographs 19/Oceania Monograph 56, pp. 43–62. Ethnology, University of Pittsburgh: Pittsburgh, Pennsylvania, United States/Oceania Publications, University of Sydney: Sydney, New South Wales, Australia.

Green, R. C., and Weisler, M. I. 2002. The Mangarevan sequence and dating of the geographic expansion into Southeast Polynesia. *Asian Perspectives* **41:** 213–241.

Green, R. C., Jones, M., and Sheppard, P. 2008. The reconstructed environment and absolute dating of SE-SZ-8 Lapita site on Nendö, Santa Cruz, Solomon Islands. *Archaeology in Oceania* **43:** 49–61.

Griffin, P. B., Riley, T., Rosendahl, P., and Tuggle, H. D. 1971. Archaeology of Halawa and Lapakahi: windward valley and leeward slope. *New Zealand Archaeological Association: Newsletter* **14:** 101–112.

Guo, Y.-P., Vogl, C., Van Loo, M., and Ehrendorfer, F. 2006. Hybrid origin and differentiation of two tetraploid *Achillea* species in East Asia: molecular, morphological and ecogeographical evidence. *Molecular Ecology* **15:** 133–144.

Guppy, H. B. 1906. *Observations of a Naturalist in the Pacific between 1896 and 1899.* Vol. 2. Plant Dispersal. Macmillan and Co.: London, United Kingdom.

Habu, J., Kim, M., Katayama, M., and Komiya, H. 2001. Jomon subsistence-settlement systems at the Sannai Maruyama site. *In* Bellwood, P., Bowdery, D., Glover, I., Hudson, M., and Keates, S. (Eds), *Indo-Pacific Prehistory: The Melaka Papers*, Bulletin of the Indo-Pacific Prehistory Association 21, pp. 9–21. Australian National University: Canberra, A.C.T., Australia.

Hagen, A. R., Giese, H., and Brochmann, C. 2001. Trans-Atlantic dispersal and phylogeography of *Cerastium arcticum* (Caryophyllaceae) inferred from RAPD and SCAR markers. *American Journal of Botany* **88:** 103–112.

Han, J.-S., Oh, D.-G., Mok, I.-G., Park, H.-G., and Kim, C. K. 2004. Efficient plant regeneration from cotyledon explants of bottle gourd (*Lagenaria siceraria* Standl.). *Plant Cell Reports* **23:** 291–296.

Han, J.-S., Kim, C. K., Park, S. H., Hirschi, K. D., and Mok, I. G. 2005. *Agrobacterium*-mediated transformation of bottle gourd (*Lagenaria siceraria* Standl.). *Plant Cell Reports* **23:** 692–698.

Harlan, J. R. 1976. Plant and animal distribution in relation to domestication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **275:** 13–25.

Harvey, C. F., Gill, G. P., Crossman, C., and Fraser, L. G. 1997. Assessing relationships of kumara cultivars by RAPD analysis. *New Zealand Journal of Botany* **35:** 479–485.

Hather, J., and Kirch, P. V. 1991. Prehistoric sweet potato (*Ipomoea batatas*) from Mangaia Island, Central Polynesia. *Antiquity* **65:** 887–893.

Hawke, D. J., and Holdaway, R. N. 2005. Avian assimilation and dispersal of carbon and nitrogen brought ashore by breeding Westland petrels (*Procellaria westlandica*): a stable isotope study. *Journal of Zoology* **266:** 419–426.

He, X.-Q., Liu, Q.-C., Ishiki, K., Zhai, H., and Wang, Y.-P. 2006. Genetic diversity and genetic relationships among Chinese sweetpotato landraces revealed by RAPD and AFLP markers. *Breeding Science* **56:** 201–207.

Heiser, C. B., Jr. 1969. Systematics and the origin of cultivated plants. *Taxon* **18:** 36–45.

———. 1973a. The penis gourd of New Guinea. *Annals of the Association of American Geographers* **63:** 312–318.

———. 1973b. Variation in the bottle gourd. *In* Meggers, B. J., Ayensu, E. S., and Duckworth, W. D. (Eds), *Tropical Forest Ecosystems in Africa and South America: A Comparative Review*, pp. 121–128. Smithsonian Institution Press: Washington, D.C., United States.

———. 1979a. *The Gourd Book*. University of Oklahoma Press: Norman, Oklahoma, United States.

———. 1979b. Origins of some cultivated New World plants. *Annual Review of Ecology and Systematics* **10:** 309–326.

———. 1985. *Of Plants and People*. University of Oklahoma Press: Norman, Oklahoma, United States.

———. 1987. Origins of *Solanum lasiocarpum* and *S. repandum*. *American Journal of Botany* **74:** 1045–1048.

———. 1989. Domestication of Cucurbitaceae: *Cucurbita* and *Lagenaria*. *In* Harris, D. R. and Hillman, G. C. (Eds), *Foraging and Farming: The Evolution of Plant Exploitation*, One World Archaeology 13, pp. 471–480. Unwin Hyman: London, United Kingdom.

———. 1990. New perspectives on the origin and evolution of New World domesticated plants: summary. *Economic Botany* **44:** 111–116.

———. 2001. Untitled. *Economic Botany* **55:** 181–182.

Hengen, P. N. 1997. Optimizing multiplex and LA-PCR with betaine. *Trends in Biochemical Sciences* **22:** 225–226.

Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S. A. 1997. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Research* **25:** 3957–3958.

Hershkovitz, M. A., and Zimmer, E. A. 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research* **24:** 2857–2867.

Heyerdahl, T. 1952. *American Indians in the Pacific: The Theory behind the Kon-Tiki Expedition*. Allen and Unwin: London, United Kingdom.

Hillis, D. M., and Bull, J. J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42:** 182–192.

Hinkle, A. E. 2007. Population structure of Pacific *Cordyline fruticosa* (Laxmanniaceae) with implications for human settlement of Polynesia. *American Journal of Botany* **94:** 828–839.

Holdaway, R. N. 1996. Arrival of rats in New Zealand. *Nature* **384:** 225–226.

Holland, B. R., Penny, D., and Hendy, M. D. 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock — a simulation study. *Systematic Biology* **52:** 229–238.

Holland, B. R., Delsuc, F., and Moulton, V. 2005. Visualizing conflicting evolutionary hypotheses in large collections of trees: using consensus networks to study the origins of placentals and hexapods. *Systematic Biology* **54:** 66–76.

Holland, B. R., Clarke, A. C., and Meudt, H. M. 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology* **57:** 347–366.

Hollingsworth, P. M., and Ennos, R. A. 2004. Neighbour joining trees, dominant markers and population genetic structure. *Heredity* **92:** 490–498.

Hong, D., Wan, L. L., Liu, P., Yang, G., and He, Q. 2006. AFLP and SCAR markers linked to the suppressor gene (*Rf*) of a dominant genetic male sterility in rapeseed (*Brassica napus* L.). *Euphytica* **151:** 401–409.

Hoppe, B. L., Conti-Tronconi, B. M., and Horton, R. M. 1992. Gel-loading dyes compatible with PCR. *BioTechniques* **12:** 679–680.

Horrocks, M., and Lawlor, I. 2006. Plant microfossil analysis of soils from Polynesian stonefields in South Auckland, New Zealand. *Journal of Archaeological Science* **33:** 200–217.

Horrocks, M., Jones, M. D., Carter, J. A., and Sutton, D. G. 2000. Pollen and phytoliths in stone mounds at Pouerua, Northland, New Zealand: implications for the study of Polynesian farming. *Antiquity* **74:** 863–872.

Horrocks, M., Jones, M. D., Beever, R. E., and Sutton, D. G. 2002. Analysis of plant microfossils in prehistoric coprolites from Harataonga Bay, Great Barrier Island, New Zealand. *Journal of the Royal Society of New Zealand* **32:** 617–628.

Horrocks, M., Smith, I. W. G., Nichol, S. L., and Wallace, R. 2008. Sediment, soil and plant microfossil analysis of Maori gardens at Anaura Bay, eastern North Island, New Zealand: comparison with descriptions made in 1769 by Captain Cook's expedition. *Journal of Archaeological Science* **35:** 2446–2464.

Howe, K. R. 2006. The last frontier. *In* Howe, K. R. (Ed.), *Vaka Moana: Voyages of the Ancestors: The Discovery and Settlement of the Pacific*, pp. 14–21. David Bateman: Auckland, New Zealand.

Hu, J., Nakatani, M., Lalusin, A. G., and Fujimura, T. 2004a. New microsatellite markers developed from reported *Ipomoea trifida* sequences and their application to sweetpotato and its related wild species. *Scientia Horticulturae* **102:** 375–386.

Hu, J., Nakatani, M., Mizuno, K., and Fujimura, T. 2004b. Development and characterization of microsatellite markers in sweetpotato. *Breeding Science* **54:** 177–188.

Huang, J. C., and Sun, M. 2000. Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theoretical and Applied Genetics* **100:** 1050–1060.

Huber, K. T., Langton, M., Penny, D., Moulton, V., and Hendy, M. 2002. Spectronet: a package for computing spectra and median networks. *Applied Bioinformatics* **1:** 159–161.

Hudson, J. L. 2004. Additional evidence for gourd floats on fishing nets. *American Antiquity* **69:** 586–587.

Hughes, C. E., Eastwood, R. J., and Bailey, C. D. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361:** 211–225.

Hunt, T. L., and Lipo, C. P. 2006. Late colonization of Easter Island. *Science* **311:** 1603–1606.

Hurles, M. E., Matisoo-Smith, E., Gray, R. D., and Penny, D. 2003a. Untangling Oceanic settlement: the edge of the knowable. *Trends in Ecology & Evolution* **18:** 531–540.

Hurles, M. E., Sykes, B. C., Jobling, M. A., and Forster, P. 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages. *American Journal of Human Genetics* **76:** 894–901.

Hurles, M. E., Maund, E., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B. C., and Jobling, M. A. 2003b. Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade. *American Journal of Human Genetics* **72:** 1282–1287.

Huson, D. H., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8:** Article 460.

Imamura, K. 1996. *Prehistoric Japan: New Perspectives on Insular East Asia*. UCL Press: London, United Kingdom.

Irwin, G. J. 1992. *The Prehistoric Exploration and Colonisation of the Pacific*. Cambridge University Press: Cambridge, Cambridgeshire, United Kingdom.

———. 2004a. Excavations and site history at Kohika. *In* Irwin, G. J. (Ed.), *Kohika: The Archaeology of a Late Maori Lake Village in the Ngati Awa Rohe, Bay of Plenty, New Zealand*, pp. 45–75. Auckland University Press: Auckland, New Zealand.

——— (Ed.). 2004b. *Kohika: The Archaeology of a Late Maori Lake Village in the Ngati Awa Rohe, Bay of Plenty, New Zealand*. Auckland University Press: Auckland, New Zealand.

———. 2006. Voyaging and settlement. *In* Howe, K. R. (Ed.), *Vaka Moana: Voyages of the Ancestors: The Discovery and Settlement of the Pacific*, pp. 54–91. David Bateman: Auckland, New Zealand.

Jarret, R. L., Gawel, N., and Whittemore, A. 1992. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. *Journal of the American Society for Horticultural Science* **117:** 633–637.

Jarret, R. L., Merrick, L. C., Holms, T., Evans, J., and Aradhya, M. K. 1997. Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai). *Genome* **40:** 433–441.

Jeffrey, C. 1967. *Cucurbitaceae*. Flora of Tropical East Africa. Crown Agents for Oversea Governments and Administrations: London, United Kingdom.

———. 1990. Appendix: an outline classification of the Cucurbitaceae. *In* Bates, D. M., Robinson, R. W., and Jeffrey, C. (Eds), *Biology and Utilization of the Cucurbitaceae*, pp. 449–463. Cornell University Press: Ithaca, New York, United States.

Jobst, J., King, K., and Hemleben, V. 1998. Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of the family Cucurbitaceae. *Molecular Phylogenetics and Evolution* **9:** 204–219.

Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A., and Karp, A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* **3:** 381–390.

Jones, T. L., and Klar, K. A. 2005. Diffusionism reconsidered: linguistic and archaeological evidence for prehistoric Polynesian contact with southern California. *American Antiquity* **70:** 457–484.

Jump, A. S., and Peñuelas, J. 2007. Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, *Fagus sylvatica*. *Molecular Ecology* **16:** 925–936.

Kardolus, J. P., van Eck, H. J., and van den Berg, R. G. 1998. The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). *Plant Systematics and Evolution* **210:** 87–103.

Katzir, N., Danin-Poleg, Y., Tzuri, G., Karchi, Z., Lavi, U., and Cregan, P. B. 1996. Length polymorphism and homologies of microsatellites in several Cucurbitaceae species. *Theoretical and Applied Genetics* **93:** 1282–1290.

Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X. Y., Nürnberg, P., Trent, R. J., and Stoneking, M. 2008. Genorne-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *American Journal of Human Genetics* **82:** 194–198.

Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A., Moyse-Faurie, C., Rutledge, R. B., Schiefenhoevel, W., Gil, D., Lin, A. A., Underhill, P. A., Oefner, P. J., Trent, R. J., and Stoneking, M. 2006. Melanesian and asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Molecular Biology and Evolution* **23:** 2234–2244.

Kennedy, J. 2008. Pacific bananas: complex origins, multiple dispersals? *Asian Perspectives* **47:** 75–94.

Kepler, A. K. 1998. *Hawaiian Heritage Plants*. Revised edition. University of Hawai‘i Press: Honolulu, Hawai‘i, United States.

Kilian, B., Özkan, H., Deusch, O., Effgen, S., Brandolini, A., Kohl, J., Martin, W., and Salamini, F. 2007. Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. *Molecular Biology and Evolution* **24:** 217–227.

Kirch, P. V. 1973. Prehistoric subsistence patterns in the northern Marquesas Islands, French Polynesia. *Archæology and Physical Anthropology in Oceania* **8:** 24–40.

———. 1985. *Feathered Gods and Fishhooks: An Introduction to Hawaiian Archaeology and Prehistory*. University of Hawai‘i Press: Honolulu, Hawai‘i, United States.

———. 1997. *The Lapita Peoples: Ancestors of the Oceanic World*. The Peoples of South-East Asia and the Pacific. Blackwell: Cambridge, Massachusetts, United States.

———. 1998. Landscapes of power: late prehistoric settlement and land use of marginal environments in the Hawaii Islands. *In* Vargas Casanova, P. (Ed.), *Easter Island and East Polynesian Prehistory*, Second International Congress on Easter Island and East Polynesian Archaeology, pp. 59–72. Instituto de Estudios Isla de Pascua, Facultad de Arquitectura y Urbanismo, Universidad de Chile: Santiago, Chile.

———. 2000. *On the Road of the Winds: An Archaeological History of the Pacific Islands before European Contact*. University of California Press: Berkeley, California, United States.

Kirch, P. V., and Kahn, J. G. 2007. Advances in Polynesian prehistory: a review and assessment of the past decade (1993–2004). *Journal of Archaeological Research* **15:** 191–238.

Kobiakova, J. A. 1930. The bottle gourd. *Bulletin of Applied Botany, Genetics and Plant Breeding* **23:** 475–520.

Koblmüller, S., Egger, B., Sturmbauer, C., and Sefc, K. M. 2007. Evolutionary history of Lake Tanganyika's scale-eating cichlid fishes. *Molecular Phylogenetics and Evolution* **44:** 1295–1305.

Koopman, W. J. M. 2005. Phylogenetic signal in AFLP data sets. *Systematic Biology* **54:** 197–217.

Koopman, W. J. M., and Gort, G. 2004. Significance tests and weighted values for AFLP similarities, based on arabidopsis *in silico* AFLP fragment length distributions. *Genetics* **167:** 1915–1928.

Kosman, E., and Leonard, K. J. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology* **14:** 415–424.

Kremer, A., Caron, H., Cavers, S., Colpaert, N., Gheysen, G., Gribel, R., Lemes, M., Lowe, A. J., Margis, R., Navarro, C., and Salgueiro, F. 2005. Monitoring genetic diversity in tropical trees with multilocus dominant markers. *Heredity* **95:** 274–280.

Kriegner, A., Cervantes, J. C., Burg, K., Mwanga, R. O. M., and Zhang, D. P. 2003. A genetic linkage map of sweetpotato [*Ipomoea batatas* (L.) Lam.] based on AFLP markers. *Molecular Breeding* **11:** 169–185.

Langdon, R. 1996. The soapberry, a neglected clue to Polynesia's prehistoric past. *Journal of the Polynesian Society* **105:** 185–200.

———. 2001. The bamboo raft as a key to the introduction of the sweet potato in prehistoric Polynesia. *Journal of Pacific History* **36:** 51–76.

Larson, G., Dobney, K., Albarella, U., Fang, M. Y., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L., and Cooper, A. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307:** 1618–1621.

Larson, G., Cucchi, T., Fujita, M., Matisoo-Smith, E., Robins, J., Anderson, A., Rolett, B., Spriggs, M., Dolman, G., Kim, T.-H., Thuy, N. T. D., Randi, E., Doherty, M., Due, R. A., Bollt, R., Djubiantono, T., Griffin, B., Intoh, M., Keane, E., Kirch, P., Li, K.-T., Morwood, M., Pedriña, L. M., Piper, P. J., Rabett, R. J., Shooter, P., Van den Bergh, G., West, E., Wickler, S., Yuan, J., Cooper, A., and Dobney, K. 2007. Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in island southeast Asia and Oceania. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 4834–4839.

Lathrap, D. W. 1977. Our father the cayman, our mother the gourd: Spinden revisited, or a unitary model for the emergence of agriculture in the New World. *In* Reed, C. A. (Ed.), *Origins of Agriculture*, World Anthropology, pp. 713–751. Mouton: The Hague, The Netherlands.

Leach, F. 1981. The prehistory of the southern Wairarapa. *Journal of the Royal Society of New Zealand* **11:** 11–33.

Leach, H. M. 1984. *1,000 Years of Gardening in New Zealand.* A. H. & A. W. Reed: Wellington, New Zealand.

Leahy, A. 1976. Whakamoenga Cave, Taupo, N94/7: a report on the ecology, economy and stratigraphy. *Records of the Auckland Institute and Museum* **13:** 29–75.

Lee, C., and Wen, J. 2004. Phylogeny of *Panax* using chloroplast *trn*C–*trn*D intergenic region and the utility of *trn*C–*trn*D in interspecific studies of plants. *Molecular Phylogenetics and Evolution* **31:** 894–903.

Lee, T., Burch, J. B., Coote, T., Fontaine, B., Gargominy, O., Pearce-Kelly, P., and Foighil, D. Ó. 2007. Prehistoric inter-archipelago trading of Polynesian tree snails leaves a conservation legacy. *Proceedings of the Royal Society B: Biological Sciences* **274:** 2907–2914.

Leonard, J. A., Wayne, R. K., Wheeler, J., Valadez, R., Guillén, S., and Vilà, C. 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science* **298:** 1613–1616.

Lewthwaite, S. L. 2004. *Storage root production in sweetpotato (*Ipomoea batatas *(L.) Lam.)*. Unpublished PhD thesis. Massey University, Palmerston North, New Zealand.

Li, L., and Garvin, D. F. 2003. Molecular mapping of *Or*, a gene inducing beta-carotene accumulation in cauliflower (*Brassica oleracea* L. var. *botrytis*). *Genome* **46:** 588–594.

Lie, B. A., Dupuy, B. M., Spurkland, A., Fernández-Viña, M. A., Hagelberg, E., and Thorsby, E. 2007. Molecular genetic studies of natives on Easter Island: evidence of an early European and Amerindian contribution to the Polynesian gene pool. *Tissue Antigens* **69:** 10–18.

Liu, A.-H., and Wang, J.-B. 2006. Genomic evolution of *Brassica* allopolyploids revealed by ISSR marker. *Genetic Resources and Crop Evolution* **53:** 603–611.

Liu, Y.-G., and Whittier, R. F. 1995. Thermal Asymmetric Interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25:** 674–681.

Liu, Y.-G., Mitsukawa, N., Oosumi, T., and Whittier, R. F. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *The Plant Journal* **8:** 457–463.

Lockhart, P. J., and McLenachan, P. A. 1997. Isolating polymorphic plant DNA fragments identified using AFLP™ technology without acrylamide gels: markers for evolutionary studies. *Focus* **19:** 70–71.

Maddison, D. R., Swofford, D. L., and Maddison, W. P. 1997. NEXUS: an extensible file format for systematic information. *Systematic Biology* **46:** 590–621.

Maingay, J. 1985. *Te Hue: People and a plant*. Unpublished PhD thesis. University of Auckland, Auckland, New Zealand.

Manfield, I. W., Pavlov, V. K., Li, J. H., Cook, H. E., Hummel, F., and Gilmartin, P. M. 2005. Molecular characterization of DNA sequences from the *Primula vulgaris S*-locus. *Journal of Experimental Botany* **56:** 1177–1188.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Mariette, S., Le Corre, V., Austerlitz, F., and Kremer, A. 2002. Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Molecular Ecology* **11:** 1145–1156.

Matisoo-Smith, E. 1994. The human colonisation of Polynesia. A novel approach: genetic analyses of the Polynesian rat (*Rattus exulans*). *Journal of the Polynesian Society* **103:** 75–87.

———. 1996. *No Hea Te Kiore: MtDNA Variation in* Rattus exulans*: A Model for Human Colonisation and Contact in Prehistoric Polynesia*. Unpublished PhD thesis. University of Auckland, Auckland, New Zealand.

———. 2002. Something old, something new: do genetic studies of contemporary populations reliably represent prehistoric populations of Pacific *Rattus exulans*? *Human Biology* **74:** 489–496.

———. 2007a. The peopling of Oceania. *In* Crawford, M. H. (Ed.), *Anthropological Genetics: Theory Methods and Applications*, pp. 409–432. Cambridge University Press: Cambridge, Cambridgeshire, United Kingdom.

———. 2007b. Animal translocations, genetic variation, and the human settlement of the Pacific. *In* Friedlaender, J. S. (Ed.), *Genes, language, and culture history in the Southwest Pacific*, Human Evolution Series, pp. 157–170. Oxford University Press: New York, New York, United States.

Matisoo-Smith, E., and Robins, J. H. 2004. Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 9167–9172.

Matisoo-Smith, E., Sutton, D. G., Ladefoged, T. N., Lambert, D. M., and Allen, J. S. 1999. Prehistoric mobility in Polynesia: mtDNA variation in *Rattus exulans* from the Chatham and Kermadec Islands. *Asian Perspectives* **38:** 186–199.

Matisoo-Smith, E., Roberts, R. M., Irwin, G. J., Allen, J. S., Penny, D., and Lambert, D. M. 1998. Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 15145–15150.

Matthews, P. J. 2003. Identification of *Benincasa hispida* (wax gourd) from the Kana archaeological site, Western Highlands Province, Papua New Guinea. *Archaeology in Oceania* **38:** 186–191.

———. 2006. Plant trails in Oceania. *In* Howe, K. R. (Ed.), *Vaka Moana: Voyages of the Ancestors: The Discovery and Settlement of the Pacific*, pp. 94–97. David Bateman: Auckland, New Zealand.

Matthews, P. J., and Gosden, C. 1997. Plant remains from waterlogged sites in the Arawe Islands, West New Britain Province, Papua New Guinea: implications for the history of plant use and domestication. *Economic Botany* **51:** 121–133.

Maunder, M., Culham, A., Bordeu, A., Allainguillaume, J., and Wilkinson, M. 1999. Genetic diversity and pedigree for *Sophora toromiro* (Leguminosae): a tree extinct in the wild. *Molecular Ecology* **8:** 725–738.

McGregor, C. E., Lambert, C. A., Greyling, M. M., Louw, J. H., and Warnich, L. 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica* **113:** 135–144.

McLenachan, P. A., Stöckler, K., Winkworth, R. C., McBreen, K., Zauner, S., and Lockhart, P. J. 2000. Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Molecular Ecology* **9:** 1899–1903.

McNab, R. 1914. *Historical Records of New Zealand*. Vol. 2 of 2. Government Printer: Wellington, New Zealand.

Mendelson, T. C., and Shaw, K. L. 2005. Use of AFLP markers in surveys of arthropod diversity. *Methods in Enzymology* **395:** 161–177.

Merriwether, D. A., Hodgson, J. A., Friedlaender, F. R., Allaby, R., Cerchio, S., Koki, G., and Friedlaender, J. S. 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 13034–13039 [Erratum: *P. Natl Acad. Sci. USA*. 2005. **102:** 16904].

Métraux, A. 1971 [1940]. *Ethnology of Easter Island*. Bernice P. Bishop Museum Bulletin 160. Bishop Museum Press: Honolulu, Hawai‘i, United States.

Meudt, H. M., and Clarke, A. C. 2007. Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends in Plant Science* **12:** 106–117.

Mogg, R., Batley, J., Hanley, S., Edwards, D., O’Sullivan, H., and Edwards, K. J. 2002. Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. *Theoretical and Applied Genetics* **105:** 532–543.

Morimoto, Y., Gikungu, M., and Maundu, P. 2004. Pollinators of the bottle gourd (*Lagenaria siceraria*) observed in Kenya. *International Journal of Tropical Insect Science* **24:** 79–86.

Morimoto, Y., Maundu, P., Fujimaki, H., and Morishima, H. 2005. Diversity of landraces of the white-flowered gourd (*Lagenaria siceraria*) and its wild relatives in Kenya: fruit and seed morphology. *Genetic Resources and Crop Evolution* **52:** 737–747.

Morimoto, Y., Maundu, P., Kawase, M., Fujimaki, H., and Morishima, H. 2006. RAPD polymorphism of the white-flowered gourd (*Lagenaria siceraria* (Molina) Standl. landraces and its wild relatives in Kenya. *Genetic Resources and Crop Evolution* **53:** 963–974.

Morin, P. A., Luikart, G., Wayne, R. K., and the SNP workshop group. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19:** 208–216.

Mort, M. E., and Crawford, D. J. 2004. The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon* **53:** 257–261.

Mueller, U. G., and Wolfenbarger, L. L. 1999. AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution* **14:** 389–394.

Muke, J., and Mandui, H. 2003. In the shadows of Kuk: evidence for prehistoric agriculture at Kana, Wahgi Valley, Papua New Guinea. *Archaeology in Oceania* **38:** 177–185.

Murray-McIntosh, R. P., Scrimshaw, B. J., Hatfield, P. J., and Penny, D. 1998. Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 9047–9052.

Negi, M. S., Devic, M., Delseny, M., and Lakshmikumaran, M. 2000. Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. *Theoretical and Applied Genetics* **101:** 146–152.

Nicod, J.-C., and Largiadèr, C. R. 2003. SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. *Nucleic Acids Research* **31:** e19.

Nik Hassan Shuhaimi. 1991. Recent research at Kuala Selinsing, Perak. *In* Bellwood, P. (Ed.), *Indo-Pacific Prehistory 1990*, Bulletin of the Indo-Pacific Prehistory Association 11, pp. 92–112. Australian National University: Canberra, A.C.T., Australia.

NIWA Science. 2001. *Climate Data: Mean Monthly Temperatures at 31 New Zealand Sites (1971-2000)* [electronic source]. http://www.niwascience.co.nz/edu/resources/climate/meanairtemp/. Last updated: 11/09/2001. Accessed: 01/08/2006. NIWA Science: Auckland, New Zealand.

Noguera, F. J., Capel, J., Alvarez, J. I., and Lozano, R. 2005. Development and mapping of a codominant SCAR marker linked to the *andromonoecious* gene of melon. *Theoretical and Applied Genetics* **110:** 714–720.

Nordyke, E. C. 1999. *Pacific Images: Views from Captain Cook's Third Voyage*. Hawaiian Historical Society: Honolulu, Hawai'i, United States.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13:** 1143–1155.

Ochman, H., Gerber, A. S., and Hartl, D. L. 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120:** 621–623.

Okamoto, H., and Hirochika, H. 2000. Efficient insertion mutagenesis of *Arabidopsis* by tissue culture-induced activation of the tobacco retrotransposon *Tto1*. *The Plant Journal* **23:** 291–304.

Olmstead, R. G., Michaels, H. J., Scott, K. M., and Palmer, J. D. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbc*L. *Annals of the Missouri Botanical Garden* **79:** 249–265.

Organ, J. 1963. *Gourds: Decorative and Edible for Garden, Craftwork and Table*. Faber and Faber: London, United Kingdom.

Osipova, E. S., Koveza, O. V., Troitskij, A. V., Dolgikh, Y. I., Shamina, Z. B., and Gostimskij, S. A. 2003. Analysis of specific RAPD and ISSR fragments in maize (*Zea mays* L.) somaclones and development of SCAR markers on their basis. *Russian Journal of Genetics* **39:** 1412–1419.

Paran, I., and Michelmore, R. W. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics* **85:** 985–993.

Parker, J. D., Rabinovitch, P. S., and Burmer, G. C. 1991. Targeted gene walking polymerase chain reaction. *Nucleic Acids Research* **19:** 3055–3060.

Penny, D., and Meyer, A. 2006. DNA and the settlement of Polynesia. *In* Howe, K. R. (Ed.), *Vaka Moana: Voyages of the Ancestors: The Discovery and Settlement of the Pacific*, pp. 98–99. David Bateman: Auckland, New Zealand.

Perrie, L. R., Brownsey, P. J., Lockhart, P. J., and Large, M. F. 2003. Evidence for an allopolyploid complex in New Zealand *Polystichum* (Dryopteridaceae). *New Zealand Journal of Botany* **41:** 189–215.

Pharmawati, M., Yan, G., and Finnegan, P. M. 2005. Molecular variation and fingerprinting of *Leucadendron* cultivars (Proteaceae) by ISSR markers. *Annals of Botany* **95:** 1163–1170.

Pickering, C. 1849. *The Races of Man and their Geographical Distribution*. John Chapman: London, United Kingdom.

Pickersgill, B. 1972. Cultivated plants as evidence for cultural contacts. *American Antiquity* **37:** 97–104.

Pickersgill, B., and Bunting, A. H. 1969. Cultivated plants and the Kon-Tiki theory. *Nature* **222:** 225–227.

Pickersgill, B., and Heiser, C. B., Jr. 1977. Origins and distribution of plants domesticated in the New World tropics. *In* Reed, C. A. (Ed.), *Origins of Agriculture*, World Anthropology, pp. 803–835. Mouton: The Hague, The Netherlands.

Pierson, M. J. 2007. *Deciphering the mtDNA Record of Prehistoric Population Movements in Oceania*. Unpublished PhD thesis. University of Canterbury, Christchurch, New Zealand.

Pierson, M. J., Martinez-Arias, R., Holland, B. R., Gemmell, N. J., Hurles, M. E., and Penny, D. 2006. Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Molecular Biology and Evolution* **23:** 1966–1975.

Piperno, D. R., and Pearsall, D. M. 1998. *The Origins of Agriculture in the Lowland Neotropics*. Academic Press: San Diego, California, United States.

Piperno, D. R., and Holst, I. 1998. The presence of starch grains on prehistoric stone tools from the humid neotropics: indications of early tuber use and agriculture in Panama. *Journal of Archaeological Science* **25:** 765–776.

Piperno, D. R., and Stothert, K. E. 2003. Phytolith evidence for early Holocene *Cucurbita* domestication in Southwest Ecuador. *Science* **299:** 1054–1057.

Piperno, D. R., Andres, T. C., and Stothert, K. E. 2000a. Phytoliths in *Cucurbita* and other neotropical Curcurbitaceae and their occurrence in early archaeological sties from the Lowland American tropics. *Journal of Archaeological Science* **27:** 193–208.

Piperno, D. R., Ranere, A. J., Holst, I., and Hansell, P. 2000b. Starch grains reveal early root crop horticulture in the Panamanian tropical forest. *Nature* **407:** 894–897.

Piperno, D. R., Holst, I., Wessel-Beaver, L., and Andres, T. C. 2002. Evidence for the control of phytolith formation in *Cucurbita* fruits by the hard rind (*Hr*) genetic locus: archaeological and ecological implications. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 10923–10928.

Piperno, D. R., Weiss, E., Holst, I., and Nadel, D. 2004. Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature* **430:** 670–673.

Piskunova, T. 2002. Status of the cucurbit collections in Russia. *In* Díez, M. J., Picó, B., and Nuez, F. (Eds), *Cucurbit Genetic Resources in Europe*, *Ad hoc* meeting, Adana, Turkey, 19 January 2002, pp. 37–38. International Plant Genetic Resources Institute: Rome, Italy.

Pissard, A., Ghislain, M., and Bertin, P. 2006. Genetic diversity of the Andean tuber-bearing species, oca (*Oxalis tuberosa* Mol.), investigated by inter-simple sequence repeats. *Genome* **49:** 8–16.

Polanco, C., González, A. I., de la Puente, R., Somalo, S., and Ruiz, M. L. 2005. Cloning of AFLP markers detected by fluorescence capillary electrophoresis. *Plant Molecular Biology Reporter* **23:** 271–277.

Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* **6:** 847–859.

Prendergast, H. D. V., and Decker-Walters, D. S. 2000. Preserving the gourd perspective (*Lagenaria siceraria* (Molina) Standl. [Cucurbitaceae]). *Economic Botany* **54:** 424–426.

Promega Corp. 1998. *SILVER SEQUENCE™ DNA Sequencing System*. Technical Manual No. 023. Promega Corp.: Madison, Wisconsin, United States.

———. 2003. *pGEM®-T and pGEM®-T Easy Vector Systems*. Technical Manual No. 042. Promega Corp.: Madison, Wisconsin, United States.

Provan, J., and Wilson, P. J. 2007. Development of microsatellites for the peat moss *Sphagnum capillifolium* using ISSR cloning. *Molecular Ecology Notes* **7:** 254–256.

Provan, J., Powell, W., and Hollingsworth, P. M. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* **16:** 142–147.

Qin, L., Prins, P., Jones, J. T., Popeijus, H., Smant, G., Bakker, J., and Helder, J. 2001. GenEST, a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP. *Nucleic Acids Research* **29:** 1616–1622.

Raina, S. N., Rani, V., Kojima, T., Ogihara, Y., Singh, K. P., and Devarumath, R. M. 2001. RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. *Genome* **44:** 763–772.

Ramírez, J. M. 1991. Transpacific contacts: the Mapuche connection. *Rapa Nui Journal* **4:** 53–55.

Ray, J. 1974 [1691]. *The Wisdom of God Manifested in the Works of the Creation*. Georg Olms Verlag: Hildesheim, Germany.

Rees, W. A., Yager, T. D., Korte, J., and von Hippel, P. H. 1993. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32:** 137–144.

Rensch, K. H. 1991. Polynesian plant names: linguistic analysis and ethnobotany, expectations and limitations. *In* Cox, P. A. and Banack, S. A. (Eds), *Islands, Plants, and Polynesians: An Introduction to Polynesian Ethnobotany*, pp. 97–111. Dioscorides Press: Portland, Oregon, United States.

Richardson, J. B., III. 1972. The pre-Columbian distribution of the bottle gourd (*Lagenaria siceraria*): a re-evaluation. *Economic Botany* **26:** 265–273.

Rieth, T. M., and Hunt, T. L. 2008. A radiocarbon chronology for Sāmoan prehistory. *Journal of Archaeological Science* **35:** 1901–1927.

Ripley, V. L., and Roslinsky, V. 2005. Identification of an ISSR marker for 2-propenyl glucosinolate content in *Brassica juncea* L. and conversion to a SCAR marker. *Molecular Breeding* **16:** 57–66.

Ritschel, P. S., Lins, T. C., Tristan, R. L., Buso, G. S. C., Buso, J. A., and Ferreira, M. E. 2004. Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). *BMC Plant Biology* **4:** Article 9.

Robinson, R. W., and Decker-Walters, D. S. 1997. *Cucurbits*. Vol. 6. Crop Production Science in Horticulture. CAB International: Wallingford, Oxfordshire, United Kingdon.

Rolett, B. V., and Conte, E. 1995. Renewed investigation of the Ha‘atuatua Dune (Nuku Hiva, Marquesas Islands): a key site in Polynesian prehistory. *Journal of the Polynesian Society* **104:** 195–228.

Ross, M. 1996. Reconstructing food plant terms and associated terminologies in Proto Oceanic. *In* Lynch, J. and Pat, F. (Eds), *Oceanic Studies: Proceedings of the First International Conference on Oceanic Linguistics*, Pacific Linguistic Series C: Vol. 133, pp. 163–221. Pacific Linguistics, Department of Linguistics, The Australian National University: Canberra, A.C.T., Australia.

Sakata, Y., Kubo, N., Morishita, M., Kitadani, E., Sugiyama, M., and Hirai, M. 2006. QTL analysis of powdery mildew resistance in cucumber (*Cucumis sativus* L.). *Theoretical and Applied Genetics* **112:** 243–250. [Erratum: *Theor. Appl. Genet.* 2006. **112:** 986].

Saliba-Colombani, V., Causse, M., Gervais, L., and Philouze, J. 2000. Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* **43:** 29–40.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74:** 5463–5467.

Savelkoul, P. H. M., Aarts, H. J. M., de Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J. L. W., Schouls, L., and Lenstra, J. A. 1999. Amplified-fragment length polymorphism analysis: the state of an art. *Journal of Clinical Microbiology* **37:** 3083–3091.

Scaglion, R. 2005. *Kumara* in the Ecuadorian Gulf of Guayaquil? *In* Ballard, C., Brown, P., Bourke, R. M., and Harwood, T. (Eds), *The Sweet Potato in Oceania: A Reappraisal*, Ethnology Monographs 19/Oceania Monograph 56, pp. 35–41. Ethnology, University of Pittsburgh: Pittsburgh, Pennsylvania, United States/Oceania Publications, University of Sydney: Sydney, New South Wales, Australia.

Schlötterer, C. 2004. The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics* **5:** 63–69.

Schoske, R. 2003. *The design, optimization and testing of Y chromosome short tandem repeat megaplexes*. Unpublished PhD thesis. American University, Washington, D.C., United States. http://www.cstl.nist.gov/div831/strbase/pub_pres/Schoske2003dis.pdf.

Schupp, J. M., Price, L. B., Klevytska, A., and Keim, P. 1999. Internal and flanking sequence from AFLP fragments using ligation-mediated suppression PCR. *BioTechniques* **26:** 905–200. [Erratum: *BioTechniques*. 1999. **27:** 340].

Schweinfurth, G. 1884a. Neue Funde auf dem Gebiete der Flora des alten Ägyptens [German]. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* **5:** 189–202.

———. 1884b. Further discoveries in the flora of Ancient Egypt. *Nature* **29:** 312–315.

Scotti, I., Vendramin, G. G., Matteotti, L. S., Scarponi, C., Sari-Gorla, M., and Binelli, G. 2000. Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Molecular Ecology* **9:** 699–708.

Seeman, B. 1865–1873. *Flora Vitiensis: A Description of the Plants of Viti or Fiji Islands with an Account of their History, Uses and Properties*. Reeve: London, United Kingdom.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M. L., Martin, L. D., Stephenson, R. O., Storer, J., Tedford, R., Zimov, S., and Cooper, A. 2004. Rise and fall of the Beringian steppe bison. *Science* **306:** 1561–1565.

Shepardson, B., Shepardson, D., Shepardson, F., Chiu, S., and Graves, M. 2008. Re-examining the evidence for late colonization on Easter Island. *Rapa Nui Journal* **22:** 97–101.

Shirasawa, K., Kishitani, S., and Nishio, T. 2004. Conversion of AFLP markers to sequence-specific markers for closely related lines in rice by use of the rice genome sequence. *Molecular Breeding* **14:** 283–292.

Shrivastava, U. 1991. Insect pollination in some Cucurbits. *In* van Heemert, C. and de Ruijter, A. (Eds), *The Sixth International Symposium on Pollination*, Tilburg, The Netherlands, pp. 445–451. Acta Horticulturae 288. Research Centre for Insect Pollination and Beekeeping, and International Society for Horticultural Science: Wageningen, The Netherlands.

Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., and Lukyanov, S. A. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* **23:** 1087–1088.

Simmons, M. P., Zhang, L.-B., Webb, C. T., and Müller, K. 2007. A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. *Molecular Phylogenetics and Evolution* **42:** 528–542.

Singh, A. K. 1990. Cytogenetics and evolution in the Cucurbitaceae. *In* Bates, D. M., Robinson, R. W., and Jeffrey, C. (Eds), *Biology and Utilization of the Cucurbitaceae*, pp. 10–28. Cornell University Press: Ithaca, New York, United States.

Small, R. L., Cronn, R. C., and Wendel, J. F. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17:** 145–170.

Smith, B. D. 1997. The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* **276:** 932–934.

———. 2001. Documenting plant domestication: the consilience of biological and archaeological approaches. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 1324–1326.

———. 2005. Reassessing Coxcatlan Cave and the early history of domesticated plants in Mesoamerica. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 9438–9445.

Smith, J. R., Carpten, J. D., Brownstein, M. J., Ghosh, S., Magnuson, V. L., Gilbert, D. A., Trent, J. M., and Collins, F. S. 1995. Approach to genotyping errors caused by nontemplated nucleotide addition by *Taq* DNA polymerase. *Genome Research* **5:** 312–317.

SoftGenetics. 2006. *GeneMarker Version 1.51 Operation Manual Version 1.5*. SoftGenetics: State College, Pennsylvania, United States.

Spooner, D. M., McLean, K., Ramsay, G., Waugh, R., and Bryan, G. J. 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 14694–14699.

Spriggs, M., and Anderson, A. J. 1993. Late colonization of East Polynesia. *Antiquity* **67:** 200–217.

Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M., and Powell, W. 2003. How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* **12:** 1339–1348.

Srisuwan, S., Sihachakr, D., and Siljak-Yakovlev, S. 2006. The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Plant Science* **171:** 424–433.

Staub, J. E., Danin-Poleg, Y., Fazio, G., Horejsi, T., Reis, N., and Katzir, N. 2000. Comparative analysis of cultivated melon groups (*Cucumis melo* L.) using random amplified polymorphic DNA and simple sequence repeat markers. *Euphytica* **115:** 225–241.

Storey, A. A., Ladefoged, T., and Matisoo-Smith, E. A. 2008a. Counting your chickens: density and distribution of chicken remains in archaeological sites of Oceania. *International Journal of Osteoarchaeology* **18:** 240–261.

Storey, A. A., Quiroz, D., Ramírez, J. M., Beavan-Athfield, N., Addison, D. J., Walter, R., Hunt, T., Athens, J. S., Huynen, L., and Matisoo-Smith, E. A. 2008b. Pre-Columbian chickens, dates, isotopes, and mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* **105:** E99.

Storey, A. A., Ramírez, J. M., Quiroz, D., Burley, D. V., Addison, D. J., Walter, R., Anderson, A. J., Hunt, T. L., Athens, J. S., Huynen, L., and Matisoo-Smith, E. A. 2007. Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 10335–10339.

Stryer, L. 1995. *Biochemistry*. 4th edition. W. H. Freeman and Company: New York, New York, United States.

Suggs, R. C. 1961. *The Archeology of Nuku Hiva, Marquesas Islands, French Polynesia*. Vol. 49, Part 1. Anthropological Papers of the American Museum of Natural History. American Museum of Natural History: New York, New York, United States.

Sullivan, J. P., Lavoué, S., Arnegard, M. E., and Hopkins, C. D. 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* **58:** 825–841.

Sunnucks, P. 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15:** 199–203.

Swofford, D. L. 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates: Sunderland, Massachusetts, United States.

Sykes, B., Leiboff, A., Lowbeer, J., Tetzner, S., and Richards, M. 1995. The origins of the Polynesians — an interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics* **57:** 1463–1475.

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105–1109.

Taylor, D. J., and Piel, W. H. 2004. An assessment of accuracy, error, and conflict with support values from genome-scale phylogenetic data. *Molecular Biology and Evolution* **21:** 1534–1537. [Erratum: *Mol. Biol. Evol.* 2004. **21:** 1812].

Terauchi, R., and Kahl, G. 2000. Rapid isolation of promoter sequences by TAIL-PCR: the 5′-flanking regions of *Pal* and *Pgi* genes from yams (*Dioscorea*). *Molecular and General Genetics* **263:** 554–560.

Towle, M. A. 1952. The pre-Columbian occurrence of *Lagenaria* seeds in coastal Peru. *Botanical Museum Leaflets — Harvard University* **15:** 171–184.

———. 1961. *The Ethnobotany of Pre-Columbian Peru*. Viking Fund Publications in Anthropology No. 30. Aldine: Chicago, Illinois, United States.

Trejaut, J. A., Kivisild, T., Loo, J. H., Lee, C. L., He, C. L., Hsu, C. J., Li, Z. Y., and Lin, M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biology* **3:** 1362–1372. [Erratum: *PLoS Biol.* 2005. **3:** 1838].

Tsugeki, R., Kochieva, E. Z., and Fedoroff, N. V. 1996. A transposon insertion in the *Arabidopsis SSR16* gene causes an embryo-defective lethal mutation. *The Plant Journal* **10:** 479–489.

van der Nest, M. A., Steenkamp, E. T., Wingfield, B. D., and Wingfield, M. J. 2000. Development of simple sequence repeat (SSR) markers in *Eucalyptus* from amplified inter-simple sequence repeats (ISSR). *Plant Breeding* **119:** 433–436.

van Eijk, M. J. T. 2006. *W410: Complexity Reduction of Polymorphic Sequences (CRoPS): A Novel Approach For High Througput [sic] Polymorphism Discovery* [electronic source]. http://www.intl-pag.org/14/abstracts/PAG14_W410.html. Last updated: 2006. Accessed: 15/11/2007. XIV International Plant & Animal Genome Conference, 14–18 January 2006, Town & Country Convention Center, San Diego, California, United States.

Vavilov, N. I. 1951. *The Origin, Variation, Immunity and Breeding of Cultivated Plants*. Chronica Botanica, Vol. 13, Number 1/6. Chronica Botanica Company: Waltham, Massachusetts, United States.

Vekemans, X., Beauwens, T., Lemaire, M., and Roldán-Ruiz, I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* **11:** 139–151.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23:** 4407–4414.

Walters, T. W. 1989. Historical overview on domesticated plants in China with special emphasis on the Cucurbitaceae. *Economic Botany* **43:** 297–313.

Walters, T. W., and Decker-Walters, D. S. 1989. Systematic re-evaluation of *Benincasa hispida* (Cucurbitaceae). *Economic Botany* **43:** 274–278.

Ward, R. G., and Brookfield, M. 1992. The dispersal of the coconut: did it float or was it carried to Panama? *Journal of Biogeography* **19:** 467–480.

Wayne, R. K., Leonard, J. A., and Vilà, C. 2006. Genetic analysis of dog domestication. *In* Zeder, M. A., Bradley, D. G., Emshwiller, E., and Smith, B. D. (Eds), *Documenting Domestication: New Genetic and Archaeological Paradigms*, pp. 279–293. University of California Press: Berkeley, California, United States.

Weisler, M. I., and Green, R. C. 2008. The many sides of Polynesian archaeology in reference to the colonization process in southeast Polynesia. *Rapa Nui Journal* **22:** 85–87.

Weissensteiner, T., and Lanchbury, J. S. 1996. Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. *BioTechniques* **21:** 1102–1108.

Welsh, J., and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18:** 7213–7218.

Whistler, W. A. 1988. Ethnobotany of Tokelau: the plants, their Tokelau names, and their uses. *Economic Botany* **42:** 155–176.

———. 1990. The other Polynesian gourd. *Pacific Science* **44:** 115–122.

———. 1991. Polynesian plant introductions. *In* Cox, P. A. and Banack, S. A. (Eds), *Islands, Plants, and Polynesians: An Introduction to Polynesian Ethnobotany*, pp. 41–66. Dioscorides Press: Portland, Oregon, United States.

Whitaker, T. W. 1948. *Lagenaria*: a pre-Columbian cultivated plant in the Americas. *Southwestern Journal of Anthropology* **4:** 49–68.

———. 1971. Endemism and pre-Columbian migration of the bottle gourd, *Lagenaria siceraria* (Mol.) Standl. *In* Riley, C. L., Kelley, J. C., Pennington, C. W., and Rands, R. L. (Eds), *Man across the Sea: Problems of Pre-Columbian Contacts*, pp. 320–327. University of Texas Press: Austin, Texas, United States.

———. 1983. Cucurbits in Andean prehistory. *American Antiquity* **48:** 576–585.

Whitaker, T. W., and Bird, J. B. 1949. *Identification and Significance of the Cucurbit Materials from Huaca Prieta, Peru*. American Museum Novitates 1426. American Museum of Natural History: New York, New York, United States.

Whitaker, T. W., and Carter, G. F. 1954. Oceanic drift of gourds — experimental observations. *American Journal of Botany* **41:** 697–700.

———. 1961. A note on the longevity of seed of *Lagenaria siceraria* (Mol.) Standl. after floating in sea water. *Bulletin of the Torrey Botanical Club* **88:** 104–106.

Whitaker, T. W., and Davis, G. N. 1962. *Cucurbits: Botany, Cultivation, and Utilization*. World Crops Books. Leonard Hill: London, United Kingdom.

Whitaker, T. W., and Cutler, H. C. 1971. Prehistoric cucurbits from the valley of Oaxaca. *Economic Botany* **25:** 123–127.

Whitaker, T. W., Cutler, H. C., and MacNeish, R. S. 1957. Cucurbit materials from three caves near Ocampo, Tamaulipas. *American Antiquity* **22:** 352–358.

White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J. (Eds), *PCR Protocols: A Guide to Methods and Applications*, pp. 315–322. Academic Press: San Diego, California, United States.

Wilkins-Ellert, M. 2003. The discovery of wild bottle gourd (*Lagenaria siceraria*). *The Cucurbit Network News* **10:** 1, 7.

Williams, A., and Walton, T. 2003. *Early Landuse Patterns in the Lake Taupo Area*. Science for Conservation 222. New Zealand Department of Conservation: Wellington, New Zealand.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18:** 6531–6535.

Wilmshurst, J. M., Anderson, A. J., Higham, T. F. G., and Worthy, T. H. 2008. Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 7676–7680.

Wolfe, A. D., Xiang, Q.-Y., and Kephart, S. R. 1998a. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proceedings of the National Academy of Sciences of the United States of America* **95:** 5112–5115.

———. 1998b. Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Molecular Ecology* **7:** 1107–1125.

Wong, A., Forbes, M. R., and Smith, M. L. 2001. Characterization of AFLP markers in damselflies: prevalence of codominant markers and implications for population genetic applications. *Genome* **44:** 677–684.

Woodhead, M., Russell, J., Squirrell, J., Hollingsworth, P. M., Mackenzie, K., Gibby, M., and Powell, W. 2005. Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Molecular Ecology* **14:** 1681–1695.

Xu, M., Huaracha, E., and Korban, S. S. 2001. Development of sequence-characterized amplified regions (SCARs) from amplified fragment length polymorphism (AFLP) markers tightly linked to the *Vf* gene in apple. *Genome* **44:** 63–70.

Yacovleff, E., and Herrera, F. L. 1934–1935. El Mundo Vegetal de los Antiguos Peruanos [Spanish]. *Revista del Museo Nacional — Lima, Peru* **3:** 243–322 [Part 1]; **4:** 31–102 [Part 2].

Yang, H., Sweetingham, M. W., Cowling, W. A., and Smith, P. M. C. 2001. DNA fingerprinting based on microsatellite-anchored fragment length polymorphisms, and isolation of sequence-specific PCR markers in lupin (*Lupinus angustifolius* L.). *Molecular Breeding* **7:** 203–209.

Yen, D. E. 1963. The New Zealand kumara or sweet potato. *Economic Botany* **17:** 31–45.

———. 1973. Ethnobotany from the voyages of Mendaña and Quiros in the Pacific. *World Archaeology* **5:** 32–43.

———. 1974. *The Sweet Potato and Oceania: An Essay in Ethnobotany*. Bernice P. Bishop Museum Bulletin 236. Bishop Museum Press: Honolulu, Hawai‘i, United States.

———. 1977. Hoabinhian horticulture? The evidence and the questions from northwest Thailand. *In* Allen, J., Golson, J., and Jones, R. (Eds), *Sunda and Sahul: Prehistoric Studies in Southeast Asia, Melanesia and Australia*, pp. 567–599. Academic Press: London, United Kingdom.

———. 1991. Polynesian cultigens and cultivars: the questions of origin. *In* Cox, P. A. and Banack, S. A. (Eds), *Islands, Plants, and Polynesians: An Introduction to Polynesian Ethnobotany*, pp. 67–95. Dioscorides Press: Portland, Oregon, United States.

———. 1995. The development of Sahul agriculture with Australia as bystander. *Antiquity* **69:** 831–847.

———. 1996. Melanesian arboriculture: historical perspectives with emphasis on the genus *Canarium*. *In* Stevens, M. L., Bourke, R. M., and Evans, B. R. (Eds), *South Pacific Indigenous Nuts*, Le Lagon Resort, Port Vila, Vanuatu, 31 October–4 November 1994, pp. 36–44. Australian Centre for International Agricultural Research: Canberra, A.C.T., Australia.

———. 1998. Subsistence to commerce in Pacific agriculture: some four thousand years of plant exchange. *In* Prendergast, H. D. V., Etkin, N. L., Harris, D. R., and Houghton, P. J. (Eds), *Plants for Food and Medicine*, pp. 161–183. Royal Botanic Gardens Press: Kew, Richmond, Surrey, United Kingdom.

———. 2005. Reflection, refraction and recombination. *In* Ballard, C., Brown, P., Bourke, R. M., and Harwood, T. (Eds), *The Sweet Potato in Oceania: A Reappraisal*, Ethnology Monographs 19/Oceania Monograph 56, pp. 181–187. Ethnology, University of Pittsburgh: Pittsburgh, Pennsylvania, United States/Oceania Publications, University of Sydney: Sydney, New South Wales, Australia.

Yen, D. E., and Head, J. 1993. Kumara remains in Pit O at P5/288. *In* Sutton, D. G. (Ed.), *The archaeology of the peripheral pa at Pouerua, Northland, New Zealand*, pp. 56–64. Auckland University Press: Auckland, New Zealand.

Yetisir, H., and Sari, N. 2003. Effect of different rootstock on plant growth, yield and quality of watermelon. *Australian Journal of Experimental Agriculture* **43:** 1269–1274.

Zane, L., Bargelloni, L., and Patarnello, T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* **11:** 1–16.

Zeder, M. A. 2006. Central questions in the domestication of plants and animals. *Evolutionary Anthropology* **15:** 105–117.

Zeder, M. A., Emshwiller, E., Smith, B. D., and Bradley, D. G. 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends in Genetics* **22:** 139–155.

Zerega, N. J. C., Ragone, D., and Motley, T. J. 2004. Complex origins of breadfruit (*Artocarpus altilis*, Moraceae): implications for human migrations in Oceania. *American Journal of Botany* **91:** 760–766.

———. 2005. Systematics and species limits of breadfruit (*Artocarpus*, Moraceae). *Systematic Botany* **30:** 603–615.

Zhang, D.-X., and Hewitt, G. M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12:** 563–584.

Zhang, D. P., Rossel, G., Kriegner, A., and Hijmans, R. 2004. AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: its implications on the dispersal of the crop. *Genetic Resources and Crop Evolution* **51:** 115–120.

Zhang, D. P., Ghislain, M., Huamán, Z., Cervantes, J. C., and Carey, E. E. 1998. AFLP assessment of sweetpotato genetic diversity in four tropical American regions. *In CIP Program Report 1997–1998*, pp. 303–310. International Potato Center: Lima, Peru.

Zhang, D. P., Cervantes, J., Huamán, Z., Carey, E., and Ghislain, M. 2000a. Assessing genetic diversity of sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars from tropical America using AFLP. *Genetic Resources and Crop Evolution* **47:** 659–665.

Zhang, D. P., Carbajulca, D., Ojeda, L., Rossel, G., Milla, S., Herrera, C., and Ghislain, M. 2000b. Microsatellite analysis of genetic diversity in sweetpotato varieties from Latin America. *In CIP Program Report 1999–2000*, pp. 295–301: Lima, Peru.

Zhang, Y., and Stommel, J. R. 2001. Development of SCAR and CAPS markers linked to the *Beta* gene in tomato. *Crop Science* **41:** 1602–1608.

Zietkiewicz, E., Rafalski, A., and Labuda, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20:** 176–183.

Zink, A., and Nerlich, A. G. 2003. Molecular analyses of the “Pharaos:” Feasibility of molecular studies in ancient Egyptian material. *American Journal of Physical Anthropology* **121:** 109–111.

Zizumbo-Villarreal, D., and Quero, H. J. 1998. Re-evaluation of early observations on coconut in the New World. *Economic Botany* **52:** 68–77.

APPENDICES

APPENDIX 1: MARKERS SCREENED FOR BOTTLE GOURD

Table A1.1
Chloroplast Markers Screened for Use in Bottle Gourd

Plant taxa for which markers have previously been used	Reference	Marker	Successfully Amplified?	PCR conditions[a,b]	Size (bp)	No. cultivars sequenced	Poly-morphic?
All plants	Olmstead *et al.* (1992)	*rbc*L	Yes	50ºC; 1.5 mM $MgCl_2$	1400	4	No
All plants	Taberlet *et al.* (1991)	*trn*L–F	Yes	60ºC; 1.5 mM $MgCl_2$	850	5	No
Angiosperms	Lee & Wen (2004)	*trn*C–D	Yes	56ºC; 1.5 mM $MgCl_2$	2800	5	Yes
Angiosperms (incl. Cucurbitaceae)	Chung & Staub (2003), Chung *et al.* (2003)	ccSSR01	Yes	56ºC; 1.5 mM $MgCl_2$	175	11	No
		ccSSR02	No				
		ccSSR03	No				
		ccSSR04	No				
		ccSSR05	No				
		ccSSR06	Yes	64ºC; 1.5 mM $MgCl_2$	300	11	No
		ccSSR07	Yes	64ºC; 1.5 mM $MgCl_2$	350	11	No
		ccSSR08	Yes	64ºC; 1.5 mM $MgCl_2$	250	11	No
		ccSSR09	No				
		ccSSR10	No				
		ccSSR11	Yes	56ºC; 1.5 mM $MgCl_2$	165	11	No
		ccSSR12	Yes	56ºC; 1.5 mM $MgCl_2$	250	10	No
		ccSSR13	Yes	54ºC; 1.5 mM $MgCl_2$	300	11	No
		ccSSR15	Yes	56ºC; 1.5 mM $MgCl_2$	265	11	No
		ccSSR16	Yes	56ºC; 1.5 mM $MgCl_2$	225	11	No
		ccSSR17	Yes	56ºC; 1.5 mM $MgCl_2$	250	11	No
		ccSSR18	Yes	56ºC; 1.5 mM $MgCl_2$	275	11	No
		ccSSR20	Yes	64ºC; 1.5 mM $MgCl_2$	350	7	Yes
		ccSSR21	Yes	62ºC; 1.5 mM $MgCl_2$	280	11	No
		ccSSR22	Yes	56ºC; 1.5 mM $MgCl_2$	190	11	No

[a] Other PCR conditions were standard; extension time = 40 s kb^{-1}.
[b] The extension temperature for *trn*C–D was lowered to 68°C.

Table A1.2
Nuclear Markers Screened for Use in Bottle Gourd

Plant taxa for which markers have previously been used	Reference	Marker	Successfully Amplified?	PCR conditions[a]	Size (bp)	No. cultivars sequenced	Poly-morphic?
All plants	Baldwin (1992), Hershkovitz & Zimmer (1996)	ITS	Yes	48ºC; 1.5 mM $MgCl_2$	700	4	No
Watermelon (*Citrullus lanatus*)	Jarret *et al.* (1997)	CL1-06	No				
		CL1-12	No				
		CL1-20	Yes	56ºC; 1.5 mM $MgCl_2$	150	11	No
		CL1-21	No				
		CL2-23	Yes	56ºC; 1.5 mM $MgCl_2$	175	8	No
		CL2-61	No				
		CL2-140	No				
Rockmelon (*Cucumis melo*)	Katzir *et al.* (1996)	CMTC13	No				
		CMAG59	No				
Cucumber (*Cucumis sativus*)	Katzir *et al.* (1996)	CSLHCPA	No				
		CSHPRAG	Yes	56ºC; 2.5 mM $MgCl_2$	50	None	Unknown
Rockmelon (*Cucumis melo*)	Chiba *et al.* (2003)	CMMS01-3	No				
		CMMS02-3	No				
		CMMS03-1	No				
		CMMS04-3	No				
		CMMS12-6	No				
		CMMS15-4	No				
		CMMS27-1	No				
		CMMS30-3	Yes	60ºC; 1.5 mM $MgCl_2$	300	7	Yes
Rockmelon (*Cucumis melo*)	Ritschel *et al.* (2004)	CMBR18	No				
		CMBR23	No				
		CMBR56	No				

[a] Other PCR conditions were standard; extension time = 40 s kb^{-1}.

APPENDIX 2: BOTTLE GOURD PCR AND SEQUENCING PRIMERS

Table A2.1
pGEM®-T Easy PCR Primers

Primer Name	Sequence (5′–3′)
M13F	CCCAGTCACGACGTTGTAAAACG
M13R	AGCGGATAACAATTTCACACAGG
T7F	GTAATACGACTCACTATAGGG

Table A2.2
TAIL PCR Primers

Type	Marker	Primer Name	Sequence (5′–3′)[a]
AD (Arbitrary Degenerate)[b]	–	TAIL-AD1	NGTCGASWGANAWGAA
	–	TAIL-AD2	GTNCGASWCANAWGTT
		TAIL-AD3	WGTGNAGWANCANAGA
SP (Specific)	BR01_19	SP19-1L	AAAAGTTCGCCCCCAGCCG
		SP19-2L	GCCGAAAATGCCAAATCACCA
		SP19-3L	TCGGATGAAACTACTGAAATGA
		SP19-1R	GTGAGGAAAGGAAGAGAGAG
		SP19-2R	CGACTTTTCCGTGTGTGCGA
		SP19-3R	TGGTGATTTGGCATTTTCGGC
	BOP19_27	SP27-1L	GTCTGACTTCATCATTTTTATTTA
		SP27-2L	AACTATTTCAAAACACTTTAGAAG
		SP27-3L	TAAACCAAACACACCTTGAATAA
		SP27-1R	TGAAGTTTTCGGGTAGAGGG
		SP27-2R	TTATTCAAGGTGTGTTTGGTTTA
		SP27-3R	CTTCTAAAGTGTTTTGAAATAGTT

Table continued next page...

Table continued...

Type	Marker	Primer Name	Sequence (5′–3′)[a]
SP (Specific)	BOP19_31	SP31-1L	TTGTATCCATTCAAAAGTTTCC
		SP31-2L	CAATAACAAAGAGTGGACTGG
		SP31-3L	TGATTACAGGAAACGAGGAGT
		SP31-1R	CCAGTCCACTCTTTGTTATTG
		SP31-2R	GGAAACTTTTGAATGGATACAA
		SP31-3R	TTTTGTTACTCTGCCACGGTT
	BOP19_35	SP35-1L	TGGTATAAAAGTAAGGACTCTA
		SP35-2L	GTCTTTTGAAGTAGCGTCGG
		SP35-3L	GGGCTTGTTTTATCTACTTGTT
		SP35-1R	TAGAGTCCTTACTTTTATACCA
		SP35-2R	TTCTCGTTTTCCTTTCTCCATT
		SP35-3R	CCTCCCTCGGTCATCTTCTA
	MR06_24	SP24-1L	TTACTGTCTGCTCCTTCAAATC
		SP24-2L	AATAGAAATAAAGAGAGACCGA
		SP24-3L	CTGCTTCTATGGCTTTCTTCTT
		SP24-1R	TCGGGGAACTGGAGATTGTT
		SP24-2R	TCAAGGGAGAGAGTAAATGTTA
		SP24-3R	GAGGAGAAGGATAAAAACATAC

[a] N = A, G, C and T; W = A and T; S = G and C.
[b] AD primers from Okamoto & Hirochika (2000).

Table A2.3
Chloroplast and Nuclear PCR and Sequencing Primers

Location	Marker	Primer Name	Sequence (5′–3′)	Use[a]
Chloroplast	***trn*****C–*****trn*****D**[b]	trnC_F[b]	CCAGTTCAAATCTGGGTGTC	P
		trnD_R[b]	GGGATTGTAGTTCAATTGGT	P
		psbM_2R[b]	TTCTTGCATTTATTGCTACTGC	S
		psbM_3F	TGCTTTTCATTTTTCTTATCTTC	S
	trn**S–*****trn*****G**	ccSSR02F[c]	AATCCTGGACGTGAAGAATAA	P
		trnG_R	AAACTATATCCGCTACAATGC	P, S
Nuclear	**BR01_19**	BR01_19_L	CCCTCTTCACCATCTTCTTC	P
		BR01_19_R	GAAATATCGTGCCTGTAAAATAT	P, S
	BOP19_27	BOP19_27_L	CAGATGTTTTGGTTTGGGGAT	P
		BOP19_27_R	CTCACTCCTTTTCCATACCAT	P, S
	BOP19_31	BOP19_31_L	GATAGGGAAGAAAAATAGAAAAG	P, S
		BOP19_31_R	CGTGAAGAAACAAAAAGGAAC	P
	BOP19_35	BOP19_35_L	GAGTGAGATGAACAAAGAAAGA	P, S
		BOP19_35_R	TCCAGACAAACCAAGAAACCA	P
	MR06_24	MR06_24_L	CCATTTTGACAGTATGCCATCT	P, S
		MR06_24_R	GTGCTGCTGCTTCAGTTTTCA	P

[a] P = PCR primer; S = sequencing primer.
[b] From Lee & Wen (2004).
[c] From Chung & Staub (2003).

APPENDIX 3: BOTTLE GOURD ISSR SEQUENCES

Underlining denotes ISSR primer sequence. Primers are from UBC Primer Set #9.

ISSR_19_CON (357 bp)

ISSR Primer: #808 $(AG)_8C$

Derived Marker: BR01_19

AGAGAGAGAGAGAGAGCGAGACGAGAGTGATTGTGAGTGAGGAAAGGAAGAGAGAGCGAGATCG
ACTTTTCCGTGTGTGCGAGTTCATTTCAGTAGTTTCATCCGAACTTGATGTCRGCAAATGGTGA
TTTGGCATTTTCGGCCCAATTTTTTAGTTATCCGTACAAATTTCCCTTATTTGAAGTTGAAGTT
GAAGTTGGTGTTAAGATTTGCAGATTGTTTTAGCCCCAACACATTTGGCTATAATGATGGAATA
TGCGGCTGGGGGCGAACTTTTCGAAAGAATTTGCAAGGCTGGACGCTTTAGTGAAGATGAAGTA
TGCTATCTACTCTCTCCTTTGCTCTCTCTCTCTCT

ISSR_27_CON (367 bp)

ISSR Primer: #823 $(TC)_8C$

Derived Marker: BOP19_27

TCTCTCTCTCTCTCTCCCTATGTAAAAGCTTGTGAAGTTTTCGGGTAGAGGGTCCTAGTCTAGA
GGGTCCATGAAGTCCTTCAGTTTTCAATATAATTTCCAAGGAATAGRAGAAAAAACAGAGAGGG
AGCTTAATGAAGGCAAGTTATTCAAGGTGTGTTTGGTTTAACTATTTAAGTATTTAATTTTGAA
AATAAATCAGTGGTTAAAATCACTCTAAACAACTTTTAAACACTTCTAAAGTGTTTTGAAATAG
TTTTTATCAAAAGAGTTTAAATAAAAATGATGAAGTCAGACTGGAGGGACCTCTTAGAATTTTC
GGTTAATTTCCAAGAAATGGAAGAAAAATCGGAGAGAGAGAGAGAGA

ISSR_31_CON (520 bp)

ISSR Primer: #846 $(CA)_8RT$

Derived Marker: BOP19_31

CACACACACACACACAGTCATCTTGCAGGGCTTGCTGAATCGTCTACGAGCTTTTTGCCATTAT
CATCTCTCTGAAATCTCCTTGATTTCCCCAATTACCAGGTACGTCTCTGAATTGGACTCCTCGT
TTCCTGTAATCAATTGCCGAATCGATTCTGGCCTTCATTGAGCCCAGTCCACTCTTTGTTATTG
GGATTTCCACAAACTCTTTTTTCTCTTAGGTTGTTTGGAAACTTTTGAATGGATACAAATAATT
GGAGACCTACTCAAGGTGGAGAACCCGGAATCGAGGCCGGGGATTGGAGGTCTCAATTGCAGCC
CGATTCTCGACAACGAATTGTCAACAAAATGTATGATTTCTTACATTTTTCTTGCAGCCCGATT
CTTGTAGGTTGTTTTGTTACTCTGCCACGGTTTTGTGAATTAAAACTATGAACAGCACAACTTT
GTAGATATAGGACTGTCTTCGACCTTTTTCTTATTTATTCTTAACATTTTCCTAGATTGTGTGT
GTGTGTGT

ISSR_35_CON (656 bp)

ISSR Primer: #812 $(GA)_8A$

Derived Marker: BOP19_35

GAGAGAGAGAGAGAGAATAATTAGATTCTATTGTGGTGTGTATTACAAAGGATCATATACCACT
ATTTATAGGACATAAGATAGTGTAGGTTACAAGATTGAATTCAATGGGGGTAAAATTCAAGGAG
AGCTTTATGATATTTTGTAACCTATATAGATTATGGATATCTACATTTATAATAAATGAAAAAT
ATTCAATTAATTATTTAAATTTCACCCTCAAAATTTAGGTTCTACAAAATTGTGACCACGTGGA
TCTTTTTAACAAGTAGATAAAACAAGCCCTTCAAGCATTCACGGCTTCGCCTGACCCGACGCTA
CTTCAAAAGACCGTACGCAATTATGCGATAACGTAACAATAATTATAATTTTGCAATAGAGTCC
TTACTTTTATACCATAATTACAATTCGGCCGCCTTATTTCATGGTAATTACATTCCGTTCTCAC
GATATAAATCTATCCCGAGCAGCAATCCAATTCTCGTTTTCCTTTCTCCATTTCTCTTTCTCTG
TTCCTTTCATTTGGGCCTCCCTCGGTCATCTTCTAGCGACGAACTCAAGCGACTTTTACAGACT
CATTAAGCTTCCTCCGATCCTCACTCGTCTTGGTTTATGAGGTTTCGATCTCTTCCTTTTCTTT
TCTCTCTCTCTCTCTC

ISSR_24_CON (548 bp)

ISSR Primer: #823 $(TC)_8C$

Derived Marker: MR06_24

TCTCTCTCTCTCTCTCCTTCTCACTTCTCTTCTTTACTCTCTTTCTTTGGTTTCTAGCAAAAGA
TGGTATTGCAGGTTCTAACAATGGCTGATGACAAGGCAAAGAAGAAAGCCATAGAAGCAGGTGC
AGAAATATTCGGTCTCTCTTTATTTCTATTAGGTTTGGGGATTTTAATGGGGTGTGACATTTTG
ATCAATAGGGATGGATTGCCTCCATTGCAGCTGATTTGAAGGAGCAGACAGTAAAAGCTATGGG
TTCTACAGTTGCAATGGTGAAGAAGATGAAGAACGTCGGGGAACTGGAGATTGTTTTCTTTCTA
GTACAACAAGAACCTCCGACCTCAAGGGAGAGAGTAAATGTTAATTACTGATGAGCTATGGGAA
AATAGAGATGGTTTCAGTTGGTCCAAGGAGGAGAAGGATAAAAACATACAAAAGAACAGAGATG
ATAAAACATGACTTCTATCAAACCAGACCCAACTAAATAGGAAAGAAGAAAGGGAGAAAGTTTT
AGGGAGGAACTAAAGTTTTGGAGAGAGAGAGAGAGA

APPENDIX 4: YEN SWEET POTATO ACCESSIONS

Table A4.1
Yen Sweet Potato Collection (June 2004)

Accession No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1010976	172715	Y-004	Thailand	Central Plateau	–
1005792	91383	Y-005	Thailand	Central Plateau	–
1005793	171719	Y-006	Thailand	Central Plateau	–
1005794	172716	Y-010	Thailand	Central Plateau	–
1005795	171720	Y-015	China	South	–
1005796	171721	Y-017	China	South	–
1005798	171723	Y-020	China	South	–
1005799	171724	Y-021	China	South	–
1005800	171725	Y-022	China	South	–
1005802	171727	Y-027	China	Hong Kong	–
1005804	171729	Y-034	Japan	Okinawa	Ishigaki
1005805	91384	Y-036	Japan	Okinawa	Ishigaki
1005806	171730	Y-037	Japan	Okinawa	Ishigaki
1005807	91385	Y-039	Japan	Okinawa	Ishigaki
1010977	172717	Y-040	Japan	Okinawa	Kohama
1005808	171731	Y-041	Japan	Okinawa	Kohama
1005809	171732	Y-046	Japan	Okinawa	Kohama
1005810	171733	Y-047	Japan	Okinawa	Kohama
1010978	172718	Y-050	Philippines	Ifugao	Bayninan-Poitaan
1010979	172719	Y-053	Philippines	Ifugao	Bayninan-Poitaan
1010980	172720	Y-054	Philippines	Ifugao	Bayninan-Poitaan
1010981	172721	Y-057	Philippines	Ifugao	Bayninan-Poitaan
1005811	171734	Y-063	Philippines	Ifugao	Bublei-Liwang
1005812	171735	Y-064	Philippines	Ifugao	Bublei-Liwang
1005813	91386	Y-065	Philippines	Ifugao	Bublei-Liwang
1005814	172724	Y-066	Philippines	Ifugao	Bublei-Liwang
1005815	172725	Y-068	Philippines	Ifugao	Bublei-Liwang
1005816	91387	Y-069	Philippines	Ifugao	Bublei-Liwang

Original ID	Cultivation	Local Cultivar Name	Notes
TH-1	Pot	–	–
5	Pot	–	–
6	Field	–	–
10	Pot	–	–
15	Field	–	Collected by W.R. Geddes (from 1965 list)
17	Field	–	Collected by W.R. Geddes (from 1965 list)
20	Field	–	Collected by W.R. Geddes (from 1965 list)
21	Field	–	Collected by W.R. Geddes (from 1965 list)
22	Field	–	Collected by W.R. Geddes (from 1965 list)
27	Pot	–	–
34	Pot	*telmā*	–
36	Field	*hyakugo*	Oldest variety in Ishigaki (from 1965 list)
37	Field & Pot	*sekai-ichizo*	–
39	Pot	*haya-innayo*	–
JAP9	Pot	*tumai-kurū*	–
41	Field & Pot	*bōko*	–
46	Field & Pot	*sekai-ichigo*	–
47	Field	*baka*	–
1PH	Pot	*buqagen*	–
4PH	Pot	*magyano*	–
5PH	Pot	*dumāqin*	–
8PH	Pot	*kayqūhin*	–
63	Field & Pot	*ballaan*	"American" from Kiangau (from 1965 list)
64	Pot	*lota-an*	from Balik (from 1965 list)
65	Field & Pot	*kawitan*	from Henyuong (from 1965 list)
66	Pot	*baqnat/pukin*	Old type (from 1965 list)
68	Pot	*ballaan*	"the real Ballaan" (from 1965 list)
69	Field & Pot	UNKNOWN	–

Table continued next page...

Table continued…

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1005817	171736	Y-070	Philippines	Ifugao	Liwang-Mayaoyao
1010984	172726	Y-071	Philippines	Ifugao	Liwang-Mayaoyao
1005818	171737	Y-073	Philippines	Ifugao	Liwang-Mayaoyao
1005819	91388	Y-074	Philippines	Ifugao	Liwang-Mayaoyao
1005821	91390	Y-079	Philippines	Ifugao	Liwang-Mayaoyao
1005822	171738	Y-081	Philippines	Ifugao	Liwang-Mayaoyao
1005823	91391	Y-085	Philippines	Ifugao	Liwang-Mayaoyao
1010988	172730	Y-086	Philippines	Ifugao	Liwang-Mayaoyao
1010989	172731	Y-088	Philippines	Ifugao	Liwang-Mayaoyao
1005824	171739	Y-090	Philippines	Ifugao	Liwang-Mayaoyao
1010990	172732	Y-092	Philippines	Ifugao	Liwang-Mayaoyao
1005825	171740	Y-093	Philippines	Ifugao	Liwang-Mayaoyao
1005826	171741	Y-095	Philippines	Ifugao	Liwang-Mayaoyao
1005827	171742	Y-098	Philippines	Ifugao	Liwang-Mayaoyao
1005828	91392	Y-102	Philippines	Ifugao	W. Bagninan
1010991	172733	Y-103	Philippines	Ifugao	Gohan
1005829	91393	Y-106	Philippines	Ifugao	Gohan
1005830	171743	Y-112	Philippines	Ifugao	Hapo
1010993	172735	Y-115	Philippines	Ifugao	Lūgu
1010994	172736	Y-118	Philippines	Ifugao	Lūgu
1005831	171744	Y-119	Philippines	Ifugao	Lūgu
1005832	171745	Y-121	Philippines	Ifugao	Lūgu
1010996	172738	Y-125	Philippines	Ifugao	Lūgu
1005834	171746	Y-126	Philippines	Ifugao	Lūgu
1010997	172739	Y-127	Philippines	Ifugao	Lūgu
1010998	172740	Y-128	Philippines	Ifugao	Lūgu
1010999	172741	Y-129	Philippines	Ifugao	Piwoŋ, Lagawi
1005835	171747	Y-130	Philippines	Ifugao	Piwoŋ, Lagawi
1005836	171748	Y-135	Philippines	Ifugao	Piwoŋ, Lagawi
1005837	91394	Y-136	Philippines	Ifugao	Piwoŋ, Lagawi
1005838	171749	Y-137	Philippines	Ifugao	Piwoŋ, Lagawi
1005841	171750	Y-141	Philippines	Ifugao	Piwoŋ, Lagawi
1005842	171751	Y-142	Philippines	Ifugao	Piwoŋ, Lagawi

Original ID	Cultivation	Local Cultivar Name	Notes
70	Field & Pot	*pūkin*	–
29PH	Pot	*luglughi*	–
73	Field	*qūdanan*	–
74	Field & Pot	*qinhud*	–
79	Field & Pot	UNKNOWN	–
81	Field & Pot	*koyyōkoy 2*	–
85	Field & Pot	*kawitan*	–
48PH	Pot	*qiggoltud*	–
52PH	Pot	*pattuki*	–
90	Field	*qimbāqao*	–
57PH	Pot	*quhohōba*	–
93	Field	*tuttōlid*	–
95	Pot	*qimbuqdag*	–
98	Field & Pot	*qubūwan*	–
102	Field	*qaggetet*	–
132PH	Pot	*dumāqan*	–
106	Field & Pot	*bannāqo*	–
112	Pot	*gohaŋ*	–
108PH	Pot	*dutdūtan*	–
111PH	Pot	*baknat*	–
119	Field & Pot	*kawitan qad hiŋyon*	Recent intro. (5 years) (from 1965 list)
121	Field	*yāwa*	–
123PH	Pot	*geŋhew*	–
126	Pot	*qahqahin*	–
126PH	Pot	*daqdaqane*	–
127PH	Pot	*qiyambābag*	–
71PH	Pot	*lumnin*	–
130	Field & Pot	*hiŋyon*	–
135	Field	*qontig*	–
136	Field & Pot	*gulguldin*	–
137	Field & Pot	*gimpāyan*	–
141	Field & Pot	*hobūl*	–
142	Field	*bayāni*	–

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1005843	91395	Y-145	Philippines	Ifugao	Kiangan
1005844	91396	Y-146	Philippines	Ifugao	Kiangan
1005845	91397	Y-147	Philippines	Ifugao	Kiangan
1005847	91399	Y-150	Philippines	Bontoc	Bayyu
1011000	172744	Y-151	Philippines	Bontoc	Bayyu
1005848	171752	Y-156	Philippines	Sagada	–
1005849	171753	Y-157	Philippines	Sagada	–
1005850	171754	Y-158	Philippines	Sagada	–
1005851	171755	Y-159	Philippines	Sagada	–
1011001	172745	Y-160	Philippines	Sagada	–
1011002	172746	Y-162	Philippines	Alab	–
1005852	91400	Y-164	Philippines	Alab	–
1005853	91401	Y-166	Philippines	Alab	–
1011004	172748	Y-171	Philippines	Manila	–
1011005	174256	Y-172	Philippines	Basilan	–
1005855	171757	Y-173	Philippines	Ilocos N.	–
1005856	171758	Y-174	Philippines	Ilocos N.	–
1011007	174258	Y-177	Philippines	Ilocos	–
1005858	171759	Y-179	Philippines	Ilocos	–
1005859	91402	Y-180	Philippines	Ifugao	Gohaŋ
1005860	171760	Y-181	Philippines	Ilocos	–
1005861	171761	Y-182	Unknown	–	–
1011009	174259	Y-185	East Timor	Dili	–
1011010	174260	Y-188	East Timor	Liquica	–
1011011	174261	Y-189	East Timor	Liquica	–
1011012	174262	Y-191	East Timor	Liquica	–
1011014	174264	Y-193	East Timor	Liquica	–
?	?	Y-196	East Timor	Dili	–
1011017	174267	Y-197	East Timor	Dili	–
1011019	174307	Y-210	Indonesia	Irian Jaya	Wissellakes
1011020	174269	Y-211	Indonesia	Irian Jaya	Wissellakes
1005862	171762	Y-212	Indonesia	Irian Jaya	Wissellakes
1005863	171763	Y-213	Indonesia	Irian Jaya	Merauke

Original ID	Cultivation	Local Cultivar Name	Notes
145	Pot	*qitbuy*	–
146	Field & Pot	*talindak*	–
147	Field	*taŋŋāya*	–
150	Pot	*qinaŋay*	–
141PH	Pot	*pattuki*	Old type (from 1965 list)
156	Field & Pot	*qabbaw*	Good for storing (from 1965 list)
157	Field	*qīlin*	New from Tadian (from 1965 list)
158	Field & Pot	*kinagayqan*	New from Benguet (from 1965 list)
159	Field & Pot	*qallusīqis*	Old type (from 1965 list)
162PH	Pot	*balo*	Newer type (from 1965 list)
166PH	Pot	*kinagāyan*	–
164	Field & Pot	*butāgan*	–
166	Field & Pot	*budbuduwan*	–
M5	Pot	UNKNOWN	–
178PH	Pot	*dabaw*	–
173	Pot	*immubi*	–
174	Field & Pot	*baisiŋ*	–
261PH	Pot	*illagan*	–
179	Field & Pot	*ubi*	–
180	Pot	*gilayan*	–
181	Field & Pot	UNKNOWN	–
182	Field & Pot	–	–
601	Pot	*mautema*	–
604	Pot	–	–
605	Pot	–	–
607	Pot	–	–
610	Pot	–	–
613	Pot	–	–
614	Pot	–	–
DNG 1	Pot	*notta dere*	from D. Ag. Hollandia, 1957 (from 1965 list)
DNG2	Pot	*notta molodoeroe*	from D. Ag. Hollandia, 1957 (from 1965 list)
212	Field & Pot	*notta waikio*	from D. Ag. Hollandia, 1957 (from 1965 list)
213	Field	–	from D. Ag. Hollandia, 1957 (from 1965 list)

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1005864	171764	Y-214-1	Indonesia	Irian Jaya	Merauke
1005865	171765	Y-214-2	Indonesia	Irian Jaya	Merauke
?	?	Y-215	Indonesia	Irian Jaya	Merauke
1005866	171766	Y-216	Indonesia	Irian Jaya	Merauke
1005867	171767	Y-217	Indonesia	Irian Jaya	Merauke
1005868	91403	Y-218	Indonesia	Irian Jaya	Merauke
1005869	171768	Y-219	Indonesia	Irian Jaya	Hollandia
1011023	174272	Y-226	Indonesia	Irian Jaya	Baliem
1011025	174274	Y-230	Indonesia	Irian Jaya	Baliem
1005871	171770	Y-231	Indonesia	Irian Jaya	Baliem
1011026	174275	Y-232	Indonesia	Irian Jaya	Baliem
1011027	174276	Y-233	Indonesia	Irian Jaya	Baliem
1011028	174277	Y-234	Indonesia	Irian Jaya	Baliem
1011029	174278	Y-235	Indonesia	Irian Jaya	Baliem
1005872	91404	Y-240	Indonesia	Irian Jaya	Biak
?	?	Y-240-1	Indonesia	Irian Jaya	Biak
?	?	Y-240-2	Indonesia	Irian Jaya	Biak
1011030	174279	Y-245	PNG	Port Moresby	–
1011031	174280	Y-246	PNG	Port Moresby	–
1005873	172750	Y-249	PNG	Port Moresby	–
1005874	91405	Y-252	PNG	Port Moresby	–
1005875	172751	Y-253	PNG	Port Moresby	–
1005876	171771	Y-255	PNG	Port Moresby	–
1011032	174281	Y-257	PNG	Port Moresby	–
1005877	91406	Y-260	PNG	Lae	–
1005878	91407	Y-262	PNG	Lae	–
1005879	171772	Y-263	PNG	Lae	–
1005880	91408	Y-264	PNG	Lae	–
1005881	171773	Y-266	PNG	Lae	–
1011033	174282	Y-267	PNG	E. Highlands	Watabung
1005882	91409	Y-269	PNG	E. Highlands	Watabung
1011034	174283	Y-270	PNG	E. Highlands	Watabung
?	?	Y-271	PNG	E. Highlands	Watabung

Original ID	Cultivation	Local Cultivar Name	Notes
214-1	Field	(Y-214)	from D. Ag. Hollandia, 1957 (from 1965 list)
214-2	Field	(Y-214)	from D. Ag. Hollandia, 1957 (from 1965 list)
215	Pot	–	from D. Ag. Hollandia, 1957 (from 1965 list)
216	Field	–	from D. Ag. Hollandia, 1957 (from 1965 list)
217	Field	–	from D. Ag. Hollandia, 1957 (from 1965 list)
218	Field	–	from D. Ag. Hollandia, 1957 (from 1965 list)
219	Field	*menes poetih*	from D. Ag. Hollandia, 1957 (from 1965 list)
DNG33	Pot	*lereika*	from D. Ag. Hollandia, 1957 (from 1965 list)
DNG37	Pot	*helemaiu*	–
231	Field	–	–
DNG39	Pot	*hulok*	–
DNG40	Pot	*werene*	–
DNG41	Pot	*hiwesa*	–
DNG42	Pot	*lyoka*	–
240	Field	*ansio berik*	–
240	Pot	*ansio berik* (Y-240)	–
240	Pot	*ansio berik* (Y-240)	–
ANG1	Pot	–	from D. AGR. (from 1965 list)
ANG6	Pot	–	from D. AGR. (from 1965 list)
249	Pot	–	from D. AGR. (from 1965 list)
252	Field & Pot	–	from D. AGR. (from 1965 list)
253	Pot	–	from D. AGR. (from 1965 list)
255	Field	–	from D. AGR. (from 1965 list)
ANG18	Pot	–	from D. AGR. (from 1965 list)
260	Field	*imabepu*	–
262	Field & Pot	*wanat*	–
263	Field	*dunkum*	–
264	Field	*bunwat*	–
266	Field	*kwankat*	–
ANG28	Pot	*foravena*	–
269	Pot	*fatena*	–
ANG31	Pot	*obume*	–
ANG32	Pot	*kiravena*	–

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1011035	174284	Y-271-1	PNG	E. Highlands	Watabung
1005884	91410	Y-275	PNG	E. Highlands	Aiyura
?	?	Y-275(2)	PNG	E. Highlands	Aiyura
1005885	91411	Y-276	PNG	E. Highlands	Namura
1005886	91412	Y-277	PNG	E. Highlands	Namura
1005887	91413	Y-280	PNG	E. Highlands	Namura
1011036	174285	Y-281	PNG	E. Highlands	Namura
1005888	171775	Y-282	PNG	E. Highlands	Namura
1011037	174286	Y-283	PNG	E. Highlands	Gitunu
1011038	174287	Y-287	PNG	W. Highlands	Gunts.
1005889	171776	Y-288	PNG	W. Highlands	Gunts.
1011039	174308	Y-289	PNG	W. Highlands	Gunts.
1011040	174288	Y-292	PNG	W. Highlands	Kaugel V.
1005890	91414	Y-293	PNG	W. Highlands	Kaugel V.
1005891	171777	Y-294	PNG	W. Highlands	Kaugel V.
1005892	172752	Y-295	PNG	W. Highlands	Kaugel V.
1005893	171778	Y-297	PNG	W. Highlands	Kaugel V.
1011041	174289	Y-298	PNG	W. Highlands	Kaugel V.
1011043	174291	Y-300	PNG	W. Highlands	Kaugel V.
1005894	91415	Y-301	PNG	W. Highlands	Kaugel V.
1011044	174292	Y-302	PNG	W. Highlands	Kaugel V.
1011045	174293	Y-304	PNG	W. Highlands	Kaugel V.
1005895	171779	Y-307	PNG	W. Highlands	Kaugel V.
1011048	174295	Y-309	PNG	W. Highlands	Kaugel V.
1011050	174297	Y-311	PNG	W. Highlands	Kaugel V.
1005896	91416	Y-312	PNG	W. Highlands	Kaugel V.
1005897	171780	Y-316	PNG	W. Highlands	Kaugel V.
1005898	91417	Y-322	PNG	W. Highlands	Kaugel V.
1005899	91418	Y-323	PNG	W. Highlands	Kaugel V.
1005900	172753	Y-324	PNG	W. Highlands	Kaugel V.
1005901	171781	Y-327	PNG	W. Highlands	Kaugel V.
1005903	91419	Y-333	PNG	W. Highlands	Kaironk V.
1005904	171783	Y-337	PNG	W. Highlands	Kaironk V.

Original ID	Cultivation	Local Cultivar Name	Notes
ANG32	Pot	*kiravena* (Y-271)	–
275	Pot	*serenta*	–
–	Pot	*serenta* (Y-275)	–
276	Pot	*sigarup*	–
277	Field	*hagesya*	–
280	Field & Pot	*kanifuta*	–
ANG43	Pot	*yuka*	–
282	Field	*afaioea*	–
ANG45	Pot	*mamat*	–
632	Pot	*kruo*	–
288	Field & Pot	*neq Int.*	–
635	Pot	UNKNOWN	–
663	Pot	–	from N. Bowers (from 1965 list)
293	Pot	–	from N. Bowers (from 1965 list)
294	Field	–	from N. Bowers (from 1965 list)
295	Pot	–	from N. Bowers (from 1965 list)
297	Field	–	from N. Bowers (from 1965 list)
ANG61	Pot	*momə*	from Ialibu; from N. Bowers (from 1965 list)
ANG63	Pot	*komeya*	pre-European; from N. Bowers (from 1965 list)
301	Pot	*taŋgambitlyə*	from Papua 4 years; from N. Bowers (from 1965 list)
ANG65	Pot	*arakoyə*	from Wabag via Tendep & Tegep; from N. Bowers (from 1965 list)
ANG67	Pot	*unju ka*	from Neibilyer V., near Hagan; from N. Bowers (from 1965 list)
307	Field	*konōma*	probably from “Okinawa type”; from N. Bowers (from 1965 list)
ANG72	Pot	*kipepi*	pre-European; from N. Bowers (from 1965 list)
ANG74	Pot	*tegamə kaini*	pre-European; from N. Bowers (from 1965 list)
312	Field & Pot	*moro piyə*	pre-European; from N. Bowers (from 1965 list)
316	Field & Pot	*maŋgo pora*	from Neibilyer Valley; from N. Bowers (from 1965 list)
322	Field	*keratl*	pre-European; from N. Bowers (from 1965 list)
323	Pot	*maiga*	pre-European; from N. Bowers (from 1965 list)
324	Pot	*mamale*	pre-European; from N. Bowers (from 1965 list)
327	Field & Pot	*simbu aŋgu*	introduced; from N. Bowers (from 1965 list)
333	Field & Pot	*mlkbj*	from R. Bulmer (from 1965 list)
337	Field & Pot	*kōnmay*	from R. Bulmer (from 1965 list)

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1005905	171784	Y-338	PNG	W. Highlands	Kaironk V.
1005906	171785	Y-341	PNG	W. Highlands	Kaironk V.
1011056	174302	Y-342	PNG	W. Highlands	Kaironk V.
1005907	171786	Y-343	PNG	W. Highlands	Kaironk V.
1005908	173099	Y-346	PNG	W. Highlands	Kaironk V.
1011057	174303	Y-358	PNG	New Britain	Rabaul
1005909	173100	Y-361	PNG	New Britain	Rabaul
1011058	174304	Y-362	PNG	New Britain	Rabaul
1005910	172754	Y-365	Solomon Islands	Malaita	–
1011059	174305	Y-367	Solomon Islands	Malaita	–
1005912	173101	Y-368	Solomon Islands	Malaita	–
1005913	173102	Y-375	Vanuatu	Santo	–
1005914	173103	Y-376	Vanuatu	Santo	–
1011060	91505	Y-377	Vanuatu	Santo	–
1005916	173105	Y-379	Vanuatu	Santo	–
1005917	173106	Y-381	Vanuatu	Santo	–
1011061	174311	Y-382	Vanuatu	Santo	–
1005919	172755	Y-385	Vanuatu	Santo	–
1005920	173108	Y-386	Vanuatu	Santo	–
1005921	173109	Y-387	Vanuatu	West Santo	–
1011063	174313	Y-389	Vanuatu	West Santo	–
1005923	171787	Y-391	Vanuatu	West Santo	–
1005925	173111	Y-397	Vanuatu	Vila	–
1005926	173112	Y-398	Vanuatu	Vila	–
1005927	172756	Y-399	Vanuatu	Vila	–
1005928	91420	Y-400	New Caledonia	Chabwen	–
1005929	173113	Y-404	New Caledonia	Chabwen	–
?	?	Y-410	New Caledonia	Balade	–
1011067	174317	Y-410A	New Caledonia	Balade	–
1011068	174318	Y-410B	New Caledonia	Balade	–
1005930	173114	Y-411	New Caledonia	Balade	–
1005931	172757	Y-415	New Caledonia	Balade	–
1005932	173115	Y-425	Fiji	–	–

Original ID	Cultivation	Local Cultivar Name	Notes
338	Field & Pot	*bajj*	from R. Bulmer (from 1965 list)
341	Field & Pot	*acic*	from R. Bulmer (from 1965 list)
16/B	Pot	*wŋawŋ*	from R. Bulmer (from 1965 list)
343	Field	*snamin*	from R. Bulmer (from 1965 list)
346	Field & Pot	*konm tud*	from R. Bulmer (from 1965 list)
ANG55	Pot	*tapuka*	–
361	Field	*buka*	–
ANG59	Pot	*kambubu*	–
365	Pot	*bilegalele*	recent introduction (from 1965 list)
685	Pot	*sangobolu*	recent (from 1965 list)
368	Field & Pot	UNKNOWN	'found' 1963 (from 1965 list)
375	Field & Pot	–	–
376	Field	–	–
696	Pot	–	–
379	Field & Pot	–	–
381	Field & Pot	–	–
701	Pot	–	–
385	Pot	–	–
386	Field & Pot	–	–
387	Field	–	–
709	Pot	–	–
391	Field & Pot	–	–
397	Field & Pot	–	–
398	Field & Pot	–	–
399	Pot	–	–
400	Pot	*kumala miia rose*	pre-European (from 1965 list)
404	Pot	*kumala belep*	pre-European, introduced from Belep (from 1965 list)
–	Pot	*taabo*	–
–	Pot	*taabo* (Y-410)	Derived from Y-410
–	Pot	*taabo* (Y-410)	Derived from Y-410
411	Field & Pot	–	grown by migrants of Loyalty Islands (from 1965 list)
415	Pot	–	grown by migrants of Loyalty Islands (from 1965 list)
425	Field	*red timala*	intro. from Niue (from 1965 list)

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1005933	173116	Y-426	Fiji	–	–
1005934	172758	Y-427	Fiji	–	–
1005935	173117	Y-428	Fiji	–	–
1005936	174319	Y-429	Fiji	–	–
1005937	173118	Y-432	Fiji	–	–
1005938	173119	Y-443	Tonga	Nukua‘lofa	–
1011070	174320	Y-451	Tonga	Nukua‘lofa	–
?	?	Y-452	Tonga	Nukua‘lofa	–
1005939	173193	Y-452(TAN)A	Tonga	Nukua‘lofa	–
1014593	173194	Y-452(TAN)B	Tonga	Nukua‘lofa	–
1011072	91506	Y-458	Tonga	Nukua‘lofa	–
1005940	173120	Y-465	Samoa	Aleisa	Opolu
1011073	174322	Y-466	Samoa	Aleisa	Opolu
1005942	172759	Y-482	Cook Islands	Aitutaki	–
1005943	173122	Y-484	Cook Islands	Aorangi	–
1005944	173123	Y-485	Cook Islands	Rarotonga	–
1011076	174324	Y-490	Cook Islands	Mangaia	–
1005945	173124	Y-491	Cook Islands	Mangaia	–
1005946	173125	Y-500	New Zealand	Auckland	–
1005947	172760	Y-501-1	New Zealand	Auckland	–
1005948	173126	Y-501-2	New Zealand	Auckland	–
1005949	173127	Y-502	New Zealand	Auckland	–
1005950	173128	Y-503	New Zealand	Taranaki	–
1005952	173130	Y-507	New Zealand	Kawhia	–
1005953	173131	Y-508	New Zealand	Tauranga	–
1005954	173132	Y-512	New Zealand	Tauranga	–
1005956	173133	Y-513-2	New Zealand	Paihia	–
1005957	173134	Y-521	USA	Hawai‘i	–
1005958	173135	Y-525	French Polynesia	Society Islands	Moorea
1005959	173136	Y-528	French Polynesia	Society Islands	Moorea
1005960	171788	Y-532	French Polynesia	Society Islands	Borabora
1005961	173137	Y-533	French Polynesia	Society Islands	Borabora
1011077	174325	Y-535	French Polynesia	Society Islands	Borabora

Original ID	Cultivation	Local Cultivar Name	Notes
426	Field & Pot	*white timala*	intro. from Niue (from 1965 list)
427	Pot	*samoa*	intro. from W. Samoa (from 1965 list)
428	Field & Pot	*V52*	unknown source (from 1965 list)
429	Pot	*navuso*	local (from 1965 list)
432	Field & Pot	*cocolevu*	local (from 1965 list)
443	Field & Pot	*kumala lau manioke*	"leaves of the manioc" (from 1965 list)
903	Pot	*kumala teiko*	–
452	Pot	*kumala tolumahina*	–
452	Field & Pot	*kumala tolumahina* (Y-452)	–
452	Field & Pot	*kumala tolumahina* (Y-452)	–
912	Pot	UNKNOWN	new spontaneous variety (from 1965 list)
465	Field & Pot	UNKNOWN	"wild" (from 1965 list)
881	Pot	UNKNOWN	"wild" (from 1965 list)
482	Field & Pot	*tauranga/torumarama*	–
484	Pot	*pipi 2*	–
485	Field & Pot	*rokini 1*	–
CI22	Pot	–	–
491	Field & Pot	–	–
500	Field	*owairaka red*	–
501-1	Pot	*tauranga red* (Y-501)	–
501-2	Field	*tauranga red* (Y-501)	–
502	Field & Pot	*gisborne red*	–
503	Field	*waina*	–
507	Field	*kotepo*	–
508	Field & Pot	*rekamaroa*	–
512	Pot	*hutihuti*	mutant from Y-511 (Hutihuti), 1958 (from 1965 list)
513-2	Field	*taputini* (Y-513)	–
521	Field & Pot	*kanaka*	from U. California (from 1965 list)
525	Field	–	–
528	Field	–	–
532	Field & Pot	*umara vario*	–
533	Field & Pot	*umara hererei*	–
865	Pot	*umara tara vteute*	–

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1011078	91508	Y-536	French Polynesia	Society Islands	Borabora
1005962	173138	Y-537	French Polynesia	Society Islands	Raiatea
1005963	173139	Y-539	French Polynesia	Society Islands	Raiatea
1011079	171789	Y-540	French Polynesia	Society Islands	Raiatea
1005964	172762	Y-541	French Polynesia	Society Islands	Raiatea
1005965	173140	Y-542	French Polynesia	Society Islands	Tahiti
1011080	171790	Y-544	French Polynesia	Society Islands	Moorea
1005966	173141	Y-545	French Polynesia	Society Islands	Moorea
1011081	171791	Y-552	French Polynesia	Marquesas Islands	Ua-huka
1005967	173142	Y-553	French Polynesia	Marquesas Islands	Fatu-hiva
1005968	171792	Y-556	French Polynesia	Marquesas Islands	Nukuhiva
1011082	171793	Y-557	French Polynesia	Marquesas Islands	Nukuhiva
1005969	171794	Y-561	French Polynesia	Marquesas Islands	Va-pau
1005970	171795	Y-563	French Polynesia	Marquesas Islands	Va-pau
1011084	171796	Y-570-2	Chile	Easter Is.	Hangaroa
1005971	171803	Y-594	Peru	Piura V.	–
1011093	171804	Y-595	Peru	Piura V.	–
1011094	171805	Y-597	Peru	Piura V.	–
1005973	171808	Y-600	Peru	Piura V.	–
1005975	171811	Y-607	Peru	Chicama V.	–
1005976	171812	Y-609	Peru	Chicama V.	–
1011097	171813	Y-610	Peru	Chipillico V.	–
1005977	171814	Y-611-1	Peru	Chipillico V.	–
1005978	171815	Y-611-2	Peru	Chipillico V.	–
1011098	171816	Y-614	Peru	Chira V.	–
1005979	173143	Y-615	Peru	Chira V.	–
?	?	Y-620(1)	Peru	Chira V.	–
?	?	Y-620(2)	Peru	Chira V.	–
1011100	171818	Y-622	Peru	Chira V.	–
1005981	173145	Y-623	Peru	Chira V.	–
1005983	171820	Y-626	Peru	Chira V.	–
1011101	174327	Y-628	Peru	la Encalada	Trujillo
?	?	Y-630	Peru	Moche V.	–

Original ID	Cultivation	Local Cultivar Name	Notes
866	Pot	*umara tara uouo*	–
537	Field	*umara pitanya*	–
539	Field	–	–
876	Pot	–	–
541	Pot	–	–
542	Field & Pot	–	–
821	Pot	–	–
545	Field & Pot	–	–
844	Pot	*maita*	–
553	Field	*‘a‘poa*	–
556	Field	*papau*	–
851	Pot	*pikino*	–
561	Field & Pot	*‘a‘poa*	–
563	Field & Pot	*hou*	–
–	Pot	*arenga riki riki* (Y-570)	derived from Y-570
594	Field & Pot	*ariqueno*	–
A6	Pot	*pedro*	–
A10	Pot	*hoji blanco*	–
600	Field & Pot	–	–
607	Field	*esperma morado*	–
609	Field & Pot	*pierma de viuda*	–
D1	Pot	*otero*	–
611-1	Field & Pot	*pava blanca* (Y-611)	–
611-2	Pot	*pava blanca* (Y-611)	–
F3	Pot	–	–
1217	Field	*porto viejo*	–
–	Pot	*pedro* (Y-620)	–
–	Pot	*pedro* (Y-620)	–
PC4	Pot	*porto viejo*	–
623	Field	*cuarenta*	–
626	Field & Pot	*paramonguina*	–
T2	Pot	*boca de chisco*	–
–	Pot	*blanco esperma*	–

Table continued next page...

Table continued...

Access-ion No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1011104	174329	Y-630B	Peru	Moche V.	–
1011105	171823	Y-634	Peru	Huacho	–
1005984	171825	Y-636	Peru	Huacho	–
1005985	171826	Y-637	Peru	Huacho	–
1005986	171827	Y-638	Peru	Huacho	–
1005987	171828	Y-639	Peru	Huacho	–
1011107	171829	Y-641	Unknown	Huacho	–
1005988	171831	Y-644	Peru	Lima	–
1005989	171832	Y-645	Peru	Lima	–
1005991	173146	Y-657	Peru	Julian	Callejeon de Huaylas
1005992	171836	Y-660	Peru	Julian	Callejeon de Huaylas
1005993	171837	Y-662	Peru	Recuay	Ancash
1011112	174330	Y-665-1	Peru	Recuay	Ancash
?	?	Y-665-2	Peru	Recuay	Ancash
1005994	171838	Y-666	Peru	Recuay	Ancash
?	?	Y-667	Unknown	–	–
1011114	171840	Y-670	Colombia	Bogata	–
1011115	171841	Y-671	Colombia	Don Martias	–
1011116	171842	Y-672	Colombia	Felidia	–
1005995	91421	Y-677	Colombia	Palmira	–
1011118	171843	Y-679	Colombia	Palmira	–
1011119	171844	Y-680	Colombia	Palmira	–
1005996	91422	Y-682	Colombia	Palmira	–
?	?	Y-682(1)	Colombia	Palmira	–
1005997	172763	Y-684	Colombia	Retirode de los Indios	–
1011120	171846	Y-686	Colombia	Retirode de los Indios	–
1011121	171847	Y-687	Colombia	Retirode de los Indios	–
1011122	171848	Y-688	Colombia	Retirode de los Indios	–
1011123	171849	Y-690	Ecuador	Equator	–
1011124	171850	Y-691	Ecuador	Equator	–
1005998	171851	Y-692	Ecuador	Guayaquil	–
1011125	171852	Y-695	Ecuador	San Horca	–
1005999	171853	Y-702	USA	Louisiana	–

Original ID	Cultivation	Local Cultivar Name	Notes
–	Pot	*blanco esperma* (Y-630)	Derived from Y-630
HU3	Pot	–	–
636	Pot	–	–
637	Field & Pot	–	–
638	Field & Pot	–	–
639	Field & Pot	–	–
HU13	Pot	–	–
644	Field & Pot	–	–
645	Field & Pot	–	–
657	Field	–	–
660	Field & Pot	–	–
662	Pot	–	–
–	Pot	(Y-665)	Derived from Y-665
–	–	(Y-665)	–
666	Pot	–	–
–	Pot	–	–
BG1	Pot	–	–
DM2	Pot	–	–
DM2	Pot	–	–
667	Field	–	–
PR7	Pot	–	–
PR8	Pot	–	–
682	Field	–	–
–	Pot	–	–
684	Pot	–	–
ER4	Pot	–	–
ER5	Pot	–	–
ER6	Pot	–	–
ECE1	Pot	–	–
ECE4	Pot	–	–
692	Field	–	–
ECQ2	Pot	–	–
702	Field	*Myers Early*	–

Table continued next page...

Table continued...

Accession No	JP-No	Cultivar Name	Origin 1	Origin 2	Origin 3
1006000	171854	Y-704	USA	Louisiana	–
1006001	171855	Y-705	USA	Louisiana	–
1006002	171856	Y-710	Mexico	–	–
1006003	91423	Y-712	USA	Puerto Rico	–

Original ID	Cultivation	Local Cultivar Name	Notes
704	Field & Pot	seedling 12	–
705	Field	Acadian	–
710	Field & Pot	–	sent by USDA (from 1965 list)
712	Field	–	sent by USDA (from 1965 list)

APPENDIX 5: AFLP OLIGONUCLEOTIDES

Table A5.1
Oligonucleotides Used in AFLP

Class		Name	Sequence (5′–3′)[a,b]
Adapter	***Eco*R I Linker**	Eco Linker I	CTCGTAGACTGCGTACC
		Eco Linker II	AATTGGTACGCAGTCTAC
	***Mse* I Linker**	Mse Linker I	GACGATGAGTCCTGAG
		Mse Linker II	TACTCAGGACTCAT
Pre-selective amplification	***Eco*R I Primer**	Eco+A	GACTGCGTACCAATTCA
	***Mse* I Primer**	Mse+C	GATGAGTCCTGAGTAAC
Selective amplification	***Eco*R I Primer**	6FAM[c]-Eco+ACT	GACTGCGTACCAATTCACT
		VIC[c]-Eco+AGC	GACTGCGTACCAATTCAGC
		NED[c]-Eco+ATA	GACTGCGTACCAATTCATA
		PET[c]-Eco+AAG	GACTGCGTACCAATTCAAG
	***Mse* I Primer**	Mse+CAA	GATGAGTCCTGAGTAACAA
		Mse+CAT	GATGAGTCCTGAGTAACAT
		Mse+CAC	GATGAGTCCTGAGTAACAC
		Mse+CAG	GATGAGTCCTGAGTAACAG
		Mse+CTT	GATGAGTCCTGAGTAACTT
		Mse+CTG	GATGAGTCCTGAGTAACTG
		Mse+CCC	GATGAGTCCTGAGTAACCC
		Mse+CGG	GATGAGTCCTGAGTAACGG

[a] Based on Vos *et al.* (1995).
[b] Underlining denotes selective bases.
[c] Fluorescent label (see Table 4.3 and Table 4.4).

APPENDIX 6: AFLP SCORING PARAMETERS

Table A6.1
AFLP Scoring Parameters in GeneMapper® v. 3.7[a]

Parameter	Setting
Size Standard	GS500(−250)LIZ (Factory Provided)
Allele (AFLP Analysis)	
Analyze Dyes	Blue; Green; Yellow; Red
Analysis Range (bp)	
Start	50.0
End	500.0
Normalization Scope	Project
Normalization Method	Sum of signal
Panel	Generate panel using samples
Bin width (bp)	0.5
Use all samples	Yes
Allele Calling	Name alleles using labels
Threshold Value Type	Absolute
Labels	$< 50.0 = 0; \geq 50.0 = 1$
Delete common alleles	No
Peak Detector	
Peak Detection Algorithm	Advanced
Ranges	
Analysis	Full Range
Sizing	All Sizes
Smoothing and Baselining	
Smoothing	Light
Baseline Window	51 pts
Size Calling Method	Local Southern Method

Table continued next page...

Table continued...

Parameter	Setting
Peak Detector (cont…)	
Peak Detection	
Peak Amplitude Thresholds	
B	50
G	50
Y	50
R	50
O	50
Min. Peak Half Width (pts)	2
Polynomial Degree	3
Peak Window size (pts)	15
Slope Threshold	
Peak Start	0.0
Peak End	0.0
Peak Quality	
Peak morphology	
Max peak width (bp)	1.5
Pull-up peak	
Pull-up ratio	0.1
Pull-up scan	1
Quality Flags	
Quality Flag Settings	
Spectral Pull-Up	0.5
Broad Peak	0.5
Off-scale	0.5
PQV Thresholds	
Sizing Quality	
Pass Range	0.75 to 1.0
Low Quality Range	0.0 to 0.25
Genotype Quality	
Pass Range	0.75 to 1.0
Low Quality Range	0.0 to 0.25

[a] Applied Biosystems.

In all Ages wherein Learning hath Flouriſhed, complaint hath been made of the Itch of Writing, and the multitude of worthleſs Books, wherewith importunate Scriblers have peſtered the World, Scribimus indoctі doctíq;

John Ray
The Wisdom of God Manifested in the Works of the Creation (1974 [1691])

Appendix D

MASSEY UNIVERSITY
Application for Approval of Request to Embargo a Thesis
(Pursuant to AC98/168 (Revised 2), Approved by Academic Board 17/02/99)

Name of Candidate: Andrew Christopher Clarke ID Number: 98041621

Degree: Doctor of Philosophy Dept/Institute/School: Molecular BioSciences

Thesis title: Origins and Dispersal of the Sweet Potato and Bottle Gourd in Oceania: Implications for Prehistoric Human Mobility

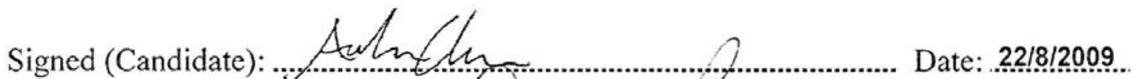
Name of Chief Supervisor: Professor David Penny Telephone Ext: 5033

As author of the above named thesis, I request that my thesis be embargoed from public access until (date) 1 July 2010 for the following reasons:

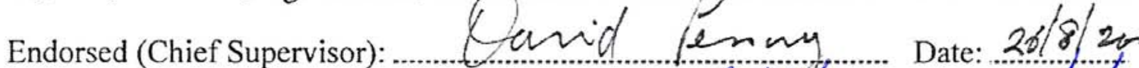
- ☐ Thesis contains commercially sensitive information.
- ☐ Thesis contains information which is personal or private and/or which was given on the basis that it not be disclosed.
- ☑ Immediate disclosure of thesis contents would not allow the author a reasonable opportunity to publish all or part of the thesis.
- ☐ Other (specify):

Please explain here why you think this request is justified:

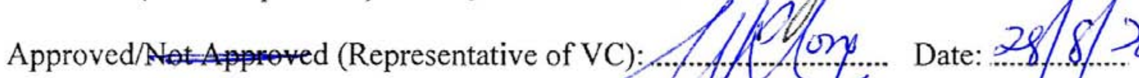
Those sections of the thesis not already published are included in several manuscripts that are currently in preparation. Allowing these sections to become publicly accessible would compromise their publication in journals that consider public access a form of unacceptable "pre-publication".



Signed (Candidate): Date: 22/8/2009



Endorsed (Chief Supervisor): Date: 26/8/2009



Approved/~~Not Approved~~ (Representative of VC): Date: 28/8/2009

Note: Copies of this form, once approved by the representative of the Vice-Chancellor, must be bound into every copy of the thesis.