

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.

NGĀ URI O KARAKA:  
A GENETIC STUDY OF THE KARAKA/KŌPI TREE  
IN AOTEAROA/NEW ZEALAND

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

Doctor of Philosophy in Genetics

at Massey University/Te Kunenga ki Pūrehuroa,  
Palmerston North/Te Papaioea, New Zealand/Aotearoa.

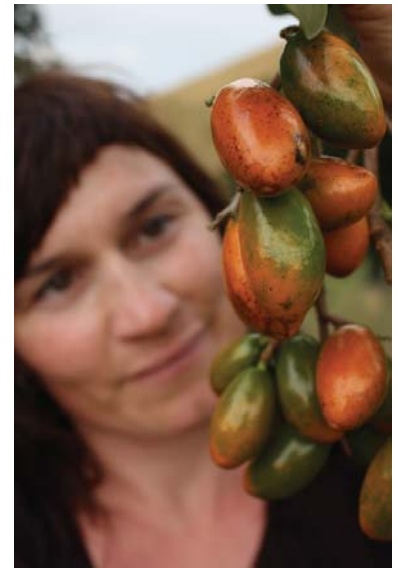
Robin Amber Atherton

July 2014



MIHI

Ko Cymru te whenua  
Ko Eryri te maunga  
Ko Banwy te awa  
Ko Vyrnwy te moana  
Ko Robin Amber Atherton tōku ingoa



I was born in England to my father, whose father was a Yorkshireman, and whose mother was a Welshwoman from Anglesey, and my mother, whose parents were both from Yorkshire. I was raised partly in South Africa, but mostly in Cymru (Wales) in a small village called Y Foel nestled in the hills in the mid-central part of the principality. At primary school I learnt Welsh in full-immersion and delved into the Welsh world feet first, learning to recite Welsh poetry, sing Welsh songs and participating in cultural competitions, known as *eisteddfod*. My roots are firmly planted in the alluvial soils of the Banwy region, it is where I feel empowered and connected; it is my foundation, my home, my *tūrangawaewae*.

My love for *Papa-tū-a-nuku* (Mother Earth), the world around us, and my interest in languages and travel, brought me to Aotearoa to continue my studies. It felt comfortable here, like a second home, and I started to learn Te Reo Māori. Being the mother of a Māori child, my world and Te Ao Māori (the Māori world) become closer with each passing day.

My PhD research has taken me all over this beautiful land, collecting leaf samples and measuring karaka/kōpi tree trunks. I am fortunate to have seen hidden coves and inlets, cliffs and coastal banks, isolated hilltops and bluffs, that few others have. Through my study of the karaka tree, my roots have sunk deep into Papa-tū-a-nuku, and Aotearoa is now my home.

*Mā te rongo, ka mōhio;  
Mā te mōhio, ka mārama;  
Mā te mārama, ka mātau;  
Mā te mātau, ka ora.*

*Through resonance comes cognisance;  
through cognisance comes understanding;  
through understanding comes knowledge;  
through knowledge comes life and well-being.*

## ABSTRACT

---

Polynesians translocated a number of plant species around the Pacific region. Many of these tropical crops were probably introduced to New Zealand, however, only a few survived owing to the cooler climate. Compensating for the loss of introduced crops, Māori cultivated endemic species they discovered in New Zealand. This project focuses on cultural and evolutionary aspects of the cultivation of one of these, karaka (*Corynocarpus laevigatus* Forst. & Forst.), which was cultivated for its highly nutritious kernel. Originally it is thought to have been restricted to the northern North Island. Its occurrence in the southern North Island, the South Island, Chatham and Kermadec Islands is strongly associated with Māori and Moriori archaeological sites and considered to have resulted from translocations as part of its cultivation. For this project, hypotheses were formulated based on existing written accounts of oral histories, published studies on karaka and informal observations and recollections. Oral histories exist regarding the origins of some translocated populations and have the potential to play an important role in tracing the history of karaka.

The relationships among the five *Corynocarpus* species were investigated by analyzing DNA sequences amplified using universal nuclear and chloroplast markers to test hypotheses of the inter- and intraspecific relationships of the genus. Nuclear markers suggest a closer relationship between *C. laevigatus* and *C. dissimilis* whereas the interpretation from chloroplast markers is less clear. This is indicated by the *rbcL* and *trnL-trnF* networks, which both show a reticulation suggesting support for both *C. laevigatus* and *C. similis* being more closely related to each other and *C. laevigatus* and *C. dissimilis* being more closely related. Nevertheless, in all cases, all markers suggest a close relationship between *C. laevigatus* and *Corynocarpus* species to the north of New Zealand (*C. dissimilis* in New Caledonia and *C. similis* in Vanuatu).

Using universal primers, intraspecific variation within karaka was found to be too low for studying translocation histories within New Zealand and extensive marker development was necessary. The first step in the development of chloroplast markers was characterisation of the chloroplast genome as a reference for different strategies in

molecular marker identification. A protocol was developed for the isolation of chloroplasts and the sequencing of the chloroplast genome using the Illumina Genome Analyser II. This protocol was also shown to be effective in the characterisation of chloroplast genomes in other elements of the New Zealand flora.

The sequence variability of the karaka chloroplast genome was investigated as a potential source for seed dispersal markers. A set of seven chloroplast molecular markers was developed and evaluated in terms of their potential for elucidating the history of karaka translocation during Māori settlement of New Zealand. Long-range polymerase chain reaction products were amplified from the chloroplast genome sequenced using Illumina Genome Analyser II, which enabled the identification of 48 putative chloroplast single nucleotide polymorphisms (SNPs). Sanger sequencing validated 16 of these detected SNPs. High resolution melting (HRM) was evaluated as an accurate, sensitive and fast PCR-based method to screen SNP variations in the chloroplast genome of karaka. Sufficient resolution in the data enabled an evaluation of the phylogeographic distribution of karaka to provide insight into the extent of human-mediated dispersal of the tree in New Zealand.

The results of the analysis of species-specific markers show the potential of the chloroplast genome to study recent events in plant history, and the use of HRM to assay several hundred accessions for a suite of chloroplast SNPs. They show an interesting relationship between Kermadec Island karaka and mainland karaka, and between Rekohu/Chatham Islands karaka and mainland karaka. To be able to pinpoint the location of the source for Rekohu/Chatham Islands karaka, more genetic work is required. However, these results are promising in their ability to trace the translocation of one of New Zealand's most important ethnobotanical species. By developing a more detailed picture of the genetic variation of karaka, this work has the potential to be the foundation for a deeper study into the translocation of the species. This has implications for further understanding the level of domestication in karaka, which at present cannot be ascertained.

## ACKNOWLEDGMENTS

---

***Ehara taku toa, he taki tahi, he toa taki tini***

*My success should not be bestowed onto me alone,  
as it was not individual success, but success of a collective*

---

It would not have been possible to write this doctoral thesis without the help and support of some wonderful and kind people around me. I thank everyone who has offered help and advice over the last five years.

Above all, I would like to thank my principal supervisor Dr Lara Shepherd for her vision, hard work, personal support and great patience at all times. My supervisor Professor Peter Lockhart who has given me his unequivocal support throughout, as always, for which this expression of thanks does not suffice. This thesis would not have been possible without the help, support and patience of my third supervisor Dr Nick Roskrige (Te Ātiawa), not to mention his advice and unsurpassed knowledge of Māori customs, culture and protocol. Trish McLenachan, no words can express my gratitude for all your tireless efforts to help me in the lab and for all the sequencing you did for me. Thanks too for all your emotional support and comments on each chapter.

I am most grateful to Peter de Lange for providing me hundreds of karaka leaf and herbarium accessions from some far flung places on these two islands, the collection of which has been valuable, time-saving and at times, for Peter, both exciting and dangerous. Peter, you are like a mountain goat, I am sincerely thankful for all your efforts. Thank you too to the following people who have also provided karaka samples or assisted with sample collection: Lara Shepherd, Leon Perrie, Jeremy Rolfe, Patrick Brownsey, Barry Sneddon, Jean-Claude Stahl, Bill Wallace, Simon Cox, David Havell, Janene Collings, Geoff Wall, Kay Kitchener, Craig McGill, Anna McNaughton, Patricia Aspinal, Kevin Matthews, Z Stevenson, Joseph Potangaroa, Eleanor Burton, Mike Shepherd, Stephen King and Jill Rapson. Thank you Bex Smith and Josie Monaghan who both helped me when I was heavily pregnant and still out collecting samples!



The good advice, support (especially through those last tough weeks) and friendship of my dear friend Dr Andrew Clarke, has been invaluable on both an academic and a personal level, for which I am extremely grateful. Andrew, you have been a pleasure to come to know, and, besides me, you are the funniest person I know! Big thanks to Chris Stowe whose Masters thesis was a fantastic source of information, references and location data. I was fortunate enough to accidentally meet Chris at a beach house on the Bay of Plenty coast (my brother is friends with Chris's brother) so we got to swap notes on karaka.

This project would not have been able to get off the ground without close contact with those involved in Te Ao Māori and Te Ao Moriori (the Māori World and the Moriori World). I have learnt so much about tikanga Māori and have a deep respect for Māori and Moriori culture and traditions. I would like to sincerely thank Maui Solomon (Hokotehi Moriori Trust) and Susan Thorpe, tēnā kōrua, your acceptance, love, help, guidance and support has been second to none, thank you for accepting both Kōpi and me into your whānau, and for supporting me naming Kōpi after the kōpi groves. Many thanks to Tom Lanauze, a fount of knowledge; Horipo (Dane) Rimene and Joseph Potangaroa (Rangitāne ō Wairarapa); Haami Te Whaiti (Ngāti Hinewaka); Henare Manaene (Ngāti Kahungunu ki Wairarapa); Clive Stone (Ngāti Wai); Alex Nathan (Te Rōroa Whatu Ora Trust); Barney Haami (Te Rūnanga o Tamaupoko); Utiku Potaka (Ngāti Hauiti); Jon Proctor (Rangitāne o Manawatu); Marty Davis (Te Kahui ō Rauru); Luana Pirihi (Patuharakeke Trust Board), Frances and Greg White (Ngāti Tama), Michelle Wi (Te Aupōuri); Victor Holloway and Alan Hetaraka (Ngāti Kahu); Rongo Bentson and Ani Walker (Te Rarawa); Jonda Subritsky (Ngāti Kuri); Rachel Puentener (Ngai Tahu); Mark Te One (Te Ātiawa, Taranaki Whanui); Paula Wilson (Pātaka Komiti); Bobby Morehu (Ngāti Tūwharetoa); Ngāti Rārua, Te Ātiawa, Ngāti Tama, Wakatu Inc and Ngāti Rārua Ātiawa Iwi Trust (Tiakina te Taiao Ltd.).

Sincere thanks to the following landowners, who allowed me access to their land to collect karaka samples: Hugh Wilson, Ōtanerito Arboretum Hinewai Reserve; Bryan and Helen Hocken (QE2 covenant), Tarata; Murray and Pahau Thacker, Okains Bay; Stan and Mary, Marakopa; John and Mary McGuiness, Flatpoint Station, Wairarapa; Jo Tuanui, Waihi, Rekohu/Chatham Islands; Arthur Bowen, Mahanga. Thank you to the wonderful people who picked me up hitchhiking around Banks Peninsula when I forgot

my driving license and couldn't hire a car. Thank you Bill Wallace (QE2 Trust), Tony Silbery, Rod Wallace, John Ogden, Mere Roberts, Helen Leach and Michael Taylor for your support and advice.

I would like to acknowledge the financial, academic and technical support of the Allan Wilson Centre and Massey University, Palmerston North and its staff, particularly in the award of an Allan Wilson Centre scholarship, a Massey Doctoral Research Scholarship and an Institute of Fundamental Sciences departmental scholarship that provided the necessary stipend support for this research. A Royal Society of New Zealand Marsden grant (MAU 07-MAU-090) financed the lab consumables and fieldtrip costs. A JP Skipworth Scholarship for Plant Biology, a Heseltine Bursary and two travel awards from the Institute of Fundamental Sciences funded three separate trips to Rekohu/Chatham Islands. Many thanks to Ann Truter, Cynthia Cresswell and Joy Wood for secretarial support and Katrina Ross for advice and guidance throughout this time. I owe a debt of gratitude to the ladies at the International Student Office at Massey University: Natalia Benquet, Olive Pimentel, Diane Reilley and Sylvia Hooker.

For technical support, Dr Lesley Collins, Senior Research Fellow, Massey University, thank you for suggesting edits to all chapters and for your emotional support, patience and comments. Thank you Dr Patrick Biggs, Senior Lecturer in Computational Biology, mEpilab and Infectious Disease Research Centre, Massey University, for all your help and for producing the Circos plot in Figure 4.6, Chapter 4. Huge thanks to Dr Andrew Clarke, Research Fellow at University of Warwick, for producing Figure 1.9 in Chapter 1, to Matt Irwin, GIS/Remote Support Officer, Institute of Agriculture & Environment at Massey University, for producing Figure 4.6 in Chapter 4 and to Rachael Ouwejan for producing Figure 4.7 in Chapter 4.

Special thanks also to all my graduate friends, especially PLEB, Farside and Phoenix lab members: Lizzie Daly, Dr Barbara Schönfeld, Dr Bennet McComish, Juan Carlos Garcia-Ramirez, Dr Simon Hills, Dr Gillian Gibb, Dr Nick Albert, Simon Cox, Tariq Mahmood, Josie Monaghan, Ibrar Ahmed and Dr Jian Han for sharing the literature and invaluable assistance. Thank you to Nick, Andrew, Barbara and Gillian for comments on some of the chapters. I really appreciated comments, advice and corrections from all

three examiners: Dr Phil Wilcox (Scion), Professor David Penny (Massey University) and Dr Michael Knapp (Bangor University).

Without the support of friends, I am not sure I would have made it to publishing this thesis. Jade, my soul-sister, thank you for all the love and support you gave to me, your friendship is pure gold. Peter Horsley, thank you from the bottom of my heart, you showed me such kindness and compassion, I truly appreciate the refuge of your lovely big house, where Kōpi was born, where I spent several months sitting in the garden listening to the birds. Thanks to Amy and Pete for letting us live with them for six weeks; to Jess and Richard for providing a refuge for us at Westoe for the last five months of thesis writing; thanks to Tabitha who looked after Kōpi one day a week for months on end so I could write; to Debbie, Kristal, Nirbha, Rachel H, Rachel A, Michelle, Bianca, Jules, Sophie, Annie, Susan, Heather, Rebekah, Lucy and Lynlee. Huge thanks to my sister Melanie, for standing by my side through all the important and life-changing parts. It really does take a village to raise a child (and support a mother writing a thesis at the same time). Mum and dad, I know you have no idea if I actually DO anything important, but I know you believed in my ability to nail this beast, and that's all I needed.

I dedicate this doctoral thesis to my son Kōpi, named for the kōpi trees on Rekohu/Chatham Islands. Having a small child AND a thesis to write is like having twins: both demand your constant attention and keep you up all night worrying. Kōpi, you have been such a patient and happy child, and although it has been anything but plain sailing, it has been a joy to share this journey with you. You have been a constant reminder of why I needed to complete this thesis. I love you with all my heart.

# TABLE OF CONTENTS

---

PREFACE	i
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
LIST OF FIGURES	xiii
LIST OF TABLES	xiv
RESEARCH OBJECTIVES	xv
THESIS LAYOUT	xvi
 1. CHAPTER ONE: GENERAL INTRODUCTION	
1.1 Chapter overview	1
1.2 Introduction	2
1.2.1 Corynocarpaceae	2
1.2.1.1 Taxonomy	2
1.2.1.2 Distribution	3
1.2.1.3 <i>Corynocarpus</i> in New Zealand	7
1.3 The Biology of karaka	10
1.3.1 Phenology	10
1.3.2 Pollination biology	10
1.3.3 Life-cycle strategy	11
1.3.4 Dispersal of karaka	12
1.4 The cultural significance of karaka	17
1.5 Corynocarpus in the Pacific region	17
1.5.1 The name karaka and its cognates in the Pacific region	18
1.6 The ‘introduction’ of karaka to New Zealand	20
1.7 The cultivation of karaka	22
1.8 Incipient domestication	24
1.8.1 Domestication defined	24
1.8.2 Genetic diversity	24
1.8.3 Domestication model	26
1.9 References	29
 2. CHAPTER TWO: ORIGINS OF KARAKA IN NEW ZEALAND	
2.1 Chapter overview	37
2.2 A note on attribution	37

2.3 Abstract	38
2.4 Introduction	38
2.4.1 Corynocarpaceae	38
2.4.2 Vegetation history of lowland species in New Zealand	39
2.4.3 What was the pre-human distribution of karaka in New Zealand, based upon what we know of other lowland species?	41
2.4.4 Molecular systematics of karaka	44
2.5 Methods	45
2.5.1 Sample Collection	45
2.5.2 DNA extraction, polymerase chain reaction amplification and sequencing	45
2.5.3 Data analysis	48
2.5.4 Dating ITS sequences	48
2.6 Results	49
2.6.1 ITS sequences	49
2.6.1.1 Dating ITS sequence divergence between the Three Kings Islands and mainland karaka	49
2.6.2 WAXY sequences	51
2.6.3 <i>rbcL</i> sequences	52
2.6.4 <i>trnL-trnF</i> sequences	53
2.7 Discussion	60
2.7.1 ITS and WAXY sequences	60
2.7.2 <i>rbcL</i> and <i>trnL-trnF</i> sequences	61
2.8 Conclusion	62
2.9 References	64
 3. CHAPTER THREE: WHOLE GENOME SEQUENCING OF ENRICHED CHLOROPLAST DNA USING THE ILLUMINA GAII PLATFORM	
Preamble	69
3.1 Utility/evaluation of the chloroplast as a molecule for a high-resolution study of translocation	69
3.2 References	73
Atherton <i>et al</i> Plant Methods paper	76
Statement of contribution to Atherton <i>et al</i> 2010	82
 4. CHAPTER FOUR: SNP MARKERS FOR KARAKA ASSAYED USING HIGH-RESOLUTION MELT ANALYSIS	
4.1 Chapter overview	83

4.2 A note on attribution	83
4.3 Abstract	84
4.4 Introduction	85
4.4.1 Background	85
4.5 Translocation of karaka	87
4.5.1 Genetic study	88
4.5.2 Molecular methods	88
4.6 Materials and methods	93
4.6.1 Sample collection and DNA extraction	93
4.6.2 Preparation of short-range amplicons for Sanger sequencing using universal primers	95
4.6.3 Preparation of long-range amplicons for next-generation sequencing using species-specific primers	96
4.6.4 Illumina sequencing, mapping and visualisation of SNPs	97
4.6.5 Sanger-based SNP validation	98
4.6.6 High-resolution melting PCR design and optimisation	98
4.6.7 Phylogenetic analysis	99
4.6.8 Using a previously discarded SNP for further resolution in the data set	99
4.6.9 Comparison with the spatial and climate data of the distribution of karaka	100
4.7 Results	100
4.7.1 Initial chloroplast investigations using universal primers	101
4.7.2 SNPs	101
4.7.3 HRM marker optimisation	101
4.7.4 HRM screening of the karaka population	103
4.7.5 HRM method compared with Sanger sequencing	104
4.7.6 Exploring the distribution of karaka in New Zealand	107
4.7.6.1 Chlorotypes and their relationships	107
4.7.6.2 Distribution of chlorotypes in New Zealand	108
4.7.6.2 Comparison between chlorotype distribution and spatial and climate data of the distribution of karaka in New Zealand.	109
4.8 Discussion	109
4.8.1 SNP discovery and verification	109
4.8.2 Effectiveness of HRM profiling	110
4.8.3 Evolution and distribution of karaka chlorotypes in New Zealand	113
4.9 Alternative approaches	115
4.10 Conclusion	118

4.11 References	119
5. CHAPTER FIVE: THESIS SUMMATION AND FUTURE DIRECTIONS	
5.1 Thesis summary	127
5.1.1 General summary	127
5.1.2 Chloroplast isolation	128
5.1.3 HRM screening	129
5.2 Future directions	131
5.2.1 Development of microsatellite markers	131
5.2.2 Double digest restriction associated DNA sequencing (DDRadSeq)	132
5.2.3 Circos plots and hotspot regions	133
5.2.4 Amplifying chloroplast genomes using Replphi™ PHI29 DNA polymerase	133
5.2.5 Exome capture	134
5.2.6 Oral histories	134
5.3 References	135
APPENDICES	
Appendix 1: Sampling strategy and consultation with Māori	137
Appendix 2: List of accessions	141
Appendix 3: List of sequencing primers	155
Appendix 4: SNP marker development table	159
Appendix 5: Table of comparison of genotyping results using HRM and Sanger sequencing	163
Appendix 6: Full data set of 360 accessions genotyped with seven SNPs	165
Appendix 7: Paper reprint: Zhong <i>et al</i> 2011	175
Statement of contribution to Zhong <i>et al</i> 2011	185
Oxford University Press license to reprint Zhong <i>et al</i> 2011	186
Appendix 8: Paper reprint: Goremykin <i>et al</i> 2013	187
Statement of contribution to Goremykin <i>et al</i> 2013	199
Oxford University Press license to reprint Goremykin <i>et al</i> 2013	200
Appendix 9: Nexus files used to make haplotype networks (on cd)	CD
Appendix 10: Sequence alignments of SNPs	CD

## LIST OF FIGURES

---

FIGURE 1.1: Karaka ( <i>Corynocarpus laevigatus</i> ) in fruit	3
FIGURE 1.2: Distributions and chromosome numbers of <i>Corynocarpus</i> species	4
FIGURE 1.3: Relationships within <i>Corynocarpus</i>	5
FIGURE 1.4: <i>Corynocarpus</i> species endemic to regions outside New Zealand	6
FIGURE 1.5: Distribution of <i>Corynocarpus laevigatus</i> in New Zealand	9
FIGURE 1.6: Gynodioecy in <i>Corynocarpus laevigatus</i>	11
FIGURE 1.7: (A) Kererū eating karaka fruit; (B) Ripe fruit of tawapou	13
FIGURE 1.8: Cassowary dung containing the large seeds of <i>Elaeocarpus bancroftii</i>	16
FIGURE 1.9: Map of the Pacific region	27
FIGURE 1.10: Map of New Zealand	28
FIGURE 2.1: Distribution of <i>Corynocarpus laevigatus</i> in New Zealand	47
FIGURE 2.2: NEIGHBORNET splits graph of aligned ITS DNA sequences	54
FIGURE 2.3: NEIGHBORNET splits graph of aligned WAXY sequences	55
FIGURE 2.4: NEIGHBORNET splits graph of aligned <i>rbcL</i> sequences (no sequences removed)	56
FIGURE 2.5: NEIGHBORNET splits graph of aligned <i>rbcL</i> sequences (two sequences removed)	57
FIGURE 2.6: NEIGHBORNET splits graph of aligned <i>trnL-trnF</i> sequences	59
FIGURE 4.1: Methodology used to identify SNP markers	92
FIGURE 4.2: Distribution of <i>Corynocarpus laevigatus</i> in New Zealand	94
FIGURE 4.3: The chloroplast genome of karaka showing putative SNPs	102
FIGURE 4.4A: High Resolution Melting analysis of SNPS 1, 3 and 8	105
FIGURE 4.4B: High Resolution Melting analysis of SNPs 16, 41 and 49	106
FIGURE 4.5: NEIGHBORNET splits graph of karaka chlorotypes	108
FIGURE 4.6: Distribution and genetic variation in karaka	111
FIGURE 4.7: Distribution of cultural and non-cultural karaka	112
FIGURE 4.8: Circos plot of karaka chloroplast genome	117
FIGURE 5.1: Double digest RAD sequencing methodology	133



## LIST OF TABLES

---

TABLE 1.1: Seed dispersing birds in New Zealand forests.	15
TABLE 2.1: Fifty variable sites defining <i>Corynocarpus ITS</i> haplotypes, with their sequence alignment position indicated	50
TABLE 2.2: Twenty four variable sites defining <i>Corynocarpus WAXY</i> haplotypes, with their sequence alignment position indicated.	51
TABLE 2.3: Twenty eight variable sites defining <i>Corynocarpus rbcL</i> haplotypes, with their sequence alignment position indicated.	52
TABLE 2.4: Twelve variable sites defining <i>Corynocarpus trnL-trnF</i> haplotypes, with their sequence alignment position indicated	53
TABLE 2.5: Incompatible parsimony site patterns in the <i>rbcL</i> alignment.	58
TABLE 4.1: Geographic location of six karaka samples sequenced with universal primers.	95
TABLE 4.2: An evaluation of the concordance of HRM profiling with Sanger sequencing	104
TABLE 4.3: Summary of chloroplast polymorphisms distinguishing chlorotypes.	107
TABLE A2.1: List of karaka accessions	141
TABLE A3.1: List of sequencing primers	155
TABLE A4.1: SNP marker development table	159
TABLE A5.1: Comparison of genotyping results using HRM and Sanger sequencing	163
TABLE A6.1: Full data set of 360 accessions genotyped with seven SNPs	165