

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.

# *The Chloroplast Genome of Arthropodium bifurcatum.*

---



*A thesis presented in partial fulfillment of the requirements  
for the degree of  
Master of Science in Biological Sciences  
Massey University  
Palmerston North, New Zealand*

***Simon James Lethbridge Cox  
2010***

---

# Abstract

---

This thesis describes the application of high throughput (Illumina) short read sequencing and analyses to obtain the chloroplast genome sequence of *Arthropodium bifurcatum* and chloroplast genome markers for future testing of hypotheses that explain geographic distributions of Rengarenga – the name Maori give to species of *Arthropodium* in New Zealand.

It has been proposed that *A.cirratum* was translocated from regions in the north of New Zealand to zones further south due to its value as a food crop. In order to develop markers to test this hypothesis, the chloroplast genome of the closely related *A.bifurcatum* was sequenced and annotated. A range of tools were used to handle the large quantities of data produced by the Illumina GAIIx. Programs included the de novo assembler Velvet, alignment tools BWA and Bowtie, the viewer Tablet and the quality control program SolexaQA.

The *A.bifurcatum* genome was then used as a reference to align long range PCR products amplified from multiple accessions of *A.cirratum* and *A.bifurcatum* sampled from a range of geographic locations. From this alignment variable SNP markers were identified.

Putative gene annotations for *A.bifurcatum* were compared to published chloroplast genomes from the Magnoliids and Monocot clades. Interesting similarities and differences have been detected and these have been discussed.

# *Acknowledgements*

---

I would like to extend many thanks to many people-my supervisors- Lara Shepherd and Pete Lockhart, my co-chloroplast extractor Robin Atherton. Thanks to Trish McLenachan and Olga Kardailsky for their help in the lab and for matters chemical. To the many people on the fifth floor of Science Tower D for answering my questions- Patrick Biggs, Lesley Collins, Bennet McComish, Jing Wang, Murray Cox, Ibrar Ahmed, Nicole Grunheit & Oliver Deusch. Also thanks to Lorraine Cook and Richard Fong from the genome service for help in understanding the Illumina process.

Thanks also goes to the people who took the time to gather Rengarenga, as well as the aforementioned Lara and Robin, a big thank you goes to P. de Lange, P.A. Aspin, J. Collins, J. Hobbs, D. Blake, J. Rolfe and R. Stone.

Thank you to Rob Hallam for the hours of formatting he saved me.

Thank you to Rititia for her love and support and to my parents Michael and Jenny.

---

# Contents

---

Abstract .....	2
Acknowledgements .....	3
Tables.....	6
Figures .....	7
Introduction.....	8
Project Outline.....	8
Background.....	9
Translocation .....	9
Molecular marker development .....	10
Taxonomy .....	11
<i>Arthropodium</i> .....	12
<i>Arthropodium</i> in New Zealand .....	12
Biogeography.....	13
Ecology .....	14
Reproductive Biology.....	14
Distribution of New Zealand <i>Arthropodium</i> .....	15
Maori, Rengarenga and Oral History .....	16
Next Generation Sequencing (NGS) .....	19
Materials and Methods .....	21
Samples and Collection .....	21
Genomic DNA Extraction Protocol .....	21
Chloroplast Isolation Protocol .....	21
Multiply-primed rolling circle amplification .....	22
EcoR1 Digestion.....	23
Illumina GAII Sequencing.....	23
Evaluation of Raw Read Data Quality.....	24
Assembly of Chloroplast Contigs and Genome .....	25
PCR and Long-range PCR .....	27
Sequencing and Mapping of PCR Products to <i>A.bifurcatum</i> Chloroplast Genome.....	28
Results .....	29
Chloroplast preparation .....	29
Sucrose gradient.....	29
Chloroplast visualization on agarose gel .....	29

SolexaQA and de Novo Assembly.....	30
SolexaQA .....	30
de Novo Assembly .....	33
Annotated <i>A.bifurcatum</i> .....	35
<i>A.bifurcatum</i> Chloroplast Genome Annotated .....	35
<i>A.bifurcatum</i> Chloroplast Genome with SNP's.....	36
Discussion .....	39
SolexaQA and de Novo Assembly.....	39
SolexaQA .....	39
de Novo Assembly .....	39
Chloroplast Comparisons .....	41
Interesting Features .....	42
SNP's .....	44
Illumina Sequencing .....	46
Future Steps .....	48
Bibliography.....	49
Appendix.....	57
Table 1 – Location of Accessions .....	57
Table 2 – Long-range Primers.....	60
Table 3 – Maximising Coverage Cutoff.....	61
Coverage Cutoff for Kmer=47.....	61
Coverage Cutoff for Kmer=49.....	62
Table 4 – List of accessions that were used in long range PCR .....	64
Table 5- SNP Positions .....	65
Figure 1- Alignment of <i>A. bifurcatum</i> , NC 0007499 <i>Phalaenopsis aphrodite</i> subsp. <i>formosana</i> and NC 014056 <i>Oncidium Gower Ramsey</i> . .....	67
Table 7- SNP's in coding regions of genes .....	68

---

# Tables

---

(Excluding Appendix)

Table 1: Shows short range primers used to fill the gaps.....	27
Table 2: Shows results from chloroplast distribution in sugar gradients. ....	29
Table 3: Shows results from differing kmer lengths. ....	33
Table 4: Shows long range PCR products that successfully amplified. ....	37
Table 5: Whole chloroplast sequences most closely related to that obtained from <i>A.bifurcatum</i> as discovered through the NCBI database. ....	42
Table 6: Shows the presence/absence and orientation of dehydrogenase subunit genes in <i>Phalaenopsis aphrodite subsp. formosana</i> , <i>Oncidium Gower Ramsey</i> and <i>A.bifurcatum</i> .....	43
Table 7: Location of <i>Arthropodium</i> SNPs in putative hot spot regionsBorsch & Quandt, 2009.....	45
Table 8: The number of SNP's in 5,000 &10,000 bp zones into the single copy regions from the 4 junction regions. ....	45
Table 9: Studies using high throughput sequencing for SNP discovery.....	47

---

# Figures

---

(Excluding Appendix)

Figure 1: Map of New Zealand showing the distribution of <i>A.cirratum</i> (left) and <i>A.bifurcatum</i> (right) based on herbarium records and personal observations (L.Shepherd). ....	9
Figure 2: Drawing of a panel depicting the curved anthers typical of Rengarenga (Harris, 1996).....	17
Figure 3: A simplified schematic of sequence construction using Velvet (Zerbino & Birney, 2008). ....	26
Figure 4: Shows the results of EcoR1 digestion. ....	30
Figure 5: Heat Map – shows quality degenerating near the end of the reads .....	31
Figure 6: Nucleotide mean quality .....	32
Figure 7: Distribution of longest reads that pass quality assessment .....	32
Figure 8: Shows the alignment of the supercontigs to reference genome NC 008326.....	34
Figure 9: Shows the annotated <i>A.bifurcatum</i> genome .....	35
Figure 10: Example of a putative SNP visualized using Tablet.....	36
Figure 11: shows <i>A.bifurcatum</i> genome with SNP's .....	38
Figure 12: Shows alignment of the rpl23 genes of <i>A. bifurcatum</i> , NC 0007499 <i>Phalaenopsis aphrodite subsp. formosana</i> and NC 014056 <i>Oncidium Gower Ramsey</i> .....	41
Figure 13: Comparison of SNP transitions and transversions in 18 accessions of <i>Arthropodium</i> chloroplast genomes.....	45
Figure 14: Relative coverage of mapped reads to the <i>A.bifurcatum</i> genome visualized in Tablet.....	46