Imperial College London

Uncovering The Mechanisms Of Sympatric Speciation In The *Howea* Palms

Matthew James Coathup

2023

A thesis submitted for the degree of Doctor of Philosophy

Department of Life Sciences,

Imperial College London

Abstract

The *Howea* palms, comprising the sister species *H. belmoreana* and *H. forsteriana* — endemic to Lord Howe Island (LHI) — represent perhaps the strongest known example of sympatric speciation; a contested area of evolutionary biology whereby an ancestral population diverges into two or more reproductively isolated species in the presence of gene flow. Whilst theoretical studies suggest that speciation can occur in sympatry, very few convincing examples exist in nature, and strong empirical evidence of the specific mechanisms involved in facilitating sympatric speciation is even more limited. In this thesis, I explore such mechanisms in the *Howea* palms.

Phenotyping experiments using thousands of *Arabidopsis thaliana* plants with knockouts of candidate *Howea* 'speciation genes' revealed multiple loci with the potential to have driven sympatric speciation. Secondly, differential expression analysis of the two *Howea* species and their F1 hybrids presented strong evidence of hybrid misexpression. Juvenile hybrids are found at a greater frequency than adults, and such misexpression gives a viable explanation for selection against hybrids, reinforcing the speciation process through increased reproductive isolation.

Finally, root symbioses of parental species and F1 hybrids with arbuscular mycorrhizal fungi (AMF) were investigated, as previous work has indicated that AMF associations may have contributed to niche divergence. Hybrid palms were found to differ significantly in beta diversity from *H. forsteriana* growing on calcareous soils. Additionally, correlation analyses of gene expression against AMF alpha diversity identified 1,522 genes in root tissues with a Spearman Correlation Coefficient >0.8. The subset of these genes with a positive correlation was significantly enriched for 18 gene ontology terms including 'response to rapamycin', with target of rapamycin (*TOR*) genes identified as potentially being key to AMF symbioses in *Howea*.

Together, results provide strong evidence of specific mechanisms which could have facilitated initial sympatric divergence in *Howea*, and driven speciation to completion by reinforcement.

Table of Contents

List of Figures 7

List of Tables 9

Copyright Declaration 10

Declaration of Originality 11

Acknowledgements 12

Chapter One - Introduction

1.1 Introduction and Literature Review 14

1.1.1 Questions in Speciation Biology 14

1.1.2 The *Howea* Palms and Sympatric Speciation 17

1.1.3 The Lord Howe Island Scenario 19

1.1.4 Reproductive Isolation in *Howea* 20

1.1.5 Mechanisms of Ecological Speciation 22

1.2 Aims and Outline of the Thesis 26

1.2.1 Broad Aims 26

1.2.2 Chapter Two 27

1.2.3 Chapter Three 28

1.2.4 Chapter Four 28

1.2.5 Chapter Five 29

Chapter Two - Investigating Pleiotropic Speciation Genes in the *Howea* Palms

2.1 Introduction 32

2.1.1 The Search for Magic Traits 32

2.1.2 Candidate *Howea* Speciation Genes 33

2.1.3 Large-Scale Phenotyping of Speciation Genes 33

2.2 Methods 34

2.2.1 Plant Material 38

2.2.2 Growth Conditions 38
2.2.3 Stress Treatments 39
2.2.4 Experimental Setup 39
2.2.5 Plant Phenotyping 40
2.2.6 Statistical Analysis 41
2.3 Results 42
2.3.1 Flowering Time 42
2.3.2 Phenotypic Variation 43
2.3.3 Knockout Stress Response 46
2.4 Discussion 51
2.4.1 Strong Evidence of Multiple Magic Trait Speciation Genes 51
2.4.2 Evaluation of Candidate Speciation Genes 51
2.4.3 *Howea* Homologs of *SAL1* and *SIZ1* Present the Strongest Evidence for Major Effect Magic Traits During LHI Speciation 58
2.4.4 Caveats and Follow-up Research 60
2.4.5 Conclusion and Wider Context 61

Chapter Three - Hybrid Fitness, Misexpression, and Genomic Incompatibility
3.1 Introduction 63
3.1.1 The Role of Hybrid Inviability During Speciation With Gene Flow 63
3.1.2 Post-Zygotic Isolation in the *Howea* Palms 65
3.1.3 Investigating Hybrid Misexpression in *Howea* 66
3.2 Methods 67
3.2.1 Tissue Sampling and RNA Sequencing 67
3.2.2 Bioinformatic Processing 67
3.2.3 Classification of Hybrids 68
3.2.4 Differential Expression Analysis 68
3.2.5 Hybrid Transgressive Expression 68
3.2.6 GO Enrichment Analysis 69

3.3 Results 70
3.3.1 Gene Expression Profiles 70
3.3.2 Differential Expression 71
3.3.3 GO Enrichment of Differentially Expressed Genes 73
3.3.4 Hybrid Transgressive Expression 75
3.4 Discussion 77
3.4.1 Differentially Expressed Genes are Up-regulated in Hybrids 77
3.4.2 Transgressive Hybrid Expression at a Relatively Small Number of Loci Associated with Development and Stress Response 78
3.4.3 Differential Expression of Genes Involves in Arbuscular Mycorrhizal Symbiosis Associated with Soil Type in *H. forsteriana* 80
3.4.4 Conclusion and Wider Context 81

Chapter Four - The Role of Arbuscular Mycorrhizal Fungi in Sympatric Niche Divergence
4.1 Introduction 83
4.1.1 Niche Differentiation in the *Howea* Palms 83
4.1.2 Arbuscular Mycorrhizal Fungi Promote Coexistence 84
4.1.3 The Role of AMF in *Howea* Warrants Further Study 85
4.2 Methods 87
4.2.1 Tissue Sampling and RNA-seq Libraries 87
4.2.2 DNA Extraction and Sequencing 87
4.2.3 Bioinformatic Processing 89
4.2.4 Statistical Analysis of AMF Communities 89
4.2.5 Correlation of Gene Expression and AMF Alpha Diversity 90
4.2.6 GO Enrichment Analysis 91
4.3 Results 91
4.3.1 Sequencing Output and Taxon Assignment 91
4.3.2 Arbuscular Mycorrhizal Fungi Diversity 94
4.3.3 Differentially Abundant ESVs 95

4.3.4 Correlation of Gene Expression and AMF Alpha Diversity 96
4.3.5 GO Enrichment Analysis 98
4.4 Discussion 99
4.4.1 No Evidence for Depleted AMF in *H. forsteriana* on Volcanic Soils 99
4.4.2 AMF Communities in Hybrid Palms Differ from *H. forsteriana* Growing on Calcareous Soils 101
4.4.3 AMF Symbiosis Strongly Affects Gene Expression in Root Tissues 102
4.4.4 Further Work Investigating AMF Diversity in *Howea* is Needed 103
4.4.5 Conclusion and Wider Context 105

Chapter Five - General Discussion
5.1 Key Findings and Significance 107
5.2 Limitations and the Scope for Future Work 112
5.3 Concluding Remarks 114

Bibliography 116
Appendices 137
Appendix I. Chapter Two Supplementary Materials 137
Appendix II. Chapter Three Supplementary Materials 158
Appendix III. Chapter Four Supplementary Materials 161
Appendix IV. Copyright Licenses 163

List of Figures

Figure 1.1. A) Geological Map of Lord Howe Island. B) Photographs of *Howea* palms, including *H. belmoreana*, *H. forsteriana*, and a hybrid 18

Figure 1.2. A) Hypothesised speciation scenario in *Howea.* B) Alternative genetic mechanisms in sympatric speciation 21

Figure 1.3. Photograph of *Howea* palms growing on Lord Howe Island, including a putative hybrid alongside parental species 30

Figure 1.4. Photograph of Lord Howe Island 30

Figure 1.5. Photograph of Lord Howe Island 31

Figure 1.6. Photograph of Lord Howe Island 31

Figure 2.1. Photograph of *A. thaliana* plants inside the growth chamber at the National Plant Phenotyping Infrastructure at the University of Helsinki 35

Figure 2.2. Representative phenotypes of wild type and knockout mutant plants under different stress treatments 44

Figure 2.3. Time-series measurements of phenotyping parameters for *sal1* knockouts and wild type plants during stress treatments 44

Figure 3.1. Clustering of *Howea* species groups by gene expression profile revealed by Principal Component Analysis 70

Figure 3.2. Heat maps of differentially expressed genes in different *Howea* tissues 71

Figure 3.3. Numbers of differentially expressed genes up and down-regulated in each *Howea* group pairwise comparison 72

Figure 3.4. Numbers of significantly enriched GO terms for differentially expressed genes in each *Howea* group pairwise comparison 74

Figure 4.1. Rarefaction curves for *Howea* metabarcoding libraries 92

Figure 4.2. Family-level AMF community composition for *Howea* metabarcoding libraries 92

Figure 4.3. Bray-Curtis dissimilarity MDS plots for each *Howea* pairwise comparison 93

Figure 4.4. Alpha diversity plots of A) Shannon and B) Gini-Simpson indices for each *Howea* species group 94

Figure 4.5. Histograms showing the frequency distribution of Spearman correlation coefficients of gene expression versus AMF alpha diversity in each *Howea* tissue 97

List of Tables

Table 2.1. List of *Howea* candidate speciation genes 37

Table 2.2. Estimated least squares means flowering time statistics for wild type and knockout mutant plants 42

Table 2.3. Test statistics for ANOVA interactions tests of QY Max measurements between wild type and knockout mutants under stress conditions 48

Table 2.4. Test statistics for ANOVA interactions tests of Mean Leaf Temperature measurements between wild type and knockout mutants under stress conditions 49

Table 2.5. Test statistics for ANOVA interactions tests of Rosette Area measurements between wild type and knockout mutants under stress conditions 50

Table 4.1. Numbers of differentially abundant AMF ESVs in each *Howea* species group pairwise comparison 95

Table 5.1. Selected putative examples of magic traits involved in speciation 108

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Declaration of Originality

Unless otherwise stated, the work presented in this thesis is entirely my own, with any collaborative work detailed below and/or referenced in the text. Those who contributed less formal advice and assistance are listed in the acknowledgements section overleaf.

In Chapter Two, the day-to-day running and set up of the phenotyping experiments at the University of Helsinki was carried out by Dr. Kristiina Himanen, Dr. Katriina Mouhu and myself. Kristiina Himanen and Katriina Mouhu also offered advice on experimental design. All other preparation for the experiments, including laboratory work, generation of mutant seeds, and the majority of experimental design, was carried out by myself. All statistical analysis was carried out by myself. Dr. Luke Dunning generated *Howea* transcriptomes which contributed to ongoing follow-up work to Chapter Two - this is detailed further in Chapter 2.4.4.

In Chapter Three, I used RNA-seq libraries and transcriptomes first generated by the authors of Dunning et al. (2016) and Osborne et al. (2019). All analysis was carried out by myself.

In Chaper Four, some of the same RNA-seq libraries as above were used, as well as raw tissue samples taken from LHI by the authors of Dunning et al. (2016). Laboratory work was all carried out by myself, with the exception of a sequencing run carried out by Azenta (USA). After sequencing, the initial pipeline of read filtering, quality control and taxon assignment was in collaboration with Jake Williams and Dr. Emma Ransome, who provided the DADA2 workflow. All following analysis was carried out by myself.

Throughout the thesis, all writing, analysis, and creation of figures was carried out by myself (with the exception of two figures adapted from previous publications, which are referenced in the text). All photographs were taken by myself.

Acknowledgements

I wish to thank everybody who has offered help, advice or guidance throughout this project. In particular, my supervisors Prof. Vincent Savolainen and Prof. Colin Turnbull, whose time, knowledge and assistance I am incredibly grateful for. This project certainly wouldn't have been possible without them. Special thanks also to Dr. Luke Dunning and Dr. Owen Osborne, who offered both hands-on help and invaluable technical expertise at various points.

I was incredibly impressed with both Prof. Tim Barraclough and Dr. Giovanni Senna, for their ability to critique, evaluate, and offer outstanding advice on the state of my thesis during progression assessments and checkpoints. They were always able to spot exactly where I knew there were potential holes in the work, and often where I didn't! Thank you also to Prof. Mick Crawley for exceptional statistical knowledge and guidance, and to Jake Williams and Dr. Emma Ransome for their willing help and involvement in metabarcoding analysis.

At various stages, advice and help with laboratory work was exceptionally appreciated, with particular mentions to Prof. Kristiina Himanen, Dr. Katriina Mouhu, and the technical team at the Plant Phenotyping Infrastructure at the University of Helsinki, for all they did to make the very long and arduous phenotyping experiments achievable. Also to Katia Hougaard and Yasin Tumtas for help with molecular cloning, and a particular thank you to the Silwood technicians Dr. Martin Selby and Paul Beales, who set up so many greenhouse rooms and fixed so many pieces of broken equipment for me I could not count.

Special mention must go to everybody on Lord Howe Island who made our field trip so welcoming, including the Lord Howe Island Board, and especially Ian Hutton. Ian, as has been the case for many years, was so happy to go out of his way to make our trip as successful as possible. Notably, by leading us impressively through the jungle to some of the most inaccessible parts of the island!

Finally, the most significant mention must go to the rest of the Savolainen lab group. There are too many people to list names, but they have been an exceptionally intelligent, hilarious, and genuinely good group of people to work alongside. It was an absolute pleasure.

Apologies if I have missed anybody from this list, there are so many people who have offered help during this project that I undoubtedly will have done so.

This study was funded by a BBSRC DTP Studentship. Thank you also to the European Plant Phenotyping Network, who provided an EPPN2020 Transnational Access Grant, funding my trip to Helsinki and all of the phenotyping experiments carried out there.

Chapter One - Introduction

1.1 Introduction and Literature Review

1.1.1 Questions in Speciation Biology

Mechanisms of speciation, and the conditions required for these mechanisms to occur, remain highly debated areas of evolutionary biology. Since Darwin first wrote about the evolution of new species through natural selection and adaptive divergence (Darwin, 1859), biologists have theorised over the precise scenarios in which speciation occurs. Within this field, perhaps the most prominent topic of discussion has been the role of gene flow, or lack thereof, in the formation of new species.

Gene flow between populations can be described as a continuum, ranging from complete separation with no exchange of genetic information, through to complete panmixis, where all individuals in a population are free to interbreed randomly and uniformly. Species divergence in the former scenario, known as allopatric speciation, is often argued to be the most common speciation scenario, and regarded as the null model of speciation. Allopatric speciation is typically characterised by a geographical divide arising between populations of an ancestral species, for example following the formation of a mountain range or river. Such extrinsic prezygotic barriers to reproduction prohibit gene flow, and in the newly isolated populations, the combination of novel mutations, different selective pressures, and genetic drift leads to divergence of the gene pools, eventually leading to speciation.

The adaptive radiation of the Galapagos Finches, also known as Darwin's Finches, provides classical examples of allopatric speciation. This group of almost 20 finch-like species is endemic to the Galapagos Islands (with the exception of one species which is restricted to Cocos Island in the Pacific) (Grant, 1999; Sato et al., 2001), where an adaptive radiation involving the evolution of a wide array of beak sizes and morphologies has driven species divergence in populations isolated across different islands and niches (Grant, 1999; Schluter, 2000). The variation in beak morphology is known to have been driven by natural selection in response to differing ecological conditions such as food availability, and is associated with the

diverse range of feeding niches the species occupy (Podos & Nowicki, 2004; Podos et al., 2004; Grant & Grant, 2006).

Intriguingly, Darwin's Finches also play host to an additional form of speciation, known as homoploid hybrid speciation. This term describes the formation of a new species by hybridisation between two existing species, without a change in chromosome number. Such hybridisation occurs in plants, but is also known to have occurred in a small range of animals, including an unnamed Darwin's Finch species found on the island of Daphne Major (Lamichhaney et al., 2018). This form was likely established by hybridisation between a male Española cactus finch from Española Island and a female medium ground finch from Daphne Major, which appears to have generated almost immediate reproductive isolation from the parental species over the course of just a few generations, and seems to be ecologically viable (Lamichhaney et al., 2018).

Speciation by hybridisation can also involve genome duplication, known as polyploid speciation. A range of potential mechanisms exists for the generation of polyploidy, but when it occurs as a result of hybridisation between two distinct taxa, it is known as allopolyploidy. Mechanistically, this most often occurs due to a hybrid inheriting unreduced gametes from a parent, perhaps as a result of chromosome duplication (Ramsey, 1998). Without this duplication, many F1 hybrids between genetically diverged taxa face issues in the pairing of chromosomes during meiosis; however, inheriting duplicate genomes from parents can actually restore the ability to undergo bivalent meiotic pairing, by providing duplicated chromosomes with their own homologues (Ramsey, 1998). F1 hybrids may then be able to produce fertile gametes with a greater ploidy level than either of the parental species, generating immediate reproductive isolation. Allopolyploid speciation is relatively common in plants, with many species of wheat being an example. Indeed, classic bread wheat is a hexaploid (six sets of chromosomes) hybrid between emmer wheat (itself a tetraploid hybrid; four sets of chromosomes) and a diploid species of goatgrass, and evolved in farmers' fields in the 'Fertile Crescent' of the Middle East (Gohar et al., 2022).

Stepping away from speciation by hybridisation and returning to the gene flow continuum described previously, if we move along the continuum away from allopatric speciation we encounter parapatric speciation. This term describes the situation in which a

sub-population colonises an adjacent niche, leading to largely non-overlapping ranges with a narrow contact zone. In these scenarios, there is a degree of ongoing gene flow, but the gene flow is unequal, with non-random mating (Slatkin, 1982). If gene flow can be interrupted enough to allow stable and largely reproductively isolated populations to diverge, speciation can take place, perhaps associated with a hybrid zone where contact remains. Arguably the most well-known example of ongoing parapatric speciation is in populations of the grass species *Agrostis tenui.* Adjacent populations of the species near derelict mine workings inhabit different soils, with some individuals able to tolerate soils contaminated with heavy metals, which has been shown to be a heritable trait (McNeilly, 1968; McNeilly & Bradshaw, 1968). However, these individuals grow more poorly on normal soils than individuals without heavy metal tolerance. Whilst there is ongoing gene flow between populations, hybridisation is limited by differences in flowering time between locations (McNeilly & Antonovics, 1968). Over time, this divergence may lead to increasing reproductive isolation and the completion of speciation.

Progressing yet further along gene flow continuum, at the far end we reach sympatric speciation — a contested form of speciation and the focus of this thesis — which involves divergence of an ancestral population into multiple new species without any geographical separation of the incipient lineages. It is characterised by overlapping ranges and random mating, with ongoing gene flow during the initial stages of speciation, but its prevalence remains a significant area of debate.

As a result of this continuing debate, in recent years there has been a large focus on understanding the genetic bases of speciation (Wu, 2001; Butlin, 2008; Jones et al., 2012; Cruickshank & Hahn, 2014; Riesch et al., 2017), and on establishing clarity over the prevalence of speciation in the presence of gene flow (Coyne & Orr, 2004; Barluenga et al., 2006; Savolainen et al., 2006a; Nosil, 2008; Smadja & Butlin, 2011; Martin et al., 2013; Osborne et al., 2013). Such questions are tightly interlinked, as whilst theoretical studies suggest that speciation can occur in sympatry with initial gene flow (Dieckmann & Doebeli, 1999; Kondrashov & Kondrashov, 1999; Doebeli et al., 2005; Bolnick & Fitzpatrick, 2007; Gavrilets & Vose, 2007), few convincing examples exist in nature. It is, therefore, important to investigate the mechanisms by which adaptive divergence can generate reproductive isolation,

and how such mechanisms could facilitate speciation with gene flow in the absence of a geographical barrier.

Indeed, divergent selection has the potential to lead to reproductive isolation, even in the presence of gene flow, when a genetic mechanism exists to link the trait(s) under selection with assortative mating (Schluter, 2000, 2001, 2009; Kirkpatrick & Ravigné, 2002; Doebeli et al., 2005; Rundle & Nosil, 2005; Nosil, 2012; Feder et al., 2012). One scenario in which this may occur is that of ecological speciation, whereby initial local adaption to different environments can lead to divergent selection and ultimately reproductive isolation (McKinnon et al., 2004; Rundle & Nosil, 2005; Funk et al., 2006; Schluter, 2009; Nosil, 2012; Osborne et al., 2013). There are three broad mechanisms by which this could occur. Firstly, colonisation of a new environment may lead to local adaptation of a trait at a single locus, which pleiotropically also affects reproductive traits. So-called 'magic traits' (see Figure 1.2), which lead to adaptive divergence and contribute to non-random mating, may be frequently involved in speciation processes (Gavrilets, 2004; Servedio et al., 2011). Secondly, loci involved in adaptive divergence in an incipient species may be non-randomly associated with reproductive traits through linkage disequilibrium (Servedio et al., 2011, Figure 1.2). Finally, colonisation of a new environment may trigger a plastic shift in a trait associated with non-random mating, leading to an immediate reduction in gene flow between populations. For example, substrate concentrations of cadmium and zinc have been shown to alter flowering time in *Arabidopsis*; such a plastic shift in flowering time associated with soil type may be common in plants (Levin, 2009).

1.1.2 The *Howea* Palms and Sympatric Speciation

One likely candidate for ecological speciation is the *Howea* palm genus, where divergence between the two extant species appears to have been the result of adaptation to soil type and an associated shift in flowering time (Savolainen et al., 2006a). The two species, *Howea belmoreana* (the curly palm) and *Howea forsteriana* (the kentia or thatch palm), also represent one of the strongest known cases of homoploid sympatric speciation (Savolainen et al., 2006a; Gavrilets & Vose, 2007), and due to the nature of the tiny Lord Howe Island on which they are endemic, offer an ideal system for the study of various components of the sympatric speciation process.

Coyne and Orr (2004) set out four criteria which must be satisfied to identify potential cases of sympatric speciation: 1) the species must be currently sympatric, 2) they must be sister taxa, 3) they must be reproductively isolated, and 4) an initial period of allopatry during divergence must be unlikely. Many potential examples of sympatric speciation fail to satisfy all four criteria; for example, speciation events via plant-host-shift in Hawthorn flies (Forbes et al., 2009), and rapid divergence in three-spined sticklebacks (Marques et al., 2016), fail to rule out a previous allopatric phase. Until recently, some of the strongest examples of sympatric speciation were those of crater lake cichlids (Schwliewen et al., 1994; Barluenga et al., 2006;

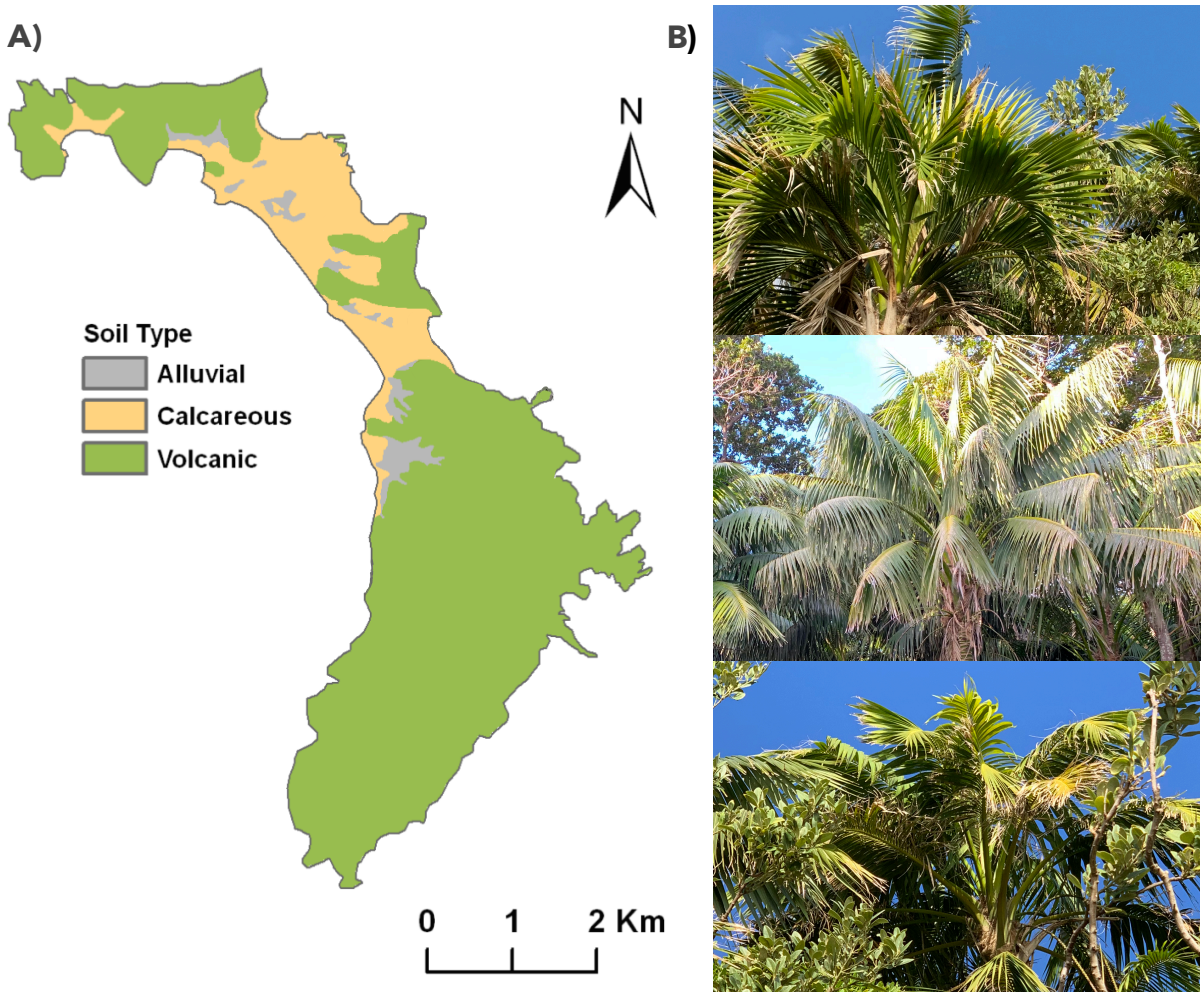


Figure 1.1. A) Geological map of Lord Howe Island. Map shows the distribution of soil types on the island, with volcanic soils in green, calcareous soils in yellow, and alluvial soils in grey. Adapted from Osborne et al. (2018). Image reproduced with permission of the rights holder, Oxford University Press. **B)** From top to bottom, photographs showing palms of *H. belmoreana*, *H. forsteriana*, and a hybrid between the two species, growing on Lord Howe Island. Notice the curly palm, *H. belmoreana,* is characterised by recurved fronds with upturned leaflets. It also bears a single spike in each inflorescence. Meanwhile, the kentia palm, *H. forsteriana*, possesses straight leaves and downturned leaflets, and has multiple spikes in each inflorescence. Hybrid palms typically bear flatter fronds that are not consistently up or downturned, and often appear noticeably ragged.

Malinsky et al., 2015); however new evidence has cast strong doubt upon true sister-relationships in cichlid radiations, with multiple instances of recolonisation and introgression through secondary contact likely (Malinsky et al, 2015; Martin et al., 2015; Kautt et al., 2016; Meier et al., 2017; Poelstra et al., 2018, Kautt et al., 2020). In Cameroon, all four cichlid radiations involved secondary gene flow, with some species shown to be more closely related to river populations than other species in the lake (Martin et al., 2015). In addition, in one particular lake, a genomic region containing olfactory genes involved in mate choice in cichlids was shown to have been introgressed prior to speciation (Poelstra et al., 2018). Furthermore, Richards et al. (2018) performed an in-depth analysis studying introgression and the history of gene flow between the Cameroonian Barombi Mbo crater lake cichlids and nearby riverine populations, and found evidence of a complex history of introgression through secondary contact.

1.1.3 The Lord Howe Island Scenario

In light of such evidence, further controversy has been cast over the prevalence of sympatric speciation in nature, leaving the *Howea* palms as perhaps the most convincing case. The two species, *H. belmoreana* and *H. forsteriana,* are found only on the minute Lord Howe Island (ca. 15km^2), which was formed around 6.9 million years ago (McDougall et al., 1981), and is located in the Tasman Sea 580km from the eastern coast of Australia, which is the closest land mass. Other than Ball's Pyramid, a sea stack located 23km southeast of Lord Howe Island which supports limited plant life (Savolainen et al., 2006a), there are no islands nearby. This geographical setting makes for an ideal system to study sympatric speciation, as the small size of the island, and its isolation from other land masses, make an allopatric phase during speciation unlikely (McDougall et al., 1981; Savolainen et al., 2006a, 2006b; Papadopulos et al., 2014).

Upon Lord Howe Island, *H. belmoreana* and *H. forsteriana* have overlapping distributions, dictated by the two predominant soil types on the island (Figure 1.1). Both *H. belmoreana* and *H. forsteriana* are found on volcanic soils across the island, whereas *H. forsteriana* also inhabits more recently deposited calcareous soils (Brooke et al., 2003), where *H. belmoreana* is absent. The two soil types are highly intercalated, with the species often occurring in sympatry (Pickard, 1983). Furthermore, both species are known to be wind-pollinated, with weak genetic structure (Babik et al., 2009) — this provides more evidence that an allopatric phase

during their speciation, driven by geographical isolation, was highly unlikely. Phylogenetic studies have also shown that the species are sister taxa (Savolainen et al., 2006a, Osborne et al., 2019), and that they likely diverged between 1 and 3.5 million years ago, within the age of Lord Howe Island, (Savolainen et al., 2006a, Osborne et al., 2019; Papadopulos et al., 2019b), with no secondary gene flow into *Howea* following their colonisation of the island (Osborne et al., 2019). Additionally, coalescence-based demographic analyses using ddRAD sequencing support a scenario in which initial divergence in *Howea* occurred in the presence of high gene flow (Papadopulos et al., 2019b).

1.1.4 Reproductive Isolation in *Howea*

Together, the above studies have provided strong evidence to fulfil three of Coyne and Orr's four conditions for sympatric speciation (2004); demonstrating a sister relationship, absence of an allopatric phase during speciation, and that the species currently exist in sympatry. The last condition, namely that the species are reproductively isolated, has been examined in detail at both the pre-zygotic and post-zygotic stages. Savolainen et al. (2006a) found that the two species differ in flowering time, with *H. forsteriana* flowering earlier than *H. belmoreana,* and peak flowering time of each species being separated by approximately 6 weeks. Interestingly, whilst *H. belmoreana* has synchronous peak male and female flowering, *H. forsteriana* is protandrous on calcareous substrates, with male flowers maturing two weeks before peak female receptivity. However, on volcanic substrates, such asynchronous maturation is not observed (Savolainen et al., 2006a). This suggests that soil type may have had a direct impact on flowering time during the speciation process in *Howea.* These flowering time differences have been shown to provide a high level of pre-zygotic isolation, with gene flow reduction from *H. forsteriana* to *H. belmoreana* due to flowering phenologies measured at 97%, and from *H. belmoreana* to *H. fosteriana* at 80% (Hipperson et al., 2016). Asymmetry in reduction of gene flow is due to *H. forsteriana* flowering earlier, and being protandrous (Hipperson et al., 2016). Similar levels of pre-zygotic isolation have been measured between other closely related plant species (Rieseberg & Willis, 2007; Widmer et al., 2009), and such barriers are likely to arise early during the speciation process, and act as greater barriers to gene flow than mechanisms of post-zygotic isolation (Martin & Willis, 2007; Rieseberg & Willis, 2007; Lowry et al., 2008).

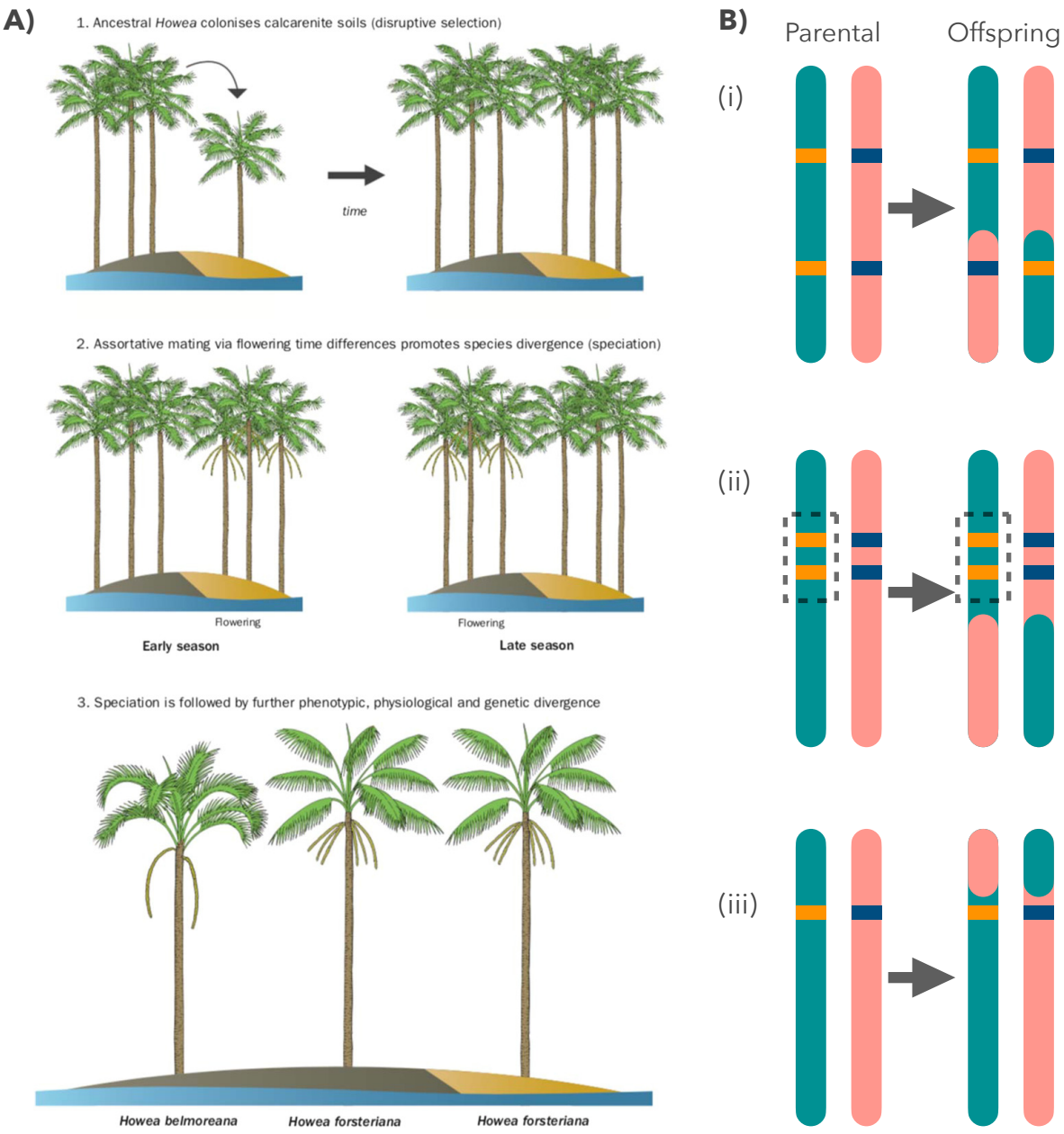


Figure 1.2. A) Speciation scenario in *Howea* palms. Adapted from Papadopulos et al. (2019a). Image reproduced with permission of the rights holder, Oxford University Press. (1) Ancestral *Howea* colonises calcarenite deposits, resulting in disruptive selection via adaptation to environmental stresses and triggering flowering time differences. (2) Assortative mating via displacement of flowering phenology promotes reproductive isolation. (3) Further divergence arises after speciation. **B)** Alternative genetic mechanisms of sympatric speciation in *Howea* palms. (i) Alleles for stress tolerance and altered flowering time (orange) are located at independent loci in parental chromosome. Recombination (grey arrow) breaks up the allele combination in gametes, homogenising populations and preventing speciation. (ii) Alleles for stress tolerance and altered flowering time (orange) are located at tightly linked loci (linkage disequilibrium). Recombination fails to segregate alleles, leading to divergence and potential sympatric speciation. (iii) A single allele (orange) pleiotropically confers stress tolerance and flowering time displacement (magic trait). Traits cannot be segregated during meiosis, leading to ecological divergence and potential sympatric speciation.

Despite this, pre-zygotic isolation is not complete in *Howea*, with hybrids existing at very low frequencies (Savolainen et al., 2006a; Babik et al., 2009; Papadopolous et al., 2013; Hipperson et al., 2016). Of the low number of hybrids detected, seedlings occur at a higher frequency than adult trees, suggesting that there is also post-zygotic isolation in *Howea,* with selection acting between the seedling and adult stages against hybrids*;* though this isolation is incomplete (Hipperson et al., 2016). Between this evidence, and the pre-zygotic barriers enforced by flowering time differences, is it clear that strong reproductive isolation exists, satisfying the last of Coyne and Orr's criteria for sympatric speciation (2004).

1.1.5 Mechanism of Ecological Speciation

What these criteria do not cover, however, is the mechanism by which ecological speciation in sympatry may have occurred in *Howea*. As aforementioned, the two species are distributed based upon soil type, with *H. belmoreana* found only on volcanic substrates, whilst *H. forsteriana* is found on both volcanic and calcareous soils. These soils have been found to differ in a range of characteristics: for example, calcarenite has higher concentrations of arsenic, boron, calcium, cadmium, sodium, phosphorous and sulfur than volcanic soil, and also possesses a higher pH and lower water content than volcanic soil (Osborne et al., 2019). Conversely, volcanic soils contain higher concentrations of aluminium, cobalt, chromium, copper, iron, potassium, manganese, nickel and zinc (Osborne et al., 2019). It is hypothesised that colonisation of calcarenite by *H. forsteriana* may have led to divergence from *H. belmoreana* through local adaptation to exogenous edaphic stress (Papadopolous et al., 2013; see Figure 1.2), which was also associated with a shift in flowering time, generating immediate reproductive isolation (Savolainen et al., 2006a; Dunning et al., 2016; Hipperson et al., 2016). This could have occurred through the mechanisms mentioned previously; with pleiotropy, linkage disequilibrium, or plasticity triggering a flowering time shift. Post-zygotic isolation, facilitated by reinforcement through selection against hybrids (Hipperson et al., 2016), may have strengthened the speciation process.

There are, in fact, several lines of evidence to support such a scenario. Firstly, Savolainen et al. (2006a), used amplified fragment-length polymorphism loci to show that genetic divergence (F_{ST}) is low across most loci between *H. belmoreana* and *H. forsteriana*, with only 4 of 274 sampled markers showing divergence significantly stronger than that expected under neutrality. These four loci were also the only markers that were fixed in the two species

— consistent with local adaptation in sympatric speciation, where differentiation is expected to be far greater for a small number of loci under divergent selection than for the majority of loci in neutral regions (Wu, 2001). More rigorous analysis carried out by Papadopulos et al. (2019a), using ddRAD genotyping to identify genomic islands of differentiation, supports this finding. In this study, the authors identified 38 genomic islands with high levels of F_{ST}, covering 3.3% of the genome; and only eight islands with elevated levels of both F_{ST} and d_{XY}, which accounted for 0.8% of the genome. It is thought that during speciation with gene flow, loci under divergent selection would act as 'barrier loci', resisting the effect of gene flow which continues to homogenise the remainder of the genome (Nosil et al., 2008; Soria-Carrasco et al., 2014). As a consequence, elevated divergence may be expected in regions containing barrier loci (Martin et al. 2013; Renaut et al. 2013; Poelstra et al. 2014).

There is also a range of evidence showing that edaphic preferences, and adaption to substrates, differ strongly between the species. Notably, adult hybrids have not been identified inhabiting calcareous soils (Hipperson et al., 2016), and common garden experiments have shown that seedlings of *H. belmoreana* perform as well on calcarenite as on volcanic soils, despite no adult *H. belmoreana* trees being found on calcarenite (Hipperson et al., 2016). This suggests that *H. belmoreana* is a specialist on volcanic soil, with strong selection acting against *H. belmoreana* on calcarenite at a late stage, whilst *H. forsteriana* is better adapted to calcarenite, and is a generalist capable of inhabiting both soil types. One may, therefore, expect *H. forsteriana* to displace *H. belmoreana* completely; however, *H. belmoreana* is the more numerous species on volcanic soils (Savolainen et al., 2006a; Osborne et al., 2018). An analysis of arbuscular mycorrhizal fungi (AMF) symbioses provides a possible explanation, with *H. forsteriana* found to have a significantly lower level of AMF on volcanic soil, relative to both *H. belmoreana* on volcanic soil and *H. forsteriana* on calcarenite (Osborne et al., 2018). Additionally, survival experiments using sterilised soil showed reduced fitness of seedlings in all species-soil combinations, except *H. forsteriana* on volcanic soils; and AMF-associated genes show signatures of divergent selection between *H. belmoreana* and *H. forsteriana*. This provides further evidence for strong local adaption to soil type.

Finally, some of the strongest evidence to support sympatric speciation facilitated by ecological divergence based on soil type, and an associated flowering time change, comes from analyses of differences in gene expression and of coding sequence divergence. Firstly,

the aforementioned genomic islands identified in Papadopulos et al. (2019a) were enriched for gene ontology (GO) terms associated with environmental stresses and flowering time. Secondly, Dunning et al. (2016) used RNA-seq data from a range of individuals and tissues in both species to generate a reference transcriptome, in order to examine differential gene expression and genetic divergence, and investigate which genes were under positive selection. They found that 1320 genes, that is 9% of their reference transcriptome, showed fixed differences between the species (i.e. $F_{ST} = 1$), and that the gene ontologies of these loci were enriched for categories associated with water deprivation, flower development and phosphate starvation, which is linked to pH. At higher pH levels, such as those associated with calcareous soils, phosphorous forms insoluble salts with calcium and magnesium, reducing its availability to plants (Yi et al., 2005). Dunning et al. (2016) also found a range of genes potentially evolving under positive selection, and found that within differentially expressed genes, there were enriched GO terms associated with abiotic stress. Based upon these genes, they then looked for loci with the potential to pleiotropically affect both soil adaptation and flowering time, and found six candidate 'speciation genes' with inferred functions relevant to both criteria. To verify the phenotypic effects of these candidates, *Arabidopsis* knockout mutants were used to measure flowering time differences, with several of the loci found to significantly alter flowering time (Dunning et al., 2016).

There was, however, no investigation of the phenotypic effect of these knockouts on soil adaptation; and a recent, more thorough, assembly of a reference transcriptome has provided greater insight into potential 'speciation genes'. Building on the transcriptome used by Dunning et al. (2016), Osborne et al. (2019) produced a reference transcriptome with 88.7% completeness, compared to 77.5% completeness for that used by Dunning et al. (2016). Similar methods were then applied to look for genes that were either under adaptive selection, or were differentially expressed between the species, with the potential to link ecological adaptation and non-random mating through pleiotropy or plasticity. Specifically, genes were identified as potentially involved in flowering time if the *Arabidopsis thaliana* homologue was in the FLOR-ID flowering time database, or annotated with GO terms related to flower development. Genes potentially involved in soil adaptation were identified through several criteria: firstly, if they had significantly differential expression with regard to soil chemistry in *H. forsteriana*; secondly, if the *A. thaliana* homologues were in the DroughtDB drought gene database; and thirdly, if the genes were annotated with GO terms relevant to abiotic stresses

associated with soil chemistry analysis. This included GO terms relevant to AMF associations, which were shown as likely to affect the fitness of the two species on different soils (Osborne et al., 2018). Genes annotated as potentially involved in both flowering time and soil adaptation were only considered candidate 'speciation genes' if they had published evidence of a mutant flowering phenotype, and either published evidence of a mutant phenotype relating to Lord Howe Island soil type differences, or differential expression in *Howea* according to soil type. Nine genes were identified in total, with two of these candidate genes amongst the six compiled by Dunning et al. (2016). Finally, Osborne et al. (2019) produced a list of 13 'speciation genes' (Table 2.1), based on those identified in both the Osborne et al. (2019) and Dunning et al. (2016) studies.

Investigating the function of these 'speciation genes' provides an opportunity to study the genetic mechanisms by which speciation with gene flow may occur, for which strong empirical evidence is very rare. Indeed, recent reviews (Foote, 2018; Richards et al., 2019) have highlighted a renewed drive to unravel the processes underlying sympatric speciation in the genomic era; and multiple recent studies have attempted to identify specific genetic mechanisms behind sympatric speciation. In the Nicaraguan crater lake cichlids, genome-wide association mapping and quantitative trait locus mapping identified multiple loci associated with polygenic traits related to ecological divergence which may have been involved in sympatric speciation in two lakes (Kautt et al., 2020). However, demographic models again found that the most likely population history scenarios in both lakes involved at least one instance of secondary contact.

Perhaps more compelling evidence for a plausible genetic mechanism of sympatric speciation comes from wild emmer wheat populations, which appear to be undergoing sympatric speciation inside a single canyon in Israel. Here, multiple populations are in the process of diverging, likely driven by similar ecological pressures as the divergence in *Howea* palms, with stress tolerance including fungal disease resistance and irradiance tolerance, alongside divergent flowering time, potentially driving ecological speciation (Wang et al., 2020). In addition, one population possesses a series of homozygous chromosomal rearrangements, which cause post-zygotic reproductive isolation with a sympatrically diverging population, located within 30m, through hybrid sterility (Wang et al., 2020).

Whilst such a chromosomal rearrangement provides an interesting and plausible genetic mechanism for sympatric speciation, strong empirical evidence for the involvement of magic traits in sympatric speciation remains absent. Both Foote (2018) and Richards et al. (2019) reviewing sympatric speciation remarked upon the need to address this:

"A key focus of future studies should be to investigate the concordance of empirical studies with prediction of theoretical models of sympatric speciation. In particular, the role of pleiotropy and so-called magic traits." - Foote (2018).

"Despite extensive searches for examples of sympatric speciation in the wild, there are few convincing case studies due to the difficulty of ruling out historical allopatric scenarios and ruling out a role of introgression in speciation. Furthermore, the role of magic traits or matching vs preference/trait mechanisms is not fully understood in any existing case study. Thus, we still have very limited empirical tests of an extensive theoretical literature and diverse competing models of the notoriously difficult process of sympatric speciation." - Richards et al. (2019).

1.2 Aims and Outline of the Thesis

1.2.1 Broad Aims

The overarching aim of this thesis is to investigate the specific mechanisms which have been involved in facilitating sympatric speciation in the *Howea* palms. With such investigation, the work presented will endeavour to further existing knowledge in two key areas. Firstly, whilst the *Howea* palms already represent a convincing case of sympatric speciation, a knowledge gap remains over the specific detail of how this speciation was initially facilitated, and how it was then driven to completion. Scepticism will remain over the status of *Howea* as a fully convincing example of sympatric speciation until this detail is uncovered. Secondly, as discussed in the opening of this chapter, such questions extend beyond just the *Howea* palms, with debate ongoing about the mechanisms involved in speciation biology as a whole. By demonstrating in depth the processes which have driven speciation on LHI, I aim to advance our wider understanding of how species originate and diverge — a central theme of evolutionary biology. Specifically, the thesis will aim to answer the following questions:

1. Could pleiotropic 'speciation genes', acting as magic traits by facilitating adaptive divergence (in the form of soil stress tolerance) and non-random mating (in the form of altered flowering time), have driven sympatric speciation in the *Howea* palms?
2. Is there evidence of gene misexpression in hybrid palms, which could have led to post-zygotic isolation driven by selection against hybrids, and contributed to the divergence of the *Howea* palms? Such a scenario could have played a mechanistic role in driving speciation to completion through reinforcement.
3. What, if any, mechanistic role did arbuscular mycorrhizal fungi play in the palms' divergence? Is the ability to form AMF associations key to the relative fitness of each *Howea* species on volcanic and calcareous substrates? What AMF associations do hybrid palms form? Does this shed insight on selection against hybrids, and therefore the reproductive isolation of the parental species? Finally, what affect do AMF associations and gene expression in the palms have on one another?

1.2.2 Chapter Two

To answer these questions, the first key focus of my research is to investigate magic traits in *Howea,* with the objective of identifying loci which pleiotropically contribute to adaptive divergence through soil stress tolerance, and simultaneously contribute to non-random mating through flowering time displacement. This is the subject of discussion in Chapter Two.

If such a mechanism can be shown to have been involved in the speciation process in *Howea,* it would not only solidify the argument for sympatric speciation in this genus, but also renew interest in the wider prevalence of sympatric speciation, which may be common even within the fauna of Lord Howe Island itself (Papadopolous et al., 2011; Osborne et al., 2020). The candidate 'speciation genes' identified by Dunning et al. (2016) and Osborne et al. (2019) are relevant to a scenario whereby soil adaptation was linked to a shift in flowering time through pleiotropy, but do not account for genes which may have been involved through linkage disequilibrium. Without detailed genetic mapping and quantitative trait locus information, which is highly laborious (Cruickshank & Hahn, 2014), it would be difficult to disentangle the role of linkage disequilibrium in the *Howea* speciation process. However, investigating the role of pleiotropy is nonetheless highly relevant to the speciation of *Howea*.

To study these candidate genes and verify a potential role in facilitating sympatric speciation, one must demonstrate a pleiotropic affect on both soil adaptation and flowering time under conditions consistent with the LHI speciation scenario. Whilst functional experiments using mutant *Howea* palms would be ideal, the generation time makes this impractical, and so extensive assays using knockout mutants in the model plant species *Arabidopsis thaliana* form the basis of Chapter Two.

1.2.3 Chapter Three

In Chapter Three, I then investigate how reinforcement may have driven speciation to completion via post-zygotic isolation and selection against hybrids. As highlighted by McGirr and Martin (2020), the role of hybrid inviability in generating post-zygotic isolation during speciation with gene flow has not been well-studied, despite the fact that hybrid gene misexpression is expected to disrupt developmental processes and reduce fitness (Signor & Nuzhdin, 2018), and the plethora of studies suggesting that such misexpression may contribute to strong post-zygotic isolation (Landry et al., 2007; Ortíz-Barrientos et al., 2007; Mack et al., 2016). By examining differential gene expression between *Howea* hybrids and the parental species, and identifying genes demonstrating hybrid misexpression — along with functional characterisation of such genes — I aim to uncover the mechanisms contributing to reinforcement of speciation on LHI, where hybrids only exist in very low frequencies, and are more numerous as juveniles than adults (Babik et al., 2009; Hipperson et al., 2016).

1.2.4 Chapter Four

The final area of research presented in this thesis focuses on AMF symbioses in *Howea* palms. As discussed above, previous work has indicated a potentially significant role for AMF in the niche divergence of *Howea*, with *H. forsteriana* palms growing on volcanic soils having depleted levels of AMF relative to both *H. belmoreana* and *H. forsteriana* on calcareous soil (Osborne et al., 2018). This is consistent with a scenario in which ancestral *H. forsteriana* became able to colonise calcarenite, potentially through adaptation in order to optimise AMF symbioses with the native microbiome of calcareous soils, concurrently associated with a reduced ability to form AMF relationships on volcanic soil. Given that the presence of AMF is likely key to plant fitness, including in palms (Moora et al., 2011; Al-Karaki, 2013), such symbioses may be vital in the niche differentiation seen during *Howea* speciation.

In Chapter Four, I expand on the research carried out by Osborne et al. (2018), with three key advancements on this previous work. Firstly, Osborne et al. amplified the ITS2 gene to assess the fungal microbiome in *Howea* roots — whilst ITS primers are commonly used to investigate broad fungal taxonomic richness, they are not optimised for assessment of Glomeromycota (the phylum to which AMF belong). In the present work, I use AMF specific 18s rRNA primers, which are likely to detect significantly more AMF taxa (Suzuki et al., 2020). Secondly, I include hybrid *Howea* palms in metabarcoding analysis, which has not been done previously. As above, hybrids appear on LHI at higher frequencies as juveniles than adults, suggesting selection against them. Here, I explore whether AMF community composition differs in hybrid palms relative to either of the parental species, which could provide insight into hybrid fitness. Finally, I combine gene expression data with AMF metagenomics, analysing the relationships between expression levels in different *Howea* tissues with AMF diversity. Through this, I aim to identify how AMF community composition and host plant gene expression affect one another, and to identify key loci involved in the formation of AMF symbioses in *Howea*.

1.2.5 Chapter Five

Having explored the novel research contributing to this thesis in Chapters Two through Four, in Chapter Five I provide concluding discussion of the findings and significance of the work. Summarisation of the key results is given, alongside reflection of whether the aims of the thesis have been achieved. Importantly, the context of the results is discussed in terms of furthering our knowledge of both the *Howea* speciation scenario, and of the wider field of speciation biology. Finally, thought is given to important avenues for both ongoing and potential future lines of research, which would further build upon the results presented here.



Figure 1.3. Photograph showing *H. belmoreana* and *H. forsteriana* individuals growing alongside a suspected hybrid on LHI. Notice the flatter leaves of the hybrid palm.



Figure 1.4. Photograph of Lord Howe Island – a World Heritage Site – looking south with a view to Mount Lidgbird, and beyond to Mount Gower, at the southern tip of the island. At a height of 875m, the summit of Mount Gower is the highest point on the island.



Figure 1.5. Photograph of Lord Howe Island, looking north from the steep slopes of Mount Lidgbird, demonstrating the extensive undisturbed jungle which covers the majority of the island.



Figure 1.6. Photograph from within the palm forests of Lord Howe Island, featuring a number of *H. belmoreana* individuals (notice the upturned leaves). Photograph was taken upon the slopes of Mount Lidgbird.

Chapter Two - Investigating Pleiotropic Speciation Genes in the *Howea* Palms

2.1 Introduction

2.1.1 The Search for Magic Traits

In the field of speciation biology, and particularly in the study of speciation with gene flow, the major zeitgeist of current research is around identifying the precise genomic mechanisms underpinning species divergence (Foote, 2018; Richards et al., 2019). Studies in the genomic era suggest that during speciation with gene flow, regions of differentiation between species are most commonly restricted to a small number of loci (Chapman et al., 2013; Poelstra et al., 2014; Seehausen et al., 2014; Malinsky et al., 2015; Marques et al., 2016; Westram et al., 2018; Papadoulos et al., 2019a; Kautt et al., 2020). This is consistent with the theoretical prediction that a limited number of major effect loci, which link assortative mating and adaptive divergence, are more likely to facilitate speciation with gene flow than a larger number of small effect loci acting cumulatively (Gavrilets, 2004; Servedio et al., 2011; Smadja & Butlin, 2011). It should, however, be acknowledged that theoretical work has also suggested that many weakly-selected loci acting together could rapidly lead to genome-wide barriers to gene flow (Feder et al, 2014; Flaxman et al., 2014); and recent work in neotropical cichlid fishes showed that widespread genomic divergence is present in species which have diverged in polygenic traits during speciation with gene flow (Kautt et al., 2020).

Nonetheless, the search for major effect loci remains pivotal in attempts to further our understanding of the mechanisms behind sympatric speciation. One such avenue is the search for speciation with gene flow driven by magic traits. The term 'magic trait' refers to the scenario in which multiple phenotypic traits, which drive both divergent selection and assortative mating, are underpinned by the same pleiotropic gene (Servedio et al., 2011). For clarity, I here refer to a pleiotropic gene as one where a single genomic locus results in distinct phenotypic effects at two or more different traits; regardless of whether the gene in question directly regulates the trait, or is indirectly associated with a trait due to its affect on

downstream pathways. In either scenario, two or more phenotypic traits are dependent on the pleiotropic gene in question. And, indeed, theoretical models have shown that sympatric speciation may be induced far more readily where a pleiotropic magic trait is present (Dieckmann & Doebeli, 1999).

Increasingly, the idea that such magic traits may not be rare is becoming accepted, with multiple examples having existed in the literature for some time. For example, colour patterning in *Heliconius* butterflies (Jiggins et al., 2001) and beak size and morphology in Darwin's finches (Podos et al., 2004). However, demonstrating the role of magic traits in a convincing example of sympatric speciation — the most extreme form of speciation with gene flow — has remained challenging, with little empirical evidence. Here, I investigate the role of potential magic traits in the speciation of the *Howea* palms.

2.1.2 Candidate *Howea* Speciation Genes

Two previous studies investigating differential gene expression and coding sequence divergence in *Howea* identified a range of candidate 'speciation genes' (Dunning et al., 2016; Osborne et al., 2019), with the potential to act as magic traits by simultaneously contributing to soil substrate adaptation and displaced flowering time during *Howea* speciation. The specific criteria used to identify these genes were discussed in Chapter 1.1.5, and summaries for each gene are shown in Table 2.1.

For each of these 13 candidate speciation genes, it has already been shown that they have likely undergone adaptive divergence in *Howea* (i.e., significantly differential expression between the two species, or evidence that the genes are under positive selection). Therefore, should it also be clearly demonstrated that these genes act as major effect pleiotropic loci affecting both soil stress tolerance and flowering time under conditions consistent with the LHI scenario; it would present strong affirmation of a viable role for magic traits in facilitating sympatric speciation.

2.1.3 Large-Scale Phenotyping of Speciation Genes

To investigate the function of these candidate speciation genes in the ecological speciation of *Howea*, one would ideally carry out loss-of-function and/or over-expression assays in *Howea* palms grown on different soil substrates. However, the generation time of the

palms — many years — would make this extremely challenging. Instead, here I carried out high-throughput phenotyping experiments using almost 2,000 *Arabidopsis thaliana* plants carrying T-DNA knockouts in orthologs of the candidate speciation genes. *A. thaliana* is an angiosperm in the mustard family, and is a widely used model organism, owing to its small size, short generation time (an entire life cycle is completed within two months), high degree of self-fertilisation, and thoroughly-studied genomics (Meinke et al., 1998). In addition, thousands of transgenic *A. thaliana* lines carrying T-DNA knockout insertions in a vast number of genes are available in public stock centres. Taken together, these characteristics make robustly studying the function of multiple genes in parallel, under a range of different treatment conditions, achievable in *A. thaliana* at a scale which would not be feasible within the *Howea* palms themselves.

Specifically, T-DNA knockouts of 11 *A. thaliana* orthologs of the *Howea* speciation genes (two pairs of the *Howea* candidates map to single *Arabidopsis* loci, meaning only 11 *A. thaliana* genes) were grown under drought, salt, and cadmium stress treatments — consistent with the abiotic stresses present in calcareous soils on LHI (Osborne et al., 2019) — at the Finnish National Plant Phenotyping Infrastructure at the University of Helsinki (Figure 2.1). State-of-the-art technology at the centre, including automated RGB, infrared, and chlorophyll fluorescence imaging systems, make it possible to track a broad range of morphometric parameters and measure fine-scale stress response in extreme detail. By combining this high-throughput phenotyping with measurements of flowering time, rigorous assessment was made of the pleiotropic role that each of the candidate speciation genes plays, in both soil stress tolerance and assortative mating through flowering time displacement.

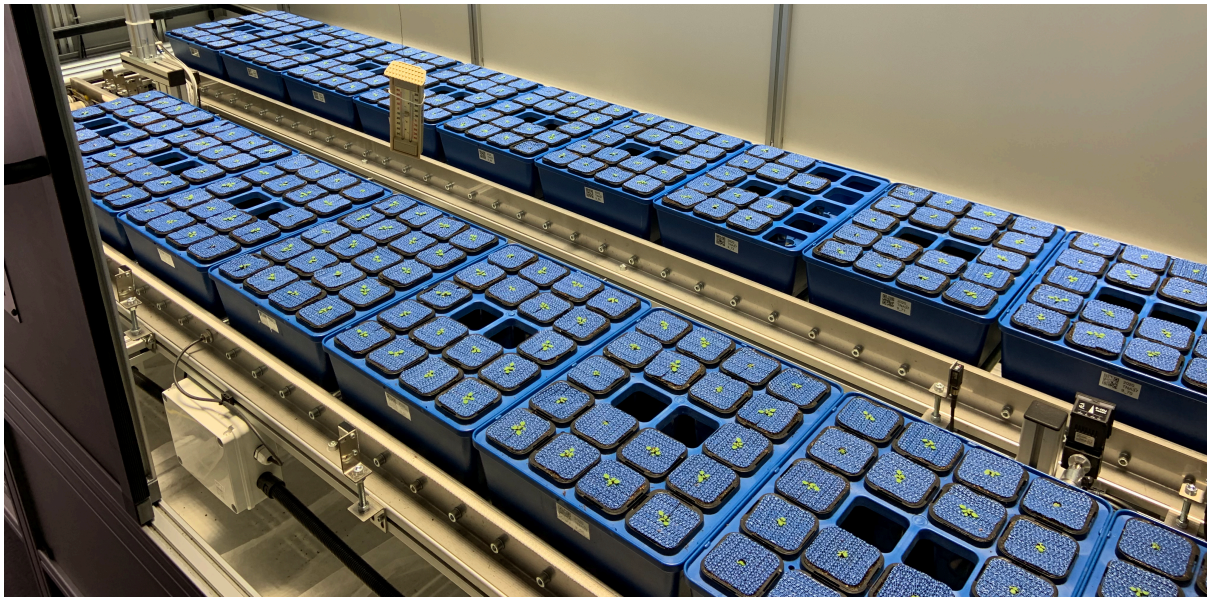


Figure 2.1. Photograph showing *A. thaliana* plants inside the growth chamber at the National Plant Phenotyping Infrastructure at the University of Helsinki. Hundreds of plants can be grown simultaneously, on a conveyor belt system which facilitates automated weighing, watering and imaging.

ID	*Arabidopsis* homologue	Gene name	Evidence of adaptive evolution in *Howea* genes	Potential for pleiotropy
1	AT1G32230	*RCD1*	Positive selection in *H. belmoreana* (branch-site test)	*Arabidopsis* mutant is early flowering (Ahlfors et al.. 2004). Involved in salt (Katiyar-Agarwal et al., 2006) and drought (Ahlfors et al., 2004) tolerance in *Arabidopsis*
2	AT1G70170	*At2-MMP*	Shifted expression in *H. belmoreana* leaves	*Arabidopsis* mutant is late flowering (Golldack et al., 2002). Expressed in response to cadmium and salt treatment in *Arabidopsis* (Golldack et al., 2002)
3	AT5G60410	*SIZ1*	Shifted expression in *H. forsteriana* inflorescence	*Arabidopsis* mutant is early flowering (Jin et al., 2008). Involved in phosphate starvation (Miura et al., 2011) and drought tolerance (Catala et al., 2007) in *Arabidopsis*.
4	AT4G30200	*VEL1*	Shifted expression in *H. forsteriana* inflorescence	*Arabidopsis* mutant is late flowering (Kim and Sung, 2010). Upregulated under drought stress in *Brassica rapa* (Lee and Park, 2017). Differentially expressed in *H. forsteriana* leaf according to soil PC1 in *Howea*.
5	AT5G60410	*SIZ1*	Shifted expression in *H. forsteriana* inflorescence	See above
6	AT2G30140	*UGT87A2*	Outiler (top 5% of F_{ST} and d_{XY})	*Arabidopsis* mutant is late flowering (Wang et al., 2012). Involved in salt and drought stress in *Arabidopsis* (Li et al., 2017)
7	AT1G01040	*DCL1*	Shifted expression in *H. belmoreana* leaves	*Arabidopsis* mutant is late flowering (Schmitz et al., 2007). Expressed in response to drought stress in *Solanum lycopersicum* (Bai et al., 2012).

8	AT2G37770	*AKR4C9*	Shifted expression in *H. forsteriana* roots (Osborne et al., 2019); Differentially expressed in roots (Dunning et al., 2016)	*Arabidopsis* mutant is late flowering (Dunning et al., 2016). Involved in cadmium stress in *Hordeum vulgare* (Éva et al., 2014), salt and drought stress in *Arabidopsis* (Simpson et al., 2009).
9	AT3G22380	*TIC*	Positive selection (dN/dS > 1 in Dunning et al., 2016)	*Arabidopsis* mutant is late flowering in Sanchez-Villarreal et al. (2013), but early flowering in Hall et al. (2003) and has reduced drought tolerance (Sanchez-Villarreal et al., 2013).
10	AT5G20320	*DCL4*	Shifted expression in *H. forsteriana* inflorescence and root; differentially expressed in inflorescence and root (Dunning et al., 2016).	Mutant is late flowering in *Arabidopsis* (Dunning et al., 2016). Down-regulated in response to drought stress in *Oryza sativa* (Liu et al., 2009). Under selection in *Sisymbrium austriacum* during adaptation to novel environments (Vandepitte et al., 2014).
11	AT1G01040	*DCL1*	Positive selection (dN/dS > 1 in Dunning et al., 2016)	See above
12	AT4G27410	*NAC072*	Differentially expressed in roots (Dunning et al., 2016)	*Arabidopsis* mutant is late flowering (Dunning et al., 2016). Expressed in response to desiccation and salt stress in *Jatropha curcas* (Zhang et al., 2014). Overexpression improves drought tolerance in *Arabidopsis* (Tran, 2004).
13	AT5G63980	*SAL1*	Differentially expressed in leaves (Dunning et al., 2016)	*Arabidopsis* mutant is late flowering (Xiong et al., 2001; Dunning et al., 2016). *Arabidopsis* mutant has reduced salt tolerance (Chen et al., 2011) and increased drought tolerance (Wilson et al., 2009).

Table 2.1. List of candidate *Howea* speciation genes, identified by Dunning et al. (2016) and Osborne et al. (2019). *Arabidopsis* homologs, evidence of adaptive evolution, and evidence for a pleiotropic role in stress tolerance and flowering time are shown. Bold gene 'ID' indicates that the locus was first identified in Dunning et al. (2016).

2.2 Methods

2.2.1 Plant Material

Mutant and Col-0 ecotype wild type *A. thaliana* lines were obtained from the European *Arabidopsis* Stock Centre (NASC). All lines were selected from the SALK collection, with T-DNA insertion mutants chosen from the confirmed homozygous SALK stock in a Col-0 background (Supplementary Table I.1). One knockout line per candidate gene was chosen, for a total of 11 mutant accessions in addition to wild type.

Homozygosity of T-DNA insertions was confirmed via PCR using SALK T-DNA specific primers, following DNA extraction using Qiagen DNEasy extraction kits. When visualised on agarose gels, PCR products homozygous for the T-DNA insert produce a single band at least 200 base pairs shorter than wild type controls, with heterozygous lines producing both fragments. Primer pairs are shown in Supplementary Table I.2.

Knockout of expression of the wild type gene for each mutant line was confirmed by RT-PCR following RNA extractions using Qiagen OneStep RT-PCR kits and purification with Invitrogren Turbo DNA-*free*. When visualised on agarose gels, PCR products for lines in which expression of the relevant wild type gene has been suppressed display no fragment, as opposed to wild type controls where the fragment is present. All mutant lines were confirmed as homozygous for the T-DNA insert with expression of the wild type gene knocked-out.

2.2.2 Growth Conditions

Seeds were sown in bulk on Kekkilä Young Plant Mix soil and vernalised at 4°C for 3 nights in order to ensure synchronous germination. After vernalisation, seedlings were transferred to the PlantScreen™ growth chamber at the Finnish National Plant Phenotyping Infrastructure at the University of Helsinki, and grown under long day lengths (16 hour days, 8 hour nights), at 23°C, 60% humidity, and photon irradiance of 130 µmol m^{-2} s^{-1}. Eight days after stratification (DAS), seedlings were transplanted to individual pots filled with 50g Kekkilä soil.

2.2.3 Stress Treatments

For each experimental run, soil field capacity was determined by fully saturating 10 pots filled with soil, weighing the pots, allowing them to drain for two hours, and then drying overnight at 80°C before reweighing. For control, salt and cadmium treatments, pots were dried to 70% field capacity and maintained at this moisture content throughout the duration of the experiment by automated daily weighing and watering of individual pots. Additionally, for salt treatment plants, 20ml of 250mM NaCl solution was added to pots using a syringe 10 DAS. For cadmium treatment plants, 20ml of $CdCl_2$ solution was instead added 10 DAS at a concentration of 140mg/L to produce estimated final soil concentration of 100mg/kg. Control plants were supplemented with 20ml of water. Drought treatment plants were instead left to dry further, until they reached 25% field capacity. Daily weighing and watering maintained drought plants at this level.

2.2.4 Experimental Setup

The PlantScreen™ analysis chamber at the National Plant Phenotyping Infrastructure allows for up to 18 trays of plants to be grown simultaneously, on a conveyor belt system which facilitates automated weighing, watering and imaging. Each of these 18 trays can hold up to 20 individual plant pots. We chose to use 18 plants per tray with the two centre positions left empty, as this improves air circulation which is beneficial for protocols which include water deprivation treatments. A total of up to to 324 plants could therefore be assayed per round.

Experiments were carried out over six rounds, with genotypes split into three groups, and each group used in two experimental rounds (Col-0 wild type plants were included in every round). Group 1 contained Col-0 plants (hereafter referred to as Col-0 1), alongside *siz1*, *rcd1*, *dcl1*, *ugt87a2* and *tic* knockout mutants; under control, drought and salt treatments. Group 2 contained Col-0 (hereafter Col-0 2), alongside *dcl4*, *sal1*, *nac072* and *vel1* knockout lines; also under control, drought and salt treatments. Group 3 contained Col-0 (hereafter Col-0 3), with *akr4c9* and *at2-mmp* knockout lines; grown under control, drought, salt and cadmium treatments. Group 1 plants were distributed in a randomised block design, with one plant of every genotype*treatment combination positioned randomly in every tray. Groups 2 and 3 used completely randomised designs across all trays. The total number of replicates for

every genotype*treatment combination was between 35 and 50 (each Col-0 group treated as an independent genotype in this regard, rather than one pooled Col-0 group across all rounds).

2.2.5 Plant Phenotyping

Flowering time for every plant was assessed by the number of leaves at bolting, according to Pavicic et al. (2017). Bolting transition, as measured by rosette leaf numbers, is a robust indicator of the start of reproductive phase change in *A. thaliana* (Pouteau & Albertini, 2009), with fewer rosette leaves at bolting equating to an earlier shift into the reproductive phase.

To quantify stress response, every tray was imaged daily for 10 days following the imposition of stress treatments at 10 DAS by top-view RGB, Infrared and Chlorophyll Fluorescence cameras. PlantScreen^{TM} software uses a binary mask to distinguish the plant from the background soil and tray, and uses RGB images to automatically calculate morphometric parameters of Rosette Area, Perimeter, Isotropy, Eccentricity, Rotational Mass Symmetry (RMS), Slenderness Of Leaves (SOL), Compactness, Roundness 1 and Roundness 2 (PlantScreen^{TM} analyzer, PSI, Czech R.). These parameters are described in detail by Pavicic et al. (2017), but in brief: Rosette Area and Perimeter are calculated based on raw pixel counts taken from binary images and translated to millimetres; Isotropy, Roundness 1 and Roundness 2 describe the circularity of plants; Eccentricity and RMS describe plant symmetry; Compactness gives information about petiole length and leaf width; SOL describes the sharpness of leaves. Infrared imaging captures the mean surface temperature of the plants, which can be used as an indicator to detect mutants with altered stomatal control, as transpiration causes leaf cooling (Merlot et al., 2002). For example, plants produce the hormone abscisic acid in response to drought stress, which triggers closure of stomatal pores in order to conserve water by reducing transpiration (Merlot et al., 2002). A reduced ability to close stomata is therefore characterised by a lower temperature. Chlorophyll fluorescence imaging can be used to measure the activity of photosystem II in a dark-adapted state through QY Max (also known as Fv/Fm). This represents maximum quantum yield of photosystem II, and is a common measure of plant stress (Baker, 2008; Woo et al, 2008), with lowered values typically indicating stress, (Kitajima & Butler, 1975; Maxwell & Johnson, 2000).

2.2.6 Statistical Analysis

All statistical analysis was carried out using R software, version 4.0.2 (R Core Team, 2022). In all analysis, knockout genotypes were compared only with Col-0 plants from the same experimental group, in order to control for variation between rounds. Flowering time was analysed by fitting a linear mixed model to the leaf number data using the *lmer* function in the R package lme4 (Bates et al., 2015) with genotype as a fixed effect, and treatment, experimental round, tray position in the growth chamber, and position of the plant within the tray as random effects. The contrast function in the package lsmeans (Lenth, 2016) was then used to make pairwise comparisons of means between knockout lines and wild type with Dunnett correction for multiple testing.

Stress response was analysed by fitting linear mixed effects models to $\log_e(x)$ transformed data from the final day of imaging (20 DAS), for each phenotyping parameter and each combination of stress treatment plus control, and single knockout plus wild type, in turn. For example, one model for Rosette Area under drought and control treatments for a*kr4c9* and Col-0 3 plants, another for Compactness under salt and control treatments for *siz1* and Col-0 1 plants, etc. Models were produced using the *lmer* function from the lmerTest package (Kuznetsova et al., 2017), with genotype, treatment and their interaction as fixed effects, and tray position and position of plants within the tray as random effects.

Such models allow computation of ANOVA interaction terms to assess whether the response of a trait under stress versus control treatment differs for a given genotype in comparison to wild type plants, taking into account the value of the relevant phenotypic trait under control conditions. This is necessary due to phenotypic differences existing among genotypes even under control conditions. Therefore, simply measuring differences under stress conditions without taking pre-existing differences into account would fail to measure stress response. Additionally, by first $\log_e(x)$ transforming data, additive interaction terms in the transformed data are equivalent to multiplicative interactions in the non-transformed dataset. This means that significant $\log_e(x)$ interactions describe differences in the proportional change in the value of a trait between treatments for a given knockout and wild type, thereby controlling for pre-existing phenotypic differences.

2.3 Results

2.3.1 Flowering Time

Of the 11 knockout lines investigated, four were found to flower significantly earlier than wild type plants, as measured through number of rosette leaves at bolting. The four genotypes were *dcl1* ($p<0.0001$), *at2-mmp* ($p=0.0001$), *rcd1* ($p<0.0001$) and *siz1* ($p<0.0001$). One knockout genotype, *sal1*, showed extremely delayed flowering under all conditions, with a mean of 29.67 leaves under control conditions, in comparison to 12.88 leaves for Col-0 2 plants under control conditions. With such a clear displacement in flowering time, *sal1* was excluded from the linear model fitted to leaf number data in order to improve fit of the model. No significant difference was found in flowering time between Col-0 and any of the remaining knockout lines. Least squared means and test statistics for all lines other than *sal1* are shown in Table 2.2.

Genotype	Least Squares Mean No. Rosette Leaves	Standard Error	Corrected P-value
Col-0 1	11.57	0.424	NA
dcl1	10.06	0.425	**<0.0001***
rcd1	10.74	0.425	**<0.0001***
siz1	9.42	0.425	**<0.0001***
tic	11.59	0.425	0.9998
ugt87a2	11.49	0.424	0.9905
Col-0 2	12.31	0.422	NA
dcl4	12.14	0.423	0.8064
nac072	12.11	0.423	0.6933
vel1	12.15	0.422	0.8331
Col-0 3	12.84	0.416	NA
akr4c9	12.53	0.416	0.0626
at2-mmp	12.32	0.416	**0.0001***

Table 2.2. Estimated least squares means for number of rosette leaves at bolting for all knockout lines and Col-0 plants excluding *sal1*. P-values represent contrasts of means for each knockout genotype against Col-0 plants from the relevant experimental group. * Represents groups with flowering time significantly different to wild type.

2.3.2 Phenotypic Variation

Several of the knockout lines demonstrated phenotypes notably different to wild type across each of the treatment conditions (Figure 2.2). For example, *sal1* showed a markedly different phenotype in comparison to Col-0 2 across all treatments, characterised by a condensed rosette with rounder, waxy-appearing leaves.

For all genotypes, Rosette Area was smaller under drought than control conditions, and smaller still under salt conditions. Rosette Area under cadmium treatment was similar to that in drought conditions. Plant Perimeter followed an equivalent pattern. Rosette Area and Perimeter increased exponentially over time in all treatments and genotypes, with the exception of *sal1,* which showed flatter growth, especially under salt conditions.

For wild type plants under control and drought conditions, QY Max increased over time, indicative of increasing maximum quantum yield of photosystem II. Under salt conditions, QY Max in Col-0 increased sharply over the first 3-4 days (13-14 DAS) following the start of treatments, then plateaued until around 18-19 DAS, before sharply rising again. Peak QY max under salt conditions was higher than control, drought and cadmium treatments in all genotypes other than *akr4c9* (where it was higher than control, but lower than drought), and *sal1,* where peak QY Max was lowest under salt conditions. For wild type plants, QY Max decreased slightly over time under cadmium treatment, with a peak markedly lower than all other conditions.

For all genotypes and treatments, Mean Leaf Temperature rose sharply over the first three days following imposition of stress treatments at 10 DAS. Temperature continued to rise further for plants grown in drought conditions, with a peak far higher than other treatments at around 16 DAS, indicating closure of stomata and reduced transpiration. Figure 2.3 and Supplementary Figures I.1-I.10 show time-series measurements for each knockout genotype against wild type over 10 days following stress imposition, for each of the phenotyped parameters other than RMS.

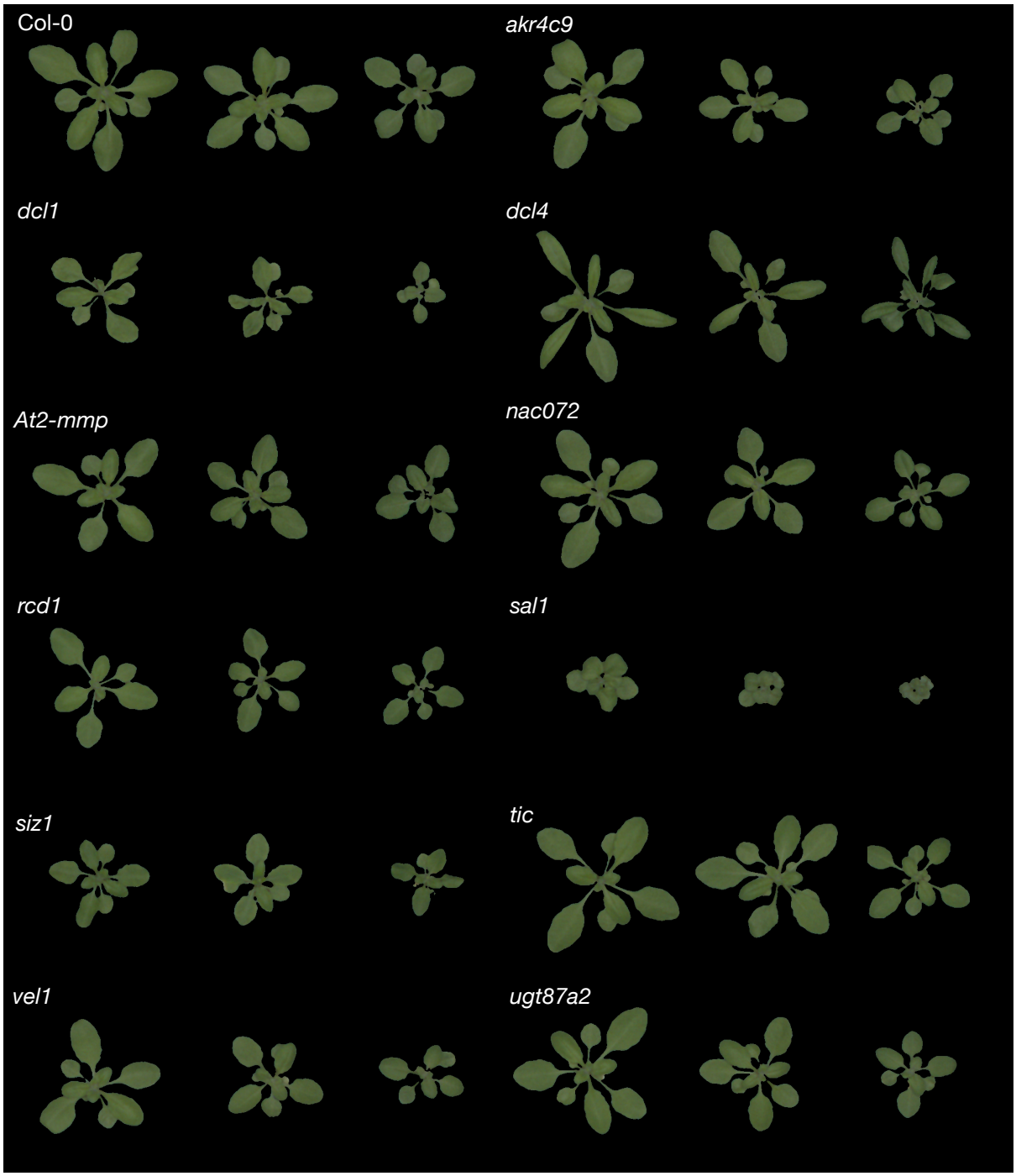
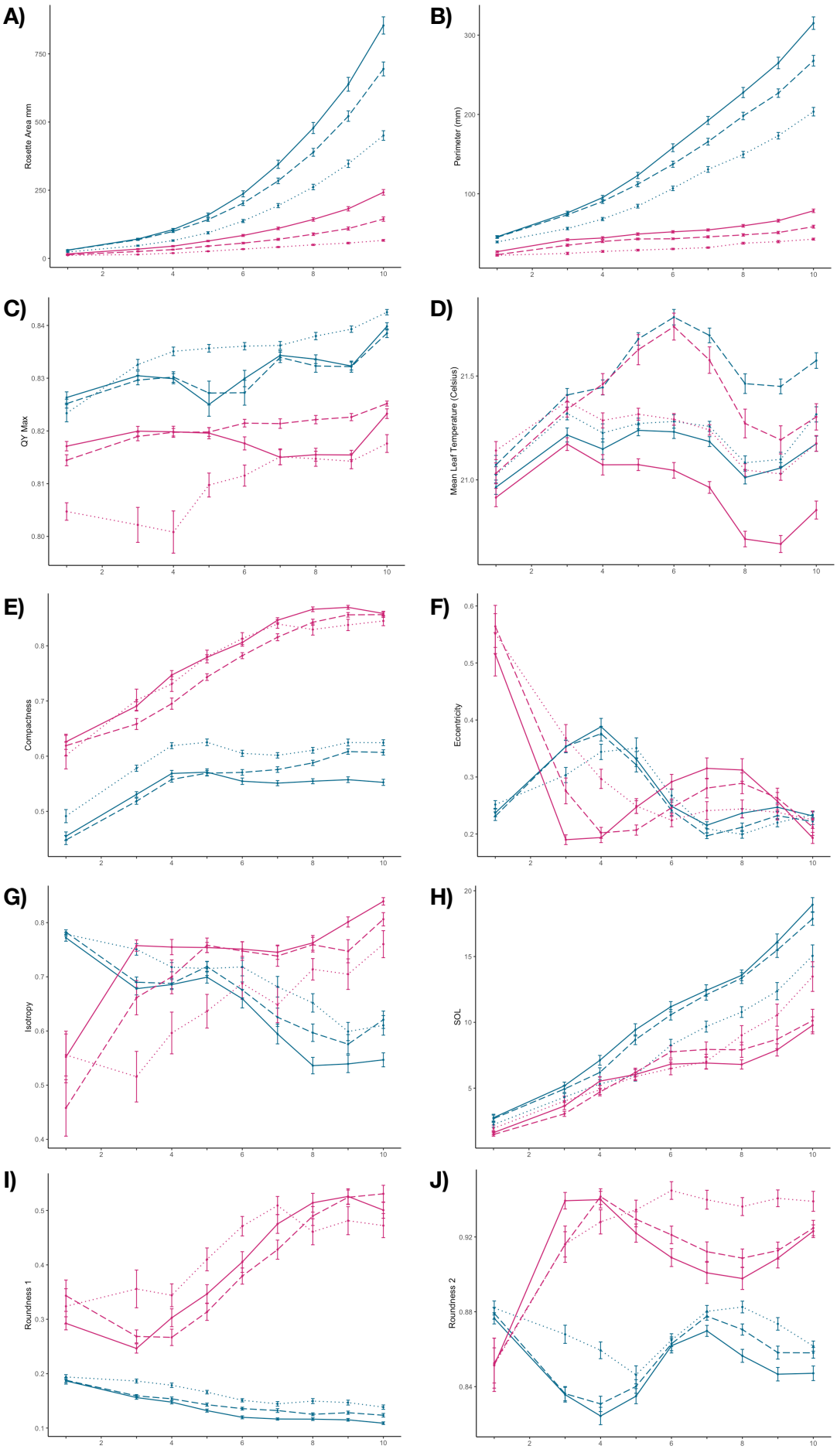


Figure 2.2. Representative phenotypes of knockout lines and Col-0 wild type under control, drought and salt conditions. For all genotypes the first column shows an example plant under control conditions, second column shows drought, and third column shows salt conditions.

Figure 2.3 (overleaf). Measurements of phenotypic traits over time for *sal1* and Col-0 2 wild type plants. In all charts, *sal1* is represented by pink lines, with Col-0 2 shown in blue. In all charts, solid lines represent plants grown under control treatment; dotted lines represent salt treatment; dashed lines represent drought treatment. Panels show phenotypic traits as follows: **A)** Rosette Area (mm^2); **B)** Perimeter (mm); **C)** QY Max; **D)** Mean Leaf Temperature (°C); **E)** Compactness; **F)** Eccentricity; **G)** Isotropy; **H)** SOL; **I)** Roundness 1; **J)** Roundness 2. RMS measurements are not shown. In all charts, X axes show days from imposition of stress treatments. Error bars show standard error.

Figure 2.3


A)
Rosette Area mm
B)
Perimeter (mm)
C)
QY Max
D)
Mean Leaf Temperature (Celsius)
E)
Compactness
F)
Eccentricity
G)
Isotropy
H)
SOL
I)
Roundness 1
J)
Roundness 2

2.3.3 Knockout Stress Response

At 20 DAS, significant multiplicative interactions in at least one measured trait across control and stress conditions were found between Col-0 and every knockout line other than *nac072*. Tables 2.3-2.5 show F and p-values for interactions tests for QY Max, Mean Leaf Temperature and Rosette Area (mm^2), for all combinations of knockout and stress treatments compared against Col-0 and control conditions. Test statistics for the remaining phenotypic parameters can be see in Supplementary Tables I.2-I.10.

For QY Max, seven knockout genotypes showed significant multiplicative interaction in at least one stress treatment versus Col-0 and control conditions. There were four such genotypes for each of Perimeter, Compactness and Roundness 1; three for Rosette Area and SOL; two for Mean Leaf Temperature and Roundness 2; and one for Isotropy and Eccentricity. No significant multiplicative interactions were found for RMS between knockout genotypes and Col-0 in any stress treatment versus control.

SAL1 and SIZ1 knockouts differed most from Col-0 with respect to significantly different proportional changes across a range of phenotypic traits between control and stress conditions. Significant multiplicative interactions between *sal1 and* Col-0 2 were found for QY Max (p=0.01867), Rosette Area (p=0.000265), Perimeter (p=0.0242), Compactness (p=6.098e-10) and Isotropy (p=8.133e-05) between control and drought conditions; and for QY Max (p=5.419e-06), Mean Leaf Temperature (p=0.04075), Rosette Area (p=9.161e-13), Perimeter (p=0.003542), Compactness (p=2.841e-13), SOL (p=2.777e-05) and Roundness 1 (p=9.459e-06) between control and salt conditions.

The directions of these significant interactions are as follows: under drought treatment, QY Max in *sal1* was higher than in control conditions, whilst for Col-0 2, QY Max was marginally lower under drought than control conditions. For salt versus control, the opposite was true, with QY Max in Col-0 2 higher under salt conditions, and QY Max in *sal1* lower under salt conditions. Rosette Area and Perimeter were lower for both *sal1* and Col-0 2 under both drought and salt conditions than under control conditions. However, in both stress treatments, *sal1* had a significantly greater proportional decrease in both Rosette Area and Perimeter than Col-0 2 in comparison with control treatment. Under drought conditions, Compactness increased for Col-0 2 relative to control conditions, but was almost identical for

sal1 between the two conditions. A similar pattern was seen between salt and control conditions, with Compactness increasing under salt treatment for Col-0 2, but decreasing slightly for *sal1*. Isotropy was higher for Col-0 2 under drought than control conditions, but lower for *sal1*. Under salt conditions, Mean Leaf Temperature was higher for both Col-0 2 and *sal1* than under control conditions, but the increase was proportionally greater for *sal1*. SOL decreased for Col-0 2 under salt conditions relative to control, but increased for *sal1*. Roundness 1 increased for Col-0 2 under salt conditions relative to control, but decreased for *sal1*. Figure 2.3 shows time series data of phenotypic measurements for *sal1* versus Col-0 2 under all treatment conditions.

For *siz1* versus Col-0 1, significant interactions were found for QY Max (p=0.04969), Rosette Area (p=0.01611), Compactness (p=0.02484) and Roundness 1 (p=0.0007965) between control and drought conditions; and for Roundness 1 (p=0.006767) and Roundness 2 (p=0.01217) between control and salt conditions. Directions of significant interactions as follows: QY Max for Col-0 1 plants was slightly lower under drought conditions versus control, but markedly lower for *siz1*. Rosette Area was lower for both Col-0 1 and *siz1* under drought conditions than control, but the proportional decrease was smaller for *siz1*. Compactness increased for both *siz1* and Col-0 1 under drought relative to control, but the increase for *siz1* was proportionally greater. Roundness 1 was higher for both *siz1* and Col-0 1 under both drought and salt treatments in comparison with control, but the increase was proportionally greater for *siz1* in both stress treatments. Roundness 2 was higher for Col-0 1 under salt than control conditions, but lower for *siz1*. Details of the phenotypic stress responses and significant multiplicative interactions for the remaining knockout lines can be seen in Tables 2.3-2.5, and in Supplementary Tables I.2-I.10 as well as Supplementary Figures I.1-I.10.

QY Max

Genotype	Treatment	F value	P-value
akr4c9	Cadmium	3.9503	**0.04827 ***
	Drought	18.453	**2.760e-05 *****
	Salt	1.2956	0.2565103
dcl1	Drought	6.7573	**0.0104510 ***
	Salt	0.8707	0.3524
dcl4	Drought	0.7895	0.3758
	Salt	0.9571	0.3296
at2-mmp	Cadmium	3.7171	0.05531
	Drought	0.0621	0.803551
	Salt	0.2016	0.654007
nac072	Drought	0.0022	0.96223
	Salt	0.0678	0.7950
rcd1	Drought	0.0244	0.8760
	Salt	12.935	**0.0004716 *****
sal1	Drought	5.6640	**0.01867 ***
	Salt	22.3317	**5.419e-06 *****
siz1	Drought	3.9300	**0.04969 ***
	Salt	0.4169	0.51962
tic	Drought	0.0113	0.9157
	Salt	7.2990	**0.007827 ****
ugt87a2	Drought	0.0014	0.9707
	Salt	7.6520	**0.006463 ****
vel1	Drought	1.4493	0.2306
	Salt	0.3494	0.5554

Table 2.3. F and p-values for ANOVA interactions tests of $\log_e(x)$ transformed QY Max measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Mean Leaf Temperature (°C)

Genotype	Treatment	F value	P-value
akr4c9	Cadmium	0.0455	0.8314029
	Drought	0.0285	0.8661
	Salt	0.8225	0.3658
dcl1	Drought	0.0275	0.8685
	Salt	0.6229	0.43135
dcl4	Drought	5.1715	**0.02451 ***
	Salt	0.1970	0.65786
at2-mmp	Cadmium	0.8194	0.3666534
	Drought	0.7047	0.4024
	Salt	0.4596	0.4988
nac072	Drought	0.0077	0.9303
	Salt	0.4827	0.4883
rcd1	Drought	0.0809	0.7765
	Salt	0.0011	0.97409
sal1	Drought	0.0359	0.8499
	Salt	4.2625	**0.04075 ***
siz1	Drought	0.1528	0.6965
	Salt	0.0162	0.8991
tic	Drought	0.4983	0.4817
	Salt	0.0447	0.83285
ugt87a2	Drought	0.1260	0.7232
	Salt	0.1141	0.73603
vel1	Drought	0.6347	0.42688
	Salt	0.0067	0.9349

Table 2.4. F and p-values for ANOVA interactions tests of $\log_e(x)$ transformed Mean Leaf Temperature (°C) measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Rosette Area (mm^2)

Genotype	Treatment	F value	P-value
akr4c9	Cadmium	0.2606	0.6103
	Drought	2.8941	0.090565
	Salt	0.0008	0.9774
dcl1	Drought	0.0593	0.8079
	Salt	2.269	0.1345
dcl4	Drought	0.243	0.6228
	Salt	0.6347	0.4269
at2-mmp	Cadmium	5.4318	**0.02082 ***
	Drought	2.5599	0.1113
	Salt	1.0464	0.3076
nac072	Drought	0.0192	0.89002
	Salt	0.8381	0.36141
rcd1	Drought	0.025	0.8745
	Salt	3.1805	0.07707
sal1	Drought	13.988	**0.000265 *** **
	Salt	60.884	**9.161e-13 *** **
siz1	Drought	5.9622	**0.01611 ***
	Salt	1.0837	0.2999
tic	Drought	2.0181	0.15791
	Salt	0.6938	0.4064
ugt87a2	Drought	0.0309	0.8607
	Salt	1.7344	0.1903
vel1	Drought	1.1422	0.2869
	Salt	0.3133	0.5765181

Table 2.5. F and p-values for ANOVA interactions tests of $\log_e(x)$ transformed Rosette Area (mm^2) measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

2.4 Discussion

2.4.1 Strong Evidence of Multiple Magic Trait Speciation Genes

Several loci were identified which show evidence of displaced flowering time and altered soil stress response acting simultaneously in *A. thaliana,* under conditions consistent with the speciation of *Howea*. The strength of evidence for each gene is evaluated in detail below, but, in brief, *At2-MMP*, *DCL1*, *RCD1*, *SAL1* and *SIZ1* knockouts displayed pleiotropic phenotypes pointing to a potential role as 'speciation genes'. Two *Howea* candidate genes mapped to each of *DCL1* and *SIZ1*, producing a list of seven *Howea* magic trait genes.

2.4.2 Evaluation of Candidate Speciation Genes

AKR4C9

We found that *AKR4C9* plays a potential role in response to both drought and cadmium stress, with the proportional change in QY Max significantly different to Col-0 3 in both of these treatments compared to control. Of particular note is the fact that QY Max in *akr4c9* decreased under cadmium treatment significantly more than in Col-0 3, with absolute QY Max 20 DAS far lower in *akr4c9* than Col-0, indicating plant stress and a reduction in photosynthetic activity. These results are concordant with two studies carried out by Evá et al. (2014; 2016) who found that over-expression of the *Arabidopsis AKR4C9* in *Hordeum vulgare* barley improves cadmium tolerance and leads to significantly higher QY Max than wild type plants in response to cadmium treatment.

We did not, however, find that *akr4c9* mutants possess a significantly different flowering time than wild type in our growth conditions. Dunning et al. (2016) found that two *akr4c9* knockout lines (SALK_119576, which is the same knockout line used here, and SAIL_71_E07) produced more rosette leaves at bolting than wild type plants under long day lengths. We have not been able to validate this flowering phenotype in the present study.

At2-MMP

Our results indicate a role for *At2-MMP* in response to cadmium induced stress, with proportional changes in Rosette Area, Perimeter and SOL being significantly different between *at2-mmp* and Col-0 3 between control and cadmium treatments. All three parameters

were lower for *at2-mmp* than Col-0 3 under cadmium treatment, but marginally higher than Col-0 3 under control conditions. These results show consistency with Golldack et al. (2002), who found that in rosette leaves, exposure to cadmium increased the transcript level of *At2-MMP*, whilst expression was unaffected by NaCl.

Golldack et al. (2002) also showed that *at2-mmp* knockout mutants flowered significantly later than wild type under short-day conditions, and Zhao et al. (2017) found that overexpression mutants of *At2-MMP* flowered earlier than wild type under short-day conditions. Conversely, we found that *at2-mmp* flowered earlier than Col-0 under long-day conditions, measured by number of rosette leaves at bolting. Whilst these results appear contradictory, flowering time in *A. thaliana* is regulated by a range of complex pathways, including a photoperiod response pathway (Simpson & Dean, 2002; Luo et al., 2022). As a facultative long day flowering plant, photoperiod significantly alters flowering time, and, therefore, contrasting responses to differing photoperiod regimes are not unexpected. For example, *vel1* mutants (see below) have been shown to display delayed flowering in short day conditions, but flower similarly to wild type in long days (Kim and Sung, 2010). Overall, our results show a clear pleiotropic role for *At2-MMP* in both flowering time and response to abiotic cadmium stress.

DCL1

Similarly to the phenotype demonstrated by Gasciolli et al. (2005), knockout *dcl1* plants produced smaller rosettes than wild type plants, characterised by highly warped leaf morphology (Figure 2.2). We found evidence for the involvement of *DCL1* in both drought and salt stress response, particularly through *dcl1* mutants possessing significantly lower QY Max than Col-0 1 under drought conditions, despite showing similar QY Max under control conditions — indicative of greater drought stress in *dcl1*. This is consistent with Bai et al. (2012), who found that the *DCL1* homolog in *Solanum lycopersicum* was expressed in response to drought stress.

Interestingly, in our growth conditions, *dcl1* mutants possessed significantly fewer leaves than Col-0 1 at bolting ($p<0.0001$), indicative of earlier flowering time. Fewer leaves were present in all treatment conditions. Previous studies have shown that in the Landsberg accession (rather than Col-0), mutations in *DCL1* alone instead result in a slight delay in

flowering time (Ray et al. 1996; Liu et al. 2004), whilst in the Col accession, Schmitz et al. found no significant difference in flowering time caused by a single *dcl1* mutant (though a *dcl1;dcl3* double mutant caused significantly delayed flowering time, but only when plants were grown without prior vernalisation). Whilst our results are not entirely consistent with previous studies with regards to the role of *DCL1* in flowering time, vernalisation is, like photoperiod, a key regulator of flowering time in *A. thaliana* (Simpson & Dean, 2022), and therefore observing differing phenotypes with and without vernalisation would not be surprising. Additionally, differences in alleles which affect flowering time are known to exist between the Landsberg and Col-0 genetic backgrounds (Lee et al., 1994), and contrasting reports on flowering time in different backgrounds have been seen previously. For example, Hall et al. (2003) reported that *tic* mutants (see below) in the C24 ecotype flowered earlier than wild type plants under both long and short day conditions, whereas Sanchez-Villarreal et al. (2013) showed that *tic* mutants in a Col background flowered significantly later than wild type under long day conditions. Further investigation into these contrasting results may be prudent — however, regardless, the combination of evidence for a role in drought tolerance, and the earlier flowering time found here, strongly indicates a pleiotropic effect for *DCL1* relevant to the *Howea* speciation scenario.

DCL4

Knockout *dcl4* mutants possessed a phenotype with clear differences in comparison to wild type plants (Figure 2.2), characterised by a less compact rosette with more slender leaves, consistent with phenotypes shown by Gasciolli et al. (2005). Possible roles for *DCL4* in both drought and salt stress response were found, through significantly different proportional differences in Mean Leaf Temperature and Perimeter compared with Col-0 2 between control and drought, and control and salt, conditions respectively. Of particular note is the proportionally greater increase in Mean Leaf Temperature of *dcl4* plants under drought conditions than Col-0 2 (and higher absolute Mean Leaf Temperature than Col-0 2 under drought conditions), which is indicative of an increased ability to close stomatal pores to conserve water. This is consistent with Liu et al. (2009), who found that the *DCL4* homolog in *Oryza sativa* is down-regulated in response to drought stress. Together these results suggest that *DCL4* negatively affects drought tolerance.

In our growth conditions, we did not find any significant difference in flowering time between *dcl4* and wild type plants. This is opposed to Dunning et al. (2016), who found that *dcl4* knockout mutants from the same T-DNA line as used here (SALK_206940) produced more rosette leaves at bolting than wild type plants, indicative of delayed flowering. However, Dunning et al. had a sample size of 15 *dcl4* plants, far smaller than used here. Should further work highlight a displaced flowering time for *dcl4*, then in combination with its likely role in drought tolerance, it would represent an interesting candidate 'speciation gene'. Currently, however, we have not been able to validate a role in flowering time.

NAC072

We found no significant difference in flowering time between *nac072* and Col-0 2 measured through number of leaves at bolting, as opposed to Dunning et al. (2016), who previously reported delayed flowering time in *nac072* using the same knockout line (SALK_063576) as used here. However, Dunning et al. also compared a second *NAC072* knockout line (SALK_090148) against wild type and found no significant difference in number of leaves at bolting. Additionally, sample sizes were small for the knockout lines in Dunning et al., at 12 for SALK_063576 and 5 for SALK_090148.

Tran (2004) showed that *NAC072* expression was strongly induced by dehydration treatment, and that four-week-old overexpression mutants of *NAC072* showed significantly enhanced survival compared to wild type after being subject to water deprivation for nine days. Whilst we did not find any significantly different proportional response to stress treatment in *nac072* knockout plants compared with wild type, our assay conditions were notably different to those used by Tran (2004); particularly regarding the age of plants subject to stress treatment (10 days old here versus 4 weeks in their study).

RCD1

Consistent with Ahlfors et al. (2004), we found that *rcd1* mutants were smaller than wild type plants, possessed altered rosette morphology, and flowered earlier than Col-0. Ahlfors et al. (2004) also reported a possible role for *RCD1* in drought tolerance through interaction with partner proteins involved in gene expression changes in response to drought. Whilst the phenotype of *rcd1* mutants in our experiments differed in several traits from that of Col-0 under drought conditions (see Supplementary Figure I.6), our measured parameters did not

identify a significantly different proportional change in phenotype between control and drought conditions in *rcd1* compared with Col-0 1.

We did, however, find proportional changes in QY Max and Perimeter between control and salt treatments which differed significantly between *rcd1* and Col-0 1, indicating a role for *rcd1* in salt stress response. Katiyar-Agarwal et al. (2006) found that germination in *rcd1* mutants was inhibited under NaCl induced salt stress in comparison to Col-0, and that *rcd1* seedlings grown on agar media supplemented with 75mM NaCl exhibited stunted growth relative to Col-0. Between these results and the earlier flowering of *rcd1* mutants, *RCD1* is a strong candidate 'speciation gene'.

SAL1

Knockouts of *SAL1* produced strikingly distinct phenotypes (Figure 2.2), and, along with *SIZ1* (see below), *SAL1* likely represents the strongest candidate for a large-effect magic trait. Indeed, knockouts flowered several weeks later than Col-0 2 in all treatment conditions, consistent with the late flowering phenotypes reported by Xiong et al. (2001) and Dunning et al. (2016). Alongside this, *sal1* plants possessed a distinct rosette morphology, being far smaller and more compact than wild type plants, and having far more rounded leaves.

We also found that *sal1* mutants differed from Col-0 2 in proportional change across a wide range of traits between control and both drought and salt conditions, suggesting that *SAL1* plays a role in both drought and salt stress response in *A. thaliana*. Notably, we found evidence consistent with Wilson et al. (2009), who reported increased drought tolerance in the same *sal1* T-DNA knockout line as used here (SALK_020882).

Under drought conditions, QY Max in *sal1* was actually higher than under control conditions, whilst for wild type, QY Max was marginally lower than under control conditions. We also found evidence to suggest that *SAL1* plays a role in salt stress tolerance, consistent with Quintero et al. (1996). QY Max for *sal1* was significantly lower under salt conditions than control conditions, whilst for all other genotypes QY Max was higher under salt conditions. Together with the displaced flowering time, these results make *SAL1* an important candidate to study further.

SIZ1

Our results highlight *SIZ1* as a very strong candidate for a pleiotropic gene which could have facilitated speciation on LHI. Flowering was significantly earlier in *siz1* than Col-0 1 plants, consistent with the results of Jin et al. (2008), who found that *siz1* mutants flowered earlier than wild type under both long and short day lengths. Knockout *siz1* plants also demonstrated a clearly altered morphology consistent with both Catala et al. (2007) and Jin et al. (2008), characterised most noticeably by a far smaller rosette and smaller leaves.

Additionally, our results highlight a role for *SIZ1* in both salt and drought response, with proportional changes in multiple traits significantly different to Col-0 1 in both treatment conditions compared to control. Particularly, we found evidence that *SIZ1* plays a role in drought tolerance, with both absolute and proportional QY Max lower under drought conditions for *siz1* than wild type. Survival assays carried out by Catala et al. (2007) found that *siz1* mutants were far less resilient to water deprivation than wild type plants, and together with our results and the earlier flowering time of *siz1,* this represents a very strong candidate speciation gene.

TIC

We found no evidence that *tic* mutants flower differently than wild type. Hall et al. (2003) previously reported that *tic* mutants in a C24 ecotype background flowered earlier than C24 wild type plants under both long and short day conditions as measured through number of leaves at bolting, whilst Sanchez-Villarreal et al. (2013) reported that *tic* mutants in a Col background flowered significantly later than wild type under long day conditions.

Sanchez-Villarreal et al. (2013) also reported that *tic* mutants were tolerant to drought stress, with growth and survival in mutants affected far less by water deprivation treatment than wild type plants. We found some evidence that *TIC* plays a role in both drought and salt response through differential proportional change than Col-0 1 in SOL and QY Max respectively. Knockout *tic* plants were also larger than wild type under drought conditions, but no significant difference in proportional change between control and drought conditions was found. However, overall evidence for a role in drought tolerance in the present study is limited.

UGT87A2

We found some evidence to suggest that *UGT87A2* plays a role in salt stress response, with a significant interaction between *ugt87a2* and Col-0 1 in QY Max between control and salt conditions. For both genotypes, QY Max was higher under salt conditions than under control conditions, but the proportional increase was smaller for *ugt87a2* than Col-0 1.

However, our most interesting finding concerning *ugt87a2* is the lack of any significant flowering time difference compared with Col-0. Previously, Wang et al. (2012) described significantly delayed flowering times for *ugt87a2*, with mutants flowering far later under both long and short day conditions. Under long day conditions, they reported that mutant plants began to bolt around 45 days after sowing, compared to 26 days for wild type plants. Additionally, one of the T-DNA knockout lines used by Wang et al., SALK_061574, is the same as the line used in this study.

Whilst other genotypes in this study have displayed flowering times not entirely consistent with those previously described, these discrepancies have arisen over genotypes with far smaller reported flowering time displacements, often using different knockout methods/lines, or genotypes with contradictory previous studies. The results reported in Wang et al. (2012) are of greater significance due to the size of the discrepancy reported, using similar growth conditions and an identical knockout line. Wang et al. further confirmed their results with a second T-DNA knockout line, SALK_124038.

In order to investigate this further, additional flowering time assays using both knockout lines (SALK_061574 and SALK_124038) alongside Col-0 wild type were carried out. Full details of the methodology can be found in Appendix I Supplementary Information, however neither knockout line demonstrated flowering time significantly different from wild type, in stark contrast to the results reported by Wang et al. (2012).

VEL1

Knockouts of *VEL1* did not possess a significantly altered flowering time under our growth conditions. Previous research found that *vel1* mutants displayed delayed flowering in short day conditions, but flowered similarly to the wild type in long days (Kim and Sung, 2010), consistent with our results.

We did, however, identify a likely role for *VEL1* in salt stress response, with proportional change in rosette morphology across multiple traits significantly different when compared to Col-0 2 between control and salt conditions.

2.4.3 *Howea* Homologs of *SAL1* and *SIZ1* Present the Strongest Evidence for Major Effect Magic Traits During LHI Speciation

In the search for magic traits capable of facilitating sympatric speciation, it is important to consider the effect size of a locus, in terms of the strength of divergent selection and the degree of assortative mating imposed by the locus during ongoing gene flow (Servedio et al., 2011; Haller et al., 2012). Even when a magic trait is present, if the effect size is weak, its role in driving speciation may be trivial or non-existent (Hallet et al., 2012). Whilst knockout mutants of *At2-MMP, DCL1, RCD1, SAL1* and *SIZ1* have all demonstrated displaced flowering time as well as providing evidence for a role in abiotic stress response, effect sizes varied between genes.

SAL1 and *SIZ1* knockouts demonstrated effect sizes noticeably larger than other candidate genes, possessing the latest and earliest flowering phenotypes, respectively. Both loci also showed clearly altered responses to both salt and drought stress treatments across a range of measured parameters. Additionally, results in the present study were entirely consistent with previous research for both genotypes — notably, we found strong evidence for increased drought tolerance and reduced salt tolerance in *sal1*, consistent with Wilson et al. (2009) and Quintero et al. (1996); and for reduced drought tolerance in *siz1*, consistent with Catala et al. (2007).

To then further assess the role these loci may have played in facilitating sympatric speciation in *Howea*, the nature of evidence for adaptive divergence was investigated. For each of the two *Howea* candidate genes which mapped to *SIZ1*, previous work by Osborne et al. (2019) had identified a significant shift in expression in the floral tissue of *H. forsteriana*. Further differential expression analysis, detailed within Chapter Three of this thesis (see Chapter 3.2 for methodology), offered the opportunity for additional insight. For one of these two *Howea* loci (Cluster-5201.1128 in the work detailed in Chapter Three), expression in *H. forsteriana,* growing on either volcanic or calcareous soils, was found to always be higher than

that in *H. belmoreana*, in each of root, flower and leaf tissues — although these differences were not significant. For the other locus (Cluster-5201.432), expression was again always higher in *H. forsteriana*, however, in this case, differences in expression were highly significant in all tissue types. In flower tissues, Cluster-5201.432 had a 140.14-fold increase in the number of read counts per million reads in *H. forsteriana* versus *H. belmoreana.* In leaf tissue, this increase was 58.90-fold, and in roots a 34.65-fold increase. In all tissues, expression in *H. belmoreana* was close to zero, with multiple individuals possessing no counts per million reads.

Such divergent expression patterns reinforce the evidence that this locus could have facilitated sympatric speciation by acting as a magic trait. Knockout of *SIZ1* in *A. thaliana* caused reduced drought tolerance — consistent with *H. belmoreana* showing effectively zero expression of the *SIZ1* homolog Cluster-5201.432, and being incapable of colonising calcareous soils.

For *SAL1*, further investigation using the methodology detailed in Chapter Three was not possible, as the *Howea* transcriptome used in this thesis, which was produced by Osborne et al. (2019), did not resolve a putative *Howea* ortholog of *SAL1*. However, Dunning et al. (2016), using an earlier transcriptome, found significantly increased expression of a *Howea SAL1* homolog in leaf and floral tissues of *H. belmoreana* relative to *H. forsteriana.* Increased drought tolerance in *SAL1* knockouts is consistent with the LHI speciation scenario, with *H. forsteriana* demonstrating reduced expression of the *SAL1* homolog, and able colonise calcareous soils, which are likely associated with drought stress (Osborne et al., 2019). The reduced salt tolerance of *SAL1* knockouts is less consistent with the *Howea* scenario, given calcareous soils likely also cause salt-induced stress (Osborne et al., 2010). Overall evidence for this locus driving sympatric speciation is, therefore, slightly less clear cut, but still strong. It should also be acknowledged that gene function may not be totally conserved between *A. thaliana* and *Howea*. This is an important caveat discussed further below, and means that evidence for pleiotropic function, relevant to LHI speciation, should not be immediately discounted if the direction of phenotypic effects do not show complete concordance with the scenario of divergence in *Howea* — it is plausible that gain or loss-of-function coding sequence differences in the species would affect the direction of traits.

2.4.4 Caveats and Follow-up Research

As highlighted above, one important caveat of the work presented here is the possibility that function of the studied genes is not conserved between *A. thaliana* and the *Howea* palms. This has two major implications. Firstly, as mentioned, this means that for all of the knockout lines demonstrating magic traits (*At2-MMP, DCL1, RCD1, SAL1* and *SIZ1*), evidence for their roles as speciation genes need not be immediately discounted if the directions of phenotypic traits do not show complete consistence with the scenario of speciation on LHI.

Secondly, the more significant implication is the possibility that the *A. thaliana* loci simply have no role, or only a minor role, in stress tolerance and/or flowering time in *Howea.* It is challenging to speculate on the likelihood of this — however, Hong et al. (2021) identified and analysed 22,784 putative flowering time genes in 19 species, within three families of flowering plants (Malvaceae, Brassicaceae, and Solanaceae), and found that flowering time genes were highly conserved. Whilst this gives more weight to evidence presented in this thesis, it is nonetheless prudent to attempt to verify conserved function.

Indeed, this work is currently ongoing, with the aim of transforming the T-DNA knockout lines of magic trait genes with synthetic constructs of their *Howea* homologs, and conducting over-expression assays. Should, for example, a *siz1* knockout line transformed with the *H. forsteriana* allele of Cluster-5201.432 demonstrate earlier flowering time and increased drought tolerance relative to the *H. belmoreana* allele, it would provide conclusive evidence of a pleiotropic magic trait which could have driven sympatric speciation.

In order to do this, individual transcriptomes were assembled from short-read RNA-seq data for 19 *H. belmoreana* and 17 *H. forsteriana* individuals using Trinity RNA-seq (Haas et al, 2013). The *Arabidopsis* homologs of candidate *Howea* genes were then used as a target query to blast against individual transcriptomes, with returned sequences used to construct alignments of *Howea* sequences from all individuals of both species. From these alignments, alleles for both *Howea* species were phased to produce consensus coding sequences based upon the most commonly occurring alleles. SNPs between species were only included in consensus alleles if they were at least 85% fixed across all individuals in each species. Complete coding sequences could be constructed for 12 of the 13 candidate *Howea* genes. The complete coding sequence of one gene (orthologous to *Arabidopsis* locus *NAC072*), could not be constructed from the

transcriptome data. For each of the 12 recovered genes, at least one non-synonymous fixed SNP was present between the two *Howea* species, therefore producing a total of 24 unique coding sequences.

Synthetic constructs of 14 of these sequences — representing coding sequences of the seven *Howea* orthologs of *At2-MMP, DCL1, RCD1, SAL1 and SIZ1* for each *Howea* species — were cloned into the Gateway Series pENTR 11 (Invitrogen, USA) entry vector plasmid by Azenta (USA). The four coding sequences of *SIZ1* orthologs were then sub-cloned by Azenta into the pGWB502 destination vector (Nakagawa et al., 2007), which contains a constitutive 35S cauliflower mosaic virus promoter for over-expression in plants. These sequences were chosen based upon the evidence discussed above, with *Howea* homologs of *SIZ1* shown to be very strong candidates for speciation genes — further coding sequences may be added subsequently.

Following this, purified plasmids carrying the *Howea* genes and 35S CMV promoter were used to transform the GV3101 strain of *Agrobacterium tumefaciens*. *A. tumefaciens* contains a tumour inducing (Ti) plasmid which facilitates transfer of tumour-causing T-DNA to plant cells. Disarmed Ti plasmids harbouring a promoter and gene of interest can be used to induce expression of the relevant gene inside plant cells (Gelvin, 2003).

A. tumefaciens containing the genes of interest will then be used to transform *Arabidopsis* plants using the floral dip method (Clough & Bent 1998), which, in brief, requires inoculation of liquid media with the relevant *Agrobacterium* strain, followed by dipping flowering *Arabidopsis* inflorescences into the media and saturating the plant. Plants are then allowed to mature and produce seeds. Successfully transformed seeds can be selected by germinating seedlings on agar plates containing antibiotics, to which resistance is conferred by the pGWB502 plasmid. Transformed seedlings can then, in turn, be allowed to mature and produce seeds, which can be collected for phenotypic assays using over-expression mutants of *Howea* alleles alongside *Arabidopsis* knockouts. This process is ongoing.

2.4.5 Conclusion and Wider Context

The results presented in Chapter Two rigorously demonstrate that multiple loci with evidence of adaptive divergence between the *Howea* species have the potential to have driven

sympatric speciation, by pleiotropically displacing flowering time and altering abiotic stress response. In particular, one *Howea* homolog of the *A. thaliana* gene *SIZ1* was highlighted for the strong evidence of its role as a key magic trait and speciation gene. Significantly depleted expression of this locus in *H. belmoreana*, which cannot survive in soils where drought stress may be common, is consistent with the reduced drought tolerance exhibited by the *siz1* mutant.

Such evidence is significant, as the key remaining question in the *Howea* speciation scenario is the identification of a specific genetic mechanism capable of driving sympatric divergence with gene flow. Such a question also extends into the wider field of speciation biology, where empirical evidence of the genetic architecture behind sympatric speciation is scant. Examples exist within cichlid radiations — such as the impressive work carried out by Kautt et al. (2020) in identifying a range of monogenic and polygenic traits that likely drove divergence in the Midas cichlids of Central America — however, cichlid radiations have repeatedly been shown to have experienced introgression from secondary contact (Malinsky et al, 2015; Martin et al., 2015; Kautt et al., 2016; Meier et al., 2017; Poelstra et al., 2018, Kautt et al., 2020), raising the question of whether they should truly be labelled as examples of sympatric speciation. There is also a putative example from research carried out in 'Evolution Canyon' in Israel, where there is apparent ongoing sympatric divergence of emmer wheat populations, driven partly by chromosomal rearrangements (Wang et al., 2020). This work is highly interesting, however divergence between the populations into fully reproductively isolated species may yet not be complete; and evidence suggests that genetic drift during ancestral population bottlenecks may have contributed to early genome differentiation. Such genetic drift is not entirely consistent with the panmictic gene flow expected during the early stages of 'pure' sympatric speciation without any allopatric separation.

Convincingly demonstrating the role of a magic trait in *Howea* speciation would, therefore, be arguably the most robust example to date of the genetic architecture behind sympatric speciation. The work presented here makes significant progress in this area, which would only be solidified further by ongoing work to confirm conservation of function between *A. thaliana* and *Howea* 'speciation genes'.

Chapter Three - Hybrid Fitness, Misexpression, and Genomic Incompatibility

3.1 Introduction

3.1.1 The Role of Hybrid Inviability During Speciation With Gene Flow

Whilst many studies in the genomic era have sought to uncover the mechanisms governing speciation, few have thoroughly investigated the role of post-zygotic isolation driven by hybrid inviability during speciation with gene flow (Brennan et al., 2009; Renaut & Bernatchez, 2011; Brennan et al., 2014; McGirr & Martin, 2020). Instead, most studies assessing the role of hybrid inviability have focussed on highly differentiated populations (Michalak & Noor, 2003; Ranz et al., 2004; Noor & Feder, 2006; Rottscheidt & Harr, 2007), with many of these studies highlighting that gene misexpression in hybrids may result in reduced hybrid fitness and strong post-zygotic isolation (Landry et al., 2007; Ortíz-Barrientos et al., 2007; Mack et al., 2016).

As such, our understanding of the role of hybrid incompatibility during speciation with gene flow, in particular during cases of sympatric ecological speciation, is limited (McGirr & Martin, 2020). In *Senecio*, a genus of flowering plants, ecological speciation appears to have occurred between two species upon Mount Etna, which have diverged along an altitudinal cline (Osborne et al., 2013). Gene flow is ongoing between the two species, with an intermediate hybrid zone characterised by genomic incompatibility and reduced hybrid fitness (Brennan et al., 2009; 2014). Whilst *Senecio* thoroughly demonstrates a role for hybrid inviability during ecological speciation with gene flow, such rigorous evidence in a case of pure sympatric speciation is largely absent. One potential example is seen in sympatric emmer wheat populations in Israel, where chromosomal rearrangements were shown to produce hybrid inviability between two populations located within the same valley (Wang et al., 2020). The evolution of this architecture is highly interesting; however, during homoploid

speciation (as in *Howea*), hybrid inviability may more likely be caused by Dobzhansky-Muller incompatibilities (DMIs). These incompatibilities occur when alleles which have diverged between parental species, and do not cause maladaptation in their own genetic backgrounds, produce negative phenotypes when acting together in hybrid genomes (Dobzhansky, 1937; Muller, 1942). This may be the result of co-adapted genetic architectures in parental populations being broken apart during hybridisation.

In turn, these DMIs have the potential to affect gene regulation in hybrid individuals, especially where they occur in regulatory alleles, leading to transgressive hybrid expression patterns outside the typical range of parental expression (Landry et al., 2007; Ortíz-Barrientos et al., 2007; Renaut & Bernatchez, 2011; McGirr & Martin, 2020). Alternatively, hybrid misexpression may be the result of stunted development leading to differences in cell differentiation and the abundance of different cell types, which can affect expression levels (Swain Lenz et al., 2014). Similarly, differences in hybrid expression may be compensatory responses to hybrid dysfunction (Barreto et al., 2015). Whatever the cause, such gene misexpression can disrupt vital developmental processes and reduce hybrid fitness (Bedford & Hartl, 2009; Signor & Nuzhdin, 2018). However, the degree to which selection on gene regulation can contribute to reproductive isolation during cases of sympatric divergence and ecological speciation is largely unknown (Mack & Nachman, 2017), despite the fact that gene regulation is expected to be a major target of selection during adaptive ecological divergence (Manceau et al., 2011; Jones et al., 2012; Muir et al., 2013).

Indeed, the few studies which have thoroughly investigated gene regulation in hybrids within sympatric or recently diverged populations have repeatedly found evidence for misexpression of genes associated with key developmental traits. McGirr & Martin (2020) identified a range of genes misexpressed in hybrids between recently diverged sympatric populations of *Cyprinodon* pupfishes, and found that these genes were enriched for functions including head, muscle and brain development, which are vital for trophic specialisation. Genes misexpressed in hybrids of two populations of the copepod *Tigriopus californicus* were overrepresented for functions including cuticle formation, a vital part of arthropod development (Barreto et al., 2015); and in species pairs of young lineages of lake whitefish, key genes associated with embryonic development were misregulated in hybrids, leading to severe abnormalities (Renaut et al., 2009; Renaut & Bernatchez, 2011).

Taken together, these studies highlight a potentially significant role for hybrid misexpression in generating post-zygotic isolation during speciation with gene flow, which warrants further study — particularly in robust examples of sympatric speciation.

3.1.2 Post-Zygotic Isolation in the *Howea* Palms

Upon LHI, hybrid palms are found in very low frequencies (Babik et al., 2009; Hipperson et al., 2016). Accurately estimating the exact number is challenging, due to the dense nature of palm forests upon LHI, the low abundance of hybrids, and the fact that identifying hybrids morphologically is not always straightforward (see Figures 1.1 and 1.3) — this is especially true in juvenile or later generation hybrids, where intermediate leaf morphology can be less apparent. However, in one site on LHI where *H. belmoreana and H. forsteriana* grow adjacently (Transit Hill), Hipperson et al. (2016) found that 13.5% of juvenile palms were of hybrid classes. This shows that although pre-zygotic isolation driven by flowering time displacement is strong in *Howea* palms — 97% for *H. belmoreana* and 80% for *H. forsteriana*; with the asymmetry resulting from *H. forsteriana* flowering earlier than *H. belmoreana* and being protandrous — pre-zygotic isolation is not complete.

It is, therefore, worth investigating the role that post-zygotic isolation may have played in facilitating sympatric divergence on LHI. Indeed, Hipperson et al. also found that across the island, hybrids are more numerous at juvenile stages than at adult stages — for example, at Transit Hill, whilst 13.5% of juveniles were hybrid classes, no adults sampled at the site were assigned to hybrid classes. This indicates selection acting against hybrids before the adult stage, potentially during periods of stress, for example drought events which are known to occur on LHI. Few studies on hybrid viability in other long-lived plants species exist; however, the pattern seen in *Howea* is similar to oaks, where 80% more hybrids were found in the seedling stage than in adults (Curtu et al., 2009), and hybrid seedlings monitored over 4 years had a lower survival rate than that of their parental species (Lepais et al., 2013). In *Eucalyptus*, hybrids only began to show abnormal phenotypes after two years, and often died before maturity (Potts & Dungey, 2004).

There is, therefore, evidence of post-zygotic isolation in *Howea* palms driven by selection against hybrids between juvenile and adult stages. Such post-zygotic isolation may have

helped maintain species boundaries during early stages of genomic divergence, but may also have actively driven speciation to completion by reinforcement. During reinforcement, selection for pre-zygotic barriers to reproduction (for example, loci linking flowering time and soil stress tolerance in *Howea*) may actually be increased in response to the cost of hybridisation, with individuals which are more likely to hybridise producing fewer viable offspring and hence having lower fitness (Butlin & Smadja, 2018). Resultantly, mechanisms of post-zygotic isolation in *Howea* may have directly contributed to the speciation process. However, the factors contributing to selection against hybrid palms are currently unknown.

3.1.3 Investigating Hybrid Misexpression in *Howea*

Due to the long generation times of *Howea* palms, which take around ten years to reach maturity, it would be extremely challenging to empirically measure the stages at which *Howea* hybrids undergo selection. It would be even more challenging to do this whilst teasing apart the specific causes of maladaptation in hybrids — for example, by controlling for extreme weather fluctuations, and the wide range of biotic and abiotic stresses associated with the LHI speciation scenario (Osborne et al., 2018; Osborne et al., 2019). This is especially true given that seedling survival experiments, carried out over 36 months, showed that hybrid seedlings performed as well on volcanic soils as *H. belmoreana* individuals (Hipperson et al., 2016) — indicating that selection may act on hybrids at a later stage than 36 months.

What is possible, however, is investigation of gene regulation and misexpression in hybrid individuals. As aforementioned, in other examples of sympatrically diverging populations, there is clear evidence of transgressive gene expression being strongly associated with stunted developmental and morphological traits (Renaut et al., 2009; Renaut & Bernatchez, 2011; Barreto et al., 2015; McGirr & Martin, 2020). In *Howea* we may, therefore, expect to see hybrid misregulation of genes associated with development, gene expression, and functional traits relevant to the biotic and abiotic stresses found on LHI.

In this chapter, I explore hybrid gene regulation by carrying out differential expression analysis between hybrid palms and palms of each parental species growing on different soil types. Through the combination of analysis of hybrid gene expression profiles, identification of transgressive expression, and functional annotation of differentially expressed genes, light is cast on the post-zygotic isolation which may have helped drive speciation on LHI.

3.2 Methods

3.2.1 Tissue Sampling and RNA Sequencing

RNA-seq data for *H. belmoreana* and *H. forsteriana* palms was taken from Dunning et al. (2016), who collected tissues from 19 *H. belmoreana* and 17 *H. forsteriana* individuals over three field trips on LHI during 2010. This data was supplemented with that of eight F1 hybrid palms, collected by the same authors during the same field trips (unpublished data). Full sample metadata including tree coordinates is available in Supplementary Table II.1.

Sampling was carried out across seasonal variation and across the flowering phonologies of each of the *Howea* species, from three tissue types: roots, leaves, and floral inflorescences possessing mature female or male flowers. Individuals of *H. forsteriana* were collected from both volcanic and calcareous sites. Supplementary Table II.2 shows the final numbers of RNA-seq libraries for each species group and tissue type. See Dunning et al. (2016) for full methodology of RNA-seq library construction and sequencing.

3.2.2 Bioinformatic Processing and Reference Transcriptome

Initial data processing was carried out by Osborne et al. (2019), who also used RNA-seq data from Dunning et al. (2016). After primer and adaptor sequence removal, reads with an average PHRED-scaled quality score <20 were removed, before read correction using Rcorrector v.1.0.2 (Song and Florea, 2015) and trimming in Trimmomatic v.0.33 (Bolger et al., 2014) with the settings 'LEADING: 5, TRAILING: 5, SLIDINGWINDOW: 4:5, MINLEN: 75' (as recommended by MacManes, 2014).

Following this initial processing, reads for each individual and tissue were separately mapped using STAR v2.5.3a (Dobin et al., 2013) to the reference *Howea* transcriptome produced by Osborne et al. (2019). The reference transcriptome has 88.7% completeness, with transcripts clustered into putative genes using Corset v.1.06 (Davidson and Oshlack, 2014) by Osborne et al. (2019). Genes in the final transcriptome were annotated by BLAST against *A. thaliana*.

3.2.3 Classification of Hybrids

The above hybrid libraries were verified as coming from F1 hybrid palms by the authors of Dunning et al. (2016, unpublished data), by genotyping using Samtools v1.18 (Danecek et al., 2021) for the same set of SNPs used in Dunning et al. (2016) to assess genetic variation between *H. belmoreana* and *H. fosteriana*. Hybrids were classified into four possible classes (F1, F2, or backcross to each parental species) using a subset of 400 loci with an FST of 0.73–1.00 between *H. belmoreana* and *H. forsteriana*. This was done in NewHybrids (Anderson & Thompson, 2002), specifying prior allele frequencies known from the genotyping of 10 pure parental individuals of each species. All eight individuals used here were classified as F1.

3.2.4 Differential Expression Analysis

For differential expression analysis, libraries were categorised into four groups for each tissue type: *H. belmoreana* (B_{vol}), *H. forsteriana* from volcanic soils (F_{vol}), *H. forsteriana* from calcareous soils (F_{cal}), and F1 hybrids ($F1_{vol}$).

Read counts were first generated for each library using Salmon v1.5.2 (Patro et al., 2007), before analysis with the 3D RNA-seq app (Guo et al., 2020) using R v4.2.2 (R Core Team, 2022). Read counts were normalised using the TMM method (Robinson and Oshlack, 2010) to account for differences in sequencing depth, and filtered to remove putative genes with less than one count per million in at least three libraries for each tissue type. The Limma v3.54 (Ritchie et al., 2015) pipeline was used within 3D RNA-seq to generate differentially expressed genes between each group, for each tissue type. A per-tissue Benjamini-Hochberg correction with an adjusted p-value cut-off of $p<0.05$ was used.

3.2.5 Hybrid Transgressive Expression

To investigate transgressive gene expression in hybrid palms (i.e. expression levels beyond those of the parental species), two methods were applied; firstly, to identify over or under-dominant hybrid expression, and, secondly, to identify genes where variance in hybrid expression extends outside the range of parental expression.

Expression of a gene can be classed as dominant if F1 expression is equal to that of either parental species, semi-dominant or additive if expression in hybrids is intermediate

between that of both parental species, and over/under-dominant if hybrid expression is outside the range of the parents. Genes showing significantly higher expression levels (over-dominant) or significantly lower expression (under-dominant) in hybrids compared to parental species exhibit transgressive inheritance and hybrid misexpression. To identify such loci, I looked for genes with significantly differential expression between F1 hybrids and *H. belmoreana* and between hybrids and *H. forsteriana*, where mean hybrid expression above that of the mean expression of the highest expressing parental species (over-dominant), or below that of the mean expression of the lowest expressing parental species (under-dominant).

Transgressive expression can also be defined as those genes where average expression levels in F1 hybrids do not necessarily differ significantly from average expression in the parental species, but whose range of expression extends beyond that of parental values (Renaut et al., 2009). Genes were classified as demonstrating this form of transgressivity if both the highest and lowest expressing individuals for a given gene were F1 hybrids, with expression therefore extending beyond the combined range of both parental species.

3.2.6 GO Enrichment Analysis

Gene ontology (GO) terms were downloaded from TAIR (Swarbreck et al., 2007) and used to functionally annotate putative *Howea* genes with the GO terms of their *A. thaliana* homologs. GO enrichment tests were carried out using the topGO v2.50.0 (Alexa et al., 2006) package in R, to test for overrepresentation of GO terms in differentially expressed genes between each contrast group (e.g., B_{vol} vs $F1_{vol}$), in each tissue type. These tests were carried out independently for genes up and down-regulated in each comparison, and restricted to the 'Biological Process' category of GO terms. Enrichment tests were also carried out on transgressive genes identified in analysis of hybrid expression profiles for each tissue type.

The *weight* algorithm was used in topGO, with a significance cut-off of $p<0.01$ applied to identify enriched GO terms. In brief, the *weight* algorithm assigns weighted scores to genes annotated with a given GO term, taking into account the scores of neighbouring terms, and thus accounting for the hierarchical nature of gene ontologies (Alexa et al., 2006). For example, for a given ancestor term, *y*, any child terms of *y* which sit further down the GO hierarchy and have a lower enrichment score than *y*, should not be reported significant, as *y* better represents the enriched set of genes. This method helps to remove redundancy and

inter-dependence among GO terms during enrichment analysis. As per the recommendations of Alexa et al., corrections were not made for multiple testing — such corrections should be applied with caution when using the *weight* algorithm, as, for the reasons described, tests are not independent, and adjustments have already been made to enrichment scores. I also argue that GO enrichment tests should, in most instances, be considered exploratory analysis, rather than an attempt to rigorously prove or disprove a given hypothesis. In such exploratory analysis, it can be folly to conservatively overcorrect p-values, as inflating the risk of Type II errors in order to strictly control for Type I serves little function other than to mask interesting results.

3.3 Results

3.3.1 Gene Expression Profiles

In all tissues, the gene expression profile of B_{vol} trees formed a distinct cluster relative to all other groups (F_{cal}, F_{vol} and $F1_{vol}$), as revealed by Principal Component Analysis (PCA, Figure 3.1). F_{cal} and F_{vol} individuals clustered closely together in root and flower tissues, but with the exception of one outlier in F_{vol}, formed distinct clusters in leaf tissue. $F1_{vol}$ trees formed clusters intermediate between each of the parental species in all tissues, though whilst the hybrid cluster was equidistant from parental species across the first two principal components in flower and leaf tissue, it overlapped with F_{cal} and F_{vol} profiles in root tissue, indicating that global F1 hybrid DNA expression is more alike *H. forsteriana* than *H. belmoreana* in roots.

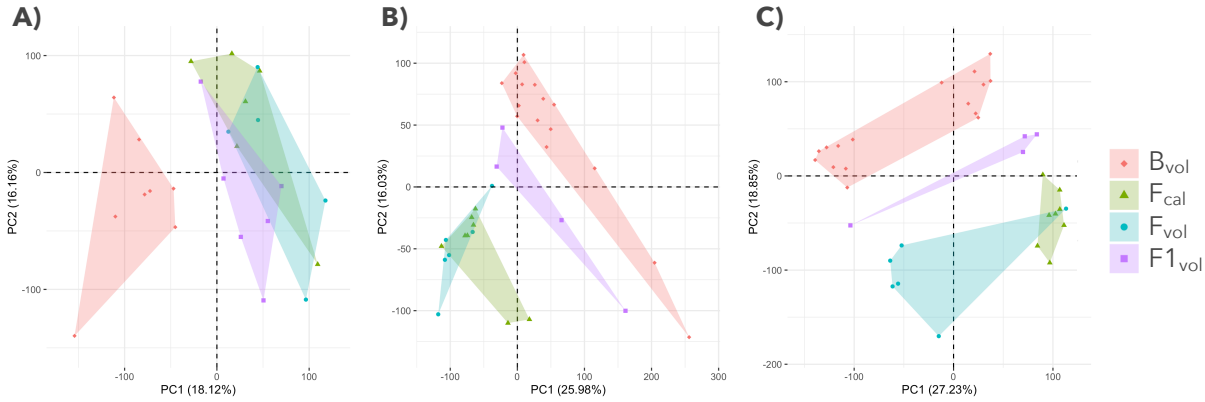


Figure 3.1. Clustering of species groups by gene expression profiles revealed by Principal Component Analyses in **A)** Root tissue; **B)** Flower tissue; **C)** Leaf tissue. In all tissues, B_{vol} formed a distinct cluster separate from F_{cal} and F_{vol}, with $F1_{vol}$ intermediate.

3.3.2 Differential Expression

After filtering for putative genes with expression levels of at least 1 count per million, in at least three libraries, for each tissue type, there were 29,079 putative genes expressed in root tissue, 29,682 in floral tissue, and 28,403 in leaf tissue.

Limma analysis revealed 6,550 genes significantly differentially expressed in at least one pairwise comparison in root tissue, 8,659 genes in floral tissue, and 11,261 in leaf tissue (Figure 3.2). In root tissues, this equated to 4,510 genes differentially expressed between B_{vol} and F_{cal}, 4,448 between B_{vol} and F_{vol}, 2805 between B_{vol} and $F1_{vol}$, 764 between F_{vol} and $F1_{vol}$, and 1187 between F_{cal} and $F1_{vol}$ (Figure 3.3a). The greater number of genes differentially expressed in root tissues between *H. belmoreana* and F1 hybrids than between either of the two *H. forsteriana* groups and F1 hybrids is consistent with PCA analysis (Figure 3.1) of gene expression, which showed F1 trees clustering more closely to *H. forsteriana* in roots than *H. belmoreana*.

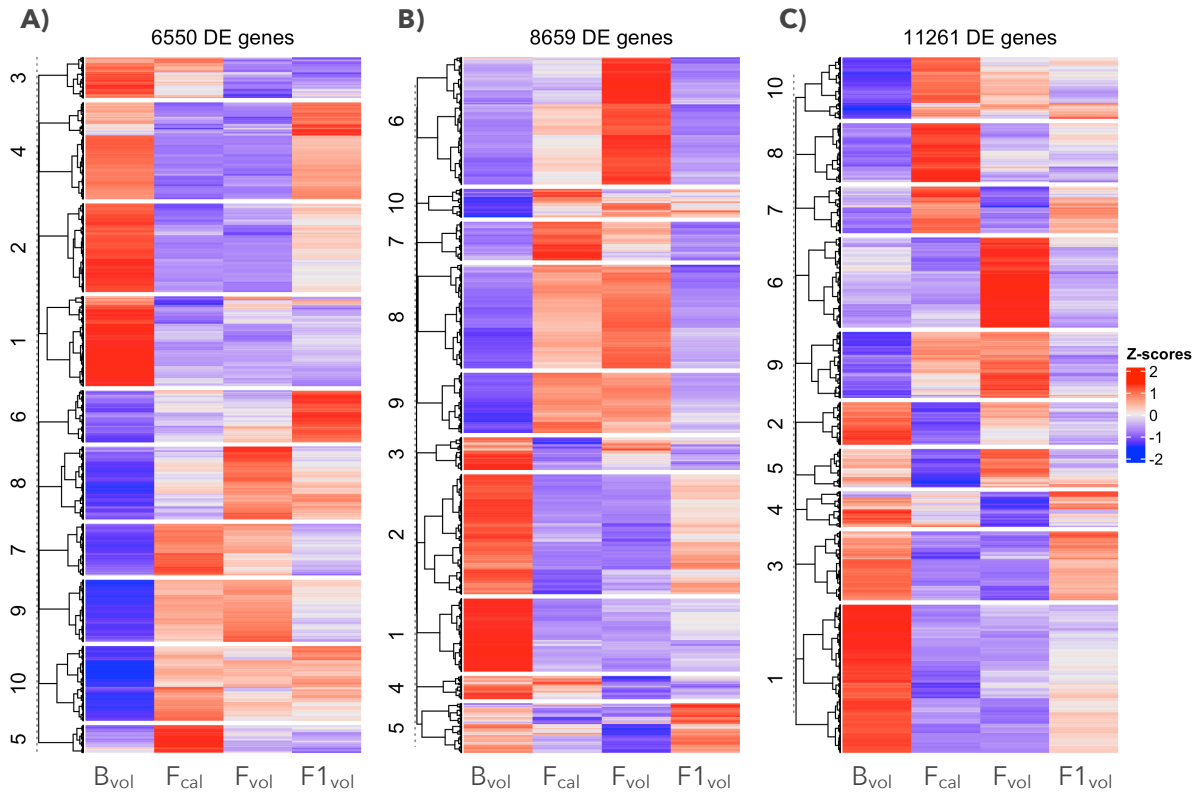


Figure 3.2. Heat maps showing relative expression levels of differentially expressed genes in **A)** Root tissue; **B)** Flower tissue; **C)** Leaf tissue. Positive Z-scores (red) indicate higher relative expression levels.

In floral tissues, 6,049 genes were differentially expressed between B_{vol} and F_{cal}, 6,945 between B_{vol} and F_{vol}, 1,137 between B_{vol} and $F1_{vol}$, 2,297 between F_{vol} and $F1_{vol}$, 1,583 between F_{cal} and $F1_{vol}$, and 3 between F_{cal} and F_{vol} (Figure 3.3b). In leaf tissues, 8,406 genes were differentially expressed between B_{vol} and F_{cal}, 6,647 between B_{vol} and F_{vol}, 1,400 between B_{vol} and $F1_{vol}$, 1,797 between F_{vol} and $F1_{vol}$, 1,439 between F_{cal} and $F1_{vol}$ and 2,190 between F_{cal} and F_{vol} (Figure 3.3c). The 2,190 genes differentially expressed between F_{cal} and F_{vol} in leaf tissues is in striking contrast to the 0 and 3 differentially expressed genes between F_{cal} and F_{vol} in root and floral tissues respectively.

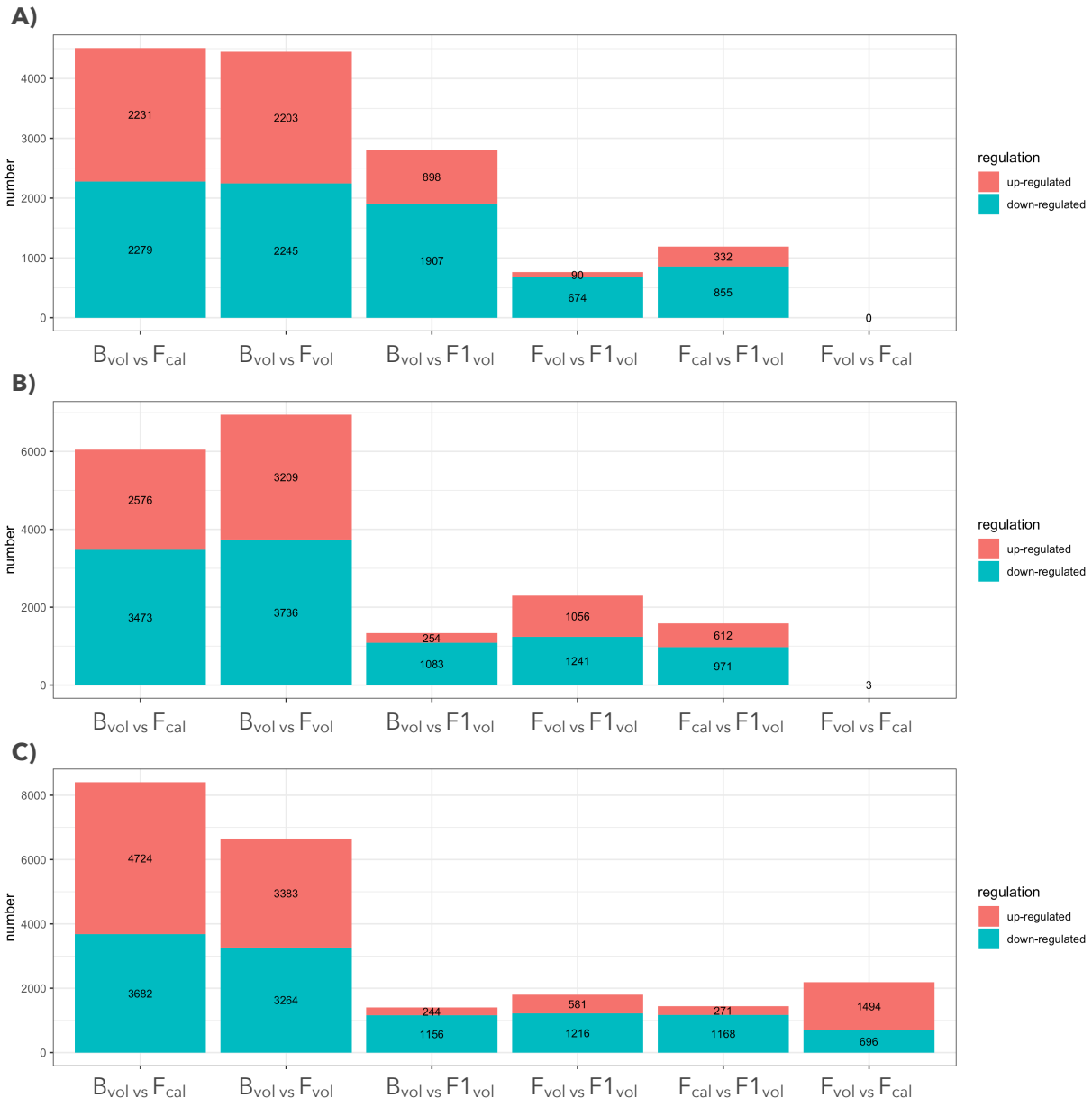


Figure 3.3. Numbers of differentially expressed genes up and down-regulated between each contrast group in **A)** Root tissue; **B)** Flower tissue; **C)** Leaf tissue. In each contrast, the first species group is the group up or down-regulated. For example, in B_{vol} vs F_{cal} in root tissue, 2231 genes are up-regulated in B_{vol} relative to F_{cal}.

In all tissues, genes differentially expressed between F1 hybrids and each of the parental groups were always significantly more likely to be up-regulated in hybrid trees than would be expected under a neutral model of 50% up-regulation and 50% down-regulation (Chi-squared tests, $p<0.002$ in all tests, see Supplementary Table II.3 for full test statistics). For example, in leaf tissues, 83% of differentially expressed genes between B_{vol} and $F1_{vol}$ were up-regulated in hybrids, 68% between F_{vol} and $F1_{vol}$, and 81% between F_{cal} and $F1_{vol}$.

3.3.3 GO Enrichment of Differentially Expressed Genes

Tests for GO enrichment among differentially expressed genes revealed 160 unique GO terms enriched in at least one test in root tissues, 192 in flower tissue, and 277 in leaf tissue. Figure 3.4 shows the numbers of enriched GO terms in each independent test. A wide range of GO terms relating to the *Howea* speciation scenario were found, with enrichment of GO terms amongst genes differentially expressed between *H. belmoreana* and *H. forsteriana* on either soil showing strong concordance with the results of Osborne et al. (2019), who used the same RNA-seq libraries for *H. belmoreana* and *H. forsteriana.* This included multiple terms associated with flowering time and abiotic stress tolerance.

Across all tissues, 87 different GO terms were significantly enriched in genes down-regulated in $F1_{vol}$ relative to F_{vol} or F_{cal}, that were not also enriched in genes down-regulated in B_{vol} relative to F_{vol} or F_{cal} in the same tissue. This included multiple terms associated with the *Howea* speciation scenario, such as "regulation of response to water deprivation", "maintenance of floral meristem identity", "regulation of stomatal opening" and "defence response to fungus". Several enriched terms also indicated the potential for hybrid misexpression, including "DNA-templated transcription", "regulation of DNA-templated transcription", "regulation of transcription from RNA polymerase II promoter in response to oxidative stress" and "mRNA export from nucleus".

Genes down-regulated in $F1_{vol}$ relative to B_{vol} were enriched for 41 terms that were not also enriched in genes down-regulated in F_{vol} or F_{cal} relative to B_{vol} in the same tissue. As above, this included multiple terms which are associated with the *Howea* speciation scenario or indicative of a potential cause of misexpression, including: "response to osmotic stress", "cellular response to calcium ion", "regulation of transcription by RNA polymerase II",

"negative regulation of nucleic acid-templated transcription", and "regulation of mRNA polyadenylation".

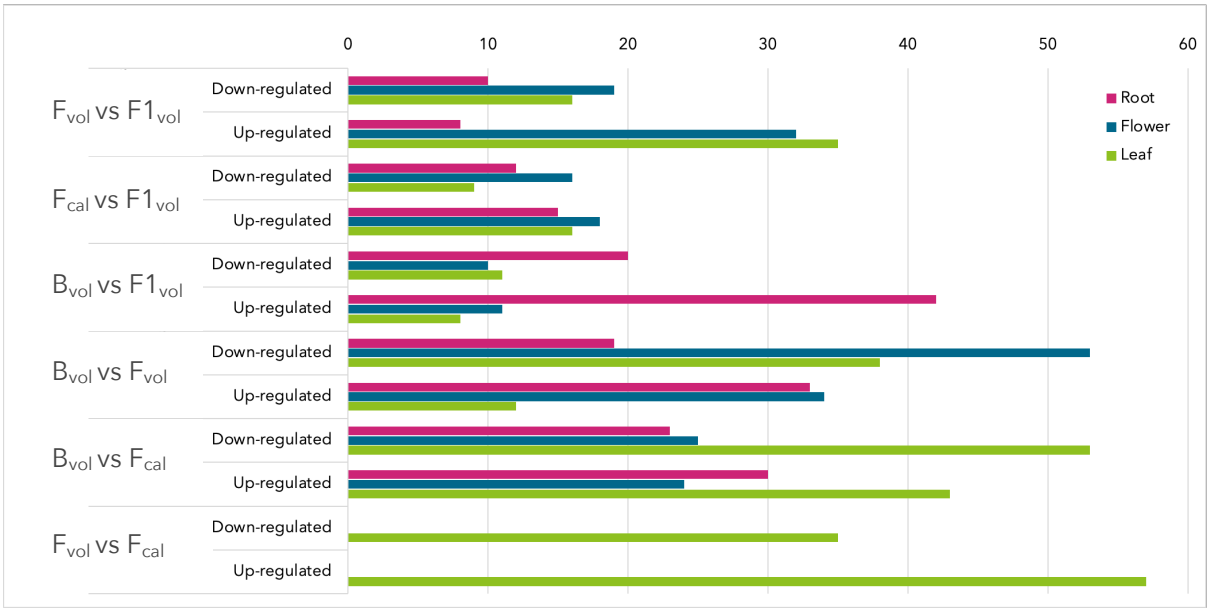


Figure 3.4. The numbers of significantly enriched GO terms in differentially expressed genes between each contrast group and tissue. In all cases, the first species group is the group up or down-regulated. For example, in F_{vol} vs $F1_{vol}$ in root tissue, 10 GO terms are enriched in genes down-regulated in F_{vol} relative to $F1_{vol}$.

For genes up-regulated in $F1_{vol}$ relative to F_{vol} or F_{cal}, 52 GO terms were enriched that were not also enriched in genes up-regulated in B_{vol} relative to F_{vol} or F_{cal} in the same tissue. These terms included "negative regulation of flower development", "meristem determinacy", "zinc ion transport" and "response to karrikin", which was enriched in genes up-regulated in $F1_{vol}$ relative to F_{cal}. Karrikin response genes are known to be involved in the formation of arbuscular mycorrhizal fungi (AMF) symbioses in rice (Gutjahr et al., 2015), and Osborne et al. (2018) showed that arbuscular mycorrhizal fungi (AMF) may have played an important role in the *Howea* speciation process, with AMF significantly depleted in *H. forsteriana* growing on volcanic soils relative to both *H. forsteriana* growing on calcareous soils and *H. belmoreana*. Again, a variety of terms associated with gene expression were also enriched, such as "DNA methylation or demethylation", "epigenetic regulation of gene expression", "alternative mRNA splicing, via spliceosome", "regulation of DNA replication", "mRNA polyadenylation" and "tRNA export from nucleus".

28 GO terms enriched for genes up-regulated in $F1_{vol}$ relative to B_{vol}, which were not also enriched in genes up-regulated in F_{vol} or F_{cal} relative to B_{vol} in the same tissue, followed the same pattern. Here, terms included “regulation of flower development”, “manganese ion transport”, “sodium ion import across plasma membrane”, “fruit replum development”, "double-strand break repair via homologous recombination”, and “interstrand cross-link repair”.

Finally, in leaf tissues, genes up-regulated in F_{vol} relative to F_{cal} were enriched for 57 GO terms, and genes down-regulated in F_{vol} relative to F_{cal} were enriched for 35 GO terms. Again, these terms contained a variety of functions associated with the *Howea* speciation hypothesis, but, notably, the GO term “response to fungus” was enriched in genes up-regulated in F_{vol}, and the term “response to symbiotic fungus” was enriched in genes down-regulated in F_{vol}. Additionally, Osborne et al. (2018) highlighted that two genes involved in the formation of AMF symbioses, *CCD8a* and *AMT2,* show signatures of divergent selection between *H. belmoreana* and *H. forsteriana. AMT2* is an ammonium transporter known to be specific to nitrogen uptake from mycorrhizae in the legume *Lotus japonicus* (Guether et al., 2009), and differential expression analysis here revealed that one putative *Howea* ortholog of *AMT2* was significantly down-regulated in the leaves of *H. forsteriana* growing on volcanic soil relative to *H. forsteriana* growing on calcareous soil. This may offer further support to the findings of Osborne et al. (2018) — that AMF are significantly depleted in *H. forsteriana* growing on volcanic soils, and that AMF symbioses may have played a significant role during *Howea* speciation.

3.3.4 Hybrid Transgressive Expression

In root tissues, 44 putative genes were identified that showed significant over-dominance in $F1_{vol}$. There were also seven over-dominant genes in flower tissue, and five in leaf tissue. Only seven significantly under-dominant genes were identified, four in roots and three in flower tissue. The over and under-dominant genes in each tissue were unique — i.e. over or under-dominance was never found for the same gene in two tissues. A number of transgressive genes were identified in each tissue where variance of expression levels in hybrids extended outside the parental range: 192 in roots, 224 in flowers and 124 in leaves.

Genes with a range of expression in $F1_{vol}$ roots extending outside the range of parental expression were significantly enriched for one GO term with a p-value <0.01 — "lagging strand elongation" — associated with DNA replication. However, I wish to also highlight that the terms "cell cycle G1/S phase transition" and "regulation of meiotic nuclear division" were enriched with significance levels of p=0.0125 and p=0.0186, respectively. Differences in growth, cell differentiation, and abundance of different cell types caused by abnormal development in hybrids can produce transgressive expression (Swain Lenz et al., 2014; McGirr & Martin, 2020), and so enrichment of such terms is notable.

In floral tissues, 10 terms were significantly enriched. This included terms associated with the LHI speciation scenario — "response to pH" and "drought recovery" — and, again, terms associated with gene expression and development: "formation of translation preinitiation complex", "cell morphogenesis involved in differentiation" and "lignin metabolic process". Lignin metabolism plays a key role in plant growth and development, with interference in this process often leading to deformity and inhibition of growth (Liu et al., 2018).

The pattern remained consistent in leaves, where six GO terms were enriched in genes for which $F1_{vol}$ expression extends beyond the range of the parents. The terms included "regulation of cell cycle G1/S phase transition", "G1/S transition of mitotic cell cycle", "regulation of cell growth", "translational elongation", and "circadian rhythm". Whilst not significant to p<0.01, the three terms closest to the p<0.01 threshold should again be highlighted: "asymmetric cell division" (p=0.0132), "cytokinesis by cell plate formation" (p=0.01707) and "regulation of nuclear division" (p=0.01781).

GO annotations of over-dominant genes only reinforced this pattern. Of the 56 genes identified as over-dominant, 40 were annotated with GO terms in the 'Biological Process' category. Strikingly, 24 of these 40 genes were annotated with terms directly associated with response to abiotic and/or biotic stimuli relevant to the *Howea* speciation scenario, including 14 possessing terms related to drought stress, such as "response to water deprivation" and "response to osmotic stress"; and 11 with terms related to response to biotic stimuli, such as "defence response to fungus" and "defence response to bacterium". In addition to differences in AMF symbioses between *Howea* species growing on different soils, Osborne et al. (2018)

showed that both fungal and bacterial plant pathogens are differentially abundant between volcanic and calcareous soils; and differentially abundant between the two parental species — even when both growing on volcanic soil. There were also 13 genes annotated with terms directly related to embryonic and organ development, including "lateral root development", "regulation of seedling development" and "flower development". In the seven under-dominant genes, there were three genes involved in abiotic stress tolerance, and two involved in gene expression — one gene annotated with the terms "DNA synthesis involved in DNA repair" and "DNA-templated DNA replication"; and another with the term 'negative regulation of translation'.

3.4 Discussion

3.4.1 Differentially Expressed Genes are Up-regulated in Hybrids

Across all tissues, genes differentially expressed between hybrids and any of the parental groups were significantly more likely to be up-regulated in hybrids (Chi-squared tests, $p<0.002$ in all tests). Between $F1_{vol}$ and B_{vol}, 74.8% of differentially expressed genes were up-regulated in hybrids, between $F1_{vol}$ and F_{vol} it was 64.5%, and between $F1_{vol}$ and F_{cal}, 71.1%. Meanwhile, between *H. belmoreana* and either *H. forsteriana* group, almost exactly 50% of differentially expressed genes were each up and down-regulated (49.7% of differentially expressed genes down-regulated in B_{vol} relative to F_{cal}, and 51.2% relative to F_{vol}), suggesting no overall directional shift in expression levels in either parental species. The fact that differentially expressed genes in hybrids were typically up-regulated relative to both parental species also indicates that hybrid expression patterns are not as a result of dominance relationships between alleles of *H. belmoreana* and *H. forsteriana.* This has been suggested in interpopulation hybrids of *Tigriopus californicus* copepods, where differentially expressed genes in hybrids were significantly more likely to be up-regulated relative to one parental population, but down-regulated relative to the other (Barreto et al., 2015).

Instead, significant up-regulation in hybrid *Howea* palms may indicate incompatibilities in regulatory alleles associated with restricting gene expression, and/or may indicate compensatory expression, with elevated expression levels attempting to make up for aberrant development or dysfunction in hybrid palms. Indeed, GO enrichment analysis of differentially expressed genes provides evidence for both scenarios.

Regardless of the parental group being compared against, and regardless of whether genes were up or down-regulated, differentially expressed genes in hybrids were consistently enriched for GO terms associated with organ development and with stress response relevant to the LHI speciation scenario. In particular, development of floral organs, and response to water deprivation and metal ions. It is plausible that hybrid dysfunction, for example due to incompatibilities or misexpression affecting development, reduces hybrid tolerance to common abiotic stresses found on LHI, thereby enforcing compensatory expression through up-regulation of stress response loci.

Simultaneously, evidence for misregulation of expression was also common in differentially expressed genes. Again, regardless of the pairwise comparison, and regardless of whether genes were up or down-regulated, GO terms were always found relating to gene expression and/or DNA repair. Most notably, however, genes down-regulated in hybrids relative to F_{vol} and/or F_{cal} were enriched for the terms "regulation of DNA-templated transcription" and "regulation of transcription from RNA polymerase II promoter in response to oxidative stress"; and genes down-regulated in hybrids relative to *H. belmoreana* were enriched for the term "negative regulation of nucleic acid-templated transcription". Down-regulation of genes involved in controlling expression in response to oxidative stress, and down-regulation of genes involved in suppressing transcription, further supports the scenario in which incompatibilities affecting expression levels, particularly related to abiotic stress response and development, may in turn lead to compensatory up-regulation in hybrid palms.

3.4.2 Transgressive Hybrid Expression at a Relatively Small Number of Loci Associated with Development and Stress Response

Two methods were used to identify loci demonstrating transgressive expression in hybrids, and both methods highlighted relatively few genes with expression beyond parental levels. In root tissues, 192 genes were identified where hybrid expression extended both above and below the combined range of the parental species — 0.66% of all genes expressed in root tissues. There were 224 such genes in flowers, and 124 in leaves, equating to 0.75% and 0.44% of genes expressed in those tissues respectively. GO enrichment amongst these genes was almost exclusively restricted to terms involved in development, cell division and

differentiation, gene expression and response to abiotic stress. This only reinforces the patterns seen in genes differentially expressed in hybrid palms, especially as differences in cell differentiation and relative abundances of cell types are known to affect gene expression (Swain Lenz et al., 2014).

The second method used to identify transgressive expression was looking for genes with significant over or under-dominance in hybrid palms. Even fewer genes demonstrated this form of misexpression — with a total of 56 over-dominant and only seven under-dominant genes identified across all three tissues, together representing only 0.22% of expressed genes. As was the case with genes demonstrating transgressive expression extending both above and below the range of parental individuals, over-dominant genes were strikingly associated with developmental processes and stress response, particularly response to water deprivation. Meanwhile, in consistence with differentially expressed genes in hybrids that are down-regulated relative to any parental group (discussed above), the seven genes showing significantly under-dominant hybrid expression included two associated with regulation of gene expression. One of these genes, an ortholog of the *Arabidopsis* gene *UL13Z*, was annotated with the term "negative regulation of translation" — significant down-regulation in such a gene could contribute to the over-expression seen in the majority of genes differentially expressed in hybrids.

Whilst the number of genes demonstrating transgressive expression is low, the functions of these genes show remarkable consistency with the prediction that reduced hybrid fitness in *Howea* would be associated with aberrant development and reduced tolerance to the variety of biotic and abiotic stresses found on LHI. Evidence is also consistent with a scenario where dysfunction in hybrids — potentially caused by the breakup of co-adapted genetic architectures at regulatory alleles, leading to misregulation of expression — has in turn produced compensatory up-regulation of expression in hybrids, at key loci involved in development and stress response.

The results presented here also show strong concordance with the results of Barreto et al. (2015). In many studies investigating genome-wide divergence in expression levels of interspecific hybrids, under or over-dominant transgressive expression has often been found in a large portion of the genome, from around 10% (Dion-Côté et al., 2014; Mack et al., 2016;

McGirr & Martin, 2020) up to 80% of genes (Ranz et al. 2004; McManus et al. 2010; Coolon et al. 2014). In contrast, Barreto et al. (2015), studying hybridisation between two populations of the copepod species *T. californicus*, found that 1.2% of genes showed over or under-dominance in hybrids; far closer to the numbers observed here. Notably, they also found that around 80% of genes demonstrating this form of transgressive inheritance were over-dominant, consistent with the 89% found in the present work. Again, this is in contrast to many other studies, which have typically found significant bias towards under-dominance in hybrids (Landry et al. 2007; McManus et al. 2010; Coloon et al. 2014). Finally, Barreto et al. found that none of the genes showing transgressive expression in the copepod hybrids were also differentially expressed between parental populations; and this is exactly what was found here.

As argued by Barreto et al. (2015) and McGirr & Martin (2020), these contrasting patterns likely reflect differences in the regulatory mechanisms and selective pressures underlying gene expression. When directional selection affects gene regulation, or where there are relaxed constraints on expression, one would expect to see hybrid misexpression in genes which are also differentially expressed between parental species. Alternatively, where stabilising selection facilitates optimal levels of gene expression shared across species, misexpression in hybrids is more likely to arise from novel epistatic effects of allele variants involved in compensatory regulation (Barreto et al., 2015; McGirr & Martin, 2020). This provides further evidence to suggest that much of the transgressive up-regulation in hybrid palms may be a compensatory physiological response, consistent with the findings of Barreto et al. (2015).

3.4.3 Differential Expression of Genes Involved in Arbuscular Mycorhizzal Symbiosis Associated with Soil Type in *H. forsteriana*

Whilst the focus of Chapter Three is investigation of misexpression and potential causes of reduced hybrid fitness in Howea, the methods used also allowed investigation of differential expression between *H. forsteriana* individuals growing on volcanic soils and individuals growing on calcareous soils. Osborne et al. (2019) have previously investigated differential expression in *H. forsteriana* based on soil type, but not in exhaustive detail. As would be expected, concordance was found in the relative numbers of genes in each tissue showing differential expression based on soil type, with Osborne et al. also finding very few

genes with differential expression in roots and flower tissue, but far more in leaves. Here, however, genes involved in the response to symbiotic fungus were shown to be significantly over-represented in genes down-regulated in F_{vol} relative to F_{cal}. Additionally, a *Howea* ortholog of *AMT2*, a key ammonium transporter involved in nitrogren uptake during AMF symbiosis (Guether et al., 2009), was shown to be down-regulated in *H. forsteriana* on volcanic soils. This is indicative of either reduced AMF colonisation in roots on volcanic soils, or a decreased ability for *H. forsteriana* palms to uptake nutrients from AMF on volcanic soils. Either way, the results provide support to the findings of Osborne et al. (2018), who showed that *H. forsteriana* palms growing on volcanic soils benefit from AMF symbioses significantly less than *H. belmoreana* and *H. forsteriana* growing on calcareous soils.

3.4.4 Conclusion and Wider Context

Whilst the number of loci exhibiting transgressive hybrid expression in *Howea* is low, repeated lines of evidence showed striking overrepresentation of genes involved in development and stress response among differentially expressed and misexpressed loci. This is consistent with the prediction that selection against hybrids before the adult stage would likely reflect abnormal development and inability to tolerate the wide range of stresses found on LHI — for example, periods of severe drought. The fact that many genes involved in these functions showed significant over-expression in hybrids suggests a scenario in which initial regulatory incompatibility may have produced hybrid dysfunction, such as reduced drought tolerance, with compensatory up-regulation then necessary to try and mitigate maladapted phenotypes.

As discussed by McGirr & Martin (2020), incompatibilities associated with ecological selection (for example, the divergent selection acting on soil stress tolerance between *H. forsteriana* and *H. belmoreana*) and reduced hybrid fitness have the potential to promote rapid speciation through post-zygotic isolation (Kulmuni & Westram, 2017) and reinforcement (Butlin & Smadja, 2018); encouraging lineage-specific assortative mating and adaptive divergence, in a manner similar to magic traits (Servedio et al., 2011). Until now, investigation of such mechanisms in a convincing and well-understood example of pure sympatric speciation has been absent, but the present work does just that, and provides strong evidence for the plausibility of such a scenario in facilitating the speciation of *Howea* palms. Of course, demonstrating that misexpression is directly causative of reduced hybrid fitness in *Howea* is

challenging, especially due to the extreme difficulty in studying selection against hybrid palms empirically. However, the consistency of evidence presented for transgressive hybrid expression, associated with development, gene regulation, and stress tolerance, convincingly demonstrates a role for genomic incompatibility and hybrid misexpression in the post-zygotic isolation and speciation of *Howea*.

Chapter Four - The Role Of Arbuscular Mycorrhizal Fungi In Sympatric Niche Divergence

4.1 Introduction

4.1.1 Niche Differentiation in the *Howea* Palms

When two species occupy the same niche, it is commonly accepted that over time one species will outcompete the other, to the point of extinction, unless the two species undergo niche divergence to allow coexistence (MacArthur & Levins, 1967; Chesson, 2000). Upon LHI, whilst *H. forsteriana* is able to inhabit the calcareous soils on which *H. belmoreana* cannot survive, both species occur intermingled on volcanic soils across the island, with a significant overlap in ranges. One would, therefore, expect competition between the species; and, if *H. forsteriana* were outcompeting *H. belmoreana*, it would likely drive *H. belmoreana* to extinction.

However, *H. belmoreana* palms are found to be more numerous on volcanic soils than *H. forsteriana* (Savolainen et al., 2006a), and have been shown to have higher rates of germination and seedling survival on volcanic soils than *H. forsteriana* (Hipperson et al., 2016). This evidence suggests that, despite extensive range overlap, adaptive niche divergence has taken place, which allows *H. forsteriana* to colonise calcareous soils and simultaneously allows *H. belmoreana* to maintain a competitive advantage on volcanic soils — leading to apparent stable coexistence of the two species. Such niche differentiation is, therefore, key to the speciation of *Howea*, with soil substrate differences likely driving initial divergence between the two species (Savolainen et al., 2006a; Dunning et al., 2016), but also vital to maintaining species boundaries and facilitating continued coexistence. Understanding the mechanisms behind this niche divergence, and particularly the mechanisms affording a competitive advantage to *H. belmoreana* on volcanic soils, are therefore key to understanding the sympatric speciation and continued adjacent coexistence of *Howea*.

4.1.2 Arbuscular Mycorrhizal Fungi Promote Coexistence

Previous research carried out by Osborne et al. (2018) has, in fact, proposed an intriguing explanation for stable niche divergence on LHI. Using ribosomal-based metabarcoding, the authors examined the fungal and prokaryotic microbiomes found in the roots and surrounding soils of both *H. belmoreana* and *H. forsteriana*, investigating the biotic selection pressures which may exist on LHI. The results were highly interesting, finding that diversity of both prokaryotes and fungi differed significantly between soil types, and that a range of bacterial and fungal plant pathogens vary between soil types and species. More notably, however, arbuscular mycorrhizal fungi (AMF) were found to be significantly depleted in *H. forsteriana* growing on volcanic soil, relative to both *H. belmoreana* and *H. forsteriana* growing on calcareous soil. Overall species richness of AMF was lower in *H. forsteriana* growing on volcanic soils, and significantly differentially abundant AMF taxa between *H. belmoreana* and *H. forsteriana* on volcanic soils; and between *H. forsteriana* on calcareous soils versus *H. forsteriana* on volcanic soils, were always less abundant in *H. forsteriana* on volcanic soils.

Furthermore, seedling survival assays on sterilised versus unsterilised soil showed a significant decrease in survival for *H. belmoreana* seedlings growing on both calcareous and volcanic sterilised soils; but for *H. forsteriana*, a decrease in survival was only seen in seedlings growing on sterilised calcareous soils. For *H. forsteriana* seedlings on volcanic soil, sterilisation had no effect on survival, convincingly demonstrating that *H. forsteriana* does not benefit from the microbiome of volcanic soils. Finally, Osborne et al. (2018) examined *Howea* RNA-seq data from Dunning et al. (2016), and found evidence of adaptive divergence between the two species in the coding sequences of multiple genes associated with AMF symbioses. Firstly, in the *CCD8a* gene, which is part of the strigolactone biogenesis pathway (Gomez-Roldan et al., 2008). Strigolactones released from plant roots are known to be key to forming AMF symbioses (Akiyama et al., 2005). Secondly, the ammonium transporter *AMT2* showed evidence of adaptive divergence between the two species, which is known to be involved in nitrogen uptake from AMF symbionts in plants (Guether et al., 2009).

These differences are notable, as over 80% of vascular plant species form AMF symbioses, which are likely key to plant fitness, and probably played a significant role in the evolutionary success of land plants (Smith & Read, 1997; Brundrett, 2002; Field et al., 2015).

In return for carbohydrates, AMF provide plants with a variety of nutrients, such as phosphorous, sulfur and nitrogen; and these mutualistic relationships are known to benefit other palm species (Moora et al., 2011; Al-Karaki, 2013). Given this, the results of Osborne et al. (2018) provide a strong argument to suggest that *H. belmoreana* has a competitive advantage on volcanic soils, due to the fact that *H. forsteriana* individuals do not benefit from AMF when growing on volcanic soils.

This may have affected the speciation process in two key ways. Firstly, the initial colonisation of calcareous soils by the ancestor of *H. forsteriana* may have involved a trade-off in the benefit *H. forsteriana* gains from the native microbiome in each soil type. Novel adaptation to optimise benefit from the calcarenite microbiome, allowing *H. forsteriana* to survive on those soils — where *H. belmoreana* and hybrid palms cannot — may have been associated with a simultaneous loss of ability to form mutualistic relationships with AMF on volcanic soils. For example, as hypothesised by Osborne et al. (2018), due to a change in chemical triggers needed to produce the root exudates that trigger AMF colonization, such as a change in the threshold for phosphate starvation. The change in microbiome/soil type could, in turn, have driven the flowering time divergence seen in *Howea*. In fact, several studies have demonstrated that microbiota (likely including AMF) can affect flowering time in plants (Gaur et al., 2000; Korves & Bergelson, 2003; Salvioli et al., 2012; Wagner et al., 2014; Jin et al., 2015; Lyons et al., 2015; Panke-Buisse et al., 2015). In this scenario, AMF may have directly contributed to ecological divergence, acting as a key driver of sympatric speciation. Alternatively, or concurrently, the results of Osborne et al. highlight a clear role for differences in AMF symbiosis in maintaining species boundaries and allowing prolonged coexistence, despite apparently large niche overlap. In both scenarios (driving initial divergence, and/or promoting coexistence), the evidence is strong that AMF associations have played a key mechanistic role in facilitating sympatric speciation.

4.1.3 The Role of AMF in *Howea* Warrants Further Study

Whilst the results of Osborne et al. (2018) are highly convincing, there are caveats to the work, and unexplored avenues of further research which would build further upon the study. Firstly, the study carried out by Osborne et al. amplified the ITS2 gene to assess the fungal microbiome in *Howea* roots. The Internal Transcribed Spacer (ITS) region of ribosomal DNA is commonly used to investigate broad fungal taxonomic richness, owing to its suitability in

detecting a wide range of fungi (Schoch et al., 2012), however, it is not optimised for assessment of Glomeromycota (the phylum to which AMF belong). Given the clear role of AMF, rather than other fungal taxa, highlighted by Osborne et al., it is prudent to further investigate AMF diversity and differentiation between *Howea* species. In the present work, I use AMF specific 18s rRNA primers, which are likely to detect significantly more AMF taxa (Suzuki et al., 2020). Through such methods, I aim to both confirm the results of the previous study, and potentially elucidate further detail on differential AMF symbioses between the *Howea* species.

Secondly, the work by Osborne et al. did not investigate the AMF symbioses formed by hybrid palms. As discussed in Chapter Three, hybrids are more commonly found as juveniles than as adults, indicating selection acting against them before the adult stage (Babik et al., 2009; Hipperson et al., 2016). Additionally, adult hybrids have never been found on calcareous soils (Hipperson et al., 2016). It is possible that AMF symbiosis, or lack thereof, in hybrid palms contributes to these patterns — for example, depleted AMF in hybrids may contribute to reduced fitness. Post-zygotic isolation driven by low hybrid fitness could have been key to *Howea* speciation, so the potential role of AMF in this post-zygotic isolation is explored in this chapter.

Finally, I delve further into the role of AMF in *Howea* by combining gene expression data with metagenomics. There are a range of scenarios in which a relationship between gene expression and AMF may be present, with a plethora of studies showing that gene expression can be drastically altered by AMF symbiosis (e.g.; Liu et al., 2007; Ortu et al., 2012; Cervantes-Gámez et al., 2016; Vangelisti et al., 2018). Firstly, external stimuli may trigger expression changes in plants which lead to initiation of AMF symbiosis — for example, strigolactones, which are key to AMF symbioses, are secreted by plant roots in response to nutrient deficiencies, such as phosphate starvation (Jamil et al., 2014). In this case, expression of genes associated with the formation of AMF symbiosis may be expected to be positively correlated with AMF presence. Secondly, once AMF associations have formed, plants may alter gene expression in order to maximise benefit obtained from AMF symbionts, for example by altering expression of genes involved in nutrient uptake — again, one may therefore expect a positive correlation between AMF presence and expression of genes

involved in nutrient transport. Additionally, signalling from AMF symbionts may also induce AMF-mediated changes in gene expression.

Through correlation analysis of expression levels of every *Howea* gene, in multiple tissues, against AMF diversity, the present work casts light on the complex relationship between AMF diversity and gene expression in *Howea*. Analysis reveals how mycorrhizal fungi affect gene expression differently in different plant organs, uncovers genes which may be key to AMF symbiosis in *Howea,* and explores the functions of genes for which regulation is affected by AMF colonisation.

4.2 Methods

4.2.1 Tissue Sampling and RNA-seq Libraries

During three field trips on LHI in 2010, the authors of Dunning et al. (2016) collected root tissue samples from 20 *Howea* trees, comprising eight *H. belmoreana* trees from volcanic soils (B_{vol}), three *H. forsteriana* individuals growing on volcanic soils (F_{vol}), five *H. forsteriana* individuals from calcareous soils (F_{cal}), and four F1 hybrids ($F1_{vol}$) from volcanic soils. Root tissue was stored in RNAlater (Sigma-Aldrich, USA) and frozen at -80°C. Sample collection coordinates and metadata are available in Supplementary Table III.1. The four hybrids used here are among those previously classified as F1 individuals by genotyping, as described in Chapter 3.2.

For 18 of these root samples, RNA-seq libraries (as described in Chapter 3.2) from the same individual were also available from at least one tissue (root, flower or leaf). This produced datasets of 10 individuals for which root samples and RNA-seq libraries from root tissue were available, 13 for root samples and RNA-seq data from flower tissues, and 14 for root samples and leaf tissue RNA-seq libraries (Supplementary Table III.2).

4.2.2 DNA Extraction and Sequencing

For all 20 root samples, total DNA extractions were carried out using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germany) with c. 0.1g of root tissue and following the manufacturer's instructions. Tissue disruption was carried out by hand using pestle and

mortar, and optional RNase-A was added to degrade unwanted RNA. The purity and concentration of final DNA elutes was assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and Qubit 4.0 Fluorometer with the dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, USA) respectively.

A nested-PCR approach was used to amplify the Glomeromycotan 18s rRNA gene, using the primer pairs AML1 (5'-ATCAACTTTCGATGGTAGGATAGA-3') and AML2 (5'-GAACCCAAACACTTTGGTTTCC-3') (Lee et al., 2008) to first amplify a region of approximately 800bp. The resultant amplicons were then used as the template for a second PCR reaction using the primer pair AMV4.5NF (5'-AAGCTCGTAGTTGAATTTCG-3') and AMDGR (5'-CCCAACTATCCCTATTAATCAT-3') (Sato et al., 2005) — which generates amplicons of approximately 300bp. Illumina overhang adapter sequences were added to the AMV4.5NF and AMDGR primers. Out of six primer combinations studied by Suzuki et al. (2020), this nested method was found to produce the highest frequency of AMF sequences in amplicons (as opposed to non-AMF taxa), combining the high AMF specificity of AML1/AML2 (Van Geel et al., 2014) with the reduced amplicon size of AMV4.5NF/AMDGR, which is compatible with the high-throughput Illumina Miseq sequencer (Illumina, USA).

The first PCR using primer pair AML1/AML2 was carried out under the following conditions: initial denaturation for 5 minutes at 95°C, 35 cycles of 60s at 95°C, 60s at 57°C and 60s at 72°C, and final elongation for 7 minutes at 72°C. The AMV4.5NF/AMDGR reaction conditions were: initial denaturation for 5 minutes at 95°C, 35 cycles of 60s at 95°C, 60s at 58°C and 60s at 72°C, and final elongation for 7 minutes at 72°C. Both reactions were made up of 12.5µl 2x Kapa HiFi Hot Start Ready Mix (Roche, Switzerland), 2.5µl DNA template, 1µl of each of the forward and reverse primers at 10µM, and 8µl PCR-grade water. PCR product from the first reaction was diluted 1:10 in PCR-grade water to produce template for the second reaction. A final PCR reaction was then carried out to attach Illumina sequencing dual indices to DNA libraries using the Nextera XT Index Kit V2 (Illumina, USA). This reaction contained 25µl 2x Kapa HiFi Hot Start Ready Mix (Roche, Switzerland), 5µl amplicon product from the AMV4.5NF/AMDGR reaction, 5µl Nextera XT Index 1, 5µl Nextera XT Index 2, and 10µl PCR-grade water. Cycling conditions as follows: 95°C for 3 minutes, 8 cycles of 30s at 95°C, 30s at 55°C and 30s at 72°C, and final

elongation for 5 minutes at 72°C. After each of the PCR reactions, amplicons were visualised on an agarose gel and purified using Agencourt Ampure XP beads (Beckman Coulter, USA).

Final DNA libraries from each individual were then quantified using a Qubit 4.0 fluorometer (Thermo Fisher Scientific, USA) and normalised to equal concentrations, before being pooled for sequencing. Pooled libraries were sent to Azenta (USA) for sequencing on an Illumina Miseq with paired end 2 × 250 bp sequencing using the MiSeq reagent kit v2 (Illumina, USA).

4.2.3 Bioinformatic Processing

Demultiplexing and adapter-trimming was carried out by Azenta, before reads were passed to a DADA2 (Callahan et al., 2016) pipeline for quality filtering, denoising, merging of paired reads, chimera removal and calling of exact sequence variants (ESVs). After visual inspection of quality plots, forward reads were truncated at 240bp, and reverse reads truncated at 200bp. Standard filtering parameters of 'maxN=0', 'truncQ=2' and 'maxEE=2" were used for filtering and error correction. The IDTAXA classifier (Murali et al., 2018) was used for taxonomic assignment of ESVs against the SILVA reference database of SSU ribosomal RNA (Quast et al., 2012). The SILVA database was preferred to MaarjAM (Öpik et al., 2014), which is a dedicated Glomeromycota sequence database, as classification using MaarjAM appeared to over-classify edge cases as AMF taxa.

4.2.4 Statistical analysis of AMF Communities

All following analyses were carried out in R v4.2.2 (R Core Team, 2022). To first assess dissimilarities in AMF community composition (beta diversity) between species groups (i.e. B_{vol}, F_{vol}, F_{cal}, $F1_{vol}$), read counts were normalised to within-sample relative abundance in order to account for differences in sequencing depth, and ESVs were filtered to those for which there were at least five reads in at least three individuals. Bray–Curtis dissimilarity (Bray & Curtis, 1957) matrices were then calculated in the R package phyloseq v1.42.0 (McMurdie & Holmes, 2013) for each comparison: B_{vol} vs F_{vol}, B_{vol} vs F_{cal}, B_{vol} vs $F1_{vol}$, F_{vol} vs F_{cal}, F_{vol} vs $F1_{vol}$, and F_{cal} vs $F1_{vol}$. Multi-dimensional scaling analysis (MDS) was carried out based on the distance matrices of each comparison in order to visualise dissimilarities, and a permutation based analysis of variance test (PERMANOVA) was carried out for each

comparison, using the *adonis2* function in the R package Vegan v2.6.4 (Oksanen et al., 2013), to test for the significance of differences in each comparison. Each test was carried out with 999 permutations, and a threshold of $p<0.05$ was used to identify significant dissimilarity. To ensure that significant differences were not the result of differential dispersion between groups, permutation-based tests for homogeneity of dispersion (PERMDISP) were also performed, using the *betadisper* function in Vegan with ANOVA tests to check for significance of differences in dispersion.

Two indices were then used to investigate within-sample alpha diversity — the Shannon Index (Shannon, 1948), and the Gini-Simpson Index (Rao, 1982), which estimate alpha diversity by taking into account both sample richness and relative abundance. Firstly, in brief, the Shannon Index measures the uncertainty about the identity of unknown individuals within the system — in more diverse systems, predicting the identity of an unknown individual is more uncertain than in less diverse systems (Morris et al, 2014). The Gini-Simpson Index instead measures the probability that two randomly chosen individuals belong to the same species (Morris et al, 2014), and is regarded as a dominance index, giving more weight to more common taxa. These measures are more robust to sampling bias than indicators such as Chao1 (Chao, 1984; Gotelli & Colwell, 2001), which are based only on species richness (Roswell et al., 2021). Pairwise Mann-Whitney U tests with Holm-Bonferroni corrections for multiple testing (Holm, 1979) were used to test for significant differences in alpha diversity in each species group comparison (e.g. B_{vol} vs F_{vol}).

Finally, the DESeq2 package (Love et al., 2014) was used to identify ESVs differing significantly in relative abundance between each species group comparison; this method is known to be more powerful than other common approaches — such as rarefaction-based methods — for assessing differentially abundant taxa (McMurdie & Holmes, 2014). DESeq2 incorporates Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) to control the false discovery rate for multiple testing, and an adjusted p-value threshold of $p<0.01$ was used to identify differentially abundant ESVs.

4.2.5 Correlation of Gene Expression and AMF Alpha Diversity

For the 18 individuals for which both AMF alpha diversity and RNA-seq libraries from from at least one tissue (root, flower or leaf) were available, the strength of monotonic

associations between AMF alpha diversity and expression levels of each putative *Howea* gene were investigated. A monotonic relationship is one in which either: as one variable increases, so does the other; or, as one variable decreases, so does the other. In turn, Spearman's rank correlation coefficients were calculated for expression levels of each expressed gene in each tissue against AMF Shannon alpha diversity. P-values were computed for each test, with Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) to control the false discovery rate for multiple testing. R v4.2.2 (R Core Team, 2022) was used for all analysis.

4.2.6 GO Enrichment Analysis

As described in Chapter 3.2.7, gene ontology (GO) terms were downloaded from TAIR (Swarbreck et al., 2007) and used to functionally annotate putative *Howea* genes with the GO terms of their *A. thaliana* homologs. GO enrichment tests for genes possessing very strong Spearman's rank correlations (r_s >= +0.8, or <= -0.8) as described above, independently for each tissue, were carried out using the topGO v2.50.0 (Alexa et al., 2006) package in R v4.2.2. As discussed in Chapter 3.2.7, the *weight* algorithm with an unadjusted p-value cut-off of p<0.01 was used to identify significantly enriched GO terms.

4.3 Results

4.3.1 Sequencing Output and Taxon Assignment

Across the 20 root samples, the number of merged reads ranged from 26,248 to 182,498. In total, 170,513 reads were then removed as suspected chimeras (an average of 8,526 per sample), producing libraries ranging from 24,854 to 159,251 reads. These reads were assigned to 326 unique Glomeromycota ESVs, and when rarefaction curves of reads against ESV number were plotted for each sample, a clear plateau was reached for almost all samples (Figure 4.1), indicating that sequencing depth was great enough to survey the existing AMF diversity. After filtering to ESVs with at least five reads in at least three samples, and normalisation to relative abundance, 93.05% of reads belonged to ESVs which could be assigned to the family level (Figure 4.2), 1.35% were assigned to the order level (all 1.35% belonged to the order Glomerales), and the remaining 5.60% were unassigned beyond Glomeromycota. ESVs in the family Glomeraceae were by far the most abundant, accounting for 88.15% of all filtered and normalised reads.

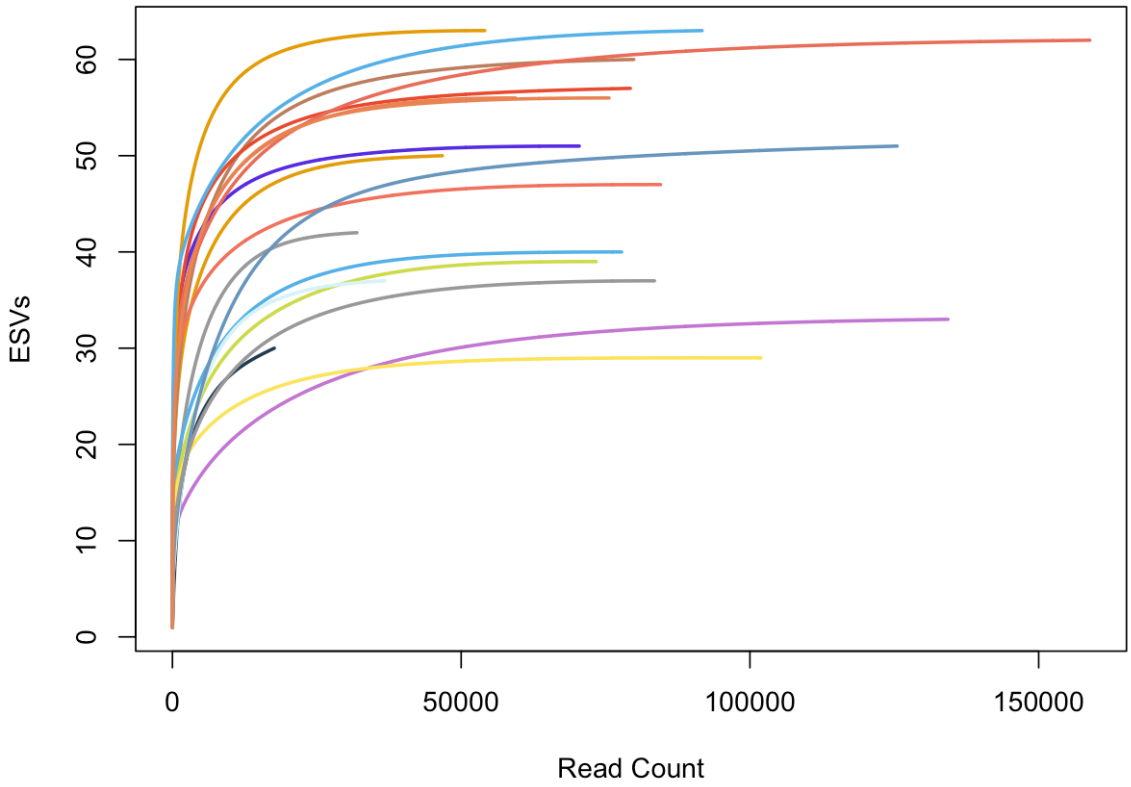


Figure 4.1. Rarefaction curves for each sample, showing unique ESV number plotted against read count. In almost all cases, the curve reaches a clear plateau, indicating that read count numbers were sufficient to capture the existing AMF diversity.

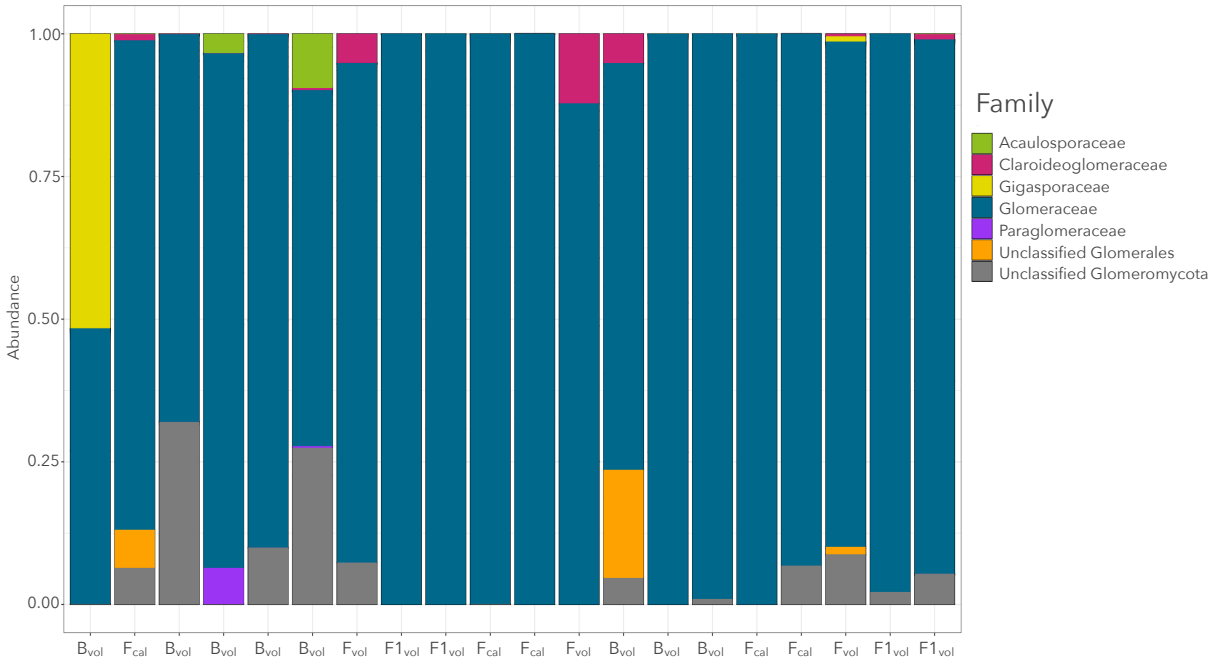


Figure 4.2. Community composition of each sample after filtering and normalisation of reads to relative abundance. ESVs assigned to the family level, with the exception of reads which could not be assigned beyond the Glomerales order level (orange), or which could not be assigned beyond Glomeromycota (grey).

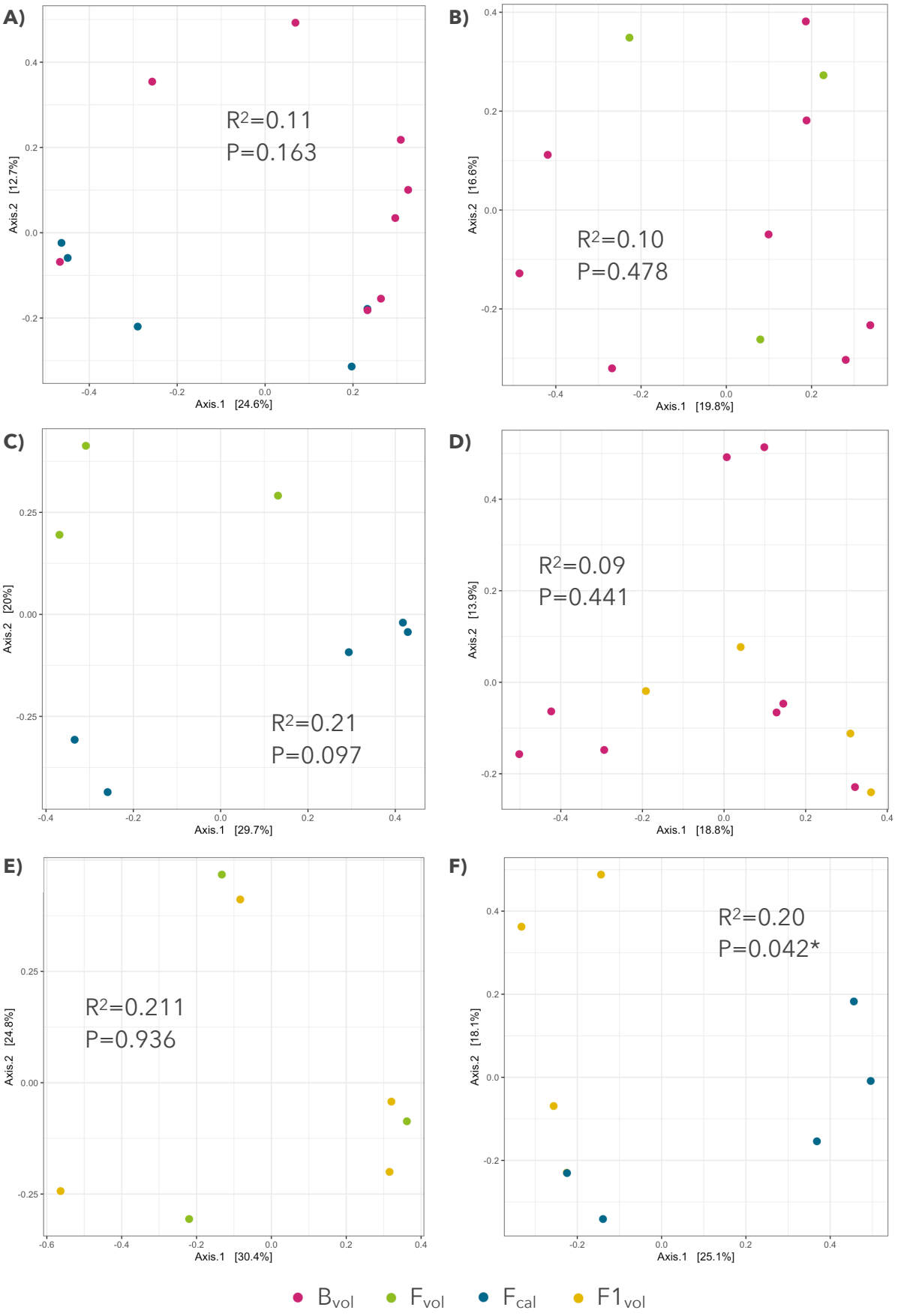


Figure 4.3. MDS plots of Bray-Curtis dissimilarity in **A)** B_{vol} vs F_{cal}; **B)** B_{vol} vs F_{vol}; **C)** F_{vol} vs F_{cal}; **D)** B_{vol} vs $F1_{vol}$; **E)** F_{vol} vs $F1_{vol}$; **F)** F_{cal} vs $F1_{vol}$. P-values and R^2 values for PERMANOVA tests are shown; only F_{cal} vs $F1_{vol}$ is significant.

4.3.2 Arbuscular Mycorrhizal Fungi Diversity

F_{cal} vs $F1_{vol}$ was the only pairwise comparison with significant dissimilarity in AMF community composition, as assessed by PERMANOVA tests of Bray-Curtis beta diversity (R^2=0.20, p=0.042). PERMDISP revealed no significant difference in dispersal between F_{cal} and $F1_{vol}$, meaning we can be confident in the results of the PERMANOVA test. In five of the six pairwise comparisons, there was no significant dissimilarity between species groups (Figure 4.3). However, MDS plots revealed clear separate clusters between F_{vol} and F_{cal}, although the dissimilarity was not significant (R^2=0.21, p=0.097). No significant differences were found for any pairwise comparison in either Shannon or Gini-Simpson alpha diversity (Figure 4.4).

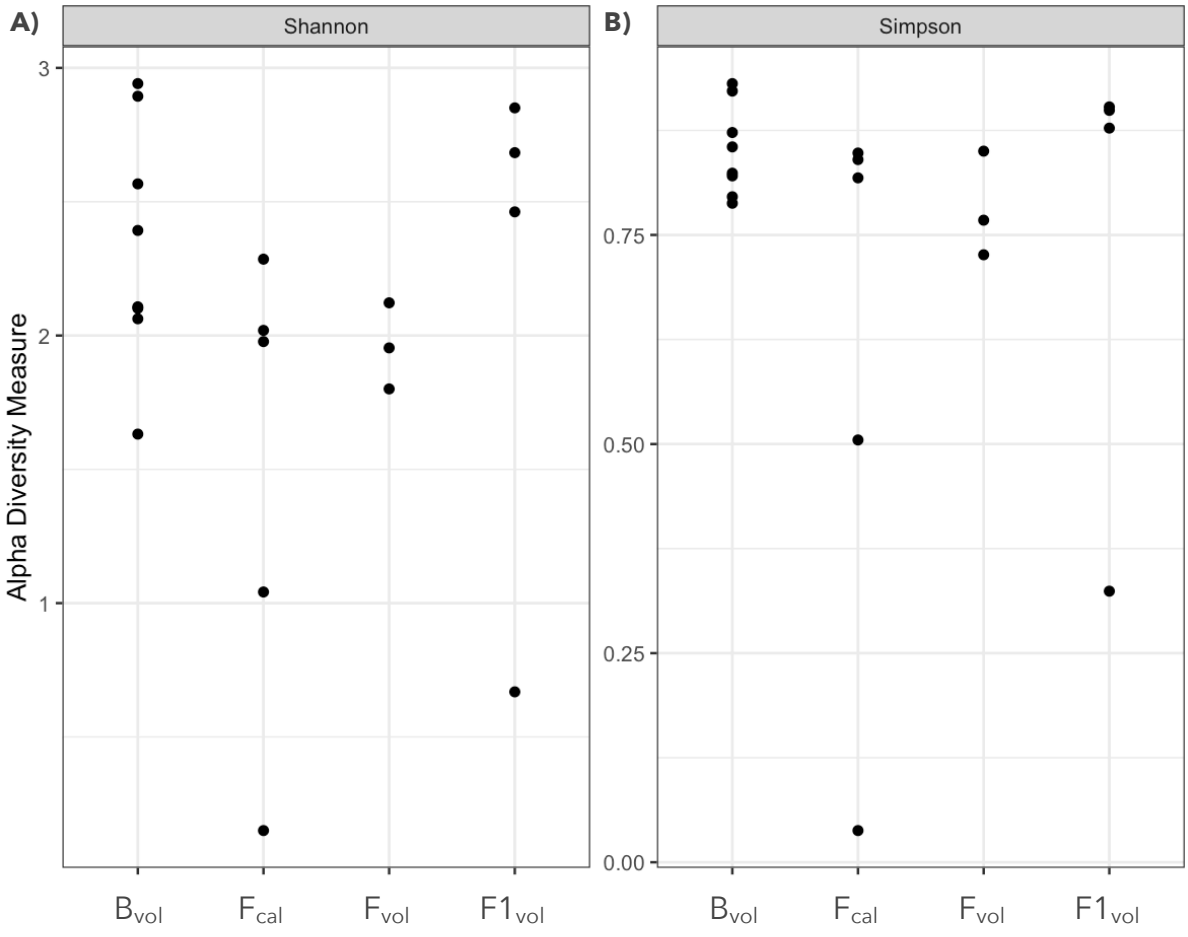


Figure 4.4. Alpha diversity plots, as measured by **A)** Shannon Index; and **B)** Gini-Simpson Index. Dots represent each individual tree. There was no significant difference in either of the alpha diversity measures for any pairwise comparisons of species groups (e.g. B_{vol} vs F_{vol}).

4.3.3 Differentially Abundant ESVs

A number of ESVs which differed significantly in relative abundance were found in each species group comparison, ranging from 18 in B_{vol} vs $F1_{vol}$ to 55 in F_{vol} vs F_{cal}. Interestingly, 55 is more than twice the number of differentially abundant ESVs as the comparison with the next largest number — 22 in F_{cal} vs B_{vol}. In five of the six comparisons, the proportion of differentially abundant ESVs with higher abundance in each of the two species groups within the comparison did not differ significantly from a 50:50 ratio (Chi-squared tests, Table 4.1) — i.e. for F_{vol} vs B_{vol}, the likelihood of differentially abundant ESVs having higher abundance in either of F_{vol} or B_{vol} was not significantly different to 50%. This was not the case for F_{cal} vs $F1_{vol}$, where four times as many differentially abundant ESVs had higher abundance in $F1_{vol}$ than F_{cal} — this did differ significantly from a 50:50 ratio (X^2=7.200, p=0.007).

Comparison	Highest Abundance	Number of ESVs	Chi-squared (X^2)
F_{vol} vs B_{vol}	F_{vol}	12	X^2=0.800, p=0.371
	B_{vol}	8	
F_{vol} vs F_{cal}	F_{vol}	30	X^2=0.455, p=0.500
	F_{cal}	25	
F_{vol} vs $F1_{vol}$	F_{vol}	8	X^2=1.191, p=0.275
	$F1_{vol}$	13	
F_{cal} vs $F1_{vol}$	F_{cal}	4	X^2=7.200, p=0.007*
	$F1_{vol}$	16	
F_{cal} vs B_{vol}	F_{cal}	8	X^2=1.636, p=0.201
	B_{vol}	14	
B_{vol} vs $F1_{vol}$	B_{vol}	5	X^2=3.556, p=0.059
	$F1_{vol}$	13	

Table 4.1. The numbers of ESVs differing significantly in relative abundance in each pairwise species group comparison, showing the numbers that are more abundant in each species group within the comparison. Test statistics and p-values for Chi-squared tests are also shown, with tests applied to investigate whether the proportion of ESVs with higher abundance in each species group within the comparison differed significantly from a 50:50 ratio.

4.3.4 Correlation of Gene Expression and AMF Alpha Diversity

There was a striking difference in the frequency of genes showing high levels of correlation between expression levels and AMF Shannon alpha diversity across different tissues (Figure 4.5). In root tissues, 191 genes possessed a Spearman correlation coefficient greater than plus or minus 0.9, and a further 1331 greater than plus or minus 0.8, indicative of very strong monotonic association. In contrast, in flower tissues, only 24 genes demonstrated a correlation coefficient greater than plus or minus 0.8, with a further 173 greater than plus or minus 0.7. This, in turn, is in contrast to leaf tissues, where only five genes showed a correlation coefficient greater than plus or minus 0.8, and only a further 48 greater than plus or minus 0.7.

Prior to correction for multiple testing, the correlation coefficients for 6,237 genes were significant to $p<0.05$ in root tissues. Whilst this number was reduced to 81 genes after Benjamini-Hochberg correction, 6,237 equates to 21.45% of all genes expressed in root tissue. With a p-value threshold of 0.05, under neutrality one would expect 5% of test results to give false positive Type I errors — far below the 21.45% seen here, suggesting a large number of genes with monotonic relationships between expression levels and AMF alpha diversity that are not the result of chance.

Conversely, in flower tissues, 1952 genes possessed correlations significant to $p<0.05$ before correction (zero following correction), equating to 6.58% of all genes expressed. In leaf tissues, only 648 genes demonstrated correlations significant to $p<0.05$ prior to correction (again, zero following correction), equivalent to only 2.28% of all genes expressed in leaf tissues.

In each tissue, the number of genes for which expression correlated positively with alpha diversity was similar to the number which correlated negatively. In roots, 14,163 correlated positively, 14,916 negatively; in flowers 15,007 positively, 14,518 negatively and 157 with zero correlation; and in leaves 14,431 positive versus 13,972 negative. However, in all three cases, the ratio of the number of positive correlations versus negative correlations did differ significantly from 50:50 (Chi-squared tests; in roots, $X^2=19.499$, $p<0.001$; in flowers, $X^2=8.099$, $p=0.004$; in leaves, $X^2=7.418$, $p=0.006$).

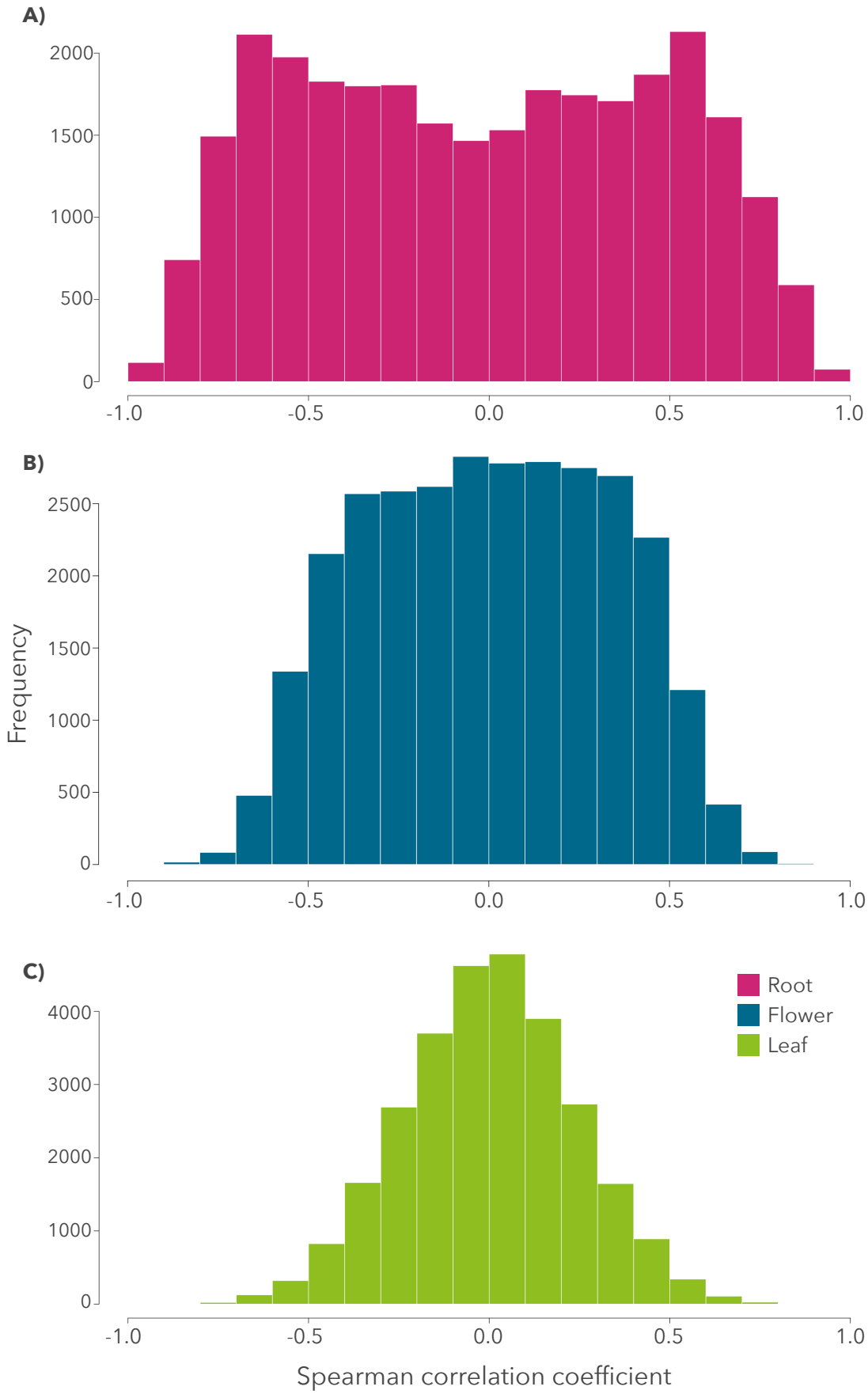


Figure 4.5. Histograms showing the frequency distribution of Spearman correlation coefficients for expression levels of each gene against AMF Shannon alpha diversity in **A)** root tissue; **B)** flower tissue; and **C)** leaf tissue.

4.3.5 GO Enrichment Analysis

Genes in root tissue whose expression levels demonstrated a very strong negative correlation — between -0.8 and -1 — against alpha diversity, were significantly enriched for 24 different GO terms with $p<0.01$, covering a variety of biological functions. Notably, ten of these terms were directly associated with cell division, movement and organisation, including terms such as 'mitotic spindle assembly', 'meiosis I', 'microtubule-based movement' and 'cell division'.

In the subset of genes expressed in root tissue for which a positive correlation of between 0.8 and 1 was observed, 18 GO terms were significantly enriched with $p<0.01$. Again, these terms were associated with a variety of biological functions, including, for example, multiple terms involved in metabolic processes. Perhaps of greatest interest, however, is enrichment of the terms 'response to rapamycin', 'response to sucrose' and 'phloem loading'. Rapamycin is an anti-fungal antibiotic, originally discovered being produced by soil bacteria on Easter Island, and is an inhibitor of the Target of Rapamycin (TOR) growth factor and related pathways (Li et al., 2014). In the common bean (*Phaseolus vulgaris*), knockdown of *TOR* transcripts in hairy roots was shown to significantly suppress AMF-induced lateral root formation, AMF colonisation, and AMF arbuscule development and maturation (Arthikala et al., 2021). Additionally, Arthikala et al. found that multiple sucrose synthases and sucrose-cleaving enzyme invertases had altered expression in *TOR* knockdowns. Such genes are involved in the increase in root sink strength associated with AMF colonisation in plants, which allows greater unloading of sucrose from the phloem in order to be transported to AMF symbionts (Schaarschmidt et al., 2006; Boldt et al., 2011; Arthikala et al., 2021). It is, therefore, notable that the terms 'response to sucrose' and 'phloem loading' are enriched alongside 'response to rapamycin'.

Not only this, but three putative *Howea* genes annotated with the *A. thaliana AtTOR* gene demonstrated strong positive correlation with AMF alpha diversity (Cluster-6827.0, $r_s=0.867$; Cluster-9248.1, $r_s=0.818$; Cluster-9248.2, $r_s=0.721$), along with two putative homologs of the *A. thaliana* gene *RAPTOR1* (a binding partner of the TOR kinase protein; Cluster-11002.0, $r_s=0.782$; Cluster-11002.1, $r_s=0.758$). No strong correlations (positive or negative) between any of these five putative *Howea* genes and alpha diversity were found in flower or leaf tissue. However, Cluster-9248.1 was found to be amongst the differentially expressed genes reported

in Chapter Three, for multiple pairwise species groups comparisons in all three tissues. The gene was always up-regulated in B_{vol} relative to each of F_{vol} and F_{cal} in all three tissues. In root tissues, it was also up-regulated in B_{vol} relative to $F1_{vol}$, and in $F1_{vol}$ relative to F_{vol}. In flower tissues, it up-regulated in $F1_{vol}$ relative to both F_{vol} and F_{cal}.

4.4 Discussion

4.4.1 No Evidence for Depleted AMF in *H. forsteriana* on Volcanic Soils

In contrast to the results of Osborne et al. (2018), the present study found no evidence for depleted AMF in the roots of *H. forsteriana* growing on volcanic soils. Previously, Osborne et al. (2018) — using ITS2 primers — had found that AMF species richness, measured by the number of OTUs scaled by library size, was significantly lower in *H. forsteriana* on volcanic soils than both *H. belmoreana* and *H. forsteriana* growing on on calcareous soils. In the two measures of alpha diversity used here (the Shannon and Gini-Simpson indices), no significant differences were found in any pairwise comparison between all species groups. To offer a more direct comparison to the results of Osborne et al., the number of unique AMF ESVs for each group was normalised by library size and compared with Mann-Whitney U tests — again, there were no significant differences in any pairwise comparison.

Additionally, Osborne et al. (2018) had found that AMF OTUs which differed significantly in relative abundance between F_{vol} and F_{cal}, and between F_{vol} and B_{vol}, were always less abundant in F_{vol}. Again, no evidence of this was found here — in every pairwise comparison involving F_{vol} (F_{vol} vs F_{cal}, F_{vol} vs $F1_{vol}$, F_{vol} vs B_{vol}), the number of ESVs with higher abundance in each of the two groups in the comparison did not differ significantly from 50% (Chi-squared tests).

There was, however, tentative evidence of significant dissimilarity between F_{vol} and F_{cal} AMF communities. Osborne et al. had shown that beta diversity of fungal communities, measured by PERMANOVA following calculation of Bray-Curtis distance matrices, showed significant dissimilarity between F_{vol} and F_{cal}. Whilst this result applied to the entire fungal communities in the roots of the two groups, the same method of analysis here, restricted to

AMF, resolved F_{vol} and F_{cal} as distinct clusters in MDS plots (Figure 4.3). A PERMANOVA test of this dissimilarity was not significant (p=0.097), however, small sample sizes likely restrict the power of this analysis (discussed further below). Additionally, the number of differentially abundant ESVs in F_{vol} versus F_{cal} (55) was more than twice the number of any other comparison. Taken together, these results highlight that F_{vol} and F_{cal} may differ significantly in the community composition of their AMF biomes; however, they do not suggest the AMF are depleted in the roots of *H. forsteriana* when growing on volcanic soils.

The results of Osborne et al. were convincing, especially when combined with their seedling survival experiments, which showed that soil sterilisation uniquely has no affect on germination and survival of *H. forsteriana* on volcanic soils. Additionally, differential expression analysis in Chapter Three of this thesis showed that genes involved in plant response to symbiotic fungi were down-regulated in F_{vol} relative to F_{cal}, including a *Howea* ortholog of the *AMT2* gene, known to be an important ammonium transporter during AMF symbiosis (Guether et al., 2009) — though significant down-regulation was only found in leaf tissues, not roots. Regardless, the contrast in results seen in Chapter Four is, therefore, notable. There are factors which may contribute to these differences — as mentioned above, sample sizes in the present study are small, with only three F_{vol} individuals, four $F1_{vol}$, and five F_{cal}. The work by Osborne et al. only used four independent samples for F_{cal} and five for F_{vol}, however each of these samples contained root DNA pooled from five trees collected at each sample site. Therefore, the community composition expected in each 'sample' may differ significantly between the two studies — one would expect greater diversity in each sample using the pooled method of Osborne et al., which may give more statistical power, especially in analysis of differentially abundant ESVs. Such an approach would not be possible in this study, due to the inclusion of rare hybrid trees.

Secondly, the primer pair used by Osborne et al. is less optimised for metabarcoding of AMF taxa than the primers used in this study. In one analysis of primer pair suitability for AMF, only 0.04% of amplified sequences from soil samples were AMF taxa when using the universal fungal primer pair ITS1-FKYO1/ITS2-KYO1 (Suzuki et al., 2020). In contrast, the authors found that 98% of amplified sequences were AMF specific when using the same primer pair and nested PCR method as the present study. However, there is no inherent reason to believe that the differences in detection rate of different primer pairs would be

significantly biased between the *Howea* species groups, especially as the vast majority of AMF sequences detected here for each group belong to the same family, Glomeraceae. Therefore, the differences in results compared to Osborne et al. clearly warrant further study, which is discussed in more detail below.

4.4.2 AMF Communities in Hybrid Palms Differ from *H. forsteriana* Growing on Calcareous Soils

There are several lines of evidence to suggest that AMF communities in hybrid palms differ significantly from those of *H. forsteriana* growing on calcareous soils. Firstly, PERMANOVA tests of Bray-Curtis beta diversity found that the two communities differ significantly in composition (p=0.042). Whilst there was no difference between the two groups in alpha diversity, F_{cal} vs $F1_{vol}$ was the only pairwise comparison where the ratio of differentially abundant ESVs that had greater abundance in either group differed significantly from 50:50, with four times as many differentially abundant ESVs having higher abundance in $F1_{vol}$ than in F_{cal}. Additionally, differential expression analysis in Chapter Three found that genes up-regulared in $F1_{vol}$ relative to F_{cal} were enriched for the GO term "response to karrikin". Karrikin response genes are known to be involved in the formation of arbuscular mycorrhizal fungi (AMF) symbioses in rice (Gutjahr et al., 2015).

The biological significance of these differences is difficult to determine without follow-up work (discussed further below) to tease apart the relative benefits that *H. forsteriana* and hybrid palms each derive from AMF symbiosis. However, of the 16 ESVs with greater abundance in $F1_{vol}$ relative to F_{cal}, 10 were among the 14 ESVs with greater relative abundance in B_{vol} than F_{cal}. This suggests that many of the ESVs with increased abundance in hybrids relative to F_{cal} may have little biological significance beyond the differences existing between B_{vol} and F_{cal}.

Intriguingly, though, all four of the ESVs with greater abundance in F_{cal} than $F1_{vol}$ were among the five ESVs that had greater abundance in B_{vol} than $F1_{vol}$, and among the eight ESVs with greater abundance in F_{vol} than $F1_{vol}$. Whilst there is no overall evidence to suggest that AMF are depleted in hybrid palms, the fact that these four ESVs have reduced abundance in hybrids relative to all other groups is interesting. If reduced hybrid fitness is

associated with reduced benefit from AMF symbiosis, these four ESVs may represent the most likely causative taxa. Follow-up work to determine whether hybrid fitness is affected by AMF colonisation is required to investigate this further, and is discussed below.

4.4.3 AMF Symbiosis Strongly Affects Gene Expression in Root Tissues

Correlation analysis of gene expression with AMF Shannon alpha diversity strikingly demonstrated that AMF have a far greater effect on gene expression in root tissues than in flower or leaf tissues. Remarkably, 1,552 genes expressed in root tissues had a Spearman correlation coefficient greater than plus or minus 0.8 between expression levels and AMF alpha diversity — 5.2% of all expressed genes. In flower tissues, only 24 genes had such a strong correlation coefficient, and in leaf tissue, only five genes. This presents clear-cut evidence that AMF colonisation has a widespread impact on *Howea* gene expression in root tissues, and strengthens the hypothesis that AMF symbiosis may be key to *Howea* fitness.

In root tissues, the subset of these genes with a positive correlation was enriched for a range of functions, including several GO terms associated with metabolism. AMF colonisation is known to affect plant metabolism; for example AMF have been shown to positively induce biosynthesis of metabolites such as isoprenoids (Walter et al., 2000; Strack & Fester, 2006), a large class of organic plant molecules. Indeed, in the present study, genes expressed in roots with a positive correlation coefficient >0.8 were enriched for the term "positive regulation of isoprenoid metabolism". Most notably, however, several *Howea* homologs of the *TOR* and *RAPTOR1* genes were found to have a very strong positive correlation between root expression and AMF diversity. In the common bean (*Phaseolus vulgaris*), the *TOR* gene has been shown to be key in AMF colonisation and AMF arbuscule development and maturation (Arthikala et al., 2021), and the results here highlight a possible role for the same pathway in the formation of AMF symbiosis in *Howea.* This is especially true given that genes with a strong positive correlation in roots were also enriched for the GO terms "response to sucrose" and "phloem loading". Expression of sucrose synthases and sucrose-cleaving enzymes, involved in the transport of sucrose from the phloem to AMF symbionts, was shown to be dependent on *TOR* expression in *P. vulgaris* (Schaarschmidt et al., 2006; Boldt et al., 2011; Arthikala et al., 2021).

Together, these results strongly indicate a role for *TOR* in AMF symbiosis in *Howea.* Additionally, one *Howea* ortholog of *TOR* was found to be significantly up-regulated in *H. belmoreana* relative to *H. forsteriana,* in all tissues, regardless of the soil substrate on which *H. forsteriana* was growing. Given the previously discussed findings of Osborne et al. (2018), showing that *H. forsteriana* appears to have depleted AMF on volcanic soils, it is plausible that up-regulation of this gene in *H. belmoreana* is pivotal to its ability to form relationships with the AMF biome of volcanic soils. Of course, the results presented earlier in this chapter are not concordant with the finding that AMF are depleted in *H. forsteriana* on volcanic soils, but this is subject to further investigation, discussed below.

It is also worth highlighting that strong negative correlations between gene expression and AMF diversity were actually marginally more common than strong positive correlations in root tissues, and in general, negative correlations were just as numerous as positive correlations. Genes in root tissues with a negative correlation coefficient with AMF alpha diversity of 0.8 or greater were particularly enriched for GO terms associated with cell division and movement. Whether or not these negative correlations represent adaptive, maladaptive, or neutral phenotypes is difficult to speculate upon, and likely varies across different genes. This is an area that may warrant further exploration, as many studies have investigated genes up-regulated in response to AMF colonisation, but fewer have investigated the converse.

Indeed, to my knowledge, only one study has previously analysed the relationship between global gene expression and AMF community composition (Gao et al., 2022), and the work presented here is the first to examine the differential effect that AMF colonisation has on global gene expression in different plant organs. Due to the large number of studies investigating AMF symbioses, however, this is difficult to verify. Regardless, the results here highlight that such analysis may be of value to a wide range of plant species.

4.4.4 Further Work Investigating AMF Diversity in *Howea* is Needed

Whilst the results presented in this chapter are interesting with regards to the differences, or lack thereof, in AMF communities seen between different *Howea* species groups, they raise several questions which require further work. This should take the form of three main avenues of investigation.

Firstly, perhaps with the exception of correlation analysis between gene expression and AMF diversity, the statistical power and reliability of the work presented is limited by the sample sizes used. The results here, suggesting that *H. forsteriana* does not have depleted AMF levels, are in stark contrast to previous research which convincingly demonstrated that *H. forsteriana* does not benefit from the microbiome of volcanic soils, almost certainly due to depleted AMF. Before doubting the findings of this previous work, additional sampling is prudent, involving trees from a wider variety of sites on LHI, along with soil sampling, to ensure differences seen are not simply the result of differences present in soil across the island. Until recently, further sampling was not possible, as Covid-19 restrictions prevented travel to Lord Howe Island. However, restrictions were lifted in summer, and during fieldwork on LHI in late August, new root and soil samples were collected from 17 trees, made up of eight putative hybrids, three *H. belmoreana* individuals, three *H. forsteriana* individuals from volcanic soils, and three from calcareous soils. It has not yet been possible to carry out library preparation, sequencing, and verification of hybrid identity; however, these additional samples will allow more rigorous analysis, which is needed to robustly challenge the results of Osborne et al. (2018). Should this further analysis, using AMF specific primers rather than ITS primers, confirm the work in this thesis, it would cast doubt on whether AMF are causative of the reduced ability of *H. forsteriana* on volcanic soils to benefit from the soil microbiome. Alternative explanations, such as differential tolerance to pathogens found in the different soil types (Osborne et al., 2018), would then require further exploration.

Secondly, the results presented here give evidence to suggest that hybrids differ in AMF composition from *H. forsteriana* on calcareous soils. There is also evidence of specific AMF taxa which are always less abundant in hybrid *Howea* palms. However, the biological significance of these findings is difficult to determine, and should be investigated further through soil sterilisation assays involving hybrid seedlings on both soil types. Such assays would shed light on whether biotic stimuli contribute to reduced hybrid fitness — for example, if soil sterilisation does not reduce hybrid seedling survival, we could conclude that hybrid palms do not benefit from the soil biome, offering explanation of the reduced hybrid fitness seen on LHI. One mechanism through which this could happen is hybrid palms not benefitting from symbios with AMF, which are known to improve drought tolerance (Wu et al., 2013; Pavithra & Yapa, 2018; Begum et al., 2019). In turn, this may increase hybrid susceptibility to periods of extreme drought, which are known to occur on LHI. Again,

carrying out these survival assays was not possible due to Covid-19 restrictions, with longer-term access to LHI required to gather hybrid seeds and conduct survival experiments over a period of months, or even years.

Finally, building on additional survival assays, a more challenging — yet more informative — approach, would be the manipulation of specific microbial and fungal taxa added to sterilised volcanic and calcareous soils. Through survival assays on these manipulated soil substrates, confirmation of the precise biotic factors behind the reduced benefit *H. forsteriana* derives from volcanic soils could be unearthed, alongside a more detailed investigation of whether biotic stimuli contribute to reduced hybrid fitness on LHI.

4.4.5 Conclusion and Wider Context

The results presented in this chapter are highly interesting on a variety of fronts. Firstly, they cast doubt upon the previous finding that AMF are significantly depleted in *H. forsteriana* growing on volcanic soils. Confirmatory work is vital in order to verify this, however, if verified, an alternative explanation would be required as to how *H. belmoreana* gains a competitive advantage over *H. forsteriana* on volcanic soils — for example, through increased resistance to pathogens found on volcanic soils. Such work is key to understanding the niche divergence and co-existence of the two *Howea* species, as without a competitive advantage on volcanic soils, *H. belmoreana* may have been driven to extinction.

Secondly, there is evidence to suggest that hybrids are depleted in a small number of AMF taxa. Whilst the biological significance of this is difficult to speculate upon without further investigation, hybrids are known to have reduced fitness, with selection acting against them between the juvenile and adult stage (Hipperson et al., 2016). This hybrid inviability may have been important to the speciation process, with post-zygotic isolation and reinforcement driving speciation to completion. As such, understanding the causes of reduced hybrid fitness is central to our understanding of sympatric speciation in *Howea*, and, therefore, the present results highlight that further investigation into whether hybrid palms benefit from AMF symbiosis is warranted.

Finally, correlation analysis between gene expression and AMF alpha diversity strikingly demonstrated that AMF colonisation has a widespread affect on root gene expression. The

extent of this effect supports the hypothesis that AMF are likely key to *Howea* fitness (Osborne et al., 2018), and reinforces the need to understand the role AMF have played in facilitating speciation and promoting coexistence in *Howea.* Such analysis also uncovered a role for *TOR* gene expression in the formation of AMF symbioses in *Howea*, with up-regulation of a *TOR* ortholog in *H. belmoreana* offering a potential explanation as to how *H. belmoreana* forms beneficial AMF relationships on volcanic soil, where *H. forsteriana* may not be able to do so (Osborne et al., 2018). This is significant, as it provides detail over a specific genetic mechanism which could have facilitated niche divergence and sympatric speciation in *Howea*; something which is largely absent in the sympatric speciation literature.

Chapter Five - General Discussion

5.1 Key Findings and Significance

In the introduction to this thesis, I highlighted that I would endeavour to further existing knowledge in two key ways: by addressing the knowledge gap in *Howea* over the specific mechanisms which could have facilitated speciation; and, as a consequence, improve understanding in the wider field of speciation biology about the forces governing speciation, which has been a central area of focus in the genomic era (Wu, 2001; Butlin, 2008; Jones et al., 2012; Cruickshank & Hahn, 2014; Riesch et al., 2017). This is especially true given *Howea's* status as perhaps the most convincing known example of sympatric speciation, with the mechanisms governing this form of speciation particularly debated and scant of evidence.

In Chapter Two, extensive phenotyping experiments using the model plant species *A. thaliana* robustly demonstrated a pleiotropic role in both flowering time and soil stress tolerance for five knockout orthologs of candidate *Howea* speciation genes: *at2-mmp*, *dcl1*, *dcl4*, *sal1*, and *siz1*, under conditions consistent with the LHI speciation scenario. Two of these five loci each mapped to two *Howea* candidate speciation genes, producing a list of seven *Howea* genes which show evidence of adaptive divergence between *H. belmoreana* and *H. forsteriana*, and could have driven initial sympatric divergence by acting as magic traits (Servedio et al., 2011).

In particular, two knockouts (*sal1* and *siz1*) demonstrated major effect sizes with regards to both stress tolerance and flowering time. Two orthologs of *SIZ1* were among the list of candidate genes, giving three *Howea* loci that represent the most likely drivers of sympatric speciation. Amongst these three, one gene, a *Howea* ortholog of *SIZ1*, was found to be almost entirely unexpressed in *H. belmoreana*. *H. belmoreana* cannot survive on calcareous soils, which are associated with abiotic stresses such as drought, and the knockout of *SIZ1* in *A. thaliana* showed significantly reduced drought tolerance, consistent with the lack of expression seen in *H. belmoreana*. There is, therefore, strong evidence that this locus could have driven initial divergence on LHI.

Organism	Magic Trait	Reference
Gasterosteus sticklebacks	Body size affecting foraging and competition in limnetic and benthic forms	Nagel & Schluter, 1998
Heliconius butterflies	Colour pattern affecting mimicry and predation among mimetic forms	Jiggins et al., 2001
Amphilophus Midas cichlids	Dark/gold coloration and hypertrophied lips associated with, for example, diet and trophic specialisation	Kautt et al., 2020
Geospiza Darwin's finches	Beak morphology affecting foraging/diet	Podos et al., 2004
Mimulus monkeyflowers	Flower colour with divergent pollinator types	Bradshaw & Schemske, 2003
Dendrobates pumilio poison dart frogs	Colour pattern affecting mimicry and predation among mimetic forms	Reynolds & Fitzpatrick, 2007
Littorina saxtalis snail	Body size affecting crab predation versus wave resistance in upper and lower shore ecotypes	Rolan-Alvarez, 2007
Ophrys orchids	Flower scent compounds with divergent pollinator types	Gervasi et al., 2017
Neodiprion sawfly	Body size associated with differing pine host needle thickness	Glover et al., 2023

Table 5.1. Selected putative examples of 'magic traits' involved in speciation, through a hypothesised pleiotropic role in both assortative mating and adaptive divergence.

Whilst the number of putative examples of magic traits involved in speciation is ever-increasing (Table 5.1), very few have been able to identify a specific pleiotropic locus underlying the trait, especially in a convincing case of sympatric speciation. In Midas cichlids which have diverged sympatrically in crater lakes (though with likely secondary contact and introgression), recent work applying genome-wide association mapping and quantitative trait locus analysis, with 453 cichlid genomes, highlighted monogenic loci controlling the dark/gold colouration morphs and the hypertrophied lip trait (Kautt et al., 2020). These features likely act as magic traits, contributing to trophic specialisation and assortative mating. However, the authors found that population divergence between these ecotypes was very low and likely unstable, and concluded that these traits probably did not drive sympatric speciation.

Evidence of specific pleiotropic loci behind major effect magic traits in other fully convincing cases of sympatric speciation is absent; and, therefore, results here provide this empirical evidence for the first time. This is notable, as whether or not magic traits can drive

speciation, and particularly sympatric speciation, has repeatedly been highlighted as a significant and unanswered question. For example, both Foote (2018) and Richards et al. (2019), reviewing sympatric speciation, singled out the need to address the conspicuous lack of empirical evidence for the role of magic traits. With follow-up work currently ongoing to verify conserved function between *A. thaliana* and *Howea* (detailed below) — the final stage required to conclusively demonstrate the role of magic traits in *Howea* — the findings of this thesis will likely be of notable significance to the field of speciation biology.

Following this investigation of magic traits, the focus of the thesis moved to hybrid palms, which exist in low frequencies on LHI (Babik et al., 2009; Hipperson et al., 2016), and appear to suffer from reduced fitness, with selection acting against them before they reach adulthood (Hipperson et al., 2016). Post-zygotic isolation due to reduced hybrid fitness may have played a significant role in the *Howea* speciation process, by maintaining species boundaries and potentially driving speciation to rapid completion by reinforcement. Indeed, in other studies of ecological speciation with gene flow in plants, strong selection against hybrids may be critical to maintaining species boundaries, for example in the *Senecio* ragworts of Mount Etna (Wong et al., 2020), and in hybrid zones of European *Populus* poplars (Christe et al., 2016).

Through examination of gene misexpression in hybrid palms, consistent evidence highlighted how reduced hybrid fitness is likely a result of aberrant developmental processes and enhanced susceptibility to the wide range of biotic and abiotic stresses which exist on LHI (Osborne et al., 2018; Osborne et al., 2019). Interestingly, in contrast to many studies investigating misexpression in hybrids (Ranz et al. 2004; McManus et al. 2010; Coloon et al. 2014; Dion-Côté et al., 2014; Mack et al., 2016; McGirr & Martin, 2020), results here identified a relatively small number of loci demonstrating transgressive expression levels. However, repeated lines of evidence strikingly showed how differentially expressed and misexpressed genes in hybrids were associated with development, gene expression, and stress response.

Results also showed that many of the misexpressed loci were significantly up-regulated in hybrids, indicative of compensatory expression, likely as a mechanism to make up for aberrant development and reduced stress tolerance. Indeed, these results showed remarkable

consistency with a previous study investigating hybrid inviability between two sympatric copepod populations (Barreto et al., 2015), where hybrids suffered from stunted development and exhibited large degrees of compensatory up-regulation. A similar phenomenon is also seen in *Senecio squalidus* — a ragwort species found in the UK, which originated from hybridisation between the aforementioned Mount Etna ragworts. Here, two transgressively expressed genes appear to be up-regulated in response to the sulfur deficiency of UK soils relative to the soils of Mount Etna (Hegarty et al., 2009).

Given the apparent selection against hybrids on LHI (Hipperson et al., 2016), the evidence presented in this thesis supports a scenario in which genomic incompatibility in hybrids leads to maladapted phenotypes, particularly through reduced tolerance to the range of biotic and abiotic stresses found on LHI. Post-zygotic isolation and reinforcement caused by hybrid inviability could have rapidly driven *Howea* speciation to completion, and the misexpression identified here provides a plausible causative mechanism for this. No previous studies have examined such mixespression in an example of completed sympatric speciation (e.g., reproductively isolated and distinct species, rather than apparently sympatrically diverging populations), and, therefore, the results presented here uncover for the first time how genomic incompatibility can lead to hybrid misexpression during sympatric speciation. Functional characterisation of misexpressed genes also provides the first mechanistic evidence of how selection against hybrids, driven by transgressive expression, stunted development, and increased susceptibility to stresses, can promote sympatric speciation.

Finally, Chapter Four examined the role of arbuscular mycorrhizal fungi in *Howea* speciation, with earlier research having shown that differences in AMF symbioses are likely key to niche differentiation and coexistence of the two species (Osborne et al., 2018). This previous work convincingly demonstrated that *H. forsteriana* does not benefit from the microbiome of volcanic soils, probably due to significantly depleted levels of AMF (Osborne et al., 2018). Such a scenario may be key to the continued coexistence of the two *Howea* species, with *H. belmoreana* retaining a competitive advantage on volcanic soils, and thereby preventing a situation in which it is out-competed and driven to extinction by *H. forsteriana*. However, the work detailed in this thesis found no evidence for reduced AMF diversity in the roots of *H. forsteriana* growing on volcanic soils, relative to either *H. belmoreana* or *H. forsteriana* growing on calcareous soils. This is significant, as if AMF are not depleted in the roots of *H.*

forsteriama, alternative explanation would be needed as to why *H. forsteriana* does not benefit from the microbiome of volcanic soils, and as to precisely how *H. belmoreana* realises a competitive advantage on volcanic soils.

Additionally, I hypothesised that hybrid palms may have a reduced ability to form mutualistic relationships with AMF, in turn leading to the reduced hybrid fitness seen on LHI (Hipperson et al., 2016). There was no evidence for overall depleted levels of AMF in hybrid palms. However, the same four AMF taxa were found to be significantly less abundant in hybrid roots than each of *H. belmoreana*, *H. forsteriana* growing on volcanic soils, and *H. forsteriana* growing on calcareous soil. Further investigation is needed to elucidate whether this is of biological significance, and whether hybrid fitness is affected by the soil microbiome.

It must also be acknowledged that a particular limitation of the present work is the small sample sizes used for metabarcoding analysis, and the fact that further seedling survival experiments to build upon those carried out by Osborne et al. were not achievable. Unfortunately, this is primarily due to Covid-19 restrictions which prevented travel to LHI until summer 2022. Regardless, these limitations do significantly reduce the ability to draw meaningful conclusions around the role of AMF in the speciation process — especially when challenging previous findings. Instead, results here open further questions. Such questions are themselves interesting, however, and avenues of further work relating to this are discussed below.

Finally, a less notable, but still interesting, contribution of the work presented here is the novel investigation of how AMF colonisation affects gene expression across different plant organs. Correlation analysis remarkably highlighted the wide-ranging effect of AMF colonisation on expression levels in *Howea* root tissues; with over 5% of genes expressed in roots demonstrating correlation coefficients greater than 0.8. In contrast, correlation coefficients >0.8 were only found in 0.08% of genes expressed in flower tissues, and 0.02% in leaf tissues. To my knowledge, this is the first study to examine the relationship between AMF colonisation and global gene expression in different plant organs, and highlights that AMF symbioses have a significant effect on expression in *Howea,* supporting the hypothesis that these mutualistic relationships are likely key to *Howea* fitness (Osborne et al., 2018).

Alongside this, correlation analysis uncovered strong positive relationships between AMF diversity and expression levels of multiple *Howea* orthologs of both the *TOR* gene and the *RAPTOR1* gene. The RAPTOR protein is a partner protein involved in the TOR pathway, which is known to be key to the formation and development of AMF symbiosis (Arthikala et al., 2021). Additionally, one *Howea* ortholog of *TOR* is significantly up-regulated in all tissues of *H. belmoreana* relative to *H. forsteriana.* It is plausible that this gene is vital for the initiation of AMF symbiosis on volcanic soils, offering a potential explanation as to why *H. forsteriana* has significantly depleted AMF on volcanic soils, as reported by Osborne et al. (2018). Of course, it should be acknowledged that the results in this thesis challenge the findings of Osborne et al. Together, the results from correlation analysis highlight how this method can be used to identify key genes involved in AMF symbiosis, and reveal the impact of AMF colonisation on host plant gene expression, which may be a powerful approach useful to other study systems.

5.2 Limitations and the Scope for Future Work

Whilst the results discussed above are of great interest, there are limitations to the work presented here, which should be recognised and, where appropriate, supplemented with further research. Firstly, conservation of function between *Howea* genes and their *A. thaliana* orthologs, which were investigated in knockout experiments, is not guaranteed. Given that multiple 'speciation genes' were found — which have evidence of adaptive divergence between the *Howea* species and acted as magic traits in *Arabidopsis* — the evidence is nonetheless convincing that such loci could have driven initial sympatric divergence. However, without confirmation of conserved function, there is scope to challenge this conclusion.

Indeed, work to verify conservation of function is currently ongoing, and was discussed in detail in Chapter 2.4.4. Briefly, however, it will involve transformation of *Arabidopsis* knockouts with synthetic constructs of both *H. belmoreana* and *H. forsteriana* alleles of magic trait genes. Phenotyping experiments with over-expression of these alleles, driven by a constitutive 35S cauliflower mosaic virus promoter, will allow us to confirm the functions of the *Howea* loci. Together with the strong evidence already presented in this thesis, if

conservation of function is demonstrated, it would provide conclusive evidence of the role of magic traits in *Howea.*

Secondly, the results from Chapter Four are mixed with regards to the aims outlined. There was no strong evidence of a role for AMF in reduced hybrid fitness, and the results contradicted previous findings suggesting that niche divergence in *Howea* may have been facilitated by a reduced ability to form AMF symbioses in *H. forsteriana* on volcanic soils (Osborne et al., 2018). As aforementioned, to delve further into this, the first area of focus should be an improved sampling effort, to carry out metabarcoding using AMF specific primers with greater statistical power in the analysis. Samples sizes used in this thesis were small, and so results should be interpreted with caution.

Again, this follow-up work is ongoing, with further sampling carried out on LHI in August 2022. Additionally, Savolainen plans to expand this work to a larger scale, which is only more prudent on the back of the results shown here. Larger scale experiments will aim to combine AMF metabarcoding with seedling survival assays, whilst manipulating the microbiome of soils on which seedlings are grown. Through such methods, clarity may be given on the causative taxa behind the benefit *H. belmoreana* derives on volcanic soils relative to *H. forsteriana.* Such experiments should incorporate hybrid palms, in order to unpick whether reduced hybrid fitness is influenced by biotic stimuli. Answering these questions is key to our understanding of adaptive divergence by soil type and the continued stable coexistence of the two *Howea* species.

Seedling survival experiments involving hybrids could also be extended to assays of abiotic and biotic stress tolerance, as although the evidence of misexpression in hybrids being associated with development and stress tolerance was highly consistent, it cannot be guaranteed that such misexpression is directly causative of the reduced hybrid fitness seen on LHI. Because hybrids exist in such low frequencies, and because generation times are so long, empirically testing the causes of selection against hybrids between the juvenile and adult stages is challenging. However, stress tolerance in hybrid seedlings may nonetheless provide insight.

5.3 Concluding Remarks

Throughout this thesis, the work presented has aimed to identify, at a detailed level, the mechanisms which could have facilitated sympatric speciation in *Howea*. This has taken three broad approaches:

1) An attempt to rigorously demonstrate whether genes showing adaptive divergence between the *Howea* species could have driven sympatric speciation by acting as magic traits
2) Investigation of hybrid gene expression, to identify causes of reduced hybrid fitness and post-zygotic isolation, which could have driven speciation to completion by reinforcement
3) Analysis of the role of arbuscular mycorrhizal fungi in adaptive niche divergence, species coexistence, and hybrid fitness

On the whole, I believe the work presented has made significant progress in each of these areas, or, at least, uncovered important questions for further research. Strong evidence was found for a pleiotropic role in both flowering time and stress tolerance in a range of candidate speciation genes; clear patterns of hybrid misexpression relating to development and stress tolerance were uncovered; and the importance of AMF colonisation on *Howea* gene expression was demonstrated, as well identification of loci likely key to the formation of mycorrhizal associations. These findings not only demonstrate specific molecular mechanisms involved in *Howea* speciation, but are largely novel in the context of sympatric speciation, and alongside ongoing follow-up work, promise to be noteworthy contributions to the wider field of speciation biology.

Indeed, with the ever-expanding suit of tools available in the study of genomics and speciation biology, more and more detail is able to be uncovered on the precise evolutionary mechanisms driving species divergence and coexistence. Understanding these forces at a detailed level is also becoming increasingly important, given the global challenges facing humanity in terms of biodiversity loss, climate change, new pandemics, and huge pressure on increasing food production.

Whilst the findings of this thesis are minor in comparison to the scale of these challenges, they nonetheless provide novel and significant insight into the genetic architecture behind how species originate and are maintained. Further, Lord Howe Island represents an outstanding natural laboratory for studies of species divergence, with a vast range of endemic flora and fauna, largely untouched by human civilisation. Future work should make use of this environment, where the 'rare' phenomenon of sympatric speciation may in fact be common (Papadopolous et al., 2011; Osborne et al., 2020). To conclude, it is an exciting, and challenging, time to work in the field, which promises to be central to solving some of our biggest problems.

Bibliography

Ahlfors R, Lång S, Overmyer K, Jaspers P, Brosché M, Tauriainen A, Kollist H, Tuominen H, Belles-Boix E, Piippo M, Inzeé D, Palva ET, Kangasjärvi J. 2004. *Arabidopsis* RADICAL-INDUCED CELL DEATH1 belongs to the WWE protein–protein interaction domain protein family and modulates abscisic acid, ethylene, and methyl jasmonate responses. *Plant Cell* 16:1925–1937.

Akiyama K, Matsuzaki K, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435: 824–827.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22:1600-1607.

Al-Karaki GN. 2013. Application of mycorrhizae in sustainable date palm cultivation. *Emir. J. Food Agric.* 25:854-862.

Anderson EC, Thompson E. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217-1229.

Arthikala MK, Nanjareddy K, Blanco L, Alvarado-Affantranger X, Lara M. 2021. Target of rapamycin, PvTOR, is a key regulator of arbuscule development during mycorrhizal symbiosis in *Phaseolus*. *Sci. Rep.* 11:1-14.

Babik W, Butlin RK, Baker WJ, Papadopulos AST, Boulesteix M, Anstett MC, Lexer C, Hutton I, Savolainen V. 2009. How sympatric is speciation in the Howea palms of Lord Howe Island? *Mol. Ecol.* 18: 3629–3638.

Bai M, Yang G-S, Chen W-T, Mao Z-C, Kang H-X, Chen G-H, Yang Y-H, Xie B-Y. 2012. Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene* 501:52–62.

Baker NR. 2008. Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu. Rev. Plant Biol.* 59:89–113.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439:719–723.

Barreto FS, Pereira RJ, Burton RS. 2015. Hybrid dysfunction and physiological compensation in gene expression. *Mol. Biol. Evol.* 32:613-622.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1-48.

Bedford T, Hartl DL. 2009. Optimization of gene expression by natural selection. *Proc. Nat. Acad. Sci.* 106:1133-1138.

Begum N, Ahanger MA, Su Y, Lei Y, Mustafa NSA, Ahmad P, Zhang L. 2019. Improved drought tolerance by AMF inoculation in maize (*Zea mays*) involves physiological and biochemical implications. *Plants* 8:579.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57:289-300.

Boldt K, Pörs Y, Haupt B, Bitterlich M, Kühn C, Grimm B, Franken, P. 2011. Photochemical processes, carbon assimilation and RNA accumulation of sucrose transporter genes in tomato arbuscular mycorrhiza. *J. Plant. Physiol.* 168:1256-1263.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* 38:459–487.

Bradshaw Jr HD, Schemske DW. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426:176-178.

Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27:326-349.

Brennan AC, Bridle JR, Wang AL, Hiscock SJ, Abbott RJ. 2009. Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily. *New Phytol.* 183:702-717.

Brennan AC, Hiscock SJ, Abbott RJ. 2014. Interspecific crossing and genetic mapping reveal intrinsic genomic incompatibility between two *Senecio* species that form a hybrid zone on Mount Etna, Sicily. *Heredity* 113:195-204.

Brooke BP, Woodroffe CD, Murray-Wallace CV, Heijnis H, Jones BG. 2003. Quaternary calcarenite stratigraphy on Lord Howe Island, southwestern Pacific Ocean and the record of coastal carbonate deposition. *Quat. Sci. Rev.* 22:859–880.

Brundrett MC. 2002. Coevolution of roots and mycorrhiza of land plants. *New Phytol.* 154:275–304.

Butlin RK. 2008. Population genomics and speciation. *Genetica* 138:409–418.

Butlin RK, Smadja CM. 2018. Coupling, reinforcement, and speciation. *Am. Nat.* 191:155-172.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13:581-583.

Catala R, Ouyang J, Abreu IA, Hu Y, Seo H, Zhang X, Chua N-H. 2007. The *Arabidopsis* E3 SUMO ligase SIZ1 regulates plant growth and drought responses. *Plant Cell* 19:2952–2966.

Cervantes-Gámez RG, Bueno-Ibarra MA, Cruz-Mendívil A, Calderón-Vázquez CL, Ramírez-Douriet CM, Maldonado-Mendoza IE, Villalobos-López MÁ, Valdez-Ortíz Á, López-Meyer, M. 2016. Arbuscular mycorrhizal symbiosis-induced expression changes in *Solanum lycopersicum* leaves revealed by RNA-seq analysis. *Plant Mol. Biol. Rep.* 34:89-102.

Chao A. 1984. Nonparametric estimation of the number of classes in a population. *Scand. J. Stat*. 265-270.

Chapman MA, Hiscock SJ, Filatov DA. 2013. Genomic divergence during speciation driven by adaptation to altitude. *Mol. Biol. Evol.* 30:2553-2567.

Chen H, Zhang B, Hicks LM, Xiong L. 2011. A nucleotide metabolite controls stress-responsive gene expression and plant development. *PLoS One* 6:26661.

Chesson P. 2000. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. System.* 31: 343–366.

Christe C, Stölting KN, Bresadola L, Fussi B, Heinze B, Wegmann D, Lexer C. 2016. Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Mol. Ecol.* 25:2482-2498.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana. Plant J.* 16:735-743.

Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* 24:797-808.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23:3133-3157.

Curtu AL, Gailing O, Finkeldey R. 2009. Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. *BMC Evol. Biol.* 9:1-9.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008.

Darwin C. 1859. On the origin of species. London: John Murray.

Davidson NM, Oshlack A. 2014. Corset: enabling differential gene expression analysis for de novoassembled transcriptomes. *Genome Biol.* 15:1-14.

Dieckmann U, Doebeli MO. 1999. On the origin of species by sympatric speciation. *Nature* 400:354–357.

Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol. Biol. Evol.* 31:1188-1199.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21.

Dobzhansky T. 1937. Genetic nature of species differences. *Am. Nat.* 71:404-420.

Doebeli M, Dieckmann U, Metz JA, Tautz D. 2005. What we have also learned: adaptive speciation is theoretically plausible. *Evolution* 59:691–699.

Dunning LT, Hipperson H, Baker WJ, Butlin RK, Devaux C, Hutton I, Igea J, Papadopulos AST, Quan X, Smadja CM, Turnbull CGN, Savolainen V. 2016. Ecological speciation in sympatric palms: 1. Gene expression, selection and pleiotropy. *J. Evol. Biol.* 29:1472–1487.

Éva C, Tóth G, Oszvald M, Tamás L. 2014. Overproduction of an *Arabidopsis* aldo-keto reductase increases barley tolerance to oxidative and cadmium stress by an in vivo reactive aldehyde detoxification. *Plant Growth Regul.* 74:55–63.

Éva C, Solti Á, Oszvald M, Tömösközi-Farkas R, Nagy B, Horváth GV, Tamás L. 2016. Improved reactive aldehyde, salt and cadmium tolerance of transgenic barley due to the expression of aldo–keto reductase genes. *Acta Physiol. Plant.* 38:99.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* 28:342-350.

Feder JL, Nosil P, Wacholder AC, Egan SP, Berlocher SH, Flaxman SM. 2014. Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *J. Hered.* 105:810-820.

Field KJ, Pressel S, Duckett JG, Rimington WR, Bidartondo MI. 2015. Symbiotic options for the conquest of land. *Trends Ecol. Evol.* 30:477–486.

Flaxman SM, Wacholder AC, Feder JL, Nosil P. 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* 23:4074-4088.

Foote AD. 2018. Sympatric speciation in the genomic era. *Trends Ecol. Evol.* 33:85-95.

Forbes AA, Powell THQ, Stelinski LL, Smith JJ, Feder JL. 2009. Sequential sympatric speciation across trophic levels. *Science* 323:776–779.

Funk DJ, Nosil P, Etges WJ. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci.* 103:3209-3213.

Gao C, Courty PE, Varoquaux N, Cole B, Montoya L, Xu L, Purdom E, Vogel J, Hutmacher RB, Dahlberg JA, Coleman-Derr D. 2022. Successional adaptive strategies revealed by correlating arbuscular mycorrhizal fungal abundance with host plant gene expression. *Mol. Ecol.* 1-14.

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.* 15:1494-500.

Gaur A, Gaur A, Adholeya A. 2000. Growth and flowering in *Petunia hybrida*, *Callistephus chinensis* and *Impatiens balsamina* inoculated with mixed AM inocula or chemical fertilizers in a soil of low P fertility. *Sci. Hortic.* 84:151–162.

Gavrilets S. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton.

Gavrilets S, Vose A. 2007. Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Mol. Ecol.* 16: 2910–2921.

Gelvin SB. 2003. Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev*. 67:16-37.

Gervasi DD, Selosse MA, Sauve M, Francke W, Vereecken NJ, Cozzolino S, Schiestl FP. 2017. Floral scent and species divergence in a pair of sexually deceptive orchids. *Ecol. Evol.* 7:6023-6034.

Gohar S, Sajjad M, Zulfiqar S, Liu J, Wu J, Rahman MU. 2022. Domestication of newly evolved hexaploid wheat—A journey of wild grass to cultivated wheat. *Front. Genet.* 13*:*1022931.

Glover AN, Bendall EE, Terbot JW, Payne N, Webb A, Filbeck A, Norman G, Linnen CR. 2023. Body size as a magic trait in two plant-feeding insect species. *Evolution* 77:437-453.

Golldack D, Popova OV, Dietz KJ. 2002. Mutation of the matrix metalloproteinase *At2-MMP* inhibits growth and causes late flowering and early senescence in *Arabidopsis*. *J. Biol. Chem.* 277:5541–5547.

Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pagès V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC, Bouwmeester H. 2008. Strigolactone inhibition of shoot branching. *Nature* 455:189-194.

Gotelli NJ, Colwell RK. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* 4:379-391.

Grant PR. 1999. Ecology and evolution of Darwin's finches. Princeton (NJ): Princeton University Press.

Grant PR, Grant BR. 2006. Evolution of character displacement in Darwin's finches. *Science* 313:224-226.

Guether M, Neuhäuser B, Balestrini R, Dynowski M, Ludewig U, Bonfante P. 2009. A mycorrhizal-specific ammonium transporter from *Lotus japonicus* acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant Physiol.* 150:73-83.

Gutjahr C, Gobbato E, Choi J, Riemann M, Johnston MG, Summers W, Carbonnel S, Mansfield C, Yang SY, Nadal M, Acosta I. 2015. Rice perception of symbiotic arbuscular mycorrhizal fungi requires the karrikin receptor complex. *Science* 350:1521-1524.

Guo W, Tzioutziou NA, Stephen G, Milne I, Calixto CP, Waugh R, Brown JW, Zhang R. 2021. 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential

expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biol.* 18:574-1587.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* 8:1494-512.

Hall A, Bastow RM, Davis SJ, Hanano S, McWatters HG, Hibberd V, Doyle MR, Sung S, Halliday KJ, Amasino RM, Millar AJ, 2003. The TIME FOR COFFEE gene maintains the mmplitude and timing of *Arabidopsis* circadian clocks. *Plant Cell* 16:183–192.

Haller BC, De Leon LF, Rolshausen G, Gotanda KM, Hendry AP. 2012. Magic traits: distinguishing the important from the trivial. *Smithsonian* 3:2006.

Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. 2009. Extreme changes to gene expression associated with homoploid hybrid speciation. *Mol. Ecol.* 18:877-889.

Hipperson H, Dunning LT, Baker WJ, Butlin RK, Hutton I, Papadopulos AST, Smadja CM, Wilson TC, Devaux C, Savolainen V. 2016. Ecological speciation in sympatric palms: 2. Pre- and post-zygotic isolation. *J. Evol. Biol.* 29:2143–2156.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 65-70.

Hong S, Lim YP, Kwon SY, Shin AY, Kim YM. 2021. Genome-wide comparative analysis of flowering-time genes; insights on the gene family expansion and evolutionary perspective. *Front. Plant Sci.* 12:702243.

Jamil M, Charnikhova T, Verstappen F, Ali Z, Wainwright H, Bouwmeester HJ. 2014. Effect of phosphate-based seed priming on strigolactone production and *Striga hermonthica* infection in cereals. *Weed Res.* 54:307-313.

Jiggins CD, Naisbit RE, Coe RL, Mallet J. 2001. Reproductive isolation caused by colour pattern mimicry. *Nature* 411:302-305.

Jin JB, Jin YH, Lee J, Miura K, Yoo CY, Kim WY, Van Oosten M, Hyun Y, Somers DE, Lee I, Yun DJ, Bressan RA, Hasegawa PM. 2008. The SUMO E3 ligase, *AtSIZ1*, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through effects on *FLC* chromatin structure. *Plant J.* 53:530–540.

Jin Z, Li J, Li Y. 2015. Interactive effects of arbuscular mycorrhizal fungi and copper stress on flowering phenology and reproduction of *Elsholtzia splendens*. *PLoS One* 10:1–10.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55.

Katiyar-Agarwal S, Zhu J, Kim K, Agarwal M, Fu X, Huang A, Zhu J-K. 2006. The plasma membrane Na+/H+ antiporter *SOS1* interacts with *RCD1* and functions in oxidative stress tolerance in *Arabidopsis*. *Proc. Natl. Acad. Sci.* 103:18816–18821.

Kautt AF, Machado-Schiaffino G, Torres-Dowdall J, Meyer A. 2016. Incipient sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater lakes in Nicaragua due to differential use of the benthic and limnetic habitats? *Ecol. Evol.* 6:5342–5357.

Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, Torres-Dowdall J, Härer A, Hulsey CD, Franchini P, Pippel M. 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* 588:106-11.

Kim D-H, Sung S. 2010. The Plant Homeo Domain finger protein, *VIN3*-LIKE 2, is necessary for photoperiod-mediated epigenetic regulation of the floral repressor, *MAF5*. *Proc. Natl. Acad. Sci.* 107:17029–17034.

Kirkpatrick M. 2001. Reinforcement during ecological speciation. *Proc. R. Soc. B Biol. Sci.* 268:1259–1263.

Kirkpatrick M, Ravigne V. 2002. Speciation by natural and sexual selection: models and experiments. *Am. Nat.* 159:S22–S35.

Kitajima MBWL, Butler WL. 1975. Quenching of chlorophyll fluorescence and primary photochemistry in chloroplasts by dibromothymoquinone. *Biochim. Biophys. Acta-Bioenerg.* 376:105-115.

Kondrashov AS, Kondrashov FA. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* 400:351–354.

Korves TM, Bergelson J. 2003. A developmental response to pathogen infection in *Arabidopsis*. *Plant Physiol.* 133:339–347.

Kulmuni J, Westram, AM. 2017. Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Mol. Ecol.* 26:3093-3103.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82:1-26.

Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid speciation in Darwin's finches. *Science* 359:224-228.

Landry CR, Hartl DL, Ranz JM. 2007. Genome clashes in hybrids: Insights from gene expression. *Heredity* 99:483–493.

Lee J, Lee S, Young JPW. 2008. Improved PCR primers for the detection and identification of arbuscular mycorrhizal fungi. *FEMS Microbiol. Ecol.* 65:339-349.

Lee G, Park Y. 2017. A co-expression network of drought stress-related genes in chinese cabbage. *Hortic. Sci. Technol.* 35:243–251.

Lenth RV. 2016. Least-Squares Means: The R package lsmeans. *J. Stat. Softw.* 69:1-33.

Lepais O, Roussel G, Hubert F, Kremer A, Gerber S. 2013. Strength and variability of postmating reproductive isolating barriers between four European white oak species. *Tree Genet. Genomes* 9:841-853.

Levin DA. 2009. Flowering-time plasticity facilitates niche shifts in adjacent populations. *New Phytol.* 183:661–666.

Li P, Li YJ, Wang B, Yu HM, Li Q, Hou BK. 2017. The *Arabidopsis UGT87A2*, a stress-inducible family 1 glycosyltransferase, is involved in the plant adaptation to abiotic stresses. *Physiol. Plant.* 159:416–432.

Liu J, He Y, Amasino R, Chen X. 2004. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev*. 18:2873-8.

Liu J, Maldonado-Mendoza I, Lopez-Meyer M, Cheung F, Town CD, Harrison MJ. 2007. Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *Plant J.* 50:529-544.

Liu Q, Feng Y, Zhu Z. 2009. Dicer-like (*DCL*) proteins in plants. *Funct. Integr. Genomics* 9:277–286.

Liu, Q, Luo L, Zheng L. 2018. Lignins: biosynthesis and biological functions in plants. *Int. J Mol Sci.* 19:335.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:1-21.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. Royal Soc. B* 363:3009-3021.

Luo X, Yin M, He Y. 2022. Molecular genetic understanding of photoperiodic regulation of flowering time in *Arabidopsis* and soybean. *Int. J. Mol. Sci.* 23:466.

Lyons R, Rusu A, Stiller J, Powell J, Manners JM, Kazan K. 2015. Investigating the association between flowering time and defense in the *Arabidopsis thaliana–Fusarium oxysporum* interaction. *PLoS One* 10:e0127699.

Mack KL, Campbell P, Nachman MW. 2016. Gene regulation and speciation in house mice. *Genome Res.* 26:451–461.

MacArthur R, Levins R. 1967. The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.* 101:377-385.

Mack KL, Nachman MW. 2017. Gene regulation and speciation. *Trends Genet.* 33:68-80.

MacManes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* 5:13.

Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate Cichlid ectomorphs in an East African crater lake. *Science* 350:1493–1498.

Manceau M, Domingues VS, Mallarino R, Hoekstra HE. 2011. The developmental role of Agouti in color pattern evolution. *Science* 331:1062-1065.

Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics* 12:1005887.

Martin CH, Cutler JS, Friel JP, Touokong CD, Coop G, Wainwright PC. 2015. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution* 69:1406–1422.

Martin NH, Willis JH. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61:68-82.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res.* 23:1817–1828.

Maxwell K, Johnson GN. 2000. Chlorophyll fluorescence—a practical guide. *J. Exp. Bot.* 51:659-668.

McDougall I, Embleton BJJ, Stone DBI. 1981. Origin and evolution of Lord Howe Island, Southwest Pacific Ocean. *J. Geol. Soc. Aust.* 28:155–176.

McGirr JA, Martin CH. 2020. Ecological divergence in sympatry causes gene misexpression in hybrids. *Mol. Ecol.* 29:2707-2721.

McKinnon JS, Mori S, Blackman BK, David L, Kingsley DM, Jamieson L, Chou J, Schluter D. 2004. Evidence for ecology's role in speciation. *Nature* 429:294.

McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 20:816-825.

McMurdie PJ, Holmes S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 8:e61217.

McNeilly T. 1968. Evolution in closely adjacent plant populations III. *Agrostis tenuis* on a small copper mine. *Heredity* 23:99-10.

McNeilly T, Antonovics J. 1968. Evolution in closely adjacent plant populations IV. Barriers to gene flow. *Heredity* 23:205–218.

McNeilly T, Bradshaw AD. 1968. Evolutionary processes in populations of copper tolerant *Agrostis tenuis* Sibth. *Evolution* 22:108-118.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:1–11.

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M.1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282:662-682.

Merlot S, Mustilli AC, Genty B, North H, Lefebvre V, Sotta B, Vavasseur A, Giraudat J. 2002. Use of infrared thermal imaging to isolate *Arabidopsis* mutants defective in stomatal regulation. *Plant J.* 30:601-609.

Michalak P, Noor MA. 2003. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol. Biol. Evol.* 20:1070-1076.

Miura K, Lee J, Gong Q, Ma S, Jin JB, Yoo CY, Miura T, Sato A, Bohnert HJ, Hasegawa PM. 2011. SIZ1 regulation of phosphate starvation-induced root architecture remodeling involves the control of auxin accumulation. *Plant Physiol.* 155:1000–1012.

Moora M, Berger S, Davison J, Öpik M, Bommarco R, Bruelheide H, Kühn I, Kunin WE, Metsis M, Rortais A, Vanatoa A. 2011. Alien plants associate with widespread generalist arbuscular mycorrhizal fungal taxa: evidence from a continental-scale study using massively parallel 454 sequencing. *J. Biogeogr.* 38:1305-1317.

Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, Meiners T, Müller C, Obermaier E, Prati D, Socher SA. 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* 4:3514-3524.

Muir G, Osborne OG, Sarasa J, Hiscock SJ, Filatov DA. 2013. Recent ecological selection on regulatory divergence is shaping clinal variation in *Senecio* on Mount Etna. *Evolution* 67:3032-3042.

Muller HJ. 1942. Isolating mechanisms, evolution, and temperature. *Biol. Symp.* 6:71-125.

Murali A, Bhargava A, Wright ES. 2018. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* 6:1-14.

Nagel L, Schluter D. 1998. Body size, natural selection, and speciation in sticklebacks. *Evolution* 52:209-218.

Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y. 2007. Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* 71:2095-2100.

Noor MA, Feder JL. 2006. Speciation genetics: evolving approaches. *Nat. Rev. Genet.* 7:851-861.

Nosil P. 2008. Speciation with gene flow could be common. *Mol. ecol.* 17:2103-2106.

Nosil P. 2012. Ecological speciation. Oxford University Press, Oxford.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H, Oksanen MJ. 2013. Package 'vegan'. *Community Ecology Package, Version* 2:1-295.

Öpik M, Vanatoa A, Vanatoa E, Moora M, Davison J, Kalwij JM, Reier Ü, Zobel M. 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (*Glomeromycota*). *New Phytol.* 188:223-241.

Ortíz-Barrientos D, Counterman BA, Noor MAF. 2007. Gene expression divergence and the origin of hybrid dysfunctions. *Genetica* 129:71–81.

Ortu G, Balestrini R, Pereira PA, Becker JD, Küster H, Bonfante P. 2012. Plant genes related to gibberellin biosynthesis and signaling are differentially regulated during the early stages of AM fungal interactions. *Mol. Plant* 5:951-954.

Osborne OG, Batstone TE, Hiscock SJ, Filatov DA. 2013. Rapid speciation with gene flow following the formation of Mt. Etna. *Genome Biol. Evol.* 5:1704-1715.

Osborne OG, De-Kayne R, Bidartondo MI, Hutton I, Baker WJ, Turnbull CGN, Savolainen V. 2018. Arbuscular mycorrhizal fungi promote coexistence and niche divergence of sympatric palm species on a remote oceanic island. *New Phytol.* 217:1254–1266.

Osborne OG, Ciezarek A, Wilson T, Crayn D, Hutton I, Baker WJ, Turnbull CGN, Savolainen V. 2019. Ecological speciation in Howea palms occurred in sympatry and was associated with edaphic and phenological adaptation. *Mol. Biol. Evol.*

Osborne OG, Kafle T, Brewer T, Dobreva MP, Hutton I, Savolainen V. 2020. Sympatric speciation in mountain roses (Metrosideros) on an oceanic island. *Phil. Trans. R. Soc. B.* 375:20190542.

Panke-Buisse K, Poole AC, Goodrich JK, Ley RE, Kao-Kniffin J. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* 9:980–989.

Papadopulos AST, Baker WJ, Crayn D, Butlin RK, Kynast RG, Hutton I, Savolainen V. 2011. Speciation with gene flow on Lord Howe Island. *Proc. Natl. Acad. Sci.* 108:13188–13193.

Papadopulos AST, Price Z, Devaux C, Hipperson H, Smadja CM, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2013. A comparative analysis of the mechanisms underlying speciation on Lord Howe Island. *J. Evol. Biol.* 26:733–745.

Papadopulos AST, Kaye M, Devaux C, Hipperson H, Lighten J, Dunning LT, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2014. Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 1648: 20130342.

Papadopulos AS, Igea J, Dunning LT, Osborne OG, Quan X, Pellicer J, Turnbull C, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2019a. Ecological speciation in sympatric palms: 3. Genetic map reveals genomic islands underlying species divergence in *Howea*. *Evolution*. 73:1986-1995.

Papadopulos AS, Igea J, Smith TP, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2019b. Ecological speciation in sympatric palms: 4. Demographic analyses support speciation of *Howea* in the face of high gene flow. *Evolution.* 73:1996-2002.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14:417-419.

Pavicic M, Mouhu K, Wang F, Bilicka M, Chovanček E, Himanen K. 2017. Genomic and phenomic screens for flower related RING type ubiquitin E3 ligases in *Arabidopsis*. *Front. Plant Sci*. 28:416.

Pavithra D, Yapa N. 2018. Arbuscular mycorrhizal fungi inoculation enhances drought stress tolerance of plants. *Groundw. Sustain. Dev.* 7:490-494.

Pickard J. 1983. Vegetation of Lord Howe Island. *Cunninghamia* 1:133–265.

Podos J, Nowicki S. 2004. Beaks, adaptation, and vocal evolution in Darwin's finches. *BioScience* 54:501–510.

Podos J, Southall JA, Rossi-Santos M. 2004. Vocal mechanics in Darwin's finches: correlation of beak gape and song frequency. *J. Exp. Biol*. 207:607–619.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf JB, 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410-1414.

Poelstra JW, Richards EJ, Martin CH. 2018. Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Mol. Ecol.* 27:4270-4288.

Potts BM, Dungey HS. 2004. Interspecific hybridization of *Eucalyptus*: key issues for breeders and geneticists. *New Forests* 27:115-138.

Pouteau S, Albertini C. 2009. The significance of bolting and floral transitions as indicators of reproductive phase change in *Arabidopsis*. *J. Exp. Bot.* 60:3367-3377.

Przedpelska-Wasowicz E, Wasowicz P. 2013. Does zinc concentration in the substrate influence the onset of flowering in *Arabidopsis arenosa* (Brassicaceae)? *Plant Growth Regul.* 69:87–97.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 41:D590-D596.

Quintero FJ, Garciadeblas B, Rodríguez-Navarro A. 1996. The SAL1 gene of *Arabidopsis*, encoding an enzyme with 3'(2'), 5'-bisphosphate nucleotidase and inositol polyphosphate 1-phosphatase activities, increases salt tolerance in yeast. *Plant Cell* 8:529-537.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Evol. Syst.* 29:467-501.

Ranz JM, Namgyal K, Gibson G, Hartl DL. 2004. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res.* 14:373-379.

Rao CR. 1982. Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* 21:24–43.

Ray A, Lang JD, Golden T, Ray S. 1996. SHORT INTEGUMENT (SIN1), a gene required for ovule development in *Arabidopsis*, also controls flowering time. *Development*. 122:2631-2638.

Renaut S, Nolte AW, Bernatchez L. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Biol. Evol.* 26:925-936.

Renaut S, Bernatchez L. 2011. Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity* 106:1003-1011.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun*. 4:1827.

Reynolds RG, Fitzpatrick BM. 2007. Assortative mating in poison-dart frogs based on an ecologically important trait. *Evolution* 61:2253-2259.

Richards EJ, Poelstra JW, Martin CH. 2018. Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evol. Lett.* 2:524-40.

Richards EJ, Servedio MR, Martin CH. 2019. Searching for sympatric speciation in the genomic era. *BioEssays*. 41:1900047.

Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, Lucek K, Hellen E, Soria-Carrasco V, Dennis SR, de Carvalho CF, Safran RJ, Sandoval CP, Feder J, Gries R, Crespi BJ, Gries G, Gompert Z, Nosil P. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nat. Ecol. Evol.* 1:0082.

Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* 317:910-914.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:1-9.

Rolán-Alvarez E. 2007. Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. *J. Molluscan Stud.* 73:1-10.

Roswell M, Dushoff J, Winfree R. 2021. A conceptual guide to measuring species diversity. *Oikos* 130:321-338.

Rottscheidt R, Harr B. 2007. Extensive additivity of gene expression differentiates subspecies of the house mouse. *Genetics* 177:1553-1567.

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecol. Lett*. 8:336–352.

Salvioli A, Zouari I, Chalot M, Bonfante P. 2012. The arbuscular mycorrhizal status has an impact on the transcriptome profile and amino acid composition of tomato fruit. *BMC Plant Biol.* 12:44.

Sanchez-Villarreal A, Shin J, Bujdoso N, Obata T, Neumann U, Du SX, Ding Z, Davis AM, Shindo T, Schmelzer E, Sulpice R, Nunes-Nesi A, Stitt M, Fernie AR, Davis SJ. 2013. TIME for COFFEE is an essential component in the maintenance of metabolic homeostasis in *Arabidopsis thaliana*. *Plant J.* 76:188–200.

Sato A, Tichy H, O'hUigin C, Grant PR, Grant BR, Klein J. 2001. On the O

origin of Darwin's finches. *Mol. Biol. Evol.* 18:299–311.

Sato K, Suyama Y, Saito M, Sugawara K. 2005. A new primer for discrimination of arbuscular mycorrhizal fungi with polymerase chain reaction-denature gradient gel electrophoresis. *Grassl. Sci.* 51:179-181.

Savolainen V, Anstett M-C, Lexer C, Hutton I, Clarkson JJ, Norup M V, Powell MP, Springate D, Salamin N, Baker WJ. 2006a. Sympatric speciation in palms on an oceanic island. *Nature* 441:210–213.

Savolainen V, Lexer C, Anstett MC, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Bake WJ. 2006b. Evolutionary biology: Sympatric plant speciation in islands?(Reply). *Nature* 443:E12.

Schaarschmidt S, Roitsch T, Hause B. 2006. Arbuscular mycorrhiza induces gene expression of the apoplastic invertase LIN6 in tomato (*Lycopersicon esculentum*) roots. *J. Exp. Bot.* 57:4015-4023.

Schliewen UK, Tautz D, Pääbo S. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368:629–632.

Schluter D. 2000. The ecology of adaptive radiation. Oxford: Oxford University Press.

Schluter D. 2001. Ecology and the origin of species. *Trends Ecol. Evol*. 16: 372–380.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323:737-741.

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28:1086–1092.

Schmitz RJ, Hong L, Fitzpatrick KE, Amasino RM. 2007. DICER-LIKE 1 and DICER-LIKE 3 redundantly act to promote flowering via repression of FLOWERING LOCUS C in *Arabidopsis thaliana*. *Genetics* 176:1359–1362.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List, Bolchacova E, Voigt K. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Nat. Acad. Sci.* 109:6241-6246.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19:198–207.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brännström Å, Brelsford A. 2014. Genomics and the origin of species. *Nature Rev. Genet.* 15:176-192.

Seehausen O. 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* 207:304–312.

Servedio MR, Doorn GSV, Kopp M, Frame AM, Nosil P. 2011. Magic traits in speciation: "magic" but not rare? *Trends Ecol. Evol.* 26:389–397.

Shannon C. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423.

Signor SA, Nuzhdin SV. 2018. The evolution of gene expression in cis and trans. *Trends Genet.* 34:532-544.

Simpson GG, Dean C. 2002. *Arabidopsis*, the Rosetta stone of flowering time? *Science* 296:285-289.

Simpson PJ, Tantitadapitak C, Reed AM, Mather OC, Bunce CM, White SA, Ride JP. 2009. Characterization of two novel aldo-keto reductases from *Arabidopsis*: expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism following stress. *J. Mol. Biol.* 392:465–480.

Slatkin M. 1982. Pleiotropy and parapatric speciation. *Evolution* 32:263-270.

Smadja CM, Butlin RK. 2011. A framework for comparing processes of speciation in the presence of gene flow. *Mol. Ecol.* 20:5123–5140.

Smith SE, Read D. 1997. Mycorrhizal symbiosis. London: Academic Press.

Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4:13742-015.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, Egan SP. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344:738-742.

Suzuki K, Takahashi K, Harada N. 2020. Evaluation of primer pairs for studying arbuscular mycorrhizal fungal community compositions using a MiSeq platform. *Biol. Fertil. Soils* 56:853-858.

Strack D, Fester T. 2006. Isoprenoid metabolism and plastid reorganization in arbuscular mycorrhizal roots. *New Phytol.* 172:22-34.

Swain Lenz D, Riles L, Fay JC. 2014. Heterochronic meiotic misexpression in an interspecific yeast hybrid. *Mol. Biol. Evol.* 31:1333-1342.

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A. 2007. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36:D1009-D1014.

Tran L-SP. 2004. Isolation and functional analysis of *Arabidopsis s*tress-inducible *NAC* transcription factors that bind to a drought-responsive *cis-e*lement in the *early responsive to dehydration stress 1* promoter. *Plant Cell* 16:2481–2498.

Vandepitte K, De Meyer T, Helsen K, Van Acker K, Roldán-Ruiz I, Mergeay J, Honnay O. 2014. Rapid genetic adaptation precedes the spread of an exotic plant species. *Mol. Ecol.* 23:2157–2164.

Van Doorn GS, Weissing FJ. 2002. Ecological versus sexual selection models of sympatric speciation: a synthesis. *Selection* 2:17-40.

Van Geel M, Busschaert P, Honnay O, Lievens B. 2014. Evaluation of six primer pairs targeting the nuclear rRNA operon for characterization of arbuscular mycorrhizal fungal (AMF) communities using 454 pyrosequencing. *J. Microbiol. Methods* 106:93-100.

Vangelisti A, Natali L, Bernardi R, Sbrana C, Turrini A, Hassani-Pak K, Hughes D, Cavallini A, Giovannetti M, Giordani T. 2018. Transcriptome changes induced by arbuscular mycorrhizal fungi in sunflower (*Helianthus annuus L.*) roots. *Sci. Rep.* 8:4.

Wagner MR, Lundberg DS, Coleman-Derr D, Tringe SG, Dangl JL, Mitchell-Olds T. 2014. Natural soil microbes alter flowering phenology and the intensity of selection on flowering time in a wild *Arabidopsis* relative. *Ecol. Lett.* 17:717–726.

Walter MH, Fester T, Strack, D. 2000. Arbuscular mycorrhizal fungi induce the non-mevalonate methylerythritol phosphate pathway of isoprenoid biosynthesis correlated with accumulation of the 'yellow pigment' and other apocarotenoids. *Plant J.* 21:571-578.

Wang WY, Xu J, Liu XJ, Yu Y, Ge Q. 2012. Cadmium induces early flowering in *Arabidopsis. Biol. Plant*. 56:117–120.

Wang H, Yin H, Jiao C, Fang X, Wang G, Li G, Ni F, Li P, Su P, Ge W, Lyu Z. 2020. Sympatric speciation of wild emmer wheat driven by ecology and chromosomal rearrangements. *Proc. Natl. Acad. Sci.* 117:5955-63.

Westram AM, Rafajlović M, Chaube P, Faria R, Larsson T, Panova M, Ravinet M, Blomberg A, Mehlig B, Johannesson K, Butlin R. 2018. Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evol. Lett.* 2:297-309.

Widmer A, Lexer C, Cozzolino S. 2009. Evolution of reproductive isolation in plants. *Heredity* 102:31.

Wilson PB, Estavillo GM, Field KJ, Pornsiriwong W, Carroll AJ, Howell KA, Woo NS, Lake JA, Smith SM, Harvey Millar A, Von Caemmerer S, Pogson BJ. 2009. The nucleotidase/phosphatase *SAL1* is a negative regulator of drought tolerance in *Arabidopsis. Plant J.* 58:299–317.

Wong EL, Nevado B, Osborne OG, Papadopulos AS, Bridle JR, Hiscock SJ, Filatov DA. 2020. Strong divergent selection at multiple loci in two closely related species of ragworts adapted to high and low elevations on Mount Etna. *Mol. Ecol.* 29:394-412.

Woo NS, Badger MR, Pogson BJ. 2008. A rapid, non-invasive procedure for quantitative assessment of drought survival using chlorophyll fluorescence. *Plant Methods* 4:1-4.

Wu CI. 2001. The genic view of the process of speciation. *J. Evol. Biol.* 14:851-865.

Wu QS, Srivastava AK, Zou YN. 2013. AMF-induced tolerance to drought stress in citrus: A review. *Sci. Hortic.* 164:77-87.

Xiong L, Lee BH, Ishitani M, Lee H, Zhang C, Zhu JK. 2001. *FIERY1* encoding an inositol polyphosphate 1-phosphatase is negative regulator of abscisic acid and stress signaling in *Arabidopsis*. *Genes Dev*. 15:1971–1984.

Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y, Wu P. 2005. OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol.* 138:2087–2096.

Zhang L, Zhang C, Wu P, Chen Y, Li M, Jiang H, Wu G. 2014. Global analysis of gene expression profiles in physic nut (*Jatropha curcas* L.) seedlings exposed to salt stress. *PLoS One* 9:97878.

Zhao P, Zhang F, Liu D, Imani J, Langen G, Kogel KH. 2017. Matrix metalloproteinases operate redundantly in *Arabidopsis* immunity against necrotrophic and biotrophic fungal pathogens. *PloS One* 12:e0183577.

Appendices

Appendix I. Chapter Two Supplementary Materials

Arabidopsis	Gene	Stock Number	Germplasm	Polymorphism
AT1G32230	RCD1	N923498	CS923498	SALKseq_046241.1
AT1G70170	MMP	N689592	SALK_201873C	SALK_201873
AT5G60410	SIZ1	N680159	SALK_023805C	SALK_023805.18.15.x
AT4G30200	VEL1	N869906	CS869906	SALKseq_137263.0
AT2G30140	UGT87A2	N680914	SALK_061574C	SALK_061574
AT1G01040	DCL1	N671419	SALK_056243C	SALK_056243.55.75.X
AT2G37770	AKR4C9	N673308	SALK_119576C	SALK_119576.26.60.x
AT3G22380	TIC	N675986	SALK_083184C	SALK_083184.30.65.x
AT5G20320	DCL4	N696794	SALK_206940C	SALK_206940
AT4G27410	NAC072	N662720	SALK_063576C	SALK_063576
AT5G63980	SAL1	N520882	SALK_020882	SALK_020882.55.50., fry1-6

Supplementary Table I.1. Knockout *A. thaliana* lines used for phenotyping experiments, acquired from the European *Arabidopsis* Stock Centre (NASC).

Arabidopsis	Gene	Forward Primer	Reverse Primer
AT1G32230	RCD1	ACGTTACATGGTTTTATGCCG	AAAGACCCAACGACATCATTG
AT1G70170	MMP	CCCAAAATATTTTACTGTGCTTTTG	AAGTACCGTTGACCACGTCAG
AT5G60410	SIZ1	AAAGAGAGAGTGAGCGAAGGG	TTCATACCCAGAAGCAACGAC
AT4G30200	VEL1	TTTGGCATCGTAAGGTACCAG	GTCTCTGCAACCTCTTTGTGC
AT2G30140	UGT87A2	GCTTCACATAAGCGATCGATC	TAGGGTTGATGTGTCCTCGAC
AT1G01040	DCL1	TTATGGGAAAGAAGTTGCACG	ATTGTTTGGTAAGCCTCCACC
AT2G37770	AKR4C9	CATCGTTCCATCTCAAACTC	TCAGGTCGAATGTCATCCTTC
AT3G22380	TIC	TTCGATATTGAACCCCATACG	TAACATGAAAGGAGGTGGTGG
AT5G20320	DCL4	TCACCAACTTTAGTGGCAACC	GAATGCATCTGACCTCCTCAG
AT4G27410	NAC072	AGTGATCGAGTGCTTCAGGAC	ACTCGTGCATAATCCAGTTGG
AT5G63980	SAL1	CTTCGAATGACTCGAAGAACG	TTTTCGATTCAATCATGACCC
LBb1.3		ATTTTGCCGATTTCGGAAC	

Supplementary Table I.2. Primer pairs used to verify homozygosity of knockout mutants. Includes the LBb1.3 SALK T-DNA specific forward primer

Supplementary Figures I.1-I.10. The following pages contain charts showing measurements of phenotypic traits over time for all knockout lines (other than *sal1*) alongside wild type plants. In all charts, knockout genotypes are represented by pink lines, with Col-0 from the relevant experimental round group shown in blue. In all charts, solid lines represent plants grown under control treatment; dotted lines represent salt treatment; dashed lines represent drought treatment. In Figures S1 and S4, dashed-dotted lines additionally represent cadmium treatment. In all Figures, panels show phenotypic traits as follows: **A)** Rosette Area (mm^2); **B)** Perimeter (mm); **C)** QY Max; **D)** Mean Leaf Temperature (°C); **E)** Compactness; **F)** Eccentricity; **G)** Isotropy; **H)** SOL; **I)** Roundness 1; **J)** Roundness 2. RMS measurements are not shown. In all charts, X axes show days from imposition of stress treatments. Error bars show standard error.

Supplementary Figure I.1. *akr4c9*. Phenotypic traits of *akr4c9* (pink) versus Col-0 3 (blue) over time.

Supplementary Figure I.2. *dcl1*. Phenotypic traits of *dcl1* (pink) versus Col-0 1 (blue) over time.

Supplementary Figure I.3. *dcl4*. Phenotypic traits of *dcl4* (pink) versus Col-0 2 (blue) over time.

Supplementary Figure I.4. *At2-mmp*. Phenotypic traits of *At2-mmp* (pink) versus Col-0 3 (blue) over time.

Supplementary Figure I.5. *nac072*. Phenotypic traits of *nac072* (pink) versus Col-0 2 (blue) over time.

Supplementary Figure I.6. *rcd1*. Phenotypic traits of *rcd1* (pink) versus Col-0 1 (blue) over time.

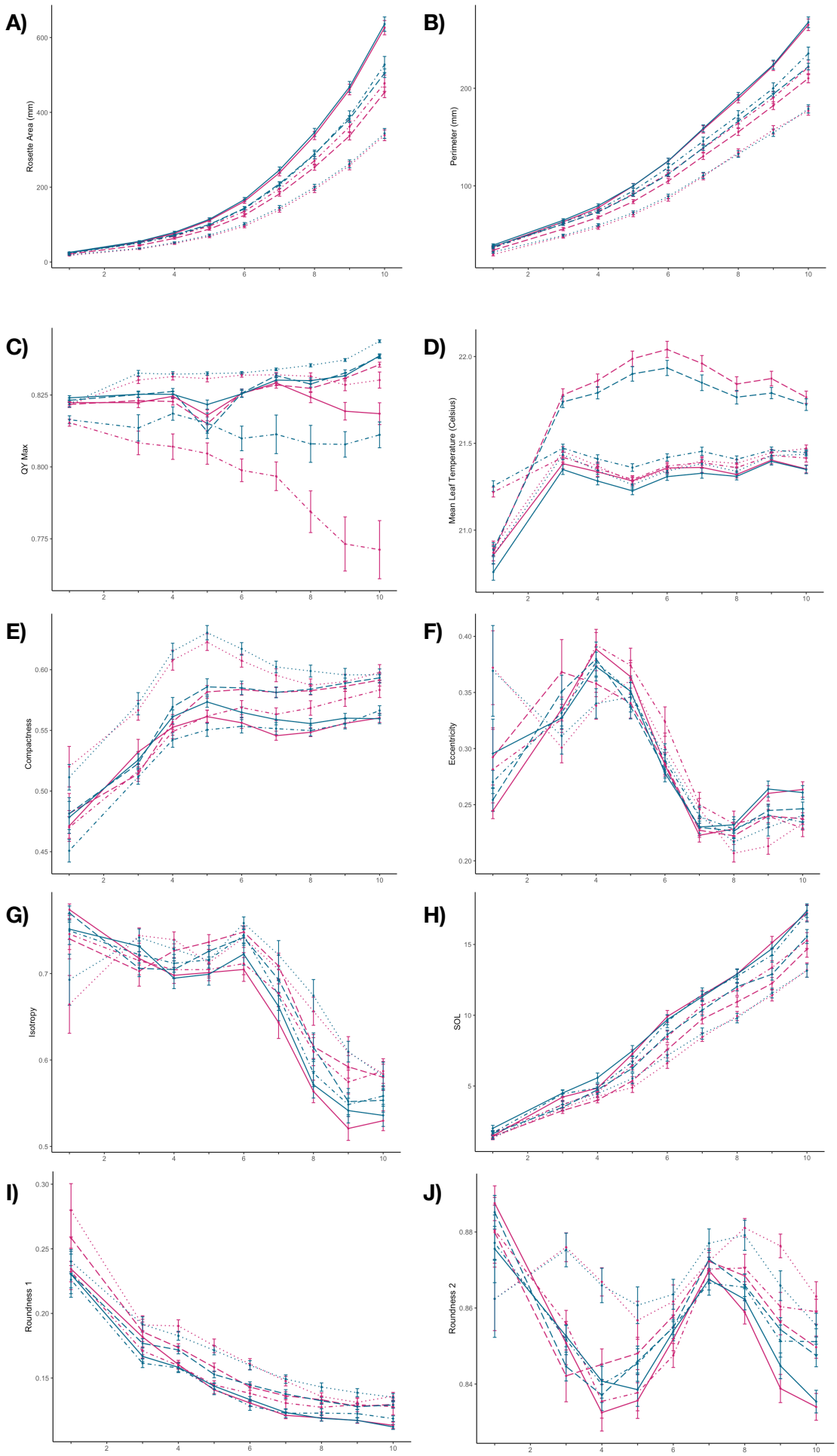
Supplementary Figure I.7. *siz1*. Phenotypic traits of *siz1* (pink) versus Col-0 1 (blue) over time.

Supplementary Figure I.8. *tic*. Phenotypic traits of *tic* (pink) versus Col-0 1 (blue) over time.

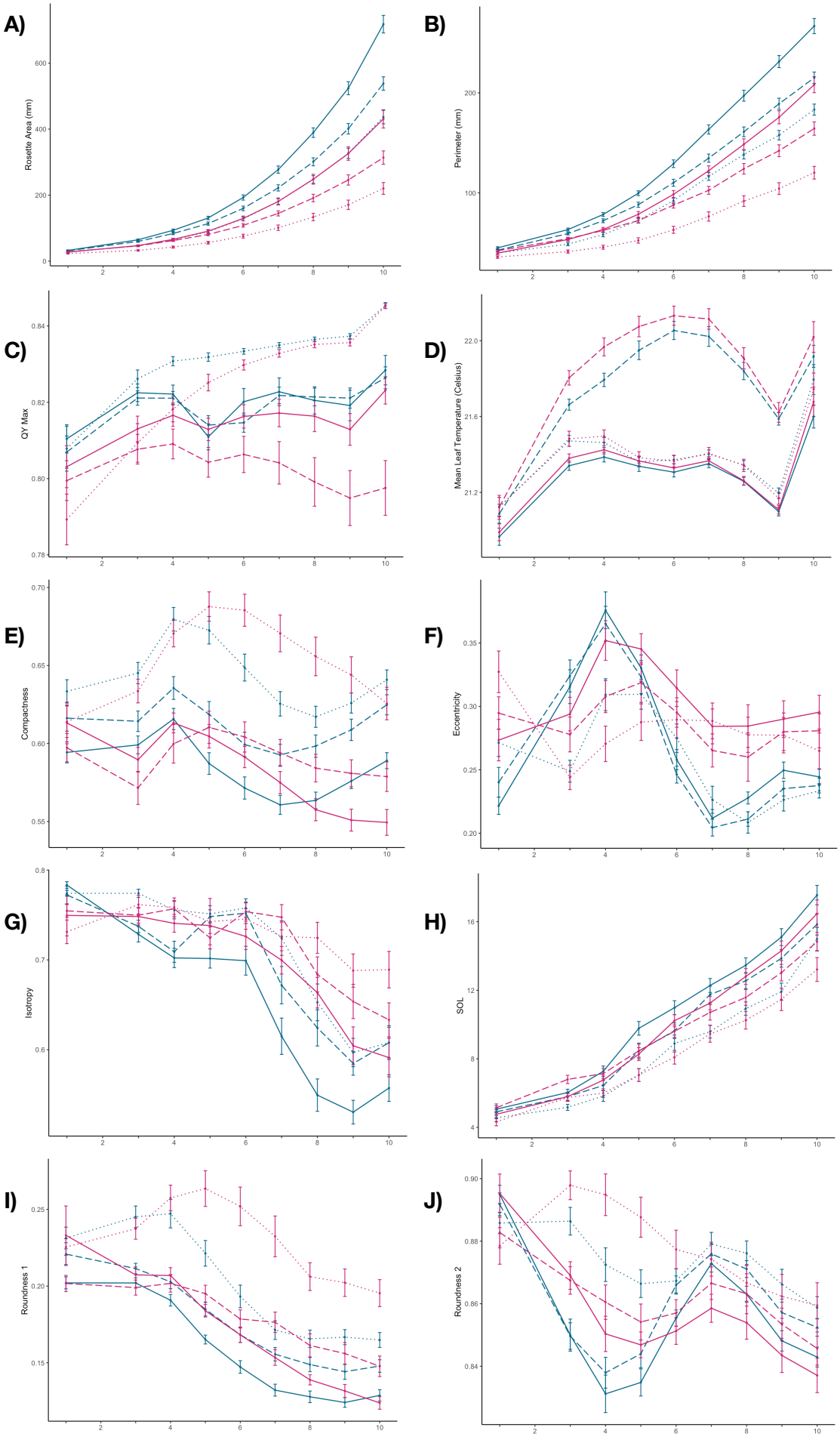
Supplementary Figure I.9. *ugt87a2*. Phenotypic traits of *ugt87a2* (pink) versus Col-0 1 (blue) over time.

Supplementary Figure I.10. *vel1*. Phenotypic traits of *vel1* (pink) versus Col-0 2 (blue) over time.

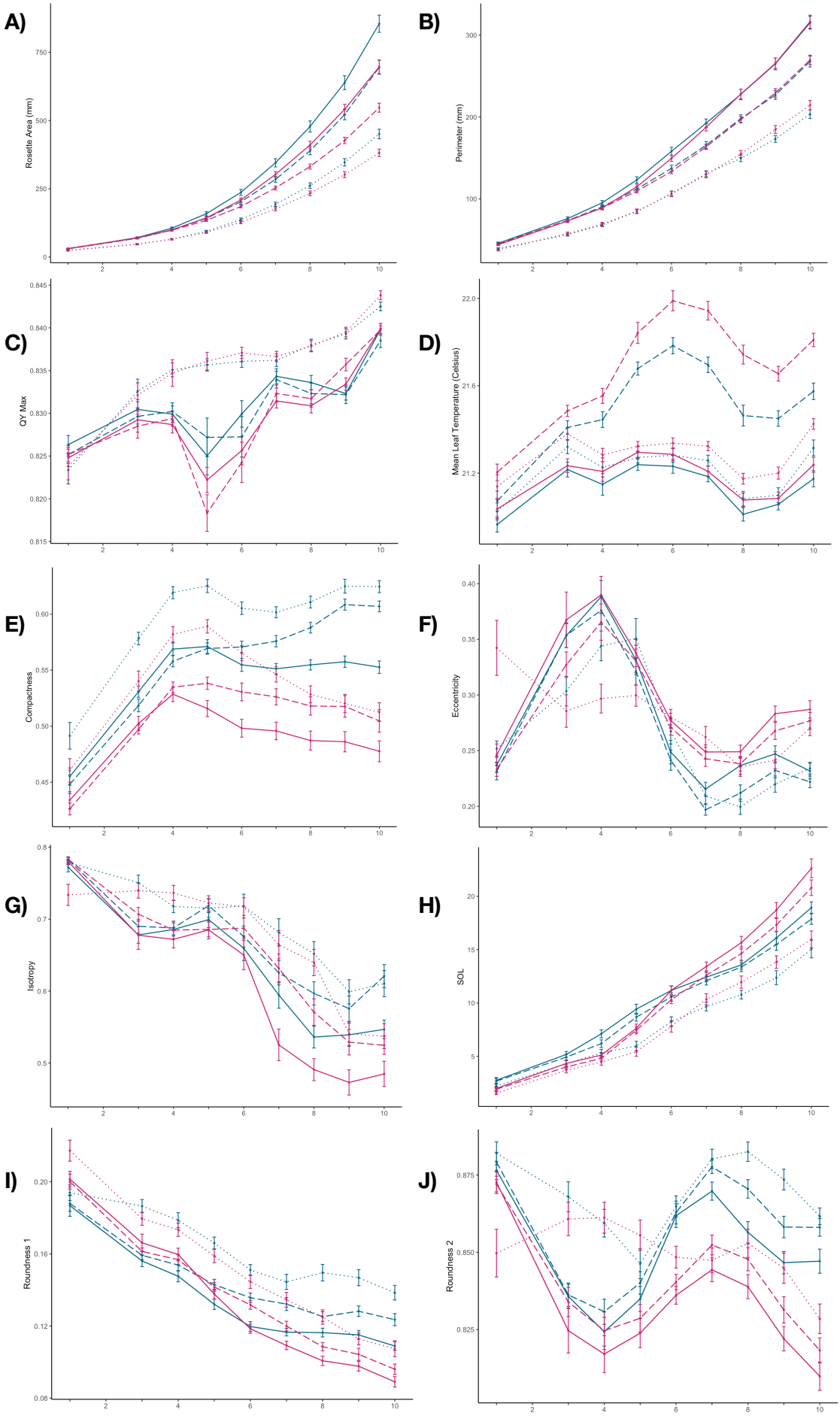
Supplementary Figure I.1. *akr4c9*



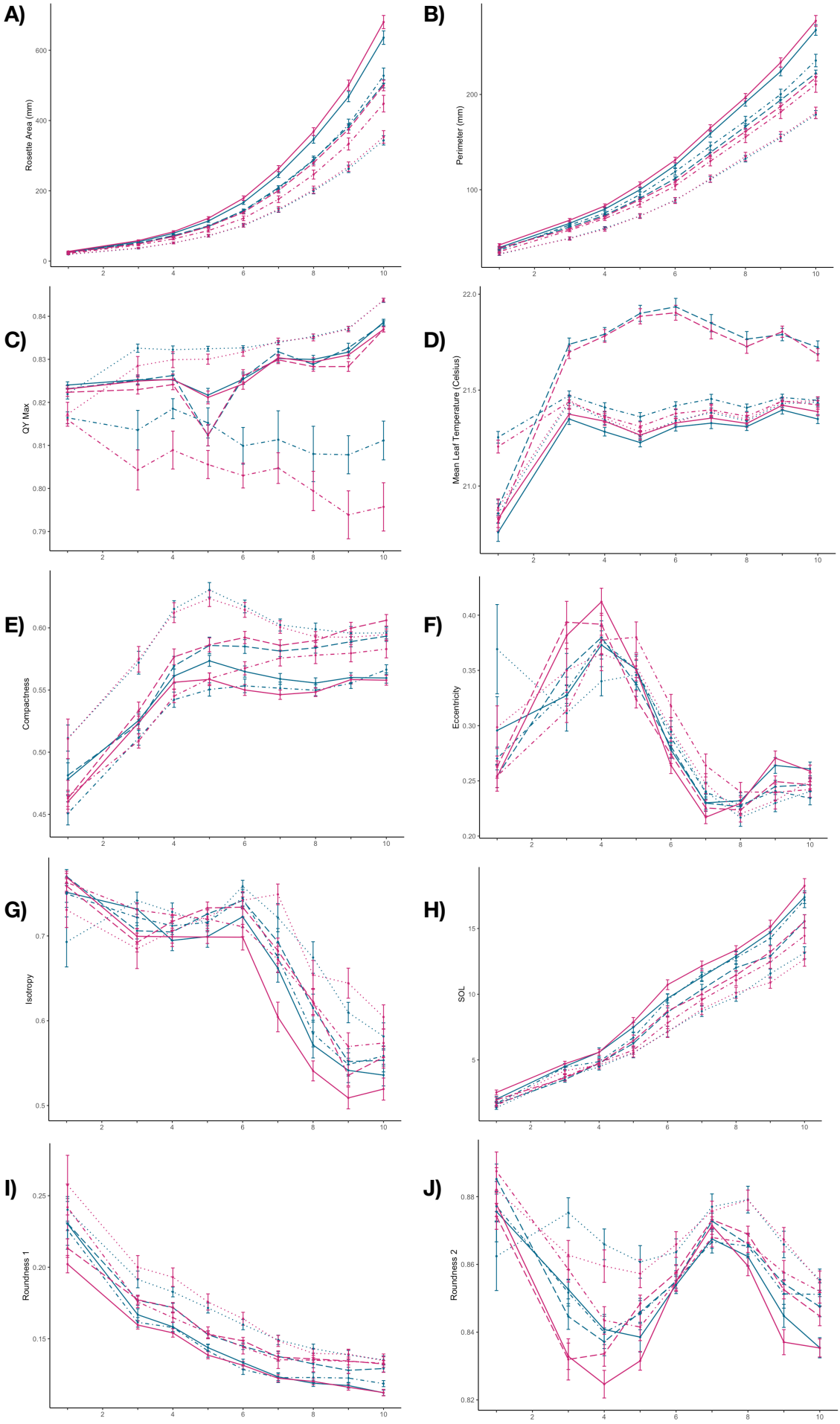
Supplementary Figure I.2. *dcl1*



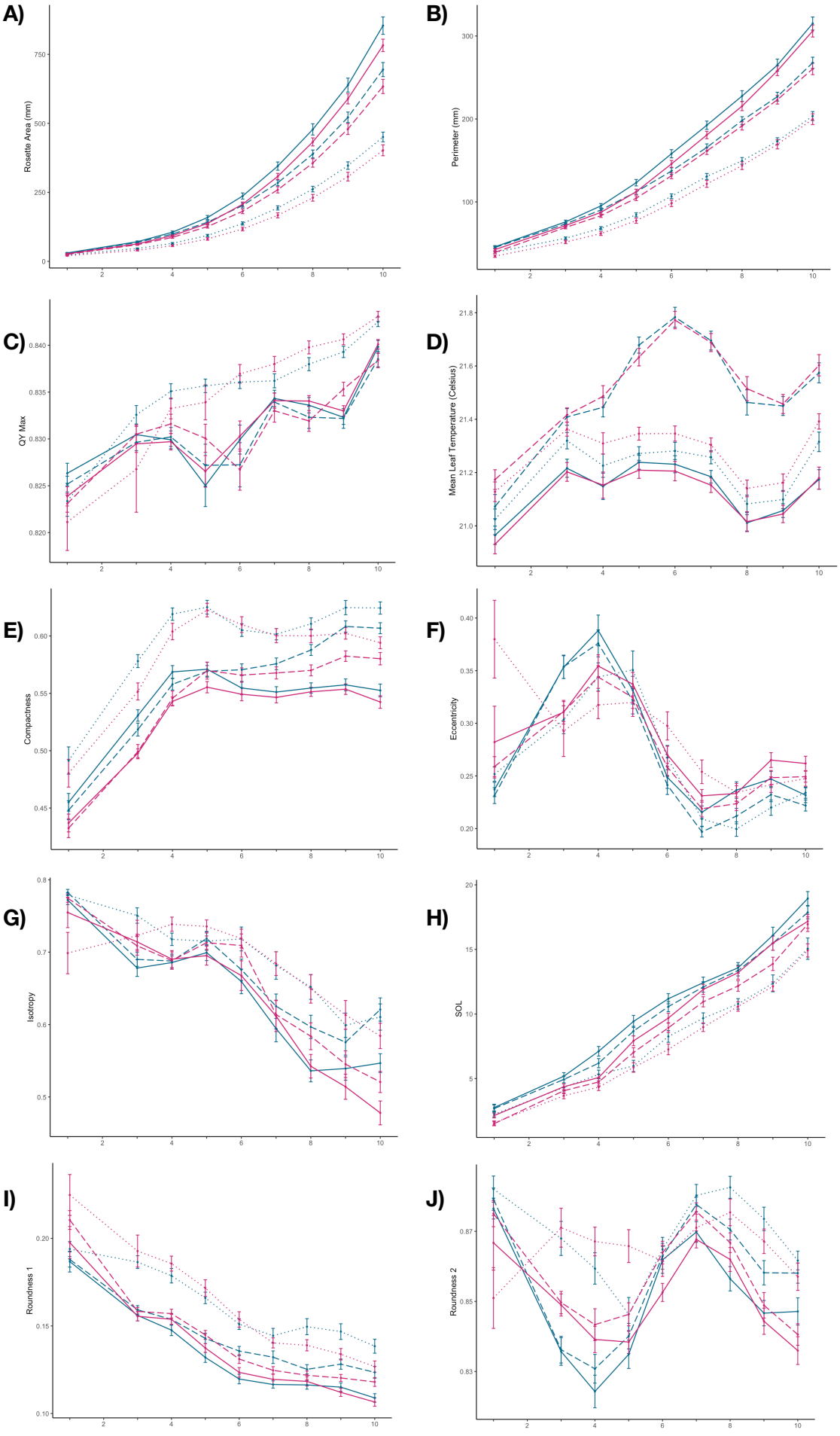
Supplementary Figure I.3. *dcl4*



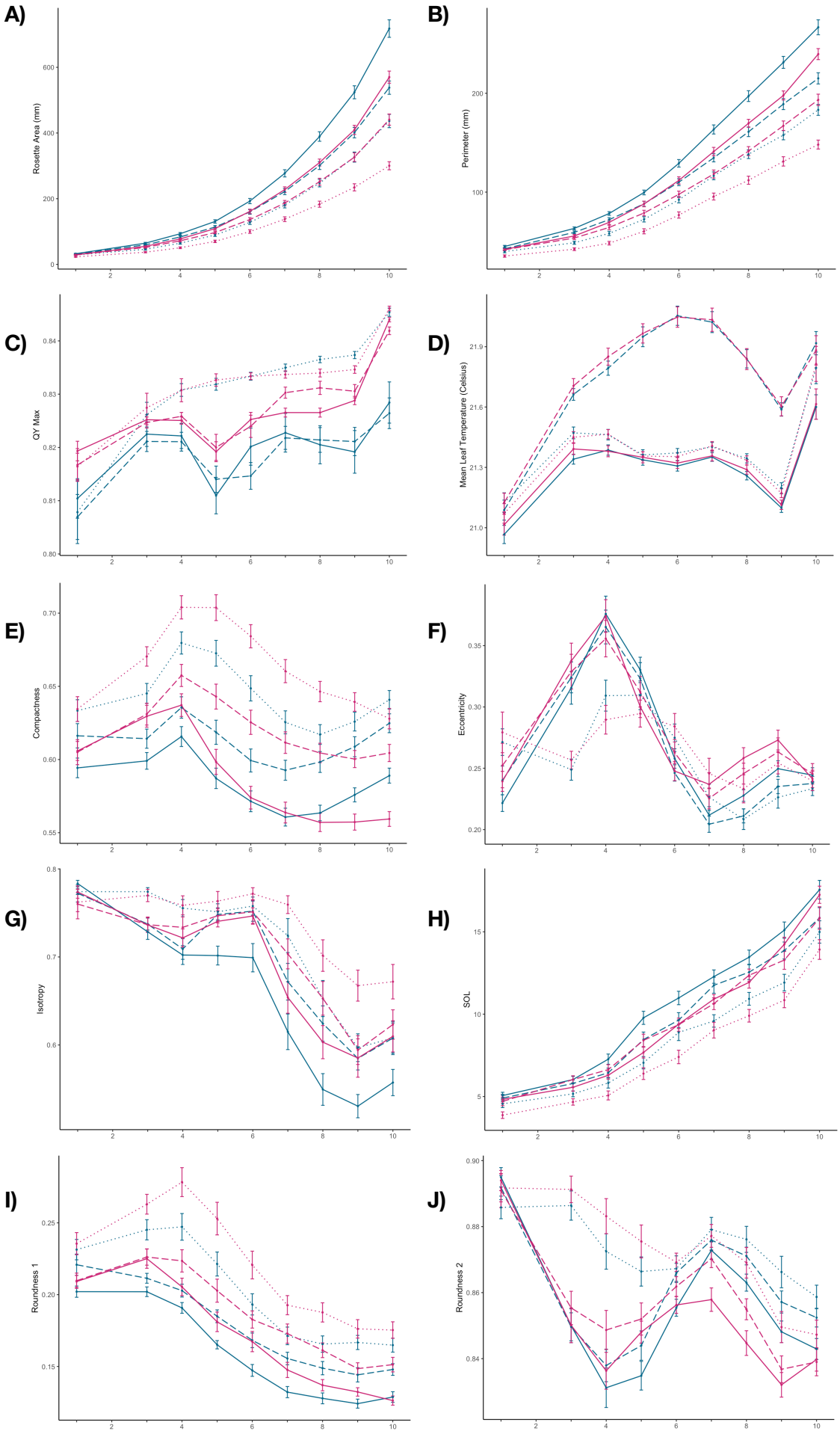
Supplementary Figure I.4. *At2-mmp*



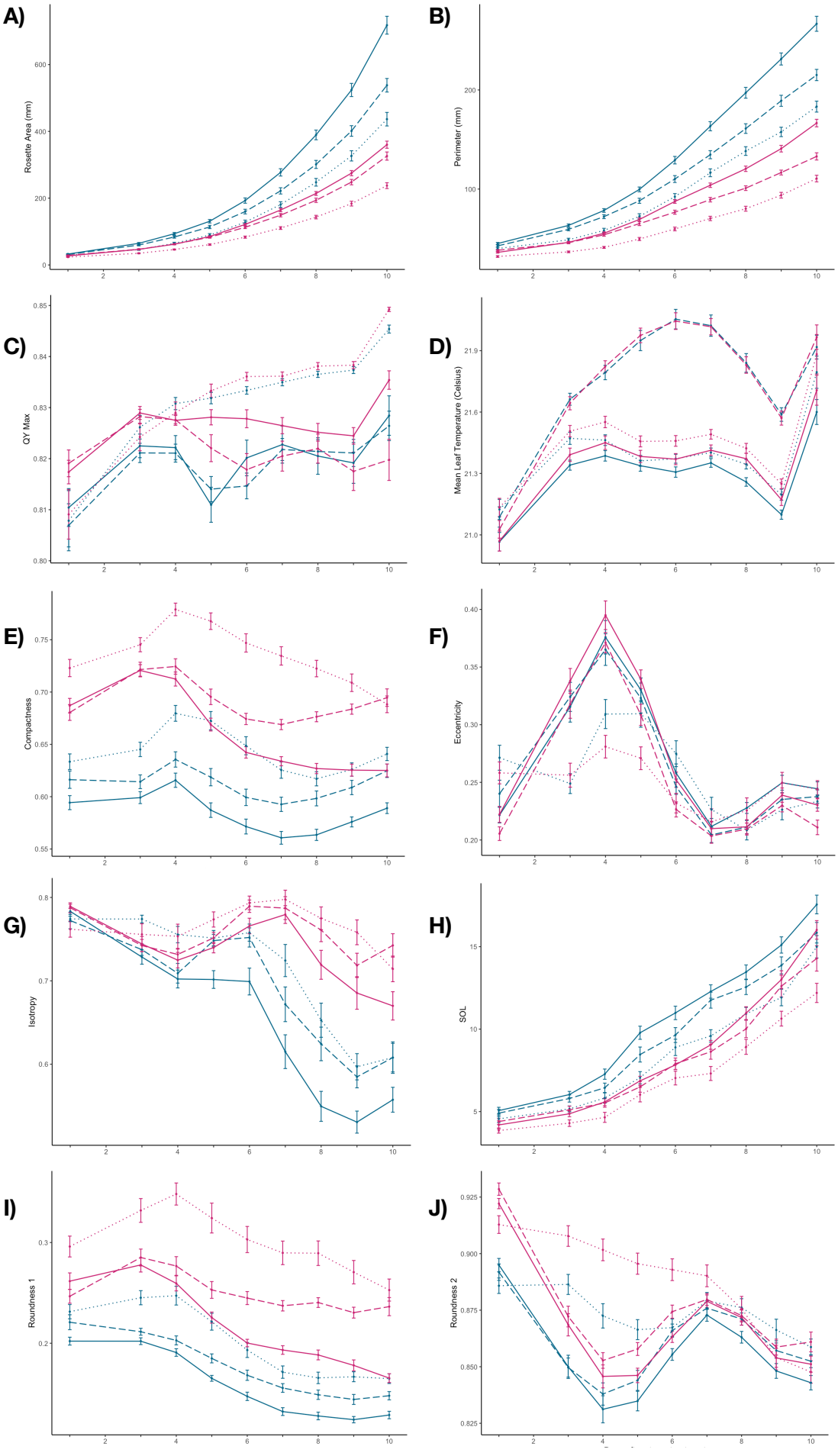
Supplementary Figure I.5. ***nac072***



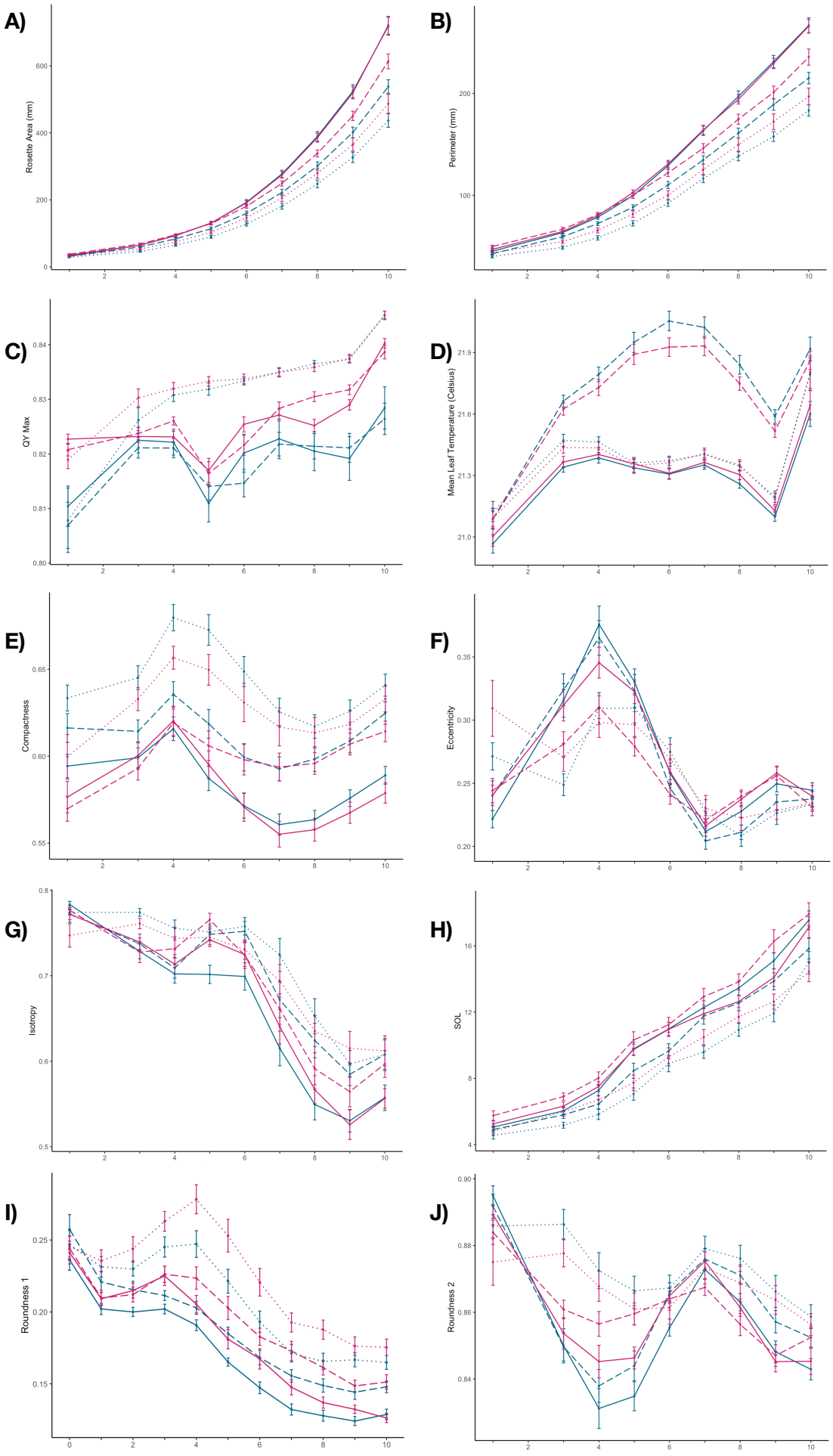
Supplementary Figure I.6. ***rcd1***



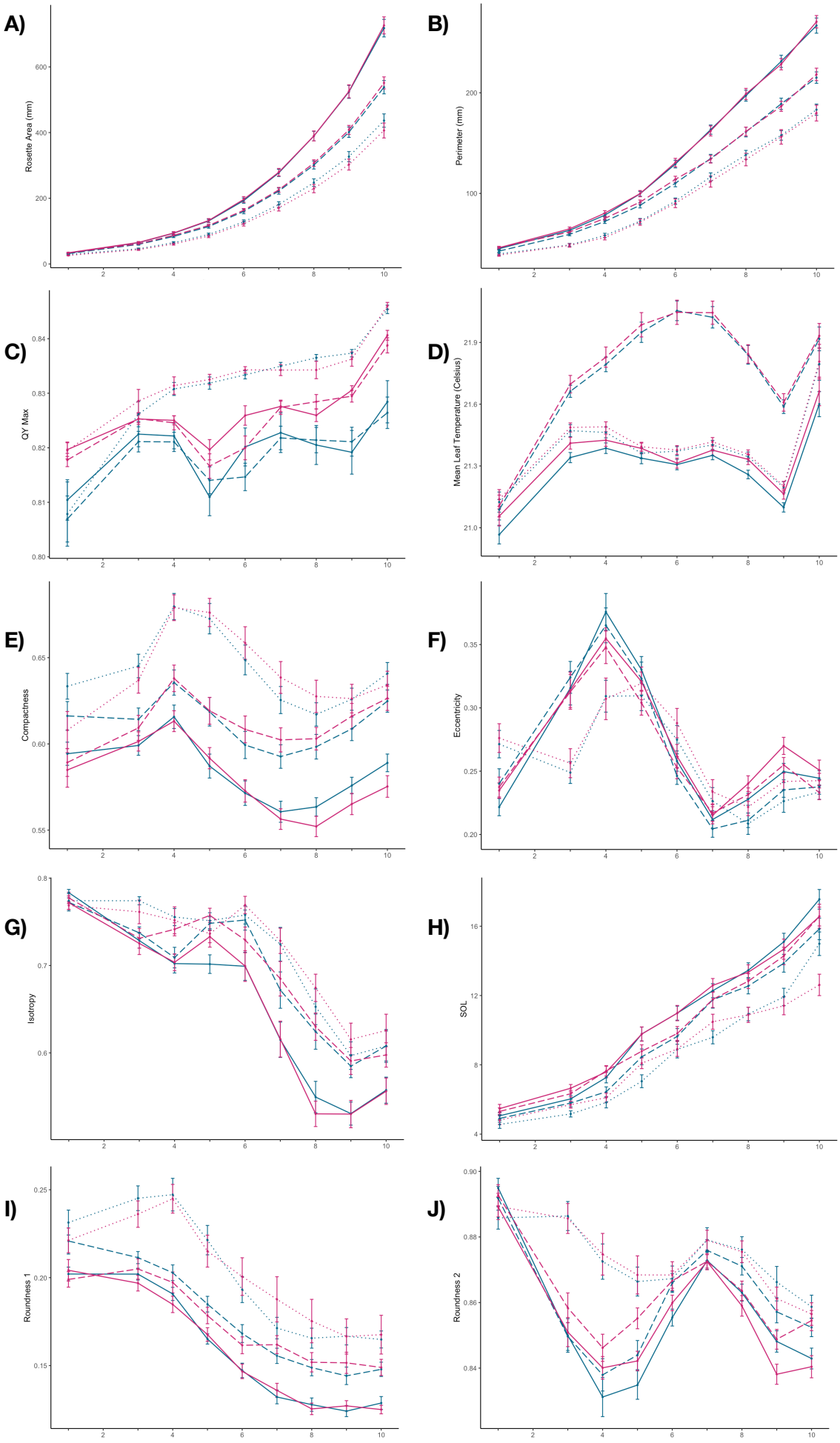
Supplementary Figure I.7. *siz1*



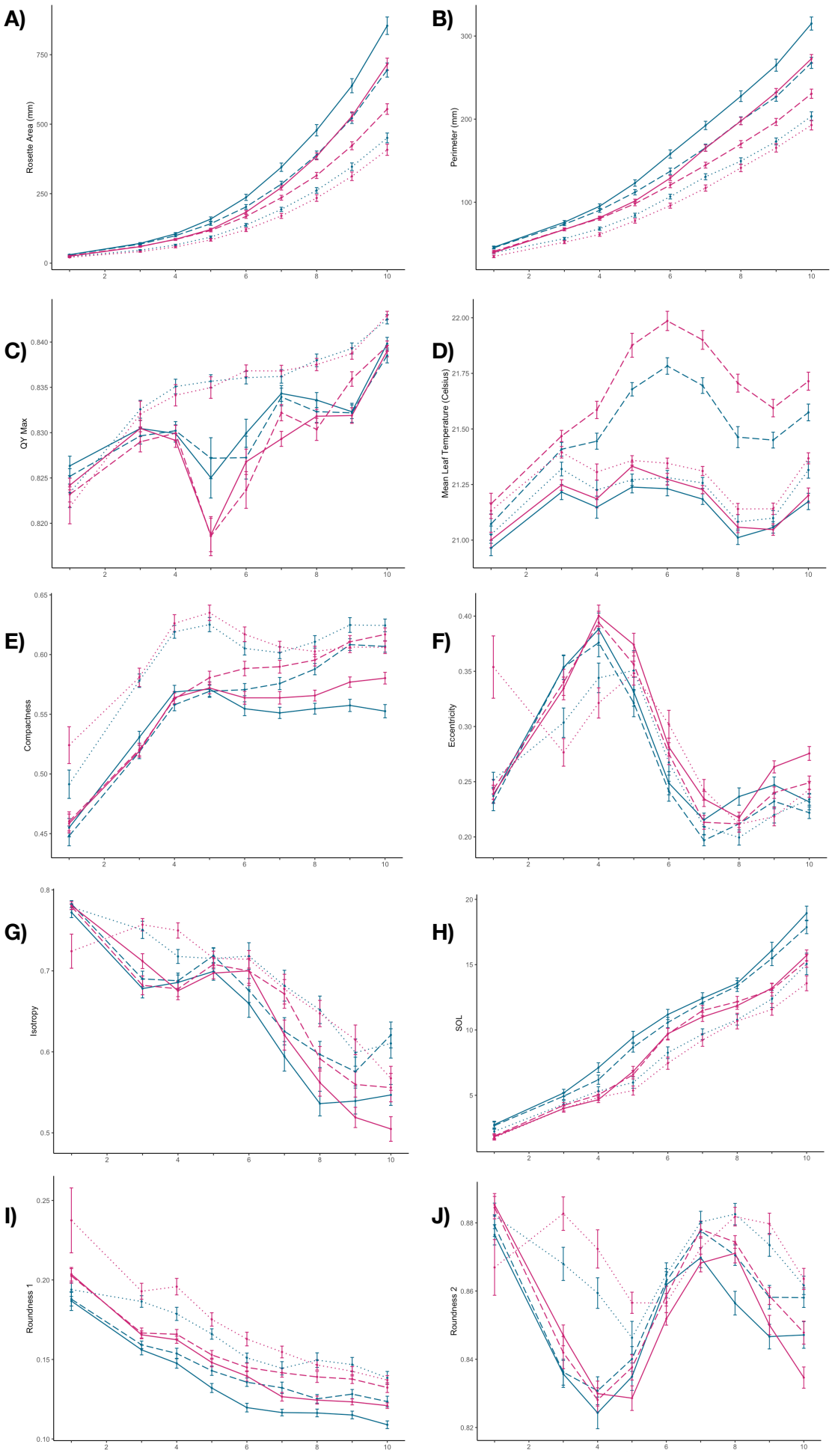
Supplementary Figure I.8. *tic*



Supplementary Figure I.9. *ugt87a2*



Supplementary Figure I.10. *vel1*



Perimeter (mm)

Genotype	Treatment	F value	P value
akr4c9	Cadmium	0.2506	0.6172
	Drought	1.8564	0.17468
	Salt	0.0097	0.9217
dcl1	Drought	0.2987	0.5856
	Salt	7.2069	**0.008277 ****
dcl4	Drought	0.0011	0.9741
	Salt	1.9149	0.1685
At2-mmp	Cadmium	4.9768	**0.02686 ***
	Drought	2.9146	0.08946
	Salt	0.6254	0.4300
nac072	Drought	0.0236	0.8780
	Salt	0.0815	0.7757
rcd1	Drought	0.0394	0.842991
	Salt	4.0459	**0.04653 ***
sal1	Drought	5.1901	**0.0242 ***
	Salt	8.7754	**0.003542 ****
siz1	Drought	0.0233	0.8791
	Salt	0.5537	0.4583
tic	Drought	2.4262	0.12197
	Salt	0.6971	0.4054
ugt87a2	Drought	0.0029	0.9572
	Salt	1.1231	0.2914
vel1	Drought	0.2689	0.6048
	Salt	2.1617	0.1435898

Supplementary Table I.3. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Perimeter (mm) measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Compactness

Genotype	Treatment	F value	P value
akr4c9	Cadmium	2.0665	0.152291
	Drought	0.0801	0.7774
	Salt	0.0783	0.7799
dcl1	Drought	0.0029	0.957
	Salt	3.3081	0.0712645
dcl4	Drought	2.0859	0.1509
	Salt	3.9169	**0.04965 ***
At2-mmp	Cadmium	3.2180	0.0744294
	Drought	3.0002	0.08489
	Salt	0.0750	0.7845
nac072	Drought	2.436	0.120743
	Salt	2.979	0.08640
rcd1	Drought	0.7506	0.388
	Salt	2.9731	0.0872796
sal1	Drought	43.838	**6.098e-10 *** **
	Salt	64.025	**2.841e-13 *** **
siz1	Drought	5.1621	**0.02484 ***
	Salt	0.3448	0.5581
tic	Drought	0.2505	0.61766
	Salt	0.1530	0.6964
ugt87a2	Drought	1.0707	0.3029
	Salt	0.5559	0.45735
vel1	Drought	2.3151	0.1302317
	Salt	19.7187	**1.742e-05 *** **

Supplementary Table I.4. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Compactness measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Isotropy

Genotype	Treatment	F value	P value
akr4c9	Cadmium	0.5016	0.221916
	Drought	1.7911	0.18239
	Salt	0.0977	0.7549203
dcl1	Drought	0.0117	0.91419
	Salt	1.5740	0.2117996
dcl4	Drought	0.3022	0.5833149
	Salt	0.0001	0.9930153
At2-mmp	Cadmium	0.9349	0.334833
	Drought	0.8005	0.37205
	Salt	2.0407	0.1548
nac072	Drought	0.2737	0.6016311
	Salt	2.6879	0.103197
rcd1	Drought	1.2964	0.25688
	Salt	0.1396	0.7093274
sal1	Drought	16.3981	**8.133e-05 *** **
	Salt	1.8964	0.1705
siz1	Drought	0.2261	0.6352418
	Salt	0.0777	0.780923
tic	Drought	0.3019	0.583678
	Salt	0.0000	0.99806
ugt87a2	Drought	0.1039	0.747701
	Salt	0.4116	0.5222192
vel1	Drought	0.2396	0.6251590
	Salt	0.0777	0.780811

Supplementary Table I.5. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Isotropy measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Eccentricity

Genotype	Treatment	F value	P value
akr4c9	Cadmium	0.7232	0.3962
	Drought	0.9640	0.327397
	Salt	0.7214	0.3968
dcl1	Drought	0.1043	0.7472
	Salt	1.0189	0.314612
dcl4	Drought	0.0004	0.9842
	Salt	2.1136	0.1481
At2-mmp	Cadmium	0.7494	0.3877940
	Drought	0.0162	0.89887
	Salt	0.7901	0.375227
nac072	Drought	0.0304	0.8619
	Salt	1.8918	0.171038
rcd1	Drought	0.6884	0.4083
	Salt	0.3446	0.5583
sal1	Drought	0.9278	0.3369950
	Salt	0.9280	0.3369036
siz1	Drought	1.4356	0.23304
	Salt	3.6288	0.05896
tic	Drought	0.0005	0.9813
	Salt	0.4426	0.5070
ugt87a2	Drought	0.4819	0.48881
	Salt	0.1869	0.6663
vel1	Drought	1.5904	0.209164
	Salt	8.104	**0.005011 ****

Supplementary Table I.6. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Eccentricity measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

RMS

Genotype	Treatment	F value	P value
akr4c9	Cadmium	0.2921	0.5895
	Drought	1.3047	0.25476
	Salt	2.4408	0.1199
dcl1	Drought	1.2813	0.2597
	Salt	3.0003	0.08553
dcl4	Drought	0.1851	0.6677
	Salt	0.6910	0.4072
At2-mmp	Cadmium	2.0685	0.152
	Drought	0.0195	0.889023
	Salt	1.4748	0.2261
nac072	Drought	0.7301	0.394181
	Salt	0.2223	0.638006
rcd1	Drought	0.3277	0.5680532
	Salt	1.1802	0.279284
sal1	Drought	0.1433	0.7056
	Salt	0.3699	0.5440
siz1	Drought	1.0064	0.31760
	Salt	2.8565	0.09337
tic	Drought	2.3967	0.12447
	Salt	1.7143	0.19268
ugt87a2	Drought	0.0919	0.762187
	Salt	1.8291	0.178672
vel1	Drought	2.3124	0.13033
	Salt	0.0137	0.9070642

Supplementary Table I.7. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Rotational Mass Symmetry (RMS) measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

SOL

Genotype	Treatment	F value	P value
akr4c9	Cadmium	1.6611	0.19899
	Drought	0.6267	0.4295
	Salt	0.1015	0.7504
dcl1	Drought	0.0296	0.863687
	Salt	0.4089	0.52365
dcl4	Drought	0.1310	0.71794
	Salt	0.9728	0.32558
At2-mmp	Cadmium	9.5320	**0.0023145 ****
	Drought	0.3956	0.5301
	Salt	2.0766	0.1512
nac072	Drought	0.9137	0.34065
	Salt	2.6736	0.1041
rcd1	Drought	0.0060	0.93828
	Salt	0.2819	0.5963
sal1	Drought	0.0697	0.7922
	Salt	18.6765	**2.777e-05 *****
siz1	Drought	0.0770	0.781957
	Salt	1.7316	0.1905038
tic	Drought	4.3493	**0.03916 ***
	Salt	0.0002	0.9888
ugt87a2	Drought	1.8189	0.1797
	Salt	2.1065	0.148978
vel1	Drought	0.1553	0.6940
	Salt	2.1328	0.146222

Supplementary Table I.8. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Slenderness Of Leaves (SOL) measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Roundness 1

Genotype	Treatment	F value	P value
akr4c9	Cadmium	0.1876	0.66541
	Drought	0.0005	0.9821
	Salt	0.0269	0.8698
dcl1	Drought	0.3934	0.5317
	Salt	11.901	**0.0007731 *****
dcl4	Drought	0.4855	0.4870398
	Salt	1.6179	0.2054
At2-mmp	Cadmium	2.4668	0.1179
	Drought	0.8124	0.3686
	Salt	0.0027	0.9589
nac072	Drought	0.1665	0.6839
	Salt	1.7324	0.19015
rcd1	Drought	0.6325	0.4279
	Salt	1.8735	0.1737
sal1	Drought	1.2172	0.271669
	Salt	21.0042	**9.459e-06 *****
siz1	Drought	11.846	**0.0007965 *****
	Salt	7.6073	**0.006767 ****
tic	Drought	1.9299	0.1675
	Salt	0.3723	0.5429
ugt87a2	Drought	0.0302	0.8624
	Salt	0.1015	0.7505
vel1	Drought	0.5301	0.4677
	Salt	6.7913	**0.01010 ***

Supplementary Table I.9. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Roundness 1 measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

Roundness 2

Genotype	Treatment	F value	P value
akr4c9	Cadmium	2.0722	0.1517
	Drought	0.3474	0.5563
	Salt	1.6959	0.1944
dcl1	Drought	0.1029	0.74885
	Salt	0.1766	0.6750
dcl4	Drought	0.0820	0.7750
	Salt	0.3001	0.5846
At2-mmp	Cadmium	0.0041	0.9489
	Drought	0.2259	0.6351513
	Salt	0.0115	0.9145
nac072	Drought	0.7990	0.37278
	Salt	0.7925	0.37477
rcd1	Drought	2.4232	0.12207
	Salt	1.0963	0.297290
sal1	Drought	1.9438	0.16528
	Salt	0.0033	0.9542034
siz1	Drought	0.0003	0.985773
	Salt	6.4748	**0.01217 ***
tic	Drought	0.1210	0.72851
	Salt	0.5322	0.4669932
ugt87a2	Drought	0.6216	0.4319408
	Salt	0.0031	0.9557
vel1	Drought	0.1255	0.7236523
	Salt	4.8322	**0.02945 ***

Supplementary Table I.10. F and P values for ANOVA interactions tests of $\log_e(x)$ transformed Roundness 2 measurements taken 10 days after stress imposition. Values shown are comparisons between each knockout genotype and Col-0 wild type, and between control and either cadmium, drought or salt treatment.

I.1 Further Evaluation of Flowering Time in *ugt87a2*

As highlighted in Chapter Two, flowering times for *ugt87a2* mutants reported by Wang et al. (2012) show extreme differences to those identified in phenotyping experiments in this thesis, with their results highlighting notably delayed flowering times for *ugt87a2* compared to wild type plants, measured through number of days to bolting.

Phenotyping experiments in Chapter Two involved the T-DNA knockout line SALK_061574, which is one of two T-DNA lines used by Wang et al. to produce their reported results. The other is SALK_124038, and to further investigate the discrepancy, additional flowering time assays using both knockout lines and Col-0 were carried out.

24 plants of each of the three genotypes were grown under long day lengths (16 hour days, 8 hour nights) in individual pots filled with John Innes No 1 soil, in a controlled environment growth chamber at 23°C, 60% humidity, and a photon irradiance of 130 µmol m^{-2} s^{-1}. Flowering time was measured by both number of leaves at bolting, and days to bolting. ANOVA tests revealed that neither of the knockout lines differed significantly from Col-0 in either measure of flowering time.

Appendix II. Chapter Three Supplementary Materials

Supplementary Table II.1. RNA-seq libraries sample metadata. Tissues 'F', 'L', and 'R' refer flower, leaf and root tissues.

Soil	Species	Site	Tree	Tissues	Latitude	Longitude
V	*belmoreana*	Transit Hill Top	HO-BEL-2371	F, L, R	-31.53188	159.07628
V	*belmoreana*	Transit Hill Top	HO-BEL-2378	F, L	-31.53208	159.07683
V	*belmoreana*	Gower Saddle	HO-BEL-2524	L	-31.58088	159.08162
V	*belmoreana*	Gower Saddle	HO-BEL-2529	F, L	-31.58092	159.08167
V	*belmoreana*	Far Flats	HO-BEL-2536	F, L	-31.565	159.07587
V	*belmoreana*	Intersection Rocky Run	HO-BEL-2603	F	-31.55563	159.09278
V	*belmoreana*	Mutton Bird Point	HO-BEL-2621	F, L, R	-31.54933	159.09253
V	*belmoreana*	Mutton Bird Point	HO-BEL-2623	L, R	-31.54908	159.09253
V	*belmoreana*	Mutton Bird Point	HO-BEL-2637	F, L	-31.54517	159.0858
V	*belmoreana*	Far Flats	HO-BEL-3329	F, L, R	-31.56572	159.07588
V	*belmoreana*	Mutton Bird track	HO-BEL-4704	F, L, R	-31.54497	159.08542
V	*belmoreana*	Mutton Bird track	HO-BEL-4705	F, L	-31.54497	159.08542
V	*belmoreana*	Barrow Flats	HO-BEL-4716	F, L	-31.57131	159.07717
V	*belmoreana*	Little Island	HO-BEL-4725	F, L	-31.56831	159.07636
V	*belmoreana*	Little Island	HO-BEL-5091	F, L. R	-31.56842	159.076
V	*belmoreana*	Little Island	HO-BEL-5093	F	-31.56868	159.07603
V	*belmoreana*	Smoking Tree ridge	HO-BEL-5100	F, R	-31.55722	159.08645
V	*belmoreana*	Grey Face	HO-BEL-5106	L	-31.56182	159.08447
V	*belmoreana*	Grey Face	HO-BEL-5107	F, L, R	-31.56178	159.08447
C	*forsteriana*	Transit Hill down	HO-FOS-2362	F, L	-31.52833	159.07033
C	*forsteriana*	Transit hill Middle	HO-FOS-2386	R	-31.53122	159.07538
C	*forsteriana*	Transit hill Middle	HO-FOS-2388	F, L, R	-31.53087	159.07488
V	*forsteriana*	Far Flats	HO-FOS-2539	F, L, R	-31.56497	159.07595
C	*forsteriana*	Neds Beach	HO-FOS-2560	L	-31.51905	159.06422
C	*forsteriana*	Neds Beach	HO-FOS-2564	F, R	-31.51933	159.06505
C	*forsteriana*	Neds Beach	HO-FOS-2566	F, L, R	-31.51942	159.065
V	*forsteriana*	Boat Harbour	HO-FOS-2605	R	-31.55958	159.09833
V	*forsteriana*	Boat Harbour	HO-FOS-2612	F, L	-31.5595	159.0977
V	*forsteriana*	Mutton Bird Point	HO-FOS-2628	F, L	-31.5488	159.09217

C	*forsteriana*	Hunter Bay	HO-FOS-4691	F, L, R	-31.5205	159.05806
C	*forsteriana*	Hunter Bay	HO-FOS-4692	F, L, R	-31.52047	159.05806
C	*forsteriana*	Blinky Beach	HO-FOS-4694	F, L	-31.53869	159.0805
C	*forsteriana*	Transit Hill down	HO-FOS-4698	F, L	-31.52792	159.07061
V	*forsteriana*	Grey Face	HO-FOS-5101	F, L, R	-31.56202	159.08453
V	*forsteriana*	Grey Face	HO-FOS-5102	F, L, R	-31.56203	159.08417
V	*forsteriana*	Grey Face	HO-FOS-5104	F, L, R	-31.56198	159.08492
V	F1	Far Flats	HO-XF1-2530	F, L, R	-31.56497	159.07585
V	F1	Far Flats	HO-XF1-2541	R	-31.56497	159.07585
V	F1	Far Flats	HO-XF1-5095	R	-31.56497	159.07585
V	F1	Salmon Beach	HO-XF1-5109	F, L	-31.5585	159.0765
V	F1	Airport	HO-XF1-5119	R	-31.53586	159.07219
V	F1	Mutton Bird Track	HO-XF1-4706	F, L	-31.54511	159.08558
V	F1	Grey Face	HO-XF1-4722	R	-31.56213	159.08583
V	F1	Mutton Bird Point	HO-XF1-4964	F, L, R	-31.54886	159.09197

Supplementary Table II.2. Final numbers of RNA-seq libraries for each species group and tissue type.

Tissue	*H. belmoreana*	*H. forsteriana* (volcanic)	*H. forsteriana* (calcareous)	F1
Root	8	5	6	6
Flower	16	6	8	4
Leaf	16	6	8	4

Supplementary Table II.3. Chi-squared test statistics for each tissue and species group comparison, testing whether the percentage of genes up or down-regulated in each species group within the comparison differs significantly from 50%. Includes Bonferroni correction for multiple testing.

Tissue	Comparison	X^2	p-value	p-value adjusted
Root	Bvol vs Fvol	0.5108647	0.4747648	1
	Bvol vs Fcal	0.3965827	0.5288594	1
	Bvol vs F1vol	362.9522282	6.41E-81	1.03E-79
	Fvol vs F1vol	446.408377	4.36E-99	6.98E-98
	Fcal vs F1vol	230.4372367	4.79E-52	7.66E-51
Flower	Bvol vs Fvol	129.1653581	6.24E-30	9.98E-29
	Bvol vs Fcal	2.1304348	0.1443998	1
	Bvol vs F1vol	594.1028571	3.21E-131	0
	Fvol vs F1vol	224.3878687	9.98E-51	1.60E-49
	Fcal vs F1vol	559.1445448	1.29E-123	0
Leaf	Bvol vs Fvol	290.7780822	3.37E-65	5.38E-64
	Bvol vs Fcal	133.0152091	8.97E-31	1.44E-29
	Bvol vs F1vol	39.9897768	2.55E-10	4.08E-09
	Fvol vs F1vol	514.0172027	8.48E-114	0
	Fcal vs F1vol	14.8998694	0.000113371	0.001813933
	Fvol vs Fcal	81.4156665	1.83E-19	2.93E-18

Appendix III. Chapter Four Supplementary Materials

Supplementary Table III.1. Sample metadata for 18s AMF metabarcoding libraries.

Soil	Species	Site	Tree	Latitude	Longitude
V	*belmoreana*	Gower Saddle	HO-BEL-2529	-31.58092	159.08167
C	*forsteriana*	Neds Beach	HO-FOS-2566	-31.51942	159.065
V	*belmoreana*	Far Flats	HO-BEL-2536	-31.565	159.07587
V	*belmoreana*	Transit Hill Top	HO-BEL-2371	-31.53188	159.07628
C	*forsteriana*	Neds Beach	HO-FOS-2564	-31.51933	159.06505
C	*forsteriana*	Transit hill Middle	HO-FOS-2386	-31.53122	159.07538
V	*forsteriana*	Far Flats	HO-FOS-2539	-31.56497	159.07595
V	F1	Mutton Bird Track	HO-XF1-4706	-31.54511	159.08558
V	F1	Airport	HO-XF1-5119	-31.53586	159.07219
C	*forsteriana*	Transit Hill down	HO-FOS-4698	-31.52792	159.07061
V	*belmoreana*	Mutton Bird track	HO-BEL-4704	-31.54497	159.08542
V	*belmoreana*	Grey Face	HO-BEL-5106	-31.56182	159.08447
V	*belmoreana*	Mutton Bird track	HO-BEL-4705	-31.54497	159.08542
V	*belmoreana*	Little Island	HO-BEL-5091	-31.56842	159.076
V	*belmoreana*	Gower Saddle	HO-BEL-2524	-31.58088	159.08162
V	*forsteriana*	Grey Face	HO-FOS-5102	-31.56203	159.08417
V	F1	Far Flats	HO-XF1-5095	-31.56497	159.07585
V	F1	Grey Face	HO-XF1-4722	-31.56213	159.08583
C	*forsteriana*	Transit Hill down	HO-FOS-2363	-31.52797	159.07059
V	*forsteriana*	Mutton Bird Point	HO-FOS-2630	-31.54519	159.0854
V	*belmoreana*	Mutton Bird Point	HO-BEL-2637	-31.54517	159.0858

Supplementary Table III.2. *Howea* individuals, for each tissue, with both RNA-seq libraries and root AMF metabarcoding, used for correlation analysis of gene expression and AMF alpha diversity.

Root		**Flower**		**Leaf**	
Soil	Tree	Soil	Tree	Soil	Tree
C	HO-FOS-2566	V	HO-BEL-2529	V	HO-BEL-2529
V	HO-BEL-2371	C	HO-FOS-2566	C	HO-FOS-2566
C	HO-FOS-2564	V	HO-BEL-2536	V	HO-BEL-2536
C	HO-FOS-2386	V	HO-BEL-2371	V	HO-BEL-2371
V	HO-FOS-2539	C	HO-FOS-2564	V	HO-FOS-2539
V	HO-XF1-5119	V	HO-FOS-2539	V	HO-XF1-4706
V	HO-BEL-4704	V	HO-XF1-4706	C	HO-FOS-4698
V	HO-BEL-5091	C	HO-FOS-4698	V	HO-BEL-4704
V	HO-FOS-5102	V	HO-BEL-4704	V	HO-BEL-5106
V	HO-XF1-5095	V	HO-BEL-4705	V	HO-BEL-4705
		V	HO-BEL-5091	V	HO-BEL-5091
		V	HO-FOS-5102	V	HO-BEL-2524
		V	HO-BEL-2637	V	HO-FOS-5102
				V	HO-BEL-2637

Appendix IV. Copyright Licenses

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Feb 10, 2023

This Agreement between Mr. Matthew Coathup ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5479470527164
License date	Jan 31, 2023
Licensed content publisher	Oxford University Press
Licensed content publication	Evolution
Licensed content title	Ecological speciation in sympatric palms: 3. Genetic map reveals genomic islands underlying species divergence in *Howea*
Licensed content author	Papadopulos, Alexander S. T.; Igea, Javier
Licensed content date	Sep 1, 2019
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	PhD Thesis: Uncovering the mechanisms of speciation in the Howea palms
Publisher of your work	Imperial College London
Expected publication date	Feb 2023
Permissions cost	0.00 GBP
Value added tax	0.00 GBP
Total	0.00 GBP
Title	PhD Thesis: Uncovering the mechanisms of speciation in the Howea palms
Institution name	Imperial College London
Expected presentation date	Feb 2023
Portions	Figure 1
Requestor Location	Mr. Matthew Coathup Silwood Park, Imperial College London Ascot, SL5 7PY United Kingdom Attn: Mr. Matthew Coathup
Publisher Tax ID	GB125506730
Total	0.00 GBP

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

Questions? customercare@copyright.com.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Feb 10, 2023

This Agreement between Mr. Matthew Coathup ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5479470328530
License date	Jan 31, 2023
Licensed content publisher	Oxford University Press
Licensed content publication	Molecular Biology and Evolution
Licensed content title	Speciation in *Howea* Palms Occurred in Sympatry, Was Preceded by Ancestral Admixture, and Was Associated with Edaphic and Phenological Adaptation
Licensed content author	Osborne, Owen G; Ciezarek, Adam
Licensed content date	Jul 18, 2019
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	PhD Thesis: Uncovering the mechanisms of speciation in the Howea palms
Publisher of your work	Imperial College London
Expected publication date	Feb 2023
Permissions cost	0.00 GBP
Value added tax	0.00 GBP
Total	0.00 GBP
Title	PhD Thesis: Uncovering the mechanisms of speciation in the Howea palms
Institution name	Imperial College London
Expected presentation date	Feb 2023
Portions	Map from Figure 5
Requestor Location	Mr. Matthew Coathup Silwood Park, Imperial College London Ascot, SL5 7PY United Kingdom Attn: Mr. Matthew Coathup
Publisher Tax ID	GB125506730
Total	0.00 GBP

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

Questions? customercare@copyright.com.