University of Southampton

Faculty of Environmental and Life Sciences

School of Biological Sciences

Adaptation and Diversification in the Canary Island endemic *Descurainia* (Brassicaceae)

by

Amy Clare Jackson

ORCID ID 0000-0003-0660-0181

Thesis for the degree of Doctor of Philosophy (PhD)

January 2023

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

School of Biological Sciences

Thesis for the degree of Doctor of Philosophy (PhD)

Adaptation and Diversification in the Canary Island endemic *Descurainia* (Brassicaceae)

by

Amy Clare Jackson

Oceanic archipelagos provide striking examples of plant lineages that have radiated rapidly over steep ecological gradients, often demonstrating remarkable adaptations. Here, we investigate the evolutionary processes behind the diversification and adaptation of the Canary Island *Descurainia* (Brassicaceae), an island cousin of *Arabidopsis*. Canary Island *Descurainia* is comprised of seven species that have rapidly radiated across islands and along an elevation cline, with several species occupying high-elevation habitats. Genotyping-by-sequencing (GBS) was employed to infer the roles ecology, geographical isolation, and hybridisation have played in the diversification of these seven species. At least three geographical shifts and one ecological shift between high- and low-elevation species were identified, as well as a putative hybrid origin for one species. A high-quality reference genome was generated for *Descurainia millefolia*, with a final size of 180 megabases and an N50 of 290 k. Through comparative phylogenomic, we confirmed the placement of *Descurainia* within Lineage I of Brassicaceae. Through whole-genome re-sequencing, representing the seven species of Canary Island *Descurainia*, we compare the low-elevation species and high-elevation species to identify candidate genes that may incur adaptation to sub-alpine environments. A Gene Ontology (GO) enrichment analysis on positively selected genes (PSGs) revealed genes associated with changes in flowering times and genes likely related to adaptation to the harsher high-elevation environment (for instance, cold and UV tolerance). We further provided evidence that one of our candidate genes incurs drought tolerance through T-DNA knockouts and a greenhouse water stress experiment. Overall, using Canary Island *Descurainia* as a model, our analyses have helped to understand the evolutionary processes behind the diversification of island plants, as well as broader questions surrounding adaptation.

Table of Contents

Table of Tables

Table of Tables

Table of Figures

Table of Appendix Tables

Table of Appendix Tables

Table of Appendix Figures

Research Thesis: Declaration of Authorship

Print name: Amy Clare Jackson

Title of thesis: Adaptation and diversification of the Canary Island endemic *Descurainia* (Brassicaceae)

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: ..... Date:.....

Acknowledgements

I would first like to thank my supervisors, Mark Chapman and Mark Carine, for their guidance and support throughout my PhD. They have allowed me to pursue my interests and, with their invaluable advice and suggestions, have shaped this thesis.

Having never conducted fieldwork before my PhD, the success of my sampling in the Canary Island would have been a different story without the help of Mark Carine, whose patience and support were unwavering. The knowledge and assistance of my local collaborators, Alfredo Reyes-Betancort and Arnoldo Santos-Guerra, were indispensable and made collecting *Descurainia* a much easier task.

I am grateful for the enjoyable times I have had with other students in the Natural History Museum and the University of Southampton (NOC and Highfield). Perusing a PhD is often a lonely task, but the network of friends I have made across my four years has provided me with support and silliness to keep me going. Indeed, undertaking a PhD during the COVID-19 pandemic was unexpected and unquestionably not something I would recommend. During these dark and challenging times, a few people provided laughter and distraction from world events. Namely, I want to thank the boys at Padwell Road for adopting me during lockdowns when I was living alone. The boardgame, film and cocktail nights, D&D, and endless homemade bread made the repetition of lockdown more bearable.

A heartfelt thanks and appreciation to Joe, who has provided me with the happiest moments and picked me up throughout the lows.

I would like to thank the support of my family, my brothers Chris and Ben, as well as Dylan the Dog. Foremost, I would like to thank my parents for their unwavering support and kindness. I am forever thankful and could not have done it without you both.

Definitions and Abbreviations

BM..... British Museum

bp..... Base pair

cp ..... Chloroplast

DNA..... Deoxyribonucleic acid

ITS ..... Internal Transcribed Spacer

Kb ..... Kilobases

Mb..... Megabases

MCMC..... Markov chain Monte Carlo

Mya ..... Million years ago

Nr ..... Nuclear

PCA..... Principal component analysis

PCR..... Polymerase Chain Reaction

RNA ..... Ribonucleic acid

SNP..... Single Nucleotide Polymorphism

Sp. ..... Species

Spp. ..... Species (pl.)

T-DNA..... Transfer DNA

Chapter 1 Introduction

1.1 Summary

This thesis aims to use Canary Island endemic *Descurainia* (Brassicaceae) as a model to investigate the evolutionary processes that underlie diversification and adaptation, using Next-generation sequencing (NGS) and Whole-genome sequencing (WGS) technology along with modern computational techniques. This chapter introduces a background review of the main themes surrounding this thesis. Larger themes are discussed first, an overview of oceanic islands and the occurrences of evolutionary processes, followed by the introduction of the first model systems used in this thesis, the oceanic archipelago of the Canary Islands. The main techniques used in this thesis are discussed next, genomic datasets and the utilisation of genomics within speciation research. The introduction of our second study system, the Canary Islands endemic group *Descurainia*, includes previous studies and our justification for using *Descurainia* as a model for evolution. Lastly, the key questions we seek to address are outlined and a summary of the chapter contents.

1.2 Oceanic Islands for Studying Speciation

Ever since the observations made by Charles Darwin (1859) and Joseph Dalton Hooker (1867), oceanic islands have emerged as ideal systems for unravelling evolutionary mechanisms (Losos and Ricklefs, 2009a; Helmus et al., 2014; Patiño et al., 2017). Despite covering 3% of the Earth's land surface, islands foster 25% of the world's species richness, positioning them as hotspots of global biodiversity (Kreft et al., 2008). Notably, the exclusive richness of plants and vertebrates on oceanic islands surpasses that of the mainland by a factor of 9.5 and 8.1, respectively (Kier et al., 2009). A significant proportion of these island species have undergone rapid evolutionary transformations through *in situ* speciation, rendering islands as natural laboratories of evolution.

Oceanic archipelagos and their large array of endemic biota (i.e., native to a specific area) have been the subject of numerous evolutionary research and have made a significant contribution to our understanding of evolution and speciation, including biogeography theories (e.g., Whittaker et al., 2008, 2017; Losos and Ricklefs, 2009), speciation processes (e.g., Savolainen et al., 2006; Givnish et al., 2009; Kisel and Barraclough, 2010; Losos, 2011), and ecological interactions (e.g.,

Losos, 1992; Gillespie, 2004; Rayner et al., 2007; Kueffer et al., 2010). Numerous characteristics of these islands have made islands a focal point for evolution studies: (1) as oceanic islands are formed through volcanic activity in the seabed, the islands have never been connected to other land masses. This means repeat colonisation of oceanic islands is unlikely due to its isolation, and we can assume endemic species most likely arose through *in situ* speciation; (2) due to the oceanic barriers between islands and the continent, gene flow between individuals is significantly reduced; (3) oceanic island often exhibits considerable habitat diversity within a small geographical space; and (4) islands present discrete geographical land components with known geological ages.

1.2.1 Common Modes of Speciation on Oceanic Islands

1.2.1.1 Geographic Isolation

Species divergence by geographic isolation (allopatric speciation, i.e., without gene flow) is widely accepted as one of the most common modes of speciation on oceanic islands (Mayr, 1982; Coyne et al., 2004), a theory dating back to the modern evolution synthesis (i.e., Mayr 1942). The reduction of gene flow by physical barriers (e.g., oceans) leads to genetic divergence by genetic drift and mutation, resulting in incompatible populations (Coyne et al., 2004).

The biota of oceanic islands offers many examples of geographic isolation. Many endemics first colonised an island from the mainland and subsequently diversified due to the oceanic barrier causing a reduction in gene flow (Cherry, 1981; Juan et al., 2000; Carine et al., 2004; Bellemain and Ricklefs, 2008; García-Verdugo et al., 2019). In particular, younger islands near the continent often experience more colonisation events in comparison with remote and older archipelagos whose biota is a result of a few original colonists (Paulay, 1994). The probability of a species colonising islands from the mainland often depends on the dispersal ability, population size and environmental conditions (Bellemain and Ricklefs, 2008). The importance of geographical isolation is again seen within archipelago systems, where once colonised, taxa diversify following dispersals between islands (inter-island) into similar habitats (Losos and Ricklefs, 2009b; Kisel and Barraclough, 2010; Mayr, 2013).

Overall, it is estimated geographical isolation accounts for 22% of oceanic island diversity and is responsible for several large notable radiations, for example, the Hawaiian *Orsonwelles* spiders (Hormiga et al., 2003) and *Laupala* crickets (Mendelson and Shaw, 2005) and the monarch flycatchers from northern Melanesia (Uy et al., 2009). Higher rates are found on islands such as Ullung (88%), St. Helena (53%) and Ogasawara (53%) (Stuessy et al., 2006). The prevalence of species diversification through geographical isolation often depends on many characteristics related to the island and species, including gene flow between islands (Whittaker et al., 2008), taxa dispersal ability (Bellemain and Ricklefs, 2008), island size (Cherry, 1981; Whittaker et al., 2008; Kisel and Barraclough, 2010), and environmental characteristics of the island. For instance, inter-island dispersal tends to be the more common mode of speciation in uniform islands, i.e., islands with lower elevations and habitat heterogeneity (Stuessy et al., 2006).

1.2.1.2 Ecological Adaptation

Ecological speciation occurs through the adaptation of populations to different environments, with reproductive isolation evolving directly or indirectly through increasing incompatibility of genetic regions (Mayr, 1947; Schluter, 2000, 2009; Nosil and Schluter, 2011; Nosil, 2012). Early laboratory experiments demonstrated the occurrence of ecological speciation (reviewed in Rice and Hostert, 1993). Evidence for ecological speciation has also been extensively documented in wild species (e.g., Schluter and McPhail, 1992; Jiggins et al., 2001; McKinnon et al., 2004; Dunning et al., 2016; Meier et al., 2018).

Ecological adaptation within oceanic islands often occurs through habitat shifts (Otte and Endler, 1989), whereby species diversify following the colonisation of new and unoccupied habitats. Examples of habitat shifts are frequently exemplified in lineages that have undergone adaptive radiation. Adaptive radiation is whereby a species rapidly radiates into a wide variety of habitats by obtaining various exploitative morphological and/or physiological traits (Schluter, 2000). A study of angiosperm flora across 13 oceanic archipelagos found that adaptive radiations explained 75% of endemic species diversity (Stuessy et al., 2006). Several classic examples are Darwin's finches from the Galapagos islands (Petren et al., 2005; Almén et al., 2016; Chaves et al., 2016), *Anolis* lizards from Caribbean islands (Losos et al., 1997, 2001), Hawaiian Drosophilidae (Kambysellis et al., 1995; Magnacca and Price, 2015), and Hawaiian silverswords (Baldwin, 1997; Baldwin and Sanderson, 1998). However, many radiation events likely contain nonadaptive (i.e., geographic isolation) and adaptive diversification elements.

Several characteristics of oceanic islands have been noted to promote adaptive radiations. Larger island size is often noted for its correlation with increased potential for adaptive radiations, which is believed to be due to more available habitats and geographical features (Lomolino, 2001; Kreft et al., 2008; Kisel and Barraclough, 2010). Similarly, large elevation gradients are often exhibited on islands, which promote ecological speciation through increasing the variety of habitats, as well as isolation from lower elevations (Lomolino, 2001; Losos and Ricklefs, 2009a; Kisel and Barraclough, 2010; Steinbauer et al., 2016). Other factors include volcanic activity and landslides, which stimulate heterogeneity by continuously creating new habitats and causing extinction events leading to the opening of previously occupied niches (Otto et al., 2016). For instance, Ullung Island is an ecologically uniform and low-elevation island off the coast of Korea where only 12% of the island's endemic species exhibit adaptive evolution (Stuessy et al., 2006). In contrast, the Hawaiian and Canarian Archipelagos are rich in diverse environments with large elevations, where adaptive radiation is documented in 93% and 84% of endemic species.

1.2.1.3 Hybridisation and Introgression

Hybridisation (i.e., reproduction of individuals between species producing offspring) is a common phenomenon within plants; an estimated 25% of plant species are known to hybridise (Barton and Hewitt, 1985). In recent years, it has become apparent that hybridisation can have adaptive potential and promote diversification by serving as a source of genetic variation through novel combinations of genotypes (Grant and Grant, 1994; Seehausen, 2004; Hegarty and Hiscock, 2005; Mallet, 2007; Abbott et al., 2013). Hybridisation can give rise to new species through the recombination of parental genomes (e.g., Howarth and Baum, 2005; Rieseberg and Willis, 2007; Mavárez and Linares, 2008; White et al., 2018; Sun et al., 2020).

Hybridisation is widespread among rapidly radiating groups (Mallet, 2007; Seehausen, 2013; Meier et al., 2017; Lamichhaney et al., 2018; Svardal et al., 2020). However, studies on the frequency of hybridisation between insular endemics range from common (Mayer, 1991; Baldwin et al., 1998; Carine et al., 2007; Lamichhaney et al., 2018; Meudt et al., 2021) to rare and less frequent than on the continent (Ellstrand et al., 1996). One study suggests that polyploids (species with three or more sets of chromosomes) are common within oceanic islands (with 50%, 46% and 46% of native flora in Hawaii, New Zealand and the Galápagos Islands being polyploids, respectively; Meudt et al., 2021). Notable examples of hybridising endemic lineages include

Macaronesian *Argyranthemum* (Brochmann et al., 2000; Fjellheim et al., 2009; White et al., 2018) and the Hawaiian *Scaevola* (Howarth and Baum, 2005; Bramwell and Caujapé-Castells, 2011). Certainly, characteristics of island biota and the islands provide opportunities for hybridisation to be commonality; for instance, the few intrinsic reproductive barriers exhibited in oceanic flora (Lowrey, 1986; Mayer, 1991; Brochmann et al., 2000) and the dynamic ecological landscape of oceanic islands. The natural disturbances through volcanic activity, as well as anthropogenic disturbances (e.g., paths and roads), can bring species into contact and provide potential habitats for hybrids (van Hengstum et al., 2012; Crawford and Stuessy, 2016; Kerbs et al., 2017).

Many questions remain about the extent to which hybridisation contributes to the diversification of insular lineages. Carlquist (1966) argues that hybridisation was a significant element of the success of early radiations as it provided recombination of genetic variations, a theory supported by later studies (e.g., Koblmüller et al., 2010; Meier et al., 2017; Marques et al., 2019; Choi et al., 2021). In one study, greater levels of polyploidy directly promoted island endemic diversity through increased genetic diversity (Meudt et al., 2021). In other cases, diversification by hybridisation is often only inferred indirectly through the conflict between nuclear and chloroplast phylogenies (e.g., Grant and Grant, 1994; Francisco-Ortega et al., 1996; Goodson et al., 2006). Only recently, with the advent of more modern sequencing techniques that enable whole-genome insights into gene flow between populations and species, studies are starting to emerge which reveal hybrid speciation events in oceanic lineages, for example, in Darwin's finches (Grant et al., 2004; Lamichhaney et al., 2018) and the Hawaiian *Metrosideros* (Choi et al., 2021).

1.3 Study System: The Canary Islands

1.3.1 The Canary Islands Formation, Climate and Vegetation

The Canary Island is an oceanic archipelago located 100 km west of Morocco in the northern Atlantic. There are eight main islands – Tenerife, Gran Canaria, Fuerteventura, Lanzarote, La Palma, La Gomera and La Graciosa (As of 2023) – and several small islets. The islands were formed through volcanic activity, and the islands vary in age, with the older island located in the east (21 million years for Fuerteventura) and the youngest in the west (0.8 Mya for El Hierro; Schmincke, 1976; Figure 1.1A). The islands are at different stages of the volcanic cycle. Fuerteventura (< 807 m in elevation) and Lanzarote (< 681 m) are low-elevation plains as they are mostly eroded. Gran Canaria (< 1,950 m) and La Gomera (< 1,484 m) are beyond their maximum and are currently in the erosion phase. Tenerife (< 3,718 m), La Palma (< 2,426 m), and El Hierro (< 1,501 m) represent the taller islands which are still growing through volcanic activity (Schmincke, 1976; Kunkel, 2012). Tenerife first existed as three separate islands, dating back to the Late Miocene: Roque del Conde (11.0-8.9 Ma) in the southwest, Teno (6.2-5.6 Ma) in the northwest, and Anaga (4.9-3.9 Ma) in the northeast. Volcanic episodes fused these islands 3.5 Mya and gave rise to the present island of Tenerife.

A typical Mediterranean climate dominates in the Canary Islands, with key features such as low seasonal variation in temperatures and low precipitation, primarily confined to winter, with summer drought (Cropper and Hanna, 2014; Del Arco and Rodríguez-Delgado, 2018). However, each island exhibits different climatic variations due to their elevation and topography, coupled with the trade winds, which play a crucial role in delimiting climatic zones.

The Canary Islands are subjected to the cool and humid northeast (NE) trade winds, which create a cloud layer at mid-elevations on the northern slopes, particularly on the taller islands (i.e., Tenerife and La Palma; del Arco Aguilar and Rodríguez Delgado, 2018; Garzón-Machado et al., 2014). As a result, the north side of the islands, between 400 and 1,200 m, is humid, whereas the southern slopes are warmer and drier. The NE trade winds do not generate a cloud layer on the slopes of Lanzarote and Fuerteventura (excluding Jandía mountain, with an elevation of 807 m) due to their low elevation, resulting in arid islands (Bramwell, 1990).

The NE trade winds, in combination with the elevational range of the islands, have produced several distinct ecological zones (Bramwell, 1976; del-Arco et al., 2006; del Arco Aguilar and Rodríguez Delgado, 2018). From low- to high-elevation, they are sub-desert coastal scrub; thermophilous woodlands; laurel forest (on northern slopes); pine forest; summit scrub (Figure 1.1B & C).

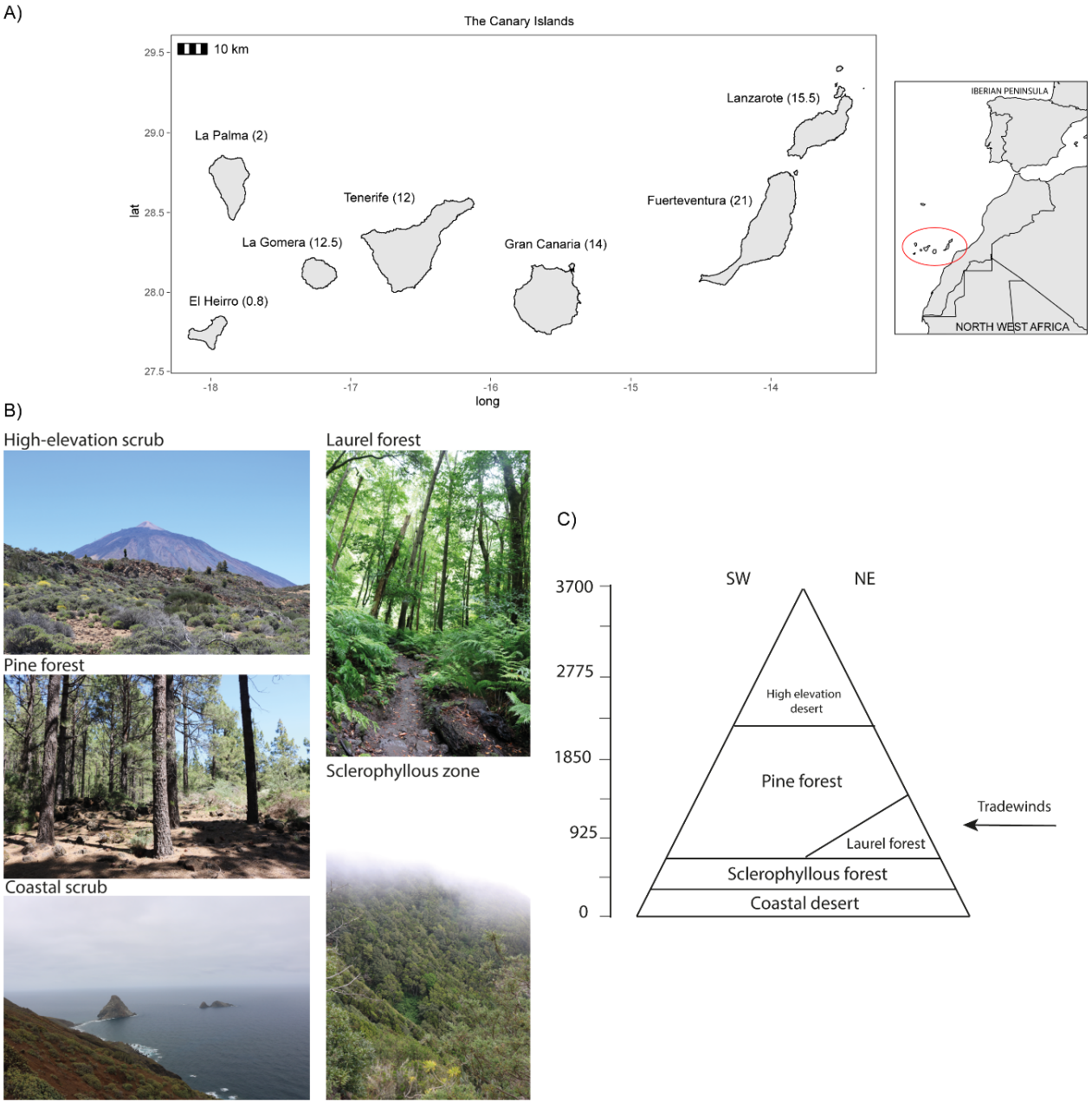


Figure 1.1: A) A Map of the Canary Islands with the ages of the islands in brackets Image on the right shows the location of the Canary Islands relative to North Africa and Iberia (Schmincke, 1976). B) Example pictures of the five main ecological zones on the Canary Islands along the elevational cline (m); C) Schematic of the ecological zone distributed along the elevational gradient.

The low-elevation habitats found on the Canary Islands are often represented below the cloud layer, between 200 and 1,000 m, which include the coastal scrub and thermophilus/sclerophyllous woodlands. The coastal scrub habitats are primarily xerophytic scrublands composed of succulent, shrubby and woody plants, which comprise the largest proportion of the Canary Island's vegetation, ca. 51.5% (Fernández-Palacios et al., 2021). Lanzarote and Fuerteventura, the lowest elevation islands, are almost entirely composed of *Euphoria* scrub. Most species are immigrants from low-elevation coastal plains of North and West-African regions or more recent colonizers (e.g., *Euphorbia canariensis*, *E. balsamifera*, *E. obtusifolia*; Euphorbiaceae; Fernández-Palacios and de Nicolás, 1995). The climate here remains relatively constant throughout the year, with an average mean temperature between 13 – 15°C and annual precipitation > 700 m, and minimal frost or drought (del Arco Aguilar and Rodríguez Delgado, 2018). Anthropogenic influences are very high in this zone, with agriculture (e.g., bananas and tomatoes) and dense settlements spread over the landscape.

Above the coastal plains is the thermophilus woodlands, which primarily consist of perennial herbaceous communities. Tree species found here include *Juniperus turbinata subsp. canariensis*, *Phoenix canariensis* and shrubs such as *Retama rhodorhizoides* and *Salvia canariensis*. Similarly, with the coastal zone, settlement and agriculture have meant only remnants of the natural vegetation remain.

A temperature inversion heavily influences the environmental differentiation on the taller islands at ca. 1,200 m asl (Bramwell, 1976; del-Arco et al., 2006; del Arco Aguilar and Rodríguez Delgado, 2018). Due to the inversion, the rising moist air carried by the prevailing NE trade winds leads to the accumulation of clouds. This 'cloud layer' subsequently decreases radiation and results in a higher moisture level. A hyper-humid evergreen laurel forest is restricted to the cloud layer on the northern slopes of the western Canaries (ca. 600 – 1,500 m). The laurel forest has a constant annual temperature between 14.5 to 16.5°C and a mean yearly rainfall of 700 m (Morales et al., 1996). This habitat is rich in endemic flora, such as *Digitalis canariensis* (Plantaginaceae) and *Laurus novocanariensis* (Lauraceae; Bramwell, 1976).

Above the cloud layer is the pine forest belt (ca. 1,500 – 2,000 m), a woodland composed of predominately *Pinus canariensis* (Pinaceae), with mean annual temperatures between 14.5 and 15.5°C and annual rainfall between 400-650 mm. We find the sub-alpine zone at the end of the gradient, above the pine forests.

Sub-alpine high-elevation habitats are distributed in Tenerife and La Palma, ca, 1,800 – 2,300 m. Temperatures here are relatively low, with an average annual temperature of 6-11°C, but they also experience significant diurnal changes. For instance, Las Cañadas (Tenerife) can vary by as much as 25°C per day (Bramwell and Caujapé-Castells, 2011). There are also frost periods which extend from October to May, as well as snow in the winter (del-Arco et al., 2006). In contrast to most continental mountains, rainfall in Tenerife decreases with elevations above 2,000 m due to the cloud layer (Fernández-Palacios and de Nicolás, 1995). As a result, rainfall here is comparatively low, between 350 – 500 mm annually. The porous volcanic soil and dry winds further reduce water availability. The radiation is also very intense (sunshine > 3,700 h/year) with low humidity (< 30%). The flora here is relatively species-poor but highly specialised, with many exhibiting adaptations consistent with high-altitude adaptation (Steinbauer et al., 2012; Fernández-Palacios et al., 2014; Irl et al., 2015, 2017). Typical flora endemic to these sub-alpine habitats include *Descurainia bourgaeana* (Figure 1.2)*, Echium wildpretii* and *Argyranthemum teneriffae.*

Figure 1.2: Picture of El Teide, Tenerife (ca. 2,000 m) sub-alpine landscape with *Descurainia bourgaeana* at the forefront.

1.3.2 Why Study Canary Island Diversity

Flora and fauna of the Canary Islands have been the subject of numerous studies in recent years. The complex topography, combined with diverse habitats along an elevational gradient (ca. < 3,750 m), has promoted high levels of diversification (Bramwell, 1976; Francisco-Ortega et al., 2000; Francisco-Ortega et al., 2000; Reyes-Betancort et al., 2008; Caujapé-Castells et al., 2017). In many instances, the species arising from these diversification processes are restricted to a single island and distributed within strict ecological zones due to *in situ* speciation (Whittaker et al., 2008). The spatial distribution of the endemic lineages can provide valuable information on the drivers of diversification and, thus, evolutionary processors (Richardson and Whittaker, 2010).

Some of the most notable cases of diversification can be found within Canarian flora, with approximately 630 plant species endemic to the islands, representing 50% of total flora (Francisco-Ortega et al., 2000). This is exemplified by the numerous large plant lineages, most notably, *Aeonium* alliance (Crassulaceae; ca. 61 species) (Jorgensen and Frydenberg, 1999; Brilhante et al., 2021), *Echium* (27 species) (Graham et al., 2021), *Sideritis* (Lamiaceae; 23 species) (Barber et al., 2000, 2002), *Crambe* (Brassicaceae; 14 species) (Francisco-Ortega et al., 1999, 2002), and *Sonchus* alliance (Asteraceae; ca. 31 species) (Caujapé-Castells et al., 2017). Many of these radiations present striking morphological adaptations that differ significantly from their continental relatives. For instance, insular woodiness (i.e., the transition from herbaceousness to woodiness for colonised plants) is a common trait among oceanic endemics, as seen in Canarian *Sonchus* and *Echium* alliances (Panero et al., 1999; van Huysduynen et al., 2020). In addition, particular habitats on the Canary Islands impose strict ecological constraints (such as the laurel forest and sub-alpine desert), resulting in specialised adaptations among plant groups (Steinbauer et al., 2012; Fernández-Palacios et al., 2014).

1.3.3 Evolutionary patterns driving species richness and endemism in the Canary Islands

It is estimated that most Canary Island endemic lineages have derived from a single colonisation event, revealed through the prevalence of strongly supported monophyletic lineages (Silvertown et al., 2005), usually from an African or Mediterranean relative (Carine et al., 2004; Vargas, 2007; García-Verdugo et al., 2019). As seen in, for example, *Sideritis* (Barber et al., 2000, 2002), *Bencomia* (Rosaceae; Helfgott et al., 2000), *Sonchus* (Kim et al., 1996), *Gonospermum* (Asteraceae; Francisco-Ortega et al., 2001), *Pericallis* (Asteraceae; Panero et al., 1999; Jones et al., 2014), and

Crambe (Francisco-Ortega et al., 1999, 2002). Although less common, congeneric colonisations into the Canary Islands have been demonstrated in several studies (Carine et al., 2004; Meimberg et al., 2006; Puppo et al., 2015).

Several evolutionary processes are hypothesised to underlie the rapid diversification of insular lineages following colonisation. Geographic isolation (i.e., speciation through isolation following inter-island dispersal between similar ecological habitats) and habitat shifts (i.e., speciation associated with the shifts of lineages to different ecological niches), are the most prominent processors used to explain the high levels of diversity within the Canary Island (and other oceanic systems). In most cases, endemic radiations involve a relative combination of the two. Geographically isolation is noted as the primary mode of diversification in, for example, *Pericallis* (Jones et al., 2014), *Lotus* (Fabaceae; Allan et al., 2004) and *Rumex* (Polygonaceae; Talavera et al., 2013). Albeit rare, there are examples of niche conservatism, which can be described as the tendency for a groups ecological preferences to remain unchained through time and space (Wiens and Graham, 2005), as seen in the Canary Island *Helianthemum* (Cistaceae; Albaladejo et al., 2021). In contrast, habitat shifts were cited as the most important evolutionary processes in the diversification of Canary Island *Sonchus* (Kim et al., 1999; Lee et al., 2005) and *Aeonium* (Jorgensen and Olesen, 2001; Jorgensen, 2002; Brilhante et al., 2021).

Canary Island flora also exhibit high rates of adaptive radiation, with 84% of endemic species displaying patterns of adaptive radiation (Stuessy et al., 2006). Impressive examples that occupy a variety of habitats and demonstrate a range of phenotypic variations include Canary Island *Lotus* (Allan et al., 2004), *Aeonium* (Jorgensen and Frydenberg, 1999; Mort et al., 2002), *Crambe* (Francisco-Ortega et al., 2002), *Pericallis* (Jones et al., 2014), *Sonchus* (Kim et al., 1996) and *Tolpis* (Asteraceae-Lactuceae; Crawford et al., 2015; Mort et al., 2022). The prevalence of adaptive radiation in the Canaries is often explained by the extent of ecological opportunities (Schluter, 2000). Within the Canary Islands, the climatic niche heterogeneity (mainly temperature and rainfall seasonality variables) has widely been used to explain habitat shifts in several lineages (Irl et al., 2015; Halbritter et al., 2018; Brilhante et al., 2021; Fernández-Palacios et al., 2021), as well as a likely explanation for the high levels of endemicity and functional diversity (Irl et al., 2015, 2017; Hanz et al., 2022). For example, at mid-elevations in the cloud layer, the precipitation levels are greatest and constant year-round, prompting the laurel forest (restricted to the cloud layer) as a hotspot for endemism (Steinbauer et al., 2012; Fernández-Palacios et al., 2014).

Recently, the importance of elevation for speciation and driving endemism has been identified within the Canary Islands (Steinbauer and Beierkuhnlein, 2010; Steinbauer et al., 2012, 2013; Irl, 2016). High-elevational clines represent large geographical distances and provide habitat heterogeneity, and thus it is postulated that speciation increases with elevation due to increasing environmental and geographical isolation. Indeed, 31.7% and 34.9% of plant species distributed in the Tenerife and La Palma sub-alpine zones, respectively, are endemics (Steinbauer et al., 2013, 2016). In addition, most species colonising the islands are adapted to coastal or low-elevation habitats, with few species pre-adapted to tolerate the harsh conditions found at high elevations (i.e., the sub-alpine zone; Wilson, 1961). This is because the nearest coastal region to the most easterly islands, e.g., Fuerteventura, is Morocco, ca. 65 km. The closest comparable elevations to the taller islands of the Canaries (Tenerife and La Palma; < 3,750 m) are the High Atlas Mountains, located ca. 800 km away. Thus, it is believed that the colonisation of high-elevation habitats is generally a result of up-slope migration instead of dispersing between sub-alpine zones (Graham et al., in prep).

Hybridisation and introgression are also important processes for diversification within oceanic archipelagos (Crawford and Stuessy, 1997; Mallet, 2007). Indeed, there are many documented examples of hybridisation occurring within Canarian flora (e.g., Brochmann et al., 2000; Herben et al., 2005; Barber et al., 2007; Jones et al., 2014; White et al., 2020). Yet, the relative importance of hybridisation in driving diversification is lesser known than geographic isolation and habitat shifts. While there are a few examples of recognised homoploid hybrid speciation events (Brochmann et al., 2000; White et al., 2018), often limitations within molecular datasets are unable to resolve hybridisation. More recent examinations of Canary Island flora, which utilise modern sequencing techniques, such as Next-generation sequencing (NGS), are starting to emerge and provide more robust estimates of gene flow between species. For example, D-statistics (method for detecting gene flow) on NGS datasets have revealed evidence of intra-island hybridisation between Canary Island *Argyranthemum* species (White et al., 2020) and intra- and inter-island introgression between species of Canary Island *Micromeria* (Curto et al., 2018). A 'combinatorial mechanism' has been proposed to explain rapid speciation; the old genetic variation often derived from hybridisation has contributed to an increase in genetic variation, enabling rapid adaptation (Marques et al., 2019). This method has been proposed to explain the rapid diversification in *Argyranthemum* (White et al., 2020) and *Helianthemum* (Albaladejo et al., 2021).

1.3.4 Future Research in Canary Island flora

Our knowledge of the evolutionary processes that underlie diversification and adaptation in the insular oceanic systems primarily stems from phylogenetic studies (Reviewed by Juan et al., 2000; Emerson, 2002; Parent et al., 2008; Brown et al., 2013). Many have provided fundamental principles on island diversification, for example, colonisation patterns (Silvertown, 2004; Silvertown et al., 2005). However, due to the small number of nuclear and/or chloroplast markers often used, many phylogenies exhibit low resolution and are poorly supported with short branch lengths (e.g., Böhle et al., 1996; Francisco-Ortega et al., 1996, 1999; Goodson et al., 2006). Issues are often exacerbated by limited sampling. Inferences regarding diversification processes on Canary Island flora can be challenging due to their close relationships, mainly when discerning incomplete lineage sorting (ILS) from other evolutionary processes such as hybridisation. However, advances in sequencing technology (such as NGS) have generated large genomic datasets consisting of thousands of genetic markers that have significantly improved the field of phylogenetics (Lemmon and Lemmon, 2013; Dodsworth et al., 2019).

Several studies have utilised next-generation technology to revisit iconic radiations, resulting in revised phylogenetic relationships of Canary Island flora, for example, in *Echium* (Graham et al., 2021)*, Argyranthemum* (White et al., 2020)*, Helianthemum* (Martín-Hernanz et al., 2019) and *Micromeria* (Lamiaceae; Curto et al., 2018). NGS studies have improved the resolution of species relationships, allowing more robust inferences regarding the frequency of geographic and ecological speciation events within the diversification of island lineages. In addition, large genomic datasets enable better detection of hybridisation events (Rannala and Yang, 2008; Curto et al., 2018; Mort et al., 2022). However, it is still an emerging field, and Stuessy et al. (2006) stated the difficulty in establishing general patterns of speciation on islands due to the lack of robust molecular studies.

Despite the recent advances in sequencing, genomic-level studies examing diversification on a molecular scale in Canary Island flora are rare. Currently, high-quality reference genomes of Canary Island endemics are restricted to animals, which include the Canary Island spider *Dysdera sylvatica* (Escuer et al., 2022) and the Canary Island barn owls (*Tyto alba*) (Cumer et al., 2022). Outside of the Canaries, examples of genomic studies focusing on island plant lineage include the Hawaiian *Metrosideros* radiation and the Galápagos *Scalesia atractyloides* radiation (Choi et al., 2021; Cerca et al., 2022), which have provided promising results in furthering our understanding

of oceanic island diversification. Both provide results which reveal the molecular basis of the evolutionary mechanisms behind adaptive radiation and have highlighted the pivotal role of hybridisation in facilitating rapid speciation and the colonisation of ecological niches (Seehausen, 2013; Berner and Salzburger, 2015; Marques et al., 2019; Choi et al., 2021). No genomic-level studies of plant radiation in the Canary Islands exist. Still, including Canary Island plant radiation genome datasets would only contribute to these genomic developments in explaining island diversification.

1.4 Genomic Datasets: Next-Generation and Third-Generation Sequencing

1.4.1 Conventional Sequencing

Before the availability of commercial next-generation sequencing (NGS) in 2005, speciation research often relied on Sanger sequencing on a small number of loci, frequently nuclear, ribosomal and chloroplast DNA (Schuster, 2008; Straiton et al., 2019). Despite many successful uses of Sanger technology in sequencing the model organisms (reviewed in Straiton et al., 2019), there were many limitations. Firstly, a single or limited number of genes provide little genetic variation, resulting in insufficient signals for robust estimates. Secondly, the use of chloroplast data can present erroneous species trees due to hybridisation and chloroplast capture, bringing uncertainty to species-level relationships (Degnan and Rosenberg, 2009).

1.4.2 Genotyping-by-Sequencing Datasets

The high demand for new and improved sequencing technologies for low-cost sequence data has driven the development of high-throughput sequencing technology, namely, next-generation sequencing (NGS). NGS is a parallel high-throughput sequencing approach, which can produce thousands or millions of sequences concurrently in the range of 25 and 700 bp in length from multiple samples at a significantly reduced cost compared to Sanger sequencing (Reuter et al., 2015).

The NGS approach has led to the development of many molecular sequencing methods (Slatko et al., 2018). One such method is reduced-representation libraries, such as Genotyping-by-sequencing (GBS), which have been the most widely used for species-level phylogenetic studies (Lemmon and Lemmon, 2013). GBS datasets are generated by first generating a reduced representation of the genome by restriction digestion, followed by pooling the samples and sequencing the fragments on an Illumina NGS platform (Elshire et al., 2011). The key benefit of this method is that it can produce thousands of single nucleotide polymorphisms (SNPs) per sample without the need for a reference genome at a relatively low-cost (Davey et al., 2011; Andrews et al., 2016). GBS datasets have been successfully applied to non-model plant and animal groups, capturing fine-scale population structure and proving well-resolve and robust phylogenies (e.g., Lerner et al., 2011; McCormack et al., 2012; Rubin et al., 2012; Straub et al., 2012; Wagner

et al., 2018). Moreover, these methods have enhanced the accuracy of demographic inference (Shafer et al., 2015; Le Moan et al., 2016; Rougemont et al., 2017).

1.4.3 Whole Genome Datasets and Third-Generation Sequencing

While reduce-representation datasets, such as GBS, are helpful for taxon delimitation and characterisation of demographic events, whole-genome datasets would provide the ultimate tool for understanding the genetics behind speciation and adaptation (Campbell et al., 2018). These genome datasets can unlock loci positional and functional information and characterise the genetic variation within and between species (Bentley, 2006; da Fonseca et al., 2016). With high-throughput sequencing enhancing sequencing ability and advances in computational power, generating whole-genome datasets are increasingly accessible for non-model organisms.

Typically, when utilising whole-genome datasets, a reference genome is needed, which represents a high-continuous and accurate representative sequence of the species genome. A single individual is sequenced at high read depth (the same nucleotide sequenced multiple times), usually with coverage of >90x. Many reference genomes use high-quality short NGS sequencing fragments (ca. 150 bps), such as those generated by the Illumina platform (Schatz et al., 2010). These sequencing reads are then aligned to create contigs, i.e., overlapping sequences representing a consensus DNA region. To bridge the gaps between contigs, a contig assembly can be “scaffolded” to create a more continuous assembly (Hunt et al., 2014).

Since 2005, most genomes have been assembled using only NGS libraries. However, genome sequencing has recently benefitted from developments in third-generation sequencing (i.e. long-read sequencing; Check Hayden, 2009). The short read lengths of NGS are often a shortcoming due to challenges in assembling repetitive regions and the characterisation of structural variants (da Fonseca et al., 2016). Longer reads libraries (5 – 15 k bp for Pacific Biosciences and Oxford Nanopore sequencing; Branton et al., 2008; Rhoads and Au, 2015) can alleviate many of these issues. As long-read data often exhibit high error rates, they are often combined with the more accurate NGS reads for a hybrid approach (Amarasinghe et al., 2020). Nonetheless, due to the advent of long-read sequencing, there has been a rapid increase in genome assemblies with larger contigs and of higher quality (Koren et al., 2012).

Once a reference genome is available, the resequencing genomes in multiple related individuals becomes a financially feasible method of generating large genome datasets (Bentley, 2006). Whole genome resequencing technology produces low-coverage (e.g., 10×) genome sequences, in many cases for multiple individuals, which are then mapped a reference genome (Poelstra et al., 2014). Informative genetic markers can then be extracted to provide large marker datasets (Li et al., 2009), which offer enormous amounts of information across the entire genome. Successful utilisation of resequencing datasets is seen in many crop domestication/improvement (e.g., Lu et al., 2019; Qin et al., 2014; Rubin et al., 2010; Verde et al., 2013) and selection and adaptation studies (e.g., Axelsson et al., 2013; Burke et al., 2010; Kim et al., 2014; Rubin et al., 2012; Turner et al., 2010).

1.5 Application of Genomics in Speciation Research

The molecular tools used to study genetic patterns across a wide variety of non-model taxa are beginning to focus on utilising whole-genome datasets to answer a broad range of questions (Ekblom and Galindo, 2011). With the falling cost of sequencing and the development of computational tools, genomic methods for non-model organisms are becoming increasingly more accessible and transforming our understanding of ecology, evolution, and patterns of diversification (Schuster, 2008; da Fonseca et al., 2016).

Two common questions within speciation and adaptation research where genomic techniques are employed are: (1) what is the genetic basis of reproductive isolation, and (2) are signatures of ancient and recent gene flow detectable? Addressing these questions has led to debates on the importance of geographic isolation (Gavrilets, 2003; Barluenga et al., 2006; Butlin et al., 2008; Kisel and Barraclough, 2010), ecological divergence (Schluter, 2000, 2009; Gavrilets, 2003; Doebeli et al., 2005; Rundle and Nosil, 2005; Nosil et al., 2009), and speciation with gene flow (Feder et al., 2012; Richards et al., 2019) in facilitating or limiting reproductive isolation.

1.5.1 Genomics of Adaptation

Understanding the genetic basis of adaptive traits has been a central objective within evolution biology for decades (Orr and Coyne, 1992). The availability of whole-scale datasets has made it more feasible to gain molecular-level insights into adaptation through the identification of candidate genes, i.e., genes that underlie a phenotypic or physiological response of adaptive significance (e.g., Swanson and Vacquier, 2002; Rieseberg and Blackman, 2010; Meier et al., 2018; Fang et al., 2020; De La Torre et al., 2021). For example, Poelstra et al. (2014) identified variants of putatively divergent genes potentially involved in the *cis*-regulation of feather pigmentation and visual perception between Hooded and Carrion crows. Thus, they are considered candidate loci potentially involved in premating reproductive isolation. One of the most popular and reliable methods is a *dN/dS* analysis, which measures the selection on protein-coding genes by the ratio of nonsynonymous (*dN*) and synonymous substitution per site (*dS*) between species (Kryazhimskiy and Plotkin, 2008). For example, *dN/dS* analyses found key genes that underlie shifts in host-plants in butterflies (Allio et al., 2021) and local adaptation to drought in wild tomatoes (Xia et al., 2010)

Identifying candidate genes that promote divergence is challenging (Orr et al., 2004; Wu and Ting, 2004; Rieseberg and Blackman, 2010; Nosil and Schluter, 2011). Firstly, it requires extensive high-quality genomic resources and extensive bioinformatic processing. Secondly, the success of correctly identifying genes under selection varies depending on taxonomic scale, sequencing quality, and annotation, among other factors. Thirdly, an adaptive phenotype often involves complex interactions between many genes, which can be difficult to detect (Wu and Ting, 2004). In addition, further work is often needed to confirm the involvement of the candidate genes as an adaptive trait. The most robust studies include experimental manipulation through transgenic or gene knockouts and expression and phenotype analyses to confirm their function and physiological effects (Nosil and Schluter, 2011).

1.5.1.1 Functional Genomics of Adaptation

Many genomic studies identify candidate genes, i.e., regions with high divergence or associated with reproductive isolation (Tabor et al., 2002; Zhu and Zhao, 2007). However, the function of these genes is often predicted through a combination of a*b initio* gene prediction methods which detects genomic signatures of protein-coding genes (such as long open reading frames and transcriptional initiation sequencing) and empirical methods using homology-based evidence. The success of gene prediction is often limited to the quality of the gene sequence and sequence database (Zhu and Zhao, 2007).

Experimental functional genomic studies can distinguish between alternative hypotheses by providing functional links between genes and phenotypes and, thus, provide more robust inferences of their involvement in speciation. Transgenics provides a means to test the function of candidate genes with the insertion or deletion of a gene to give a loss-of-function mutant individual (Østergaard and Yanofsky, 2004), which has been utilised in many studies to successfully identify the function of genes (e.g., Lee et al., 1994; Ito and Meyerowitz, 2000; Filleur et al., 2001; Koh et al., 2007).

1.5.2 Genomics of Hybridisation and Introgression

Hybridisation and introgression are increasingly seen to play a role in speciation (Seehausen, 2004, 2004; Soltis and Soltis, 2009; Feder et al., 2012; Abbott et al., 2013). With recent advances

in sequencing enabling the generation of thousands of loci, phylogenetic networks (i.e. networks) are an increasingly popular genomic method to detect hybridisation events between species. These networks infer reticulate evolution (merging of two lineages) through the conflicting tree topologies of, typically, thousands of individual loci (Huson and Bryant, 2006; Nakhleh, 2011; Wen et al., 2016). Studies have successfully identified hybridisation events through phylogenetic networks in a number of plants and animals, for example, the Neotropical genus *Lachemilla* (Rosaceae; Morales-Briones et al., 2018), whiptail lizards (Barley et al., 2022), Hawaiian *Cyrtandra* (Gesneriaceae; Kleinkopf et al., 2019) and bear species (V. Kumar et al., 2017). Alternative statistical tests have also been developed for genome-wide datasets which do not rely on the generation of gene trees, for instance, Treemix (Pickrell and Pritchard, 2012) and D-statistics (Patterson et al., 2012), which are now routinely used to detect introgression (e.g., Eaton and Ree, 2013; Brandvain et al., 2014; White et al., 2020).

1.6 Study System: Canary Island *Descurainia*

Descurainia (Brassicaceae) Webb & Berthel is a genus of 45 species distributed throughout America, Eurasia, and the Canary Islands (Bramwell, 1977; Goodson et al., 2011). The Canary Island clade of *Descurainia* comprises seven species distributed across four islands (Tenerife, La Palma, La Gomera, and Gran Canaria), relatively large-flowered woody perennials. Five species (*D. gilva*, *D. gonzalezii*, *D. preauxiana*, *D. artemisioides*, and *D. lemsii*) are single island endemics (SIEs), and two species (*D. millefolia* and *D. bourgaeana*) are multi-island endemics (MIEs) found on three and two islands, respectively (Bramwell, 1977; Goodson et al., 2006; Figure 1.3; Table 1.1). Canary Island *Descurainia* has received lesser attention than the larger iconic radiations within the Canary Island, for example, *Echium, Sonchus and Argyranthemum*. However, the few available studies described the group as endemic radiation distributed across multiple islands and ecological zones along an elevational cline (Goodson et al., 2006). A summary of their habitats, elevation, and islands they occupy is described in table 1.1.

Table 1.1: Overview of the seven species of endemic Canary Island *Descurainia* (Brassicaceae). Habitat descriptions and elevation range from Bramwell (1977). Images of each species can be found in Figure 1.3.

Species	Authority	Island	Habitat and distribution description	Elevation range (m)
D. artemisioides	Svent.	Gran Canaria	Restricted to shady ravines and cliffs in the west (Corral and Cáceres, 2006)	400 – 800
D. bourgaeana	Webb ex O.E. Schulz	Tenerife, La Palma	Sub-alpine scrub, rocky volcanic terrain on Las Cañadas.	2,000 – 2,300
D. gilva	Svent.	La Palma	Outer boundaries of the pine forests and sub-alpine habitat of the Caldera de Taburiente in the North Central regions	1,700 – 2,100
D. gonzalezii	Svent.	Tenerife	South-west slope in pine forests	1,800 – 2,100
D. lemsii	Bramwell	Tenerife	Outer boundaries on the pine forest on the north-central rim of the Las Cañadas Caldera, . Some overlapping distributions with *D. bourgaeana*.	1,750 – 2,000
D. millefolia	(Jacq.) Webb & Berth.	Tenerife, La Palma, and Gran Canaria	Lower cliffs, rocks, and walls in the shade in the low-elevation scrub	50 – 1,000
D. preauxiana	(Webb) Webb ex O.E. Schulz	Gran Canaria	Widespread distribution in central and eastern Gran Canaria	400 – 900 m

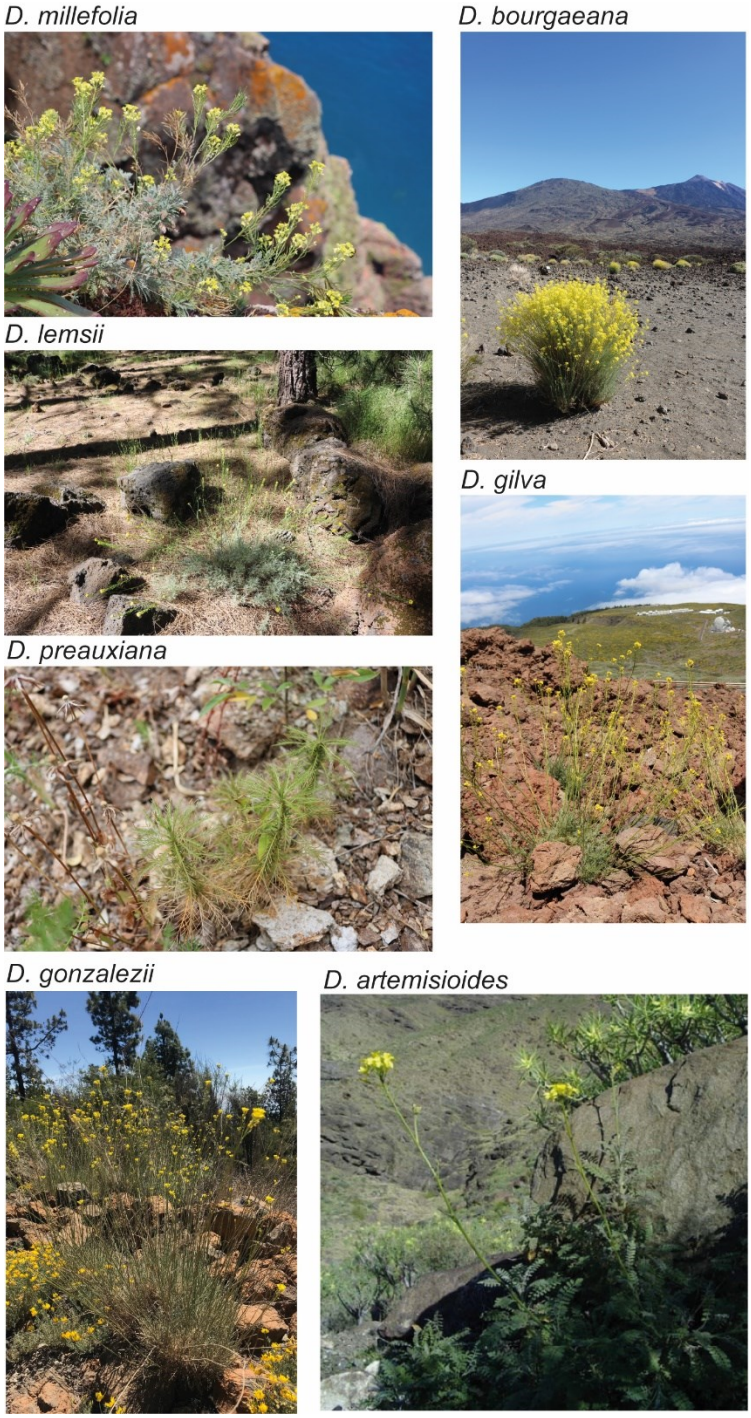


Figure 1.3: The seven species of *Descurainia* endemic to the Canary Islands. Picture of *D. artemisioides* sourced from Corral and Cáceres (2006).

1.6.1 Previous Research on Canary Island *Descurainia*

A phylogenetic framework for understanding the diversification of *Descurainia* within the Canary Islands was provided by Goodson et al. (2006), who addressed taxonomic and biogeographical questions using nuclear Internal transcribed spacer (ITS) and non-coding chloroplast DNA (cpDNA) for phylogenetic reconstructions. The ITS data, which contained 128 variable sites and 47 parsimony informative sites, resulted in an unresolved phylogeny and failed to provide information about relationships between Canary Island taxa. However, the phylogeny strongly supported the monophyly of Canary Island *Descurainia* and identified *Descurainia tanacetifolia* (previously *Hugueninia tanacetifolia* (L.) *Rchb.*), a species restricted to Italian and Swiss Alps and the Pyrenees, as a sister group. This suggests that Canary Island *Descurainia* colonised the Canary Islands through a single colonisation event from a continental relative. Using a combined data set of 44 informative characters of rapidly evolving non-coding chloroplast (cp) data, additional phylogenetic reconstructions support Canary Island *Descurainia* as monophyletic and sister to *D. tanacetifolia*. Specifically, *D. tanacetifolia* spp. *Suffruticosa,* a sub-species found only in the Pyrenees, is more closely related to the Canarian species than that spp. *Tanacetifolia* (Goodson et al., 2006). The study further suggested that the Canary Islands were colonised 1 – 2.8 Mya from the Old World, most likely from the Iberian Peninsula, which acted as a refugium during Pleistocene glacial maxima.

The cp data successfully recovered species-level taxonomic relationships and evolutionary patterns in Canary Island *Descurainia* (Goodson et al., 2006). Two major clades were identified: the “Tenerife” clade, restricted to Tenerife, and the “mixed” clade, including taxa from four islands. The study found that *D. millefolia,* the most widespread species, was polyphyletic. *Descurainia millefolia* exhibits wide variation in several morphological traits, as well as occupying a range of habitat types. Schulz (1924) described three subspecies of (spp. *brachycarpa* (west of La Palma), spp. *Marcrocarpa* (Tenerife, La Palma, and La Gomera), and spp. *Sabinalis* (Tenerife)) to account for the variation observed although these were not recognised by Bramwell (1977). The Goodson et al. (2006) results raise the possibility that these may, indeed, be genetically distinct. However, population genetic and morphological studies would be needed to understand relationships within *D millefolia*. *Descuraina gonzalezii* was resolved as paraphyletic, but given the short branch lengths, this was thought most likely to reflect recent chloroplast capture from *D. bourgaeana.*

Based on their phylogeny, Goodson et al. (2006) proposed an initial colonisation of Tenerife, with subsequent dispersal to La Palma, dispersal events from La Palma to La Gomera and Gran Canaria, followed by a back-colonisation from La Palma to Tenerife.

Seven endemic species distributed across four islands suggests that intra- and inter-island speciation has occurred in Canary Island *Descurainia*. Ancestral habitat reconstructions presented suggested habitat shifts following the colonisation of from low-elevation scrub to pine forest and sub-alpine scrub on Tenerife in *D. gonzalezii, D. lemsii* and *D. bourgaeana* (Goodson et al., 2006), whereas allopatric speciation between comparable habitats on Tenerife/La Palma and Gran Canaria, gave rise to *D. artemisoides* and *D. preauxiana*. A habitat shift was also inferred on La Palma for *D. gilva*. Using the phylogenetic framework of Goodson et al. (2006), Herben et al. (2014) subsequently detected a phylogenetic signal within vegetative traits of Canary Islands *Descurainia*, indicating a selection for traits involved in habitat shifts. Therefore, it is believed that abiotic habitat shifts combined with island isolation have contributed to speciation within Canary Island *Descurainia*.

Earlier studies have provided a framework for discussing phylogenetic relationships among Canary Island *Descurainia* taxa (Goodson et al., 2006, 2011), but knowledge gaps persist. For example, the ITS phylogenetic tree of Canarian Island *Descurainia* (above) failed to resolve relationships within *Descurainia* because the ITS data set contained no parsimony-informative characters (Goodson et al., 2006). Meanwhile, nodes in the cpDNA tree are well-supported. Still, the tree lacks resolution, and the unusual placements of individuals, for example, the paraphyly of *D. millefolia* and *D. gonzalezii,* may be due to the chloroplast markers failing to provide enough variation for effective phylogenetic reconstructions. In addition, these placements strongly suggest hybridisation between species, which can result in a transfer of cp DNA from one species into another, obscuring relationships.

Further, hybrids between *D. bourgaeana* and *D. lemsii* and *D. bourgaeana* and *D. gonzalezii* have been identified (Goodson et al., 2006). Canary Island *Descurainia* frequently occur in close geographical proximity with overlapping flowering periods, which could allow gene flow (Goodson 2008). There is also no variation in ploidy level (Bramwell 1976), suggesting a lack of karyotypic barriers to interbreeding (Ramsey and Schemske, 1998). Therefore, gene flow and hybridisation between individuals are feasible and must be accounted for when constructing phylogenies.

1.6.2 Canary Island *Descurainia* as a Model Organism for Studying Speciation

Over the past few decades, many evolutionary hypotheses have been proposed to explain the diversity of Canary Island flora (e.g., Bramwell, 1976; Juan et al., 2000; Reyes-Betancort et al., 2008; Bramwell and Caujapé-Castells, 2011; Caujapé-Castells et al., 2017). Adaptation to local environments has played a significant role in their evolutionary history, as has been inferred for speciation processes more widely (Rundle and Nosil, 2005). Yet very few studies have been undertaken to characterise these adaptations and understand the genetic basis of the divergence of the species. The lack of genomic resources to address these questions is most likely the biggest hindrance due to the large amount of genomic information needed (Wolf et al., 2010). This issue is not restricted to the Canarian flora but is indicative of the current cost of sequencing and the lack of suitable systems needed to address these problems. Therefore, much progress in addressing these questions has been restricted to traditional models systems such as *Arabidopsis* (Stinchcombe et al., 2004; Bikard et al., 2009; Hu et al., 2011; Schmickl and Koch, 2011; Hämälä and Savolainen, 2019) and *Drosophila* (Ayala et al., 1974; Coyne and Orr, 1989; Noor, 1995; Kelly and Noor, 1996; Presgraves, 2008; Garrigan et al., 2012).

Canary Island *Descurainia* offers an opportunity to study ecological speciation as it has a wide array of biological characteristics that make it an emerging model for ecological and evolutionary functional genomics, as follows:

- *Genome size*: Plants have an extensive range of genome sizes, ranging from 125 Mb to 120,000 Mb (Pellicer and Leitch, 2020). The sequencing cost and computational complexity increase with genome size. The seven species of Canary Island *Descurainia* are diploid (2n =14; Bramwell, 1977), and their genome sizes are estimated to be relatively small for plants, ranging between ca. 195 – 225 Mb (Pellicer and Leitch, 2020). Small and diploid genomes provide a cost-effective sequencing project - a rare opportunity within non-model organisms. Particularly in plants which are often ploidy with large genomes, provide sequencing and computational complexities (Li and Harkess, 2018).
- *Ecological significance*: Canary Island *Descurainia* is distributed across various habitat types (e.g., pine forests, coastal scrub, sub-alpine desert), suggesting genetic differentiation resulting from ecological adaptation (Goodson et al., 2006). In particular, several species of

Canary Island *Descurainia* (*D. bourgaeana, D. gilva, D. gonzalezii,* and *D. gilva*) occupy high-elevation habitats of the taller islands of Tenerife and La Palma, which could provide an invaluable opportunity for studying adaptation to these challenging environments.

- *Gene discovery:* The availability of genetic resources for species in Brassicaceae is extensive. This is primarily due to extensive research on economically important *Brassica* crops and model species in the family. For instance, *Arabidops*is *thaliana* is one of the most well-studied organisms, and a significant amount of research has been done to understand the function of its 25,498 genes (The Arabidopsis Genome Initiative, 2000). Limited knowledge of the function of genes is one of the biggest hindrances in disentangling genotype-phenotype in non-model species. This issue is mitigated for *Descurainia* by using the well-characterised genomes of *Arabidopsis* and other Brassicaceae species from which the *Descurainia* genome can be annotated.

The combination of the above factors highlights the suitability of Canary Island *Descurainia* to develop our understanding of speciation processes, particularly ecological adaptation, within an oceanic island context.

1.7 Summary of Thesis

1.7.1 Aims and Objectives

This thesis aims to investigate the evolutionary processes underlying the diversification and adaptation of the seven species of *Descurainia* (Brassicaceae) that are endemic to the Canary Islands. I will attempt to address broader questions regarding ecological adaptation in plants using genomic and more traditional greenhouse methods. The three main objectives of this thesis are as follows:

1. To revisit the phylogenetic relationships of Canary Island *Descurainia* using next-generation sequencing technologies to investigate the roles that ecology, geography and hybridisation have played in the diversification of this lineage.
2. To establish a high-quality reference genome to be utilised as a tool to address evolutionary and ecological questions and understand the relationship of *Descurainia* within the family Brassicaceae.
3. To identify candidate genes involved in the adaption to high-elevation habitats within Canary Island *Descurainia* and assess the level of genomic convergence between *Descurainia* and other species in Brassicaceae adapted to high-elevation or high-latitude environments.

1.7.2 Outline of Thesis

The chapters of this thesis are presented in chronological order in which the data were produced and analysed.

In **Chapter 2,** I explore alternative phylogenetic hypothesises for Canary Island *Descurainia* using genotyping-by-sequencing (GBS) technology. The phylogenetic reconstructions serve as a framework for inferring the group's evolutionary history. Multiple phylogenetic and genomics techniques were undertaken to identify potential hybridisation events between species.

In **Chapter 3,** I assembled and annotated a reference genome of *Descurainia millefolia* using a hybrid assembly approach. The aim is to generate a highly-complete genome to enable further downstream comparison genomic analysis. The *Descurainia* genome is then compared with other species in the Brassicaceae in a genomic and phylogenetic context.

In **Chapter 4,** I use comparative genomic techniques to explore the molecular basis for adaptation to high-elevation environments within the high-elevation Canary Island *Descurainia* species. Genomes were re-sequenced for all seven species and the continental outgroup to identify

genome-wide patterns of positive selection using the genome generated from chapter three. These genes were also examined for molecular convergence between other species in Brassicaceae that are distributed in high-elevation habitats. Candidate genes were then tested for their function and effects on fitness using transgenic *Arabidopsis thaliana* within greenhouse experiments.

Chapter 2 The diversification of the Canary Islands endemic *Descurainia* (Brassicaceae) using genotyping-by-sequencing (GBS)

2.1 Abstract

Oceanic islands offer the opportunity to understand evolutionary processes underlying the rapid diversification of plant lineages. Along with geographic isolation and ecological shifts, a growing body of genomic evidence has suggested that hybridisation can play an essential role in the evolution of these insular groups. However, hybridisation is challenging to detect due to confounding evolutionary processes, such as incomplete lineage sorting (ILS), and methods that differentiate between ILS and hybridisation in phylogenetic studies need to be utilised. We carried out genotyping-by-sequencing (GBS) for 18 individuals from all seven Canary Island *Descurainia* (Brassicaceae) species and two outgroups. Phylogenetic analyses of the GBS data were performed using a supermatrix approach and individual gene trees to reconstruct species networks to infer evolutionary relationships. Hybridisation events inferred in the gene tree analysis were further tested using gene flow tests and modelling approaches. Climatic data was gathered to describe ecological niches for each taxon and examine the relationship between ecological niches and diversification. Analysis of the supermatrix dataset resulted in a fully resolved phylogeny. Species networks suggest a hybrid origin for one taxon, with these results being supported by additional hybridisation tests. Strong phylogenetic signals for bioclimatic variables (temperature and precipitation) indicate only one major ecological shift within Canary Island *Descurainia*. They suggest that inter- and intra-island dispersal between similar habitats has played an important role in the evolution of the group. Inter-island dispersal played a significant role in the diversification of *Descurainia*, with evidence of only limited shifts in climate preferences. Despite weak reproductive barriers and known interspecific hybrids, hybridization appears to have played only a limited role in the diversification of the group, with a single instance detected. The results highlight the need to use phylogenetic network approaches that can simultaneously accommodate ILS and gene flow when studying groups prone to hybridisation patterns that might otherwise be obscured in species trees.

2.2 Introduction

Island systems have played a fundamental role in providing insights into the processes underlying the evolution of plant life (Warren et al., 2015). The Canary Islands, an oceanic archipelago comprised of seven islands, is a hot-spot for biodiversity and exhibits high levels of endemism (Reyes-Betancort et al., 2008; Bramwell and Caujapé-Castells, 2011). Approximately 570 plant species are endemic to the Canary Islands (25% of the entire flora; Bramwell, 1976), many of which are found in recent and rapidly diversifying lineages, and the archipelago's flora has been used to investigate evolution, speciation, and adaptation (Juan et al., 2000). Three main processes have been identified to explain the high levels of endemism found across these islands: (1) allopatric speciation arising through geographical isolation, typically after inter-island dispersal between similar habitats (Francisco-Ortega et al., 2000; Juan et al., 2000); (2) ecological speciation due to speciation into new ecological niches (Silvertown et al., 2005); and (3) hybrid speciation owing to the weak reproductive barriers, dynamic landscapes and the close proximity of related species with different ecological niches (Silvertown, 2004; Fjellheim et al., 2009; White et al., 2018). A significant challenge when constructing the evolutionary history of Canary Islands lineages is untangling the relative influence of the different evolutionary processes involved.

Interspecific hybridisation has long been viewed as an essential driver of speciation in plants (Rieseberg and Willis, 2007). Nevertheless, despite numerous studies demonstrating hybridisation and gene flow between Canary Island lineages (Kim et al., 1996; Francisco-Ortega et al., 1999; Brochmann et al., 2000; Carine et al., 2007; Fjellheim et al., 2009). Most molecular studies have focused only on the relative roles of ecological speciation and inter-island dispersal in generating the high levels of species diversity observed (Jorgensen and Frydenberg, 1999; Barber et al., 2000; Francisco-Ortega et al., 2002; Allan et al., 2004). Hybridisation as a potential evolutionary force is often overlooked, likely due to the difficulty of confirming that hybridisation has occurred when using small numbers of genetic markers. The low variation uncovered when using a limited number of markers in recent species radiations generally results in poorly resolved phylogenies. Even when phylogenetic incongruence is identified, for example, when chloroplast [cpDNA] and nuclear [nDNA] give different phylogenetic hypotheses (Francisco-Ortega et al., 1996; Barber et al., 2000; Mort et al., 2002; Jones et al., 2014), the use of only a small number of markers makes it difficult to determine the cause of the incongruence.

Reduced-representation sequencing methods, such as Restriction site Association DNA Sequencing (RAD-seq; Eaton and Ree, 2013) and Genotyping-by-Sequencing (GBS; Elshire et al., 2011), using high-throughput sequencing technologies to sample hundreds or thousands of loci throughout the genome simultaneously (Baird et al., 2008; Rowe et al., 2011; Eaton and Ree, 2013; Fernández-Mazuecos et al., 2018). These methods generate large numbers of phylogenetically informative markers, which can overcome the limitations of traditional marker methods. As a result, RAD-seq and GBS data sets provide new insights into the evolution of the Canary Islands flora, as demonstrated in *Micromeria* (Lamiaceae; Puppo et al., 2015) and *Argyranthemum* (Asteraceae; (White et al., 2018, 2020). Combined with the development of new statistical methods (Green et al., 2010; Patterson et al., 2012; Sankararaman et al., 2014; Vernot and Akey, 2014), studies are also starting to emerge which offer more comprehensive insight into hybridisation in oceanic lineages (Puppo et al., 2015; White et al., 2020). These developments reinforce the notion that gene discordance due to hybridisation is not only common but is shaping the plant diversity found across the Canary Islands.

Descurainia (Brassicaceae) Webb & Berthel is a genus of 45 species distributed throughout America, Eurasia, and the Canary Islands (Bramwell, 1977; Goodson et al., 2011). The seven species in the Canary Islands form a monophyletic group (Goodson et al., 2006) distributed across four islands (Figure 2.1). Five species (*D. gilva* Svent., *D. gonzalezii* Svent., *D. preauxiana* (Webb) Webb ex O.E.Schulz, *D. artemisioides* Svent., and *D. lemsii* Bramwell.) are single island endemics (SIEs), and two species (*D. millefolia* Webb & Berthel. and *D. bourgaeana* (E.Fourn.) Webb ex O.E.Schulz.) are multi-island endemics (MIEs) found on three and two islands, respectively (Bramwell, 1977; Goodson et al., 2006). The species are found in a range of different habitats, and Goodson et al. (2006) broadly categorised the ecological zones for Canary Island *Descurainia* as lowland scrub and, more rarely, the thermo-sclerophilous woodland (250 - 700 m) for *D. millefolia*, *D. preauxiana*, and *D. artemisioides*, pine forest (600 – 1,200 m) for *D. gilva*, *D. gonzalezii* and *D. lemsii*, and sub-alpine scrub (1,900 – 2,100 m) for *D. bourgaeana* (Bramwell, 1977; Goodson et al., 2006). Putative hybrids have been documented in Tenerife (e.g., *D. bourgaeana* x *D. lemsii* and *D. gonzalezii* x *D. bourgaeana*; Goodson et al., 2006). Given the distribution of Canary Island *Descurainia* across multiple islands and ecological zones following a single colonisation of the archipelago (Goodson et al., 2006, 2011), the group offers an opportunity to examine the roles of different evolutionary processes underlying diversification within this archipelago.

Previous phylogenetic analyses of Canary Island *Descurainia* applied nDNA (ITS, the internal transcribed spacer region) and cpDNA sequences and resolved the Euro-Siberian *Descurainia tanacetifolia* as the continental sister species to the island clade (Goodson et al., 2006, 2011). The ITS phylogeny failed to resolve relationships within the Canary Island clade. However, the cpDNA phylogeny revealed two clades. One clade was restricted to Tenerife (*D. gonzalezii*, *D. lemsii* and some *D. millefolia* accessions), and the other comprised taxa from across the four islands (*D. bourgaeana* on Tenerife, *D. artemisioides* and *D. preauxiana* on Gran Canaria, *D. gilva* on La Palma, and *D. millefolia* from La Palma and La Gomera). *Descurainia millefolia* was resolved in both clades, and *D. gonzalezii* was found to be polyphyletic in the Tenerife clade.

Goodson et al. (2006) suggested that the drivers of diversification in Canary Island *Descurainia* were intra-island ecological speciation and inter-island colonisation. Studies that have subsequently used the Goodson et al. (2006) phylogenetic hypothesis to examine the adaptive radiation of *Descurainia* provide conflicting results on the role of climate and habitats (Herben et al., 2014; Steinbauer et al., 2016). The polyphyly of *D. millefolia* and *D. gonzalezii* revealed by their study could suggest taxonomic revision is needed. However, it also raises questions about the influence of hybridisation on the relationships inferred using chloroplast markers (Goodson et al., 2006). Chloroplast capture can give rise to topologies that do not reflect species relationships, leading to incorrect species trees (Rieseberg and Soltis, 1991). This is backed up by field observations of putative hybrids (Goodson et al., 2006). There is also the potential for gene flow as many species occur in close geographical proximity, with a lack of karyotypic barriers, similar floral morphologies, and long overlapping flowering periods (Bramwell, 1977).

Here, we use Genotyping-by-Sequencing (GBS) data in conjunction with ecological and distribution data across all seven Canary Island *Descurainia* species and two continental outgroups to resolve the evolutionary history of the group. Our objectives are to (1) investigate phylogenetic relationships using both concatenated data and coalescence-based species networks and using an Approximate Bayesian Computation (ABC) modelling analysis (Cornuet et al., 2008) to examine alternate phylogenies, (2) quantify the extent of hybridisation between taxa using D-statistics (Eaton and Ree, 2013), and species networks, (3) infer the role of ecology in the diversification of Canary Island *Descurainia* using spatial climatic variables, and (4) in so doing, establish the relative role of geographical isolation, hybridisation and ecological divergence in the diversification of the group.

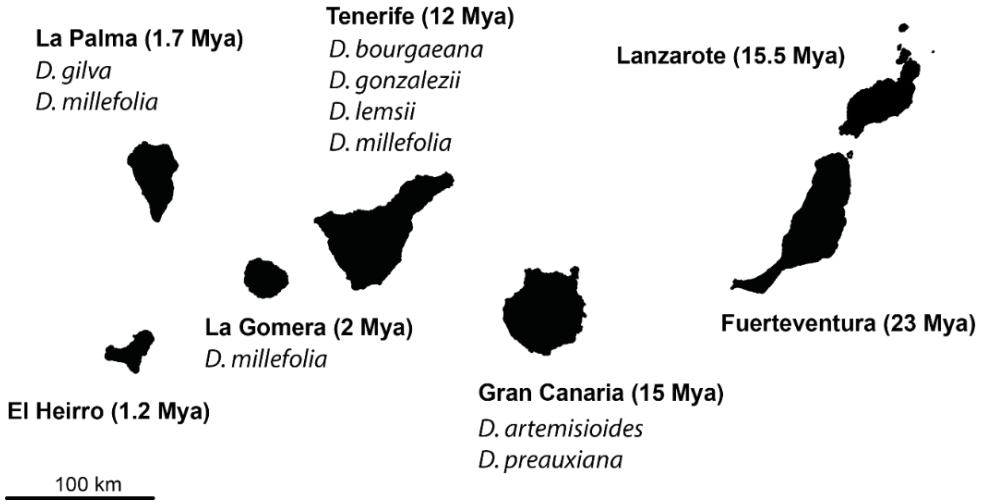


Figure 2.1: Map of the Canary Islands (age of islands in brackets; Schmincke, 1976) and the distribution of each island's seven endemic Canary Island *Descurainia*.

2.3 Material and Methods

2.3.1 Sampling, DNA extraction, and Genotyping-by-sequencing (GBS)

A total of 18 individuals from eight populations were used to represent the seven Canary Island species of *Descurainia* (Appendix Table A.1). Sampling comprised five specimens of *D. bourgaeana*, two each of *D. gilva*, *D. gonzalezii*, *D. artemisioides* and *D. preauxiana*, and four of *D. millefolia* (two from La Palma and two from Tenerife), and one individual of *D. lemsii*. *Descurainia millefolia* populations from La Gomera and *D. bourgaeana* populations from La Palma were not sampled. *Descurainia tanacetifolia* and *D. depressa* were sampled as outgroups.

DNA was isolated from silica-dried leaves collected from wild plant material (collected in 2016; eight samples, permit numbers in the Acknowledgements section) and from fresh leaves of plants grown from seed from the Banco de Germoplasma Vegetal (BGV) seed bank (four samples; Appendix Table A.1). DNA isolations were performed using a modified CTAB protocol (Doyle, 1991). An additional eight DNA samples used in the Goodson et al. (2006) study were supplied by R. Jansen (University of Texas).

DNA samples were sequenced using Genotyping-by-Sequencing (GBS) at Novogene (Hong Kong). DNA was digested with *Mse*I and *Hae*II, and fragments were sequenced for 144 cycles on an Illumina HiSeq.

2.3.2 GBS assembly

The GBS data were quality filtered and assembled using ipyrad v.0.7.23 (Eaton and Ree, 2013; Eaton and Overcast, 2020). Raw sequence data were demultiplexed for each sample using barcode sequences with no mismatches allowed. Low-quality base scores (< 33) were converted to N, and reads with more than five Ns were discarded. Adapter sequences were filtered using the strict setting (2) as recommended for GBS data. The *de novo*-reference assembly method was used to remove reads which mapped to the chloroplast or mitochondrial genomes of *Arabidopsis thaliana* (L.) Heynh. (NCBI Reference Sequence NC_000932.1 and NC_001284.2, respectively) whilst assembling unmapped nuclear reads *de novo*. Reads were assembled to identify consensus allele sequences within individuals using three similarity clustering thresholds (80%, 85%, and

90%) before clustering consensus allele sequences across samples to identify loci. For the three levels of clustering employed, loci were filtered for a minimum number of samples per locus of 8 or 10, equivalent to 60% and 50% missing data, respectively. A minimum depth of six reads was required for base calling, and any locus with a shared heterozygous site in 20% or more of samples was removed as potentially comprising paralogs. Therefore, six data sets were generated with three clustering thresholds (hereafter referred to as c80, c85, c90) and two minimum samples per locus threshold (hereafter m8, m10). For each assembly, ipyrad produces a custom "loci" format, which was used for all subsequent analyses.

2.3.3 Phylogenetic Reconstructions 1 – Concatenated Approach

A supermatrix approach was first used for phylogenetic reconstructions to infer evolutionary relationships and to allow comparison between GBS assemblies. This involves concatenating loci from the GBS assembly data set into a single alignment. For each of the six assemblies, the optimal model of sequence evolution was identified using ModelTest-NG v.0.1.5 (DARRIBA et al., 2020; Appendix Table A.2), and a Maximum likelihood (ML) tree was generated using RAxML Next-Generation v.0.6.0 (Stamatakis, 2014), with bootstrap support estimated from 1,000 replicate searches from random starting trees.

Five of the six GBS data sets gave identical phylogenetic topologies (see results). Of these five, the data set with a clustering threshold of 85% and a minimum sample number of 8 (c85m8) was selected for all subsequent analyses as it had the most significant number of SNPs.

2.3.4 Phylogenetic Reconstructions 2 – Coalescent Approach

The second phylogenetic approach uses a coalescent-based method which can accommodate reticulated evolutionary histories and incomplete lineage sorting (ILS). Here, we generated species networks modelling hybridisation and ILS using Phylonet v.3.5 (Than et al., 2008). To prepare the data, the unlinked SNP data set generated by ipyrad was transformed into bi-allelic markers. *Descurainia depressa* was removed from the analysis because of its more distant relationship to the Canary Islands ingroup (see results). The function *MLE_BiMarkers* in Phylonet was used to generate a pseudo-likelihood ML estimation of phylogenetic networks with the "-diploid" and "-pseudo" parameters. The maximum number of reticulation nodes in the phylogenetic networks

explored during the search was limited to 8. This number was set to a small number to reduce the overall complexity of a network. The number of runs was set to 50 to reduce memory load. Where multiple individuals of a species were present, these were mapped to a single species to reduce computational time to establish a species tree. The remaining parameters were set to their default settings.

2.3.5 Approximate Bayesian Computation (ABC)

We used an ABC approach (Beaumont et al., 2002) to assess the posterior probabilities of the evolutionary scenarios suggested by our analyses (concatenated vs coalescent approaches; see results). DIYABC v.2.1 (Cornuet et al., 2008) was used to compare two scenarios: (1) a fully bifurcating phylogeny representing the topology generated using ML analysis of the concatenated data set; and (2) a scenario including a hybrid origin for *D. gilva* between *D. millefolia* and *D. gonzalezii* as inferred in the coalescent analysis.

The unlinked SNP dataset was transformed into a DIYABC-friendly format using the python script (vcf2DIYABC.py, available from https://github.com/loire/ [accessed July 2019]).

Population sizes and divergence rates were set between 10^2 - 10^7, and admixture rates were 0.001-0.999. A total of 10^7 simulations were performed for each of the two scenarios. Summary statistics were incorporated to compare the observed and simulated data comparisons, which comprised the mean of genetic diversity, pairwise sample F_{st} and Nei's distance, and admixture summary statistics (Hudson et al., 1992; Choisy et al., 2004). The most probable scenario was identified using the posterior probability for each scenario, computed using a direct approach and Logistic regression. The scenario probabilities were calculated using 500 simulations for the direct estimate and 10,000 for the Logistic regression.

Once the most probable scenario was identified, posterior analyses were carried out to evaluate the robustness of the simulation of the selected scenario with our GBS data set. As DIYABC is a computationally and memory-intensive method, it was necessary to perform the remaining post-simulation analyses on a subset of 1,000 simulations. Firstly, the prior distribution of the scenario was performed using a logistic regression, from which false positives and false negatives in the

choice of the scenario were estimated. False positives and negatives were calculated by measuring the proportion of the simulated data set for the best scenario when assigned to the other scenario and the proportion of data sets simulated under the other scenario that was assigned to the best scenario, respectively. Second, the goodness of fit of the selected scenario was performed using DIYABC model-checking analysis. This assesses how close the simulated data set fits our GBS data set by producing summary statistics and ranking the observed value among the values obtained with simulated data sets.

2.3.6 Patterson's D Statistics

Patterson's D-statistics (also known as ABBA-BABA tests; Durand et al., 2011) is used to infer introgression between two lineages in a given phylogeny by testing for shared derived alleles between taxa. ABBA-BABA tests are strictly a tree-based test from which species are assigned to the topology (((P1, P2), P3), O). P1 and P2 are species that belong to a monophyletic group, P3 corresponds to a taxon from a different ingroup species, and O is the outgroup. The D-statistic measures the asymmetry in the number of alleles supporting "ABBA" and "BABA" patterns of allele distributions (where A is ancestral, and B is derived). If the proportion of alleles is equal (ABBA = BABA), then we cannot rule out ILS as the likely cause of gene incongruence. In contrast, a significant asymmetry between the number of alleles for each topology suggests introgression has occurred between P3 and either P2 (ABBA > BABA) or P1 (BABA > ABBA).

We used Patterson's D-statistic to test whether there is an excess of shared allele between *D. millefolia* and *D. gilva*, as indicated in our species network (see results). Due to the tree's topology, it was impossible to test whether there is gene flow between *D. gonzalezii* and *D. gilva,* and *D. millefolia.*

SNPs from the c85m8 assembly were used to generate the D-statistic using the *baba* tool in ipyrad v.0.9.16 (Eaton and Overcast, 2020). A P value was calculated from the Z-score and adjusted for multiple comparisons using Bonferroni correction.

2.3.7 Ecological Niche Occupation and Conservatisms

GPS coordinates were taken from 216 Canary Island *Descurainia* herbarium specimens held at the Natural History Museum, London (BM). Climatic variables for these localities were obtained from WorldClim v.2.1 (years 1970-2000; Fick and Hijmans, 2017), using 19 bioclimatic (bioclim) variables derived from temperature and rainfall values at a spatial resolution of 30 seconds (~1 km^2). The GIS datasets were used to interpolate climate values for all individuals using ArcMap, and the means for each bioclimatic variable per species were calculated. The *pairs* function in R was used to identify correlated variables. Variables were assessed for correlation in R and only uncorrelated variables were retained for further analysis. A phylogenetic PCA (pPCA) was generated with the R package Phytools (Revell, 2012), using the phylogeny generated from the analysis of the concatenated data set and the five bioclim (annual precipitation, temperature seasonality, annual precipitation, precipitation in the wettest quarter, and precipitation in the coldest quarter) variables retained. The averages for each climatic variable were represented in a dotTree using Phytools, along with the results of the phylogeny PCA (PC1 and PC2).

We tested for phylogenetic signal in each trait using Blomberg's K statistic (Blomberg et al., 2003), as implemented with the function *phylosig* in the R package Phytools. K values closer to zero correspond to random or convergent patterns of evolution, while values greater than 1 indicate a strong phylogenetic signal and conservatism of traits.

2.4 Results

2.4.1 GBS Assembly

Between 0.46 and 2.14 million reads were generated for each sample. After quality control and the removal of poor-quality reads, six data sets were assembled from the GBS raw reads (Table 2.1). Between 113,107 and 122,724 loci were generated in each assembly (Table 2.1). An increase in the clustering threshold (c; 80%, 85%, and 90%) resulted in a larger number of loci. Increasing the minimum sample coverage (m) from 8 to 10 resulted in a considerable decrease in the number of total SNPs and Parsimony informative (PI) SNPs available (Table 2.1).

Table 2.1: Summary of the results of filtering and clustering of GBS raw sequences using ipyrad (Eaton, 2014). GBS reads included 18 individuals from seven Canary Island *Descurainia* species and two continental *Descurainia* individuals. Loci = unique GBS DNA cluster; PI = Parsimony informative; SNPs = PI SNPs and autapomorphies; PI SNPs = minor allele in >1 sample; PI uSNPs = unlinked PI SNP

Assembly	Clustering Threshold (%)	Minimum Taxon Coverage	N loci	SNPs	PI SNPs	PI uSNPs
c80m8	80	8	113,107	69,339	29,037	12,779
c80m10	80	10	113,107	41,235	18,002	7,201
c85m8	85	8	118,553	75,603	31,714	13,797
c85m10	85	10	118,553	45,347	19,794	7,828
c90m8	90	8	122,724	76,875	32,488	14,652
c90m10	90	10	122,724	45,526	19,981	8,174

2.4.2 Phylogenetic Reconstructions 1 – Concatenated Approach

The GTR+I+G model of nucleotide evolution was found to be the best model for all six assemblies (Appendix S2) and was employed for maximum likelihood (ML) phylogeny reconstruction.

All phylogenies resolved two main clades (BS = 100%), “clade A” and “clade B” (Figure 2.2A, Appendix Figure A.1). Clade A is composed of *D. preauxiana*, *D. millefolia* and *D. artemisioides,*

and in five of the six phylogenies (i.e., excluding c90m8), *D. millefolia* is resolved as monophyletic (BS = 54-82%). *Descurainia millefolia* is sister to *D. artemisioides*, with *D. preauxiana* resolved as sister to this pair. Within *D. millefolia*, Tenerife populations are resolved as paraphyletic with respect to the La Palma accessions that are resolved as monophyletic (BS = 100%). For the data set with the alternate topology (c90m8), *D. millefolia* is resolved as polyphyletic with one individual nested within *D. artemisioides* (BS = 48%).

The clade B topology was consistent for all six assemblies and is composed of subclades B1 (*D. gilva* and *D. gonzalezii*) and B2 (*D. bourgaeana* and *D. lemsii*), both of which have maximum support (BS = 100%). In B1, both *D. gonzalezii* and *D. gilva* are resolved as monophyletic (BS > 96%). Within Clade B, *D. lemsii* is found nested within *D. bourgaeana*, although the position is variable and weakly supported (BS = 42-79%).

2.4.3 Phylogenetic Reconstructions 2 – Coalescent Approach

A total of 13,797 unlinked SNPs were used as input data for Phylonet. The phylogenetic network with the lowest log-likelihood score from all 50 runs resulted in a network with one reticulation (Figure 2.2B). The resulting topology was similar to that of the concatenated approached, but *D. gilva* is suggested to be of hybrid origin between *D. gonzalezii* (clade B) and *D. millefolia* (clade A).

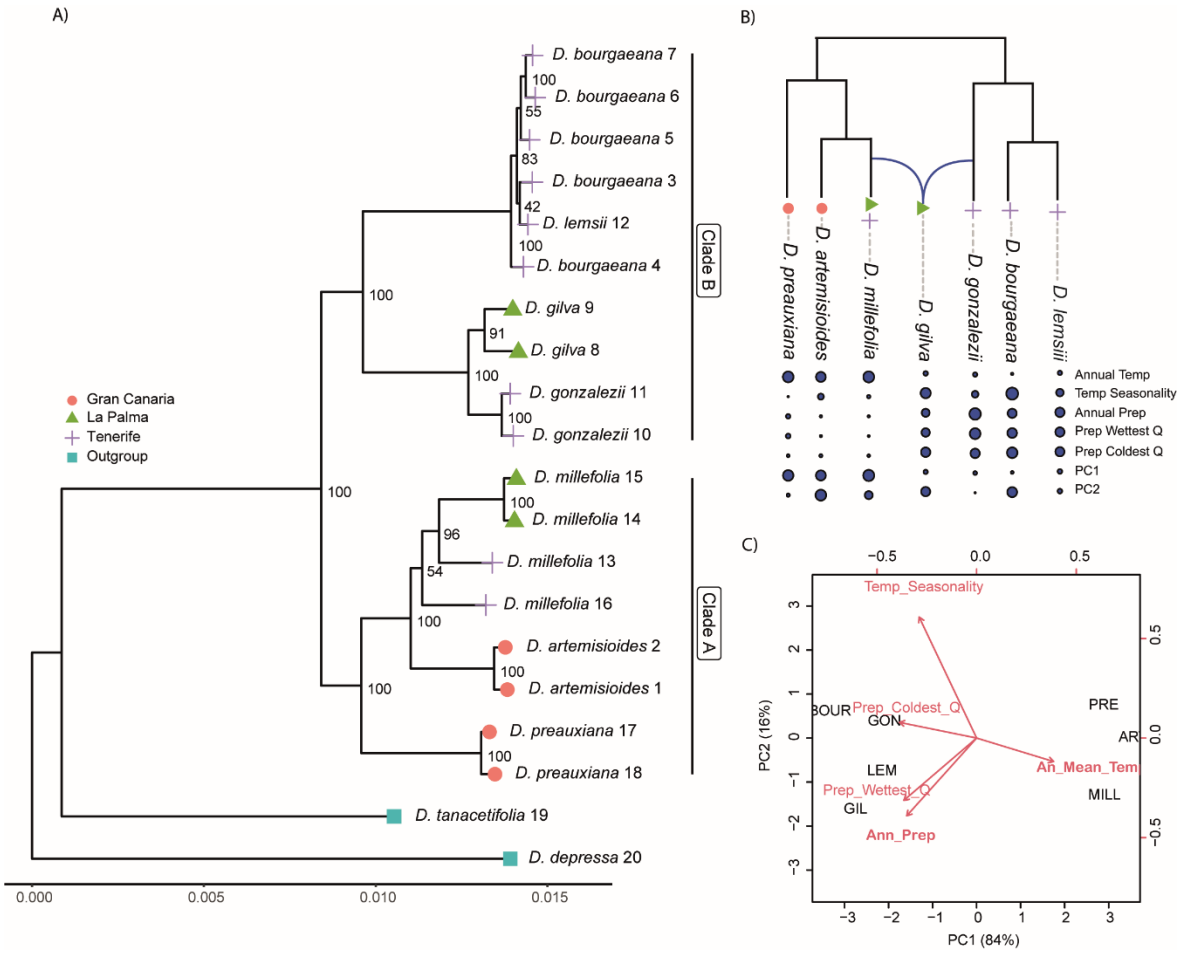


Figure 2.2: (A) Maximum-likelihood phylogenetic tree inferred from GBS data set (c85m8) for 18 Canary Island *Descurainia* individuals and two continental relatives. Numbers represent bootstrap values as inferred from 1,000 bootstraps (BS) repetitions. Branch lengths are represented by the bottom bar. The continental relatives, *D. depressa* and *D. tanacetifolia,* are the outgroups. ID numbers next to taxa names represent samples in Appendix Table A.1, and shapes represent the island of origin. (B) Schematic representing the phylogenetic network generated by Phylonet under Pseudo-likelihood when max reticulates is set to eight. Dots represent the average value of the bioclimatic variable and pPCA loads for each species (C) Phylogenetic principal component analysis (pPCA) plots of PC1 and PC2 from five bioclimatic variables. Arrows in bold indicate a significant phylogenetic signal ($K > 1$ at $P < 0.05$).

2.4.4 Approximate Bayesian Computation (ABC)

The c85m8 SNP data set was filtered for at least one individual per species at each SNP site (as required by DIYABC), after which 1,722 SNPs remained. Scenario 2, in which *D. gilva* is of hybrid origin between *D. millefolia* and *D. gonzalezii*, had the highest posterior probability (0.77, 95% confidence interval 0.43 – 1.00; Table 2). Scenario 1, which represented the bifurcating phylogeny without hybridisation, had a much lower posterior probability (0.21; confidence interval 0.00 – 0.57).

The summary statistics calculated from the scenario 2 simulated data set were found to be close to the summary statistics calculated with our GBS data set (Appendix Table A.3). Estimation of error rates provided confidence in scenario choice; the false positive and false negative error rates were estimated to be 10.2% and 0.1%, respectively.

Our analysis also indicated that the parental contributions to the genetic composition of the hybrid *D. gilva* were not equal (Appendix Table A.3). There was a significant bias in contribution, with *D. gonzalezii* contributing 98% (95% CI: 98.4-99.6) and *D. millefolia* 2%. The parameter estimation for the time of this hybridisation event was 3.7 ± 2.1 Mya.

Table 2.2: Posterior probability of parameters, and their confidence intervals, for each four scenarios tested with ABC in DIYABC (Cornuet et al., 2008). The corresponding phylogeny or network for each scenario is found in Figure 2.2.

Scenario	Test Description	Posterior Probability	95% Confidence Interval
1	Bifurcating phylogeny	0.214	[0.000, 0.574]
2	*D. gilva* is a hybrid origin between parents *D. millefolia* and *D. gonzalezii*.	0.786	[0.427, 1.000]

2.4.5 Patterson's D-statistics

Patterson's D-statistic examining gene flow between *D. gilva* (P3) and *D. millefolia* (P2) resulted in a significant D-statistic (BABA > ABBA) when both *D. artemisioides* and *D. preauxiana* were used as P1 (Table 2.3; $P < 0.005$, after multiple corrections).

Table 2.3: Patterson's D-statistic (ABBA-BABA), which compares *D. gilva* (gil; P3) with *D. millefolia* (mil; P2) and its sister taxa (P1), either *D. artemisioides* (art) or *D. preauxiana* (pre). *Descurainia tanacetifolia* and *D. depressa* were selected as the outgroups (O). D-statistics in bold represent significant P values at < 0.05.

Test	D stat	Z	ABBA	BABA	# loci	P value
art (P1) mil (P2) gil (P3)	-0.121	3.929	1031.256	1313.998	5,817	**0.000**
pre (P1) mil (P2) gil (P3)	-0.132	3.221	610.708	796	3,530	**0.001**

2.4.6 Ecological Niche Occupation and Conservatism

After filtering the most highly correlated variables, five BioClim variables were included in the ecological analyses: Annual Temperature, Temperature Seasonality, Annual Precipitation, Precipitation in the Wettest Quarter, and Precipitation in the Coldest Quarter (Appendix Table A.4). The phylogenetically corrected PCA's first two axes, PC1 and PC2, explained 84% and 16% of the variance in climatic variables across the seven species (Table 2.4). Precipitation in the Coldest

Quarter (-0.99) and Annual Mean Temperature (0.98) have the highest loads for PC1. *Descurainia bourgaeana* (-3.45) was on one extreme of PC1, followed by *D. gilva*, *D. lemsii* and *D. gonzalezii,* all demonstrating high average precipitation, whereas *D. artemisioides* (3.44), *D. millefolia* (2.77) and *D. preauxiana* (2.77) are on the opposite end of the PC1 axis and represent higher annual mean temperature (Figure 2.2B). Temperature seasonality (0.67) and annual precipitation (-0.43) represented the highest loads for PC2, with *D. gilva* (-1.59) and *D. preauxiana* (0.72) on the extremes of this axis.

The K statistic indicated a strong phylogenetic signal for all bioclimatic variables ($K > 1$; Table 2.4). Annual mean temperature, annual precipitation and precipitation in the wettest quarter were significant ($P < 0.05$), indicating these traits have a non-random phylogenetic signal.

Table 2.4: The loadings of each trait for the first two axes from the pPCA and K statistic under a Brownian motion model of evolution for our five bioclimatic variables. Traits which are significant (bold) where P < 0.05 have a non-random phylogenetic signal. K-values greater than 1 indicate a strong phylogenetic signal and phylogenetically conservation.

Bio Variable	PC1 (84%)	PC2 (16%)	K	K P value
Annual Mean Temp (BIO1)	0.989	-0.131	1.365	**0.045**
Temp Seasonality (BIO4)	-0.738	0.672	1.447	**0.037**
Annual Precipitation (BIO12)	-0.898	-0.432	1.150	0.116
Precipitation in Wettest Quarter (BIO16)	-0.935	-0.346	1.395	**0.041**
Precipitation in Coldest Quarter (BIO19)	-0.992	0.087	1.530	0.053

2.5 Discussion

The Genotyping-by-Sequencing (GBS) analysis has provided a well-resolved phylogenetic relationship among Canary Island *Descurainia* (Brassicaceae), for which standard sequence markers previously failed (e.g., Goodson et al., 2006) and revealed evolutionary processes underpinning the group's diversification. Phylogenetic analyses of the concatenated data set revealed two major and well-resolved clades, with ABBA-BABA tests and species networks supporting a role for hybridization in the group's diversification. Our results suggest that *D. gilva* has a hybrid origin between *D. millefolia* and *D. gonzalezii*. Further, the results indicate that diversification of the group has involved a single shift in climatic preferences associated with a shift in elevation, with the low-elevation Clade A species occupying high temperature and low precipitation environments and the higher-elevation Clade B species occupying lower temperature and higher precipitation environments. The study, therefore, demonstrates the importance of geographic isolation, hybridisation, and ecological diversification in driving the diversification of Canary Island *Descurainia*.

2.6 Inferring Species Relationships Using Coalescent versus Concatenated Methods

Previous studies have suggested that phylogenies based on concatenated GBS data sets suffer strong systematic bias because high statistical support for multiple alternative topologies depending on clustering threshold values used in the sequence analysis is often found (Fernández-Mazuecos et al., 2018). Here our GBS assemblies show congruent and well-supported topologies across five of six data sets. The incongruent data set differed only in the resolution of *D. millefolia.* Phylogenetic analysis of all concatenated data sets revealed two major clades (A and B) and two sub-clades within B (B1 and B2) with maximum bootstrap support. This suggests the different clustering and minimum sample coverage thresholds used have little impact on phylogenetic inferences, providing confidence in the results.

In the topology resolved in five of the six analyses, *D. bourgaeana* was the only species revealed as non-monophyletic, with *D. lemsii* nested within this species. However, branch lengths in this clade were short, and relationships were not well supported. The occurrence of paraphyly such as

this is often explained by taxonomic error, ILS, or hybridisation (Schrempf and Szöllösi, 2020). Morphologically, *D. bourgaeana* and *D. lemsii* may be distinguished since the former has decurrent leaves and ascending siliquae, whereas the latter lacks decurrent leaf-segments and has erect siliquae (Bramwell, 1977). However, hybridisation has played a role in Canary Island *Descurainia* more generally (see below), and putative hybrids of *D. bourgaeana* and *D. lemsii* have been observed in the wild (Goodson, 2007; ACJ personal field observations). Due to our phylogeny's topology, it was impossible to test for hybridisation (versus ILS) involving these two taxa with Patterson's D-statistics. Given that neither ILS nor hybridisation can be ruled out, further morphological and genetic studies, with a greater sampling of *D. lemsii*, are necessary to resolve the status of these two taxa.

The topology of our concatenated phylogeny was incongruent with the previous phylogenetic reconstructions of Goodson et al. (2006) based on 44 cpDNA characters. In particular, our analysis did not support the polyphyly of *D. gonzalezii* and *D. millefolia* found by Goodson et al. (2006), wherein both were resolved as monophyletic. The Goodson et al. (2006) phylogeny may be impacted by chloroplast (cp) capture (the transfer of cpDNA between species due to hybridisation). Given that our results indicate hybridisation in the evolutionary history of the Canary Islands *Descurainia*, cpDNA capture likely explains the significant differences between the results of Goodson et al. (2006) and those presented here. Further, our phylogenetic reconstructions were generated using a significantly larger number of genetic markers due to our use of our GBS datasets generated through next-generation sequencing technology. For instance, our ML phylogeny was generated using 76,875 SNPs (c85m8 GBS assembly; Figure 2.2), which significantly exceeds the 44 cpDNA characters employed in the Goodson et al. (2006) study. Our datasets provide a far more powerful phylogenetic signal to resolve the relationships within Canary Island *Descurainia* and likely explains the conflicting topologies between studies.

Evolutionary networks have recently been explored as an alternative method for phylogenetic reconstructions (McCluskey and Postlethwait, 2015; Blanco-Pastor et al., 2019). Since reduced-representation data sets (i.e., RADseq and GBS) sample data from hundreds to tens of thousands of loci across the entire genome, these sequencing methods can be applied to generate multilocus species trees. While Phylonet is a promising tool for estimating hybridisation between taxa (Hibbins and Hahn, 2022), few studies have attempted to use reduced-representation data sets to estimate evolutionary networks (also see Blanco-Pastor et al., 2019; Eaton and Ree, 2013). Several issues have been highlighted with this approach; notably, the short length of each locus

means that there is variable phylogenetic information between loci, and the data sets may be subjected to phylogenetically structured patterns of missing data (Salichos and Rokas, 2013b). However, a coalescent analysis of reduced-representation data allows for variation across loci in a genealogy, whereas a concatenated approach assumes a shared phylogenetic history for all genes (Rokas et al., 2003). As a result, a coalescent approach can be useful for understanding hybridisation.

Here, the non-reticulated species network demonstrated a similar topology to that of the concatenated analysis. However, we found strong evidence that Canary Island *Descurainia*'s most likely evolutionary scenario includes reticulated evolution. The reticulated phylogenies generated by Phylonet were also supported by D-statistics and DIYABC models. This implies that the concatenated approach, despite the larger number of loci included and the high bootstrap support, has masked incongruence between loci within the data set.

Our analyses support a hybrid origin for *D. gilva* between *D. gonzalezii* and *D. millefolia*. This is at odds with the hypothesis of Bramwell (1977), who, based on morphological similarities, considered *D. gilva* to be a vicariant of *D. bourgaeana*, a species which also occupies high altitudes.

Our ABC analysis indicated a large bias in putative parental contributions, with 98% contribution from *D. gonzalezii* and 2% from *D. millefolia.* Inter-island hybridisation has been documented in other Canary island endemic plant lineages (van Hengstum et al., 2012; Puppo et al., 2015; White et al., 2020). However, it seems most likely that *D. gilva* is instead a vicariant form of *D. gonzalezii,* which has been exposed to gene flow from *D. millefolia* upon establishing itself on La Palma. Notably, *D. gilva and D. gonzalezii* are somewhat ecologically distinct: *D. gilva* habitats exhibit an Annual Precipitation and Precipitation in the Wettest Quarter of 241 and 215 mm, respectively, 86% and 90% higher than the habitats of *D. gonzalezii* (Figure 2.2B; Appendix Table A.4).

2.7 Diversification and the role of Ecology and Geography

Inter-island dispersal and ecological shifts are the two processes most commonly invoked to explain the high levels of diversity seen within Canary Island lineages (Jorgensen and Frydenberg, 1999; Francisco-Ortega et al., 2002; Mort et al., 2002; Allan et al., 2004; Jones et al., 2014). Previous phylogenetic analysis of Canary Island *Descurainia* suggested that inter-island dispersal was more influential than ecological speciation (Goodson et al., 2006), with three ecological shifts inferred alongside multiple inter-island colonisation events. Our phylogeny also supports the importance of geographic speciation in the evolution of Canary Island *Descurainia*. Within clade B1, there is a dispersal event from Tenerife to La Palma for *D. gilva,* and in clade A, there are likely two dispersal events, one from Gran Canaria to Tenerife, followed by another from Tenerife to La Palma.

Climatic variation within the Canary Islands has often been seen as a major driver of diversification within endemic plant lineages (Irl et al., 2015, 2017). Climatic variation, especially along elevation gradients, promotes ecological shifts and speciation (Hua and Wiens, 2013). For example, precipitation gradients have played a significant role in the radiation of *Aeonium* (Crassulaceae; Harter et al., 2015). Our results indicate that precipitation and temperature are strongly linked phylogenetically for *Descurainia*. We observe only one shift in climatic preferences, supporting the idea of niche conservatism between closely related *Descurainia* species. Taxa in each of our two clades occupy similar ecological zones. *Descurainia gonzalezii, D. lemsii* and *D. bourgaeana* in clade B are species that inhabit Tenerife with parallel environmental pPCA axes. They are mainly allopatric, suggesting intra-island geographic diversification without large ecological divergences.

Similarly, *D. artemisioides* and *D. preauxiana,* endemic to Gran Canaria, occupy similar habitats but are largely allopatric: *D. artemisioides* is restricted to the northwest of the island and *D. preauxiana* is more widespread, but they do not overlap geographically. Inter- and intra-island isolation within similar habitats has played a more significant role in the diversification of the group than ecological shifts based on the data to hand. However, we acknowledge that other factors, for example, wind exposure and soil types, should be examined.

The phylogenetic framework of Goodson et al. (2006) has been used to investigate the role of adaptive divergence in the evolution of Canary Island *Descurainia.* Steinbauer et al. (2016) examined climatic niche differentiation between pairs of species existing in sympatry with that for pairs of species in allopatry for a range of Canary Island radiations, including *Descurainia*. They concluded that *Descurainia* showed niche conservatism for temperature. In contrast, Herben et al. (2014) proposed that morphological traits related to water availability were not phylogenetically linked and therefore suggested adaptive divergence to differing habitats based on water availability. However, both these studies used the Goodson et al. (2006) phylogenetic framework, which conflicts with our phylogeny and could underlie the differences between studies.

Whilst many Macaronesian plant radiations have been subject to molecular phylogenetic analysis, many would benefit from further analyses with more extensive molecular sampling. Where reduced representation sequencing approaches have been used, for example, to investigate Macaronesian endemics *Argyranthemum*, *Micromeria* and *Lavatera* (Curto et al., 2018; Villa-Machío et al., 2020; White et al., 2020) better-resolved phylogenies have resulted, and more complex patterns of relationships than earlier assessments suggested have been revealed.

Our results suggest that ecological shifts are not absent from the diversification of Canary Island *Descurainia*. A significant ecological distinction is found between the two clades. Our pPCA results show that species from clade A are found in warm and arid habitats, and species from clade B are found in cooler, wetter habitats. Upslope migrations, involving adaptation to higher altitudes, are a common observation in Canary Island lineages, for example, in *Helianthemum* (Cistaceae) and *Echium* (Albaladejo et al., 2021; Graham et al., 2021). This fits with the finding that the initial colonisation of the Canary Islands is often at low altitudes where more diverse habitats are available, followed by diversification upslope to more specialised high-altitude habitats (Steinbauer et al., 2012). However, in *Descurainia*, the sister group of the Canary Islands clade, *D. tanacetifolia* (Goodson et al., 2006), is a species restricted to montane regions of northern Iberia and the Alps. Given our data, we cannot rule out 'downslope' migration from cooler, wetter habitats to warmer, drier habitats. A broader analysis of the evolution of climatic preferences in the genus would be necessary to provide insights into this.

2.8 Conclusions

While our phylogeny would benefit from a greater sampling of individuals, particularly with respect to *D. lemsii* and La Gomera populations of *D. millefolia*, our phylogenetic reconstructions utilising GBS sequencing generated a well-resolved and robust evolutionary relationship. Our new multilocus framework for Canary Island *Descurainia* indicates that hybridisation has occurred within this group, likely explaining species polyphyly identified in an earlier study (Goodson et al., 2006). Island dispersal, and to a lesser extent, ecological shifts, and hybridisation, are all implicated in the diversification of Canary Island *Descurainia*. The findings of this study reinforce the notion that hybridisation is one of the drivers contributing to the complex history of Canary Island flora. It also highlights how phylogenetic analyses of island lineages should employ multiple analytical approaches to test for alternative scenarios.

2.9 Acknowledgements

The authors thank Rachael Graham and Alfredo Reyes-Bentacort for the Canary Island field sampling. We also thank Barbara Goodson and Bob Jansen (University of Texas) for providing DNA samples and the Banco de Germoplasma Vegetal (BGV) Seedbank for providing *Descurainia* seeds. We are grateful to the national parks for permits to collect. Permit numbers are as follows: Cabildo de Tenerife (no. 22835), Parque Nacional del Teide (no. 24339), Cabildo de La Palma (no. 2016005709), Parque Nacional de la Caldera de Taburiente (no. 671303), Parque Nacional Garajonay (no. 25771), Cabildo de Gran Canaria (no. 16267), Gobierno de Canarias (no. 671303). This work was supported by a PhD studentship to ACJ from the Natural Environmental Research Council [grant number NE/L002531/1]. We acknowledge the use of the IRIDIS High-Performance Computing Facility and associated support services at the University of Southampton in the completion of this work.

2.10 Author Contributions

ACJ, MCarine and MChapman designed the study; MChapman extracted DNA, OWW assembled the raw GBS data with ACJ performing the rest of the analyses; ACJ wrote the manuscript with contributions from all authors.

2.11 Data Accessibility

Climatic data for Canary Island *Descurainia* individuals and raw Genotyping-by-sequencing data will be deposited on Dyrad.

Chapter 3 The genome assembly and annotation of the oceanic island endemic *Descurainia millefolia* (Brassicaceae) and a comparison with other Brassicaceae

3.1 Abstract

Descurainia millefolia (Brassicaceae) is endemic to the Canary Islands oceanic archipelago and belongs to a group that has rapidly diversified across habitats and islands. As such, there is potential for *D. millefolia* to serve as a model to address questions on the underlying evolutionary processes that have led to the large endemic diversity found in the Canary Islands. There is also potential to contribute to answering larger questions surrounding adaptation to climate change. Here, we produce a high-quality *de novo* scaffold-level assembly of the *D. millefolia* genome using a hybrid assembly approach with a combination of long-read PacBio sequencing and short-read Illumina sequencing. The final draft genome is 180 Mb, with an N50 size of 280 kb. We annotated 26,407 protein-coding genes with a BUSCO score (complete and single copy) of 97.6%. Phylogenetic analyses using proteomes of 13 other Brassicaceae species showed that the genera *Descurainia* form a monophyletic group within Lineage I of the Brassicaceae family, with a divergence estimate of 20.12 Mya (95% CI: 19.04-21.48) from other sampled Brassicaceae species. Gene family analysis indicated 96 gene families specific to *D. millefolia* with 1,087 and 1,469 gene expansions and contractions, respectively. We hope the genomic resources provided by our reference genome will enable future studies to investigate evolutionary and taxonomic questions.

3.2 Introduction

The flora of oceanic islands has long served as an evolutionary model for understanding the drivers of speciation and phenotypic diversification (Burns, 2019). Many of the endemic plant groups has rapidly diversified through various evolutionary processes, including dispersal events between islands and adaptation to a broad spectrum of ecological niches (Losos and Ricklefs, 2009a). Extensively studied examples of oceanic island radiations include the Hawaiian silverswords (Witter and Carr, 1988; Robichaux et al., 1990; Baldwin, 1997; Baldwin and Sanderson, 1998, 1998; Blonder et al., 2016), and the Canary Island *Aeonium* (Crassulaceae) (Jorgensen and Frydenberg, 1999; Jorgensen and Olesen, 2001). Yet, the lack of whole-genome resources often hinders our understanding of the underlying genome evolution of island groups and its ecomorphological diversification processes during speciation and adaptive radiation.

In 2000, the first genome of a flowering plant was sequenced: *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000). Since then, genome sequencing has been focused on economically important and model species of plants. As of 2021, there are 798 plant genomes (deposited within GenBank), of which 59% represent domesticated, relatives of domesticates, and model species, with the remaining being wild species (41%; Marks et al., 2021; Purugganan and Jackson, 2021). The underrepresentation of wild plants is primarily due to issues surrounding the sequencing of their genomes: firstly, the size of plant genomes – which range from ca. 10 Mb to over 100 Gb - provides financial constraints (Kersey, 2019). Secondly, complex characteristics of plant genomes, such as polyploidy (exhibited in 80% of all living plants; Meyers and Levin, 2006) and a large composition of transposable elements, leads to difficulties in bioinformatic processes.

For instance, the Brassicaceae (mustard) family comprises almost 4,000 species in 351 genera (Al-Shehbaz et al., 2006). However, much of the existing genome sequencing in this family has focused on economically important crops (e.g., *Brassica oleracea* and *Brassica napus*) and the model organism *Arabidopsis*. To date, 26 Brassicaceae species have had their genomes sequenced (Chen et al., 2022). Wild species within Brassicaceae are largely underrepresented, resulting in an untapped genomic resource for evolutionary studies.

Descurainia (commonly known as “Tansy Mustard”), within the family Brassicaceae, is an under-studied group comprising of ca. 45 species distributed through the Old and New World (Bramwell,

1977; Goodson et al., 2011). Taxa within *Descurainia* are primarily found in North and South America, with a smaller distribution in Europe and the Canary Islands. *Descurainia* is similar to most Brassicaceae species, with its characteristic small yellow flowers and long erect stems. The few phylogenetic studies utilising *Descurainia* taxa have consistently resolved the genus within 'Lineage I' of the family, a lineage known for comprising well-studied genera, namely *Arabidopsis* and *Capsella* (Nikolov et al., 2019; Walden et al., 2020; Guo et al., 2021; Liu et al., 2021). These studies suggest that the closest extant relatives of *Descurainia* are *Smelowskia* (native to the mountainous and arctic regions of Asia and North America) and *Yinshania* (native to China and Vietnam). Divergence estimates suggest a Crown group age of 14.75 million years (My; 95% CI: 10.65-18.99) and 17.0 My (95% CI: 12.95-21.46) from the stem group age (Walden et al., 2020).

A particular clade within *Descurainia* that has received attention within evolutionary research is the seven species endemic to the Canary Islands, an archipelago of seven islands found 100 km west of Africa. The seven species include *Descurainia bourgaeana* (Webb ex Christ) O.E. Schulz*, D. lemsii* Bramwell*, D. gilva* Svent.*, D. gonzalezii* Svent.*, D. millefolia* (Jacq.) Webb & Berth.*, D. artemisioides* Svent., and *D. preauxiana* (Webb) I.E. Schulz. The Canary Island *Descurainia* form a monophyletic group, the ancestor of which colonised the islands from continental Spain ca. 7 Mya (Bramwell, 1977; Goodson et al., 2006). Following initial colonisation, *Descurainia* has rapidly diversified into different habitats, covering xero-thermophilus scrub (250-600 m), pine forests (600-2,000 m), and sub-alpine scrub (2,000-2,300 m; Goodson et al., 2006). Previous research has demonstrated these seven species have diversified through ecological and geographic shifts between and within islands, as well as evidence of hybridisation (Chapter 2; Goodson et al., 2006). These evolutionary processes are commonly used to explain the high levels of endemic diversity (over 680 endemic taxa) in the Canary Islands (Reyes-Betancort et al., 2008).

Here, we aim to generate a reference genome for *Descurainia millefolia* (a multi-island endemic distributed across Tenerife, La Palma, and La Gomera; Figure 3.1). Island plant radiations, such as Canary Island *Descurainia*, provide an opportunity to understand evolutionary processes, yet, representative genome sequences are scarce (Choi et al., 2021; Bellinger et al., 2022; Cerca et al., 2022). The few existing examples have demonstrated their potential to provide insights into plant evolution. For instance, the genome of *Metrosideros* (Myrtaceae)*,* an endemic plant group to the Hawaiian Islands, provided evidence that ancestral variation drove the adaptive divergence of lineages (Choi et al., 2021), and the genome of Darwin's giant daisy (*Scalesia atractyloides*; Galápagos islands) also revealed genomic insights into the island syndrome (Cerca et al., 2022).

We aim to apply whole-genome sequencing techniques to establish a high-quality reference genome of *D. millefolia.* We used long-read Pacific Biosciences (PacBio) and short-read Illumina sequencing for a hybrid assembly approach to achieve a highly contiguous and complete genome. A hybrid assembly approach is beneficial as the long reads can resolve extensive repeating sequencing that is often exhibited in plant genomes. Due to the close taxonomic relationship of *Descurainia* with well-studied organisms within Brassicaceae (notably *Arabidopsis*), we utilise these well-established and overwhelming available genomic resources to facilitate annotation with high completeness. As this is the first examination of the *Descurainia* genome, we apply comparative phylogenetic approaches to understand relationships within Brassicaceae and shed light on *Descurainia* genomic characteristics. By generating a high-quality reference genome for an island endemic radiation, we hope it will serve as a genomic resource and model for understanding the evolutionary genomics of rapid plant diversification.

Figure 3.1: *Descurainia millefolia* on the cliffs of Chinamada, Tenerife (220 m). © Photos: ACJ.

3.3 Methods

3.3.1 Sample collection and DNA extraction

To obtain DNA for whole-genome sequencing, tissue was collected from a *Descurainia millefolia* plant cultivated within the School of Biological Sciences, University of Southampton greenhouse facilities. The plant was grown from seed material obtained from the UPM (Madrid, Spain) seed bank (seed accession number: ETSIA-246-1073-67). Fresh leaf tissue was frozen in liquid nitrogen, and a CTAB DNA extraction protocol (Porebski et al., 1997) was performed to obtain isolated DNA. The extracted DNA was treated with RNase and assessed for molecular weight and concentration.

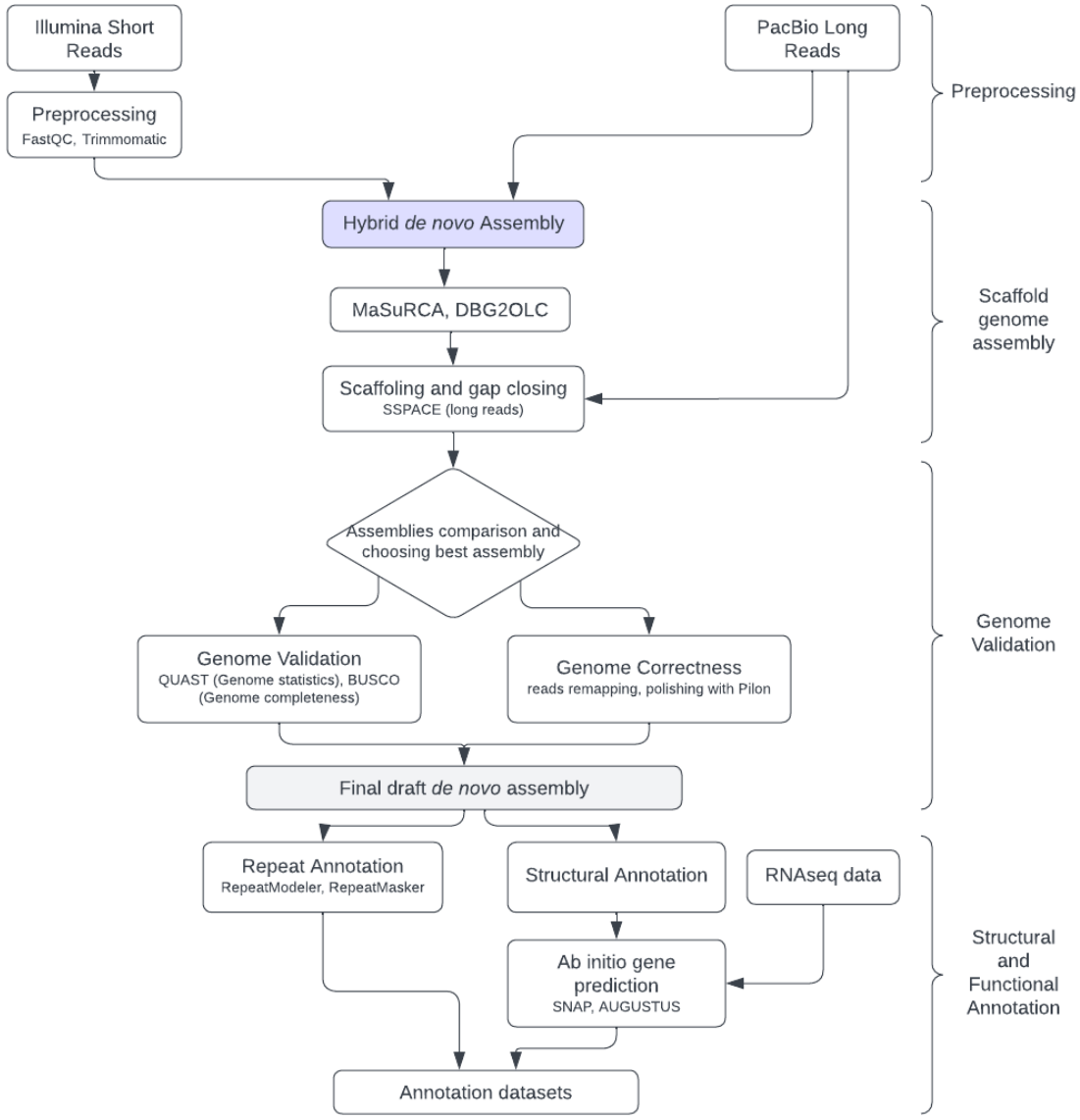


Figure 3.2: Schematic and overview of the workflow assembly and annotation process of the *Descurainia millefolia* genome.

3.3.2 Reference assembly

The draft *de novo* genome of *Descurainia millefolia* was generated using a hybrid assembly approach that combines short- (Next-Generation sequencing; NGS) and long-read (Third Generation Sequencing; 3GS) sequencing libraries (Figure 3.2). Hybrid assembly approaches involved supplementing the accurate NGS short-read sequencing data with the more error-prone 3GS long-read sequences to resolve a large and repetitive genome.

Isolated high-molecular-weight DNA was used to prepare whole genome libraries at NovoGene (Hong Kong) on a Pacific Biosciences Sequel with a 1.0 chemistry platform for long-read sequencing and another with Illumina HiSeq 3000 platform for 150-bp paired-end short-read sequencing. The raw data for short-read sequences were quality controlled through FastQC v.3.0 (Andrews, 2010) and filtered and trimmed using Trimmomatic v.0.36 (Bolger et al., 2014). Due to the noise given by the PacBio long-reads, trimming was not applicable.

To estimate the genome size for *Descurainia millefolia*, the most optimal k-mer value for both libraries (short- and long-read) was calculated by passing reads through KmerGenie v.1.7 (Chikhi and Medvedev, 2014) and compared to published *Descurainia millefolia* C-value (Lysak et al., 2009).

Two hybrid assembly software, DBG2OLC (https://github.com/yechengxi/DBG2OLC; Ye et al., 2016) and MaSuRCA v.3.3.0 (Zimin et al., 2013), were selected to ensure the best quality draft assembly possible. The purpose-built NGS and 3GS hybrid assembler, DBG2OLC, involves three main steps to generate a *de novo* hybrid assembly. First, *de Bruijn* graph (DBG) contigs from Illumina reads were created using Sparse Assembler (github.com/yechengxi/SparseAssembler), with k = 119 (based on previous k-mer estimation; see results). Second, DBG2OLC was used to map the DBG contigs generated from the previous step to the long-reads to create a backbone assembly (adaptive threshold of 1×10^{-4}). Finally, the last step uses a PacBio long-read aligner, Blasr (github.com/PacificBiosciences/blasr), to create a consensus assembly from the previously generated DBG contigs, long-reads, and the backbone assembly.

MaSuRCA can employ a hybrid assembly approach and can handle repetitive plant genomes better. Here, the k-mer size was estimated by MaSuRCA from read error rate and coverage, giving a value of 99. Due to the relatively short read length of our PacBio reads (7,086 average bp length), the coverage parameter was set at 25, 30, and 35 to find the most optimal assembly. This assembly was polished using POLCA (Zimin and Salzberg, 2020) to improve consensus accuracy.

3.3.3 Assembly validation

Genome quality for each of our assemblies was assessed using standard assembly statistics generated from QUAST (Gurevich et al., 2013), which includes N50 and contig length. In addition, Benchmarking Universal Single-Copy Orthologs v.3.0 (BUSCO; Simão et al., 2015) was used to assess genome completeness by identifying single-copy, duplicated, fragmented, or missing BUSCOs. The resulting contigs from the final draft assembly genome were checked for contamination using BLASTn (Altschul et al., 1990) against the NCBI nucleotide database. Identified contaminated contigs were removed using Picard v.2.8.3 (https://broadinstitute.github.io/picard).

The contigs from the chosen final assembly were scaffolded with our PacBio long-reads using SSPACE Long Read v.1.1 (Boetzer and Pirovano, 2014). Gap filling was performed with our Illumina short reads using SSPACE Gap Filler v.1.1. Both of our libraries of trimmed reads were mapped back to the reference genome using BWA-MEM v.0.7.11 (Li and Durbin, 2009).

3.3.4 Repeat annotation

RepeatModeler v.1.0.11 (Smith and Hubley, 2008) was used for *de novo* identification of repetitive sequences. The classified repeats were appended to the *A. thaliana* repeat library obtained from the RepBase database (2017 release), resulting in a final repeat library. This library was used to annotate repeats and transposable elements (TEs) within the scaffolds with RepeatMasker v.4.0.9 (Tarailo-Graovac and Chen, 2009) using RMBlast v.2.9.0 to blast against the RepBase (2017 release; Jurka et al., 2005) and Dfam v.3.0 database, using the “Viridiplantae” species function. Protein-level repeats were identified by searching against the TE protein Database using Repeat ProteinMask v.4.0.7 (Tarailo-Graovac and Chen, 2009). Tandem repeats were also identified using Tandem Repeat Finder v.4.09 (Benson, 1999).

3.3.5 Structural annotation

Trimmed RNA reads from the close relative *Descurainia bourgaeana* were obtained from White et al. (2016) for initial annotation input to train the *ab initio* gene prediction. The reference genome was indexed, and RNA-seq was mapped to the reference genome using STAR aligner (Dobin et al., 2013). To predict genes from the generated reference genome, MAKER v.2.31.10 (Cantarel et al., 2008) was used in conjunction with the gene prediction software Augustus v.3.3.3 (Stanke et al., 2006) to generate an initial *de novo* annotation. To obtain training models for *de novo* gene finders, we aligned *D. bourgaeana* RNA sequence data following the cufflinks protocol. Augustus and SNAP (github.com/KorfLab/SNAP) were used for *ab initio* gene training using protein datasets from four closely related organisms in the same family, using the Nikolov et al. (2019) backbone Brassicaceae phylogeny for reference, namely *Arabidopsis thaliana* (v. TAIR10; Berardini et al., 2015), *Brassica napus* (Accession: PRJNA546246; Song et al., 2020), *Brassica oleracea* (Accession: GCF_000695525.1; Parkin et al., 2014) and *Capsella rubella* (Accession: GCA_000375325.1; Slotte et al., 2013). A MAKER-derived Annotation Edit Distance (AED) score (Eilbeck et al., 2009) was assigned to each gene model. AED scores range between 0 and 1, with an AED of 0 denoting a perfect match to the evidence and a value of 1 indicating an absence of support for the annotated gene model. The quality of a set of predicted genes was assessed; if 90% of predictions display an AED below 0.5, this is considered well-annotated (Campbell et al., 2014). MAKER was run iteratively three times, providing training files for each run until the AED scores for each gene model plateaus.

3.3.6 Functional annotation

To assess putative gene function, MAKER was used to provide functional annotation for the draft assembly. Here, we used NCBI BLAST+ and the Uniprot TrEMBl protein dataset (Morgat et al., 2020) to assign putative functions to the annotated gene set. Gene Ontology (GO) terms were also annotated by searching against the UniProt database for Viridiplantae. Gene models were used in a BLAST search against the protein dataset using a cut-off of e value of 1×10^{-4}. Since our annotation pipeline uses multiple sources of protein evidence (i.e., RNA-seq and protein datasets from closely related species), each protein model was labelled with its source of evidence.

3.3.7 Comparative phylogenetics – phylogenetic reconstructions and divergence estimates

To determine the taxonomic placement of *D. millefolia* within Brassicaceae 'Lineage I', gene family composition and membership were first determined using OrthoFinder v.2.2.7 (Emms and Kelly, 2019) across the 13 species within Lineage I in Brassicaceae, including *Aethionema arabicum* which was selected as an outgroup (Appendix Table B.6). Orthofinder was run on 13 proteomes with default settings using MMseqs2 to cluster proteins by sequence similarity. To generate a species tree, STAG (Emms and Kelly, 2018) was used to create a consensus tree from individual gene trees of aligned orthogroup sequences. Only orthogroups with all species present were included. STRIDE was used to root the tree (Emms and Kelly, 2017).

The divergence times were estimated from the protein alignment generated by Orthofinder using MCMCtree in Phylogenetic Analysis by Maximum Likelihood v.4.8 (PAML; Yang, 2007). MCMCtree was run using an Approximate Maximum Likelihood method (dos Reis and Yang, 2019), whereby an empirical rate matrix and gamma rates among sites were used to generate a Hessian matrix. A Markov chain Monte Carlo (MCMC) analysis was run for 10,000 generations, using a burn-in of 1,000 iterations. The time of divergence was calibrated from the TimeTree database for each genus pair (Kumar et al., 2017; Appendix Table B.7).

3.3.8 Gene family expansions and contractions

To test for significant contractions and expansions of gene families (defined as "orthogroups" by OrthoFinder), CAFÉ v.5 (https://github.com/hahnlab/CAFE5) was used. CAFE5 implements a maximum-likelihood estimation of gene family evolutionary rates using the given phylogeny and gene family counts. To prepare the input for CAFÉ5, we obtained the species tree generated in the previous step (i.e., OrthoFinder) was transformed it into an ultrametric tree format using the R packages 'phytools' and 'phangorn' (Schliep, 2011; Revell, 2012). We also converted the orthogroup count data from our Orthofinder analysis into a CAFÉ5-friendly format. A P value was calculated for each gene family, and gene families with a P value less than 0.01 were considered to have notable gene gains or losses.

For each significantly expanded and contracted gene family in *D. millefolia,* functional information was inferred based on its ortholog in *A. thaliana*. A Kyoto Encyclopaedia of Genes and Genomes

(KEGG) annotation analysis of genes was conducted using web-based agriGO (www.systemsbiology. cau.edu.cn/agriGOv2) with the singular enrichment analysis method and TAIR10 database.

3.4 Results

3.4.1 Sampling, DNA extraction, and sequencing

Genome sequencing performed by the Illumina platform generated a total of 111 M paired-end short-read sequences (150 bp; Appendix Table B.1). These reads were then quality checked, including removing duplicated sequences and trimming the first ten bp. The genome sequencing performed by the PacBio platform generated 515 k long-reads (average subread length 7,067 bps), with an N50 of 1.1 kb.

3.4.2 Reference genome assembly

The estimated genome size for assembling parameters was obtained using *Descurainia millefolia* C-index, estimated as 0.22 pg (C) (Suda et al., 2003) or ca. 216 Mb. Using k-mer counting, we estimated that the genome size for *Descurainia millefolia* is 183 Mb for our short (Illumina) reads and 223 Mb for our long (PacBio) reads. Therefore, the estimated genome for assembly parameters was set at 200 Mb. These estimates are larger than its close relative size to *A. thaliana* (135 Mb; The Arabidopsis Genome Initiative, 2000) but similar to *Arabidopsis lyrata* at 207 Mb.

Two alternative assembly approaches were employed, DBG2OLC and MaSuRCA. The optimal parameters for DBG2OLC (Appendix Table B.2) provided a poorer draft genome assembly than the alternative MaSuRCA. The primary statistics used to assess quality are the number of contigs, length of the largest contig, N50, L50, and total length. N50 is defined as the length of the contig at 50% (the halfway point) of the assembly, and L50 is the smaller number of contigs whose length sum makes up half the genome size. All three coverage parameters in MaSuRCA (25, 30 and 35) provided a higher-quality genome assembly (greater N50 and length of largest contig, and smaller L50) and a total length closest to our genome size estimates compared to the most optimal DBG2OLC assembly (Table 3.1). While all three assemblies using MaSuRCA provided

similar genome lengths, coverage 30 provided the largest N50. Therefore, the draft assembly using the MaSuRCA assembly with 30 coverage was then used for annotation.

Table 3.1: QUAST summary statistics of the raw genome assemblies generated by DBG2OLC and MaSuRCA assembly methods (before scaffolding and space-filling) for *Descurainia millefolia*. Parameters performed using the DBG2OLC assembler are found in Appendix Table B.2.

	Assembler			
QUAST statistic	**DBG2OLC**	**MaSuRCA - 25 coverage**	**MaSuRCA - 30 coverage**	**MaSuRCA - 35 coverage**
# contigs	2,110	3,021	2,816	3,048
# contigs (>= 1,000 bp)	2,110	3,021	2,816	3,048
# contigs (>= 10,000 bp)	1,836	1,973	1,875	1,939
# contigs (>= 50,000 bp)	717	716	669	679
Largest contig (bp)	2,366,526	2,539,606	1,957,821	2,138,193
Total length (bp)	155,969,182	176,914,168	174,396,279	176,629,879
GC (%)	35.32	35.83	35.78	35.8
N50 (bp)	192,527	207,120	223,560	219,265
N75 (bp)	71,503	75,249	79,445	75,913
L50	199	193	176	183
L75	545	543	499	517

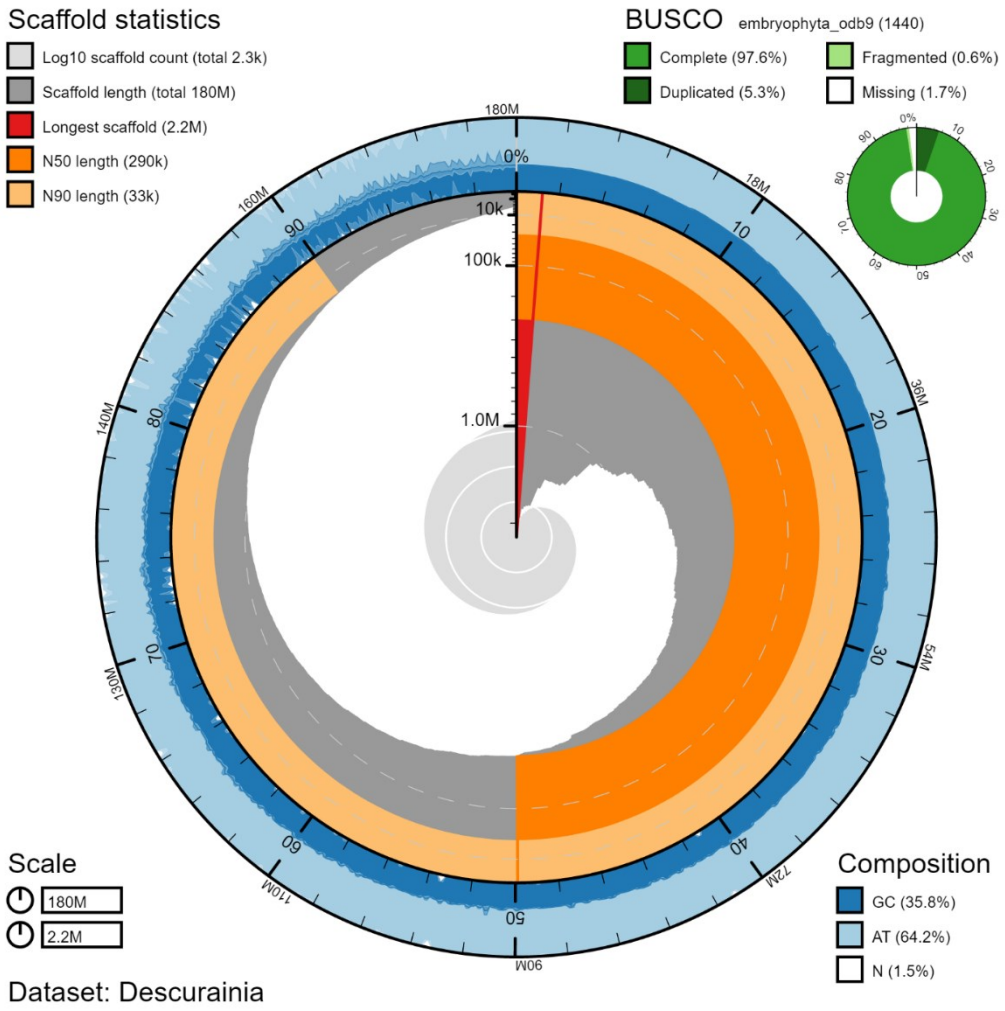


Figure 3.3: Snail plot summary of the assembly statistics and BUSCO scores of the *Descurainia millefolia* genome. The distribution of record lengths is shown in dark grey, with the plot radius scaled to the longest record present in the assembly (red). Orange and pale-orange arcs show the N50 and N90 record lengths, respectively. The pale grey spiral shows the cumulative record count on a log scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. BUSCO scores are presented in the top right, using the background genes of embryophyte (1,440). Genome assembly was visualised using BlobTools (Laetsch and Blaxter, 2017).

The gap closing and scaffolding process closed 2.3% of gaps and 6.7% of nucleotides, reducing the number of scaffolds from 2,773 to 2,274 (Appendix Table B.3). The N50 of the scaffold assembly was 289 kb, and the longest scaffold was 2.2 Mb (Figure 3.3; Table 3.2). The final assembled genome was 180 Mb in length, which is 3.8% less than our lower genome size estimate. The final draft assembly has a L50 of 183 with 1.5% of the genome missing, suggesting that the final draft is relatively fragmented but largely complete. BUSCO was used to assess the completeness of the draft genome assembly, using the embryophyte database, with 97.6% of BUSCOs being complete and single copy (Figure 3.3; Appendix Table B.5).

Table 3.2: Genome and annotation summary statistics of the final *de novo* assembly and annotation of *Descurainia millefolia,* using a hybrid assembly (MaSuRCA) approach.

Genome statistics	
# scaffolds	2,274
Total length (bp)	179,686,191
Largest contig (bp)	2,172,884
GC (%)	35.81
N50	289,995
L50	183
Ns (% genome)	1.5
Annotation Statistics	
# gene loci	26,407
Mean gene loci length (bp)	2,096.35
# transcripts	28,025
mean CDS size (bp)	1,106.32
median CDS size (bp)	1,021
mean exon size (bp)	242.9
median exon size (bp)	135
mean # exons/transcript	5.3
median # exons/transcript	4

3.4.3 Structural and functional annotation

We performed gene prediction and annotation using *de novo, ab initio* and transcriptome-based approaches. MAKER and Augustus generated a final set of 26,407 gene models at an average length of 2,096.35 bp (Table 3.2). For comparison, the most recent annotation of the *A. thaliana* draft genome had 27,655 protein-coding gene models (Araport11). Through labelling the source of the annotated genes, we can show over 99% of the gene set was annotated from protein databases and 4×10^{-3} from our RNA-seq. Less than 0.5% of the annotated genes possessed an AED score of over 0.5. The general benchmark Campbell et al. (2014) proposed for a well-

annotated genome is 90% of genes with an AED score of less than 0.5; therefore, our reference genome is of good annotation quality for downstream analyses. Through functional annotation of our gene set, 75% of the genes had functional annotations.

3.4.4 Repeat annotation

A total of 23.2 Mb (13.27%) of the assembled genome consisted of repeat regions, as indicated by RepeatMasker (Figure 3.4; Appendix Table B.4). A total of 30,931 tandem repeats were found across the 2,274 scaffolds. For comparison, the *Arabidopsis thaliana* genome is comprised of 16.5% repeat elements (Appendix Table B.6). Of the assembled genome, 7.81% were retroelements (20 k bps). The most abundant repeat element families were Copia (2.57%) and Gypsy (3.78%). DNA transposons comprise 3% of the genome, and the total interspersed repeats comprise 10.89%. Compared to other members of the Brassicaceae family, the number of repeat elements is relatively low. The repeat landscape (Figure 3.4) represents a bell curve, suggesting an equilibrium between transposition and excision over evolution time; therefore, no recent or rapid expansions are apparent.

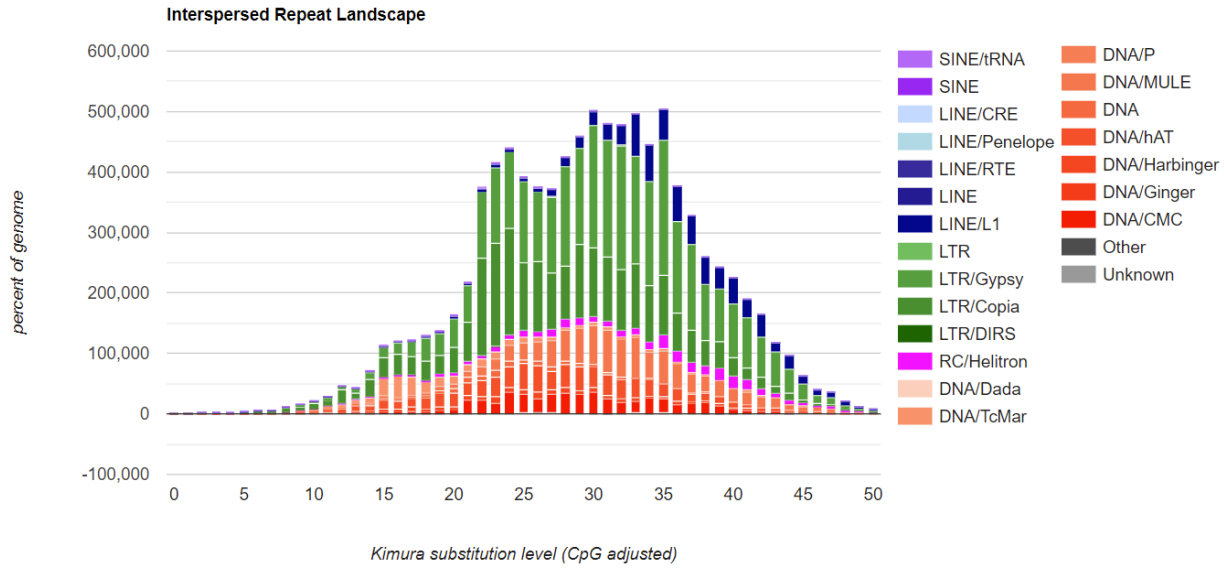


Figure 3.4: *Descurainia millefolia* transposable element divergence landscape. Transposable elements (TEs) within the *Descurainia* genome have been characterised (different classes represented by distinct colours). The plot shows the relative abundance of each class and their relative age (molecular clock estimate). Older copies of TEs are located on the right side of the graph.

3.4.5 Comparative phylogenetics – phylogenetic reconstructions and divergence estimates

A total of 33,288 orthogroups were recovered across the 13 sampled proteomes, with 1,103 orthogroups shared between all 13 Brassicaceae species. For *D. millefolia,* 97.4% of gene models were assigned to orthogroups, with 728 unassigned (2.6%) and 96 specific to *D. millefolia* (0.8%). *Descurainia millefolia* shared the largest number of orthogroups with *D. sophiodies* (18,725), closely followed by *A. thaliana* (18,077; Appendix Figure B.1).

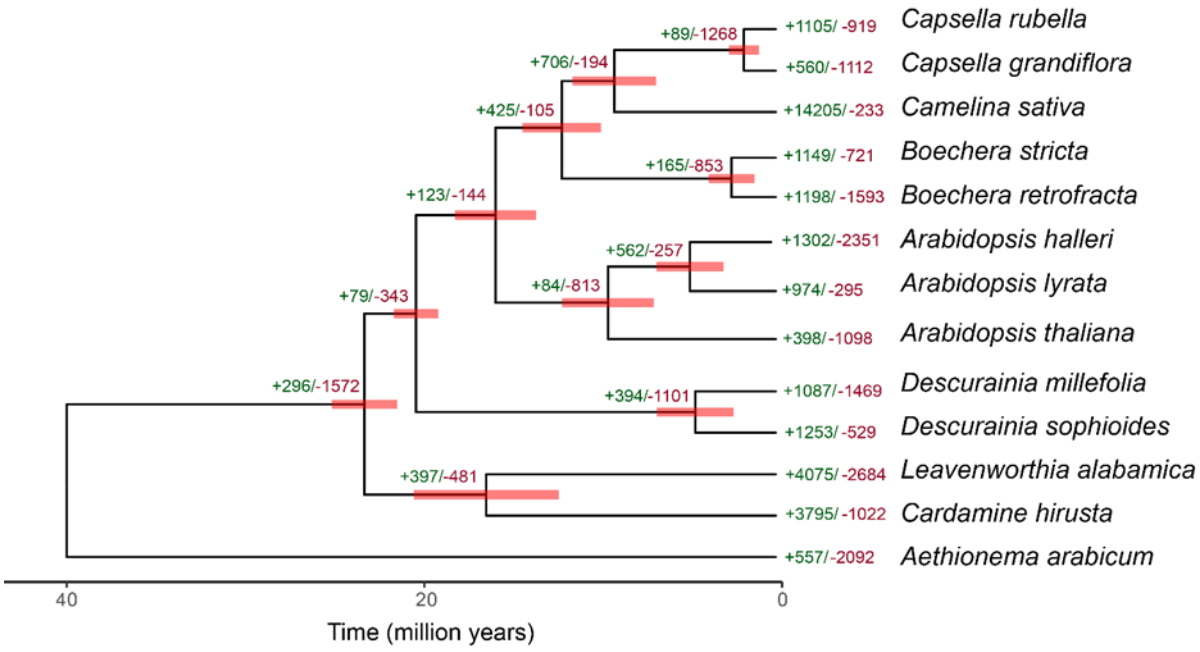


Figure 3.5: Species tree generated from a consensus of 1,103 ortholog trees for 13 species within the Brassicaceae, with *Aethionema arabicum* set as the outgroup. The red bar on the nodes represents divergence estimates (million years) from PAML MCMC. Gene expansions (green) and gene contractions (red) generated by CAFÉ5 are also provided on the nodes and species tips.

The species tree, generated from 1,103 orthogroups, consisted of five clades that form successive sister groups: (1) *Leavenworthia* and *Cardamine*; (2) *Descurainia*; (3) All four *Arabidopsis* genomes (representing three species); (4) *Camelina* and *Capsella* species; (5) *Boechera.* Our Brassicaceae phylogeny was estimated to have diverged 39.7 Mya (95% CI: 32.8-45) from our selected outgroup *Aethionema arabicum,* and the split for the two major clades (between clade 1 and clade 2 - 5) occurred 23 Mya (95% CI: 21.41-25.8). *Descurainia millefolia* and *Descurainia sophiodies* diverged 4.5 Mya (95% CI: 2.63-6.92), and the genus 20.12 Mya (95% CI: 19.04-21.48) from other sampled Brassicaceae species.

3.4.6 Gene family expansions and contractions

Descurainia millefolia and *D. sophiodies* shared 136 gene families. For the genus *Descurainia* (clade 2), there was an expansion of 1,087 gene families and a contraction of 1,469 ($P < 0.05$) for *D. millefolia* relative to *D. sophiodies*. Between the *Descurainia* clade and clades 3-5 (consisting of

Capsella, Camelina, Boechera and *Arabidopsis*), there was an increase of 79 and a decrease of 343 gene families. There were three significant (FDR < 0.05) KEGG pathways each for the contracted and expanded gene families in *D. millefolia* (Table 3.3).

Table 3.3: Functional KEGG annotation of the most significantly expansive and contract gene families in *D. millefolia.*

Gene Families	Pathway	# Genes	Pathway Genes	Fold Enrichment
Contracted	Circadian rhythm	7	39	5.34
	Spliceosome	18	187	2.86
	Ribosome	29	315	2.74
Expansion	ABC transporters	6	32	6.61
	Aminoacyl-tRNA biosynthesis	7	57	4.33
	Biosynthesis of secondary metabolites	65	1243	1.84

3.5 Discussion

As far as we are aware, this is the first draft genome assembly from a Canary Island endemic plant and the second genome within *Descurainia* (the first being *Descurainia sophiodies,* unpublished; Chen et al., 2022). Since the genome of *A. thaliana* was sequenced in 2000 (The Arabidopsis Genome Initiative, 2000), several high-quality Brassicaceae genomes have been sequenced to high-quality (plabipd.de). The latest statistics (as of 2022) suggest 43 species in Brassicaceae have a published genome. The release of reference genomes has increased the exploration and opportunities for avenues of research. There is an opportunity for the *D. millefolia* genome to be utilised for evolutionary analyses: previous research has suggested that the seven species of Canary Island *Descurainia* have undergone diversification patterns related to ecological, geographic and hybridisation evolutionary influences (Goodson et al., 2006; Jackson et al., chapter 2). The endemic Canary Island *Descurainia* occupy various ecological habitats, for example, dry and hot low-altitude scrublands and high-altitude regions which experience low temperatures and low precipitation. The genome assembly of *Descurainia millefolia* will provide genomic resources to improve our understanding of island radiations and phylogenetic understanding and aid conservation efforts. Furthermore, it opens an understanding of the molecular underpinning of adaptation and diversification to extreme environmental pressures.

3.5.1 The *Descurainia millefolia* genome

We used a combination of short (Illumina) and long (PacBio) sequencing strategies and used a hybrid assembly approach to generate a scaffold-level genome assembly for *Descurainia millefolia*. The final assembly size (180 Mb) was close to the estimated genome size (200 Mb), consisting of 2,274 scaffolds. The final genome size was smaller than the estimate. However, not within the realm of potential assembly errors. It was estimated that 0.6% of the BUSCOs were missing, so we assumed a small fraction of the genome was missing. Other estimates for species within *Descurainia* range between 0.17-0.58 pg (i.e., ca. 166 – 567 Mb; Lysak et al., 2009), so our *D. millefolia* falls within the lower end of genome size expected within *Descurainia*.

The final assembly is of high quality with an N50 of 0.29 Mb and high completeness based on BUSCO results (97.6%). Fragmentation of plant genome assemblies, mainly when only using short-read sequencing, is often attributed to a high content of repetitive DNA (Claros et al., 2012). Therefore, hybrid assembly approaches are often suitable for plant genomes which often exhibit

long repeats which are challenging to assemble (Jiao and Schneeberger, 2017). Paired end-reads are often designed to overlap approximately 20 bp; therefore, longer read approaches enable a better ordering of the initial contigs. The initial contig assembly generated a 176 Mb genome with ~3 k contigs and an N50 of 223 k bp. Our long-read PacBio reads were able to scaffold the genome from 3,034 contigs to 2,274 scaffolds to resolve part of the repetitive compartment of the genome. Scaffolding increased our total genome size from 174 Mb to 180 Mb, and the N50 improved from 223 k to 290 k, an improvement of 3.4% and 33%, respectively. Long reads to improve gap closing provided little benefit for our genome assembly, reducing the number of Ns from 2.9 to 2.7 Mb, having little effect on the resulting genome assembly quality. This could be attributed to our relatively short, long-reads averaging ca. 7,000 bps. Frequently Nanopore technology (ONT) for long-read sequencing libraries has shown to be more successful in assembling complex genomes with fewer gaps than those genome assemblies using PacBio technology, often resulting in a higher N50 (Rousseau-Gueutin et al., 2020). While our genome still demonstrated a level of fragmentation, the genome statistics suggest that the draft genome for *D. millefolia* is of high quality through high completeness, which will allow further downstream genomic analysis.

The *D. millefolia* genome annotations resulted in a highly complete genome (97.7%), which can be partly attributed to the large number of genomics resources available within Brassicaceae. A total of 11% of the genome was recovered and classed as interspersed repeats (7.8% were Retroelements). Most Brassicaceae genomes are dominated by Gypsy long terminal repeats (LTR) retrotransposons. *Descurainia millefolia* exhibits a smaller number of LTR (16 k) compared to other species in Brassicaceae, for example, repeat compositions are comparable to *Arabidopsis thaliana* (11 k).

The number of gene loci models predicted is 28,025, comparable to *Arabidopsis thaliana* (27,416) and *Capsella rubella* (26,521) (The Arabidopsis Genome Initiative, 2000; Slotte et al., 2013b). The *Descurainia sophiodies* genome has 30,141 gene models (Chen et al., 2022), 6.4% more than our *D. millefolia* genome. It is estimated that our genome is 97.7% complete, with 92.4% being single-copy BUSCOs. Therefore, the 26,407 gene model provides a good representation of the *D. millefolia* genome. More complete proteomes allow for accurate phylogenetic analysis, increase evolutionary signal, and improve node support (Hall et al., 2002). Furthermore, these annotations could add valuable information about recent duplication and gene retention bursts. Here we

compare the gene models found in *D. millefolia* with closely related species with highly-complete genomes.

3.5.2 Phylogenetic relationships

We constructed a species tree of 'Lineage I' of the Brassicaceae using the *D. millefolia* proteome with an additional 13 proteomes representing 12 species obtained from the BRAD v.3.0 databases (Chen et al., 2022). The tribe *Descurainieae* was first described by Al-Shehbaz et al. (2006), consisting of 6 genera and ca. 60 species (48 of which are *Descurainia*). Few Brassicaceae phylogenies include an extensive sampling of Descurainieae species, and earlier studies have suggested Descurainieae is polyphyletic (Beilstein et al., 2008). However, the consensus is that the tribe is monophyletic within Lineage I of the Brassicaceae phylogeny and sister to the clade consisting of tribes Smelowskieae, Smelowskieae and Yinshaniea (Franzke et al., 2011; Nikolov et al., 2019; Beric et al., 2021; Liu et al., 2021; Hendriks et al., 2022). *Descurainia millefolia* and *D. sophiodies* formed a monophyletic clade and was revealed to be sister to the *Arabidopsis* clade and *Boechera, Camelina* and *Capsella* clade, with previously published phylogenetic analyses (Al-Shehbaz et al., 2006; Nikolov et al., 2019; Liu et al., 2021).

Based on our phylogenomic analyses, we estimated the split between *Descurainia* and the rest of the sampled Brassicaceae in Lineage I was 20 Mya (95% CI: 19 – 21.5). Our estimates are similar to previous phylogenetic studies, including the *Descurainia* species. For example, using transcriptome assemblies, Guo et al. (2021) estimated the divergence of *Descurainia bourgaeana* within Brassicaceae to be 15 Mya (95% CI: 18 – 22 Mya) from *Arabidopsis* and *Capsella* clades. Our divergence results suggest the split between *D. millefolia* and *D. sophiodies* occurred 4.5 Mya (95% CI: 2.6 – 6.8 Mya). In the absence of genomic information for other *Descurainia* species, we cannot date the origin of *D. millefolia*. There are also few studies to compare divergence within *Descurainia,* and our results conflict with Goodson et al. (2001) estimates, which more extensively sampled *Descurainia* species, which exceeds ours, with a split between *D. millefolia* and *D. sophiodies* at 11.5 Mya (± 1.5) for ITS datasets and 14.5 (± 1.1) with cpDNA.

Potential explanations for these differences may include the use of ITS and cpDNA, which often fail to capture accurate divergence estimates and polygenetic relationships, inaccurate

substitution models, and there is a possibility of the use of outdated estimates for time calibration points (Olmstead and Palmer, 1994; Ballard and Rand, 2005).

A comparison of the genome of *D. millefolia* and the 12 other species in Brassicaceae revealed significant expansion or contraction in the size of gene families within *D.* millefolia. Gene expansions can often provide insights into the evolution of a species. For example, one significant KEGG pathway is the ABC transport genes, which represent proteins involved in the facilitation of movement of molecular across the cell membrane (Hollenstein et al., 2007). ABC transporters play a vital role in plant growth, development, and stress response by regulating the uptake and distribution of nutrients and by protecting against harmful substances (Kang et al., 2011; Kretzschmar et al., 2011). Contracted gene families include those related to Circadian rhythm, an internal clock within plants which anticipates diurnal changes (Venkat and Muneer, 2022). This may suggest a lesser need within *D. millefolia*, compared to their close relatives, to regulate their physiological pathways to code with diurnal environmental changes.

3.6 Conclusion and Future direction

As whole-genome sequencing becomes more economically viable, more plant genomes within the Brassicaceae family are being produced. High-quality genomes will create numerous avenues for the continued development of our understanding of evolution and Brassicaceae. Here, we present the first *de novo* sequenced and assembled high-quality scaffold-level *Descurainia millefolia* genome. There is potential for *Descurainia* to further our understanding of adaptation: within the Canary Island *Descurainia* clade, several species occupy habitats with environmental and climatic conditions that promote ecological adaptation. Recently, there has been an explosion in research on species found in extreme environments and high altitudes, utilising genome sequencing to understand adaptation at a molecular level (Guo et al., 2018; Hu et al., 2021; Nowak et al., 2021).

3.7 Data availability

The genome assembly and annotations for the *Descurainia millefolia* genome are available at: github.com/amycjack/*Descurainia*_millefolia_V.1.0 (Private repo: available on pull-request).

3.8 Acknowledgements

The authors would like to thank Anne Romero for performing the DNA isolations for our PacBio Sequencing. This work was supported by a PhD studentship to ACJ from the Natural Environmental Research Council [grant number NE/L002531/1]. We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

Chapter 4 Ecological adaptation to high-elevation in Canary Island *Descurainia* (Brassicaceae) and parallels with other Brassicaceae

4.1 Abstract

Oceanic archipelagos promote striking examples of plant lineages that have radiated rapidly over steep ecological gradients and short distances. Many have evolved adaptations that allow lineages to diversify and survive in the more extreme environmental conditions. Here, we investigate the genetic basis of ecological adaptation in the island plant lineage Canary Island *Descurainia* (Brassicaceae), an island cousin of *Arabidopsis.* This radiation provides an attractive group to study speciation as these seven endemic species have diversified *in situ* along an elevational gradient, from low-elevation scrub (ca. 500 m) to high-elevation sub-alpine desert (ca. 2,200 m). Here we employed genome re-sequencing to sample 14 genomes across the seven species and an outgroup. Phylogenomic analyses using predicted proteome datasets were consistent with previous reconstructions of Canary Island *Descurainia*. Using the branch-site *dN/dS* method, a selection analysis found 276 genes under positive selection on the branch leading to a clade solely consisting of high-elevation species of Canary Island *Descurainia*. Gene ontology (GO) enrichment analysis revealed significantly enriched GO terms associated with reproduction and abiotic stress. To provide further evidence that our positively selected genes are adaptive, we used T-DNA *Arabidopsis* knockouts for three genes to establish the function and phenotypic responses to abiotic stresses. When exposed to a water stress treatment, knockout RD29B (AT5G52300) exhibited less tolerance to water stress than the control lines. Comparing the list of positively selected genes (PSGs) to other species in Brassicaceae that inhabit high-elevation and -latitude environments, we found little evidence of widespread convergent and gene reuse across the species. There was a significant overlap with PSGs in the Arctic species *Draba nivalis* and our *Descurainia* clade, with several genes involved in tolerance to abiotic stressors that could be related to diversification into these inhospitable environments.

4.2 Introduction

Whole-genome studies have the potential to further our understanding of the mechanisms contributing to plant diversification and speciation (Flood and Hancock, 2017; Anderson et al., 2021). With these advanced sequencing tools, it is possible to assess the genomics of adaptation by identifying the genes underlying adaptive traits and metabolic pathways that play a role in a species adaptation to different habitats (Stapley et al., 2010; Strasburg et al., 2012; Savolainen et al., 2013). Along with furthering our understanding of evolutionary processes, there is also the potential to reveal key genes for resilience to climate change (Franks and Hoffmann, 2012).

Islands contribute disproportionately to global biodiversity by harbouring 25% of global endemic species richness whilst covering only 3% of the land surface area (Paulay, 1994). Oceanic island and their endemic lineages have often been the subject of evolutionary studies as they have adapted and diversified *in situ* across short space and time (Emerson, 2002; Warren et al., 2015). High-elevation oceanic islands are a particularly useful tool due to the large number of endemic species distributed along strong environmental and climatic gradients (Steinbauer et al., 2012; Halbritter et al., 2018). In addition, these endemic lineages are likely to have adapted to high-elevation environments relatively recently. Their closest relatives are often found in lower-elevation environments, providing a persuasive comparison device (Graham et al., in prep). As such, isolation, combined with environmental gradients, serve as a powerful platform to understand evolutionary processes such as adaptation, with implications for biodiversity conservation and climate change resilience.

Evolutionary research often focuses on resolving the genetic mechanisms of differential adaptation between high- and low-elevation species. However, research within plants is still in its infancy compared to animal and human studies (Cheviron and Brumfield, 2012; Qiu et al., 2012; Yu et al., 2016; Azad et al., 2017; Ma et al., 2019), with the majority focusing on continental mountain systems. High-elevation habitats on mountain-like systems often share similar environmental and climatic features, such as low temperatures, poor soil, large diurnal changes, strong winds, frost events, and high UV radiation (Nagy and Grabherr, 2009). These harsh conditions cause plants to alter their morphology and physiology to become more tolerant. The genomic studies suggest that high-elevation adaptation can leave strong signatures of selection related to these environmental conditions. For example, genomic comparisons of two *Eutrema* (Brassicaceae) species distributed at low- and high-elevation in Qinghai-Tibet Plateau identified

selection on genes associated with DNA damage repair, reproduction, and cold tolerance within the high-elevation species (Guo et al., 2018). An analysis of Maca (*Lepidium meyenii*), distributed in the Andes, has identified genes linked to the development of specialised leaf shapes (Zhang et al., 2016). Population genomic analyses of other species adapted to contrasting elevations have identified genes with the potential for involvement in adaptation to high-elevation (Chapman et al., 2013; Chen et al., 2019; Ma et al., 2019; Zeng et al., 2020; Li et al., 2021; Feng et al., 2022). The candidate genes can vary greatly, but often they are associated with similar molecular pathways related to abiotic and environmental stressors, including response to light, temperature, water availability, and influencing reproductive strategies (for instance, vernalisation and flowering time).

The Canary Islands is an archipelago of eight main islands located ca. 100 km west of North Africa. The older islands are located in the east (21 million years for Fuerteventura) and the youngest in the west (0.8 million years for El Hierro; Schmincke, 1976) and are all at different stages of volcanic cycles (Schmincke, 1976; Chapter 2, Figure 2.1). Fuerteventura (ca. < 807 m, asl) and Lanzarote (ca. < 671 m) have eroded, resulting in a relatively low altitude plain island. The erosion process in Gran Canaria (ca. < 156 m) and La Gomera (ca. < 1,487 m) had begun, while Tenerife (ca. > 3,700 m, asl), La Palma (ca. > 2,400 m, asl) and El Hierro (ca. < 1,501 m, asl) represent the taller islands and are still growing through recent volcanic activity (Schmincke, 1976).

The Canary Islands are home to many endemic lineages, with 40% of the total flora endemic to the archipelago (Reyes-Betancort et al., 2008). The rapid diversification of these lineages is repeatedly attributed to the islands broad range of environmental habitats transitioning over steep elevational gradients, as well as unique climatic conditions (such as the humid north-east trade winds) and dynamic volcanic terrain (Fernández-Palacios and de Nicolás, 1995; Stuessy et al., 2006; Steinbauer et al., 2012, 2013, 2016; Irl et al., 2015; Irl, 2016; Hanz et al., 2022; Rodríguez-Rodríguez et al., 2022). For the taller western islands (Tenerife and La Palma), the habitats transition broadly from low-elevation sub-desert scrub, through mid-elevation subtropical forests on the northern slopes, to pine forests, and into high-elevation sub-alpine scrub (González-Mancebo and Hernández-García, 1996). The floristic composition is highly stratified, and endemics are often restricted to elevational-environmental belts (Fernández-Palacios and de Nicolás, 1995). Generally, Canary Island flora migrate upslope to colonise the higher-elevation habitats, with 83.3% of sub-alpine species revealing their closest relative as a low-elevation species (Graham et al., in prep).

The harsh abiotic conditions and geographical isolation of the sub-alpine zones (Figure 4.1) have resulted in relatively few plants colonising these environments but high levels of endemism (38% endemism; Irl et al., 2015; Steinbauer et al., 2016). Temperatures here are low, with an average annual temperature of 6-11°C with significant diurnal changes. For instance, Las Canadas (Tenerife) can vary as much as 25°C (Bramwell and Caujapé-Castells, 2011). There is also a frost period which extends from October to May, as well as snow in the winter (del Arco Aguilar and Rodríguez Delgado, 2018). In contrast to continental mountains, rainfall in Tenerife decrease with elevation above 2,000 m due to the temperature inversion. As a result, rainfall at sub-alpine zones is relatively low (between 350 – 500 m) compared to mid-elevations. The availability of water is further decreased with the porous volcanic soil.

Figure 4.1: Picture of Teide National Park (Tenerife) taken at ca. 2,000 m with Mount Teide in the background. The pine forest zone on the northern slope (right), dominated by *Pinus canariensis*, meets the sub-alpine desert (left) habitat where species diversity is relatively poor.

Understanding how recently-evolved high-elevation Canary Island taxa have adapted to habitats can provide novel insights into the processes behind ecological divergence. In particular, the harsh environmental constraints within the high-elevation zone have been associated with several specialised morphological adaptations (Irl, 2016). Examples include fire resistance (Climent et al.,

2004), leaf variation (White et al., 2018), monocarpy (Jorgensen and Frydenberg, 1999; Jorgensen and Olesen, 2001), giant rosettes (Mora et al., 2019), and deep groundwater tapping roots (González-Rodríguez et al., 2017). However, many of these traits have been inferred as adaptive through phylogenetic analyses, which offers limitations. Firstly, there can be a blurring of adaptive and non-adaptive signals of divergence, most notably through hybridisation (Rundell and Price, 2009). Secondly, there can be a lack of appropriate phylogenetic resolution, a particular issue within very closely related taxa. Lastly, certain adaptations, such as changes in metabolic pathways, may be less pronounced than morphological traits. Adaptation occurs from molecular to whole-plant levels, but the complexities of plant physiology are often challenging to detect. For example, the production of antifreeze agents, the capacity to photosynthesize at low temperatures, and low respiration rate are common physiological traits within high-elevation plants (Nagy and Grabherr, 2009). Despite the abundance of studies examining speciation and adaptation within Canary Island lineages, genomic-scale analyses are limited, and even fewer have investigated adaptation at a molecular level.

Within the Canary Islands endemic flora, *Descurainia* Webb & Berthel. (Brassicaceae; Goodson et al., 2006; Chapter 2) is an example of a plant radiation with congeners occurring at different elevations. *Descurainia* comprises ca. 45 species, distributed across North America, Europe and Euro-Siberia. They are small shrubs with yellow flowers. Seven species are endemic to the Canary Islands and have radiated within and between four islands (Tenerife, La Palma, La Gomera, and Gran Canaria; Figure 4.2) and across habitats and elevations. Canary Island *Descurainia* is estimated to have arrived through a single colonisation event, 500,000 – 750,000 years ago, from a continental Mediterranean relative (Goodson, 2003). Our previous work resolved two groups of species, one distributed largely at low elevation (< 1,100 m; *D. millefolia, D. artemisioides* and *D. preauxiana*) and the other at high-elevation (typically > 1,800 m; *D. bourgaeana, D. lemsii, D. gilva* and *D. gonzalezii*; Chapter 2). Chapter two indicated a single shift between these low and high-elevation habitats, which differed in precipitation and temperature regimes. Overall, a combination of ecological, hybrid, and geographic processes has contributed to the diversification of this group (Goodson et al., 2006; Chapter 2).

Extending the phylogenetic work of Jackson et al. (Chapter 2) and the assembly and annotation of a reference genome of *Descurainia millefolia* (Chapter 3), we aimed to understand the molecular basis of ecological adaptation in Canary Island *Descurainia* using whole-genome resequencing techniques, with a focus on the divergence between the low- and high-elevation species. In

addition, we wish to identify any possible similar signatures of adaptation between Canary Island *Descurainia* and species in Brassicaceae that occur in high-elevation and high-latitude environments. A total of 14 genomes were sequenced across the seven species to enable a comparative analysis approach. To address our questions, we undergo five main aims:

(i) Previous phylogenetic reconstructions of Canary Island *Descurainia* did not resolve certain populations and species, namely *D. lemsii* and a La Gomera population of *D millefolia* (Chapter 2). Therefore, we aimed to produce an updated phylogeny of Canary Island *Descurainia* using whole-genome sequencing data and a Maximum-likelihood method.

(ii) Chapter two identified one ecological shift between our low- and high-elevation clades. We aimed to extend this analysis by identifying climatic differences between the two clades using characteristics commonly associated with high-elevation; Nagy and Grabherr, 2009) and their seasonal variation.

(iii) To identify candidate genes that may have incurred adaptive properties to high-elevation habitats, a test for positive selection was performed using a Branch-Site test (Zhang et al., 2005), which identified accelerated non-synonymous substitutions on protein-coding genes along a specific lineage. Here we test the branch leading to the clade solely consisting of our high-elevation species. As noted previously, there is evidence of hybridisation within the evolutionary history of Canary Island *Descurainia,* which has the potential to distort evolutionary signals. Our previous work indicated that the proportion of genome shared between our low- and high-elevation species is less than 2% (see results of Chapter 2). Therefore, we are confident that the effects of geneflow for the analyses within this chapter would be minimal.

(iv) To provide further evidence that the identified genes are adaptive, we selected three candidate genes to be tested for function using transfer DNA (T-DNA) *Arabidopsis thaliana* knockouts and stress-induced greenhouse experiments.

(v) Our final aim is to examine a broader question regarding high-altitude adaptation and explore whether species in Brassicaceae independently evolve similar strategies through gene reuse (i.e., genomic parallelism). We compared genes in our list of candidates in high-elevation *Descurainia* with other Brassicaceae taxa that are restricted to high-elevation or high-latitude conditions.

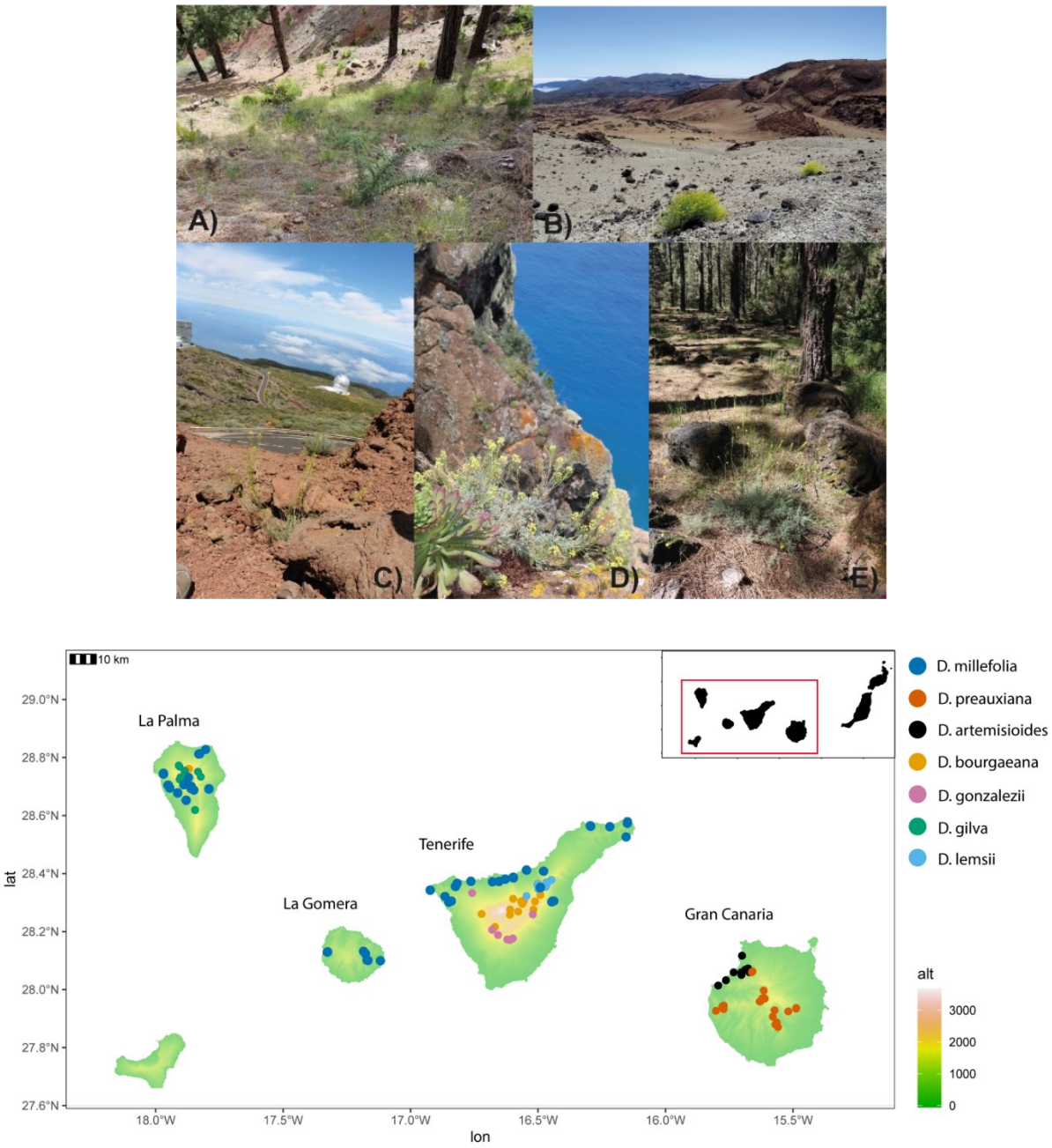


Figure 4.2: Top: A) *Descurainia gonzalezii* in a pine forest habitat in Tenerife (1,750 m), B) The sub-alpine scrub of El Teide, Tenerife (2,100 m), with *D. bourgaeana* in the foreground, C) sub-alpine scrub of La Palma where *D. gilva* is distributed (2,000 m), D) dry scrublands on Tenerife where *D. millefolia* is often found on the cliffs (250 m), E) *D. lemsii* distributed on the margins of the pine forest (2,000 m). Bottom**:** Distribution of the seven endemic Canary Island *Descurainia* species across four islands. Each point represents a GPS coordinate (261 individuals) collated from herbarium specimens deposited within the herbarium of the Natural History Museum, London (BM). Terrain colours represent the elevation (m).

4.3 Material and Methods

4.3.1 Canary Island *Descurainia* sample collection and DNA extraction

Leaf material from the seven species of Canary Island *Descurainia* was collected in 2019 and preserved in silica gel (11 accessions; Under the permits from Parque Nacional del Teide: no. 42917710D; Cabildo de la Palma: no. 2019004595; Cabildo de la Gomera: no. 9319; Cabildo de Gran Canaria: no. 14328). Herbarium vouchers were made for each and deposited within the Natural History Museum, London (BM). Additional Canary Island *Descurainia* leaf material was obtained from live plants grown from seeds obtained from a seed bank (3 accessions: Banco de Germoplasma Vegetal "César Gómez Campo"). Two individuals (representing different populations) were sampled per species, except for *D. artemisioides,* where one individual was sampled and *D. millefolia,* for which we sampled three individuals (one each from Tenerife, La Palma and La Gomera; the three islands on which this species occurs). DNA was extracted using a CTAB-based method from either silica-dried or fresh leaf material (Porebski et al., 1997). The closet extant relative of the Canary Island taxa is *D. tanacetifolia*, distributed in the Iberian Peninsula, France, Italy, and Switzerland (Goodson et al., 2006). DNA of *D. tanacetifolia* utilised in the Goodson et al. (2006) study (kindly provided by Robert K. Jansen, The University of Texas at Austin) was used as the outgroup. Paired-end sequencing libraries were generated for 15 samples by Novogene (Hong Kong) and sequenced on Illumina HISEQ X Ten Sequencer (Illumina Inc.; Table 4.1).

Table 4.1: Summary of Canary Island *Descurainia* individuals and outgroup selected for whole-genome resequencing. GC = Gran Canaria, TEN = Tenerife, LG = La Gomera, LP = La Palma. Habitat descriptions for Canary Island *Descurainia* is based on Bramwell (1977) and personal observations (ACJ).

Species	Study ID	Origin	Elevation (m)	Habitat type	Voucher ID
D. artemisioides	1	GC	410	Subdesert scrub	BM013825841
D. bourgaeana	6	TEN	2,058	Sub-alpine	BM013825795
D. bourgaeana	7	TEN	N/A	Sub-alpine	N/A [b]
D. gilva	8	LP	1,275	Sub-alpine	BM013825888
D. gilva	9	LP	2,073	Sub-alpine	BM013825861
D. gonzalezii	10	TEN	2,064	Southern-slope pine forest and pine forest/sub-alpine border	BM013825808
D. gonzalezii	15	TEN	N/A	Southern-slope pine forest and pine forest/sub-alpine border	N/A [b]
D. lemsii	2	TEN	1,963	Norther-slope pine forest/sub-alpine border	BM013825789
D. lemsii	11	TEN	1,615	Norther-slope pine forest/sub-alpine border	BM013825782
D. millefolia	3	TEN	540	Subdesert scrub	BM013825821
D. millefolia	4	LG	605	Subdesert scrub	BM013825833
D. millefolia	12	LP	432	Subdesert scrub	BM013825880
D. preauxiana	13	GC	845	Subdesert scrub	BM013825850
D. preauxiana	14	GC	N/A	Subdesert scrub	N/A [b]
D. tanacetifolia	5	[Spain]	N/A	[Mediterranean]	N/A [a]

[a] DNA obtained from Goodson et al. 2004 study; [b] DNA extracted from a seed bank

4.3.2 Whole-genome resequencing, mapping, and SNP calling and filtering

Whole-genome resequencing (WGRS) reads (150 bp paired-end) were generated for the 14 Canary Island *Descurainia* individuals across and the continental outgroup. Prior to mapping, the quality of the reads was checked using FASTQC, and reads were trimmed using Trimmomatic (Bolger et al., 2014) to discard reads with a quality score of 0.05 and shorter than 65 bps. Using our high-quality annotated genome of *Descurainia millefolia* (v.1.1; 186 Mb, N50 of 280 kb) as a reference (Chapter 3; github.com/amycjack/*Descurainia*_millefolia_V.1.1, accessed 2020), the trimmed reads were mapped using the Burrows-Wheeler Aligner (BWA) v.0.7.17 and *mem* function (Li and Durbin, 2009) with default parameters. Mapping results were then converted into the Bam format and sorted with SAMtools v.1.3.1 (H. Li et al., 2009). Unmapped reads were discarded using SAMtools. PCR duplicates were removed from the mapped reads using Picard (github.com/broadinstitute/picard, accessed 2020). The chloroplast regions were identified using the *D. millefolia* annotation GFF file and extracted and discarded from the mapped reads using SAMtools.

The aligned BAM files were processed with SAMtools v.1.15 using the *mpileup* function to identify SNPs and with BCFtools to identify INDELS, which were then converted into a VCF file. VCFtools v.0.1.13 (Danecek et al., 2011) was used to filter the VCF file to only include higher-quality SNPs using the following filters: a) a minor allele count less than 3, b) include only sites with a quality value above 30 and c) exclude sites with more than 50% missing data.

Using the annotation GFF file for our *D. millefolia* reference genome, the VCF files containing SNPs were annotated using SNPeff (Cingolani et al., 2012) for the locations of exons, introns, coding regions (CDSs), and intergenic regions.

4.3.3 Phylogenetic analysis

Phylogenetic reconstruction of the 14 genomes was generated using filtered SNPs and a Maximum-Likelihood (ML) method. SNPs were first converted into phylip format using BioPython v.1.79, and RAxML v.8.2.9 (Stamatakis, 2014) was employed with a GTRGAMMA evolutionary substitution model and 2,000 partitions. *Descurainia tanacetifolia* was set as the outgroup, and bootstraps were computed with 1,000 bootstrap iterations.

It is important to note that in Chapter 2, we identified a hybridisation event between *D. gonzalezii* and *D. millefolia,* resulting in a hybrid origin for *D. gilva*. Gene flow between these two clades can potentially obscure the detection of positively selected genes within our high-elevation clade. However, the analysis indicated that most of the genome shared was between the sister taxa *D. gonzalezii* and *D. gilva* (99%). We assumed that the 1% of the shared genome with *D. millefolia* was a local introgression event and posed little influence over our *dN*/*dS* analysis.

4.3.4 Genetic structuring

To identify genetic groupings among our 14 Canary Island *Descurainia* individuals, a Principal Component Analysis (PCA) was performed using SNPs. The PCA was performed on the filtered VCF file using the *pca* function in PLINK, and visualisation of the first two PCAs was done in the R package 'ggplot2' (Wickham, 2009).

Pairwise comparisons of the fixation index (F_{ST}; Whitlock and Mccauley, 1999) values were calculated between the Canary Island *Descurainia* species using our SNP dataset and the python script popgenWindows.py (github.com/simonhmartin/genomics_general; accessed 2020) within a 10 kb non-overlapping genomic windows.

4.3.5 Climate analysis

To determine whether there are significant differences in climate experienced between the high- and low-elevation Canary Island *Descurainia,* we first collated georeferenced localities and elevation data for individuals across the seven species. Data was collected from the combination of my data from the 2019 field trip and herbarium specimens deposited in the Natural History Museum (UK). We mapped the coordinates onto monthly bioclimatic variables at ~1 km^2 resolution, sourced from WorldClim 2.1 (Fick and Hijmans, 2017). The raster value of each data layer at each locality point was extracted in ArcGIS Pro using the *points-to-values* function. Values for individuals were grouped into their respective clades, A ('high-elevation clade') or B ('low-elevation clade'; see results), and climatic values were grouped into the four seasons. For every

climatic variable, an ANOVA test was performed between the two clades for each season, with the P value adjusted for multiple testing.

4.3.6 Identification of genes under selection

To identify candidate genes associated with high-elevation adaptation in Canary Island *Descurainia,* the program CODEML within PAML v.4.0 (Yang, 2007) was used to perform a *dN*/*dS* (ω) analysis (non-synonymous substitutions per non-synonymous site (*dN*) to synonymous substitutions per synonymous site (*dS*)). The ω ratio measures the direction and magnitude of selection on amino acid changes, with the values ω < 1, = 1, and > 1 generally assumed to indicate a negative purifying selection, neutral evolution, and positive selection, respectively. A Branch-Site Model was employed, which detects positive selection acting on a particular lineage.

To obtain coding sequences (CDS) for the CODEML analysis, we first created a consensus genome for each of the mapped 14 WGRS genomes using our BAM files and our *D. millefolia* reference genome using the SAMtool *consensus* function. These consensus sequences were subsequently converted to a FASTA file with bases with quality lower than 20 were masked to N. CDSs, and their protein counterparts were extracted from each species (eight in total) using the *D. millefolia* genome GFF transcript (Chapter 3) and the software Gffread (Pertea and Pertea, 2020) with the –w and –y parameter. Alignments were performed for the CDSs and protein sequences using BWA across the eight species.

Two inputs are required for CODEML: codon alignments and their corresponding gene tree. PAL2NAL was used to create codon alignments (Suyama et al., 2006) with aligned protein and CDS sequences as input. These resulting codon alignments were then used to generate un-rooted gene trees using RAxMLHPC.

For CODEML Branch-Site analysis, the branch of interest (the 'foreground branch') was selected as the branch leading to clade A, i.e., our high-elevation species (see results) for each of our gene trees. CODEML first constructed a null model (M1) for each gene where all branches were limited to ω = 1 (i.e., excluding positive selection). An alternative model (bsA) was then run where the foreground branch was allowed to display ω > 1 (i.e., including positive selection). CODEML was

first run with a subset of 5,000 genes with different omega parameters, and the omega value with, on average, the highest likelihood score was used in the subsequent full analysis. Due to the large number of genes, the *ete-evol* function within the python package 'ETE toolkit' (Huerta-Cepas et al., 2016) was used to automate the marking of foreground branches and perform the CODEML branch-site analysis on each gene tree. Likelihood ratio tests (LRTs) were used to compare the two implemented models (M1 vs bsA) and adjusted for multiple testing using the Bonferroni test for multiple corrections (Benjamini and Hochberg, 1995).

4.3.7 GO Enrichment analysis

We performed Gene Ontology (GO) analysis (Ashburner et al., 2000) to identify enriched GO Biological Processes, Molecular Function and Cellular Component pathways on our significant ($P < 0.05$) positively selected gene set. Our gene list was tested against annotated gene models from the *D. millefolia* genome (Chapter 3). 'ShinyGO' R package (Ge et al., 2020) was used to conduct the analysis with a False Discovery Rate (FDR) cut-off of 0.05, with the resulting P values adjusted for multiple testing using the Bonferroni test for multiple corrections.

4.3.8 Selection of candidate genes and knockout transformation

From our set of significantly enriched positively selected genes (see results), three genes related to temperature and drought responses were selected to test their function (Table 4.2). *Arabidopsis thaliana* knockout mutants were obtained for these three genes using T(transfer)-DNA lines in the Col-0 ecotype, which knocks out the targeted gene (mutant seeds were obtained from the *Arabidopsis* Biological Resource Center). To confirm that the individuals are homozygous, DNA was extracted using a CTAB protocol (Porebski et al., 1997) from plants grown in environmentally controlled chambers under short-day conditions (10 h light/14 h dark cycles) at 22°C. The presence of the transgene and homozygous mutants was confirmed from PCRs using specific and left-border primers (Appendix Table C.1). Two primer pair reactions were set up for wild-type (WT) and mutant lines, 1) left primer (LP) and right primer (RP), and 2) left border (LB) and RP. Here we used the left border primer LBb1.3 for SALK lines (3' ATTTTGCCGATTTCGGAAC 5'). Seeds from confirmed homozygous mutants were collected and stored at 4°C.

Table 4.2: Descriptions of our three candidate genes selected from our from our positively selected gene list for *Arabidopsis* knockout function experiments.

Gene Names	Araport Gene name	Biological process	References
RD29B	AT5G52300	Encodes a protein that is induced in expression in response to water deprivation, such as cold, high-salt, and desiccation. The response appears to be via abscisic acid.	(Nakashima et al., 2006)
LTI29, LTI45	AT1G20450	Encodes a gene induced by low temperature and dehydration.	(Kim and Nam, 2010)
PK1, PK6	AT3G08730	Encodes a protein-serine kinase that phosphorylates ribosomal protein in vitro.	(Mahfouz et al., 2006)

4.3.9 Assessment of water stress tolerance

Fifty individuals for both wild-type controls (*Arabidopsis thaliana* Col-0) and mutant plants (*A. thaliana* knockouts) were grown for two weeks under normal conditions and then subjected to dehydration stress by ceasing watering for 12 days. Plants were then rewatered for five days for a recovery period.

Response to water availability was assayed by the capability of plants to resume growth when they returned to normal conditions following a period of water stress. The number of surviving plants was counted after the recovery period. Rosette leaves were detached and placed on Petri dishes under dim light. Leaves were weighed at 1-hour intervals to determine water loss with three replicates for each individual. Leaf water loss was expressed as a percentage of initial fresh weight. A Chi-squared comparison of proportions was conducted to determine a significant reduction in survival compared to the control accession.

4.3.10 Quantifying the extent of parallel divergence with other Brassicaceae species

We reviewed studies that detected genes under selection within species of the Brassicaceae family distributed in high latitude and/or elevational habitats. To do so, an online search was performed in PubMed, Web of Science, and Google Scholar for articles published between 2000

and 2022. The search terms used were ‘Brassicaceae’, ‘high-elevation, ‘high-latitude’, ‘adaptation’, and ‘positively selected genes’.

Information was collected from suitable studies, including the year of publication, method of selection analysis, altitude range, location of species, species of study, taxonomic comparison level, and genes under positive selection.

Lists of PSGs were first translated and only included Araport gene names to allow cross-study comparison. To test for a higher-than-expected number of overlapping PSGs, we used Fisher's exact tests, calculated using the 'SuperExactTest' package in R (Wang et al., 2015). We calculated the probability of gene-level parallelism (i.e., gene reuse) between any significant ($P < 0.05$; see results) overlap. Gene reuse is calculated as the number of parallel candidate items divided by the total number of candidate items between them (i.e., the union of candidate lists from both lineages; Conte et al., 2012).

4.4 Results

4.4.1 Whole-genome resequencing, mapping, and variant calling

Between 73 to 96% of reads (33.4 - 49.5 M reads) were retained per sample after quality filtering (Appendix Table C.2). The whole-genome resequencing (WGRS) reads had a breadth of coverage of over 99% for all samples, with an average read depth of 32X. Of the trimmed reads, over 99% of the Canary Islands reads and 71% of the outgroup reads were mapped to the reference genome (Appendix Figure A.1). On average, 0.17% of the mapped reads were PCR duplicates, which were removed. A total of 1,483,762 variable sites were marked across the alignment, and after filtering, a dataset of 353,296 high-quality SNPs was retained for further analysis (Appendix Figure C.1).

4.4.2 Phylogenetic reconstructions

The maximum likelihood phylogenetic reconstruction using our SNP datasets provided a well-resolved phylogeny (90% of nodes show support values of 100%) and resolved two main clades (Figure 4.3A; BS = 100%), which matches the previous phylogenetic analyses (see chapter 2). Clade A comprises *D. preauxiana*, *D. millefolia,* and *D. artemisioides*. *Descurainia millefolia,* the only multiple-island species, was resolved as monophyletic, although with low bootstrap support (27%). *Descurainia artemisioides* is resolved as sister to *D. millefolia*, with *D. preauxiana* resolved as sister to this clade. Within *D. millefolia*, the La Palma and Gomera samples were resolved as a clade. All nodes have bootstrap support of 100, except for the *D. millefolia* clade noted above. Within Clade B, all species are resolved as monophyletic. *Descurainia gilva* and *D. gonzalezii* are resolved as sister taxa with *D. lemsii* as sister to this clade and *D. bourgaeana* as sister to the *D. gilva-D. gonzalezii-D. lemsii* clade).

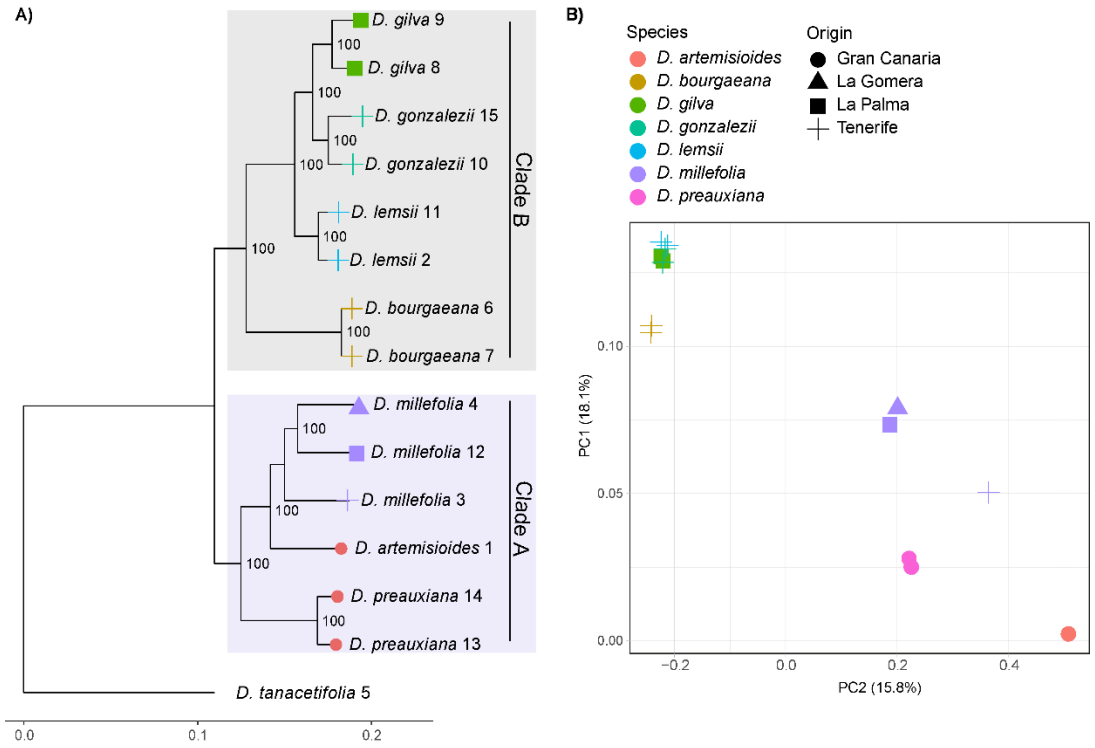


Figure 4.3: A) Maximum-likelihood phylogeny using SNPs from 14 individuals of Canary Island *Descurainia* and *D. tanacetifolia*, from the Mediterranean, as the outgroup. Species ID numbers correspond to Table 4.1 and Appendix Table C.2. Tip colours represent the island of origin. Numbers represent bootstrap values > 70%. B) A principal component analysis (PCA) representing PC1 and PC2 of the 14 individuals of seven Canary Island *Descurainia* based on SNPs. Colours represent species, and shape represents the island of origin.

4.4.3 Genetic structuring

The PCA analysis provides similar genetic structuring to our phylogenetic reconstructions, where individuals within the two clades, clade A (*D. artemisioides, D. millefolia,* and *D. preauxiana*) and clade B (*D. bourgaeana, D. gilva, D. gonzalezii,* and *D. lemsii*) are clustered together (Figure 4.3b), with a clear separation between the two groups. All individuals within the species clustered tightly except for Tenerife *D. millefolia,* which was distinct from the other *D. millefolia* individuals sampled from La Gomera and La Palma. Species that were resolved within clade A were clearly separated in the PCA plot in contrast to those from Clade B, which were more tightly clustered.

The F_{ST} values followed the structuring found in our PCA analysis. The largest average value was between *D. bourgaeana* and the two Gran Canarian endemics *D. preauxiana* and *D. artemisioides* at 23% (Appendix Table C.3), while the lowest values were between *D. gilva* and *D. gonzalezii* at 0.5%.

4.4.4 Climate analysis

Analysis of climatic variables revealed that clade B, consisting of our high-elevation species (*D. bourgaeana, D. gilva, D. gonzalezii,* and *D. lemsii*), are exposed to more significant precipitation during the winter months, solar radiation during the summer months, and wind speed all year round, compared with the low-elevation clade A species ($P < 0.05$; Figure 4.4; Appendix Table C.4). Conversely, clade A species (*D. millefolia, D. artemisioides,* and *D. preauxiana*) were consistently exposed to greater mean temperature and water vapour pressure all year round.

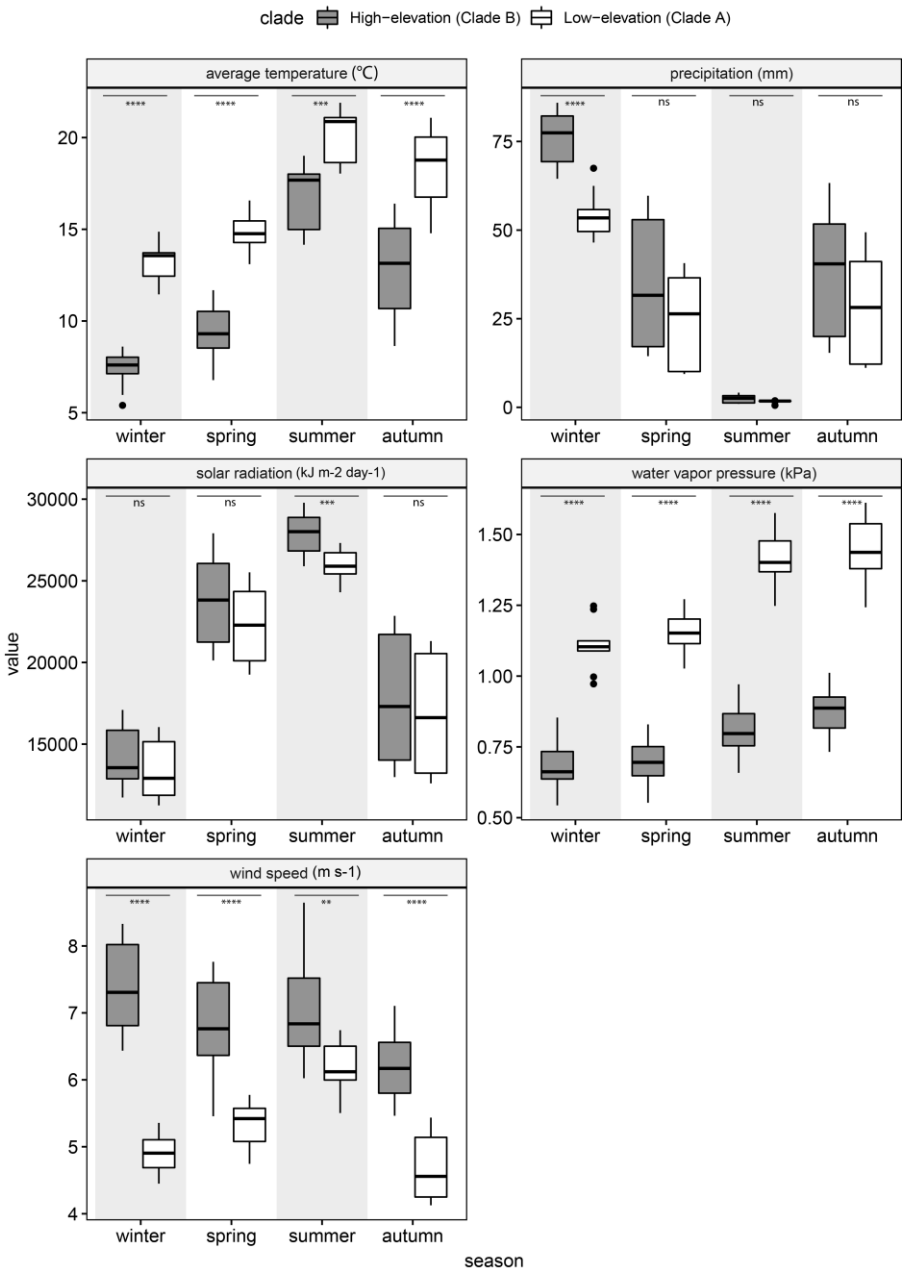


Figure 4.4: Five climate variables across the four seasons, taken from GPS coordinates of seven Canary Island *Descurainia* species, were extrapolated using location data from herbarium specimens deposited in the Natural History Museum (BM). Climatic values were grouped into species retrospective clades, A and B, which represent distinctive elevation ranges. Seasons are defined as follows: Winter = December – February; Spring = March–May; Summer = June – August; Autumn = September – November. Asterisk (*) represents significance level as determined by ANOVA. 'ns' = not significant.

4.4.5 Identification of genes under selection

After filtering for a minimum of one individual per species, 25,394 CDS regions (91%) were included in the selection analyses, from which we identified 276 positively selected genes (PSGs; ca. 1.1% of the total genes examined) on the branch leading to our high-elevation clade (Appendix Table C.5). Of these 276 PSGs, 53 (19.2%) had unknown function according to our *D. millefolia* annotations. The GO enrichment analysis found five significantly enriched Biological Processes GO terms and six significant Molecular Function pathways (FDR-corrected $P < 0.05$), compared to our background gene list (Figure 4.5; Appendix Table C.6). There were no significant Cellular Component pathways.

Four significantly enriched Biological Processes pathways (*negative regulation of flower development,* GO:0009910; *reproductive processors,* GO:0022414*; post-embryonic development,* GO:0009791; *multicellular organismal processors,* GO:0032501) share the same five genes: two MADS AFFECTING FLOWERING genes (MAF1 and MAF2/AGL31). APUM9, FRI (FLOWERING LOCUS A), and MET1 (DECREASED DNA METHYLATION 2) (Figure 4.5). These five are associated with processors that stop, prevent, or reduce the frequency, rate, or extent of flower development, which are directly relevant to typical changes found in subalpine plants. For example, FRI induces late flowering (Fournier-Level et al., 2022), and AGL31 prevents vernalisation by short cold periods (Scortecci et al., 2001; Dai et al., 2020). The fifth pathway is *cellular response to stress* (GO:0033554), which comprises 19 PSGs (full summary in Appendix Table C.6). Within this group, there are several genes with GO terms associations that may be associated with climatic stress and/or other stressors associated with high-elevation. For example, *response to heat stress* (GO:0034605), *response to cold* (GO:0009409)*, response to osmotic stress* (GO:0006970)*, response to hypoxia* (GO:0001666)*,* and *response to UV* (GO:0009411).

Four significantly enriched Molecular Function pathways (FDR < 0.05) were related to Nicotinamide adenine dinucleotide (NAD; *NAD+ nucleosidase activity,* GO:0003953*; NAD(P)+ nucleosidase activity,* GO:0050135*; NAD+ nucleotidase, cyclic ADP-ribose generating,* GO:0061809; *ADP binding,* GO:0043531). These four pathways share the same six genes, with *ADP binding* including an additional gene (Figure 4.5; Appendix Table C.6). The remaining two pathways were related to the protein kinase, which included *Protein serine kinase activity* (GO:0106310; 10 genes) and *Protein serine/threonine kinase activity* (GO:0004674, 12 genes).

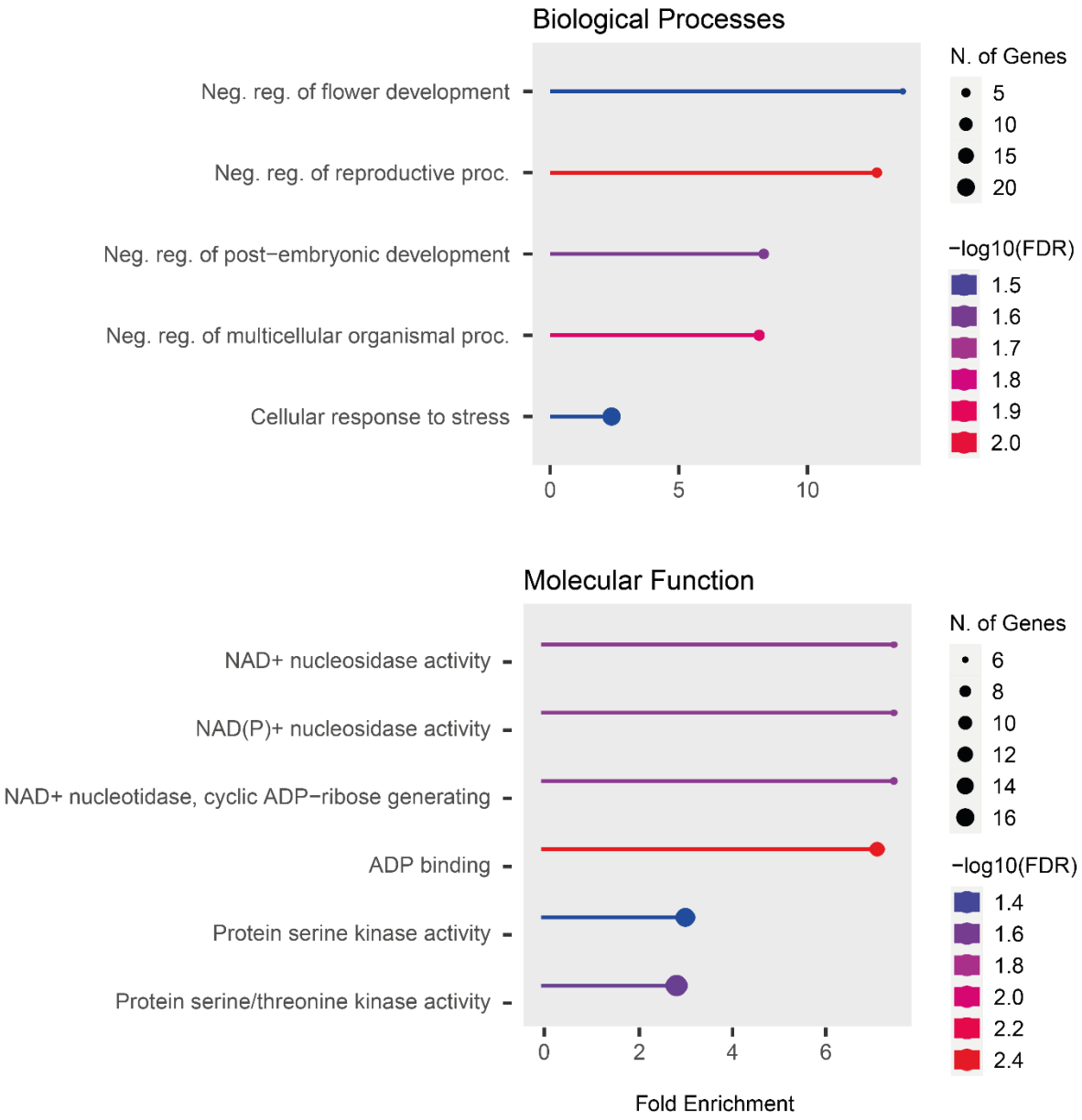


Figure 4.5: The significantly enriched Biological Processes (Top) and Molecular Function (Bottom) GO terms (FDR < 0.05) for the PSGs found within the high-elevation clade of Canary Island *Descurainia,* ranked by fold enrichment. The size of the dot represents the number of genes. Colour the intensity of the fold enrichment ratio (log10; red = high, blue = low). The X-axis is the level of fold enrichment.

4.4.6 T-DNA insertions in candidate genes and water stress experiment

After the recovery period from the drought stress, the T-DNA knockout for RB29B (AT5G52300, SALK_052958) demonstrated significant mortality ($P < 0.05$), where less than 50% of individuals survived (Figure 4.2; Appendix Table C.7). This is in comparison with our wild-type control, *Arabidopsis thaliana* (Col-0), which exhibited a 78% survival rate. While individuals for knockout for LBI29 showed no significant reduction in survival compared to the control, LBI29 knockouts had the highest level of water loss post-recovery period, followed by RB29B knockouts. The T-DNA knock for gene PK1 (SALK_112945) showed negligible differences with the level of water loss compared and similar survival levels to the control to the control.

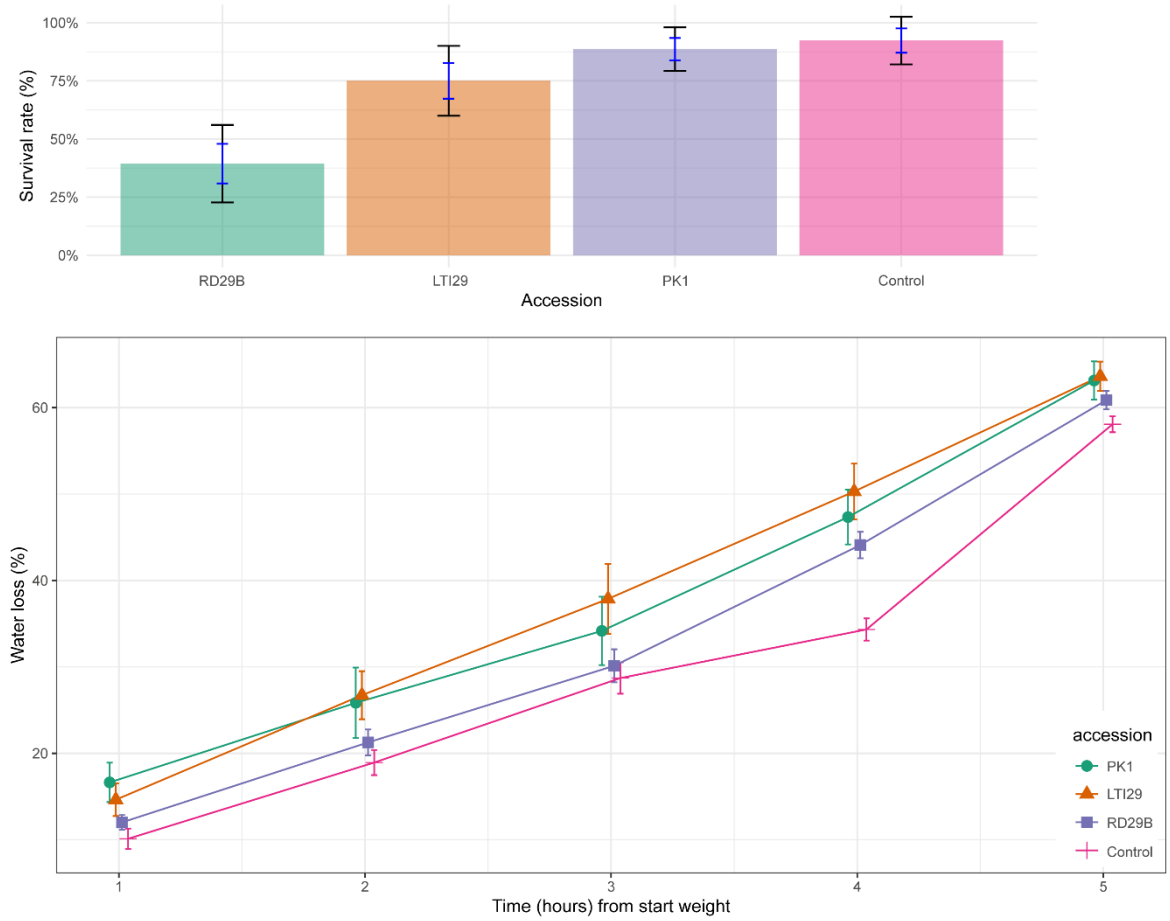


Figure 4.6: Effects of a water availability stress on three T-DNA *Arabidopsis* knockouts for three different candidate genes and one wild-type (*Arabidopsis* Col-0) representing the control. Drought treatment included withholding water for 10-days and a 5-day watered recovery period. Drought tolerance was measured by survival (top) which includes SE (blue) and SD (black) bars, and the percentage (%) of water loss from leaves (bottom) with SE bars, after 5 hours after the recovery period. Details of the accessions are found in Table 4.2.

4.4.7 Comparison with Brassicaceae studies

Using PSG sets from eight studies (including our own set from *Descurainia;* Table 4.3), a total of 28 tests were performed for 2-degree gene set comparisons. Two pairs showed significant overlap (i.e. a greater than expected overlap in the PSGs identified, Bonferroni-adjusted $P < 0.05$, Fisher's exact test) across tests (7.1% of tests; Figure 4.7; Appendix Table C.8). The comparison of PSGs in *Descurainia* and *Draba nivalis* (typically found in Arctic-alpine zones), revealed 23 significantly overlapping PSGs (Appendix Table C.9), with a gene reuse probability of 1.98%. Sixteen of these genes had Biological Process GO terms associated with them. A number can be associated with

high-elevation-like environmental conditions, such as PME41, which is associated with the GO term *response to cold* (GO:0009409; Qu et al., 2011). There is also a number related to the GO term *root development* (GO:0048364), such as LOX1 and PRX44, which are both involved in the polar expansion of roots (Nagase et al., 2001; Marzol et al., 2022). The second significant overlap was for *Arabidopsis halleri* and *D. nivalis,* where 22 genes overlapped with a gene reuse probability of 1.76%. Seventeen of these genes have Biological GO terms associated with them, including *flowering development* (GO:0009908), *hyperosmotic salinity response* (GO:0042538), and *stomatal closure* (GO:0090332).

Table 4.3: The seven studies and the positively selected genes (PSG) sets included within our comparative analysis. ω = *dN/dS*, GWAS = Genome Wide Association.

Species	Study location	Elevation (m)	Reference	Comparison level	Statistical test	N PSGs (% of proteome)
Arabidopsis halleri	Japan	200-1,377	(Honjo and Kudoh, 2019)	Population	GWAS	281 (0.86)
Arabidopsis lyrata	Norway	10-1,360	(Hämälä and Savolainen, 2019)	Population	GWAS	423 (1.29)
Arabidopsis thaliana	Italian Alps	580-2,350	(Günther et al., 2016)	Population	π and Tajima's D	19 (0.07)
Crucihimalaya himalaica	Himalayan	Ca. 4,000	(Zhang et al., 2019)	Family	ω	610 (2.26)
Crucihimalaya lasiocarpa	Qinghai–Tibet Plateau	Ca. 4,000	(Feng et al., 2022)	Genus	ω	403 (1.67)
Draba nivalis	Arctic (Alaska)	1,419-1,990	(Nowak et al., 2021)	Family	ω	969 (2.89)
Lepidium meyenii	Central Andes	2,350-4,500	(Zhang et al., 2016)	Family	ω	74 (0.07)

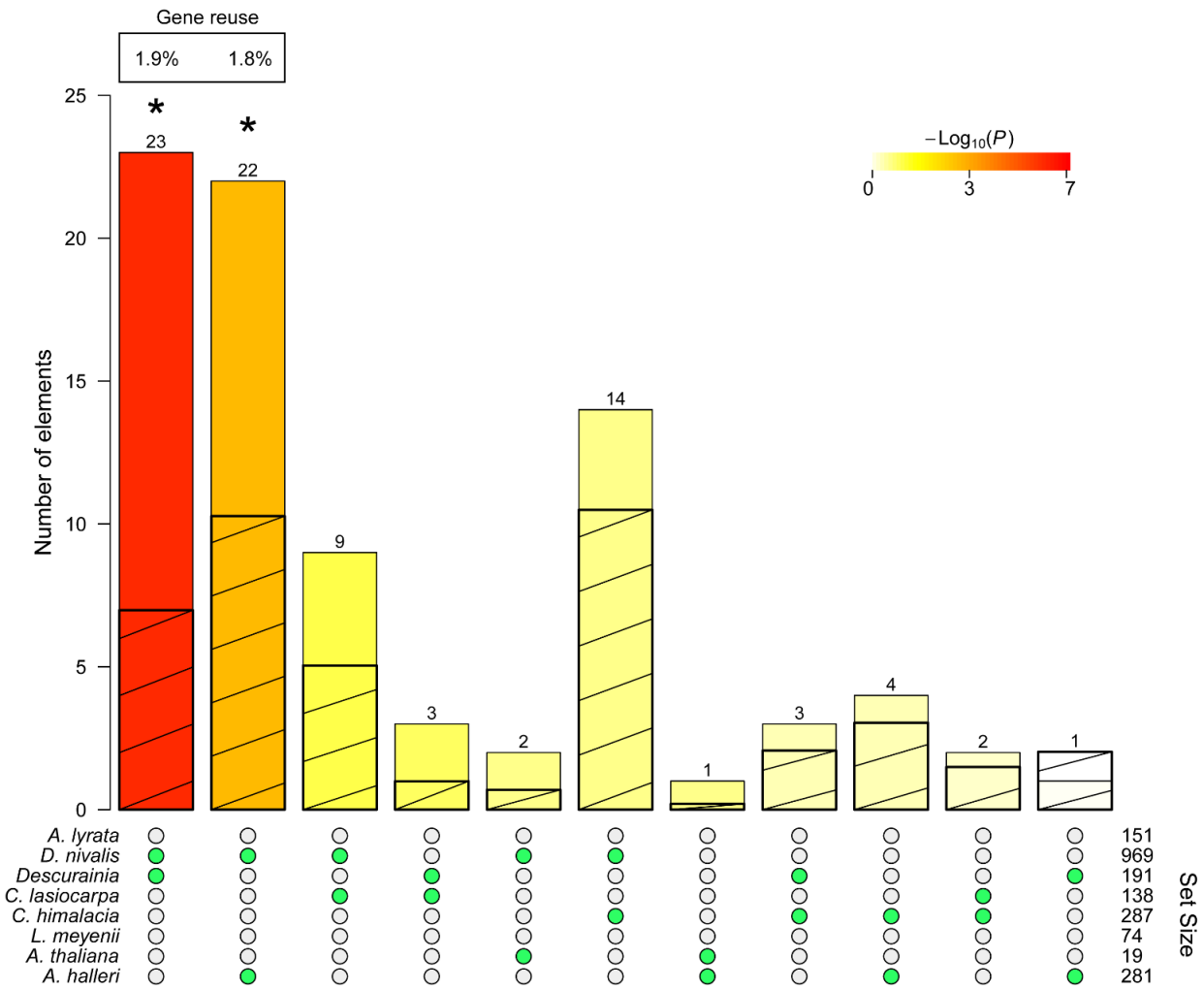


Figure 4.7**:** The pairwise comparison that provided overlapping PSGs between studies which identified PSGs in high-elevation and arctic Brassicaceae species (ordered by smallest P values). Green dots show species pairs in each case. Bars represent the observed overlap, and hatch bars represent the expected overlap. The colour intensity of the bars represents the P value significance of the intersections, and those significant (P < 0.05) intersections have * symbol. The percentage (%) of gene reuse is shown above significant overlaps.

4.5 Discussion

The diversification of plant lineages in the Canary Islands is strongly influenced by the climatic cline exhibited by the taller islands (Tenerife and La Palma). The different light, temperature, and water availability regimes between low-, mid-, and high-elevation habitats are frequent explainers of the islands high levels of diversity and endemism (Fischer et al., 2013; Zhou et al., 2014; Kubota et al., 2015; Holliday et al., 2016). These different climatic pressures can have a significant influence on plant morphological, physiological and life-history characteristics, for example, vernalisation, the timing of flowering, or bud burst (Leinonen et al., 2009; Alberto et al., 2013; Csilléry et al., 2014; Zhang et al., 2019; Mishra et al., 2020; Nowak et al., 2021). Our previous work (Chapter 2) showed that Canary Island *Descurainia* had undergone one phylogenetically linked climatic shift between the two major clades, which we hypothesise to be associated with an elevational range (i.e., Clade A species range between 50 – 1,100 m, and Clade B species range between 1,800 – 2,300 m; see results of Chapter 2). Here we analysed whole-genome resequencing data for all seven species (13 individuals) of Canary Island *Descurainia* to investigate the molecular basis of adaptation to high-elevation habitats. We compared coding sequences of the low-elevation species (Clade A; *D. millefolia, D. artemisioides* and *D.* preauxiana) with the high-elevation species (Clade B; *D. bourgaeana, D. lemsii, D. gonzalezii* and *D. gilva*) to identify putatively adaptive genes related to ecological adaptation.

Our climatic analysis demonstrated that, indeed, high-elevation habitats for Canary Island *Descurainia* were significantly different to those in low-elevation habitats in several climatic variables, including temperature, UV, precipitation, and wind speed. The selection analysis revealed candidate genes that are potentially associated with adaptation strategies related to the sub-alpine environments, including divergence in flowering time and water stress tolerance. We performed functional validation on three candidate genes related to abiotic stress by performing water availability stress experiments on *Arabidopsis* T-DNA knockouts. Two out of three of our accessions demonstrate a reduction in fitness, providing further evidence that divergent selection on these genes could incur adaptive advantages in the respective environments which differ in water availability.

When comparing our list of positively selected genes in our high-elevation species of Canary Island *Descurainia* with other species in Brassicaceae that occupy potentially similarly stressful environments, we found little evidence of convergence between species. However, there was a

significant overlap with the PSGs of Draba nivalis (Nowak et al., 2021) and Descurainia, which are associated with abiotic stressors commonly related to alpine environments, providing evidence for the occurrence of similar molecular trajectories to adaptation.

4.5.1 Evolutionary of Canary Island *Descurainia*

Our ML phylogeny, using ca. 353 k SNPs across 14 individuals, resolved two clades within the Canary Island *Descurainia* group, namely clades A (*D. millefolia, D. artemisioides,* and *D. preauxiana*) and B (*D. bourgaeana, D. gilva, D. gonzalezii,* and *D. lemsii*), consistent with our previous study derived from Genotyping-by-sequencing (Chapter 2). A notable difference between the results of these two studies was the position of *D. lemsii,* which is resolved as sister to *D. gilva* and *D. gonzalezii,* whereas in Chapter 2, *D. lemsii* was nested within *D. bourgaeana.* This supports the hypothesis that the *D. lemsii* individual sequenced in Chapter 2 was possibly a misidentification. Due to the lack of a herbarium specimen for this individual in Chapter 2, formal identification can not be made, and we cannot provide a more decisive conclusion.

Our WGRS phylogeny also adds a previously unsampled population of *D. millefolia* from La Gomera, revealing it to be sister to the La Palma population (*D. millefolia* being the only multiple islands endemic species of *Descurainia* distributed on Tenerife, La Gomera and La Palma). Whilst the individuals of most species clustered tightly, the Tenerife *D. millefolia* specimen showed genetic divergence from individuals from the other two islands in both the PCA and the phylogenomic analysis. La Gomera and La Palma represent the more Western islands and are geographically close to each other. Hence, we might expect individuals from these islands to cluster together.

4.5.2 Ecological adaptation within Canary Island *Descurainia*

Our previous work on Canary Island *Descurainia* revealed one significant phylogenetically linked ecologically shift between the two clades (Chapter 2). The results presented in this chapter confirmed the environmental divergence associated with clade A (low-elevation) and clade B (high-elevation) species (Figure 4.4) and provided a finer-scale insight into the climatic differences between low- and high-elevation *Descurainia*. The locations from which higher-elevation species exhibited lower mean temperatures all year round, lower vapour pressure all year round, higher

precipitation during winter, higher solar radiation in the summer, and highest wind speeds all year round. These results are consistent with the typical characterisation of alpine and sub-alpine habitats (Nagy and Proctor, 1997).

To investigate the genes and biological processes that underpin this significant ecological transition, we used a Branch-Site selection approach to test for signatures of selection on the branch leading to our high-elevation species. A total of 278 genes (0.98% of available genes) showed evidence for significant positive selection, with our GO (Gene Ontology) enrichment analysis suggesting biological pathways related to high-elevation adaptation, such as reproduction and development and abiotic stress. We found a relatively small number of enriched GO terms compared to similar studies; this may be due to a large portion of our PSGs with no annotations (19.2%).

4.5.2.1 Reproduction and Development

We identified 21 PSGs representing five GO terms associated with reproduction and development, notably *negative regulation of flowering development* (GO:0009910), within our high-elevation species. One of the most important components of reproduction in plants is flowering due to its direct relationship with reproductive success. The flowering season for high-elevation plants is commonly restricted to accommodate shorter day-length (Dupont et al., 2003; Halbritter et al., 2018) and to avoid the risk of frost damage (Kollas et al., 2014). Therefore, changes in reproductive strategies to accommodate elevational shifts are often expected within high-elevation floras (Leinonen et al., 2009; Cheviron and Brumfield, 2012; Feng et al., 2022). Within the Canary Islands *Descurainia* clade, *D. bourgaeana* and *D. gilva* are the species with the highest elevation distributions (1,800 – 2,300 m) and have a flowering period of one month (ca. mid-May to mid-June; Appendix Figure C.2). *Descurainia gonzalezii* and *D. lemsii* have a slightly longer flowering period (April – July). This contrasts with species distributed at lower elevations (e.g., *D. millefolia, D. artemisioides,* and *D. preauxiana*) which have flowering periods of between five and eight months (January - September; Bramwell, 1977).

Several notable flowering time genes were under positive selection within our high-elevation lineage. One such gene is *FLOWERING LOCUS A* (FRI), which is a major determinant of flowering time in *Arabidopsis* (Zhu et al., 2021; Fournier-Level et al., 2022) and in homologues across

species (Tadege et al., 2001; Schranz et al., 2002; Okazaki et al., 2007; Albani et al., 2012; Guo et al., 2012); it confers vernalisation (i.e. flowering induced by prolonged exposure to the cold) requirement causing plants to overwinter vegetatively, often significantly reducing flowering time. Our climate analysis demonstrated the higher-elevation *Descurainia* species experience significantly cold environments, particularly during the winter period, where the contrast is greatest (43% colder compared to the low-elevation species). Similarly, clinal patterns in flowering time in European *A. thaliana* populations found an increase in the expression of FRI at higher elevations (Fernández-Palacios and de Nicolás, 1995; Caicedo et al., 2004; Stinchcombe et al., 2004). In another study, FRI was found to be under positive selection in *Arabidopsis* populations that have colonised the arid Cape Verde Islands, with a functional analysis demonstrating a ca. 30-day reduction in flowering time (Fulgione et al., 2022). However, beyond *Arabidopsis,* there is little evidence for FRI being an adaptive trait in other high-elevation species, although these studies are somewhat lacking (Keller et al., 2018; Geng et al., 2021). Nonetheless, our study suggests that colder climates have driven selection on the flowering time gene, FRI.

Two genes within the *MADS AFFECTING FLOWERING* (MAF) gene family (MAF1 and MAF2/AGL31) were also found to be under positive selection, which regulates flowering time (Scortecci et al., 2001; Ratcliffe et al., 2003; Rosloski et al., 2013; Ortuño-Miquel et al., 2019). MAF1 gene has been shown to inhibit flowering and overexpression, producing late-flowering plants (Scortecci et al., 2001), and MAF2 is a transcription factor that prevents vernalisation during short cold periods, acting as a flora repressor (Ratcliffe et al., 2003). Similar results were found in a population-level study of *A. thaliana* in the Swiss Alps which found the gene expression MAF genes mediated flowering responses to vernalisation along an altitudinal gradient (Günther et al., 2016). *MADS AFFECTING FLOWERING* genes have also been proposed to facilitate rapid adaptation (Theißen et al., 2018). Our results strongly suggest that adaptation to divergent altitudes is, in part, caused by the different flowering durations necessary in the different environments.

4.5.2.2 Abiotic stress tolerance

4.5.2.2.1 Thermal

Along a large elevational cline, the temperature can vary considerably. For example, the mean annual temperature in Tenerife ranges from ca. 22°C on the south coast to 2°C at the Teide peak (Fernández-Palacios, 1992). With a decrease in temperature, the frequency of frost events also increases. The climatic analysis demonstrated the high-elevational species of Canary Island

Descurainia experience significantly lower temperatures across all seasons compared to their low-elevation conspecifics (see also Chapter 2). Adaption to lower temperatures is vital to the survival of plants as it can have a number of physiological and molecular consequences (Körner, 2016). For example, effects on the membrane fluidity, stability of RNA and DNA, and enzyme activity (Bowles et al., 2002). Low temperatures also impose dehydration stress by lowering water absorption by the root and water transport in roots (Theocharis et al., 2012).

Our results found ten PSGs within our significant enriched GO term *response to abiotic stress* (GO:0009628) that have been associated with temperature stimulus. Two notable genes encode for heat shock proteins (HSPs), a conserved protein family which are produced in response to stress. The first HSP is HSP70-2 (a member of the cytosolic heat shock 70 protein family) which negatively regulates heat tolerance by negatively repressing the heat shock factor (Tiwari et al., 2020). Among the other PSGs genes in *response to abiotic* stress is ATJ49, a chaperone cofactor-dependent protein which enables Hsp70 protein binding (Gaudet et al., 2011). Typically, HSP70 genes are known for playing a critical role in high-temperature stressors in plants, but the over-expression of HSP70s has also been found to enhance tolerance to other stresses, such as drought, cold and salt stress (Alvim et al., 2001; Ono et al., 2001; Masand and Yadav, 2016; Tang et al., 2016; Zhao et al., 2019). For example, a study of the wild crucifer *Thlaspi arvense* (Brassicaceae) was found to up-regulate HSP70-2 during cold stress (Sharma et al., 2007).

The second HSP, CLPB4 (HSP98.7, a member of the heat shock 100 protein family), confers thermotolerance to chloroplasts during heat stress (Myouga et al., 2006). Transcripts of CLPB4 accumulate dramatically at high temperatures (Lee et al., 2007). Additional genes include the MDO1 gene that rapidly responds to temperature, which stabilities telomeres in response to heat stress (Lee et al., 2016). Due to the lower temperatures exhibited in these high-elevation habitats, these PSGs associated with thermal tolerance may be down-regulated.

4.5.2.2.2 UV Radiation

For every 1,000 m in elevation, the UV radiation level will increase by about 12% as the thinner atmosphere filters out less UV radiation (Blumthaler et al., 1997). An excess of UV radiation is well known to result in DNA damage, including mutagenic and cytotoxic DNA lesions (Tevini and Teramura, 1989; Strid et al., 1994; Jansen et al., 1998; Sinha and Häder, 2002; Verdaguer et al., 2017); thus high-elevation plants often have UV-resistant genes (Zhang et al., 2019; Feng et al.,

2022). During the summer, the high-elevation *Descurainia* clade experiences significantly more solar radiation than lower elevations, notably *D. bourgaeana,* which occupies those highest elevations and exposed habitats (Chapter 2). Our selection analysis found two genes under positive selection under the significantly enriched GO term *response to abiotic* stress, which are associated with UV tolerance: The first is gene ERCC1 (UV hypersensitive 7; AtERCC1 is the *Arabidopsis* homologue), a DNA repair gene which plays a key role in nucleotide excision report to the endonuclease, which is associated with a response to UV-B and gamma radiation (Jiang et al., 1997; Hefner et al., 2003; Dubest et al., 2004). The second gene is UVSSA, a lesser-studied gene, but *Arabidopsis* knockouts demonstrate a contribution to UV tolerance believed through DNA repair (Al Khateeb et al., 2019).

4.5.2.2.3 Water availability

Water availability and drought have repeatedly been associated high elevation environments (Nagy and Grabherr, 2009), and drought tolerance has been demonstrated in a number of Canary Island plants (Lausi and Nimis, 1986; Grill et al., 2004; López et al., 2008, 2009; González-Rodríguez et al., 2017). The higher elevation clade has greater higher precipitation; however, the accessibility of water for high-elevation plants is negligible: (1) During the winter months, water is often frozen due to frequent frost periods; (2) more significant drought during the dry season; (3) The substrate on the volcanic soil is permeable (lava and pumice), which allows free drainage; and (4) significantly higher wind speeds throughout the year reduces available water. These factors lead to extensive drought stress for many plant species in the sub-alpine zone. For example, the soil moisture in the uppermost layer (0-10 cm) is very low, at 0-4% (Gieger and Leuschner, 2004; Brito et al., 2013).

Several genes associated with water availability were found under the enriched GO term *response to abiotic stress.* One such gene, Receptor-like Protein Kinase 1 (RPK1), is thought to be a general osmotic stress response gene and has been found to reduce the level of water lost during a drought (Hong et al., 1997). RPK1 is localised within the plasma membrane and functions as a regulator of abscisic acid (ABA) signalling, ABA playing a pivotal role in stress responses (Lee et al., 2011; Dai et al., 2018). *Arabidopsis* overexpressing RPK1 demonstrate increased root growth and stomatal closure, as well as reduced transpiration water loss (Osakabe et al., 2010). Similar water retention behaviour is seen in other high-elevation Canary Island species. For instance, there is a correlation between needle water retention, partial stomatal closure, and increased elevation in *Pinus canariensis* (ca. 1,600 - 2,000 m; Gieger and Leuschner, 2004).

Another PSGs was RD29B, a gene induced in response to various abiotic stressors, including cold and desiccation (Yamaguchi-Shinozaki and Shinozaki, 1993; Nakashima et al., 2006). The RD29B gene is one of 27 genes commonly induced during or repressed in water deficit studies in *Arabidopsis thaliana* (Bray, 2004). Similarly, a study in soybean also found this gene to be expressed during water stress (Bihmidine et al., 2013)

4.5.2.3 Biotic stress

Adaptation to a novel environment often leads to exposure to new diseases and pathogens. Plant pathogens and the microbial community can have multiple effects on plants, ranging from disease (Smith et al., 2006) and mycorrhizal affecting the plant life cycle (Bolan, 1991). Several genes (RLM1B, RLM1A, and RPP1) under the significantly enriched Molecular Pathway GO terms also have the annotations GO terms *defence response to bacterium* (GO:0042742) and *defence response to fungus* (GO:0050832) (Rehmany et al., 2005; Staal et al., 2006; Creff et al., 2010). Specifically, for example, the RPP1 gene incurs resistance to Downy Mildew in *Arabidopsis* (Rehmany et al., 2005), the RLM1B gene confers resistance to the pathogen *Leptosphaeria maculans* (blackleg disease; Staal et al., 2006). There is current a lack of research focusing on the microbial communities within the Canary Islands, but other studies have found associations between soil composition and plant diversification. Most notably, Osborne et al. (2018) demonstrated that plant pathogen (fungal and bacterial) abundance difference between soil types on the oceanic island of Lord Howe and the sympatric speciation of Howea palms is associated with soil preference (Savolainen et al., 2006).

4.5.3 Testing candidate genes under water availability stress

While methods such as *dN/dS* analyses can identify genes putatively under selection, there is still a need to confirm the function and role of these genes to demonstrate the divergence is due to adaptive selection (Luikart et al., 2003). Here we tested the impact on fitness through water availability stress on T-DNA *Arabidopsis* lines containing potential knockouts of three candidate genes. Two of our T-DNA knockouts for genes RD29B and LTI29 were found to be less tolerant to drought stress compared to the control. T-DNA knockouts for RD29B demonstrated the most significant intolerance in both of our measures, survival and rate of water loss in leaves. The results for these gene are consistent with previous studies examining gene functions (Nordin et

al., 1993; Yamaguchi-Shinozaki and Shinozaki, 1993; Nakashima et al., 2006). In comparison, T-DNA knockouts for LTI29 showed no significant reduction in survival but demonstrated increased water loss. LTI29 has been annotated with the GO term *response to water deprivation* through a similar study using T-DNA knockouts and water stress conditions (Kim and Nam, 2010). However, this study measured seed germination in response to stress, with T-DNA knockouts displaying reduced germination. Since our study only used seeds that had successfully germinated, it may suggest that LTI29 may play a lesser role in drought tolerance in established individuals.

The third candidate gene, PK1/PK6, demonstrated no significant difference from our control plant in survival and leaf water loss after drought treatment. The activity of the PK1 gene is affected by osmotic stress through the translational up-regulation of ribosomal proteins. Plants overexpressing PK1 are shown to be hypersensitive to osmotic stress, and water-stressed tobacco plants demonstrated significantly reduced PK1 activity (Mahfouz et al., 2006). The reduction of PK1 activity under unfavourable conditions may be a mechanism employed by plants to sustain harsh conditions. Therefore, the T-DNA knockout of this gene could have a neutral or even positive response to drought treatment and support the results of Mahfouz et al. (2006). It is likely that within our high-altitude clade, this gene may be upregulated in response to high levels of seasonal drought. A knockout in a species adapted to high elevation (as opposed to *Arabidopsis* Col-0) would allow us to assess this possibility.

By mapping reproductive isolation to candidate genes and using experimental methods, such as positional cloning, gene replacement/ knockout, gene expression assays, and transgenic manipulations to demonstrate function (Juenger et al., 2000), we can provide further evidence that these genes are adaptive in providing a fitness potential to survive in such environments. Such research can have the potential to guide additional avenues of research, namely crop improvement.

4.5.4 Evidence of convergent selection across Brassicaceae species adapted to high elevation and latitude environments

In this study, 7.4% of pairwise comparisons showed overlap between genes in high-elevation and -latitude Brassicaceae species, with two comparisons (*Draba nivalis* and *Descurainia* and*,* and *D. nivalis* and *Arabidopsis halleri*) providing a significant result. Many of the findings of genomic

parallelism are between closely related species as gene reuse decrease with increased divergence (Bohutínská et al., 2021). However, our results provide evidence of parallism between diverged individuals as *Draba nivalis* represents a different lineage within the Brassicaceae phylogeny (Lineage II) compared to *Descurainia* and *A. halleri* (lineage I; Nowak et al., 2021).

Of the 23 genes shared between the two lineages, at least three (13%) have been associated with abiotic stressors. One such gene is OSCA1, an osmosenor responsible for Ca^{2+} increase induced by osmotic stress in plants (Zhang et al., 2020; Pei et al., 2022). In addition, the LOX1 gene has been shown to be induced during root responses to osmotic stress (Chen et al., 2021), and transgenic mutants have been shown to increase the length of the primary root (Vellosillo et al., 2007). It is likely the LOX1 gene is involved in the increase of deeper root formation during a period of water stress, a common adaptation found within alpine plants (Bliss, 1962). The gene PME41 is an important cell wall enzyme that plays a role in cold and freezing tolerance (QU et al., 2011). Comparative gene expression studies indicate the response to temperature and moisture is highly conserved in certain plant families (Yeaman et al., 2014). Convergence can be expected in similar environments because they exert similar selection pressures, generating non-random and repeated outcomes in independent lineages (Stuart, 2019).

4.6 Conclusion and Future Direction

The annotated high-quality *Descurainia millefolia* reference genome generated in Chapter 3 provided a foundation for advancing hypotheses on the evolutionary processes involved in the the rapid diversification of Canary Island *Descurainia*. Here we examine the molecular basis of ecological divergence within a Canary Island radiation using whole-genome datasets. Canary Island *Descurainia* is an endemic group of seven species that have undergone rapid radiation between low-elevation generalist habitats to more specialised high-elevation habitats. Through this comparison of the high-elevation versus low-elevation species using whole-genome resequencing datasets, we identified several genes associated with many environmental stressors associated with high-elevation habitats. The more notable genes that were significantly enriched were those related to reproduction, especially flowering. Further, we present evidence of fitness effects associated with one of the genes we found under positive selection by using *A. thaliana* knockouts. Understanding the genomic basis of adaptation to extreme stressors can ultimately have larger implications and provide positive benefits in the future of an integrative strategy combining genomics, epigenetics and genome editing to establish resilient crop varieties.

4.7 Acknowledgements

The authors would like to thank Alfredo Reyes-Bentacort and Arnoldo Santos-Guerra for their assistance in Canary Island field sampling. We are grateful to the national parks for permits to collect. Permit numbers are as follows: Cabildo de Tenerife (no. 22835), Parque Nacional del Teide (no. 42917710), Cabildo de La Palma (no. 2019004595), Cabildo de Gran Canaria (no. 14328), Cabildo Insular De La Gomera (no. 4319), Área de Sostenibilidad, Medio Ambiente y Seguridad (no. 42917710D). This work was supported by a PhD studentship to ACJ from the Natural Environmental Research Council [grant number NE/L002531/1]. We acknowledge the use of the IRIDIS High-Performance Computing Facility and associated support services at the University of Southampton, in the completion of this work.

Chapter 5 Conclusions and Discussion

The endemic Canary Island clade of *Descurainia* (Brassicaceae) comprises seven species that have diversified across four islands into different habitat types and altitude ranges, from a continental, Mediterranean or Euro-Siberian lineage. This group provides an excellent model to explore speciation and adaptation processes that underlie the rapid diversification of island lineages using whole genome sequencing approaches, given its small genome (relative to other plants) and close taxonomic relationship with the well-characterised model organism *Arabidopsis thaliana*. This thesis had three main aims:

1. To revisit the evolutionary relationships of Canary Island *Descurainia* using NGS sequence data to determine the role that ecology, geography and hybridisation have played in the diversification of this lineage.
2. To generate a high-quality reference genome for *D. millefolia* so it can be utilised to address evolutionary and ecological questions and understand the relationship of *Descurainia* within the family Brassicaceae.
3. To identify candidate genes involved in adaptation to high-elevation habitats within Canary Island *Descurainia* and assess the level of convergence between other species in Brassicaceae.

5.1 Review of Genomics Datasets Generated

5.1.1 Genotyping-by-Sequencing

In chapters two to four, I utilised two different genomic datasets to address the three main aims of this thesis. The first dataset, primarily used in chapter two, was generated using the next-generation sequencing technique named genotyping-by-sequencing (GBS). Genotyping-by-sequencing is an inexpensive method that detects and sequences many SNPs across individuals without needing a reference genome. GBS studies have proven to be a powerful tool for phylogenetic analyses, particularly in non-model organisms (Elshire et al., 2011; Chung et al., 2017; Fernández-Mazuecos et al., 2018), the benefits of which are discussed in chapter two.

In chapter two, our GBS dataset for Canary Island *Descurainia* generated a large number of SNPs (ca. 350 k) which overcame the limitations of the genetic markers utilised by Goodson et al.

(2006) and resulted in hypotheses of evolutionary relationships that were incongruent with previous hypotheses. The application of increasingly reliable sequencing libraries is being applicated to oceanic flora (e.g., Dunning et al., 2016) and Canary Island flora groups (Curto et al., 2018; White et al., 2020; Graham et al., 2021; Mort et al., 2022), providing more up-to-date phylogenies which are more well-resolved and robust. Phylogenetic reconstructions are the unpinning of many evolutionary hypothesises, therefore the importance of continuing producing up-to-date phylogenies as sequencing approaches improve can not be understated.

Next-generation sequencing can provide more robust estimations of hybridisation and introgression between groups (e.g., Holliday et al., 2016; Schilling et al., 2018; Wagner et al., 2018; Hodkinson et al., 2019). In the case of *Descurainia*, we find evidence for a hybridisation event and reveal gene flow patterns previously not found by Goodson et al. (2006). Phylogenetic analyses of Canarian flora commonly infer hybridisation based on the incongruence of a few loci. However, here, GBS can assess patterns of introgression or hybridisation on a genome-wide scale. As such, extensive sampling increases genomic resolution and power that a more limited number of loci cannot provide. Our results in chapter two provide further evidence that reduce-representation datasets (in our case, GBS) can resolve reticulate evolutionary events in rapidly diversifying lineages (Blanco-Pastor et al., 2019; Mort et al., 2022).

5.1.2 Whole-genome Sequencing

The second genomic dataset was generated using whole-genome sequencing, and in chapter three, we present a 180 Mb reference genome for *D. millefolia,* utilising long and short sequencing libraries with a hybrid assembly approach. The combination of long-read sequencing and the availability of genomic resources for its close relatives (i.e., *Arabidopsis thaliana*) resulted in a high-quality (N50 280 k) genome assembly. On average, long-read sequencing technology, PacBio, provides a ~32-fold increase in mean contig N50 for long-read assembly (Marks et al., 2021). The introduction of long sequencing has allowed for the rapid accumulation of plant genomes in recent years. Our reference genome demonstrates a fragmentation level, which is explained by our long-read sequences being unable to resolve large repetitive regions due to their short length. Nevertheless, this had a limited impact on the resolution of the gene content, and the genome shows 97.7% completeness. This enabled the use of the reference genome for further downstream analyses, specifically, the detection of putative genes under selection in chapter four.

To date, ca. 1,000 plant genomes have been published, representing 788 species (Li and Harkess, 2018). However, genome sequencing efforts are not representative of the diversity of plants. There are many biases in taxonomic representation and geographical regions, which mainly focus on model organisms and economically important crops. An estimated 404 genome assemblies from wild species, of which 77 represent wild relatives of crops (Marks et al., 2021). Wild species represent untapped genomic information. Our genome of Canary Island *Descurainia* represents a wild plant species with characteristics that would be advantageous for evolutionary studies. Such as an endemic to oceanic islands and distributions along an elevational cline. Particularly with rapid biodiversity loss and climate change, it is critical we take the opportunity to learn from wild species.

5.2 Phylogenetics of Canary Island *Descurainia*

Chapter two explored the use of alternative methods for generating phylogenetic hypotheses in *Descurainia*, the first being a bifurcating phylogeny (non-reticulated) and the second being a species network (reticulated), which allows for hybridisation events. Both resolved two major clades, with one (Clade A) comprised of low-elevation species and the other (Clade B) consisting solely of high-elevation species. The bifurcating phylogeny, generated using ca. 75 k SNPs, identified one phylogenetically significant ecological shift between these two clades and at least four dispersal events between islands. An upslope migration from low-elevation habitats to high-elevation habitats is proposed, with a strong phylogenetic signal in temperature and precipitation. The bifurcating phylogeny in chapter four, generated using whole-genome re-sequencing technology of concatenated CDS regions, was consistent with our GBS phylogeny in revealing two main clades. Minor conflicts occurred in the resolution of *D. lemsii* (ID no.: 11, chapter 2), which was previously nested within *D. bourgaeana* in chapter 2 but resolved as sister to *D. bourgaeana* in chapter 4. We hypothesise that the GBS sample of *D. lemsii* has been misidentified and was most likely a sample of *D. bourgaeana*. However, a more extensive sampling of *D. lemsii* would be needed to confirm this.

The species network exploring gene discordance, also presented in chapter 2, indicated a hybrid origin event for *D. gilva* with *D. millefolia* and *D. gonzalezii* as the parents. The contribution from the parents was heavily biased, with 99% of the *D. gilva* genome shared with *D. gonzalezii*. We

propose that the most likely scenario is that *D. gilva* is a vicariant of *D. gonzalezii,* which experienced introgression with *D. millefolia*. Additional hybridisation analyses, ABBA-BABA tests and a DIYAB model supported our findings. Our findings highlight the importance of techniques which allow for gene flow events between species.

It is important to incorporate hybridisation and introgression events into our understanding of evolutionary relationships since these are increasingly viewed as playing a pivotal role in the rapid radiation of lineages. For instance, the reassembly of old gene variation into new combinations is thought to facilitate adaptive radiation (Marques et al., 2019), and ancient hybridisation events have been documented in other radiations, such as cichlid fishes (Koblmüller et al., 2010; Meier et al., 2017; Ronco et al., 2021).

Utilising our reference genome for *D. millefolia* and other Brassicaceae genomes, we found *D. millefolia* was resolved within Lineage I and sister to *Cardamine* within *Brassicaceae,* which is consistent with previous phylogenetic studies on Brassicaceae (Beilstein et al., 2006, 2006; Nikolov et al., 2019; Liu et al., 2021). However, we included only high-quality proteome datasets in the analysis, and therefore taxonomic sampling was limited. Including additional species would provide a better indication of the relationships of *Descurainia* to other species in Brassicaceae.

Overall, chapters two and four revealed the importance of geographic isolation and ecological shifts in the diversification of Canary Island *Descurainia*. The need to use methods that incorporate and account for hybridisation in rapidly evolving island lineages cannot be understated.

5.3 Genomics of Ecological Adaptation

Recent research has highlighted the significance of ecological-driven speciation within the Canary Islands flora (e.g. Irl and Beierkuhnlein, 2011; Hua and Wiens, 2013; Steinbauer et al., 2013, 2016, 2017; Irl et al., 2015; Fernández-Palacios et al., 2021; Hanz et al., 2022). In particular, the sub-alpine habitats have been noted for their extreme ecological pressures, resulting in high levels of specialised adaptions and endemism. While the physiological mechanisms of high-elevation adaptation have been extensively studied, understanding the molecular underpinning of such

adaptation is only beginning to emerge. Here, building on the results of chapter 2 and the genomic resources from chapter 3, we explored the genetic underpinning of ecological adaptation in Canary Island *Descurainia* and, more widely, in Brassicaceae in Chapter 4. Our focus was on the high-elevation species of *Descurainia* (i.e., *D. bourgaeana, D. gilva, D. gonzalezii* and *D. lemsii*), which inhabit the sub-alpine zones of Tenerife and La Palma.

Chapter four demonstrated that, indeed, the high-elevation *Descurainia* experience significantly lower temperatures, an increase in UV radiation and precipitation, and high wind speed, compared to their lower-elevation counterparts. The results are consistent with Chapter 2, which showed that temperature is phylogenetically linked between the two major clades of Canary Island *Descurainia*. To detect candidate genes that may have incurred ecological adaptation, we used 14 genomes generated by re-sequencing technology across the seven species of Canary Island *Descurainia* to detect genes under selection on the branch leading to the high-elevation clade. A branch-site selection analysis employs *dN/dS* to identify genes that have evolved faster than expected. We found 276 genes under positive selection (0.98% of available genes) on the branch leading to our high-elevation *Descurainia* species, of which 21 were significantly enriched across 5 Biological Processes GO terms representing reproduction and abiotic stressors. We hypothesise a number of these genes provide an adaptive advantage to species distributed in high-elevation habitats.

For instance, at higher elevations and latitudes, the growing season is greatly reduced due to shorter day length and cold weather, and as such, alpine and sub-alpine flowers often exhibit altered flowering times. One such gene found to be under positive selection within our high-elevation clade was the FRI gene: FRI leads to an accumulation of FLOWERING LOCUS C (FLC) mRNA, inhibiting flowering unless down-regulated by vernalisation. FRI may be an adaptive gene to moderate flowering time to accommodate for the lower temperatures exhibited at high-elevation. Changes in reproduction appear to be frequent in similar genomic studies, such presented in the high-elevation *Crucihimalaya lasiocarpa* (Feng et al., 2022).

However, there is difficulty in making comparisons with other studies when results are dependent on methodology, as well as the number of samples and level of divergence between samples. One issue with our methodology is that incongruence between species trees and individual gene trees can lead to false evidence of positive selection (Mendes and Hahn, 2016). Chapter 2 showed that

Canary Island *Descurainia* have the propensity to hybridisation, and observations in the field support this. While our CODEML *dN/dS* analyses used gene trees and detected hybridisation events were predominantly restricted within the high-altitude clade (Chapter 2), the conflict between species and gene trees is a potential pitfall for our analysis. To date, few studies have provided methodologies to address this challenge, although Graham (2018) performed two CODEML analyses, one fixed to the species tree topology and another using gene trees. Those significant PSGs using its gene tree were excluded if they conflicted with the counterpart significant PSG, which used the fixed species tree topology.

Similarities between GO terms for high-elevation *Descurainia* and the seven other high-elevation and high-latitude Brassicaceae examined in Chapter 4 are apparent. For example, several studies found enriched GO terms related to tolerance to UV (Qiu et al., 2012; Guo et al., 2018). However, we found little evidence for gene convergence related to high-elevation abiotic stresses except for the phylogenetically divergent species pair high-elevation *Descurainia* clade (i.e., *D. bourgaeana, D. gilva, D. gonzalezii,* and *D. lemsii*) and the Arctic *Draba nivalis*. Overall, there was little convergence in the genetic underpinning of adaptation to high-elevation/high-latitude habitats.

Several studies have utilised whole-genome sequencing to investigate adaptation in Brassicaceae, commonly using *dN/dS* ratios or GWAS to identify genes under selection (e.g., Guo et al., 2018; Halbritter et al., 2018; Bohutínská et al., 2021; Nowak et al., 2021; Chiou et al., 2022; Feng et al., 2022). While these genomic scans can identify genes putatively under selection, there is still a need to confirm the function of these loci to provide further evidence that the genetic divergence is due to the adaptive section. The most thorough studies have mapped reproductive isolation to a candidate gene and then used experimental methods such as positional cloning, gene replacement/knockout, gene expression assays, and transgenic manipulations to confirm the genes involved (e.g., Bradshaw & Stettler, 1995; Rieseberg et al., 1996; Bradshaw et al., 1995; Juenger, Purugganan & Mackay, 2000). In chapter four, we tested three genes that have GO terms associated with water availability. Through greenhouse water availability experiments, we demonstrate that one of these genes significantly decreased the drought resilience in *Arabidopsis* knock-outs.

Ideally, all PSGs that are found to have significantly enriched GO terms would be tested. However, the identification of T-DNA knockout and the development of greenhouse conditions that would

effectively test fitness and function was beyond the scope of the present study. We hope our methodology here will promote the inclusion of functional validation experiments in similar studies.

Understanding adaptation at a molecular level has a broader significance beyond oceanic island diversification. Climate change continues to threaten food security in a world of rising crop demand. Changes in rainfall regimes and temperature can lead to crops experiencing abiotic stress, resulting in a reduction in crop yield. The identification of genes implicated in climate adaptation may improve the breeding of crops to be more resilient, particularly with developments in genome editing technologies (e.g., CRISPR/Cas9). For instance, in recent years, several genome-edited crop plants have been commercialisation in the USA, including drought and salt-tolerant soybean (Waltz, 2018). Genomic studies of adaptation in non-model organisms can only broaden our knowledge of genetic variants that underly adapt to diverse environments, particularly as Chapter 4 highlighted indicated there is little convergence (i.e., multiple genetics pathways) for adaptive phenotypes.

5.4 Future Work

Several outstanding questions regarding Canary Island *Descurainia* taxonomy have arisen throughout our phylogenetic analyses (chapters 2 and 4). Firstly, questions still surround the taxonomic status and relationships of *D. lemsii* (discussion found in Chapter 2), and further sampling of this species would be beneficial to confirm the placement of *D. lemsii* within our WGRS phylogeny (Chapter 4). Secondly, morphological and ecological differences between Tenerife, La Palma, and La Gomera populations of *D. millefolia* have been proposed (Pitard and Proust, 1908; Bramwell, 1977), and our WGRS data also suggest genetic divergence between these (chapter 4). Several taxa within *D. millefolia* have been described: *D. millefolia* f. *branchycarpa* Bornm. (Schulz, 1924) from Western La Palma, *D. millefolia* var. *sabinalis* O.E. Schulz (Schulz, 1924) from Tenerife, and *D. millefolia* var. *macricarpa* Pitard from La Gomera, Tenerife, and La Palma. Individuals with only 2-pinnatisect lower leaves have been described as *D. millefolia* var. *sabinalis*, but in cultivation, that character was not consistent. *Descurainia millefolia* var. *macrocarpa* and *D. millefolia* f. *branchycarpa* are challenging to delimit as they fall within the standard fruit size of most populations. *Descurainia millefolia* also exhibits different varying elevational ranges between islands. Tenerife and La Gomera populations are distributed between 150 – 700 m, whereas La Palma *D. millefolia* can reach 1,500 m. More extensive sampling of

Canary Island *Descurainia*, coupled with a detailed analysis of morphology, would allow a robust re-assessment of the infraspecific taxonomy of the multi-island endemic *D. millefolia.*

In chapter four, we hypothesized that ecology has played a role in the diversification of Canary Island *Descurainia* and focused on climatic variables related to high-elevation habitats that are restricted to Tenerife and La Palma in the Canary Islands. Weather station data, even at finer scales, is widely available and, therefore, accessible ecological datasets. However, many environmental variables could have potentially influenced the diversification of *Descurainia* and/or other Canary Island radiations that were not explored in this thesis. One developing field in adaptation research is the role of soil composition and soil biomes, ecological elements which have received less attention due to difficulties in sampling and quantification. The recognition of soil characteristics in promoting speciation is well known in *Howea* palms, an endemic plant to Lord Howe Island (Australia). Sympatric speciation in the *Howea* palms has been facilitated by arbuscular mycorrhizal fungi (Savolainen et al., 2006) and the colonisation of calcareous soil from volcanic soil (Osborne et al., 2019; Papadopulos et al., 2019a, 2019b). Recent studies on the petrographic traits of the volcanic bedrock are beginnings to be seen as a vital and influential force of plant distribution within the Canary Islands (Köhler et al., 2006; Irl and Beierkuhnlein, 2011; del Arco Aguilar and Rodríguez Delgado, 2018). For instance, there is a geological divide in plant destitution on La Palma with the young, poorly developed soils in the South and better-developed soils in the North, reflecting the distribution of endemic species richness of Canarian endemics (Reyes-Betancort et al., 2008; Irl and Beierkuhnlein, 2011). Ecological adaptation events could have gone undetected within Canary Island radiations due to a lack of sampling and datasets to relate back to phylogenetic and genomic studies. Environmental sampling beyond that of climate would allow for a more comprehensive understanding of the role of ecology in the diversification of Canary Island flora.

Further, similar genomic studies assessing molecular-level adaptation in Canary Island flora would provide further insights into adaptation patterns. Many Canary Island radiations have exhibited similar patterns to *Descurainia* in which they have colonised high-elevation habitats from closely related species distributed in the lower-elevations, for example, *Echium* and *Argyranthemum* (White et al., 2020; Graham et al., 2021). There are currently very limited studies examining similarities in adaptation between lineages, and comparisons between genes under positive selection within Canary Island lineages that have colonised similar habitats could potentially provide insights into convergent adaptation.

5.5 Conclusion

This thesis aimed to address knowledge gaps in the diversification and adaptation of Canary Island *Descurainia* using next-generation sequencing technologies and computational techniques. Canary Island *Descurainia* is also a suitable model for understanding the genomics of ecological adaptation due to the cost and computational ease of generating a reference genome compared to other groups in the Canary Islands' endemic flora. The first data chapter (Chapter 2) revised the phylogeny of Canary Island *Descurainia* using a reduced-representation dataset (GBS). Here, we generated a robust and well-resolved phylogeny, providing the primary framework for generating evolutionary hypotheses. We demonstrated that one hybridisation event, one ecological shift, and at least four geographical shifts have occurred in the evolutionary history of Canary Island *Descurainia*. The second data chapter (Chapter 3) generated a high-quality reference genome for the multi-island endemic *D. millefolia*. The last data chapter (Chapter 4) utilised the reference genome with an additional 14 Descurainia genomes to identify candidate genes hypothesized to infer adaptation to the high-elevation environment for *D. bourgaeana, D. gilva, D. gonzalezii* and *D. lemsii.*

In summary, this thesis is contributed to our understanding of the speciation processes that contribute to the rapid diversification of island radiations. We hope this thesis will aid further research on island radiations, plant speciation and the molecular-level understanding of adaptation in plants.

Appendix A Chapter 2

Appendix Table A.1: Twenty Canary Island *Descurainia* individuals from eight species and their origins were used in this study. Field-collected leaf material was dried in silica gel

ID	Species	Origin	DNA Source	Collector(s)	No. GBS raw reads
1	*D. artemisioides*	GC	Pre-extracted DNA	Goodson et al. 2006	1,601,123
2	*D. artemisioides*	GC	Pre-extracted DNA	Goodson et al. 2006	1,383,489
3	*D. bourgaeana*	TEN	Field collected silica leaf	R. Graham et al. †	722,778
4	*D. bourgaeana*	TEN	Greenhouse grown fresh leaf	BGV ‡	630,969
5	*D. bourgaeana*	TEN	Field collected silica leaf	R. Graham et al. †	1,732,499
6	*D. bourgaeana*	TEN	Field collected silica leaf	R. Graham et al. †	1,414,784
7	*D. bourgaeana*	TEN	Field collected silica leaf	R. Graham et al. †	1,754,449
8	*D. gilva*	LP	Field collected silica leaf	R. Graham et al. †	459,309
9	*D. gilva*	LP	Pre-extracted DNA	Goodson et al. 2006	1,639,245
10	*D. gonzalezii*	TEN	Greenhouse grown fresh leaf	BGV ‡	656,808
11	*D. gonzalezii*	TEN	Pre-extracted DNA	Goodson *et al.* 2006	1,551,613
12	*D. lemsii*	TEN	Pre-extracted DNA	Goodson *et al.* 2006	1,131,036
13	*D. millefolia*	TEN	Greenhouse grown fresh leaf	BGV ‡	1,611,156
14	*D. millefolia*	LP	Field collected silica leaf	R. Graham *et al.* †	1,774,120
15	*D. millefolia*	LP	Field collected silica leaf	R. Graham *et al.* †	2,144,194
16	*D. millefolia*	TEN	Field collected silica leaf	R. Graham *et al.* †	561,796
17	*D. preauxiana*	GC	Pre-extracted DNA	Goodson *et al.* 2006	1,526,147
18	*D. preauxiana*	GC	Greenhouse grown fresh leaf	BGV ‡	1,595,777
19	*D. tanacetifolia*	SP	Pre-extracted DNA	Goodson et al. 2006	1,420,720
20	*D. depressa*	SA	Pre-extracted DNA	Goodson et al. 2006	1,016,702

† Collected by R. Graham, O. White, M. Carine, and A. Beancourt in 2016; ‡ GBV = Banco de Germoplasma Vegetal Seedbank

Appendix Table A.2: Most likely Maximum-Likelihood model for all six GBS assemblies using ModelTest-NG

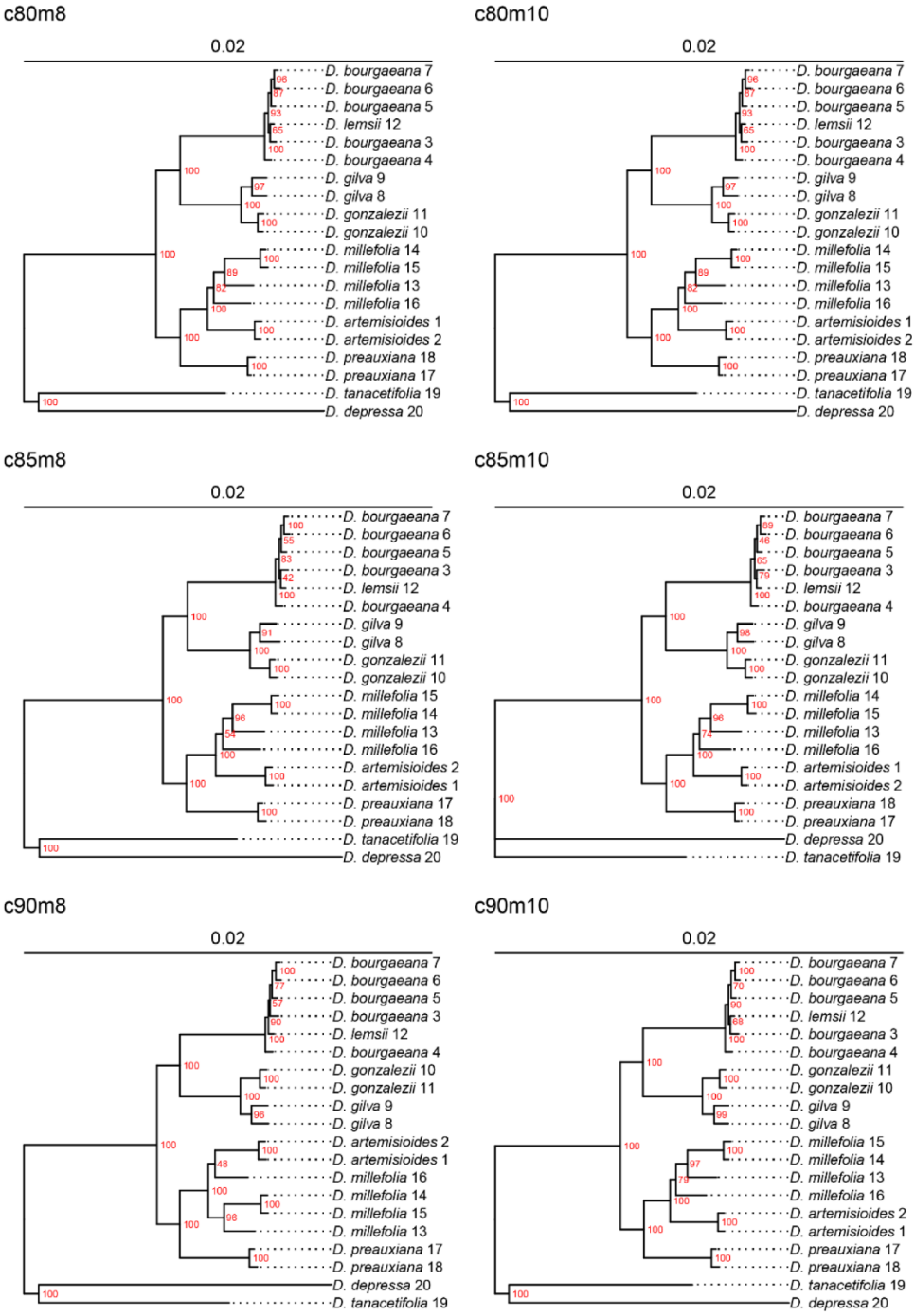
Assembly name	Clustering (%)	Minimum individuals	Model	Weight
c80m10	80	10	GTR+I+G	0.9958
c80m8	80	8	GTR+I+G	0.9879
c85m10	85	10	GTR+I+G	0.9783
c85m8	85	8	GTR+I+G	0.9859
c90m10	90	10	GTR+I+G	0.9977
c90m8	90	8	GTR+I+G	0.9943

Appendix Table A.3: The mean, median, mode and quantiles of posterior distribution samples for effective population sizes (parameter N), time (t) and admixture (r) for the tested scenario within DIYABC

Parameter	mean	median	mode	q025	q050	q250	q750	q950	q975
N1	5.49E+06	5.49E+06	5.41E+06	3.51E+06	3.85E+06	4.88E+06	6.04E+06	7.02E+06	7.81E+06
N2	1.48E+06	1.41E+06	1.34E+06	6.82E+05	8.00E+05	1.16E+06	1.69E+06	2.21E+06	2.58E+06
N3	2.09E+06	2.07E+06	2.00E+06	1.17E+06	1.38E+06	1.81E+06	2.35E+06	2.82E+06	3.03E+06
N4	4.05E+04	9.64E+03	3.93E+03	1.29E+03	1.93E+03	4.90E+03	2.22E+04	1.11E+05	2.28E+05
N5	3.60E+06	3.44E+06	3.25E+06	1.30E+06	1.61E+06	2.66E+06	4.32E+06	6.16E+06	7.40E+06
N6	3.82E+06	3.80E+06	3.74E+06	2.23E+06	2.56E+06	3.33E+06	4.26E+06	4.98E+06	5.39E+06
N7	1.51E+06	1.46E+06	1.42E+06	8.20E+05	9.28E+05	1.26E+06	1.69E+06	2.10E+06	2.37E+06
N8	9.93E+06	9.94E+06	9.95E+06	9.82E+06	9.86E+06	9.92E+06	9.96E+06	9.99E+06	9.99E+06
t7	1.35E+04	1.30E+04	1.27E+04	4.82E+03	6.43E+03	1.04E+04	1.60E+04	2.20E+04	2.54E+04
t6	3.74E+06	3.69E+06	3.67E+06	2.10E+06	2.37E+06	3.12E+06	4.28E+06	5.31E+06	5.72E+06
r	9.88E-01	9.90E-01	9.91E-01	9.73E-01	9.79E-01	9.87E-01	9.93E-01	9.96E-01	9.98E-01
t5	1.47E+07	1.48E+07	1.49E+07	1.21E+07	1.27E+07	1.40E+07	1.54E+07	1.64E+07	1.69E+07
t4	1.91E+07	1.92E+07	1.95E+07	1.64E+07	1.72E+07	1.87E+07	1.97E+07	2.05E+07	2.07E+07
t3	1.74E+07	1.75E+07	1.76E+07	1.46E+07	1.52E+07	1.68E+07	1.82E+07	1.93E+07	1.97E+07
t2	1.83E+07	1.87E+07	1.99E+07	1.30E+07	1.42E+07	1.73E+07	1.98E+07	2.08E+07	2.09E+07
t1	4.82E+07	4.84E+07	4.86E+07	4.56E+07	4.64E+07	4.79E+07	4.88E+07	4.95E+07	4.97E+07

Appendix Table A.4: Averages and standard deviation (in brackets) for bioclimatic variables for each 7 species across 147 individuals of Canary Island *Descurainia*. The bioclimatic variables are denoted as follows: AT = Average Temperature; TS = Temperature Seasonality; AP = Annual Precipitation; PWQ = Precipitation in the Wettest Quarter; PCQ = Precipitation in the Coldest Quarter. The Islands are noted as follows: GC = Gran Canaria; Tenerife = TEN; LA Palma = LP; La Gomera = LG.

Species	**Island**	**n**	**altitude**	**AT**	**TS**	**AP**	**PWQ**	**PCQ**
D. preauxiana	GC	34	802	16.224 (2.030)	356.787 (41.74)	317.941 (40.575)	156.441 (19.56)	156.441 (19.56)
D. millefolia	TEN; LP; LG	84	588	16.921 (1.87)	277.540 (37.614)	356.108 (52.99)	180.554 (24.63)	154.398 (27.02)
D. lemsii	TEN	14	1737	11.894 (0.99)	383.474 (34.427)	450.857 (9.882)	225.857 (8.511)	213.571 (18.58)
D. gonzalezii	TEN	16	1798	12.036 (1.038)	430.777 (36.55)	424.313 (5.016)	217.375 (5.852)	215.813 (10.703)
D. gilva	LP	22	1951	11.871 (1.923)	368.664 (32.603)	491.727 (38.801)	240.955 (16.752)	215.000 (19.265)
D. bourgaeana	TEN	24	2232	10.427 (0.883)	461.534 (39.904)	445.750 (32.671)	229.875 (11.207)	226.125 (10.027)
D. artemisoides	GC	23	582	17.088 (1.385)	311.273 (28.08)	316.045 (49.971)	154.591 (21.277)	153.091 (24.523)



Appendix Figure A.1: Six Maximum-Likelihood phylogenies using 18 individuals of Canary Island *Descurainia* and 2 outgroups from 6 assembly parameters using GBS data. Bootstrap values are next to branches in red. Branch length bar above phylogeny. Individual ID numbers are represented in Table 2.1.

Appendix B Chapter 3

Appendix Table B.1**:** Subread statistics for PacBio reads and Illumina reads (after filtering).

Library	**Subreads number**	**Average subreads length**	**N50**	**Coverage**
Illumina Reads	111,812,151	150	NA	83
Pac-Bio	515,738	7,067	1,118	18

Appendix Table B.2**:** The parameters for the highest-quality hybrid draft *do novo* assembly of *Descurainia millefolia* with the hybrid assembler DBG2OLC*.*

Software	**Parameter**	
Sparse Assembler	NodeCovTh	4
	EdgeCovTh	1
	k	119
	g	15
DBG2OLC	k	17
	KmerCovTh	2
	MinOverlap	20
	AdaptiveTh	0.001

Appendix Table B.3: Summary of genome statistics after gap closing.

Genome statistics	**Original**	**After gap closing**
Total number of scaffolds	2,274	2,274
Sum (bp)	179,681,242	1.8E+08
Total number of N's	2,889,089	2,709,239
Sum (bp) no N's	176,792,153	1.77E+08
GC Content	35.81%	35.81%
Max scaffold size	2,172,884	2,172,884
Min scaffold size	1,662	1,662
Average scaffold size	79,015	79,017
N25	637,255	637,255
N50	289,995	289,995
N75	109,019	109,019

Appendix B

Appendix Table B.4: Summary table generated by RepeatMasker for the draft assembly of *D. millefolia*

Repeat element	# elements	bps occupied	percentage of genome
Retroelements	20,203	13,627,141	7.81%
SINEs	933	143,935	0.08%
Penelope	119	11,012	0.01%
LINEs	3,034	1,475,491	0.85%
CRE/SLACS	1	81	0.00%
L2/CR1/Rex	0	0	0.00%
R1/LOA/Jockey	0	0	0.00%
R2/R4/NeSL	0	0	0.00%
RTE/Bov-B	4	516	0.00%
L1/CIN4	2,797	1,446,718	0.83%
LTR elements	16,236	12,007,715	6.88%
BEL/Pao	0	0	0.00%
Ty1/Copia	6,425	4,495,135	2.57%
Gypsy/DIRS1	8,954	6,592,006	3.78%
Retroviral	0	0	0.00%
DNA transposons	18,270	5,317,792	3.05%
hobo-Activator	5,536	1,403,830	0.80%
Tc1-IS630-Pogo	2,212	570,976	0.33%
En-Spm	0	0	0.00%
MuDR-IS905	0	0	0.00%
PiggyBac	0	0	0.00%
Tourist/Harbinger	1,134	307,550	0.18%
Other (Mirage, P-element, Transib)	2	138	0.00%
Rolling-circles	2,980	702,441	0.40%
Unclassified:	380	60,884	0.03%

Repeat element	# elements	bps occupied	percentage of genome
Total interspersed repeats		19,005,817	10.89%
Small RNA:	1,580	418,783	0.24%
Satellites:	709	113,140	0.06%
Simple repeats	54,905	2,373,384	1.36%
Low complexity	14,392	712,194	0.41%

Appendix Table B.5**:** BUSCO summary statistics for our assembled reference genome of *Descurainia millefolia*.

	BDG2OLC	**MaSuRCA**
BUSCO summary statistic	**Number of genes (%)**	
Complete BUSCOs	411 (96.7)	1,406 (97.7)
Complete and single-copy BUSCOs	394 (92.7)	1,330 (92.4)
Complete and duplicated BUSCOs	17 (4.0)	76 (5.3)
Fragmented BUSCOs	5 (1.2)	9 (0.6)
Missing BUSCOs	9 (2.1)	25 (1.7)
Total BUSCO groups searched	425	1,440

Appendix B

Appendix Table B.6: Summary of taxa and genome datasets for comparison genomics analyses, obtained from The Brassicaceae Database (BRAD v.3.0; http://brassicadb.cn; Chen et al., 2022).

ID	Taxon	Reference	Assembly size (Mb)	Scaffold N50 (Mb)	Protein-coding genes
1	*Aethionema arabicum*	(Fernandez-Pozo et al., 2021)	192.49	1.24	18,312
2	*Arabidopsis halleri*	(Briskine et al., 2017)	196	0.71	25,328
3	*Arabidopsis lyrata*	(Hu et al., 2011)	206.67	24.46	32,667
4	*Arabidopsis thaliana*	(Cheng et al.	120	33.10	27,379
5	*Boechera retrofracra*	(Kliver et al., 2018)	227	2.30	27,048
6	*Boechera stricta*†	(Chen et al., 2022)	N/A	N/A	N/A
7	*Camelina sativa*	(Kagale et al., 2014)	547.65	0.99	15,937
8	*Capsella grandiflora*†	(Chen et al., 2022)	N/A	N/A	N/A
9	*Capsella rubella*	(Slotte et al., 2013)	133.06	15.04	26,521
10	*Cardamine hirsuta*	(Gan et al., 2016)	198	N/A	29,458
11	*Descurainia sophiodies* †	(Chen et al., 2022)	124	1.8	28,032
12	*Leavenworthia alabamica*	(Haudry et al., 2013)	243	0.50	30,343

† Genome is unpublished

Appendix Table B.7: Time calibration used for MCMCtree (PAML), obtained from TIMETREE5 (Kumar et al., 2022)

Taxon 1	**Taxon 2**	**Time Range (MYA)**
Capsella	*Camelina*	1.1 - 9.3
Camelina	*Boechera*	10.0 - 21.2
Arabidopsis	*Boechera*	10 – 21.2
Descurainia	*Arabidopsis*	20 – 28.7
Descurainia	*Cardamine*	20 – 28.7
Descurainia	*Aethionema*	32.4-45.0

Appendix Figure B.1: Heatmap of orthologues comparisons between species pairs. Values represent the number of orthologs/1000.

Appendix C Chapter 4

Appendix Table C.1: Summary of the primers used for the T-DNA insertions for homozygous confirmation. The LBb1.3 primer was used as the left border primer for all three mutants.

Gene	Study name	chr	Name	Product Size	LP	RP
AT5G52300	A	chr5	SALK_052958	1099	AACAACACACAGTGCATTTGC	GAGCCTCTCTGCTCTCTCCTC
At1g20450	B	chr1	SALK_064253	1102	TGATGAAGAAGGTGAAGACGG	TTGAAAAGTCCATTGCCTAGTTC
AT3G08730	C	chr3	SALK_112945	1105	TCTGTATCGATTTTTCCCGTG	TTTCAGGTGCCATATACTCCG

Appendix Table C.2: Summary of the processing, filtering, and mapping of whole-genome sequencing reads.

ID	Species	Island	Trimmed Paired	% retained	Mapped reads	Breathe of Coverage	Mean Read depth	% reads mapped	% PCR duplication
1	*D. artemisioides*	GC	36,474,166	95.24	95.24	99.75	30.22	95.772	0.15
2	*D. bourgaeana*	TEN	40,218,230	73.81	73.81	99.57	37.13	99.493	0.17
3	*D. bourgaeana*	TEN	33,657,110	94.90	94.90	99.54	31.30	99.599	0.16
4	*D. gilva*	LP	38,200,740	93.39	93.39	99.60	33.80	99.732	0.20
5	*D. gilva*	LP	37,049,276	95.06	95.06	99.58	34.67	99.581	0.16
6	*D. gonzalezii*	TEN	33,477,284	94.91	94.91	99.46	30.99	99.790	0.17
7	*D. gonzalezii*	TEN	34,310,710	94.77	94.77	99.57	32.01	99.734	0.17
8	*D. lemsii*	TEN	36,050,644	94.52	94.52	99.63	32.28	99.723	0.17
9	*D. lemsii*	TEN	33,792,814	95.58	95.58	99.60	29.92	95.163	0.16
10	*D. millefolia*	TEN	34,825,938	94.81	94.81	99.83	31.70	99.839	0.18
11	*D. millefolia*	LG	37,778,182	95.07	95.07	99.78	33.90	99.064	0.16
12	*D. millefolia*	LP	34,468,666	95.79	95.79	99.77	30.98	98.915	0.16
13	*D. preauxiana*	GC	37,231,246	94.76	94.76	99.67	30.98	99.808	0.19

Appendix C

ID	Species	Island	Trimmed Paired	% retained	Mapped reads	Breathe of Coverage	Mean Read depth	% reads mapped	% PCR duplication
14	*D. preauxiana*	GC	33,387,142	95.09	95.09	99.65	30.57	99.758	0.16
15	*D. tanacetifolia*	SP	49,520,076	96.29	96.29	99.23	31.48	71.185	0.16

Appendix Table C.3: F_{st} statistics calculated from 15 whole-genome sequencing SNP dataset.

	D. bourgaeana	***D. millefolia***	***D. preauxiana***	***D. artemisioides***	***D. gilva***	***D. gonzalezi***
D. millefolia	0.180915					
D. preauxiana	0.231397	0.133029				
D. artemisioides	0.225494	0.073515	0.113833			
D. gilva	0.16554	0.138428	0.206055	0.196588		
D. gonzalezi	0.165394	0.13827	0.202379	0.193684	0.004778	
D. lemsii	0.109212	0.053	0.121264	0.08311	0.031074	0.023994

Appendix Table C.4: ANOVA comparing the average values between species in Clade A (low-altitude) and B (high-altitude) for five of our climatic variables.

climate	season	estimate	conf.low	conf.high	p.adj
average temperature	autumn	5.455	3.285	7.626	**0.000**
average temperature	spring	5.467	4.269	6.664	**0.000**
average temperature	summer	3.298	1.728	4.869	**0.000**
average temperature	winter	5.808	4.841	6.774	**0.000**
precipitation	autumn	-9.827	-24.663	5.010	0.182
precipitation	spring	-9.581	-23.850	4.688	0.176
precipitation	summer	-0.819	-1.664	0.026	0.057
precipitation	winter	-21.659	-28.374	-14.944	**0.000**
solar radiation	autumn	-862.838	-4,211.114	2,485.437	0.596
solar radiation	spring	-1,503.753	-3,840.182	832.676	0.194
solar radiation	summer	-1,989.654	-3,043.800	-935.509	**0.001**
solar radiation	winter	-739.732	-2,430.531	951.067	0.371
water vapor pressure	autumn	0.578	0.483	0.672	**0.000**
water vapor pressure	spring	0.462	0.390	0.535	**0.000**
water vapor pressure	summer	0.615	0.523	0.706	**0.000**
water vapor pressure	winter	0.430	0.346	0.514	**0.000**
wind speed	autumn	-1.553	-2.044	-1.063	**0.000**
wind speed	spring	-1.508	-2.037	-0.979	**0.000**
wind speed	summer	-0.859	-1.485	-0.233	**0.010**
wind speed	winter	-2.510	-3.031	-1.989	**0.000**

Appendix Table C.5: List of genes under positive selection on the branch leading to the high-altitude species of Canary Island *Descurainia*. * indicate P value significance level. There are 53 *Descurainia* (GMOD) genes that have no annotation.

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00000369	0.029	At4g29180 F19B15.210	membrane [GO:0016020]; ATP binding [GO:0005524]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; protein phosphorylation [GO:0006468]
GMOD_00000546	0.004	HRQ1 YDR291W	nucleus [GO:0005634]; 3'-5' DNA helicase activity [GO:0043138]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; DNA binding [GO:0003677]; DNA duplex unwinding [GO:0032508]; nucleotide-excision repair [GO:0006289]
GMOD_00000593	0.009	SQS1 YALI0F11253g	cytoplasm [GO:0005737]; nucleus [GO:0005634]; nucleic acid binding [GO:0003676]; mRNA processing [GO:0006397]; RNA splicing [GO:0008380]
GMOD_00000683	0.043	RED1 MGG_12983	oxidoreductase activity [GO:0016491]
GMOD_00000866	0.007	DAR3 At5g66640 MSN2.2	ubiquitin binding [GO:0043130]
GMOD_00002100	0.014	AUL1 At1g75310 F1B16.14 F22H5.15	cytoplasm [GO:0005737]; intracellular membrane-bounded organelle [GO:0043231]; vesicle [GO:0031982]; clathrin binding [GO:0030276]; clathrin coat disassembly [GO:0072318]; clathrin-dependent endocytosis [GO:0072583]
GMOD_00002133	0.004	At1g75040 F9E10.11	apoplast [GO:0048046]; mitochondrion [GO:0005739]; plant-type vacuole [GO:0000325]; secretory vesicle [GO:0099503]; mRNA binding [GO:0003729]; defense response [GO:0006952]; regulation of anthocyanin biosynthetic process [GO:0031540]; response to cadmium ion [GO:0046686]; response to UV-B [GO:0010224]; response to virus [GO:0009615]; systemic acquired resistance [GO:0009627]
GMOD_00002270	0.017	At3g61800 F15G16.190	chromosome [GO:0005694]; RNA polymerase II complex binding [GO:0000993]; response to UV [GO:0009411]; transcription-coupled nucleotide-excision repair [GO:0006283]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00002597	0.015	AHL17 At5g49700 K2I5.6	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; minor groove of adenine-thymine-rich DNA binding [GO:0003680]
GMOD_00002771	0.005		
GMOD_00003076	0.036	ARP REF At2g41460 T26J13.5	chloroplast nucleoid [GO:0042644]; nucleus [GO:0005634]; 3'-5' exonuclease activity [GO:0008408]; DNA binding [GO:0003677]; DNA-(apurinic or apyrimidinic site) endonuclease activity [GO:0003906]; double-stranded DNA 3'-5' exodeoxyribonuclease activity [GO:0008311]; endonuclease activity [GO:0004519]; metal ion binding [GO:0046872]; phosphatase activity [GO:0016791]; phosphodiesterase I activity [GO:0004528]; phosphoric diester hydrolase activity [GO:0008081]; base-excision repair [GO:0006284]; nucleotide-excision repair, DNA incision [GO:0033683]; positive regulation of DNA-templated transcription [GO:0045893]
GMOD_00003120	0.008	R3HDM2	nucleus [GO:0005634]; nucleic acid binding [GO:0003676]
GMOD_00003203	0.021	ESF1.1 MEG1.1 At1g10747 F20B24 T16B5	embryonic axis specification [GO:0000578]; suspensor development [GO:0010098]
GMOD_00003489	0.001		
GMOD_00003599	0.000	SULTR1;2 At1g78000 F28K19.22	plasma membrane [GO:0005886]; secondary active sulfate transmembrane transporter activity [GO:0008271]; symporter activity [GO:0015293]; cellular response to sulfate starvation [GO:0009970]
GMOD_00003600	0.013	mnmE trmE Synpcc7942_1582 sed0016	cytoplasm [GO:0005737]; GTP binding [GO:0005525]; GTPase activity [GO:0003924]; metal ion binding [GO:0046872]; tRNA modification [GO:0006400]
GMOD_00003627	0.024		
GMOD_00003771	0.011		
GMOD_00004706	0.044	AGL27 FK1 FLM MAF1 At1g77080 F22K20.15	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; DNA-binding transcription factor activity, RNA polymerase II-specific [GO:0000981]; protein dimerization activity [GO:0046983]; RNA polymerase II cis-regulatory region sequence-specific DNA binding [GO:0000978]; floral meristem determinacy [GO:0010582]; negative regulation of flower development [GO:0009910]; photoperiodism, flowering [GO:0048573]; regulation of

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			DNA-templated transcription [GO:0006355]; regulation of flower development [GO:0009909]; regulation of transcription by RNA polymerase II [GO:0006357]
GMOD_00004709	0.041	AGL31 MAF2 At5g65050 MXK3.30	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; DNA-binding transcription factor activity, RNA polymerase II-specific [GO:0000981]; protein dimerization activity [GO:0046983]; RNA polymerase II cis-regulatory region sequence-specific DNA binding [GO:0000978]; cell differentiation [GO:0030154]; floral meristem determinacy [GO:0010582]; negative regulation of flower development [GO:0009910]; negative regulation of vernalization response [GO:0010221]; positive regulation of transcription by RNA polymerase II [GO:0045944]; regulation of DNA-templated transcription [GO:0006355]; regulation of transcription by RNA polymerase II [GO:0006357]; vernalization response [GO:0010048]
GMOD_00004897	0.049	At5g63170 MDC12.14	extracellular region [GO:0005576]; lipase activity [GO:0016298]; lipid catabolic process [GO:0016042]
GMOD_00004966	0.003	PER28 P28 At3g03670 T12J13.5	extracellular region [GO:0005576]; plant-type cell wall [GO:0009505]; heme binding [GO:0020037]; lactoperoxidase activity [GO:0140825]; metal ion binding [GO:0046872]; peroxidase activity [GO:0004601]; cellular response to hypoxia [GO:0071456]; hydrogen peroxide catabolic process [GO:0042744]; response to oxidative stress [GO:0006979]
GMOD_00004988	0.001	CHAT At3g03480 T21P5.10	(Z)-3-hexen-1-ol acetyltransferase activity [GO:0102165]; acetyl CoA:(Z)-3-hexen-1-ol acetyltransferase activity [GO:0010327]; acyltransferase activity, transferring groups other than amino-acyl groups [GO:0016747]; green leaf volatile biosynthetic process [GO:0010597]
GMOD_00005015	0.004	GOT1 At3g03180 T17B22.13	cytosol [GO:0005829]; Golgi membrane [GO:0000139]; endoplasmic reticulum to Golgi vesicle-mediated transport [GO:0006888]; retrograde transport, endosome to Golgi [GO:0042147]
GMOD_00005161	0.000	IAN10 PP2A3 At2g26820 F12C20.14	carbohydrate binding [GO:0030246]; GTP binding [GO:0005525]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00005188	0.013	ARF12 At1g34310 F23M19.4	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; transcription cis-regulatory region binding [GO:0000976]; auxin-activated signaling pathway [GO:0009734]
GMOD_00005303	0.000		
GMOD_00005619	0.024	nsa2 DDB_G0290031	nucleolus [GO:0005730]; preribosome, large subunit precursor [GO:0030687]; maturation of 5.8S rRNA [GO:0000460]; maturation of LSU-rRNA [GO:0000470]
GMOD_00005687	0.002	MOT2 ST5.1 SULTR5.1 At1g80310 F5I6.6	mitochondrion [GO:0005739]; plant-type vacuole [GO:0000325]; plant-type vacuole membrane [GO:0009705]; molybdate ion transmembrane transporter activity [GO:0015098]; molybdate ion export from vacuole [GO:0090414]; molybdate ion transport [GO:0015689]
GMOD_00005729	0.009	RBK2 At3g05140 T12H1.10	cytoplasm [GO:0005737]; nucleus [GO:0005634]; ATP binding [GO:0005524]; GTPase binding [GO:0051020]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; protein phosphorylation [GO:0006468]
GMOD_00005760	0.000	ADG2 APL1 At5g19220 T24G5.120	chloroplast [GO:0009507]; chloroplast envelope [GO:0009941]; chloroplast stroma [GO:0009570]; ATP binding [GO:0005524]; glucose-1-phosphate adenylyltransferase activity [GO:0008878]; glycogen biosynthetic process [GO:0005978]; starch biosynthetic process [GO:0019252]
GMOD_00005915	0.015	At4g14190 dl3135c FCAALL.115	chloroplast [GO:0009507]
GMOD_00005927	0.000	RLP47 At4g13810 F18A5.200	plasma membrane [GO:0005886]
GMOD_00006002	0.006	At1g26930 T2P11.12	nucleus [GO:0005634]; peroxisome [GO:0005777]
GMOD_00006019	0.049		
GMOD_00006145	0.026	At1g25211 F4F7.44	
GMOD_00006158	0.026	CYTC-1 At1g22840 F19G10.20	cytosol [GO:0005829]; mitochondrial intermembrane space [GO:0005758]; mitochondrion [GO:0005739]; nucleus [GO:0005634]; plant-type vacuole [GO:0000325];

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			respirasome [GO:0070469]; copper ion binding [GO:0005507]; electron transfer activity [GO:0009055]; heme binding [GO:0020037]; gibberellic acid homeostasis [GO:0010336]; mitochondrial electron transport, cytochrome c to oxygen [GO:0006123]; mitochondrial electron transport, ubiquinol to cytochrome c [GO:0006122]
GMOD_00006397	0.000	At2g15640 F9O13.19	
GMOD_00006419	0.002	KCS9 At2g16280 F16F14.22	endoplasmic reticulum membrane [GO:0005789]; very-long-chain 3-ketoacyl-CoA synthase activity [GO:0102756]; fatty acid biosynthetic process [GO:0006633]; response to cold [GO:0009409]; response to light stimulus [GO:0009416]
GMOD_00006443	0.023		RNA binding [GO:0003723]; translation initiation factor activity [GO:0003743]
GMOD_00006463	0.004	ASK11 At4g34210 F10M10.2	cytoplasm [GO:0005737]; nucleus [GO:0005634]; SCF ubiquitin ligase complex [GO:0019005]; cullin family protein binding [GO:0097602]; auxin-activated signaling pathway [GO:0009734]; jasmonic acid mediated signaling pathway [GO:0009867]; protein ubiquitination [GO:0016567]; response to auxin [GO:0009733]; response to jasmonic acid [GO:0009753]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
GMOD_00006505	0.012	RPS6 At5g46470 K11I1.6	ADP binding [GO:0043531]; ATP binding [GO:0005524]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleotidase, cyclic ADP-ribose generating [GO:0061809]; defense response to bacterium [GO:0042742]; signal transduction [GO:0007165]
GMOD_00006642	0.014	fabG ylpF BSU15910	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity [GO:0004316]; NAD binding [GO:0051287]; oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor [GO:0016616]; fatty acid elongation [GO:0030497]
GMOD_00006916	0.018	INT3 At2g35740 T20F21.7	membrane [GO:0016020]; myo-inositol:proton symporter activity [GO:0005366]; transmembrane transporter activity [GO:0022857]; myo-inositol transport [GO:0015798]; transmembrane transport [GO:0055085]
GMOD_00006969	0.008	RUN1	cytoplasm [GO:0005737]; nucleus [GO:0005634]; ADP binding [GO:0043531]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleosidase activity [GO:0003953];

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			NAD+ nucleotidase, cyclic ADP-ribose generating [GO:0061809]; defense response to fungus [GO:0050832]; NAD catabolic process [GO:0019677]; positive regulation of programmed cell death [GO:0043068]; signal transduction [GO:0007165]
GMOD_00007066	0.021	SRG1 At1g17020 F20D23.28 F6I1.30	metal ion binding [GO:0046872]; oxidoreductase activity, acting on diphenols and related substances as donors, oxygen as acceptor [GO:0016682]; leaf senescence [GO:0010150]
GMOD_00007099	0.000	Ermp1 Fxna	endoplasmic reticulum [GO:0005783]; endoplasmic reticulum membrane [GO:0005789]; metal ion binding [GO:0046872]; metalloexopeptidase activity [GO:0008235]; cellular response to oxidative stress [GO:0034599]; endoplasmic reticulum unfolded protein response [GO:0030968]; ovarian follicle development [GO:0001541]; proteolysis [GO:0006508]
GMOD_00007111	0.033	HSP70-2 HSC70-2 HSC70-G8 MED37_3 MED37D At5g02490 T22P11.80	cytoplasm [GO:0005737]; cytosol [GO:0005829]; Golgi apparatus [GO:0005794]; nucleus [GO:0005634]; plant-type cell wall [GO:0009505]; plasma membrane [GO:0005886]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; ATP-dependent protein folding chaperone [GO:0140662]; heat shock protein binding [GO:0031072]; misfolded protein binding [GO:0051787]; protein folding chaperone [GO:0044183]; unfolded protein binding [GO:0051082]; cellular response to unfolded protein [GO:0034620]; chaperone cofactor-dependent protein refolding [GO:0051085]; protein refolding [GO:0042026]; response to bacterium [GO:0009617]; response to heat [GO:0009408]; response to virus [GO:0009615]
GMOD_00007120	0.002	POLA2	alpha DNA polymerase:primase complex [GO:0005658]; cytosol [GO:0005829]; nucleoplasm [GO:0005654]; DNA binding [GO:0003677]; DNA replication [GO:0006260]; DNA replication initiation [GO:0006270]; DNA replication, synthesis of RNA primer [GO:0006269]; protein import into nucleus [GO:0006606]
GMOD_00007172	0.031	TDR PXY At5g61480 MCI2.4	plasma membrane [GO:0005886]; ATP binding [GO:0005524]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; cell division [GO:0051301]; phloem or xylem histogenesis [GO:0010087]; procambium histogenesis [GO:0010067];

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			protein phosphorylation [GO:0006468]; secondary shoot formation [GO:0010223]; xylem development [GO:0010089]
GMOD_00007257	0.011	CDC48C EMB1354 At3g01610 F4P13.15	cytoskeleton [GO:0005856]; nucleus [GO:0005634]; phragmoplast [GO:0009524]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; preribosome binding [GO:1990275]; cell cycle [GO:0007049]; cell division [GO:0051301]; positive regulation of telomerase activity [GO:0051973]; protein transport [GO:0015031]; ribosome biogenesis [GO:0042254]
GMOD_00007296	0.015	SEC11C SEC11L3	signal peptidase complex [GO:0005787]; serine-type endopeptidase activity [GO:0004252]; signal peptide processing [GO:0006465]
GMOD_00007418	0.007	APRR3 At5g60100 MGO3.8	nucleus [GO:0005634]; cytokinin-activated signaling pathway [GO:0009736]; phosphorelay signal transduction system [GO:0000160]; rhythmic process [GO:0048511]
GMOD_00007528	0.042	sde2 zgc:112095	nucleus [GO:0005634]; DNA binding [GO:0003677]; cell cycle [GO:0007049]; cell division [GO:0051301]; DNA replication [GO:0006260]
GMOD_00007549	0.044	ATPK1 ATPK6 S6K1 At3g08730 F17O14.20	cytoplasm [GO:0005737]; nucleus [GO:0005634]; ATP binding [GO:0005524]; kinase activity [GO:0016301]; protein kinase activity [GO:0004672]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; negative regulation of cell population proliferation [GO:0008285]; peptidyl-serine phosphorylation [GO:0018105]; positive regulation of translation [GO:0045727]; protein phosphorylation [GO:0006468]; response to cold [GO:0009409]; response to osmotic stress [GO:0006970]; response to salt stress [GO:0009651]
GMOD_00007625	0.032	CALS3 GSL12 At5g13000 T24H18.170	1,3-beta-D-glucan synthase complex [GO:0000148]; plasma membrane [GO:0005886]; 1,3-beta-D-glucan synthase activity [GO:0003843]; glucosyltransferase activity [GO:0046527]; (1->3)-beta-D-glucan biosynthetic process [GO:0006075]; cell wall organization [GO:0071555]; regulation of cell shape [GO:0008360]
GMOD_00008007	0.006	At5g53635 MNC6.17	
GMOD_00008054	0.012	prmA Bxeno_A3865 Bxe_A0530	cytoplasm [GO:0005737]; protein methyltransferase activity [GO:0008276]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00008203	0.031	FER AAK1 SIR SRN At3g51550 F26O13.190	filiform apparatus [GO:0043680]; plasma membrane [GO:0005886]; plasmodesma [GO:0009506]; ATP binding [GO:0005524]; protein kinase activity [GO:0004672]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; transmembrane receptor protein tyrosine kinase activity [GO:0004714]; abscisic acid-activated signaling pathway [GO:0009738]; brassinosteroid mediated signaling pathway [GO:0009742]; circadian regulation of gene expression [GO:0032922]; defense response to fungus [GO:0050832]; ethylene-activated signaling pathway [GO:0009873]; negative regulation of abscisic acid-activated signaling pathway [GO:0009788]; negative regulation of cell growth [GO:0030308]; pollen tube reception [GO:0010483]; post-embryonic development [GO:0009791]; protein autophosphorylation [GO:0046777]; response to brassinosteroid [GO:0009741]; response to ethylene [GO:0009723]; root development [GO:0048364]; stomatal movement [GO:0010118]
GMOD_00008695	0.035	RF178 At2g38920 T7F6.9	metal ion binding [GO:0046872]; transferase activity [GO:0016740]; protein ubiquitination [GO:0016567]
GMOD_00008711	0.010	EMB2204 At1g22090 F2E2.16	
GMOD_00008906	0.008	ERCC1 RAD10 UVR7 At3g05210 T12H1.18	ERCC4-ERCC1 complex [GO:0070522]; nucleotide-excision repair factor 1 complex [GO:0000110]; 5'-flap endonuclease activity [GO:0017108]; damaged DNA binding [GO:0003684]; single-stranded DNA binding [GO:0003697]; double-strand break repair via homologous recombination [GO:0000724]; meiotic mismatch repair [GO:0000710]; mitotic recombination [GO:0006312]; non-photoreactive DNA repair [GO:0010213]; nucleotide-excision repair, DNA incision, 5'-to lesion [GO:0006296]; nucleotide-excision repair, preincision complex assembly [GO:0006294]; response to gamma radiation [GO:0010332]; response to UV-B [GO:0010224]; UV-damage excision repair [GO:0070914]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00008932	0.006	AMT1-5 At3g24290 K7M2.6	plasma membrane [GO:0005886]; ammonium transmembrane transporter activity [GO:0008519]; ammonium homeostasis [GO:0097272]; ammonium transmembrane transport [GO:0072488]
GMOD_00008972	0.024		
GMOD_00009175	0.020	PIGM PNT1 At5g22130 T6G21.34	endoplasmic reticulum membrane [GO:0005789]; glycosylphosphatidylinositol-mannosyltransferase I complex [GO:1990529]; alpha-1,4-mannosyltransferase activity [GO:0051751]; glycolipid mannosyltransferase activity [GO:0004376]; mannosyltransferase activity [GO:0000030]; GPI anchor biosynthetic process [GO:0006506]
GMOD_00009671	0.004		
GMOD_00009697	0.025	At1g64065	membrane [GO:0016020]
GMOD_00009727	0.007	RLM1A At1g64070 F22C12.17	ADP binding [GO:0043531]; ATP binding [GO:0005524]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleotidase, cyclic ADP-ribose generating [GO:0061809]; defense response to fungus [GO:0050832]; signal transduction [GO:0007165]
GMOD_00009757	0.000		
GMOD_00010067	0.041		
GMOD_00010176	0.001	At2g01390/At2 g01380 F10A8.29 F2I9.1	
GMOD_00010193	0.002		
GMOD_00010209	0.001	UTP20 YBL004W YBL0101	90S preribosome [GO:0030686]; cytoplasm [GO:0005737]; nucleolus [GO:0005730]; nucleoplasm [GO:0005654]; preribosome, small subunit precursor [GO:0030688]; small-subunit processome [GO:0032040]; mRNA binding [GO:0003729]; endonucleolytic cleavage in 5'-ETS of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000480]; endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000447]; endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000472]; maturation of SSU-rRNA [GO:0030490]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00010255	0.005	EMB175 PCMP-H19 At5g03800 F8F6_10 MED24_10	zinc ion binding [GO:0008270]; embryo development ending in seed dormancy [GO:0009793]
GMOD_00010324	0.012	PICBP At5g04020 F8F6_230	calmodulin binding [GO:0005516]; defense response [GO:0006952]
GMOD_00010409	0.001	FH8 At1g70140 F20P5.14	plasma membrane [GO:0005886]; plasmodesma [GO:0009506]; actin binding [GO:0003779]; actin filament binding [GO:0051015]; profilin binding [GO:0005522]; actin nucleation [GO:0045010]; cell tip growth [GO:0009932]
GMOD_00010434	0.011	NPF4.6 AIT1 NRT1.2 NTL1 NTR1:2 At1g69850 T17F3.12	plasma membrane [GO:0005886]; abscisic acid transmembrane transporter activity [GO:0090440]; symporter activity [GO:0015293]; abscisic acid transport [GO:0080168]; nitrate assimilation [GO:0042128]; regulation of stomatal movement [GO:0010119]; response to nematode [GO:0009624]
GMOD_00010437	0.024	WRKY36 At1g69810 T17F3.16	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; sequence-specific DNA binding [GO:0043565]
GMOD_00010461	0.047		
GMOD_00010491	0.038	EXO70H1 At3g55150 T26I12.30	cytosol [GO:0005829]; endomembrane system [GO:0012505]; exocyst [GO:0000145]; membrane [GO:0016020]; nucleus [GO:0005634]; phagocytic vesicle [GO:0045335]; vesicle [GO:0031982]; phosphatidylinositol-4,5-bisphosphate binding [GO:0005546]; defense response [GO:0006952]; exocytosis [GO:0006887]; positive regulation of defense response to bacterium [GO:1900426]; response to bacterium [GO:0009617]; response to fungus [GO:0009620]; response to molecule of bacterial origin [GO:0002237]
GMOD_00010511	0.040	GSTT3 GST10C At5g41220 K1O13.1	cytoplasm [GO:0005737]; nucleus [GO:0005634]; plastid [GO:0009536]; glutathione transferase activity [GO:0004364]; toxin catabolic process [GO:0009407]
GMOD_00010521	0.031	WDR11 BRWD2 KIAA1351 WDR15	axoneme [GO:0005930]; ciliary basal body [GO:0036064]; cytoplasm [GO:0005737]; cytoplasmic vesicle [GO:0031410]; cytosol [GO:0005829]; lysosomal membrane [GO:0005765]; membrane [GO:0016020]; microtubule cytoskeleton [GO:0015630]; nucleus [GO:0005634]; trans-Golgi network

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			[GO:0005802]; cilium assembly [GO:0060271]; head development [GO:0060322]; heart development [GO:0007507]; intracellular protein transport [GO:0006886]; multicellular organism growth [GO:0035264]; regulation of smoothened signaling pathway [GO:0008589]; vesicle tethering to Golgi [GO:0099041]
GMOD_00010733	0.045	Os08g0135800 LOC_Os08g04170 OJ1613_G04.4-1 P0680F05.43-1	DNA binding [GO:0003677]; metal ion binding [GO:0046872]
GMOD_00010784	0.018	At5g14285 F18O22_70	membrane [GO:0016020]
GMOD_00010815	0.002	Mgat3 Gnt3	Golgi membrane [GO:0000139]; lysosome [GO:0005764]; beta-1,4-mannosylglycoprotein 4-beta-N-acetylglucosaminyltransferase activity [GO:0003830]; glycosyltransferase activity [GO:0016757]; amyloid-beta metabolic process [GO:0050435]; canonical Wnt signaling pathway [GO:0060070]; cellular response to oxidative stress [GO:0034599]; cognition [GO:0050890]; lysosomal protein catabolic process [GO:1905146]; N-acetylglucosamine metabolic process [GO:0006044]; negative regulation of lysosomal protein catabolic process [GO:1905166]; positive regulation of protein localization to early endosome [GO:1902966]; protein localization [GO:0008104]; protein localization to early endosome [GO:1902946]; protein N-linked glycosylation [GO:0006487]; regulation of cell migration [GO:0030334]
GMOD_00011075	0.027	NUCL2 NUC2 PARLL1 At3g18610 K24M9.11	nucleolus [GO:0005730]; ribonucleoprotein complex [GO:1990904]; nucleosome binding [GO:0031491]; RNA binding [GO:0003723]; rRNA processing [GO:0006364]
GMOD_00011078	0.009	ATL32 At4g40070 T5J17.240	membrane [GO:0016020]; metal ion binding [GO:0046872]; transferase activity [GO:0016740]; protein ubiquitination [GO:0016567]
GMOD_00011200	0.046	CYCB2-2 CYC2B At4g35620 F8D20.130	cyclin-dependent protein kinase holoenzyme complex [GO:0000307]; cytoplasm [GO:0005737]; nucleus [GO:0005634]; cyclin-dependent protein serine/threonine kinase regulator activity [GO:0016538]; cell division

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			[GO:0051301]; mitotic cell cycle phase transition [GO:0044772]; regulation of cyclin-dependent protein serine/threonine kinase activity [GO:0000079]
GMOD_00011294	0.047	P2A12 At1g12710 T12C24.23	carbohydrate binding [GO:0030246]
GMOD_00011382	0.031	BAD1 At5g54610 MRB17.11	plasma membrane [GO:0005886]; innate immune response [GO:0045087]; response to salicylic acid [GO:0009751]
GMOD_00011448	0.040	Os08g0121900 LOC_Os08g02850 B1203H11.32 OsJ_25865 OSJNBa0073J19.9 OSJNBa0091F23.17	membrane [GO:0016020]; glycosyltransferase activity [GO:0016757]
GMOD_00011574	0.039	LRK10L-2.1 At5g38260 MXA21.5	membrane [GO:0016020]; ATP binding [GO:0005524]; polysaccharide binding [GO:0030247]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; protein phosphorylation [GO:0006468]
GMOD_00011582	0.004	LRK10L-2.7 LRK10L2 At1g66930 F1O19.1 T4O24.2	membrane [GO:0016020]; ATP binding [GO:0005524]; polysaccharide binding [GO:0030247]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; protein phosphorylation [GO:0006468]
GMOD_00011594	0.024	SWEETIE At1g67140 F5A8.5	carbohydrate metabolic process [GO:0005975]; ethylene biosynthetic process [GO:0009693]; regulation of ethylene biosynthetic process [GO:0010364]; regulation of leaf senescence [GO:1900055]; starch metabolic process [GO:0005982]; trehalose metabolic process [GO:0005991]
GMOD_00011779	0.004	PAP9 FBN7a FIB7a At3g58010 T10K17.220	chloroplast [GO:0009507]; chloroplast thylakoid membrane [GO:0009535]; cytosol [GO:0005829]; plastoglobule [GO:0010287]; thylakoid [GO:0009579]; thylakoid lumen [GO:0031977]
GMOD_00011829	0.048	PME41 ARATH41	extracellular region [GO:0005576]; aspartyl esterase activity [GO:0045330]; pectinesterase activity [GO:0030599];

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
		At4g02330 T14P8.14	pectinesterase inhibitor activity [GO:0046910]; cell wall modification [GO:0042545]; defense response to fungus [GO:0050832]; pectin catabolic process [GO:0045490]; response to brassinosteroid [GO:0009741]; response to cold [GO:0009409]; response to fungus [GO:0009620]
GMOD_00011888	0.021	EDA41 At5g52460 K24M7.21	
GMOD_00011947	0.001	IMPA6 At1g02690 T14P4.3	cytoplasm [GO:0005737]; nuclear envelope [GO:0005635]; nucleus [GO:0005634]; nuclear import signal receptor activity [GO:0061608]; nuclear localization sequence binding [GO:0008139]; NLS-bearing protein import into nucleus [GO:0006607]
GMOD_00012037	0.014	ERD10 LTI29 LTI45 At1g20450 F5M15.21 F5M15_20	cytoplasm [GO:0005737]; cytosol [GO:0005829]; extrinsic component of membrane [GO:0019898]; membrane [GO:0016020]; nucleus [GO:0005634]; plasmodesma [GO:0009506]; actin binding [GO:0003779]; copper ion binding [GO:0005507]; lipid binding [GO:0008289]; nickel cation binding [GO:0016151]; phosphatidylcholine binding [GO:0031210]; phosphatidylserine binding [GO:0001786]; protein folding chaperone [GO:0044183]; cellular water homeostasis [GO:0009992]; cold acclimation [GO:0009631]; protein stabilization [GO:0050821]; regulation of membrane permeability [GO:0090559]; regulation of seed germination [GO:0010029]; response to abscisic acid [GO:0009737]; response to cold [GO:0009409]; response to water deprivation [GO:0009414]
GMOD_00012054	0.039	RNR2B	cytoplasm [GO:0005737]; metal ion binding [GO:0046872]; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor [GO:0004748]; deoxyribonucleotide biosynthetic process [GO:0009263]; DNA replication [GO:0006260]
GMOD_00012156	0.002		
GMOD_00012237	0.015	PCMP-E42 PCMP-E60 At1g50270 F14I3.12	

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00012379	0.003	NFYC6 At5g50480 MBA10.3	nucleus [GO:0005634]; DNA-binding transcription activator activity, RNA polymerase II-specific [GO:0001228]; DNA-binding transcription factor activity [GO:0003700]; protein heterodimerization activity [GO:0046982]; transcription cis-regulatory region binding [GO:0000976]; regulation of DNA-templated transcription [GO:0006355]
GMOD_00012845	0.026	CYP716A1 At5g36110 MAB16.5	membrane [GO:0016020]; heme binding [GO:0020037]; iron ion binding [GO:0005506]; monooxygenase activity [GO:0004497]; oxidoreductase activity [GO:0016491]; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen [GO:0016709]; sterol metabolic process [GO:0016125]; triterpenoid metabolic process [GO:0006722]
GMOD_00013016	0.005	CHS1 At1g17610 F1L3.31	cytoplasm [GO:0005737]; nucleus [GO:0005634]; ADP binding [GO:0043531]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleotidase, cyclic ADP-ribose generating [GO:0061809]; borate transport [GO:0046713]; defense response [GO:0006952]; signal transduction [GO:0007165]
GMOD_00013175	0.038	At2g19130 T20K24.15	plasma membrane [GO:0005886]; ATP binding [GO:0005524]; calmodulin binding [GO:0005516]; carbohydrate binding [GO:0030246]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; ubiquitin protein ligase binding [GO:0031625]; protein phosphorylation [GO:0006468]; recognition of pollen [GO:0048544]
GMOD_00013335	0.001	LECRK11 At3g45330 F18N11.90	plasma membrane [GO:0005886]; ATP binding [GO:0005524]; carbohydrate binding [GO:0030246]; protein serine kinase activity [GO:0106310]; transmembrane receptor protein serine/threonine kinase activity [GO:0004675]; defense response to bacterium [GO:0042742]; defense response to fungus [GO:0050832]; defense response to oomycetes [GO:0002229]; protein phosphorylation [GO:0006468]
GMOD_00013370	0.031	NPF2.3 At3g45680 T6D9.10	plasma membrane [GO:0005886]; nitrate transmembrane transporter activity [GO:0015112]; cellular response to salt stress [GO:0071472]; nitrate assimilation [GO:0042128]; nitrate transmembrane transport [GO:0015706];

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			oligopeptide transport [GO:0006857]; response to nematode [GO:0009624]
GMOD_00013398	0.002	At1g67720 F12A21.30	membrane [GO:0016020]; ATP binding [GO:0005524]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; protein phosphorylation [GO:0006468]
GMOD_00013423	0.016	ABCG33 PDR5 At2g37280 F3G5.7	plasma membrane [GO:0005886]; ABC-type transporter activity [GO:0140359]; ATP binding [GO:0005524]
GMOD_00013425	0.023		
GMOD_00013757	0.001		
GMOD_00013759	0.002	ABCF4 GCN4 At3g54540 T14E10.110	cytosol [GO:0005829]; ATP binding [GO:0005524]
GMOD_00013851	0.009	GSO2 EDA23 At5g44700 K23L20.3	Casparian strip [GO:0048226]; plasma membrane [GO:0005886]; ATP binding [GO:0005524]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; cell wall organization [GO:0071555]; embryo sac development [GO:0009553]; protein phosphorylation [GO:0006468]; regulation of cell division [GO:0051302]; regulation of cell fate specification [GO:0042659]; regulation of root development [GO:2000280]; regulation of root morphogenesis [GO:2000067]; specification of plant organ axis polarity [GO:0090708]
GMOD_00013873	0.020	At3g19880 MPN9.12	
GMOD_00014090	0.008	usf	hydrolase activity [GO:0016787]
GMOD_00014100	0.014	At2g32630 T26B15.19	
GMOD_00014169	0.006		
GMOD_00014333	0.019	CLPB4 CLPB-M HSP98.7 At2g25140 F13D4.100	chloroplast [GO:0009507]; chloroplast envelope [GO:0009941]; chloroplast stroma [GO:0009570]; cytoplasm [GO:0005737]; mitochondrion [GO:0005739]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; cellular response to heat [GO:0034605]; protein refolding [GO:0042026]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00014549	0.031	UBC23 At2g16920 F12A24.10	ATP binding [GO:0005524]; ubiquitin conjugating enzyme activity [GO:0061631]
GMOD_00014630	0.022	DTX14 At1g71140 F23N20.13	membrane [GO:0016020]; plant-type vacuole [GO:0000325]; antiporter activity [GO:0015297]; transmembrane transporter activity [GO:0022857]; xenobiotic transmembrane transporter activity [GO:0042910]; xenobiotic detoxification by transmembrane export across the plasma membrane [GO:1990961]; xenobiotic transport [GO:0042908]
GMOD_00014643	0.023	WHY2 At1g71260 F3I17.9	DNA repair complex [GO:1990391]; mitochondrion [GO:0005739]; DNA binding [GO:0003677]; mRNA binding [GO:0003729]; single-stranded DNA binding [GO:0003697]; defense response [GO:0006952]; DNA repair [GO:0006281]; regulation of DNA-templated transcription [GO:0006355]
GMOD_00014713	0.011	At3g47200 F13I12.250	plasma membrane [GO:0005886]
GMOD_00014767	0.027	At4g33290 F17M5.50	
GMOD_00014856	0.036	MES2 ACL ME8 At2g23600 F26B6.25	cytosol [GO:0005829]; vacuole [GO:0005773]; carboxylic ester hydrolase activity [GO:0052689]; hydrolase activity, acting on ester bonds [GO:0016788]; methyl indole-3-acetate esterase activity [GO:0080030]; methyl jasmonate esterase activity [GO:0080032]; methyl salicylate esterase activity [GO:0080031]; jasmonic acid metabolic process [GO:0009694]; nicotinate metabolic process [GO:1901847]; oxylipin biosynthetic process [GO:0031408]; salicylic acid metabolic process [GO:0009696]
GMOD_00014876	0.032	AATP1 ASD At5g40010 MYH19.21	endoplasmic reticulum [GO:0005783]; mitochondrial membrane [GO:0031966]; mitochondrion [GO:0005739]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; fruit development [GO:0010154]; response to abscisic acid [GO:0009737]; response to cold [GO:0009409]; response to salt stress [GO:0009651]; response to water deprivation [GO:0009414]; seed maturation [GO:0010431]
GMOD_00014934	0.031	XBAT35 At3g23280 K14B15.19	cytoplasm [GO:0005737]; nucleus [GO:0005634]; metal ion binding [GO:0046872]; ubiquitin-protein transferase activity [GO:0004842]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00014992	0.002	At3g22940 F5N5.12	
GMOD_00015035	0.027	CLE43 At1g25425 F2J7	apoplast [GO:0048046]; extracellular space [GO:0005615]; receptor serine/threonine kinase binding [GO:0033612]; cell-cell signaling involved in cell fate commitment [GO:0045168]
GMOD_00015121	0.001	ABCG24 WBC25 At1g53390 F12M16.28	membrane [GO:0016020]; ABC-type transporter activity [GO:0140359]; ATP binding [GO:0005524]; ATPase-coupled transmembrane transporter activity [GO:0042626]; transmembrane transport [GO:0055085]
GMOD_00015184	0.002		
GMOD_00015203	0.034	At3g28330 MZF16.14	nucleus [GO:0005634]; zinc ion binding [GO:0008270]; cell cycle [GO:0007049]; cell division [GO:0051301]; protein ubiquitination [GO:0016567]
GMOD_00015205	0.046		ubiquitin-protein transferase activity [GO:0004842]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
GMOD_00015207	0.013		
GMOD_00015585	0.021	At3g49450 T1G12.15 T9C5.50	
GMOD_00015645	0.002	SAT	acyltransferase activity [GO:0016746]; oxidoreductase activity [GO:0016491]; alkaloid metabolic process [GO:0009820]
GMOD_00015668	0.035		
GMOD_00015682	0.015	AAP19-1 At2g17380 F5J6.14	AP-1 adaptor complex [GO:0030121]; intracellular membrane-bounded organelle [GO:0043231]; clathrin adaptor activity [GO:0035615]; intracellular protein transport [GO:0006886]; vesicle-mediated transport [GO:0016192]
GMOD_00015706	0.019	BSK12 SSP At2g17090 F6P23.23	cytoplasmic side of plasma membrane [GO:0009898]; ATP binding [GO:0005524]; protein serine/threonine kinase activity [GO:0004674]; brassinosteroid mediated signaling pathway [GO:0009742]; positive regulation of intracellular signal transduction [GO:1902533]; protein phosphorylation [GO:0006468]; suspensor development [GO:0010098]; zygote elongation [GO:0080159]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00015712	0.031	REL2 LOC_Os10g41310 Os10g0562700 OsJ_32472 OSJNBb0089A17.4	plasma membrane [GO:0005886]; cell differentiation [GO:0030154]
GMOD_00015743	0.022	yqkD BSU23640	membrane [GO:0016020]
GMOD_00015780	0.003	STARD7 GTT1	mitochondrial outer membrane [GO:0005741]; lipid binding [GO:0008289]
GMOD_00015824	0.004		
GMOD_00015840	0.001	At4g20800 F21C20.150	extracellular region [GO:0005576]; plant-type cell wall [GO:0009505]; FAD binding [GO:0071949]; oxidoreductase activity [GO:0016491]
GMOD_00015847	0.013	At4g20460 F9F13.110	Golgi cisterna membrane [GO:0032580]; UDP-arabinose 4-epimerase activity [GO:0050373]; UDP-glucose 4-epimerase activity [GO:0003978]; capsule polysaccharide biosynthetic process [GO:0045227]; galactose metabolic process [GO:0006012]; UDP-L-arabinose biosynthetic process [GO:0033358]
GMOD_00015858	0.009	VPS2.2 CHMP2-2 At5g44560 MFC16.25	ESCRT III complex [GO:0000815]; multivesicular body [GO:0005771]; endosome transport via multivesicular body sorting pathway [GO:0032509]; late endosome to vacuole transport [GO:0045324]; protein transport [GO:0015031]
GMOD_00015991	0.001	At4g00315 A_IG005I10.18 F5I10.18	negative regulation of long-day photoperiodism, flowering [GO:0048579]; regulation of short-day photoperiodism, flowering [GO:0048587]
GMOD_00015997	0.028	At4g09580 T25P22.20	endoplasmic reticulum [GO:0005783]; membrane [GO:0016020]
GMOD_00016037	0.040		cytoplasm [GO:0005737]; dynein complex [GO:0030286]; microtubule [GO:0005874]; motile cilium [GO:0031514]; microtubule-based process [GO:0007017]
GMOD_00016148	0.002	RPA1E RPA70E At4g19130 T18B16.100	DNA replication factor A complex [GO:0005662]; damaged DNA binding [GO:0003684]; single-stranded telomeric DNA binding [GO:0043047]; zinc ion binding [GO:0008270]; DNA unwinding involved in DNA replication [GO:0006268]; double-strand break repair via homologous recombination [GO:0000724]; meiotic cell cycle [GO:0051321]; nucleotide-

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			excision repair [GO:0006289]; telomere maintenance via telomerase [GO:0007004]
GMOD_00016175	0.005	NGA3 At1g01030 T25K16.3	nucleus [GO:0005634]; DNA binding [GO:0003677]; DNA-binding transcription factor activity [GO:0003700]; flower development [GO:0009908]; leaf development [GO:0048366]; regulation of DNA-templated transcription [GO:0006355]; regulation of leaf morphogenesis [GO:1901371]
GMOD_00016206	0.001	At2g05300 F5G3.20	
GMOD_00016223	0.026	At1g20795 F2D10.29	
GMOD_00016351	0.007	PER44 P44 At4g26010 F20B18.120	extracellular region [GO:0005576]; plant-type cell wall [GO:0009505]; heme binding [GO:0020037]; lactoperoxidase activity [GO:0140825]; metal ion binding [GO:0046872]; peroxidase activity [GO:0004601]; hydrogen peroxide catabolic process [GO:0042744]; response to oxidative stress [GO:0006979]
GMOD_00016420	0.018		
GMOD_00016422	0.002	SBT1.3 At5g51750 MIO24.12	extracellular region [GO:0005576]; serine-type endopeptidase activity [GO:0004252]; proteolysis [GO:0006508]
GMOD_00016520	0.003	ACD6 At4g14400 dl3240w FCAALL.190	endoplasmic reticulum membrane [GO:0005789]; membrane [GO:0016020]; plasma membrane [GO:0005886]; cell death [GO:0008219]; cellular response to salicylic acid stimulus [GO:0071446]; defense response to bacterium [GO:0042742]; positive regulation of defense response to bacterium [GO:1900426]; positive regulation of defense response to oomycetes [GO:1902290]; protein ubiquitination [GO:0016567]; regulation of defense response to fungus [GO:1900150]; regulation of salicylic acid mediated signaling pathway [GO:2000031]; response to bacterium [GO:0009617]; response to freezing [GO:0050826]; response to light stimulus [GO:0009416]; response to salicylic acid [GO:0009751]; response to virus [GO:0009615]
GMOD_00016781	0.037	RPS6	ADP binding [GO:0043531]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleotidase, cyclic ADP-ribose

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			generating [GO:0061809]; defense response [GO:0006952]; signal transduction [GO:0007165]
GMOD_00016802	0.008	TAF1 GTD1 HAC13 HAF1 TAF1A At1g32750 F6N18.13	transcription factor TFIID complex [GO:0005669]; histone acetyltransferase activity [GO:0004402]; RNA polymerase II general transcription initiation factor activity [GO:0016251]; RNA polymerase II general transcription initiation factor binding [GO:0001091]; TBP-class protein binding [GO:0017025]; chromatin organization [GO:0006325]; DNA-mediated transformation [GO:0009294]; RNA polymerase II preinitiation complex assembly [GO:0051123]
GMOD_00016979	0.004	BSU1 At1g03445 F21B7.7	cytoplasm [GO:0005737]; nucleus [GO:0005634]; plasma membrane [GO:0005886]; metal ion binding [GO:0046872]; myosin phosphatase activity [GO:0017018]; protein serine/threonine phosphatase activity [GO:0004722]; brassinosteroid mediated signaling pathway [GO:0009742]; positive regulation of brassinosteroid mediated signaling pathway [GO:1900459]; regulation of protein localization [GO:0032880]
GMOD_00017116	0.026	At1g58390 F9K23.8 X7J.2	ADP binding [GO:0043531]; ATP binding [GO:0005524]; defense response [GO:0006952]; response to other organism [GO:0051707]
GMOD_00017125	0.001		
GMOD_00017128	0.050		
GMOD_00017139	0.037	IDM1 ROS4 At3g14980 K15M2.12	nucleus [GO:0005634]; plasmodesma [GO:0009506]; double-stranded methylated DNA binding [GO:0010385]; histone acetyltransferase activity [GO:0004402]; histone binding [GO:0042393]; metal ion binding [GO:0046872]; gene silencing by RNA-directed DNA methylation [GO:0080188]; histone H3-K14 acetylation [GO:0044154]; histone H3-K18 acetylation [GO:0043971]; histone H3-K23 acetylation [GO:0043972]; regulation of DNA methylation [GO:0044030]
GMOD_00017333	0.049		
GMOD_00017334	0.000	At5g58620 MZN1.16 MZN1.8	cytoplasmic stress granule [GO:0010494]; P-body [GO:0000932]; DNA binding [GO:0003677]; DNA-binding transcription factor activity [GO:0003700]; metal ion binding [GO:0046872]; RNA binding [GO:0003723]; regulation of DNA-templated transcription [GO:0006355]; translation [GO:0006412]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00017412	0.034	At5g22670 MDJ22.9	
GMOD_00017429	0.028	PUR3 At1g31220 F28K20.18	chloroplast [GO:0009507]; cytoplasm [GO:0005737]; phosphoribosylglycinamide formyltransferase activity [GO:0004644]; 'de novo' IMP biosynthetic process [GO:0006189]
GMOD_00017505	0.040	Wdr70	nucleus [GO:0005634]; site of double-strand break [GO:0035861]; enzyme binding [GO:0019899]; regulation of DNA double-strand break processing [GO:1903775]
GMOD_00017680	0.041		
GMOD_00017769	0.000	ABCG36 PDR8 PEN3 At1g59870 F23H11.19	chloroplast [GO:0009507]; chloroplast envelope [GO:0009941]; endoplasmic reticulum [GO:0005783]; endoplasmic reticulum membrane [GO:0005789]; mitochondrion [GO:0005739]; plant-type vacuole [GO:0000325]; plasma membrane [GO:0005886]; trans-Golgi network [GO:0005802]; ABC-type transporter activity [GO:0140359]; ATP binding [GO:0005524]; auxin efflux transmembrane transporter activity [GO:0010329]; cadmium ion transmembrane transporter activity [GO:0015086]; efflux transmembrane transporter activity [GO:0015562]; mRNA binding [GO:0003729]; auxin polar transport [GO:0009926]; auxin-activated signaling pathway [GO:0009734]; cadmium ion transport [GO:0015691]; cellular detoxification [GO:1990748]; cellular response to indolebutyric acid stimulus [GO:0071366]; cotyledon development [GO:0048825]; coumarin metabolic process [GO:0009804]; defense response by callose deposition in cell wall [GO:0052544]; defense response to bacterium [GO:0042742]; defense response to fungus [GO:0050832]; defense response to oomycetes [GO:0002229]; export from cell [GO:0140352]; indole glucosinolate catabolic process [GO:0042344]; negative regulation of defense response [GO:0031348]; pathogen-associated molecular pattern receptor signaling pathway [GO:0140426]; plant-type hypersensitive response [GO:0009626]; regulation of auxin mediated signaling pathway [GO:0010928]; regulation of defense response by callose deposition [GO:2000071]; regulation of lateral root development [GO:2000023]; response to auxin [GO:0009733]; response to bacterium [GO:0009617]; response to chitin [GO:0010200]; response

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			to fungus [GO:0009620]; response to heat [GO:0009408]; response to molecule of bacterial origin [GO:0002237]; response to molecule of fungal origin [GO:0002238]; response to molecule of oomycetes origin [GO:0002240]; response to oomycetes [GO:0002239]; response to salt stress [GO:0009651]; response to water deprivation [GO:0009414]; root development [GO:0048364]; root hair cell development [GO:0080147]; sodium ion homeostasis [GO:0055078]; systemic acquired resistance [GO:0009627]; thymidine metabolic process [GO:0046104]; transmembrane transport [GO:0055085]
GMOD_00017803	0.024	APUM9 At1g35730 F14D7.3	cytoplasm [GO:0005737]; mRNA binding [GO:0003729]; negative regulation of seed dormancy process [GO:1902039]; post-transcriptional regulation of gene expression [GO:0010608]; regulation of translation [GO:0006417]
GMOD_00017869	0.018	CRSP SBT5.2 At1g20160 T20H2.6	apoplast [GO:0048046]; serine-type endopeptidase activity [GO:0004252]; serine-type peptidase activity [GO:0008236]; tripeptidyl-peptidase activity [GO:0008240]; negative regulation of defense response to bacterium [GO:1900425]; negative regulation of stomatal complex development [GO:2000122]; proteolysis [GO:0006508]; regulation of stomatal complex development [GO:2000038]; response to carbon dioxide [GO:0010037]
GMOD_00018061	0.010	At2g16450 F16F14.5	
GMOD_00018072	0.014	At2g29780 T27A16.12	
GMOD_00018103	0.007	RL2 MEE3 RMS1 At2g21650 F2G1.8	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; embryo development ending in seed dormancy [GO:0009793]; gravitropism [GO:0009630]; response to red light [GO:0010114]
GMOD_00018126	0.011	abhd17b	dendritic spine [GO:0043197]; endosome membrane [GO:0010008]; plasma membrane [GO:0005886]; postsynaptic density membrane [GO:0098839]; recycling endosome membrane [GO:0055038]; palmitoyl-(protein) hydrolase activity [GO:0008474]; protein depalmitoylation [GO:0002084]; regulation of postsynapse organization [GO:0099175]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00018148	0.018	LTI65, RD29B, At5g52300, K24M7.3	abscisic acid-activated signaling pathway [GO:0009738]; leaf senescence [GO:0010150]; response to abscisic acid [GO:0009737]; response to cold [GO:0009409]; response to salt stress [GO:0009651]; response to water deprivation [GO:0009414]
GMOD_00018172	0.010	CRRSP20 At3g21933 MZN24.6	extracellular region [GO:0005576]
GMOD_00018258	0.036	At4g11590 T5C23.20	nucleus [GO:0005634]; protein ubiquitination [GO:0016567]
GMOD_00018265	0.001		
GMOD_00018286	0.011	At2g21930 F7D8.25	
GMOD_00018319	0.003	F3H-3 F3H Os04g0667200 LOC_Os04g57160 OSJNBa0043A12.1	L-ascorbic acid binding [GO:0031418]; metal ion binding [GO:0046872]; naringenin 3-dioxygenase activity [GO:0045486]; flavonoid biosynthetic process [GO:0009813]
GMOD_00018399	0.026	RPK1 TOAD1 At1g69270 F23O10.15 F4N2.23	peroxisome [GO:0005777]; plasma membrane [GO:0005886]; ATP binding [GO:0005524]; identical protein binding [GO:0042802]; protein kinase activity [GO:0004672]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; abscisic acid-activated signaling pathway [GO:0009738]; embryonic meristem development [GO:0048508]; longitudinal axis specification [GO:0009942]; protein phosphorylation [GO:0006468]; radial axis specification [GO:0009945]; response to abscisic acid [GO:0009737]; response to cold [GO:0009409]; response to salt stress [GO:0009651]; response to water deprivation [GO:0009414]
GMOD_00018426	0.034	RPP1 At3g44480 F14L2_30	endoplasmic reticulum membrane [GO:0005789]; Golgi membrane [GO:0000139]; plasma membrane [GO:0005886]; ADP binding [GO:0043531]; ATP binding [GO:0005524]; innate immune receptor activity [GO:0140376]; LRR domain binding [GO:0030275]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleotidase, cyclic ADP-ribose generating [GO:0061809]; defense response [GO:0006952]; defense response to fungus

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			[GO:0050832]; innate immune response-activating signaling pathway [GO:0002758]; positive regulation of plant-type hypersensitive response [GO:0034052]; response to oomycetes [GO:0002239]
GMOD_00018452	0.000	At2g42480 MHK10.20	
GMOD_00018494	0.000	RD21A At1g47128 F2G19.31	apoplast [GO:0048046]; cytoplasm [GO:0005737]; cytoplasmic stress granule [GO:0010494]; extracellular space [GO:0005615]; Golgi apparatus [GO:0005794]; lysosome [GO:0005764]; nucleus [GO:0005634]; P-body [GO:0000932]; plant-type vacuole [GO:0000325]; plasma membrane [GO:0005886]; plasmodesma [GO:0009506]; secretory vesicle [GO:0099503]; vacuole [GO:0005773]; cysteine-type endopeptidase activity [GO:0004197]; cysteine-type peptidase activity [GO:0008234]; peptidase activity [GO:0008233]; defense response to fungus [GO:0050832]; protein ubiquitination [GO:0016567]; proteolysis [GO:0006508]; proteolysis involved in protein catabolic process [GO:0051603]; regulation of stomatal closure [GO:0090333]
GMOD_00018497	0.049	AGL86 At1g31630 F27M3.17	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; DNA-binding transcription factor activity, RNA polymerase II-specific [GO:0000981]; protein dimerization activity [GO:0046983]; RNA polymerase II cis-regulatory region sequence-specific DNA binding [GO:0000978]; positive regulation of transcription by RNA polymerase II [GO:0045944]; regulation of transcription by RNA polymerase II [GO:0006357]
GMOD_00018555	0.004	ABCA3 ATH2 At3g47740 T23J7.70	intracellular membrane-bounded organelle [GO:0043231]; membrane [GO:0016020]; ABC-type transporter activity [GO:0140359]; ATP binding [GO:0005524]; ATPase-coupled transmembrane transporter activity [GO:0042626]; lipid transporter activity [GO:0005319]; lipid transport [GO:0006869]
GMOD_00018575	0.013	At3g47570 F1P2.120	plasma membrane [GO:0005886]; ATP binding [GO:0005524]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; protein phosphorylation [GO:0006468]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00018628	0.000	At2g28450 T1B3.3	metal ion binding [GO:0046872]; RNA methyltransferase activity [GO:0008173]; transcription cis-regulatory region binding [GO:0000976]; RNA processing [GO:0006396]
GMOD_00019138	0.003	LRX1 At1g12040 F12F1.9	extracellular region [GO:0005576]; plant-type cell wall [GO:0009505]; plasmodesma [GO:0009506]; structural constituent of cell wall [GO:0005199]; cell morphogenesis involved in differentiation [GO:0000904]; cell wall organization [GO:0071555]; trichoblast differentiation [GO:0010054]; unidimensional cell growth [GO:0009826]
GMOD_00019304	0.000	DMT1 ATHIM DDM2 DMT01 MET1 MET2 At5g49160 K21P3.3	nucleus [GO:0005634]; chromatin binding [GO:0003682]; DNA (cytosine-5-)-methyltransferase activity [GO:0003886]; DNA binding [GO:0003677]; methyltransferase activity [GO:0008168]; DNA methylation on cytosine within a CG sequence [GO:0010424]; DNA-mediated transformation [GO:0009294]; maintenance of DNA methylation [GO:0010216]; negative regulation of flower development [GO:0009910]; regulation of gene expression by genomic imprinting [GO:0006349]; zygote asymmetric cytokinesis in embryo sac [GO:0010069]
GMOD_00019331	0.027	ATL56 At2g18670 MSF3.5	membrane [GO:0016020]; metal ion binding [GO:0046872]; transferase activity [GO:0016740]; cellular response to hypoxia [GO:0071456]; protein ubiquitination [GO:0016567]
GMOD_00019396	0.046	At3g24580 MOB24.16	
GMOD_00019465	0.003	At2g29810 T27A16.9	
GMOD_00019487	0.038	BGLU22 At1g66280 T27F4.3	endoplasmic reticulum [GO:0005783]; endoplasmic reticulum lumen [GO:0005788]; mitochondrion [GO:0005739]; plasmodesma [GO:0009506]; beta-glucosidase activity [GO:0008422]; scopolin beta-glucosidase activity [GO:0102483]; carbohydrate metabolic process [GO:0005975]; cellular response to cold [GO:0070417]; cellular response to salt stress [GO:0071472]; glucosinolate catabolic process [GO:0019762]; response to salt stress [GO:0009651]
GMOD_00019557	0.008	NAC012 NST3 SND1 At1g32770 F6N18.15	nucleus [GO:0005634]; DNA-binding transcription factor activity [GO:0003700]; sequence-specific DNA binding [GO:0043565]; transcription cis-regulatory region binding [GO:0000976]; fruit dehiscence [GO:0010047]; plant-type

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			secondary cell wall biogenesis [GO:0009834]; positive regulation of DNA-templated transcription [GO:0045893]
GMOD_00019562	0.010	At1g32140 F3C3.15	
GMOD_00019832	0.002		plasma membrane [GO:0005886]; ATP binding [GO:0005524]; protein serine kinase activity [GO:0106310]; protein serine/threonine kinase activity [GO:0004674]; defense response to bacterium [GO:0042742]; protein phosphorylation [GO:0006468]
GMOD_00019949	0.009	DTX15 At2g34360 F13P17.20	membrane [GO:0016020]; antiporter activity [GO:0015297]; transmembrane transporter activity [GO:0022857]; xenobiotic transmembrane transporter activity [GO:0042910]; xenobiotic detoxification by transmembrane export across the plasma membrane [GO:1990961]
GMOD_00020081	0.042	At3g51120 F24M12.160	DNA binding [GO:0003677]; metal ion binding [GO:0046872]
GMOD_00020176	0.005	TEN1 MDO1 At1g56260 F14G9.13	chromosome, telomeric region [GO:0000781]; CST complex [GO:1990879]; protein folding chaperone [GO:0044183]; single-stranded DNA binding [GO:0003697]; telomerase inhibitor activity [GO:0010521]; telomeric DNA binding [GO:0042162]; negative regulation of telomere maintenance via telomerase [GO:0032211]; response to heat [GO:0009408]; shoot system development [GO:0048367]; stem cell population maintenance [GO:0019827]; telomere maintenance [GO:0000723]
GMOD_00020369	0.004		
GMOD_00020378	0.008	RLP26 At2g33050 F25I18.21 T21L14	plasma membrane [GO:0005886]
GMOD_00020446	0.008	PCMP-E36 At4g21300 T6K22.30	intracellular membrane-bounded organelle [GO:0043231]; RNA binding [GO:0003723]; RNA modification [GO:0009451]
GMOD_00020468	0.045	DOF4.4 At4g21050 T13K14.210	nucleus [GO:0005634]; DNA binding [GO:0003677]; DNA-binding transcription factor activity [GO:0003700]; metal ion binding [GO:0046872]; fruit development [GO:0010154]; positive regulation of DNA-templated transcription [GO:0045893]; regulation of DNA-templated transcription

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			[GO:0006355]; regulation of secondary shoot formation [GO:2000032]
GMOD_00020490	0.015	At2g17723 T17A5	extracellular region [GO:0005576]; defense response to fungus [GO:0050832]; killing of cells of another organism [GO:0031640]
GMOD_00020503	0.038		
GMOD_00020609	0.005	APRL4 At1g34780 F11O6.7 F21H2.1	membrane [GO:0016020]
GMOD_00020716	0.000	FBX9 At2g04920 F1O13.5	
GMOD_00020890	0.003		
GMOD_00020898	0.013		
GMOD_00020924	0.002	TGA2 BZIP20 HBP1B At5g06950 MOJ9.12	cytoplasm [GO:0005737]; nucleus [GO:0005634]; RNA polymerase II transcription regulator complex [GO:0090575]; RNA polymerase II transcription repressor complex [GO:0090571]; DNA binding [GO:0003677]; DNA-binding transcription factor activity [GO:0003700]; transcription cis-regulatory region binding [GO:0000976]; negative regulation of DNA-templated transcription [GO:0045892]; plant-type hypersensitive response [GO:0009626]; positive regulation of DNA-templated transcription [GO:0045893]; response to xenobiotic stimulus [GO:0009410]; systemic acquired resistance, salicylic acid mediated signaling pathway [GO:0009862]
GMOD_00020950	0.006	PABN1 At5g51120 MWD22.6	cytoplasm [GO:0005737]; nuclear speck [GO:0016607]; RNA binding [GO:0003723]; mRNA processing [GO:0006397]
GMOD_00020954	0.005	FRI	nuclear speck [GO:0016607]; cell differentiation [GO:0030154]; flower development [GO:0009908]
GMOD_00021162	0.001	WRKY19 At4g12020 F16J13.90	chloroplast envelope [GO:0009941]; nucleus [GO:0005634]; ADP binding [GO:0043531]; ATP binding [GO:0005524]; DNA-binding transcription factor activity [GO:0003700]; protein serine/threonine kinase activity [GO:0004674]; transcription cis-regulatory region binding [GO:0000976];

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			defense response [GO:0006952]; protein phosphorylation [GO:0006468]; signal transduction [GO:0007165]
GMOD_00021223	0.004	MOD1 ENR-A ENR1 At2g05990 T6P5.19	chloroplast [GO:0009507]; chloroplast envelope [GO:0009941]; chloroplast stroma [GO:0009570]; fatty acid synthase complex [GO:0005835]; mitochondrion [GO:0005739]; thylakoid [GO:0009579]; copper ion binding [GO:0005507]; enoyl-[acyl-carrier-protein] reductase (NADH) activity [GO:0004318]; enoyl-[acyl-carrier-protein] reductase activity [GO:0016631]; fatty acid biosynthetic process [GO:0006633]
GMOD_00021286	0.024	DRT100 At3g12610 MMF12.5 T2E22.8 T2E22_107	chloroplast [GO:0009507]; nucleotide binding [GO:0000166]; DNA repair [GO:0006281]
GMOD_00021324	0.046	XAT3 Os03g0567600 LOC_Os03g37010 OsJ_11476 OSJNBa0026A15.4	Golgi membrane [GO:0000139]; arabinosyltransferase activity [GO:0052636]; glycosyltransferase activity [GO:0016757]; plant-type cell wall organization [GO:0009664]
GMOD_00021414	0.000		
GMOD_00021655	0.000	At5g36200 MAB16.15	plasmodesma [GO:0009506]
GMOD_00021811	0.001	At5g39000 MXF12.2	plasma membrane [GO:0005886]; plastid [GO:0009536]; ATP binding [GO:0005524]; protein kinase activity [GO:0004672]; protein serine/threonine kinase activity [GO:0004674]; transmembrane receptor protein tyrosine kinase activity [GO:0004714]; protein autophosphorylation [GO:0046777]; response to metal ion [GO:0010038]
GMOD_00021869	0.038		
GMOD_00021951	0.032	IAN3 At1g33890 T3M13.9	endoplasmic reticulum [GO:0005783]; plasma membrane [GO:0005886]; GTP binding [GO:0005525]; cellular response to heat [GO:0034605]; endoplasmic reticulum unfolded protein response [GO:0030968]
GMOD_00022038	0.001		
GMOD_00022089	0.003	ATJ49 C49 At5g49060	endoplasmic reticulum membrane [GO:0005789]; Hsp70 protein binding [GO:0030544]; cellular response to misfolded protein [GO:0071218]; chaperone cofactor-

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
		K19E20.16 K20J1_3	dependent protein refolding [GO:0051085]; ubiquitin-dependent ERAD pathway [GO:0030433]
GMOD_00022566	0.015	UCC1 At2g32300 T32F6	plasma membrane [GO:0005886]; electron transfer activity [GO:0009055]; metal ion binding [GO:0046872]
GMOD_00022700	0.027		
GMOD_00022719	0.001	PAB1 KLLA0C17600g	cytoplasm [GO:0005737]; nucleus [GO:0005634]; RNA binding [GO:0003723]; mRNA processing [GO:0006397]; mRNA transport [GO:0051028]; regulation of translation [GO:0006417]
GMOD_00022722	0.032	PAB2 At4g34110 F28A23.130	cytosol [GO:0005829]; nucleus [GO:0005634]; plasma membrane [GO:0005886]; ribonucleoprotein complex [GO:1990904]; mRNA 3'-UTR binding [GO:0003730]; mRNA binding [GO:0003729]; poly(A) binding [GO:0008143]; poly(U) RNA binding [GO:0008266]; RNA binding [GO:0003723]; translation initiation factor activity [GO:0003743]; nuclear-transcribed mRNA catabolic process, nonsense-mediated decay [GO:0000184]; regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay [GO:1900151]; regulation of nuclear-transcribed mRNA poly(A) tail shortening [GO:0060211]; regulation of translational initiation [GO:0006446]; translational initiation [GO:0006413]
GMOD_00022741	0.044		
GMOD_00022770	0.044		
GMOD_00022788	0.025	At2g26860 F12C20.10	
GMOD_00022797	0.000	At1g62000 F8K4.19	apoplast [GO:0048046]; mucilage biosynthetic process [GO:0010192]; seed coat development [GO:0010214]
GMOD_00022850	0.004	INVH At3g05820 F10A16.11	chloroplast [GO:0009507]; mitochondrion [GO:0005739]; glycopeptide alpha-N-acetylgalactosaminidase activity [GO:0033926]; sucrose alpha-glucosidase activity [GO:0004575]; sucrose catabolic process [GO:0005987]
GMOD_00022973	0.028	NIH At1g06670 F12K11.4	nucleus [GO:0005634]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; DNA binding [GO:0003677]; DNA helicase activity [GO:0003678]; RNA binding [GO:0003723]; RNA helicase activity [GO:0003724]; DNA metabolic process [GO:0006259]

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00023000	0.014	DRM3 At3g17310	nucleus [GO:0005634]; DNA binding [GO:0003677]; methyltransferase activity [GO:0008168]; DNA methylation [GO:0006306]
GMOD_00023005	0.032		
GMOD_00023021	0.000	DUF3 At1g26540 T1K7.9	plant-type vacuole [GO:0000325]; protein homodimerization activity [GO:0042803]
GMOD_00023070	0.008	At4g01037 A_IG002N01.30	chloroplast [GO:0009507]; mRNA binding [GO:0003729]; RNA binding [GO:0003723]; Group II intron splicing [GO:0000373]; mRNA processing [GO:0006397]; photosynthesis [GO:0015979]
GMOD_00023144	0.037	SPAC56F8.03	cytoplasm [GO:0005737]; cytosol [GO:0005829]; eukaryotic 48S preinitiation complex [GO:0033290]; mitochondrion [GO:0005739]; GTP binding [GO:0005525]; GTPase activity [GO:0003924]; metal ion binding [GO:0046872]; translation initiation factor activity [GO:0003743]; cytoplasmic translational initiation [GO:0002183]; translational initiation [GO:0006413]
GMOD_00023346	0.033	GAD4 At2g02010 F14H20.8	cytosol [GO:0005829]; mitochondrion [GO:0005739]; calmodulin binding [GO:0005516]; glutamate decarboxylase activity [GO:0004351]; pyridoxal phosphate binding [GO:0030170]; glutamate catabolic process [GO:0006538]
GMOD_00023446	0.008	ILL6 GR1 At1g44350 T18F15.9	hydrolase activity [GO:0016787]; jasmonyl-Ile conjugate hydrolase activity [GO:1990206]; metal ion binding [GO:0046872]; auxin metabolic process [GO:0009850]; jasmonic acid metabolic process [GO:0009694]; regulation of systemic acquired resistance [GO:0010112]; response to jasmonic acid [GO:0009753]
GMOD_00023487	0.028	FRS1 At4g19990 F18F4.90	nucleus [GO:0005634]; zinc ion binding [GO:0008270]; regulation of DNA-templated transcription [GO:0006355]
GMOD_00023596	0.009	At4g22180 T10I14.10	extracellular region [GO:0005576]; plasma membrane [GO:0005886]; lipid binding [GO:0008289]; lipid transport [GO:0006869]; xylem development [GO:0010089]
GMOD_00023632	0.002	At1g43190 F1I21.14	
GMOD_00023709	0.021		cytoplasm [GO:0005737]; nucleus [GO:0005634]; P-body [GO:0000932]; mRNA binding [GO:0003729]; mRNA processing [GO:0006397]; regulation of RNA splicing

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			[GO:0043484]; regulation of translation [GO:0006417]; RNA splicing [GO:0008380]
GMOD_00023767	0.045	At1g59780 F23H11.10	ADP binding [GO:0043531]; ATP binding [GO:0005524]; defense response [GO:0006952]; response to other organism [GO:0051707]
GMOD_00023827	0.016		
GMOD_00023935	0.036	CAM4 At1g66410 F28G11.13 T27F4.1	calcium ion binding [GO:0005509]
GMOD_00023951	0.004	At3g22710 MWI23.8	
GMOD_00023984	0.002	MYB104 At2g26950 T20P8.20	nucleus [GO:0005634]; DNA binding [GO:0003677]; DNA-binding transcription factor activity [GO:0003700]; regulation of DNA-templated transcription [GO:0006355]
GMOD_00024201	0.008	SBT4.13 At5g59120 MNC17.3	extracellular region [GO:0005576]; serine-type endopeptidase activity [GO:0004252]; proteolysis [GO:0006508]
GMOD_00024289	0.008	ERF042 At2g25820 F17H15.15	nucleus [GO:0005634]; DNA binding [GO:0003677]; DNA-binding transcription factor activity [GO:0003700]; ethylene-activated signaling pathway [GO:0009873]
GMOD_00024467	0.038	PYRD At4g20960 T13K14.120	chloroplast [GO:0009507]; chloroplast stroma [GO:0009570]; diaminohydroxyphosphoribosylaminopyrimidine deaminase activity [GO:0008835]; zinc ion binding [GO:0008270]; riboflavin biosynthetic process [GO:0009231]
GMOD_00024476	0.000	FAD-OXR At4g20860 T13K14.20	extracellular region [GO:0005576]; plant-type cell wall [GO:0009505]; FAD binding [GO:0071949]; oxidoreductase activity [GO:0016491]; cellular response to hypoxia [GO:0071456]; positive regulation of hydrogen peroxide biosynthetic process [GO:0010729]; response to jasmonic acid [GO:0009753]
GMOD_00024485	0.003	RID1 At1g26370 T1K7.25	nucleolus [GO:0005730]; nucleus [GO:0005634]; spliceosomal complex [GO:0005681]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; RNA binding [GO:0003723]; RNA helicase activity [GO:0003724]; mRNA processing [GO:0006397]; positive regulation of

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
			transcription by RNA polymerase I [GO:0045943]; regulation of RNA splicing [GO:0043484]; RNA splicing [GO:0008380]
GMOD_00024714	0.010		
GMOD_00024737	0.037	RLM1B At1g63880 T12P18.10	ADP binding [GO:0043531]; ATP binding [GO:0005524]; NAD(P)+ nucleosidase activity [GO:0050135]; NAD+ nucleotidase, cyclic ADP-ribose generating [GO:0061809]; defense response to fungus [GO:0050832]; signal transduction [GO:0007165]
GMOD_00024899	0.014		
GMOD_00024917	0.000		
GMOD_00024921	0.042		
GMOD_00025019	0.001	RRS1 RCH2 RRS1-R RSH4 SLH1 WRKY52	cytoplasm [GO:0005737]; nucleus [GO:0005634]; ADP binding [GO:0043531]; ATP binding [GO:0005524]; DNA-binding transcription factor activity [GO:0003700]; sequence-specific DNA binding [GO:0043565]; defense response [GO:0006952]; signal transduction [GO:0007165]
GMOD_00025088	0.007	At1g64840 F13O11.14	cytoplasm [GO:0005737]
GMOD_00025103	0.012	TIP4-1 At2g25810 F17H15.16	central vacuole [GO:0042807]; membrane [GO:0016020]; plant-type vacuole membrane [GO:0009705]; water channel activity [GO:0015250]; water transport [GO:0006833]
GMOD_00025157	0.001		
GMOD_00025172	0.039	OSCA1 OSCA1.1 At4g04340 T19B17.6	cytosol [GO:0005829]; plasma membrane [GO:0005886]; calcium activated cation channel activity [GO:0005227]; cation channel activity [GO:0005261]; cellular hyperosmotic response [GO:0071474]; protein tetramerization [GO:0051262]; regulation of calcium ion import [GO:0090279]; response to osmotic stress [GO:0006970]
GMOD_00025449	0.009	ABCG32 PDR4 PEC1 At2g26910 F12C20.5	peroxisome [GO:0005777]; plasma membrane [GO:0005886]; ABC-type transporter activity [GO:0140359]; ATP binding [GO:0005524]; cutin transport [GO:0080051]
GMOD_00025676	0.003	CAND6 At5g02630 T22P11.220	Golgi apparatus [GO:0005794]; membrane [GO:0016020]; G protein-coupled receptor activity [GO:0004930]; G protein-coupled receptor signaling pathway [GO:0007186]
GMOD_00025854	0.006	At5g41510 MBK23.3	

Desc_ID	CODEML Pvalue	Gene Names	Gene Ontology (GO)
GMOD_00025874	0.008		
GMOD_00025990	0.044		metal ion binding [GO:0046872]; metallopeptidase activity [GO:0008237]; proteolysis [GO:0006508]
GMOD_00026045	0.012		
GMOD_00026167	0.047		

Appendix Table C.6: Significantly enriched (FDR < 0.05) GO terms on our list of PSGs. n = number of genes.

Pathway Database	GO Term	Enrichment FDR	n	Pathway Genes	Fold Enrichment	Genes
Biological Pathway	Negative regulation of reproductive process	0.0083	6	68	12.6967	APUM9, MAF1, FRI, DMT1, AGL31
	Negative regulation of multicellular organismal process	0.0138	7	124	8.1232	CRSP, APUM9, MAF1, FRI, MT1, AGL31
	Negative regulation of post-embryonic development	0.0242	6	104	8.3017	CRSP, MAF1, FRI, DMT1, AGL31
	Negative regulation of developmental process	0.0242	7	151	6.6707	CRSP, APUM9, MAF1, FRI, DMT1, AGL31
	Negative regulation of flower development	0.0356	4	42	13.7044	MAF1, FRI, DMT1, AGL31
	Cellular response to stress	0.0356	21	1,265	2.3888	IAN3, ABCG36, BGLU22, WHY2, SULTR1.2, RD29B, LTI45, PK1, ATL56, CLPB4, ARP, PER28, ERCC1, TSO2, RPP1, NPF2.3, OSCA1, MED37D, TGA2, ATJ49
Molecular Function	NAD+ nucleosidase activity	0.020662	6	127	6.798228	CHS1, RLM1B, RLM1A, RPP1, WRKY19, RPS6
	NAD(P)+ nucleosidase activity	0.020662	6	127	6.798228	CHS1, RLM1B, RLM1A, RPP1, WRKY19, RPS6
	NAD+ nucleotidase, cyclic ADP-ribose generating	0.020662	6	127	6.798228	CHS1, RLM1B, RLM1A, RPP1, WRKY19, RPS6
	ADP binding	0.003627	9	200	6.475313	CHS1, RLM1B, RLM1A, RPP1, WRKY19, RRS1, RPS6
	Protein serine kinase activity	0.046527	13	673	2.779563	LRK10L-2.7, RPK1, RBK2, LECRK11, FER, CRK34, RHS16, LRK10L-2.1, GSO2, TDR
	Protein serine/threonine kinase activity	0.028852	16	882	2.610355	LRK10L-2.7, RPK1, BSK12, RBK2, LECRK11, FER, CRK34, WRKY19, RHS16, LRK10L-2.1, GSO2, TDR

Appendix Table C.7: Chi squared pairwise proportion comparisons with Bonferroni adjustment comparing the survival rate for out T-DNA knockout accessions A-C with the control (Col-0) *Arabidopsis* plants.

Null hyp.	Alt. hyp.	diff	p.value	chisq.value	df	2.50%	97.50%
A = Control	A not equal to Control	-0.529	0.001	17.412	1	-0.725	-0.333
B = Control	B not equal to Control	-0.173	0.495	3.012	1	-0.355	0.009
C = Control	C not equal to Control	-0.037	1	0.245	1	-0.176	0.102

Appendix Table C.8: Summary table for a test of significant overlap between PSGs within other Brassicaceae lineages, using the SuperExactTest method. Only those comparisons that presented an overlap are shown. Significant P values are in bold.

Intersections	Degree	Ob	Exp	FE	P	Elements
Descurainia & D_nivalis	2	23	6.9841	3.2932	**0.0000**	AT1G26930, AT1G55020, AT1G64065, AT1G64070, AT1G67720, AT2G23600, AT3G12610, AT3G49450, AT4G02330, AT4G04340, AT4G09580, AT4G14190, AT4G20460, AT4G21050, AT4G26010, AT5G19220, AT5G22670, AT5G39000, AT5G49060, AT5G49160, AT5G53635, AT5G54610, AT5G61480
A_halleri & D_nivalis	2	22	10.2751	2.1411	**0.0007**	AT1G01540, AT1G11990, AT1G25480, AT1G59610, AT1G60770, AT1G80740, AT2G17410, AT2G17790, AT2G24130, AT2G25050, AT2G46920, AT3G15130, AT3G62150, AT4G04980, AT4G08350, AT4G27040, AT4G34040, AT4G34050, AT4G34910, AT4G38950, AT5G49720, AT5G66170
C_lasiocarpa & D_nivalis	2	9	5.0461	1.7836	0.0670	AT1G02340, AT2G17760, AT3G06830, AT3G11240, AT3G24320, AT4G27430, AT5G04430, AT5G11010, AT5G23720
C_lasiocarpa & *Descurainia*	2	3	0.9946	3.0162	0.0782	AT1G25425, AT3G03670, AT4G12020
A_thaliana_LM & D_nivalis	2	2	0.6948	2.8787	0.1519	AT4G27040, AT5G17860
C_himalacia & D_nivalis	2	14	10.4945	1.3340	0.1691	AT1G06900, AT1G53570, AT1G72740, AT2G21370, AT3G47440, AT3G61860, AT3G63400, AT4G25600, AT4G35180, AT5G16020, AT5G52790, AT5G55130, AT5G57300, AT5G66850

Appendix C

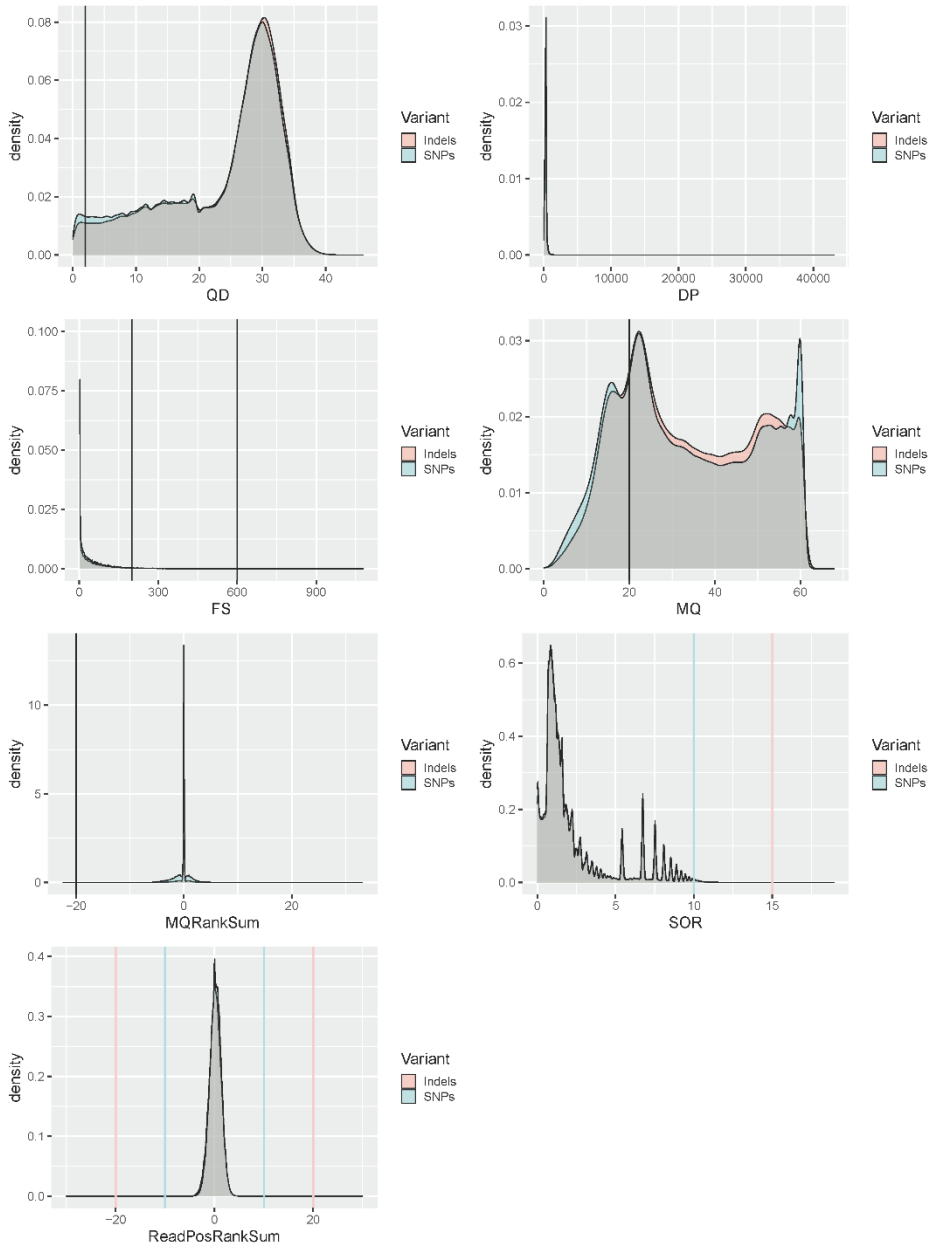
Intersections	Degree	Ob	Exp	FE	P	Elements
A_halleri & A_thaliana_LM	2	1	0.2015	4.9635	0.1834	AT4G27040
C_himalacia & *Descurainia*	2	3	2.0686	1.4503	0.3420	AT1G32770, AT3G03180, AT3G05210
A_halleri & C_himalacia	2	4	3.0433	1.3144	0.3626	AT1G27520, AT2G38823, AT2G44440, AT2G45010
C_himalacia & C_lasiocarpa	2	2	1.4946	1.3382	0.4417	AT3G11220, AT3G59960
A_halleri & *Descurainia*	2	1	2.0253	0.4937	0.8704	AT1G32750

Appendix Table C.9: List of the significant overlap of Biological Process Gene Ontology of PSGs between CI *Descurainia* lineage and the high-altitude alpine *Drava nivalis*.

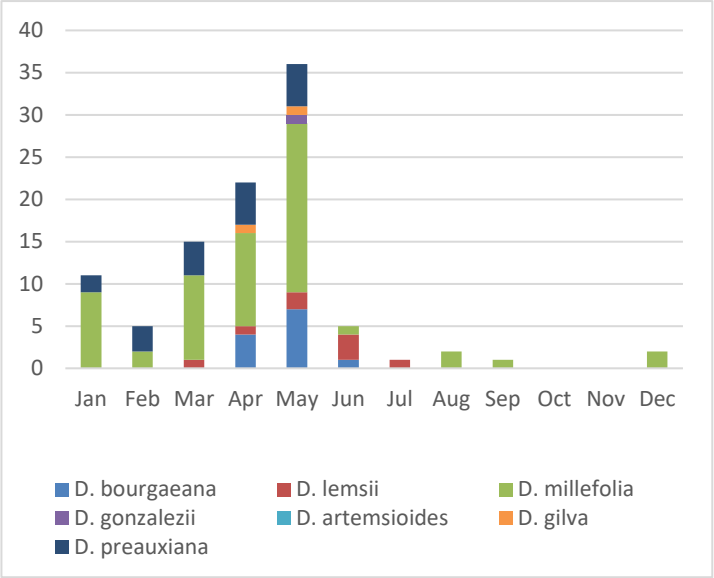
From	Protein names	Gene Names	Gene Ontology (biological process)
AT1G67720	Probable LRR receptor-like serine/threonine-protein kinase At1g67720, EC 2.7.11.1	At1g67720 F12A21.30	protein phosphorylation [GO:0006468]
AT1G64070	Disease resistance protein RML1A, EC 3.2.2.6 (Protein RESISTANCE TO LEPTOSPHAERIA MACULANS 1A)	RLM1A At1g64070 F22C12.17	defense response to fungus [GO:0050832]; signal transduction [GO:0007165]
AT4G14190	Pentatricopeptide repeat-containing protein At4g14190, chloroplastic	At4g14190 dl3135c FCAALL.115	
AT2G23600	Methylesterase 2, AtMES2, EC 3.1.1.- (Protein METHYLESTERASE 8, AtME8)	MES2 ACL ME8 At2g23600 F26B6.25	jasmonic acid metabolic process [GO:0009694]; nicotinate metabolic process [GO:1901847]; oxylipin biosynthetic process [GO:0031408]; salicylic acid metabolic process [GO:0009696]
AT5G49160	DNA (cytosine-5)-methyltransferase 1, EC 2.1.1.37 (DNA methyltransferase 01) (DNA methyltransferase 2) (DNA methyltransferase AthI, DNA Metase AthI, M.AthI) (DNA methyltransferase DDM2) (Protein DECREASED DNA METHYLATION 2)	DMT1 ATHIM DDM2 DMT01 MET1 MET2 At5g49160 K21P3.3	DNA methylation on cytosine within a CG sequence [GO:0010424]; DNA-mediated transformation [GO:0009294]; maintenance of DNA methylation [GO:0010216]; negative regulation of flower development [GO:0009910]; regulation of gene expression by genomic imprinting [GO:0006349]; zygote asymmetric cytokinesis in embryo sac [GO:0010069]
AT5G19220	Glucose-1-phosphate adenylyltransferase large subunit 1, chloroplastic, EC 2.7.7.27 (ADP-glucose pyrophosphorylase) (ADP-glucose synthase) (AGPase S) (Alpha-D-glucose-1-phosphate adenyl transferase)	ADG2 APL1 At5g19220 T24G5.120	glycogen biosynthetic process [GO:0005978]; starch biosynthetic process [GO:0019252]
AT3G12610	DNA damage-repair/toleration protein DRT100	DRT100 At3g12610 MMF12.5 T2E22.8 T2E22_107	DNA repair [GO:0006281]
AT1G55020	Linoleate 9S-lipoxygenase 1, EC 1.13.11.58 (Lipoxygenase 1, AtLOX1)	LOX1 At1g55020 F14C21.3 F14C21.54 T24C10.13	defense response [GO:0006952]; fatty acid biosynthetic process [GO:0006633]; lateral root formation [GO:0010311]; lipid oxidation [GO:0034440]; oxylipin biosynthetic process [GO:0031408]; response to abscisic acid [GO:0009737]; response to jasmonic acid [GO:0009753]; root development [GO:0048364]

From	Protein names	Gene Names	Gene Ontology (biological process)
AT1G64065	Late embryogenesis abundant protein At1g64065	At1g64065	
AT1G26930	F-box/kelch-repeat protein At1g26930	At1g26930 T2P11.12	
AT5G54610	Ankyrin repeat-containing protein BDA1 (Protein BIAN DA 1)	BAD1 At5g54610 MRB17.11	innate immune response [GO:0045087]; response to salicylic acid [GO:0009751]
AT4G09580	Uncharacterized membrane protein At4g09580	At4g09580 T25P22.20	
AT4G02330	Probable pectinesterase/pectinesterase inhibitor 41 [Includes: Pectinesterase inhibitor 41 (Pectin methylesterase inhibitor 41); Pectinesterase 41, PE 41, EC 3.1.1.11 (AtPMEpcrB) (Pectin methylesterase 41, AtPME41)]	PME41 ARATH41 At4g02330 T14P8.14	cell wall modification [GO:0042545]; defense response to fungus [GO:0050832]; pectin catabolic process [GO:0045490]; response to brassinosteroid [GO:0009741]; response to cold [GO:0009409]; response to fungus [GO:0009620]
AT4G26010	Peroxidase 44, Atperox P44, EC 1.11.1.7 (ATP35)	PER44 P44 At4g26010 F20B18.120	hydrogen peroxide catabolic process [GO:0042744]; response to oxidative stress [GO:0006979]
AT5G49060	Chaperone protein dnaJ 49, AtDjC49, AtJ49	ATJ49 C49 At5g49060 K19E20.16 K20J1_3	cellular response to misfolded protein [GO:0071218]; chaperone cofactor-dependent protein refolding [GO:0051085]; ubiquitin-dependent ERAD pathway [GO:0030433]
AT5G39000	Putative receptor-like protein kinase At5g39000, EC 2.7.11.-	At5g39000 MXF12.2	protein autophosphorylation [GO:0046777]; response to metal ion [GO:0010038]
AT5G61480	Leucine-rich repeat receptor-like protein kinase TDR, EC 2.7.11.1 (Protein PHLOEM INTERCALATED WITH XYLEM) (Tracheary element differentiation inhibitory factor receptor, AtTDR, TDIF receptor)	TDR PXY At5g61480 MCI2.4	cell division [GO:0051301]; phloem or xylem histogenesis [GO:0010087]; procambium histogenesis [GO:0010067]; protein phosphorylation [GO:0006468]; secondary shoot formation [GO:0010223]; xylem development [GO:0010089]
AT5G53635	Putative FBD-associated F-box protein At5g53635	At5g53635 MNC6.17	
AT5G22670	Putative F-box/FBD/LRR-repeat protein At5g22670	At5g22670 MDJ22.9	
AT3G49450	F-box protein At3g49450	At3g49450 T1G12.15 T9C5.50	
AT4G21050	Dof zinc finger protein DOF4.4, AtDOF4.4	DOF4.4 At4g21050 T13K14.210	fruit development [GO:0010154]; positive regulation of transcription, DNA-templated [GO:0045893]; regulation of

From	Protein names	Gene Names	Gene Ontology (biological process)
			secondary shoot formation [GO:2000032]; regulation of transcription, DNA-templated [GO:0006355]
AT4G20460	Probable UDP-arabinose 4-epimerase 3, EC 5.1.3.5 (UDP-D-xylose 4-epimerase 3)	At4g20460 F9F13.110	capsule polysaccharide biosynthetic process [GO:0045227]; galactose metabolic process [GO:0006012]; UDP-L-arabinose biosynthetic process [GO:0033358]
AT4G04340	Protein OSCA1 (CSC1-like protein At4g04340) (Hyperosmolality-gated Ca2+ permeable channel 1.1, AtOSCA1.1) (Protein reduced hyperosmolality-induced [Ca(2+)]i increase 1)	OSCA1 OSCA1.1 At4g04340 T19B17.6	cellular hyperosmotic response [GO:0071474]; protein tetramerization [GO:0051262]; regulation of calcium ion import [GO:0090279]; response to osmotic stress [GO:0006970]



Appendix Figure C.1: *DP* - combined depth per SNP across samples (24 samples in the case above) *QD* - variant confidence standardized by depth. *MQ* - Mapping quality of a SNP. *FS* - strand bias in support for REF vs ALT allele calls *SOR* - sequencing bias in which one DNA strand is favoured over the other *MQRankSum* - Rank sum test for mapping qualities of REF vs. ALT reads. *ReadPosRankSum* - do all the reads support a SNP call tend to be near the end of a read



Appendix Figure C.2: Distribution of flowering herbarium specimens (deposited within the Natural History Museum, UK) of the seven Canary Island *Descurainia* species, according to the month of collection.

Glossary of Terms

contig	Series of overlapping DNA sequences
endemic	Unique to a given locality
hybridisation	Two different species breeding
Incomplete lineage sorting	Ancestral genes that have different divergence time to that of between species
introgression	The transfer of genetic material between species as a result of hybridisation
k-mer	The substrings of lengths within a DNA sequence
L50	The smallest number of contigs whose length sum makes up half of genome size
N50	Length of the median contig within a genome assembly
phylogenetics	The study of evolutionary relatedness among various organisms
scaffold	A reconstructed DNA sequences created by chaining contigs together
speciation	The process by which a population become differentiated, genetically and morphological, from its parent population

List of References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A.W., Parisod, C., Pfennig, K., Rice, A.M., Ritchie, M.G., Seifert, B., Smadja, C.M., Stelkens, R., Szymura, J.M., Väinölä, R., Wolf, J.B.W., Zinner, D., 2013. Hybridization and speciation. Journal of Evolutionary Biology 26, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Al Khateeb, W.M., Sher, A.A., Marcus, J.M., Schroeder, D.F., 2019. UVSSA, UBP12, and RDO2/TFIIS Contribute to Arabidopsis UV Tolerance. Frontiers in Plant Science 10.

Albaladejo, R.G., Martín-Hernanz, S., Reyes-Betancort, J.A., Santos-Guerra, A., Olangua-Corral, M., Aparicio, A., 2021. Reconstruction of the spatio-temporal diversification and ecological niche evolution of Helianthemum (Cistaceae) in the Canary Islands using genotyping-by-sequencing data. Annals of Botany 127, 597–611. https://doi.org/10.1093/aob/mcaa090

Albani, M.C., Castaings, L., Wötzel, S., Mateos, J.L., Wunder, J., Wang, R., Reymond, M., Coupland, G., 2012. PEP1 of Arabis alpina Is Encoded by Two Overlapping Genes That Contribute to Natural Genetic Variation in Perennial Flowering. PLOS Genetics 8, e1003130. https://doi.org/10.1371/journal.pgen.1003130

Alberto, F.J., Derory, J., Boury, C., Frigerio, J.-M., Zimmermann, N.E., Kremer, A., 2013. Imprints of Natural Selection Along Environmental Gradients in Phenology-Related Genes of Quercus petraea. Genetics 195, 495–512. https://doi.org/10.1534/genetics.113.153783

Allan, G.J., Francisco-Ortega, J., Santos-Guerra, A., Boerner, E., Zimmer, E.A., 2004. Molecular phylogenetic evidence for the geographic origin and classification of Canary Island Lotus (Fabaceae: Loteae). Molecular Phylogenetics and Evolution 32, 123–138. https://doi.org/10.1016/j.ympev.2003.11.018

Allio, R., Nabholz, B., Wanke, S., Chomicki, G., Pérez-Escobar, O.A., Cotton, A.M., Clamens, A.-L., Kergoat, G.J., Sperling, F.A.H., Condamine, F.L., 2021. Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. Nat Commun 12, 354. https://doi.org/10.1038/s41467-020-20507-3

Almén, M.S., Lamichhaney, S., Berglund, J., Grant, B.R., Grant, P.R., Webster, M.T., Andersson, L., 2016. Adaptive radiation of Darwin's finches revisited using whole genome sequencing. BioEssays 38, 14–20. https://doi.org/10.1002/bies.201500079

Al-Shehbaz, I.A., Beilstein, M.A., Kellogg, E.A., 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst. Evol. 259, 89–120. https://doi.org/10.1007/s00606-006-0415-z

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Alvim, F.C., Carolino, S.M.B., Cascardo, J.C.M., Nunes, C.C., Martinez, C.A., Otoni, W.C., Fontes, E.P.B., 2001. Enhanced Accumulation of BiP in Transgenic Plants Confers Tolerance to Water Stress. Plant Physiology 126, 1042–1054. https://doi.org/10.1104/pp.126.3.1042

References

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biology 21, 30. https://doi.org/10.1186/s13059-020-1935-5

Anderson, G.J., Pérez De Paz, J., Anderson, M., Bernardello, G., Taylor, D.W., 2021. A leaky dimorphic sexual system and breeding system characterize a successful island colonist: the reproductive biology of Plocama pendula (Rubiaceae). Botanical Journal of the Linnean Society 196, 540–555. https://doi.org/10.1093/botlinnean/boab026

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17, 81–92. https://doi.org/10.1038/nrg.2015.28

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, Å., Lindblad-Toh, K., 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495, 360–364. https://doi.org/10.1038/nature11837

Ayala, F.J., Tracey, M.L., Hedgecock, D., Richmond, R.C., 1974. Genetic Differentiation During the Speciation Process in Drosophila. Evolution 28, 576–592. https://doi.org/10.2307/2407283

Azad, P., Stobdan, T., Zhou, D., Hartley, I., Akbari, A., Bafna, V., Haddad, G.G., 2017. High-altitude adaptation in humans: from genomics to integrative physiology. J Mol Med 95, 1269–1282. https://doi.org/10.1007/s00109-017-1584-7

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE 3, e3376. https://doi.org/10.1371/journal.pone.0003376

Baldwin, B., 1997. ADAPTIVE RADIATION OF THE HAWAIIAN SILVERSWORD ALLIANCE: CONGRUENCE AND CONFLICT OF PHYLOGENETIC EVIDENCE FROM MOLECULAR AND NON-MOLECULAR INVESTIGATIONS. pp. 103–128.

Baldwin, B.G., Crawford, D.J., Francisco-Ortega, J., Kim, S.-C., Sang, T., Stuessy, T.F., 1998. Molecular Phylogenetic Insights on the Origin and Evolution of Oceanic Island Plants, in: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II: DNA Sequencing. Springer US, Boston, MA, pp. 410–441. https://doi.org/10.1007/978-1-4615-5419-6_14

Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proceedings of the National Academy of Sciences 95, 9402–9406. https://doi.org/10.1073/pnas.95.16.9402

Ballard, J.W.O., Rand, D.M., 2005. The Population Biology of Mitochondrial DNA and Its Phylogenetic Implications. Annual Review of Ecology, Evolution, and Systematics 36, 621–642.

Barber, J.C., Finch, C.C., Francisco-Ortega, J., Santos-Guerra, A., Jansen, R.K., 2007. Hybridization in Macaronesian Sideritis (Lamiaceae): evidence from incongruence of multiple

independent nuclear and chloroplast sequence datasets. TAXON 56, 74–88. https://doi.org/10.2307/25065737

Barber, J.C., Francisco-Ortega, J., Santos-Guerra, A., Turner, K.G., Jansen, R.K., 2002. Origin of Macaronesian Sideritis L. (Lamioideae: Lamiaceae) inferred from nuclear and chloroplast sequence datasets. Molecular Phylogenetics and Evolution 23, 293–306. https://doi.org/10.1016/S1055-7903(02)00018-0

Barber, J.C., Ortega, J.F., Santos-Guerra, A., Marrero, A., Jansen, R.K., 2000. Evolution of Endemic Sideritis (Lamiaceae) in Macaronesia: Insights from a Chloroplast DNA Restriction Site Analysis. sbot 25, 633–647. https://doi.org/10.2307/2666725

Barley, A.J., Nieto-Montes de Oca, A., Manríquez-Morán, N.L., Thomson, R.C., 2022. The evolutionary network of whiptail lizards reveals predictable outcomes of hybridization. Science 377, 773–777. https://doi.org/10.1126/science.abn1593

Barluenga, M., Stölting, K.N., Salzburger, W., Muschick, M., Meyer, A., 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439, 719–723. https://doi.org/10.1038/nature04325

Barton, N.H., Hewitt, G.M., 1985. Analysis of Hybrid Zones. Annual Review of Ecology and Systematics 16, 113–148.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. Genetics 162, 2025–2035. https://doi.org/10.1093/genetics/162.4.2025

Beilstein, M.A., Al-Shehbaz, I.A., Kellogg, E.A., 2006. Brassicaceae phylogeny and trichome evolution. American Journal of Botany 93, 607–619. https://doi.org/10.3732/ajb.93.4.607

Beilstein, M.A., Al-Shehbaz, I.A., Mathews, S., Kellogg, E.A., 2008. Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. American Journal of Botany 95, 1307–1327. https://doi.org/10.3732/ajb.0800065

Bellemain, E., Ricklefs, R.E., 2008. Are islands the end of the colonization road? Trends in Ecology & Evolution 23, 461–468. https://doi.org/10.1016/j.tree.2008.05.001

Bellinger, M.R., Datlof, E.M., Selph, K.E., Gallaher, T.J., Knope, M.L., 2022. A Genome for Bidens hawaiensis: A Member of a Hexaploid Hawaiian Plant Adaptive Radiation. Journal of Heredity 113, 205–214. https://doi.org/10.1093/jhered/esab077

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27, 573–580. https://doi.org/10.1093/nar/27.2.573

Bentley, D.R., 2006. Whole-genome re-sequencing. Current Opinion in Genetics & Development, Genomes and evolution 16, 545–552. https://doi.org/10.1016/j.gde.2006.10.009

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E., 2015. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. genesis 53, 474–485. https://doi.org/10.1002/dvg.22877

Beric, A., Mabry, M.E., Harkess, A.E., Brose, J., Schranz, M.E., Conant, G.C., Edger, P.P., Meyers, B.C., Pires, J.C., 2021. Comparative phylogenetics of repetitive elements in a diverse order

of flowering plants (Brassicales). G3 Genes|Genomes|Genetics 11, jkab140. https://doi.org/10.1093/g3journal/jkab140

Berner, D., Salzburger, W., 2015. The genomics of organismal diversification illuminated by adaptive radiations. Trends in Genetics 31, 491–499. https://doi.org/10.1016/j.tig.2015.07.002

Bihmidine, S., Lin, J., Stone, J.M., Awada, T., Specht, J.E., Clemente, T.E., 2013. Activity of the Arabidopsis RD29A and RD29B promoter elements in soybean under water stress. Planta 237, 55–64. https://doi.org/10.1007/s00425-012-1740-9

Bikard, D., Patel, D., Le Metté, C., Giorgi, V., Camilleri, C., Bennett, M.J., Loudet, O., 2009. Divergent Evolution of Duplicate Genes Leads to Genetic Incompatibilities Within A. thaliana. Science 323, 623–626. https://doi.org/10.1126/science.1165917

Blanco-Pastor, J.L., Bertrand, Y.J.K., Liberal, I.M., Wei, Y., Brummer, E.C., Pfeil, B.E., 2019. Evolutionary networks from RADseq loci point to hybrid origins of Medicago carstiensis and Medicago cretacea. American Journal of Botany 106, 1219–1228. https://doi.org/10.1002/ajb2.1352

Bliss, L.C., 1962. Adaptations of Arctic and Alpine Plants to Environmental Conditions. Arctic 15, 117–144.

Blomberg, S.P., Garland JR., T., Ives, A.R., 2003. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. Evolution 57, 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Blonder, B., Baldwin, B.G., Enquist, B.J., Robichaux, R.H., 2016. Variation and macroevolution in leaf functional traits in the Hawaiian silversword alliance (Asteraceae). Journal of Ecology 104, 219–228. https://doi.org/10.1111/1365-2745.12497

Blumthaler, M., Ambach, W., Ellinger, R., 1997. Increase in solar UV radiation with altitude. Journal of Photochemistry and Photobiology B: Biology 39, 130–134. https://doi.org/10.1016/S1011-1344(96)00018-8

Boetzer, M., Pirovano, W., 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15, 211. https://doi.org/10.1186/1471-2105-15-211

Böhle, U.R., Hilger, H.H., Martin, W.F., 1996. Island colonization and evolution of the insular woody habit in Echium L. (Boraginaceae). Proceedings of the National Academy of Sciences 93, 11740–11745. https://doi.org/10.1073/pnas.93.21.11740

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., Slotte, T., Kolář, F., 2021. Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. Proceedings of the National Academy of Sciences 118, e2022713118. https://doi.org/10.1073/pnas.2022713118

Bolan, N.S., 1991. A critical review on the role of mycorrhizal fungi in the uptake of phosphorus by plants. Plant Soil 134, 189–207. https://doi.org/10.1007/BF00012037

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowles, D. J., Lillford, P.J., Rees, D.A., Shanks, I.A., Smallwood, M., Bowles, Dianna J., 2002. Plants in a cold climate. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357, 831–847. https://doi.org/10.1098/rstb.2002.1073

Bramwell, 1977. A revision of Descurainia Webb and Berth. section Sisymbriodendron (Christ) O.E. Schulz in the Canary Islands [Cruciferae]. Botanica Macaronesica.

Bramwell, D., 1990. Conserving Biodiversity in the Canary Islands. Annals of the Missouri Botanical Garden 77, 28–37. https://doi.org/10.2307/2399622

Bramwell, D., 1976. The Endemic Flora of the Canary Islands; Distribution, Relationships and Phytogeography, in: Kunkel, G. (Ed.), Biogeography and Ecology in the Canary Islands, Monographiae Biologicae. Springer Netherlands, Dordrecht, pp. 207–240. https://doi.org/10.1007/978-94-010-1566-0_6

Bramwell, D., Caujapé-Castells, J., 2011. The Biology of Island Floras. Cambridge University Press.

Brandvain, Y., Kenney, A.M., Flagel, L., Coop, G., Sweigart, A.L., 2014. Speciation and Introgression between Mimulus nasutus and Mimulus guttatus. PLOS Genetics 10, e1004410. https://doi.org/10.1371/journal.pgen.1004410

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y.V., Ramsey, J.M., Riehn, R., Soni, G.V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J.A., 2008. The potential and challenges of nanopore sequencing. Nat Biotechnol 26, 1146–1153. https://doi.org/10.1038/nbt.1495

Bray, E.A., 2004. Genes commonly regulated by water-deficit stress in Arabidopsis thaliana. Journal of Experimental Botany 55, 2331–2341. https://doi.org/10.1093/jxb/erh270

Brilhante, M., Roxo, G., Catarino, S., dos Santos, P., Reyes-Betancort, J.A., Caujapé-Castells, J., Sequeira, M.M., Talhinhas, P., Romeiras, M.M., 2021. Diversification of Aeonium Species Across Macaronesian Archipelagos: Correlations Between Genome-Size Variation and Their Conservation Status. Frontiers in Ecology and Evolution 9.

Briskine, R.V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., Shimizu, K.K., 2017. Genome assembly and annotation of Arabidopsis halleri, a model for heavy metal hyperaccumulation and evolutionary ecology. Molecular Ecology Resources 17, 1025–1036. https://doi.org/10.1111/1755-0998.12604

Brito, P., Jimenez, M.S., Morales, D., Wieser, G., 2013. Assessment of ecosystem CO2 efflux and its components in a Pinus canariensis forest at the treeline. Trees 27, 999–1009. https://doi.org/10.1007/s00468-013-0851-7

broadinstitute/picard, 2020a. . Broad Institute.

broadinstitute/picard, 2020b.

Brochmann, C., Borgen, L., Stabbetorp, O.E., 2000. Multiple diploid hybrid speciation of the Canary Island endemicArgyranthemum sundingii (Asteraceae). Pl Syst Evol 220, 77–92. https://doi.org/10.1007/BF00985372

Brown, R.M., Siler, C.D., Oliveros, C.H., Esselstyn, J.A., Diesmos, A.C., Hosner, P.A., Linkem, C.W., Barley, A.J., Oaks, J.R., Sanguila, M.B., Welton, L.J., Blackburn, D.C., Moyle, R.G., Townsend Peterson, A., Alcala, A.C., 2013. Evolutionary Processes of Diversification in a Model Island Archipelago. Annual Review of Ecology, Evolution, and Systematics 44, 411–435. https://doi.org/10.1146/annurev-ecolsys-110411-160323

Burke, M.K., Dunham, J.P., Shahrestani, P., Thornton, K.R., Rose, M.R., Long, A.D., 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467, 587–590. https://doi.org/10.1038/nature09352

References

Burns, K.C., 2019. Evolution in Isolation: The Search for an Island Syndrome in Plants. Cambridge University Press.

Butlin, R.K., Galindo, J., Grahame, J.W., 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philosophical Transactions of the Royal Society B: Biological Sciences 363, 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Caicedo, A.L., Stinchcombe, J.R., Olsen, K.M., Schmitt, J., Purugganan, M.D., 2004. Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proceedings of the National Academy of Sciences 101, 15670–15675. https://doi.org/10.1073/pnas.0406232101

Campbell, C.R., Poelstra, J.W., Yoder, A.D., 2018. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biological Journal of the Linnean Society 124, 561–583. https://doi.org/10.1093/biolinnean/bly063

Campbell, M.S., Holt, C., Moore, B., Yandell, M., 2014. Genome Annotation and Curation Using MAKER and MAKER-P. Current Protocols in Bioinformatics 48, 4.11.1-4.11.39. https://doi.org/10.1002/0471250953.bi0411s48

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., Yandell, M., 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18, 188–196. https://doi.org/10.1101/gr.6743907

Carine, M.A., Robba, L., Little, R., Russell, S., Guerra, A.S., 2007. Molecular and morphological evidence for hybridization between endemic Canary Island Convolvulus. Botanical Journal of the Linnean Society 154, 187–204. https://doi.org/10.1111/j.1095-8339.2007.00657.x

Carine, M.A., Russell, S.J., Santos-Guerra, A., Francisco-Ortega, J., 2004. Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in Convolvulus (Convolvulaceae). American Journal of Botany 91, 1070–1085. https://doi.org/10.3732/ajb.91.7.1070

Carlquist, S., 1966. The Biota of Long-Distance Dispersal. IV. Genetic Systems in the Floras of Oceanic Islands. Evolution 20, 433–455. https://doi.org/10.2307/2406583

Caujapé-Castells, J., García-Verdugo, C., Marrero-Rodríguez, Á., Fernández-Palacios, J.M., Crawford, D.J., Mort, M.E., 2017. Island ontogenies, syngameons, and the origins and evolution of genetic diversity in the Canarian endemic flora. Perspectives in Plant Ecology, Evolution and Systematics 27, 9–22. https://doi.org/10.1016/j.ppees.2017.03.003

Cerca, J., Petersen, B., Guevara, J.M.L., Rivera-Colón, A., Birkeland, S., Vizueta, J., Li, S., Loureiro, J., Kosawang, C., Díaz, P.J., Rivas-Torres, G., Fernández-Mazuecos, M., Vargas, P., McCauley, R., Petersen, G., Santos-Bay, L., Wales, N., Catchen, J., Machado, D., Nowak, M.D., Suh, A., Sinha, N., Nielsen, L.R., Seberg, O., Gilbert, M.T.P., Leebens-Mack, J.H., Rieseberg, L., Martin, M.D., 2022. The genomic basis of the plant island syndrome in Darwin's giant daisies. https://doi.org/10.1101/2022.01.26.477903

Chapman, M.A., Hiscock, S.J., Filatov, D.A., 2013. Genomic Divergence during Speciation Driven by Adaptation to Altitude. Mol Biol Evol 30, 2553–2567. https://doi.org/10.1093/molbev/mst168

Chaves, J.A., Cooper, E.A., Hendry, A.P., Podos, J., De León, L.F., Raeymaekers, J.A.M., MacMillan, W.O., Uy, J.A.C., 2016. Genomic variation at the tips of the adaptive radiation of Darwin's finches. Molecular Ecology 25, 5282–5295. https://doi.org/10.1111/mec.13743

Check Hayden, E., 2009. Genome sequencing: the third generation. Nature. https://doi.org/10.1038/news.2009.86

Chen, H., Wang, T., He, X., Cai, X., Lin, R., Liang, J., Wu, J., King, G., Wang, X., 2022. BRAD V3.0: an upgraded Brassicaceae database. Nucleic Acids Research 50, D1432–D1441. https://doi.org/10.1093/nar/gkab1057

Chen, J., Huang, Y., Brachi, B., Yun, Q., Zhang, W., Lu, W., Li, H., Li, W., Sun, X., Wang, G., He, J., Zhou, Z., Chen, K., Ji, Y., Shi, M., Sun, W., Yang, Y., Zhang, R., Abbott, R.J., Sun, H., 2019. Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. Nat Commun 10, 5230. https://doi.org/10.1038/s41467-019-13128-y

Chen, T., Cohen, D., Itkin, M., Malitsky, S., Fluhr, R., 2021. Lipoxygenase functions in 1O2 production during root responses to osmotic stress. Plant Physiology 185, 1638–1651. https://doi.org/10.1093/plphys/kiab025

Cherry, J.F., 1981. Pattern and Process in the Earliest Colonization of the Mediterranean Islands. Proceedings of the Prehistoric Society 47, 41–68. https://doi.org/10.1017/S0079497X00008859

Cheviron, Z.A., Brumfield, R.T., 2012. Genomic insights into adaptation to high-altitude environments. Heredity 108, 354–361. https://doi.org/10.1038/hdy.2011.85

Chikhi, R., Medvedev, P., 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30, 31–37. https://doi.org/10.1093/bioinformatics/btt310

Chiou, K.L., Janiak, M.C., Schneider-Crease, I.A., Sen, S., Ayele, F., Chuma, I.S., Knauf, S., Lemma, A., Signore, A.V., D'Ippolito, A.M., Abebe, B., Haile, A.A., Kebede, F., Fashing, P.J., Nguyen, N., McCann, C., Houck, M.L., Wall, J.D., Burrell, A.S., Bergey, C.M., Rogers, J., Phillips-Conroy, J.E., Jolly, C.J., Melin, A.D., Storz, J.F., Lu, A., Beehner, J.C., Bergman, T.J., Snyder-Mackler, N., 2022. Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. Nat Ecol Evol 6, 630–643. https://doi.org/10.1038/s41559-022-01703-4

Choi, J.Y., Dai, X., Alam, O., Peng, J.Z., Rughani, P., Hickey, S., Harrington, E., Juul, S., Ayroles, J.F., Purugganan, M.D., Stacy, E.A., 2021. Ancestral polymorphisms shape the adaptive radiation of Metrosideros across the Hawaiian Islands. Proceedings of the National Academy of Sciences 118, e2023801118. https://doi.org/10.1073/pnas.2023801118

Choisy, M., Franck, P., Cornuet, J.-M., 2004. Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. Molecular Ecology 13, 955–968. https://doi.org/10.1111/j.1365-294X.2004.02107.x

Chung, Y.S., Choi, S.C., Jun, T.-H., Kim, C., 2017. Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. Hortic. Environ. Biotechnol. 58, 425–431. https://doi.org/10.1007/s13580-017-0297-8

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92. https://doi.org/10.4161/fly.19695

Claros, M.G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P., Fernández-Pozo, N., 2012. Why Assembling Plant Genome Sequences Is So Challenging. Biology 1, 439–459. https://doi.org/10.3390/biology1020439

References

Climent, J., Tapias, R., Pardos, J.A., Gil, L., 2004. Fire adaptations in the Canary Islands pine (Pinus canariensis). Plant Ecology 171, 185–196. https://doi.org/10.1023/B:VEGE.0000029374.64778.68

Conte, G.L., Arnegard, M.E., Peichel, C.L., Schluter, D., 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B: Biological Sciences 279, 5039–5047. https://doi.org/10.1098/rspb.2012.2146

Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., Estoup, A., 2008a. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24, 2713–2719. https://doi.org/10.1093/bioinformatics/btn514

Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., Estoup, A., 2008b. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24, 2713–2719. https://doi.org/10.1093/bioinformatics/btn514

Corral, M., Cáceres, J., 2006. Contribucción a la corología de Descurainia artemisioides Svent. (Brassicaceae). Botánica Macaronésica, ISSN 0211-7150, Nº. 26, 2006, pags. 85-96.

Coyne, J.A., Orr, H.A., 1989. Patterns of Speciation in Drosophila. Evolution 43, 362–381. https://doi.org/10.1111/j.1558-5646.1989.tb04233.x

Coyne, J.A., Orr, H.A., Coyne, J.A., Orr, H.A., 2004. Speciation. Oxford University Press, Oxford, New York.

Crawford, D.J., Anderson, G.J., Silva, L.B., de Sequeira, M.M., Moura, M., Santos-Guerra, A., Kelly, J.K., Mort, M.E., 2015. Breeding systems in Tolpis (Asteraceae) in the Macaronesian islands: the Azores, Madeira and the Canaries. Plant Syst Evol 301, 1981–1993. https://doi.org/10.1007/s00606-015-1210-5

Crawford, D.J., Stuessy, T.F., 2016. Cryptic variation, molecular data, and the challenge of conserving plant diversity in oceanic archipelagos: the critical role of plant systematics. Korean Journal of Plant Taxonomy 46, 129–148. https://doi.org/10.11110/kjpt.2016.46.2.129

Crawford, D.J., Stuessy, T.F., 1997. Plant Speciation on Oceanic Islands, in: Iwatsuki, K., Raven, P.H. (Eds.), Evolution and Diversification of Land Plants. Springer Japan, Tokyo, pp. 249–267. https://doi.org/10.1007/978-4-431-65918-1_12

Creff, A., Sormani, R., Desnos, T., 2010. The two Arabidopsis RPS6 genes, encoding for cytoplasmic ribosomal proteins S6, are functionally equivalent. Plant Mol Biol 73, 533–546. https://doi.org/10.1007/s11103-010-9639-y

Cropper, T.E., Hanna, E., 2014. An analysis of the climate of Macaronesia, 1865–2012. International Journal of Climatology 34, 604–622. https://doi.org/10.1002/joc.3710

Csilléry, K., Lalagüe, H., Vendramin, G.G., González-Martínez, S.C., Fady, B., Oddou-Muratorio, S., 2014. Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (Fagus sylvatica) populations. Molecular Ecology 23, 4696–4708. https://doi.org/10.1111/mec.12902

Cumer, T., Machado, A.P., Siverio, F., Cherkaoui, S.I., Roque, I., Lourenço, R., Charter, M., Roulin, A., Goudet, J., 2022. Genomic basis of insularity and ecological divergence in barn owls (Tyto alba) of the Canary Islands. Heredity 129, 281–294. https://doi.org/10.1038/s41437-022-00562-w

Curto, M., Schachtler, C., Puppo, P., Meimberg, H., 2018. Using a new RAD-sequencing approach to study the evolution of Micromeria in the Canary islands. Molecular Phylogenetics and Evolution 119, 160–169. https://doi.org/10.1016/j.ympev.2017.11.005

da Fonseca, R.R., Albrechtsen, A., Themudo, G.E., Ramos-Madrigal, J., Sibbesen, J.A., Maretty, L., Zepeda-Mendoza, M.L., Campos, P.F., Heller, R., Pereira, R.J., 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Marine Genomics 30, 3–13. https://doi.org/10.1016/j.margen.2016.04.012

Dai, C., Lee, Y., Lee, I.C., Nam, H.G., Kwak, J.M., 2018. Calmodulin 1 Regulates Senescence and ABA Response in Arabidopsis. Frontiers in Plant Science 9.

Dai, G.-Y., Chen, D.-K., Sun, Y.-P., Liang, W.-Y., Liu, Y., Huang, L.-Q., Li, Y.-K., He, J.-F., Yao, N., 2020. The Arabidopsis KH-domain protein FLOWERING LOCUS Y delays flowering by upregulating FLOWERING LOCUS C family members. Plant Cell Rep 39, 1705–1717. https://doi.org/10.1007/s00299-020-02598-w

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DARRIBA, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution 37, 291–294. https://doi.org/10.1093/molbev/msz189

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2019. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. bioRxiv 612903. https://doi.org/10.1101/612903

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12, 499–510. https://doi.org/10.1038/nrg3012

De La Torre, A.R., Wilhite, B., Puiu, D., St. Clair, J.B., Crepeau, M.W., Salzberg, S.L., Langley, C.H., Allen, B., Neale, D.B., 2021. Dissecting the Polygenic Basis of Cold Adaptation Using Genome-Wide Association of Traits and Environmental Data in Douglas-fir. Genes 12, 110. https://doi.org/10.3390/genes12010110

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24, 332–340. https://doi.org/10.1016/j.tree.2009.01.009

del Arco Aguilar, M.J., Rodríguez Delgado, O., 2018. Vegetation of the Canary Islands, in: del Arco Aguilar, M.J., Rodríguez Delgado, O. (Eds.), Vegetation of the Canary Islands, Plant and Vegetation. Springer International Publishing, Cham, pp. 83–319. https://doi.org/10.1007/978-3-319-77255-4_6

del-Arco, M., Pérez-de-Paz, P.L., Acebes, J.R., González-Mancebo, J.M., Reyes-Betancort, J.A., Bermejo, J.A., de-Armas, S., González-González, R., 2006. Bioclimatology and climatophilous vegetation of Tenerife (Canary Islands). Annales Botanici Fennici 43, 167–192.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dodsworth, S., Pokorny, L., Johnson, M.G., Kim, J.T., Maurin, O., Wickett, N.J., Forest, F., Baker, W.J., 2019. Hyb-Seq for Flowering Plant Systematics. Trends in Plant Science 24, 887–891. https://doi.org/10.1016/j.tplants.2019.07.011

Doebeli, M., Dieckmann, U., Metz, J.A.J., Tautz, D., 2005. What We Have Also Learned: Adaptive Spectiation Is Theoretically Plausible. Evolution 59, 691–695. https://doi.org/10.1111/j.0014-3820.2005.tb01028.x

dos Reis, M., Yang, Z., 2019. Bayesian Molecular Clock Dating Using Genome-Scale Datasets, in: Anisimova, M. (Ed.), Evolutionary Genomics: Statistical and Computational Methods, Methods in Molecular Biology. Springer, New York, NY, pp. 309–330. https://doi.org/10.1007/978-1-4939-9074-0_10

Doyle, J., 1991. DNA Protocols for Plants, in: Hewitt, G.M., Johnston, A.W.B., Young, J.P.W. (Eds.), Molecular Techniques in Taxonomy, NATO ASI Series. Springer, Berlin, Heidelberg, pp. 283–293. https://doi.org/10.1007/978-3-642-83962-7_18

Dubest, S., Gallego, M.E., White, C.I., 2004. Roles of the AtErcc1 protein in recombination. The Plant Journal 39, 334–342. https://doi.org/10.1111/j.1365-313X.2004.02136.x

Dunning, L.T., Hipperson, H., Baker, W.J., Butlin, R.K., Devaux, C., Hutton, I., Igea, J., Papadopulos, A.S.T., Quan, X., Smadja, C.M., Turnbull, C.G.N., Savolainen, V., 2016. Ecological speciation in sympatric palms: 1. Gene expression, selection and pleiotropy. Journal of Evolutionary Biology 29, 1472–1487. https://doi.org/10.1111/jeb.12895

Dupont, Y.L., Hansen, D.M., Olesen, J.M., 2003. Structure of a plant–flower-visitor network in the high-altitude sub-alpine desert of Tenerife, Canary Islands. Ecography 26, 301–310. https://doi.org/10.1034/j.1600-0587.2003.03443.x

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for Ancient Admixture between Closely Related Populations. Molecular Biology and Evolution 28, 2239–2252. https://doi.org/10.1093/molbev/msr048

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Eaton, D.A.R., Overcast, I., 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36, 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Eaton, D.A.R., Ree, R.H., 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Systematic Biology 62, 689–706. https://doi.org/10.1093/sysbio/syt032

Eilbeck, K., Moore, B., Holt, C., Yandell, M., 2009. Quantitative measures for the management and comparison of annotated genomes. BMC Bioinformatics 10, 67. https://doi.org/10.1186/1471-2105-10-67

Ekblom, R., Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107, 1–15. https://doi.org/10.1038/hdy.2010.152

Ellstrand, N.C., Whitkus, R., Rieseberg, L.H., 1996. Distribution of spontaneous plant hybrids. Proceedings of the National Academy of Sciences 93, 5090–5093. https://doi.org/10.1073/pnas.93.10.5090

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6, e19379. https://doi.org/10.1371/journal.pone.0019379

Emerson, B.C., 2002. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. Mol Ecol 11, 951–966. https://doi.org/10.1046/j.1365-294x.2002.01507.x

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20, 238. https://doi.org/10.1186/s13059-019-1832-y

Emms, D.M., Kelly, S., 2018. STAG: Species Tree Inference from All Genes. https://doi.org/10.1101/267914

Emms, D.M., Kelly, S., 2017. STRIDE: Species Tree Root Inference from Gene Duplication Events. Molecular Biology and Evolution 34, 3267–3278. https://doi.org/10.1093/molbev/msx259

Escuer, P., Pisarenco, V.A., Fernández-Ruiz, A.A., Vizueta, J., Sánchez-Herrero, J.F., Arnedo, M.A., Sánchez-Gracia, A., Rozas, J., 2022. The chromosome-scale assembly of the Canary Islands endemic spider Dysdera silvatica (Arachnida, Araneae) sheds light on the origin and genome structure of chemoreceptor gene families in chelicerates. Molecular Ecology Resources 22, 375–390. https://doi.org/10.1111/1755-0998.13471

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., Merilä, J., 2020. On the causes of geographically heterogeneous parallel evolution in sticklebacks. Nat Ecol Evol 4, 1105–1115. https://doi.org/10.1038/s41559-020-1222-6

Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends in Genetics 28, 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feng, L., Lin, H., Kang, M., Ren, Y., Yu, X., Xu, Z., Wang, S., Li, T., Yang, W., Hu, Q., 2022. A chromosome-level genome assembly of an alpine plant Crucihimalaya lasiocarpa provides insights into high-altitude adaptation. DNA Research 29, dsac004. https://doi.org/10.1093/dnares/dsac004

Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., Glover, B.J., 2018. Resolving Recent Plant Radiations: Power and Robustness of Genotyping-by-Sequencing. Systematic Biology 67, 250–268. https://doi.org/10.1093/sysbio/syx062

Fernández-Palacios, J.M., 1992. Climatic responses of plant species on Tenerife, The Canary Islands. Journal of Vegetation Science 3, 595–603. https://doi.org/10.2307/3235826

Fernández-Palacios, J.M., de Nicolás, J.P., 1995. Altitudinal pattern of vegetation variation on Tenerife. Journal of Vegetation Science 6, 183–190. https://doi.org/10.2307/3236213

Fernández-Palacios, J.M., Otto, R., Borregaard, M.K., Kreft, H., Price, J.P., Steinbauer, M.J., Weigelt, P., Whittaker, R.J., 2021. Evolutionary winners are ecological losers among oceanic island plants. Journal of Biogeography 48, 2186–2198. https://doi.org/10.1111/jbi.14143

Fernández-Palacios, J.M., Otto, R., Thebaud, C., Price, J., 2014. Overview of Habitat History in Subtropical Oceanic Island Summit Ecosystems. Arctic, Antarctic, and Alpine Research 46, 801–809. https://doi.org/10.1657/1938-4246-46.4.801

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. https://doi.org/10.1002/joc.5086

Filleur, S., Dorbe, M.-F., Cerezo, M., Orsel, M., Granier, F., Gojon, A., Daniel-Vedele, F., 2001. An Arabidopsis T-DNA mutant affected in Nrt2 genes is impaired in nitrate uptake. FEBS Letters 489, 220–224. https://doi.org/10.1016/S0014-5793(01)02096-8

Fischer, M.C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K.K., Holderegger, R., Widmer, A., 2013. Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. Molecular Ecology 22, 5594–5607. https://doi.org/10.1111/mec.12521

Fjellheim, S., Jørgensen, M.H., Kjos, M., Borgen, L., 2009. A molecular study of hybridization and homoploid hybrid speciation in Argyranthemum (Asteraceae) on Tenerife, the Canary Islands. Botanical Journal of the Linnean Society 159, 19–31. https://doi.org/10.1111/j.1095-8339.2008.00947.x

Flood, P.J., Hancock, A.M., 2017. The genomic basis of adaptation in plants. Current Opinion in Plant Biology, 36 Genome studies and molecular genetics 36, 88–94. https://doi.org/10.1016/j.pbi.2017.02.003

Fournier-Level, A., Taylor, M.A., Paril, J.F., Martínez-Berdeja, A., Stitzer, M.C., Cooper, M.D., Roe, J.L., Wilczek, A.M., Schmitt, J., 2022. Adaptive significance of flowering time variation across natural seasonal environments in Arabidopsis thaliana. New Phytologist 234, 719–734. https://doi.org/10.1111/nph.17999

Francisco-Ortega, J., Barber, J.C., Santos-Guerra, A., Febles-Hernández, R., Jansen, R.K., 2001. Origin and evolution of the endemic genera of Gonosperminae (Asteraceae: Anthemideae) from the Canary Islands: evidence from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA. American Journal of Botany 88, 161–169. https://doi.org/10.2307/2657136

Francisco-Ortega, J., Fuertes-Aguilar, J., Gómez-Campo, C., Santos-Guerra, A., Jansen, R.K., 1999. Internal Transcribed Spacer Sequence Phylogeny ofCrambeL. (Brassicaceae): Molecular Data Reveal Two Old World Disjunctions. Molecular Phylogenetics and Evolution 11, 361–380. https://doi.org/10.1006/mpev.1998.0592

Francisco-Ortega, J., Fuertes-Aguilar, J., Kim, S.-C., Santos-Guerra, A., Crawford, D.J., Jansen, R.K., 2002. Phylogeny of the Macaronesian endemic Crambe section Dendrocrambe (Brassicaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. American Journal of Botany 89, 1984–1990. https://doi.org/10.3732/ajb.89.12.1984

Francisco-Ortega, J., Jansen, R.K., Santos-Guerra, A., 1996. Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. Proceedings of the National Academy of Sciences 93, 4085–4090. https://doi.org/10.1073/pnas.93.9.4085

Francisco-Ortega, J., Santos-Guerra, A., Kim, S.-C., Crawford, D.J., 2000. Plant genetic diversity in the Canary Islands: a conservation perspective. American Journal of Botany 87, 909–919. https://doi.org/10.2307/2656988

Francisco-Ortega, J., Santos-Guerra, A., Kim, S.-C., Crawford, D.J., 2000. Plant genetic diversity in the Canary Islands: a conservation perspective. American Journal of Botany 87, 909–919. https://doi.org/10.2307/2656988

Franks, S.J., Hoffmann, A.A., 2012. Genetics of climate change adaptation. Annu Rev Genet 46, 185–208. https://doi.org/10.1146/annurev-genet-110711-155511

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., Mummenhoff, K., 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in Plant Science 16, 108–116. https://doi.org/10.1016/j.tplants.2010.11.005

Fulgione, A., Neto, C., Elfarargi, A.F., Tergemina, E., Ansari, S., Göktay, M., Dinis, H., Döring, N., Flood, P.J., Rodriguez-Pacheco, S., Walden, N., Koch, M.A., Roux, F., Hermisson, J.,

Hancock, A.M., 2022. Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. Nat Commun 13, 1461. https://doi.org/10.1038/s41467-022-28800-z

García-Verdugo, C., Caujapé-Castells, J., Sanmartín, I., 2019. Colonization time on island settings: lessons from the Hawaiian and Canary Island floras. Botanical Journal of the Linnean Society 191, 155–163. https://doi.org/10.1093/botlinnean/boz044

Garrigan, D., Kingan, S.B., Geneva, A.J., Andolfatto, P., Clark, A.G., Thornton, K.R., Presgraves, D.C., 2012. Genome sequencing reveals complex speciation in the Drosophila simulans clade. Genome Res. 22, 1499–1511. https://doi.org/10.1101/gr.130922.111

Garzón-Machado, V., Otto, R., del Arco Aguilar, M.J., 2014. Bioclimatic and vegetation mapping of a topographically complex oceanic island applying different interpolation techniques. Int J Biometeorol 58, 887–899. https://doi.org/10.1007/s00484-013-0670-y

Gaudet, P., Livstone, M.S., Lewis, S.E., Thomas, P.D., 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Briefings in Bioinformatics 12, 449–462. https://doi.org/10.1093/bib/bbr042

Gavrilets, S., 2003. Perspective: Models of Speciation: What Have We Learned in 40 Years? Evolution 57, 2197–2215. https://doi.org/10.1111/j.0014-3820.2003.tb00233.x

Ge, S.X., Jung, D., Yao, R., 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 36, 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Geng, Y., Guan, Y., Qiong, L., Lu, S., An, M., Crabbe, M.J.C., Qi, J., Zhao, F., Qiao, Q., Zhang, T., 2021. Genomic analysis of field pennycress (Thlaspi arvense) provides insights into mechanisms of adaptation to high elevation. BMC Biol 19, 143. https://doi.org/10.1186/s12915-021-01079-0

Gieger, T., Leuschner, C., 2004. Altitudinal change in needle water relations of Pinus canariensis and possible evidence of a drought-induced alpine timberline on Mt. Teide, Tenerife. Flora - Morphology, Distribution, Functional Ecology of Plants 199, 100–109. https://doi.org/10.1078/0367-2530-00139

Gillespie, R., 2004. Community Assembly Through Adaptive Radiation in Hawaiian Spiders. Science 303, 356–359. https://doi.org/10.1126/science.1091875

Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., Wood, K.R., Sytsma, K.J., 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). Proceedings of the Royal Society B: Biological Sciences 276, 407–416. https://doi.org/10.1098/rspb.2008.1204

González-Mancebo, J.M., Hernández-García, C.D., 1996. Bryophyte life strategies along an altitudinal gradient in El Canal y Los Tiles (La Palma, Canary Islands). Journal of Bryology 19, 243–255. https://doi.org/10.1179/jbr.1996.19.2.243

González-Rodríguez, Á.M., Brito, P., Lorenzo, J.R., Gruber, A., Oberhuber, W., Wieser, G., 2017. Seasonal cycles of sap flow and stem radius variation of Spartocytisus supranubius in the alpine zone of Tenerife, Canary Islands. Alp Botany 127, 97–108. https://doi.org/10.1007/s00035-017-0189-7

Goodson, B., 2007. Molecular systematics and biogeography of Descurainia Webb & Berthel, (Brassicaceae).

References

Goodson, B.E., n.d. Molecular systematics and biogeography of Descurainia Webb and Berthel. (Brassicaceae) (Ph.D.). The University of Texas at Austin, United States -- Texas.

Goodson, B.E., Rehman, S.K., Jansen, R.K., 2011. Molecular Systematics and Biogeography of Descurainia (Brassicaceae) based on Nuclear ITS and Non-Coding Chloroplast DNA. Systematic Botany 36, 957–980. https://doi.org/10.1600/036364411X604976

Goodson, B.E., Santos-Guerra, A., Jansen, R.K., 2006a. Molecular systematics of Descurainia (Brassicaceae) in the Canary Islands: biogeographic and taxonomic implications. TAXON 55, 671–682. https://doi.org/10.2307/25065643

Goodson, B.E., Santos-Guerra, A., Jansen, R.K., 2006b. Molecular systematics of Descurainia (Brassicaceae) in the Canary Islands: biogeographic and taxonomic implications [WWW Document]. URL https://www.ingentaconnect.com/content/iapt/tax/2006/00000055/00000003/art00014 (accessed 1.1.19).

Graham, R.E., Reyes-Betancort, J.A., Chapman, M.A., Carine, M.A., 2021. Inter-island differentiation and contrasting patterns of diversity in the iconic Canary Island sub-alpine endemic Echium wildpretii (Boraginaceae). Systematics and Biodiversity 19, 507–525. https://doi.org/10.1080/14772000.2021.1877847

Grant, P.R., Grant, B.R., 1994. Phenotypic and Genetic Effects of Hybridization in Darwin's Finches. Evolution 48, 297–316. https://doi.org/10.1111/j.1558-5646.1994.tb01313.x

Grant, P.R., Grant, B.R., Markert, J.A., Keller, L.F., Petren, K., 2004. Convergent Evolution of Darwin's Finches Caused by Introgressive Hybridization and Selection. Evolution 58, 1588–1599. https://doi.org/10.1111/j.0014-3820.2004.tb01738.x

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A Draft Sequence of the Neandertal Genome. Science 328, 710–722. https://doi.org/10.1126/science.1188021

Grill, D., Tausz, M., Pöllinger, U. t. e., Jiménez, M.S., Morales, D., 2004. Effects of drought on needle anatomy of Pinus canariensis. Flora - Morphology, Distribution, Functional Ecology of Plants 199, 85–89. https://doi.org/10.1078/0367-2530-00137

Günther, T., Lampei, C., Barilar, I., Schmid, K.J., 2016. Genomic and phenotypic differentiation of Arabidopsis thaliana along altitudinal gradients in the North Italian Alps. Molecular Ecology 25, 3574–3592. https://doi.org/10.1111/mec.13705

Guo, X., Hu, Q., Hao, G., Wang, X., Zhang, D., Ma, T., Liu, J., 2018. The genomes of two Eutrema species provide insight into plant adaptation to high altitudes. DNA Research 25, 307–315. https://doi.org/10.1093/dnares/dsy003

Guo, X., Mandáková, T., Trachtová, K., Özüdoğru, B., Liu, J., Lysak, M.A., 2021. Linked by Ancestral Bonds: Multiple Whole-Genome Duplications and Reticulate Evolution in a Brassicaceae Tribe. Molecular Biology and Evolution 38, 1695–1714. https://doi.org/10.1093/molbev/msaa327

Guo, Y.-L., Todesco, M., Hagmann, J., Das, S., Weigel, D., 2012. Independent FLC Mutations as Causes of Flowering-Time Variation in Arabidopsis thaliana and Capsella rubella. Genetics 192, 729–739. https://doi.org/10.1534/genetics.112.143958

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Halbritter, A.H., Fior, S., Keller, I., Billeter, R., Edwards, P.J., Holderegger, R., Karrenberg, S., Pluess, A.R., Widmer, A., Alexander, J.M., 2018. Trait differentiation and adaptation of plants along elevation gradients. Journal of Evolutionary Biology 31, 784–800. https://doi.org/10.1111/jeb.13262

Hall, A.E., Fiebig, A., Preuss, D., 2002. Beyond the Arabidopsis Genome: Opportunities for Comparative Genomics. Plant Physiology 129, 1439–1447. https://doi.org/10.1104/pp.004051

Hämälä, T., Savolainen, O., 2019. Genomic Patterns of Local Adaptation under Gene Flow in Arabidopsis lyrata. Molecular Biology and Evolution 36, 2557–2571. https://doi.org/10.1093/molbev/msz149

Hanz, D.M., Cutts, V., Barajas-Barbosa, M.P., Algar, A.C., Beierkuhnlein, C., Fernández-Palacios, J.-M., Field, R., Kreft, H., Steinbauer, M.J., Weigelt, P., Irl, S.D.H., 2022. Climatic and biogeographical drivers of functional diversity in the flora of the Canary Islands. Global Ecology and Biogeography 31, 1313–1331. https://doi.org/10.1111/geb.13507

Harter, D.E.V., Thiv, M., Weig, A., Jentsch, A., Beierkuhnlein, C., 2015. Spatial and ecological population genetic structures within two island-endemic Aeonium species of different niche width. Ecology and Evolution 5, 4327–4344. https://doi.org/10.1002/ece3.1682

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., Dewar, K., Stinchcombe, J.R., Schoen, D.J., Wang, X., Schmutz, J., Town, C.D., Edger, P.P., Pires, J.C., Schumaker, K.S., Jarvis, D.E., Mandáková, T., Lysak, M.A., van den Bergh, E., Schranz, M.E., Harrison, P.M., Moses, A.M., Bureau, T.E., Wright, S.I., Blanchette, M., 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet 45, 891–898. https://doi.org/10.1038/ng.2684

Hefner, E., Preuss, S.B., Britt, A.B., 2003. Arabidopsis mutants sensitive to gamma radiation include the homologue of the human repair gene ERCC1. Journal of Experimental Botany 54, 669–680. https://doi.org/10.1093/jxb/erg069

Hegarty, M.J., Hiscock, S.J., 2005. Hybrid speciation in plants: new insights from molecular studies. New Phytologist 165, 411–423. https://doi.org/10.1111/j.1469-8137.2004.01253.x

Helfgott, D.M., Francisco-Ortega, J., Santos-Guerra, A., Jansen, R.K., Simpson, B.B., 2000. Biogeography and Breeding System Evolution of the Woody Bencomia Alliance (Rosaceae) in Macaronesia Based on ITS Sequence Data. sbot 25, 82–97. https://doi.org/10.2307/2666675

Helmus, M.R., Mahler, D.L., Losos, J.B., 2014. Island biogeography of the Anthropocene. Nature 513, 543–546. https://doi.org/10.1038/nature13739

Hendriks, K.P., Kiefer, C., Al-Shehbaz, I.A., Bailey, C.D., Huysduynen, A.H. van, Nikolov, L.A., Nauheimer, L., Zuntini, A.R., German, D.A., Franzke, A., Koch, M.A., Lysak, M.A., Toro-Núñez, Ó., Özüdoğru, B., Invernón, V.R., Walden, N., Maurin, O., Hay, N.M., Shushkov, P., Mandáková, T., Thulin, M., Windham, M.D., Rešetnik, I., Španiel, S., Ly, E., Pires, J.C., Harkess, A., Neuffer, B., Vogt, R., Bräuchler, C., Rainer, H., Janssens, S.B., Schmull, M.,

Forrest, A., Guggisberg, A., Zmarzty, S., Lepschi, B.J., Scarlett, N., Stauffer, F.W., Schönberger, I., Heenan, P., Baker, W.J., Forest, F., Mummenhoff, K., Lens, F., 2022. Global Phylogeny of the Brassicaceae Provides Important Insights into Gene Discordance. https://doi.org/10.1101/2022.09.01.506188

HERBEN, T., Rydlová, V., Fér, T., Suda, J., Münzbergová, Z., Wildová, R., Wild, J., 2014. Phylogenetic signal in growth and reproductive traits and in their plasticity: the Descurainia radiation in the Canary Islands. Botanical Journal of the Linnean Society 174, 384–398. https://doi.org/10.1111/boj.12139

Herben, T., Rydlová, V., Fér, T., Suda, J., Münzbergová, Z., Wildová, R., Wild, J., 2014. Phylogenetic signal in growth and reproductive traits and in their plasticity: the Descurainia radiation in the Canary Islands. Bot J Linn Soc 174, 384–398. https://doi.org/10.1111/boj.12139

Herben, T., Suda, J., Munclinger, P., 2005. The Ghost of Hybridization Past: Niche Pre-Emption Is Not the Only Explanation of Apparent Monophyly in Island Endemics. Journal of Ecology 93, 572–575.

Hibbins, M.S., Hahn, M.W., 2022. Phylogenomic approaches to detecting and characterizing introgression. Genetics 220, iyab173. https://doi.org/10.1093/genetics/iyab173

Hodkinson, T.R., Perdereau, A., Klaas, M., Cormican, P., Barth, S., 2019. Genotyping by Sequencing and Plastome Analysis Finds High Genetic Variability and Geographical Structure in Dactylis glomerata L. in Northwest Europe Despite Lack of Ploidy Variation. Agronomy 9, 342. https://doi.org/10.3390/agronomy9070342

Hollenstein, K., Dawson, R.J., Locher, K.P., 2007. Structure and mechanism of ABC transporter proteins. Current Opinion in Structural Biology, Membranes / Engineering and design 17, 412–418. https://doi.org/10.1016/j.sbi.2007.07.003

Holliday, J.A., Zhou, L., Bawa, R., Zhang, M., Oubida, R.W., 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in Populus trichocarpa. New Phytologist 209, 1240–1251. https://doi.org/10.1111/nph.13643

Hong, S.W., Jon, J.H., Kwak, J.M., Nam, H.G., 1997. Identification of a Receptor-Like Protein Kinase Gene Rapidly Induced by Abscisic Acid, Dehydration, High Salt, and Cold Treatments in Arabidopsis thaliana. Plant Physiology 113, 1203–1212. https://doi.org/10.1104/pp.113.4.1203

Honjo, M.N., Kudoh, H., 2019. Arabidopsis halleri: a perennial model system for studying population differentiation and local adaptation. AoB PLANTS 11, plz076. https://doi.org/10.1093/aobpla/plz076

Hormiga, G., Arnedo, M., Gillespie, R.G., 2003. Speciation on a Conveyor Belt: Sequential Colonization of the Hawaiian Islands by Orsonwelles Spiders (Araneae, Linyphiidae). Systematic Biology 52, 70–88. https://doi.org/10.1080/10635150390132786

Howarth, D.G., Baum, D.A., 2005. Genealogical Evidence of Homoploid Hybrid Speciation in an Adaptive Radiation of Scaevola (goodeniaceae) in the Hawaiian Islands. Evolution 59, 948–961. https://doi.org/10.1111/j.0014-3820.2005.tb01034.x

Hu, Q., Ma, Y., Mandáková, T., Shi, S., Chen, C., Sun, P., Zhang, L., Feng, L., Zheng, Y., Feng, X., Yang, W., Jiang, J., Li, T., Zhou, P., Yu, Q., Wan, D., Lysak, M.A., Xi, Z., Nevo, E., Liu, J., 2021. Genome evolution of the psammophyte Pugionium for desert adaptation and further speciation. Proceedings of the National Academy of Sciences 118, e2025711118. https://doi.org/10.1073/pnas.2025711118

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F.X., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., Guo, Y.-L., 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet 43, 476–481. https://doi.org/10.1038/ng.807

Hua, X., Wiens, J.J., 2013. How Does Climate Influence Speciation? The American Naturalist 182, 1–12. https://doi.org/10.1086/670690

Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589. https://doi.org/10.1093/genetics/132.2.583

Huerta-Cepas, J., Serra, F., Bork, P., 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Molecular Biology and Evolution 33, 1635–1638. https://doi.org/10.1093/molbev/msw046

Hunt, M., Newbold, C., Berriman, M., Otto, T.D., 2014. A comprehensive evaluation of assembly scaffolding tools. Genome Biol 15, R42. https://doi.org/10.1186/gb-2014-15-3-r42

Huson, D.H., Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution 23, 254–267. https://doi.org/10.1093/molbev/msj030

Irl, S.D.H., 2016. Plant diversity on high elevation islands – drivers of species richness and endemism. Frontiers of Biogeography 8. https://doi.org/10.21425/F58329717

Irl, S.D.H., Beierkuhnlein, C., 2011. Distribution of endemic plant species on an oceanic island – a geospatial analysis of La Palma (Canary Islands). Procedia Environmental Sciences, Spatial Statistics 2011: Mapping Global Change 7, 170–175. https://doi.org/10.1016/j.proenv.2011.07.030

Irl, S.D.H., Harter, D.E.V., Steinbauer, M.J., Gallego Puyol, D., Fernández-Palacios, J.M., Jentsch, A., Beierkuhnlein, C., 2015. Climate vs. topography – spatial patterns of plant species diversity and endemism on a high-elevation island. Journal of Ecology 103, 1621–1633. https://doi.org/10.1111/1365-2745.12463

Irl, S.D.H., Schweiger, A.H., Medina, F.M., Fernández-Palacios, J.M., Harter, D.E.V., Jentsch, A., Provenzale, A., Steinbauer, M.J., Beierkuhnlein, C., 2017. An island view of endemic rarity—Environmental drivers and consequences for nature conservation. Diversity and Distributions 23, 1132–1142. https://doi.org/10.1111/ddi.12605

Ito, T., Meyerowitz, E.M., 2000. Overexpression of a Gene Encoding a Cytochrome P450, CYP78A9, Induces Large and Seedless Fruit in Arabidopsis. The Plant Cell 12, 1541–1550. https://doi.org/10.1105/tpc.12.9.1541

Jansen, M.A.K., Gaba, V., Greenberg, B.M., 1998. Higher plants and UV-B radiation: balancing damage, repair and acclimation. Trends in Plant Science 3, 131–135. https://doi.org/10.1016/S1360-1385(98)01215-1

Jiang, C.-Z., Yen, C.-N., Cronin, K., Mitchell, D., Britt, A.B., 1997. UV- and Gamma-Radiation Sensitive Mutants of Arabidopsis thaliana. Genetics 147, 1401–1409. https://doi.org/10.1093/genetics/147.3.1401

Jiao, W.-B., Schneeberger, K., 2017. The impact of third generation genomic technologies on plant genome assembly. Current Opinion in Plant Biology, 36 Genome studies and molecular genetics 36, 64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jiggins, C.D., Naisbit, R.E., Coe, R.L., Mallet, J., 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411, 302–305. https://doi.org/10.1038/35077075

Jones, K.E., Reyes-Betancort, J.A., Hiscock, S.J., Carine, M.A., 2014. Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of Pericallis (Asteraceae), a Macaronesian endemic genus. American Journal of Botany 101, 637–651. https://doi.org/10.3732/ajb.1300390

Jorgensen, T.H., 2002. The importance of phylogeny and ecology in microgeographical variation in the morphology of four Canarian species of Aeonium (Crassulaceae). Biological Journal of the Linnean Society 76, 521–533. https://doi.org/10.1046/j.1095-8312.2002.00088.x

Jorgensen, T.H., Frydenberg, J., 1999. Diversification in insular plants: inferring the phylogenetic relationship in Aeonium (Crassulaceae) using ITS sequences of nuclear ribosomal DNA. Nordic Journal of Botany 19, 613–621. https://doi.org/10.1111/j.1756-1051.1999.tb01150.x

Jorgensen, T.H., Olesen, J.M., 2001. Adaptive radiation of island plants: evidence from Aeonium (Crassulaceae) of the Canary Islands. Perspectives in Plant Ecology, Evolution and Systematics 4, 29–42. https://doi.org/10.1078/1433-8319-00013

Juan, C., Emerson, B.C., Oromí, P., Hewitt, G.M., 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends in Ecology & Evolution 15, 104–109. https://doi.org/10.1016/S0169-5347(99)01776-0

Juenger, T., Purugganan, M., Mackay, T.F.C., 2000. Quantitative Trait Loci for Floral Morphology in Arabidopsis thaliana. Genetics 156, 1379–1392. https://doi.org/10.1093/genetics/156.3.1379

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. CGR 110, 462–467. https://doi.org/10.1159/000084979

Kambysellis, M.P., Ho, K.-F., Craddock, E.M., Piano, F., Parisi, M., Cohen, J., 1995. Pattern of ecological shifts in the diversification of Hawaiian Drosophila inferred from a molecular phylogeny. Current Biology 5, 1129–1139. https://doi.org/10.1016/S0960-9822(95)00229-6

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., Martinoia, E., 2011. Plant ABC Transporters. Arabidopsis Book 9, e0153. https://doi.org/10.1199/tab.0153

Keller, S.R., Chhatre, V.E., Fitzpatrick, M.C., 2018. Influence of Range Position on Locally Adaptive Gene–Environment Associations in Populus Flowering Time Genes. Journal of Heredity 109, 47–58. https://doi.org/10.1093/jhered/esx098

Kelly, J.K., Noor, M.A.F., 1996. Speciation by Reinforcement: A Model Derived from Studies of Drosophila. Genetics 143, 1485–1497. https://doi.org/10.1093/genetics/143.3.1485

Kerbs, B., Ressler, J., Kelly, J.K., Mort, M.E., Santos-Guerra, A., Gibson, M.J.S., Caujapé-Castells, J., Crawford, D.J., 2017. The potential role of hybridization in diversification and speciation in an insular plant lineage: insights from synthetic interspecific hybrids. AoB PLANTS 9, plx043. https://doi.org/10.1093/aobpla/plx043

Kersey, P.J., 2019. Plant genome sequences: past, present, future. Current Opinion in Plant Biology, Genome studies and molecular genetics 48, 1–8. https://doi.org/10.1016/j.pbi.2018.11.001

Kier, G., Kreft, H., Lee, T.M., Jetz, W., Ibisch, P.L., Nowicki, C., Mutke, J., Barthlott, W., 2009. A global assessment of endemism and species richness across island and mainland regions. Proceedings of the National Academy of Sciences 106, 9322–9327. https://doi.org/10.1073/pnas.0810306106

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J.M., Lee, H.-A., Seo, E., Choi, J., Cheong, K., Kim, K.-T., Jung, K., Lee, G.-W., Oh, S.-K., Bae, C., Kim, S.-B., Lee, H.-Y., Kim, S.-Y., Kim, M.-S., Kang, B.-C., Jo, Y.D., Yang, H.-B., Jeong, H.-J., Kang, W.-H., Kwon, J.-K., Shin, C., Lim, J.Y., Park, J.H., Huh, J.H., Kim, J.-S., Kim, B.-D., Cohen, O., Paran, I., Suh, M.C., Lee, S.B., Kim, Y.-K., Shin, Y., Noh, S.-J., Park, J., Seo, Y.S., Kwon, S.-Y., Kim, H.A., Park, J.M., Kim, H.-J., Choi, S.-B., Bosland, P.W., Reeves, G., Jo, S.-H., Lee, B.-W., Cho, H.-T., Choi, H.-S., Lee, M.-S., Yu, Y., Do Choi, Y., Park, B.-S., van Deynze, A., Ashrafi, H., Hill, T., Kim, W.T., Pai, H.-S., Ahn, H.K., Yeam, I., Giovannoni, J.J., Rose, J.K.C., Sørensen, I., Lee, S.-J., Kim, R.W., Choi, I.-Y., Choi, B.-S., Lim, J.-S., Lee, Y.-H., Choi, D., 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet 46, 270–278. https://doi.org/10.1038/ng.2877

Kim, S.-C., Crawford, D.J., Francisco-Ortega, J., Santos-Guerra, A., 1999. Adaptive radiation and genetic differentiation in the woodySonchus alliance (Asteraceae:Sonchinae) in the Canary Islands. Pl Syst Evol 215, 101–118. https://doi.org/10.1007/BF00984650

Kim, S.C., Crawford, D.J., Francisco-Ortega, J., Santos-Guerra, A., 1996. A common origin for woody Sonchus and five related genera in the Macaronesian islands: molecular evidence for extensive radiation. Proceedings of the National Academy of Sciences 93, 7743–7748. https://doi.org/10.1073/pnas.93.15.7743

Kim, S.Y., Nam, K.H., 2010. Physiological roles of ERD10 in abiotic stresses and seed germination of Arabidopsis. Plant Cell Rep 29, 203–209. https://doi.org/10.1007/s00299-009-0813-0

Kisel, Y., Barraclough, T.G., 2010. Speciation Has a Spatial Scale That Depends on Levels of Gene Flow. The American Naturalist 175, 316–334. https://doi.org/10.1086/650369

Kleinkopf, J.A., Roberts, W.R., Wagner, W.L., Roalson, E.H., 2019. Diversification of Hawaiian Cyrtandra (Gesneriaceae) under the influence of incomplete lineage sorting and hybridization. Journal of Systematics and Evolution 57, 561–578. https://doi.org/10.1111/jse.12519

Kliver, S., Rayko, M., Komissarov, A., Bakin, E., Zhernakova, D., Prasad, K., Rushworth, C., Baskar, R., Smetanin, D., Schmutz, J., Rokhsar, D.S., Mitchell-Olds, T., Grossniklaus, U., Brukhin, V., 2018. Assembly of the Boechera retrofracta Genome and Evolutionary Analysis of Apomixis-Associated Genes. Genes 9, 185. https://doi.org/10.3390/genes9040185

Koblmüller, S., Egger, B., Sturmbauer, C., Sefc, K.M., 2010. Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. Molecular Phylogenetics and Evolution 55, 318–334. https://doi.org/10.1016/j.ympev.2009.09.032

Koh, S., Lee, S.-C., Kim, M.-K., Koh, J.H., Lee, S., An, G., Choe, S., Kim, S.-R., 2007. T-DNA tagged knockout mutation of rice OsGSK1, an orthologue of Arabidopsis BIN2, with enhanced tolerance to various abiotic stresses. Plant Mol Biol 65, 453–466. https://doi.org/10.1007/s11103-007-9213-4

Köhler, L., Gieger, T., Leuschner, C., 2006. Altitudinal change in soil and foliar nutrient concentrations and in microclimate across the tree line on the subtropical island mountain Mt. Teide (Canary Islands). Flora - Morphology, Distribution, Functional Ecology of Plants 201, 202–214. https://doi.org/10.1016/j.flora.2005.07.003

Kollas, C., Körner, C., Randin, C.F., 2014. Spring frost and growing season length co-control the cold range limits of broad-leaved trees. Journal of Biogeography 41, 773–783. https://doi.org/10.1111/jbi.12238

Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., Phillippy, A.M., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 30, 693–700. https://doi.org/10.1038/nbt.2280

Körner, C., 2016. Plant adaptation to cold climates. F1000Res 5, F1000 Faculty Rev-2769. https://doi.org/10.12688/f1000research.9107.1

Kreft, H., Jetz, W., Mutke, J., Kier, G., Barthlott, W., 2008. Global diversity of island floras from a macroecological perspective. Ecology Letters 11, 116–127. https://doi.org/10.1111/j.1461-0248.2007.01129.x

Kretzschmar, T., Burla, B., Lee, Y., Martinoia, E., Nagy, R., 2011. Functions of ABC transporters in plants. Essays in Biochemistry 50, 145–160. https://doi.org/10.1042/bse0500145

Kryazhimskiy, S., Plotkin, J.B., 2008. The Population Genetics of dN/dS. PLOS Genetics 4, e1000304. https://doi.org/10.1371/journal.pgen.1000304

Kubota, S., Iwasaki, T., Hanada, K., Nagano, A.J., Fujiyama, A., Toyoda, A., Sugano, S., Suzuki, Y., Hikosaka, K., Ito, M., Morinaga, S.-I., 2015. A Genome Scan for Genes Underlying Microgeographic-Scale Local Adaptation in a Wild Arabidopsis Species. PLOS Genetics 11, e1005361. https://doi.org/10.1371/journal.pgen.1005361

Kueffer, C., Daehler, C.C., Torres-Santana, C.W., Lavergne, C., Meyer, J.-Y., Otto, R., Silva, L., 2010. A global comparison of plant invasions on oceanic islands. Perspectives in Plant Ecology, Evolution and Systematics, Comparative ecological research on oceanic islands 12, 145–161. https://doi.org/10.1016/j.ppees.2009.06.002

Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Molecular Biology and Evolution 34, 1812–1819. https://doi.org/10.1093/molbev/msx116

Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., Hedges, S.B., 2022. TimeTree 5: An Expanded Resource for Species Divergence Times. Molecular Biology and Evolution 39, msac174. https://doi.org/10.1093/molbev/msac174

Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M.A., Janke, A., 2017. The evolutionary history of bears is characterized by gene flow across species. Sci Rep 7, 46487. https://doi.org/10.1038/srep46487

Kunkel, G., 2012. Biogeography and Ecology in the Canary Islands. Springer Science & Business Media.

Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. https://doi.org/10.12688/f1000research.12232.1

Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., Grant, P.R., 2018. Rapid hybrid speciation in Darwin's finches. Science 359, 224–228. https://doi.org/10.1126/science.aao4593

Lausi, D., Nimis, P.L., 1986. Leaf and canopy adaptations in a high-elevation desert on Tenerife, Canary Islands. Vegetatio 68, 19–31. https://doi.org/10.1007/BF00031576

Le Moan, A., Gagnaire, P.-A., Bonhomme, F., 2016. Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Molecular Ecology 25, 3187–3202. https://doi.org/10.1111/mec.13627

Lee, C., Kim, S.-C., Lundy, K., Santos-Guerra, A., 2005. Chloroplast DNA phylogeny of the woody Sonchus alliance (Asteraceae: Sonchinae) in the Macaronesian Islands. American Journal of Botany 92, 2072–2085. https://doi.org/10.3732/ajb.92.12.2072

Lee, I., Aukerman, M.J., Gore, S.L., Lohman, K.N., Michaels, S.D., Weaver, L.M., John, M.C., Feldmann, K.A., Amasino, R.M., 1994. Isolation of LUMINIDEPENDENS: a gene involved in the control of flowering time in Arabidopsis. The Plant Cell 6, 75–83. https://doi.org/10.1105/tpc.6.1.75

Lee, I.C., Hong, S.W., Whang, S.S., Lim, P.O., Nam, H.G., Koo, J.C., 2011. Age-Dependent Action of an ABA-Inducible Receptor Kinase, RPK1, as a Positive Regulator of Senescence in Arabidopsis Leaves. Plant and Cell Physiology 52, 651–662. https://doi.org/10.1093/pcp/pcr026

Lee, J.R., Xie, X., Yang, K., Zhang, J., Lee, S.Y., Shippen, D.E., 2016. Dynamic Interactions of Arabidopsis TEN1: Stabilizing Telomeres in Response to Heat Stress. The Plant Cell 28, 2212–2224. https://doi.org/10.1105/tpc.16.00408

Lee, U., Rioflorido, I., Hong, S.-W., Larkindale, J., Waters, E.R., Vierling, E., 2007. The Arabidopsis ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development. The Plant Journal 49, 115–127. https://doi.org/10.1111/j.1365-313X.2006.02940.x

Leinonen, P.H., Sandring, S., Quilot, B., Clauss, M.J., Mitchell-Olds, T., Ågren, J., Savolainen, O., 2009. Local adaptation in European populations of Arabidopsis lyrata (Brassicaceae). American Journal of Botany 96, 1129–1137. https://doi.org/10.3732/ajb.0800080

Lemmon, E., Lemmon, A., 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44, 99–121. https://doi.org/10.1146/annurev-ecolsys-110512-135822

Lerner, H.R.L., Meyer, M., James, H.F., Hofreiter, M., Fleischer, R.C., 2011. Multilocus Resolution of Phylogeny and Timescale in the Extant Adaptive Radiation of Hawaiian Honeycreepers. Current Biology 21, 1838–1844. https://doi.org/10.1016/j.cub.2011.09.039

Li, F.-W., Harkess, A., 2018. A guide to sequence your favorite plant genomes. Applications in Plant Sciences 6, e1030. https://doi.org/10.1002/aps3.1030

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J.-L., Zhong, L.-L., Wang, J., Ma, T., Mao, K.-S., Zhang, L., 2021. Genomic insights into speciation history and local adaptation of an alpine aspen in the Qinghai–Tibet Plateau and adjacent highlands. Journal of Systematics and Evolution 59, 1220–1231. https://doi.org/10.1111/jse.12665

Li, R., Li, Y., Fang, X., Yang, H., Wang, Jian, Kristiansen, K., Wang, Jun, 2009. SNP detection for massively parallel whole-genome resequencing. Genome Res. 19, 1124–1132. https://doi.org/10.1101/gr.088013.108

Liu, L.-M., Du, X.-Y., Guo, C., Li, D.-Z., 2021. Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. Journal of Systematics and Evolution 59, 442–453. https://doi.org/10.1111/jse.12666

Lomolino, Mark.V., 2001. Elevation gradients of species-density: historical and prospective views. Global Ecology and Biogeography 10, 3–13. https://doi.org/10.1046/j.1466-822x.2001.00229.x

Lomolino, M.V., 2001. The species-area relationship: new challenges for an old pattern. Progress in Physical Geography: Earth and Environment 25, 1–21. https://doi.org/10.1177/030913330102500101

López, R., Aranda, I., Gil, L., 2009. Osmotic adjustment is a significant mechanism of drought resistance in Pinus pinaster and Pinus canariensis. Forest Systems 18, 159–166. https://doi.org/10.5424/fs/2009182-01059

López, R., Climent, J., Gil, L., 2008. From desert to cloud forest: the non-trivial phenotypic variation of Canary Island pine needles. Trees 22, 843–849. https://doi.org/10.1007/s00468-008-0245-4

Losos, J.B., 2011. Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles. Univ of California Press.

Losos, J.B., 1992. The Evolution of Convergent Structure in Caribbean Anolis Communities. Systematic Biology 41, 403–420. https://doi.org/10.1093/sysbio/41.4.403

Losos, J.B., Ricklefs, R.E., 2009a. Adaptation and diversification on islands. Nature 457, 830–836. https://doi.org/10.1038/nature07893

Losos, J.B., Ricklefs, R.E., 2009b. The Theory of Island Biogeography Revisited. Princeton University Press.

Losos, J.B., Schoener, T.W., Warheit, K.I., Creer, D., 2001. Experimental studies of adaptive differentiation in Bahamian Anolis lizards, in: Hendry, A.P., Kinnison, M.T. (Eds.), Microevolution Rate, Pattern, Process, Contemporary Issues in Genetics and Evolution. Springer Netherlands, Dordrecht, pp. 399–415. https://doi.org/10.1007/978-94-010-0585-2_24

Losos, J.B., Warheitt, K.I., Schoener, T.W., 1997. Adaptive differentiation following experimental island colonization in Anolis lizards. Nature 387, 70–73. https://doi.org/10.1038/387070a0

Lowrey, T.K., 1986. A Biosystematic Revision of Hawaiian Tetramolopium (compositae: Astereae). Allertonia 4, 203–265.

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., Zhang, C., Chen, Z., Xiao, Z., Jian, H., Cheng, F., Zhang, K., Du, H., Cheng, X., Qu, C., Qian, W., Liu, L., Wang, R., Zou, Q., Ying, J., Xu, X., Mei, J., Liang, Y., Chai, Y.-R., Tang, Z., Wan, H., Ni, Y., He, Y., Lin, N., Fan, Y., Sun, W., Li, N.-N., Zhou, G., Zheng, H., Wang, X., Paterson, A.H., Li, J., 2019. Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement. Nat Commun 10, 1154. https://doi.org/10.1038/s41467-019-09134-9

Luikart, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P., 2003. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4, 981–994. https://doi.org/10.1038/nrg1226

Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A., Leitch, I.J., 2009. The Dynamic Ups and Downs of Genome Size Evolution in Brassicaceae. Molecular Biology and Evolution 26, 85–98. https://doi.org/10.1093/molbev/msn223

Ma, Y., Wang, J., Hu, Q., Li, J., Sun, Y., Zhang, L., Abbott, R.J., Liu, J., Mao, K., 2019. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. Commun Biol 2, 1–12. https://doi.org/10.1038/s42003-019-0445-z

Ma, Y.-F., Han, X.-M., Huang, C.-P., Zhong, L., Adeola, A.C., Irwin, D.M., Xie, H.-B., Zhang, Y.-P., 2019. Population Genomics Analysis Revealed Origin and High-altitude Adaptation of Tibetan Pigs. Sci Rep 9, 11463. https://doi.org/10.1038/s41598-019-47711-6

Magnacca, K.N., Price, D.K., 2015. Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing Drosophila (Diptera: Drosophilidae). Molecular Phylogenetics and Evolution 92, 226–242. https://doi.org/10.1016/j.ympev.2015.06.014

Mahfouz, M.M., Kim, S., Delauney, A.J., Verma, D.P.S., 2006. Arabidopsis TARGET OF RAPAMYCIN Interacts with RAPTOR, Which Regulates the Activity of S6 Kinase in Response to Osmotic Stress Signals. The Plant Cell 18, 477–490. https://doi.org/10.1105/tpc.105.035931

Mallet, J., 2007. Hybrid speciation. Nature 446, 279–283. https://doi.org/10.1038/nature05706

Marks, R.A., Hotaling, S., Frandsen, P.B., VanBuren, R., 2021. Representation and participation across 20 years of plant genome sequencing. Nat. Plants 7, 1571–1578. https://doi.org/10.1038/s41477-021-01031-8

Marques, D.A., Meier, J.I., Seehausen, O., 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends in Ecology & Evolution 34, 531–544. https://doi.org/10.1016/j.tree.2019.02.008

Martín-Hernanz, S., Aparicio, A., Fernández-Mazuecos, M., Rubio, E., Reyes-Betancort, J.A., Santos-Guerra, A., Olangua-Corral, M., Albaladejo, R.G., 2019. Maximize Resolution or Minimize Error? Using Genotyping-By-Sequencing to Investigate the Recent Diversification of Helianthemum (Cistaceae). Frontiers in Plant Science 10.

Marzol, E., Borassi, C., Carignani Sardoy, M., Ranocha, P., Aptekmann, A.A., Bringas, M., Pennington, J., Paez-Valencia, J., Martínez Pacheco, J., Rodríguez-Garcia, D.R., Rondón Guerrero, Y. del C., Peralta, J.M., Fleming, M., Mishler-Elmore, J.W., Mangano, S., Blanco-Herrera, F., Bedinger, P.A., Dunand, C., Capece, L., Nadra, A.D., Held, M., Otegui, M.S., Estevez, J.M., 2022. Class III Peroxidases PRX01, PRX44, and PRX73 Control Root Hair Growth in Arabidopsis thaliana. International Journal of Molecular Sciences 23, 5375. https://doi.org/10.3390/ijms23105375

Masand, S., Yadav, S.K., 2016. Overexpression of MuHSP70 gene from Macrotyloma uniflorum confers multiple abiotic stress tolerance in transgenic Arabidopsis thaliana. Mol Biol Rep 43, 53–64. https://doi.org/10.1007/s11033-015-3938-y

Mavárez, J., Linares, M., 2008. Homoploid hybrid speciation in animals. Molecular Ecology 17, 4181–4185. https://doi.org/10.1111/j.1365-294X.2008.03898.x

Mayer, S.S., 1991. Artificial Hybridization in Hawaiian Wikstroemia (thymelaeaceae). American Journal of Botany 78, 122–130. https://doi.org/10.1002/j.1537-2197.1991.tb12578.x

References

Mayr, E., 2013. Animal Species and Evolution, Animal Species and Evolution. Harvard University Press. https://doi.org/10.4159/harvard.9780674865327

Mayr, E., 1982. Speciation and Macroevolution. Evolution 36, 1119–1132. https://doi.org/10.2307/2408147

Mayr, E., 1947. Ecological Factors in Speciation. Evolution 1, 263–288. https://doi.org/10.2307/2405327

McCluskey, B.M., Postlethwait, J.H., 2015. Phylogeny of Zebrafish, a "Model Species," within Danio, a "Model Genus." Molecular Biology and Evolution 32, 635–652. https://doi.org/10.1093/molbev/msu325

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22, 746–754. https://doi.org/10.1101/gr.125864.111

McKinnon, J.S., Mori, S., Blackman, B.K., David, L., Kingsley, D.M., Jamieson, L., Chou, J., Schluter, D., 2004. Evidence for ecology's role in speciation. Nature 429, 294–298. https://doi.org/10.1038/nature02556

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., Seehausen, O., 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat Commun 8, 14363. https://doi.org/10.1038/ncomms14363

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L., Seehausen, O., 2018. Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. Molecular Biology and Evolution 35, 1489–1506. https://doi.org/10.1093/molbev/msy051

Meimberg, H., Abele, T., Bräuchler, C., McKay, J.K., Pérez de Paz, P.L., Heubl, G., 2006. Molecular evidence for adaptive radiation of Micromeria Benth. (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. Molecular Phylogenetics and Evolution 41, 566–578. https://doi.org/10.1016/j.ympev.2006.05.037

Mendelson, T.C., Shaw, K.L., 2005. Rapid speciation in an arthropod. Nature 433, 375–376. https://doi.org/10.1038/433375a

Meudt, H.M., Albach, D.C., Tanentzap, A.J., Igea, J., Newmarch, S.C., Brandt, A.J., Lee, W.G., Tate, J.A., 2021. Polyploidy on Islands: Its Emergence and Importance for Diversification. Frontiers in Plant Science 12.

Meyers, L.A., Levin, D.A., 2006. On the Abundance of Polyploids in Flowering Plants. Evolution 60, 1198–1206. https://doi.org/10.1111/j.0014-3820.2006.tb01198.x

Mishra, B., Ploch, S., Runge, F., Schmuker, A., Xia, X., Gupta, D.K., Sharma, R., Thines, M., 2020. The Genome of Microthlaspi erraticum (Brassicaceae) Provides Insights Into the Adaptation to Highly Calcareous Soils. Frontiers in Plant Science 11.

Mora, M.A., Llambí, L.D., Ramírez, L., 2019. Giant stem rosettes have strong facilitation effects on alpine plant communities in the tropical Andes. Plant Ecology & Diversity 12, 593–606. https://doi.org/10.1080/17550874.2018.1507055

Morales, D., Jiménez, M.S., González-Rodríguez, A.M., Cermák, J., 1996. Laurel forests in Tenerife, Canary Islands. Trees 11, 34–40. https://doi.org/10.1007/s004680050055

Morales-Briones, D.F., Liston, A., Tank, D.C., 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus Lachemilla (Rosaceae). New Phytologist 218, 1668–1684. https://doi.org/10.1111/nph.15099

Morgat, A., Lombardot, T., Coudert, E., Axelsen, K., Neto, T.B., Gehant, S., Bansal, P., Bolleman, J., Gasteiger, E., de Castro, E., Baratin, D., Pozzato, M., Xenarios, I., Poux, S., Redaschi, N., Bridge, A., The UniProt Consortium, 2020. Enzyme annotation in UniProtKB using Rhea. Bioinformatics 36, 1896–1901. https://doi.org/10.1093/bioinformatics/btz817

Mort, M.E., Kerbs, B.R., Kelly, J.K., Silva, L.B., Moura, M., de Sequeira, M.M., Santos-Guerra, A., Schaefer, H., Alfredo Reyes-Betancort, J., Caujapé-Castells, J., Crawford, D.J., 2022. Multiplexed shotgun genotyping (MSG) data resolve phylogenetic relationships within and among archipelagos in Macaronesian Tolpis. American Journal of Botany 109, 952–965. https://doi.org/10.1002/ajb2.1866

Mort, M.E., Soltis, D.E., Soltis, P.S., Francisco-Ortega, J., Santos-Guerra, A., 2002. Phylogenetics and Evolution of the Macaronesian Clade of Crassulaceae Inferred from Nuclear and Chloroplast Sequence Data. sbot 27, 271–288. https://doi.org/10.1043/0363-6445-27.2.271

Myouga, F., Motohashi, R., Kuromori, T., Nagata, N., Shinozaki, K., 2006. An Arabidopsis chloroplast-targeted Hsp101 homologue, APG6, has an essential role in chloroplast development as well as heat-stress response. The Plant Journal 48, 249–260. https://doi.org/10.1111/j.1365-313X.2006.02873.x

Nagase, M., Ando, K., Nagase, T., Kaname, S., Sawamura, T., Fujita, T., 2001. Redox-Sensitive Regulation of LOX-1 Gene Expression in Vascular Endothelium. Biochemical and Biophysical Research Communications 281, 720–725. https://doi.org/10.1006/bbrc.2001.4374

Nagy, L., Grabherr, G., 2009. The Biology of Alpine Habitats. OUP Oxford.

Nagy, L., Proctor, J., 1997. Plant growth and reproduction on a toxic alpine ultramafic soil: adaptation to nutrient limitation. The New Phytologist 137, 267–274. https://doi.org/10.1046/j.1469-8137.1997.00799.x

Nakashima, K., Fujita, Y., Katsura, K., Maruyama, K., Narusaka, Y., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki, K., 2006. Transcriptional Regulation of ABI3- and ABA-responsive Genes Including RD29B and RD29A in Seeds, Germinating Embryos, and Seedlings of Arabidopsis. Plant Mol Biol 60, 51–68. https://doi.org/10.1007/s11103-005-2418-5

Nakhleh, L., 2011. Evolutionary Phylogenetic Networks: Models and Issues, in: Heath, L.S., Ramakrishnan, N. (Eds.), Problem Solving Handbook in Computational Biology and Bioinformatics. Springer US, Boston, MA, pp. 125–158. https://doi.org/10.1007/978-0-387-09760-2_7

Nikolov, L.A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I.A., Filatov, D., Bailey, C.D., Tsiantis, M., 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytologist 222, 1638–1651. https://doi.org/10.1111/nph.15732

Noor, M.A., 1995. Speciation driven by natural selection in Drosophila. Nature 375, 674–675. https://doi.org/10.1038/375674a0

Nordin, K., Vahala, T., Palva, E.T., 1993. Differential expression of two related, low-temperature-induced genes in Arabidopsis thaliana (L.) Heynh. Plant Mol Biol 21, 641–653. https://doi.org/10.1007/BF00014547

Nosil, P., 2012. Ecological Speciation. OUP Oxford.

Nosil, P., Harmon, L.J., Seehausen, O., 2009. Ecological explanations for (incomplete) speciation. Trends in Ecology & Evolution 24, 145–156. https://doi.org/10.1016/j.tree.2008.10.011

Nosil, P., Schluter, D., 2011. The genes underlying the process of speciation. Trends in Ecology & Evolution 26, 160–167. https://doi.org/10.1016/j.tree.2011.01.001

Nowak, M.D., Birkeland, S., Mandáková, T., Roy Choudhury, R., Guo, X., Gustafsson, A.L.S., Gizaw, A., Schrøder-Nielsen, A., Fracassetti, M., Brysting, A.K., Rieseberg, L., Slotte, T., Parisod, C., Lysak, M.A., Brochmann, C., 2021. The genome of Draba nivalis shows signatures of adaptation to the extreme environmental stresses of the Arctic. Molecular Ecology Resources 21, 661–676. https://doi.org/10.1111/1755-0998.13280

Okazaki, K., Sakamoto, K., Kikuchi, R., Saito, A., Togashi, E., Kuginuki, Y., Matsumoto, S., Hirai, M., 2007. Mapping and characterization of FLC homologs and QTL analysis of flowering time in Brassica oleracea. Theor Appl Genet 114, 595–608. https://doi.org/10.1007/s00122-006-0460-6

Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294. https://doi.org/10.1093/bioinformatics/btv566

Olmstead, R.G., Palmer, J.D., 1994. Chloroplast DNA systematics: a review of methods and data analysis. American Journal of Botany 81, 1205–1224. https://doi.org/10.1002/j.1537-2197.1994.tb15615.x

Ono, K., Hibino, T., Kohinata, T., Suzuki, S., Tanaka, Y., Nakamura, T., Takabe, T., Takabe, T., 2001. Overexpression of DnaK from a halotolerant cyanobacterium Aphanothece halophytica enhances the high-temperatue tolerance of tobacco during germination and early growth. Plant Science 160, 455–461. https://doi.org/10.1016/S0168-9452(00)00412-X

Orr, H.A., Coyne, J.A., 1992. The Genetics of Adaptation: A Reassessment. The American Naturalist 140, 725–742. https://doi.org/10.1086/285437

Orr, H.A., Masly, J.P., Presgraves, D.C., 2004. Speciation genes. Current Opinion in Genetics & Development 14, 675–679. https://doi.org/10.1016/j.gde.2004.08.009

Ortuño-Miquel, S., Rodríguez-Cazorla, E., Zavala-Gonzalez, E.A., Martínez-Laborda, A., Vera, A., 2019. Arabidopsis HUA ENHANCER 4 delays flowering by upregulating the MADS-box repressor genes FLC and MAF4. Sci Rep 9, 1478. https://doi.org/10.1038/s41598-018-38327-3

Osakabe, Y., Mizuno, S., Tanaka, H., Maruyama, K., Osakabe, K., Todaka, D., Fujita, Y., Kobayashi, M., Shinozaki, K., Yamaguchi-Shinozaki, K., 2010. Overproduction of the Membrane-bound Receptor-like Protein Kinase 1, RPK1, Enhances Abiotic Stress Tolerance in Arabidopsis*. Journal of Biological Chemistry 285, 9190–9201. https://doi.org/10.1074/jbc.M109.051938

Osborne, O.G., Ciezarek, A., Wilson, T., Crayn, D., Hutton, I., Baker, W.J., Turnbull, C.G.N., Savolainen, V., 2019. Speciation in Howea Palms Occurred in Sympatry, Was Preceded by Ancestral Admixture, and Was Associated with Edaphic and Phenological Adaptation. Molecular Biology and Evolution 36, 2682–2697. https://doi.org/10.1093/molbev/msz166

Osborne, O.G., De-Kayne, R., Bidartondo, M.I., Hutton, I., Baker, W.J., Turnbull, C.G.N., Savolainen, V., 2018. Arbuscular mycorrhizal fungi promote coexistence and niche divergence of

sympatric palm species on a remote oceanic island. New Phytologist 217, 1254–1266. https://doi.org/10.1111/nph.14850

Østergaard, L., Yanofsky, M.F., 2004. Establishing gene function by mutagenesis in Arabidopsis thaliana. The Plant Journal 39, 682–696. https://doi.org/10.1111/j.1365-313X.2004.02149.x

Otte, D., Endler, J.A., 1989. Speciation and its Consequences. Sinauer Associates.

Otto, R., Whittaker, R.J., von Gaisberg, M., Stierstorfer, C., Naranjo-Cigala, A., Steinbauer, M.J., Borregaard, M.K., Arévalo, J.R., Garzón-Machado, V., del Arco, M., Fernández-Palacios, J.M., 2016. Transferring and implementing the general dynamic model of oceanic island biogeography at the scale of island fragments: the roles of geological age and topography in plant diversification in the Canaries. Journal of Biogeography 43, 911–922. https://doi.org/10.1111/jbi.12684

Panero, J.L., Francisco-Ortega, J., Jansen, R.K., Santos-Guerra, A., 1999. Molecular evidence for multiple origins of woodiness and a New World biogeographic connection of the Macaronesian Island endemic Pericallis (Asteraceae: Senecioneae). Proceedings of the National Academy of Sciences 96, 13886–13891. https://doi.org/10.1073/pnas.96.24.13886

Papadopulos, A.S.T., Igea, J., Dunning, L.T., Osborne, O.G., Quan, X., Pellicer, J., Turnbull, C., Hutton, I., Baker, W.J., Butlin, R.K., Savolainen, V., 2019a. Ecological speciation in sympatric palms: 3. Genetic map reveals genomic islands underlying species divergence in Howea. Evolution 73, 1986–1995. https://doi.org/10.1111/evo.13796

Papadopulos, A.S.T., Igea, J., Smith, T.P., Hutton, I., Baker, W.J., Butlin, R.K., 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. https://doi.org/10.1111/evo.13813

Parent, C.E., Caccone, A., Petren, K., 2008. Colonization and diversification of Galápagos terrestrial fauna: a phylogenetic and biogeographical synthesis. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 3347–3361. https://doi.org/10.1098/rstb.2008.0118

PARKIN, I.A., Koh, C., Tang, H., Robinson, S.J., Kagale, S., Clarke, W.E., Town, C.D., Nixon, J., Krishnakumar, V., Bidwell, S.L., Denoeud, F., Belcram, H., Links, M.G., Just, J., Clarke, C., Bender, T., Huebert, T., Mason, A.S., Pires, J.C., Barker, G., Moore, J., Walley, P.G., Manoli, S., Batley, J., Edwards, D., Nelson, M.N., Wang, X., Paterson, A.H., King, G., Bancroft, I., Chalhoub, B., Sharpe, A.G., 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biology 15, R77. https://doi.org/10.1186/gb-2014-15-6-r77

Patiño, J., Whittaker, R.J., Borges, P.A.V., Fernández-Palacios, J.M., Ah-Peng, C., Araújo, M.B., Ávila, S.P., Cardoso, P., Cornuault, J., de Boer, E.J., de Nascimento, L., Gil, A., González-Castro, A., Gruner, D.S., Heleno, R., Hortal, J., Illera, J.C., Kaiser-Bunbury, C.N., Matthews, T.J., Papadopoulou, A., Pettorelli, N., Price, J.P., Santos, A.M.C., Steinbauer, M.J., Triantis, K.A., Valente, L., Vargas, P., Weigelt, P., Emerson, B.C., 2017. A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. Journal of Biogeography 44, 963–983. https://doi.org/10.1111/jbi.12986

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient Admixture in Human History. Genetics 192, 1065–1093. https://doi.org/10.1534/genetics.112.145037

Paulay, G., 1994. Biodiversity on Oceanic Islands: Its Origin and Extinction1. American Zoologist 34, 134–144. https://doi.org/10.1093/icb/34.1.134

Pei, S., Liu, Y., Li, W., Krichilsky, B., Dai, S., Wang, Y., Wang, X., Johnson, D.M., Crawford, B.M., Swift, G.B., Vo-Dinh, T., Pei, Z.-M., Yuan, F., 2022. OSCA1 is an osmotic specific sensor: a method to distinguish Ca2+-mediated osmotic and ionic perception. New Phytologist 235, 1665–1678. https://doi.org/10.1111/nph.18217

Pellicer, J., Leitch, I.J., n.d. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies [WWW Document].

Pertea, G., Pertea, M., 2020. GFF Utilities: GffRead and GffCompare. F1000Res 9, ISCB Comm J-304. https://doi.org/10.12688/f1000research.23297.2

Petren, K., Grant, P.R., Grant, B.R., Keller, L.F., 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Molecular Ecology 14, 2943–2957. https://doi.org/10.1111/j.1365-294X.2005.02632.x

Pickrell, J., Pritchard, J., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. Nat Prec 1–1. https://doi.org/10.1038/npre.2012.6956.1

Pitard, J., Proust, L., 1908. Les îles Canaries. Flore de l'archipel. P. Klincksieck, Paris. https://doi.org/10.5962/bhl.title.49564

Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., Wolf, J.B.W., 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344, 1410–1414. https://doi.org/10.1126/science.1253226

Porebski, S., Bailey, L.G., Baum, B.R., 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep 15, 8–15. https://doi.org/10.1007/BF02772108

Presgraves, D.C., 2008. Sex chromosomes and speciation in Drosophila. Trends in Genetics 24, 336–343. https://doi.org/10.1016/j.tig.2008.04.007

Puppo, P., Curto, M., Gusmão-Guedes, J., Cochofel, J., Pérez de Paz, P.L., Bräuchler, C., Meimberg, H., 2015. Molecular phylogenetics of Micromeria (Lamiaceae) in the Canary Islands, diversification and inter-island colonization patterns inferred from nuclear genes. Molecular Phylogenetics and Evolution 89, 160–170. https://doi.org/10.1016/j.ympev.2015.04.017

Purugganan, M.D., Jackson, S.A., 2021. Advancing crop genomics from lab to field. Nat Genet 53, 595–601. https://doi.org/10.1038/s41588-021-00866-3

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., González-Morales, S., Trejo-Saavedra, D.L., Tian, H., Tang, Xin, Zhao, M., Huang, Z., Zhou, A., Yao, X., Cui, J., Li, Wenqi, Chen, Z., Feng, Y., Niu, Y., Bi, S., Yang, X., Li, Weipeng, Cai, H., Luo, X., Montes-Hernández, S., Leyva-González, M.A., Xiong, Z., He, X., Bai, L., Tan, S., Tang, Xiangqun, Liu, D., Liu, J., Zhang, S., Chen, M., Zhang, Lu, Zhang, Li, Zhang, Yinchao, Liao, W., Zhang, Yan, Wang, M., Lv, X., Wen, B., Liu, H., Luan, H., Zhang, Yonggang, Yang, S., Wang, X., Xu, J., Li, X., Li, S., Wang, J., Palloix, A., Bosland, P.W., Li, Y., Krogh, A., Rivera-Bustamante, R.F., Herrera-Estrella, L., Yin, Y., Yu, J., Hu, K., Zhang, Z., 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proceedings of the National Academy of Sciences 111, 5135–5140. https://doi.org/10.1073/pnas.1400975111

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, Junyi, Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., Auvil, L., Capitanu, B., Ma, J., Lewin, H.A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang, K., Xia, J., Liao, S., Pan, S., Lu, X., Hou, H., Wang, Y., Zang, X., Yin, Y., Ma, H., Zhang, J., Wang, Z., Zhang, Yingmei, Zhang, D., Yonezawa, T., Hasegawa, M., Zhong, Y., Liu, W., Zhang, Yan, Huang, Z., Zhang, S., Long, R., Yang, H., Wang, Jian, Lenstra, J.A., Cooper, D.N., Wu, Y., Wang, Jun, Shi, P., Wang, Jian, Liu, J., 2012. The yak genome and adaptation to life at high altitude. Nat Genet 44, 946–949. https://doi.org/10.1038/ng.2343

Qu, T., Liu, R., Wang, W., An, L., Chen, T., Liu, G., Zhao, Z., 2011. Brassinosteroids regulate pectin methylesterase activity and AtPME41 expression in Arabidopsis under chilling stress. Cryobiology 63, 111–117. https://doi.org/10.1016/j.cryobiol.2011.07.003

QU, T., Liu, R., Wang, W., An, L., Chen, T., Liu, G., Zhao, Z., 2011. Brassinosteroids regulate pectin methylesterase activity and AtPME41 expression in Arabidopsis under chilling stress. Cryobiology 63, 111–117. https://doi.org/10.1016/j.cryobiol.2011.07.003

Ramsey, J., Schemske, D.W., 1998. Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annual Review of Ecology and Systematics 29, 467–501.

Rannala, B., Yang, Z., 2008. Phylogenetic Inference Using Whole Genomes. Annual review of genomics and human genetics 9, 217–31. https://doi.org/10.1146/annurev.genom.9.081307.164407

Ratcliffe, O.J., Kumimoto, R.W., Wong, B.J., Riechmann, J.L., 2003. Analysis of the Arabidopsis MADS AFFECTING FLOWERING Gene Family: MAF2 Prevents Vernalization by Short Periods of Cold [W]. The Plant Cell 15, 1159–1169. https://doi.org/10.1105/tpc.009506

Rayner, M.J., Hauber, M.E., Imber, M.J., Stamp, R.K., Clout, M.N., 2007. Spatial heterogeneity of mesopredator release within an oceanic island system. Proceedings of the National Academy of Sciences 104, 20862–20865. https://doi.org/10.1073/pnas.0707414105

Rehmany, A.P., Gordon, A., Rose, L.E., Allen, R.L., Armstrong, M.R., Whisson, S.C., Kamoun, S., Tyler, B.M., Birch, P.R.J., Beynon, J.L., 2005. Differential Recognition of Highly Divergent Downy Mildew Avirulence Gene Alleles by RPP1 Resistance Genes from Two Arabidopsis Lines. The Plant Cell 17, 1839–1850. https://doi.org/10.1105/tpc.105.031807

Reuter, J.A., Spacek, D.V., Snyder, M.P., 2015. High-Throughput Sequencing Technologies. Molecular Cell 58, 586–597. https://doi.org/10.1016/j.molcel.2015.05.004

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Reyes-Betancort, J.A., Guerra, A.S., Guma, I.R., Humphries, C.J., Carine, M.A., 2008. Diversity, rarity and the evolution and conservation of the Canary Islands endemic flora. Anales del Jardín Botánico de Madrid 65, 25–45. https://doi.org/10.3989/ajbm.2008.v65.i1.244

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, SI: Metagenomics of Marine Environments 13, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Rice, W.R., Hostert, E.E., 1993. Laboratory Experiments on Speciation: What Have We Learned in 40 Years? Evolution 47, 1637–1653. https://doi.org/10.1111/j.1558-5646.1993.tb01257.x

Richards, E.J., Servedio, M.R., Martin, C.H., 2019. Searching for Sympatric Speciation in the Genomic Era. BioEssays 41, 1900047. https://doi.org/10.1002/bies.201900047

Richardson, D.M., Whittaker, R.J., 2010. Conservation biogeography – foundations, concepts and challenges. Diversity and Distributions 16, 313–320. https://doi.org/10.1111/j.1472-4642.2010.00660.x

Rieseberg, L., Soltis, D., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5.

Rieseberg, L.H., Blackman, B.K., 2010. Speciation genes in plants. Annals of Botany 106, 439–455. https://doi.org/10.1093/aob/mcq126

Rieseberg, L.H., Willis, J.H., 2007. Plant Speciation. Science 317, 910–914. https://doi.org/10.1126/science.1137729

Robichaux, R.H., Carr, G.D., Liebman, M., Pearcy, R.W., 1990. Adaptive Radiation of the Hawaiian Silversword Alliance (Compositae- Madiinae): Ecological, Morphological, and Physiological Diversity. Annals of the Missouri Botanical Garden 77, 64–72. https://doi.org/10.2307/2399626

Rodríguez-Rodríguez, P., Fernández de Castro, A.G., Pérez de Paz, P.L., Curbelo, L., Palomares, Á., Mesa, R., Acevedo, A., Sosa, P.A., 2022. Evolution and conservation genetics of an insular hemiparasitic plant lineage at the limit of survival: the case of Thesium sect. Kunkeliella in the Canary Islands. American Journal of Botany 109, 419–436. https://doi.org/10.1002/ajb2.1830

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804. https://doi.org/10.1038/nature02053

Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H.H., El Taher, A., Indermaur, A., Malinsky, M., Ricci, V., Kahmen, A., Jentoft, S., Salzburger, W., 2021. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature 589, 76–81. https://doi.org/10.1038/s41586-020-2930-4

Rosloski, S.M., Singh, A., Jali, S.S., Balasubramanian, S., Weigel, D., Grbic, V., 2013. Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of Arabidopsis thaliana. Plant Mol Biol 81, 57–69. https://doi.org/10.1007/s11103-012-9982-2

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., Evanno, G., 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular Ecology 26, 142–162. https://doi.org/10.1111/mec.13664

Rousseau-Gueutin, M., Belser, C., Silva, C.D., Richard, G., Istace, B., Cruaud, C., Falentin, C., Boideau, F., Boutte, J., Delourme, R., Deniot, G., Engelen, S., de Carvalho, J.F., Lemainque, A., Maillet, L., Morice, J., Wincker, P., Denoeud, F., Chèvre, A.-M., Aury, J.-M., 2020. Long-reads assembly of the *Brassica napus* reference genome, Darmor-bzh (preprint). Genomics. https://doi.org/10.1101/2020.07.22.215749

Rowe, H.C., Renaut, S., Guggisberg, A., 2011. RAD in the realm of next-generation sequencing technologies. Molecular Ecology 20, 3499–3502. https://doi.org/10.1111/j.1365-294X.2011.05197.x

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring Phylogenies from RAD Sequence Data. PLOS ONE 7, e33394. https://doi.org/10.1371/journal.pone.0033394

Rubin, C.-J., Megens, H.-J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A.M.,

Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. Proceedings of the National Academy of Sciences 109, 19529–19536. https://doi.org/10.1073/pnas.1217149109

Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591. https://doi.org/10.1038/nature08832

Rundell, R.J., Price, T.D., 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends Ecol Evol 24, 394–399. https://doi.org/10.1016/j.tree.2009.02.007

Rundle, H.D., Nosil, P., 2005. Ecological speciation. Ecology Letters 8, 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Salichos, L., Rokas, A., 2013a. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331. https://doi.org/10.1038/nature12130

Salichos, L., Rokas, A., 2013b. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331. https://doi.org/10.1038/nature12130

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., Reich, D., 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507, 354–357. https://doi.org/10.1038/nature12961

Savolainen, O., Lascoux, M., Merilä, J., 2013. Ecological genomics of local adaptation. Nat Rev Genet 14, 807–820. https://doi.org/10.1038/nrg3522

Savolainen, V., Anstett, M.-C., Lexer, C., Hutton, I., Clarkson, J.J., Norup, M.V., Powell, M.P., Springate, D., Salamin, N., Baker, W.J., 2006. Sympatric speciation in palms on an oceanic island. Nature 441, 210–213. https://doi.org/10.1038/nature04566

Schatz, M.C., Delcher, A.L., Salzberg, S.L., 2010. Assembly of large genomes using second-generation sequencing. Genome Res. 20, 1165–1173. https://doi.org/10.1101/gr.101360.109

Schilling, M.P., Gompert, Z., Li, F.-W., Windham, M.D., Wolf, P.G., 2018. Admixture, evolution, and variation in reproductive isolation in the Boechera puberula clade. BMC Evolutionary Biology 18, 61. https://doi.org/10.1186/s12862-018-1173-6

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schluter, D., 2009. Evidence for Ecological Speciation and Its Alternative. Science 323, 737–741. https://doi.org/10.1126/science.1160006

Schluter, D., 2000. The Ecology of Adaptive Radiation. OUP Oxford.

Schluter, D., McPhail, J.D., 1992. Ecological Character Displacement and Speciation in Sticklebacks. The American Naturalist 140, 85–108. https://doi.org/10.1086/285404

Schmickl, R., Koch, M.A., 2011. Arabidopsis hybrid speciation processes. Proceedings of the National Academy of Sciences 108, 14192–14197. https://doi.org/10.1073/pnas.1104212108

Schmincke, H.-U., 1976. The Geology of the Canary Islands, in: Kunkel, G. (Ed.), Biogeography and Ecology in the Canary Islands, Monographiae Biologicae. Springer Netherlands, Dordrecht, pp. 67–184. https://doi.org/10.1007/978-94-010-1566-0_4

Schranz, M.E., Quijada, P., Sung, S.-B., Lukens, L., Amasino, R., Osborn, T.C., 2002. Characterization and Effects of the Replicated Flowering Time Gene FLC in Brassica rapa. Genetics 162, 1457–1468. https://doi.org/10.1093/genetics/162.3.1457

Schrempf, D., Szöllösi, G., 2020. The Sources of Phylogenetic Conflicts, in: Scornavacca, C., Delsuc, F., Galtier, N. (Eds.), Phylogenetics in the Genomic Era. No commercial publisher | Authors open access book, p. 3.1:1-3.1:23.

Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. Nat Methods 5, 16–18. https://doi.org/10.1038/nmeth1156

Scortecci, K.C., Michaels, S.D., Amasino, R.M., 2001. Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. Plant J 26, 229–236. https://doi.org/10.1046/j.1365-313x.2001.01024.x

Seehausen, O., 2013. Conditions when hybridization might predispose populations for adaptive radiation. Journal of Evolutionary Biology 26, 279–281. https://doi.org/10.1111/jeb.12026

Seehausen, O., 2004. Hybridization and adaptive radiation. Trends in Ecology & Evolution 19, 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Shafer, A.B.A., Gattepaille, L.M., Stewart, R.E.A., Wolf, J.B.W., 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. Molecular Ecology 24, 328–345. https://doi.org/10.1111/mec.13034

Sharma, N., Cram, D., Huebert, T., Zhou, N., Parkin, I.A.P., 2007. Exploiting the wild crucifer Thlaspi arvense to identify conserved and novel genes expressed during a plant's response to cold stress. Plant Mol Biol 63, 171–184. https://doi.org/10.1007/s11103-006-9080-4

Silvertown, J., 2004. The Ghost of Competition Past in the Phylogeny of Island Endemic Plants. Journal of Ecology 92, 168–173.

Silvertown, J., Francisco-Ortega, J., Carine, M., 2005. The Monophyly of Island Radiations: An Evaluation of Niche Pre-Emption and Some Alternative Explanations. Journal of Ecology 93, 653–657.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sinha, R.P., Häder, D.P., 2002. UV-induced DNA damage and repair: a review. Photochem Photobiol Sci 1, 225–236. https://doi.org/10.1039/b201230h

Slatko, B.E., Gardner, A.F., Ausubel, F.M., 2018. Overview of Next-Generation Sequencing Technologies. Current Protocols in Molecular Biology 122, e59. https://doi.org/10.1002/cpmb.59

Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K., Wang, W., Mandáková, T., Vello, E., Smith, L.M., Henz, S.R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T.T., Blanchette, M., Clark, R.M., Quesneville, H., Nordborg, M., Gaut, B.S., Lysak, M.A., Jenkins, J., Grimwood, J., Chapman, J., Prochnik, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D., Wright, S.I.,

2013a. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nature Genetics 45, 831–835. https://doi.org/10.1038/ng.2669

Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K., Wang, W., Mandáková, T., Vello, E., Smith, L.M., Henz, S.R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T.T., Blanchette, M., Clark, R.M., Quesneville, H., Nordborg, M., Gaut, B.S., Lysak, M.A., Jenkins, J., Grimwood, J., Chapman, J., Prochnik, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D., Wright, S.I., 2013b. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat Genet 45, 831–835. https://doi.org/10.1038/ng.2669

Smith, K.F., Sax, D.F., Lafferty, K.D., 2006. Evidence for the Role of Infectious Disease in Species Extinction and Endangerment. Conservation Biology 20, 1349–1357. https://doi.org/10.1111/j.1523-1739.2006.00524.x

Soltis, P.S., Soltis, D.E., 2009. The Role of Hybridization in Plant Speciation. Annu. Rev. Plant Biol. 60, 561–588. https://doi.org/10.1146/annurev.arplant.043008.092039

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.-Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.-Y., Chen, L.-L., Guo, L., 2020a. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat. Plants 6, 34–45. https://doi.org/10.1038/s41477-019-0577-7

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.-Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.-Y., Chen, L.-L., Guo, L., 2020b. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat. Plants 6, 34–45. https://doi.org/10.1038/s41477-019-0577-7

Staal, J., Kaliff, M., Bohman, S., Dixelius, C., 2006. Transgressive segregation reveals two Arabidopsis TIR-NB-LRR resistance genes effective against Leptosphaeria maculans, causal agent of blackleg disease. The Plant Journal 46, 218–230. https://doi.org/10.1111/j.1365-313X.2006.02688.x

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 34, W435–W439. https://doi.org/10.1093/nar/gkl200

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., Slate, J., 2010. Adaptation genomics: the next generation. Trends in Ecology & Evolution 25, 705–712. https://doi.org/10.1016/j.tree.2010.09.002

Steinbauer, M.J., Beierkuhnlein, C., 2010. Characteristic Pattern of Species Diversity on the Canary Islands. Erdkunde 64, 57–71.

Steinbauer, M.J., Field, R., Grytnes, J.-A., Trigas, P., Ah-Peng, C., Attorre, F., Birks, H.J.B., Borges, P.A.V., Cardoso, P., Chou, C.-H., De Sanctis, M., de Sequeira, M.M., Duarte, M.C., Elias, R.B., Fernández-Palacios, J.M., Gabriel, R., Gereau, R.E., Gillespie, R.G., Greimler, J., Harter, D.E.V., Huang, T.-J., Irl, S.D.H., Jeanmonod, D., Jentsch, A., Jump, A.S., Kueffer, C., Nogué, S., Otto, R., Price, J., Romeiras, M.M., Strasberg, D., Stuessy, T., Svenning, J.-C., Vetaas, O.R., Beierkuhnlein, C., 2016. Topography-driven isolation, speciation and a global increase of endemism with elevation. Global Ecology and Biogeography 25, 1097–1107. https://doi.org/10.1111/geb.12469

Steinbauer, M.J., Irl, S.D.H., Beierkuhnlein, C., 2013. Elevation-driven ecological isolation promotes diversification on Mediterranean islands. Acta Oecologica 47, 52–56. https://doi.org/10.1016/j.actao.2012.11.004

Steinbauer, M.J., Irl, S.D.H., González-Mancebo, J.M., Breiner, F.T., Hernández-Hernández, R., Hopfenmüller, S., Kidane, Y., Jentsch, A., Beierkuhnlein, C., 2017. Plant invasion and speciation along elevational gradients on the oceanic island La Palma, Canary Islands. Ecology and Evolution 7, 771–779. https://doi.org/10.1002/ece3.2640

Steinbauer, M.J., Otto, R., Naranjo-Cigala, A., Beierkuhnlein, C., Fernández-Palacios, J.-M., 2012. Increase of island endemism with altitude – speciation processes on oceanic islands. Ecography 35, 23–32. https://doi.org/10.1111/j.1600-0587.2011.07064.x

Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., Purugganan, M.D., Schmitt, J., 2004. A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. Proceedings of the National Academy of Sciences 101, 4712–4717. https://doi.org/10.1073/pnas.0306401101

Straiton, J., Free, T., Sawyer, A., Martin, J., 2019. From Sanger sequencing to genome databases and beyond. BioTechniques 66, 60–63. https://doi.org/10.2144/btn-2019-0011

Strasburg, J.L., Sherman, N.A., Wright, K.M., Moyle, L.C., Willis, J.H., Rieseberg, L.H., 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? Philosophical Transactions of the Royal Society B: Biological Sciences 367, 364–373. https://doi.org/10.1098/rstb.2011.0199

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany 99, 349–364. https://doi.org/10.3732/ajb.1100335

Strid, Å., Chow, W.S., Anderson, J.M., 1994. UV-B damage and protection at the molecular level in plants. Photosynth Res 39, 475–489. https://doi.org/10.1007/BF00014600

Stuart, Y.E., 2019. Divergent Uses of "Parallel Evolution" during the History of The American Naturalist. The American Naturalist 193, 11–19. https://doi.org/10.1086/700718

Stuessy, T.F., Jakubowsky, G., Gómez, R.S., Pfosser, M., Schlüter, P.M., Fer, T., Sun, B.-Y., Kato, H., 2006. Anagenetic evolution in island plants. Journal of Biogeography 33, 1259–1265. https://doi.org/10.1111/j.1365-2699.2006.01504.x

Suda, J., KYNCL, T., FREIOVÁ, R., 2003. Nuclear DNA Amounts in Macaronesian Angiosperms. Annals of Botany 92, 153–164. https://doi.org/10.1093/aob/mcg104

Sun, Y., Lu, Z., Zhu, X., Ma, H., 2020. Genomic basis of homoploid hybrid speciation within chestnut trees. Nat Commun 11, 3375. https://doi.org/10.1038/s41467-020-17111-w

Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research 34, W609–W612. https://doi.org/10.1093/nar/gkl315

Svardal, H., Quah, F.X., Malinsky, M., Ngatunga, B.P., Miska, E.A., Salzburger, W., Genner, M.J., Turner, G.F., Durbin, R., 2020. Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. Molecular Biology and Evolution 37, 1100–1113. https://doi.org/10.1093/molbev/msz294

Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3, 137–144. https://doi.org/10.1038/nrg733

Tabor, H.K., Risch, N.J., Myers, R.M., 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3, 391–397. https://doi.org/10.1038/nrg796

Tadege, M., Sheldon, C.C., Helliwell, C.A., Stoutjesdijk, P., Dennis, E.S., Peacock, W.J., 2001. Control of flowering time by FLC orthologues in Brassica napus. The Plant Journal 28, 545–553. https://doi.org/10.1046/j.1365-313X.2001.01182.x

Talavera, M., Navarro-Sampedro, L., Ortiz, P.L., Arista, M., 2013. Phylogeography and seed dispersal in islands: the case of Rumex bucephalophorus subsp. canariensis (Polygonaceae). Annals of Botany 111, 249–260. https://doi.org/10.1093/aob/mcs284

Tang, T., Yu, A., Li, P., Yang, H., Liu, G., Liu, L., 2016. Sequence analysis of the Hsp70 family in moss and evaluation of their functions in abiotic stress responses. Sci Rep 6, 33650. https://doi.org/10.1038/srep33650

Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current Protocols in Bioinformatics 25, 4.10.1-4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

Tevini, M., Teramura, A.H., 1989. Uv-B Effects on Terrestrial Plants. Photochemistry and Photobiology 50, 479–487. https://doi.org/10.1111/j.1751-1097.1989.tb05552.x

Than, C., Ruths, D., Nakhleh, L., 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9, 322. https://doi.org/10.1186/1471-2105-9-322

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796–815. https://doi.org/10.1038/35048692

Theißen, G., Rümpler, F., Gramzow, L., 2018. Array of MADS-Box Genes: Facilitator for Rapid Adaptation? Trends in Plant Science 23, 563–576. https://doi.org/10.1016/j.tplants.2018.04.008

Theocharis, A., Clément, C., Barka, E.A., 2012. Physiological and molecular changes in plants grown at low temperatures. Planta 235, 1091–1105. https://doi.org/10.1007/s00425-012-1641-y

Tiwari, L.D., Khungar, L., Grover, A., 2020. AtHsc70-1 negatively regulates the basal heat tolerance in Arabidopsis thaliana through affecting the activity of HsfAs and Hsp101. The Plant Journal 103, 2069–2083. https://doi.org/10.1111/tpj.14883

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T., Nuzhdin, S.V., 2010. Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. Nat Genet 42, 260–263. https://doi.org/10.1038/ng.515

Uy, J.A.C., Moyle, R.G., Filardi, C.E., 2009. Plumage and Song Differences Mediate Species Recognition Between Incipient Flycatcher Species of the Solomon Islands. Evolution 63, 153–164. https://doi.org/10.1111/j.1558-5646.2008.00530.x

van Hengstum, T., Lachmuth, S., Oostermeijer, J.G.B., den Nijs, H. (J.) C.M., Meirmans, P.G., van Tienderen, P.H., 2012. Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. Plant Syst Evol 298, 1119–1131. https://doi.org/10.1007/s00606-012-0624-6

van Huysduynen, A.H., Janssens, S., Merckx, V., Vos, R., Valente, L., Zizka, A., Larter, M., Karabayir, B., Maaskant, D., Witmer, Y., Fernández-Palacios, J.M., de Nascimento, L., Jaén, R.M.,

Castells, J.C., Marrero-Rodríguez, Á., del Arco, M., Lens, F., 2020. Multiple origins of insular woodiness on the Canary Islands are consistent with palaeoclimatic aridification (preprint). Evolutionary Biology. https://doi.org/10.1101/2020.05.09.084582

Vargas, P., 2007. Are Macaronesian islands refugia of relict plant lineages?: a molecular survey, in: Weiss, S., Ferrand, N. (Eds.), Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity. Springer Netherlands, Dordrecht, pp. 297–314. https://doi.org/10.1007/1-4020-4904-8_11

Vellosillo, T., Martínez, M., López, M.A., Vicente, J., Cascón, T., Dolan, L., Hamberg, M., Castresana, C., 2007. Oxylipins Produced by the 9-Lipoxygenase Pathway in Arabidopsis Regulate Lateral Root Development and Defense Responses through a Specific Signaling Cascade. The Plant Cell 19, 831–846. https://doi.org/10.1105/tpc.106.046052

Venkat, A., Muneer, S., 2022. Role of Circadian Rhythms in Major Plant Metabolic and Signaling Pathways. Frontiers in Plant Science 13.

Verdaguer, D., Jansen, M.A.K., Llorens, L., Morales, L.O., Neugart, S., 2017. UV-A radiation effects on higher plants: Exploring the known unknown. Plant Science 255, 72–81. https://doi.org/10.1016/j.plantsci.2016.11.014

Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L., Jenkins, J., Vendramin, E., Meisel, L.A., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., Policriti, A., Cipriani, G., Dondini, L., Ficklin, S., Goodstein, D.M., Xuan, P., Fabbro, C.D., Aramini, V., Copetti, D., Gonzalez, S., Horner, D.S., Falchi, R., Lucas, S., Mica, E., Maldonado, J., Lazzari, B., Bielenberg, D., Pirona, R., Miculan, M., Barakat, A., Testolin, R., Stella, A., Tartarini, S., Tonutti, P., Arús, P., Orellana, A., Wells, C., Main, D., Vizzotto, G., Silva, H., Salamini, F., Schmutz, J., Morgante, M., Rokhsar, D.S., 2013. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45, 487–494. https://doi.org/10.1038/ng.2586

Vernot, B., Akey, J.M., 2014. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. Science 343, 1017–1021. https://doi.org/10.1126/science.1245938

Villa-Machío, I., Fernández de Castro, A.G., Fuertes-Aguilar, J., Nieto Feliner, G., 2020. Colonization history of the Canary Islands endemic Lavatera acerifolia, (Malvaceae) unveiled with genotyping-by-sequencing data and niche modelling. Journal of Biogeography 47, 993–1005. https://doi.org/10.1111/jbi.13808

Wagner, N.D., Gramlich, S., Hörandl, E., 2018. RAD sequencing resolved phylogenetic relationships in European shrub willows (Salix L. subg. Chamaetia and subg. Vetrix) and revealed multiple evolution of dwarf shrubs. Ecology and Evolution 8, 8243–8255. https://doi.org/10.1002/ece3.4360

Walden, N., German, D.A., Wolf, E.M., Kiefer, M., Rigault, P., Huang, X.-C., Kiefer, C., Schmickl, R., Franzke, A., Neuffer, B., Mummenhoff, K., Koch, M.A., 2020. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat Commun 11, 3795. https://doi.org/10.1038/s41467-020-17605-7

Waltz, E., 2018. With a free pass, CRISPR-edited plants reach market in record time. Nature Biotechnology 36, 6–8.

Wang, M., Zhao, Y., Zhang, B., 2015. Efficient Test and Visualization of Multi-Set Intersections. Sci Rep 5, 16923. https://doi.org/10.1038/srep16923

Warren, B.H., Simberloff, D., Ricklefs, R.E., Aguilée, R., Condamine, F.L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J.M., Hengl, T., Norder, S.J., Rijsdijk, K.F., Sanmartín, I., Strasberg, D., Triantis, K.A., Valente, L.M., Whittaker, R.J., Gillespie, R.G., Emerson, B.C., Thébaud, C., 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecology Letters 18, 200–217. https://doi.org/10.1111/ele.12398

White, O.W., Doo, B., Carine, M.A., Chapman, M.A., 2016. Transcriptome sequencing and simple sequence repeat marker development for three Macaronesian endemic plant species. Applications in Plant Sciences 4, 1600050. https://doi.org/10.3732/apps.1600050

White, O.W., Reyes-Betancort, A., Chapman, M.A., Carine, M.A., 2018. Independent homoploid hybrid speciation events in the Macaronesian endemic genus Argyranthemum. Molecular Ecology 27, 4856–4874. https://doi.org/10.1111/mec.14889

White, O.W., Reyes-Betancort, J.A., Chapman, M.A., Carine, M.A., 2020. Geographical isolation, habitat shifts and hybridisation in the diversification of the Macaronesian endemic genus Argyranthemum (Asteraceae). New Phytologist 228, 1953–1971. https://doi.org/10.1111/nph.16980

Whitlock, M.C., Mccauley, D.E., 1999. Indirect measures of gene flow and migration: FST≠1/(4Nm+1). Heredity 82, 117–125. https://doi.org/10.1046/j.1365-2540.1999.00496.x

Whittaker, R.J., Fernández-Palacios, J.M., Matthews, T.J., Borregaard, M.K., Triantis, K.A., 2017. Island biogeography: Taking the long view of nature's laboratories. Science 357, eaam8326. https://doi.org/10.1126/science.aam8326

Whittaker, R.J., Triantis, K.A., Ladle, R.J., 2008. ORIGINAL ARTICLE: A general dynamic theory of oceanic island biogeography. Journal of Biogeography 35, 977–994. https://doi.org/10.1111/j.1365-2699.2008.01892.x

Wickham, H., 2009. ggplot2. Springer, New York, NY. https://doi.org/10.1007/978-0-387-98141-3

Wiens, J.J., Graham, C.H., 2005. Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. Annual Review of Ecology, Evolution, and Systematics 36, 519–539.

Wilson, J.T., 1961. Continental and Oceanic Differentiation. Nature 192, 125–128. https://doi.org/10.1038/192125b0

Witter, M.S., Carr, G.D., 1988. Adaptive Radiation and Genetic Differentiation in the Hawaiian Silversword Alliance (compositae: Madiinae). Evolution 42, 1278–1287. https://doi.org/10.1111/j.1558-5646.1988.tb04187.x

Wolf, J.B.W., Lindell, J., Backström, N., 2010. Speciation genetics: current status and evolving approaches. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 1717–1733. https://doi.org/10.1098/rstb.2010.0023

Wu, C.-I., Ting, C.-T., 2004. Genes and speciation. Nat Rev Genet 5, 114–122. https://doi.org/10.1038/nrg1269

Xia, H., Camus-Kulandaivelu, L., Stephan, W., TELlier, A., Zhang, Z., 2010. Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. Molecular Ecology 19, 4144–4154. https://doi.org/10.1111/j.1365-294X.2010.04762.x

References

Yamaguchi-Shinozaki, K., Shinozaki, K., 1993. Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants. Molec. Gen. Genet. 236, 331–340. https://doi.org/10.1007/BF00277130

Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24, 1586–1591. https://doi.org/10.1093/molbev/msm088

Ye, C., Hill, C.M., Wu, S., Ruan, J., Ma, Z. (Sam), 2016. DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. Scientific Reports 6, 31900. https://doi.org/10.1038/srep31900

Yeaman, S., Hodgins, K.A., Suren, H., Nurkowski, K.A., Rieseberg, L.H., Holliday, J.A., Aitken, S.N., 2014. Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (Pinus contorta) and interior spruce (Picea glauca × Picea engelmannii). New Phytologist 203, 578–591. https://doi.org/10.1111/nph.12819

Yu, L., Wang, G.-D., Ruan, J., Chen, Y.-B., Yang, C.-P., Cao, X., Wu, H., Liu, Y.-H., Du, Z.-L., Wang, X.-P., Yang, J., Cheng, S.-C., Zhong, L., Wang, L., Wang, X., Hu, J.-Y., Fang, L., Bai, B., Wang, K.-L., Yuan, N., Wu, S.-F., Li, B.-G., Zhang, J.-G., Yang, Y.-Q., Zhang, C.-L., Long, Y.-C., Li, H.-S., Yang, J.-Y., Irwin, D.M., Ryder, O.A., Li, Y., Wu, C.-I., Zhang, Y.-P., 2016. Genomic analysis of snub-nosed monkeys (Rhinopithecus) identifies genes and processes related to high-altitude adaptation. Nat Genet 48, 947–952. https://doi.org/10.1038/ng.3615

Zeng, X., Yuan, H., Dong, X., Peng, M., Jing, X., Xu, Q., Tang, T., Wang, Y., Zha, S., Gao, M., Li, C., Shu, C., Wei, Z., Qimei, W., Basang, Y., Dunzhu, J., Li, Z., Bai, L., Shi, J., Zheng, Z., Yu, S., Fernie, A.R., Luo, J., Nyima, T., 2020. Genome-wide Dissection of Co-selected UV-B Responsive Pathways in the UV-B Adaptation of Qingke. Molecular Plant 13, 112–127. https://doi.org/10.1016/j.molp.2019.10.009

Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Molecular Biology and Evolution 22, 2472–2479. https://doi.org/10.1093/molbev/msi237

Zhang, Jing, Tian, Y., Yan, L., Zhang, G., Wang, X., Zeng, Y., Zhang, Jiajin, Ma, X., Tan, Y., Long, N., Wang, Y., Ma, Y., He, Y., Xue, Y., Hao, S., Yang, S., Wang, W., Zhang, L., Dong, Y., Chen, W., Sheng, J., 2016. Genome of Plant Maca (Lepidium meyenii) Illuminates Genomic Basis for High-Altitude Adaptation in the Central Andes. Molecular Plant 9, 1066–1077. https://doi.org/10.1016/j.molp.2016.04.016

Zhang, S., Wu, Q.-R., Liu, L.-L., Zhang, H.-M., Gao, J.-W., Pei, Z.-M., 2020. Osmotic stress alters circadian cytosolic Ca2+ oscillations and OSCA1 is required in circadian gated stress adaptation. Plant Signaling & Behavior 15, 1836883. https://doi.org/10.1080/15592324.2020.1836883

Zhang, T., Qiao, Q., Novikova, P.Yu., Wang, Q., Yue, J., Guan, Y., Ming, S., Liu, T., De, J., Liu, Y., Al-Shehbaz, I.A., Sun, H., Van Montagu, M., Huang, J., Van de Peer, Y., Qiong, L., 2019. Genome of Crucihimalaya himalaica, a close relative of Arabidopsis, shows ecological adaptation to high altitude. Proceedings of the National Academy of Sciences 116, 7137–7146. https://doi.org/10.1073/pnas.1817580116

Zhao, D., Xia, X., Su, J., Wei, M., Wu, Y., Tao, J., 2019. Overexpression of herbaceous peony HSP70 confers high temperature tolerance. BMC Genomics 20, 70. https://doi.org/10.1186/s12864-019-5448-0

Zhou, Y., Zhang, L., Liu, J., Wu, G., Savolainen, O., 2014. Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. Molecular Ecology 23, 3504–3522. https://doi.org/10.1111/mec.12830

Zhu, M., Zhao, S., 2007. Candidate Gene Identification Approach: Progress and Challenges. Int J Biol Sci 3, 420–427.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A., 2013. The MaSuRCA genome assembler. Bioinformatics 29, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Zimin, A.V., Salzberg, S.L., 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLOS Computational Biology 16, e1007981. https://doi.org/10.1371/journal.pcbi.1007981