Hybridization in *Helianthus*:

The genomic profiles of potential and confirmed sunflower hybrid species

by

Gregory Lawrence Owens

B.Sc., The University of Victoria, 2008

M.Sc., The University of Victoria, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2016

Abstract

Hybridization is an important evolutionary force that acts in both constructive and destructive ways. It can both swamp out rare species and create new ones. To better understand these effects I studied hybridization within the sunflower genus *Helianthus* from three angles. First, I used a rich literature of artificial crossing experiments in *Helianthus* and Madiinae to ask how fast reproductive isolation evolves and what features affect its accumulation. I show that hybrid sterility can evolve quickly and is faster in annuals than in perennials. I then examine a classic case of introgression involving *Helianthus bolanderi*. I use modern genomic tools to show that it is not of hybrid origin and likely not a separate species from its congener *H. exilis*. We do however find introgression with the invading species, *H. annuus.* In agreement with theory, we find that gene flow is mainly into the invading species. Lastly, I use transcriptomic data for three established homoploid hybrid species, *H. anomalus*, *H. deserticola*, and *H. paradoxus*, and their parents *H. annuus* and *H. petiolaris* to map the genomic composition of hybrid species. I show that composition is even or biased towards *H. petiolaris*. Hybrid genomes are highly recombined but are more similar in genomic composition than expected by chance, suggesting the work of selection. Furthermore, although analyses of genetic distance between the hybrid species and their parents suggests that the hybrids are older than previously appreciated, they do

not appear to be fully stabilized. Lastly two of the species, *H. anomalus* and *H. deserticola*, may share a common origin. Future directions include mapping introgression in *H. annuus*, and modeling parental block size to determine the number of loci and strength of selection during hybrid speciation.

Preface

I designed and ran the analyses in chapter 2 with consultation from L.H. Rieseberg. All data were taken from publically available sources as listed in the Appendix A. A version of this work has been published in Evolution:

- Owens GL, Rieseberg LH (2014) Hybrid incompatibility is acquired faster in annual than in perennial species of sunflower and tarweed. *Evolution*, **68** 893-900

I designed the study, collected the data, and ran the analyses, in consultation with L.H. Rieseberg, for chapter 3. G.J. Baute and D.G. Bock provided additional sequence data. K. Samuk, K.L. Ostevik and B.T. Moyers provided help collecting seeds. T. Gulya and L. Marek supplied collection location information. Seed collections were supplied by Jake Schweitzer and the USDA GRIN. A version of this work is accepted at Molecular Ecology:

- Owens GL, Baute GJ, Rieseberg LH (2016) Revisiting a classic case of introgression: Hybridization and gene flow in the Californian sunflowers. *Molecular Ecology*, **In press**

I designed the study and ran the analyses, in consultation with L.H. Rieseberg, for chapter 4. Sequence data were taken from previously published resources within the

Rieseberg lab. Sally Otto contributed to the creation of the windowed ancestry algorithm.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of variance
DM	Dobzhansky-Muller
DNA	Deoxyribonucleic acid
cM	Centimorgan
F_{ST}	Fixation index
F_{IS}	Inbreeding coefficient
LG	Linkage group (or chromosome)
MAF	Minor allele frequency
MQ	Mapping quality
NGRP	National Genetic Resources Program
Qual	Quality
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
USDA	United States Department of Agriculture
ybp	Years before present

Acknowledgements

This project would never have been completed, or have been completed much more poorly, if it were not for the support and engagement of a number of friends and colleagues.

- My office mates Kieran Samuk* and Brook Moyers*, who spent innumerable hours discussing evolutionary biology with me over five years.
- My longtime colleague Diana Rennison, who does great science and drags me along for the ride.
- My fellow grad students Kate Ostevik*, Kathryn Turner, Greg Baute, Chris Grassa, Emily Drummond and Dan Bock.
- My colleagues Heather Rowe, Josh Chang Mell, Rose Andrew, Dan Ebert, Sebastien Renaut, Kay Hodgins, Kristin Nurkowski and Sam Yeaman.
- Of course, my lovely girlfriend, Virginia Woloshen.
- Special thanks to those * above who helped me collect seeds across California.

I thank my parents for their steadfast support and faith in my abilities and for never asking when I would graduate.

My work was partially supported by an NSERC CGS-D scholarship.

I thank my supervisory committee, Dolph Schluter, Keith Adams and Quentin Cronk. Lastly, I thank my supervisor Loren Rieseberg, who led by example.

Dedication

I dedicate this work to the march of technology.

Only through backbreaking technological advancement can we discover how nature does it so easily.

Chapter 1: Introduction

1.1 Hybridization

Hybridization was long thought of as a destructive maladaptive force that had to be overcome for diversity to increase through speciation (Darwin 1859; Dobzhansky 1940; Mayr 1963). In this view, hybrids are evolutionary dead ends and selection favors preventing their production. In contrast to this, botanists have recognized the ubiquity of hybridization in plants and its potential for providing the raw material for adaptation (Anderson 1948; Stebbins 1959). Modern theoretical and empirical work has largely supported the botanical view that hybridization can play an important role in adaptive evolution and diversification (Abbott et al. 2013), although see Servedio et al. 2013 and Barton 2013. Furthermore, genomic analyses have uncovered evidence of hybridization in the evolutionary histories of a surprisingly large and diverse array of taxa (Heliconius Genome Consortium 2012; Jónsson et al. 2014; Fontaine et al. 2015). Thus to understand the evolutionary past and predict the evolutionary future, we need to understand the different roles hybridization can play.

1.1.1 What is hybridization?

Before further discussing hybridization, it is important to define it. For the purposes of this thesis, I define hybridization as the successful mating between individuals of two different named species based on a relaxed version of the biological species concept (sensu Coyne and Orr 2004). I'm using a relaxed version of the

biological species concept because under a strict interpretation all hybrids are sterile, which is not the case for the examples I discuss. Although I am defining hybridization conservatively, I recognize that others have defined hybridization in a more inclusive way that includes inter-population crosses (e.g. Harrison 1990 and Arnold 1996). It is likely that evolutionary phenomena often associated with hybridization such as outbreeding depression, heterosis, and reinforcement, will vary in strength and/or frequency depending on the degree of divergence between the hybridizing taxa, but there is no one discrete cut off point that can be used to predict the viability, sterility, or heterosis of hybrids.

One reason I do not use the more liberal definition of Arnold is that it turns almost all long distance mating events into hybridization. Arnold defines hybridization as *successful mating between individuals of two populations or groups of populations, which are distinguishable on the basis of one or more heritable characters.* Heritable characteristics include genetic markers, like SNPs, and even populations with low overall divergence (i.e., minimal but non-zero F_{ST}) can be distinguished genetically using large amounts of genetic data in aggregate. In the case of *Helianthus bolanderi-exilis*, matings between populations would be classified as hybridizations as well as matings with the related species *H. annuus* but the interspecific crosses involve significant sterility barriers that we don't expect to find in the inter-population crosses (See chapter 3). Thus although hybridization is a continuum, I focus on one end of that

continuum to avoid confounding hybridization with more general gene flow within a species.

1.1.2 Hybridization as a destructive force

Darwin regarded hybrids as being generally sterile and unimportant (Darwin 1859). Consistent with Darwin's viewpoint, the zoological literature has long regarded hybridization as an unfortunate side effect of the speciation process that is overcome through reproductive isolation (Dobzhansky 1940; Mayr 1963). In the case of many animals this view is accurate: hybrids are completely sterile and do not contribute to future generations. If the hybrids are not completely sterile, hybrids can also have reduced fitness due to partial sterility, intrinsic (e.g. hybrid necrosis (Bomblies and Weigel 2007)) or extrinsic inviability (e.g. ecological mismatch (Schluter 2000; Rundle and Whitlock 2001)).

Sustained hybridization can result in outbreeding depression, in which hybridization reduces individual or population fitness (Frankham et al. 2011). This can occur through the breaking up of co-adapted gene complexes or the bringing together of genetic incompatibilities (e.g. Dobzhanksy-Muller incompatibilities (Bateson 1909; Dobzhansky 1936; Muller 1942)). In the case where one species involved in hybridization is rare, hybridization can bring about extinction either through demographic swamping (i.e., where hybrids are infertile and the rare taxon wastes gametes on hybrid production) or genetic swamping (i.e., where hybrids are fertile and hybrids replace pure populations) (Wolf et al. 2001). Hybridization frequency can be

increased by anthropogenic habitat changes and is recognized as a mechanism by which species may be threatened (Chunco 2014). When hybridization is maladaptive it becomes adaptive to avoid interspecific matings. This process is called reinforcement and has been studied extensively, although definitive cases remain rare (Blair 1955; Butlin 1987; Hoskin et al. 2005; Hopkins and Rausher 2012).

These forces together paint a picture of hybridization as unimportant or purely negative; a mistake that species should avoid. But, this isn't the only side to the hybridization coin.

1.1.3 Hybridization as a constructive process

In contrast to its role as a destructive force, hybridization can also supply diversity to species or populations and even facilitate the creation of new species entirely. The importance of this is best illustrated in adaptive introgression (Anderson 1949). Adaptive introgression is demonstrated when a trait that is selectively favored in one species is caused by an allele that was acquired from a separate species. This has been seen in sunflowers as well as mice, Darwin's finches and butterflies (Whitney et al. 2010; Song et al. 2011; Grant and Grant 2011; Pardo-Diaz et al. 2012). This process could be quite important because it allows for the utilization of a whole suite of new alleles found in related species. For example, depending on divergence and effective population size, a single hybridization may bring in more novel alleles than all mutations in the entire population for a generation (Hedrick 2013). Unlike new mutations, these novel alleles are pretested and can be complicated (i.e., full

haplotypes instead of individual SNPs). On the other hand, introgressed alleles may be linked to negatively selected alleles (e.g. DM incompatibilities) and they start at low frequency but still they have large potential for kick starting evolutionary change.

When species ranges overlap, hybridization can occur in the overlap region, creating a hybrid zone. Most hybrid zones are best described by the tension zone model (Barton and Hewitt 1985), in which hybrids are less fit and are maintained by continuous dispersal pressure from the parental species. If the species ranges are determined by continuous environmental variables, however, then the hybrid zone may fall in an intermediate region that is at the range edge of each species. In this case hybrids may better fit the bounded hybrid superiority model and be more fit than their parents within the intermediate habitat (Moore 1977). This is most easily thought of when hybrids are intermediate between their parents in both phenotype and habit (e.g. the hybrid of an alpine and a lowlands species that is better suited to the midlands than either parent). Support for the bounded hybrid superiority model is not widespread but has been shown in several examples (Saino and Villa 1992; Wang et al. 1997; Good et al. 2000).

Hybrids need not be intermediate between the phenotypes of the parents; they can also exceed (or be inferior) to the trait values for either parent. The former is commonly seen in heterosis, where hybrids are more vigorous than their parents. Heterosis is thought to occur from dominance (where recessive deleterious alleles are masked in the hybrid), overdominance (heterozygote superiority) or epistasis (where

alleles at different loci interact to generate hybrid superiority) (Chen 2013). Heterosis is strongest in the F1 hybrid generation, where interspecific heterozygosity is highest, but extreme phenotypes are commonly produced in advanced generation hybrids through transgressive segregation (Rieseberg et al. 1999). In transgressive segregation, combinations of alleles at different loci produce phenotypes beyond the parents' phenotype range. For example, if two species each have alleles at three independent loci that make a plant taller, some segregants will have the "tall" alleles at all six loci and produce an extremely tall plant. This type of transgressive segregation has been shown to contribute to the formation of homoploid hybrid species (Rieseberg 2003).

1.1.4 Homoploid hybrid speciation

In homoploid hybrid speciation, the hybrids of two species become reproductively isolated from their parents. Some authors further argue that the reproductive isolation must be a consequence of hybridization for it to be considered hybrid speciation (Schumer et al. 2014). Although rare, in recent years more cases have been proposed in both plants and animals (Mavarez and Linares 2008; Schumer et al. 2014). The most difficult criterion to satisfy is that the reproductive isolation is derived from hybridization. In some cases, for example, a putative hybrid species may have genetic material from two species but the introgression occurred before or after reproductive isolation was acquired.

One reason homoploid hybrid speciation is rare is that it requires the parental species to be in close proximity for hybrids to form but is also inhibited by this

proximity because it encourages hybrids to backcross into the parental lineages. To become a new species, the hybrids must interbreed and not backcross into the parents; the three leading models to accomplish this are the recombinational speciation mechanism, the 'segregation of a new type isolated by external barriers' mechanism or the 'selection against genetic incompatibilities' mechanism (Grant 1981; Templeton 1981; Schumer et al. 2015). Recombinational speciation requires the parental species to have two or more chromosomal rearrangements. Although F_1s will be chromosomally unbalanced and have reduced fertility, subsequent generations can produce novel chromosomal combinations that are reproductively isolated from both parents. The second mechanism suggests that the segregation of novel combinations of alleles will allow the hybrids to invade a new niche that is geographically or ecologically isolated from the parents. Alternatively, the novel combination of alleles may produce a trait that results in assortative mating (e.g., flowering time divergence). It is possible that both of these mechanisms act together and, indeed, the three homoploid hybrid sunflower species are both chromosomally and ecologically isolated from their parents (Rieseberg et al. 1995). The final mechanism requires an isolated hybrid population segregating for multiple adaptive or coevolving genetic incompatibility pairs (Schumer et al. 2015). Selection against genetic incompatibilities can lead to the fixation of one parental version of a given incompatibility pair. If there are multiple such pairs, versions from different parents can sometimes be fixed leading

to fixed incompatibilities isolating the hybrid population from both parental populations.

1.1.5 Allopolyploid hybrid speciation

Unlike homoploid hybrid speciation, in allopolyploid hybrid speciation reproductive isolation is instantly acquired. Allopolyploid hybrid speciation is the production of a 4x organism that contains two copies of each parental species chromosomes. This can occur through somatic chromosome doubling in a diploid hybrid, the fusion of two unreduced gametes or through a triploid bridge (Soltis et al. 2004). Allopolyploids may initially have problems with chromosomal pairing in the meiosis leading to reduced fertility and few appropriate mates (Levin 1975; Husband 2000). Despite this, polyploidy is common in plants; between 15-30% of speciation events are a result of polyploidy (Wood et al. 2009). Allopolyploidy seems to be as common as autopolyploidy (genome doubling without hybridization), suggesting that this form of hybridization is broadly important to plant evolution (Barker et al. 2015).

1.1.6 The prevalence of hybridization

I have emphasized the large potential effects of hybridization but the overall importance of hybridization in evolution is dependent on how frequent hybridization is in nature. If species barriers are inviolate, then the potential costs and benefits of hybridization are null and void. At the individual level, hybrids are rare by definition. If two taxa produce copious hybrids, they are unlikely to be classified as different species based on most species concepts. Despite this, the percentage of species that hybridize

with at least one other species is surprisingly high. Mallet (2005) surveyed the literature for studies that estimated hybridization rates and found that up to 25% of plant species and 10% of animal species produce hybrids (Mallet 2005). Considering this is based on contemporary hybridization, the percentage of species that were influenced by hybridization in their recent evolutionary past may be significantly higher.

Only in recent years has the technology been available to detect ancient hybridization. This was shown most strikingly in humans whose ancestors hybridized with Neanderthals in Europe (Green et al. 2010). In *Anopheles* mosquitos, several hybridization events across the phylogeny have led to a scenario where only a portion of the X chromosome shows the true species phylogeny and the rest of the genome shows the false signal from introgression (Fontaine et al. 2015). Similarly, introgression has also been seen in the evolutionary past of horses, butterflies and cichlids (Heliconius Genome Consortium 2012; Keller et al. 2013; Jónsson et al. 2014). As phylogenetics moves to the genomic era, it may be that ancient hybridization becomes the norm instead of the exception.

1.2 Sunflowers as models for hybridization research

Several key advances in the study of hybridization have been based on studies of the sunflower genus, *Helianthus*. This genus, within the family Asteraceae, subfamily Asterioudeae, tribe Heliantheae and subtribe Helianthineae, contains 49 species, both annual and perennial, native to central North America (Panero and Funk 2002). The

common sunflower, *H. annuus*, is the most widespread species and is also the progenitor of the domestic sunflower, thus much of the research has focused on it and its close annual relatives.

Hybridization has been exploited to breed better domestic sunflowers. Cytoplasmic male sterility and the restorer of fertility allele, two traits that are necessary for commercial hybrid seed production, were introgressed from *H. petiolaris* (Leclercq 1969). Similarly, the branching trait found in pollen production lines is derived from *H. annuus ssp. texanus* (Baute et al. 2015). Despite strong reproductive barriers, *H. annuus* has been crossed to a wide variety of species within the same genus (e.g., Heiser 1951a; Jackson and Guard 1956; Heiser 1965; Jan 1997).

Sunflower species also hybridize frequently in nature. The common sunflower, *H. annuus*, is known to hybridize with *H. bolanderi*, *H. petiolaris*, *H. argophyllus*, and *H. debilis*, across its wide range (Heiser 1947,a,b; Rieseberg et al. 1990b; Carney et al. 2000). This is seen in Texas, where the local subspecies *H. annnus ssp. texanus* is a product of adaptive introgression from the *H. debilis* (Rieseberg et al. 1990b; Whitney et al. 2010). In California, invading *H. annuus* populations have replaced native *H. bolanderi* populations, possibly through genetic swamping (Carney et al. 2000). Hybrids between other annual species have also been found, although geographic isolation prevents many combinations that are possible introgression vectors based on artificial hybridization studies (Chandler et al. 1986). Similarly, hybrids have been found between different perennial species, including several confirmed or proposed

allopolyploids (Heiser and Smith 1964; Heiser et al. 1969; Timme et al. 2007; Bock et al. 2014)

Within the genus, there are also three homoploid hybrid species, *H. anomalus*, *H. deserticola* and *H. paradoxus*. Each is a product of hybridization between *H. annuus* and *H. petiolaris* relatively recently compared to other speciation events in the genus: *H. anomalus* 116,000 to 160,000 ybp, *H. deserticola* 63,000 to 170,000 ybp, and *H. paradoxus* 75,000 to 208,000 ybp (Schwarzbach and Rieseberg 2002; Welch and Rieseberg 2002b; Gross et al. 2003). Hybrid ancestry is based on molecular markers, as well as shared chromosomal rearrangements (Rieseberg et al. 1990a; Rieseberg 1991; Rieseberg et al. 1993; 1995). Ecologically, each of the hybrid species have diverged from the preferred parental environments into more extreme habitats; sand dune for *H. anomalus*, sand sheet for *H. deserticola*, and salt marsh for *H. paradoxus* (Heiser et al. 1969). Interestingly, the genome size of each of the hybrid species has expanded considerably (Baack et al. 2005). This seems to have occurred through the proliferation of transposable elements, although the cause of this proliferation is unknown (Staton et al. 2009).

Overall, *Helianthus* is an excellent genus to explore questions about hybridization. It exemplifies both the creative (hybrid speciation and adaptive introgression) and destructive (genetic swamping) consequences of hybridization. Due to the use of wild species as genetic donors to the domestic sunflower, strong

commercial interest exists in understanding the genetic diversity among species and how that diversity is being spread through hybridization.

1.3 What we don't know

Many questions remain to be answered about hybridization's role in evolution. For example, we do not have empirical estimates of the prevalence of adaptive introgression in nature. Introgressed alleles are pre-tested in an organism and bring in more variation than *de novo* mutations, but the prevalence of hybrid incompatibilities linked to adaptive loci may determine its actual utility to species (Hedrick 2013). Similarly, we do not know how large of a role hybridization plays in speciation. Although hybridization is increasingly being found in the evolutionary past, we do not know if the hybridization played a role in the actual speciation events themselves.

With regard to species conservation, we need a better understanding of the dangers and benefits of hybridization. Hybridization can threaten rare species through outbreeding depression or swamping, but it can also effectively alleviate inbreeding depression (Rhymer and Simberloff 1996; Brennan et al. 2015). The likelihood of these outcomes will be affected by both the demography of the parental species as well as the directionality of introgression due to hybridization (Currat et al. 2008). Understanding when these alternate scenarios are likely to occur in nature will inform the design of management strategies that exploit the positive effects of hybridization while avoiding its negative effects. Furthermore, illuminating the prevalence of hybridization in evolutionary history may change management goals. For example, if a

clade frequently produced hybrid lineages in the past, then protecting rare declining taxa from hybridization at great financial cost may not be prudent use of resources.

Much about homoploid hybrid speciation still remains a mystery. We don't know its frequency in nature or the most common route(s) by which hompoloid hybrid species arise (although see Gross and Rieseberg 2005). Mathematical models and simulations have predicted what the genomic composition will be for stabilized hybrid species, but so far empirical work has used sparse marker sets (Buerkle and Rieseberg 2008). We don't know, for example, the average parental contributions to homoploid hybrid species. It can range from equivalent proportions like in an F1 hybrid to only a few loci from one species (Heliconius Genome Consortium 2012). At a deeper level, we don't know the extent of recombination in hybrid species' genomes, the rate of genome stabilization, or the relative importance of deterministic versus stochastic forces in the process. The repeatability of speciation in *Helianthus* hybrids implies that natural selection plays a crucial role in shaping the phenotype and genomic composition of hybrid lineages (Rieseberg 2003) , but disentangling the contributions of fertility and ecological selection continues to be challenging (although see Karrenberg et al. 2007). Homoploid hybrid speciation is thought to involve population bottlenecks, but we know very little about the extent of population size reductions and length of such bottlenecks or their effects on rates and patterns of genome stabilization.

1.4 Research questions

In this thesis, I aim to better understand hybridization's role in evolution by approaching the topic from three angles.

Question 1: What factors affect the rate of reproductive isolation evolution?

The evolution of reproductive isolation is a key step in speciation and plays a large role in determining the rate of post-speciation hybridization. In chapter 2, I use artificial crossing data from sunflowers and silverswords to explore how one trait, life history, affects the rate of reproductive isolation evolution.

Question 2: Is there genetic evidence of hybridization in Californian sunflowers?

Bolander's sunflower (*H. bolanderi*) in California is a classic example of a hybrid lineage arising through introgression. In chapter 3, I use next-gen sequencing data to definitively answer whether *H. bolanderi* is of hybrid origin and to explore the magnitude and direction of gene flow with invasive *H. annuus*.

Question 3: What is the genomic composition of homoploid hybrid species?

Homoploid hybrids are the most dramatic examples of the creative results of hybridization but exactly how two disparate genomes come together is still poorly understood. In chapter 4, I use transcriptomic data for three homoploid hybrid species and their parents to map parental contribution across the genome and explore questions about the origin of these hybrid species.

In chapter 5, I bring together and synthesis the results from the previous three chapters on hybridization. I discuss the strengths and weakness of the work, as well as future directions to explore.

Chapter 2: Hybrid incompatibility is acquired faster in annual than in perennial species of sunflower and tarweed.

2.1 Introduction

Speciation is characterized by the evolution of reproductive isolation. This can come in many forms including prezygotic barriers such as reproductive timing and gametic incompatibility or postzygotic barriers like hybrid viability or sterility (Coyne and Orr 2004; Rieseberg and Willis 2007). The speed with which these barriers arise and the impact of life history variation on their evolution remain poorly understood (Edmands 2002). In plants it is common for well-recognized species to be able to interbreed and produce hybrids of varying levels of fertility (Levin 1979). These intermediates can be used to study how intrinsic reproductive isolation evolves. That different plant species can interbreed is not a new discovery. This has been recognized since the 18th century and during the mid-20th century hybridization between taxa was widely employed to estimate phylogenetic relationships (Zirkle 1935; Levin 1979; Edmands 2002; Turesson 2010). Species with hybrids that had greater F1 viability or fertility were judged to be more closely related. This rich data set can be combined with modern sequencing efforts, which more precisely estimate divergence between species, to explicitly examine the relationship between genetic divergence and the strength of reproductive isolation.

In animals, it is widely accepted that reproductive isolation evolves in a relaxed clock-like manner. This has been shown in a variety of taxa including fish, birds, frogs, flies and butterflies (Sasa et al. 1998; Price and Bouvier 2002; Presgraves 2002; Russell 2003; Lijtmaer et al. 2003; Bolnick and Near 2005). In plants the relationship is less clear; a loosely clock-like relationship was found in *Silene* and *Coreopsis* but not in *Glycine, Streptanthus,* and *Frageria* (Moyle et al. 2004; Nosrati et al. 2011). This may reflect inherent differences in the genetic architecture of reproductive isolation. If many genes of small effect cause isolation, then a clear relationship will occur. Alternatively, if few genes (or chromosomal rearrangements) of large effect cause isolation, then stochastic variation among lineages may obscure any relationship (Edmands 2002).

Several biological factors have been shown to affect the rate of reproductive barrier evolution, including the degree of sympatry between species, the presence of sex chromosomes and the extent of ecological divergence (Edmands 2002; Nosil and Crespi 2006). Life history, annuals versus perennials, is associated with the evolution of reproductive isolation in the plant genus *Coreopsis* (family Asteraceae): annuals were found to accumulate hybrid incompatibilities more quickly than perennials (Archibald et al. 2005). However, this pattern hasn't been tested beyond this single genus. To determine whether this is a more general phenomenon, we analyzed the relationship between life history and the strength of hybrid sterility barriers in two independent

clades containing both extensive crossing data and life history variation, the genus *Helianthus* and subtribe Madiinae.

Helianthus (family Asteraceae) comprises 52 species, all native to North America. One of these is the common sunflower, *H. annuus*, which includes both the cultivated sunflower – an important crop – and its wild progenitor. The genus has been studied extensively for both agricultural and evolutionary purposes, resulting in a rich literature on chromosomal evolution and speciation (Rieseberg et al. 1995; Jan 1997; Archibald et al. 2005; Lai et al. 2005). Subtribe Madiinae (family Asteraceae) contains 24 genera and 121 species. This includes the tarweeds of California and silverswords of the Hawaiian Islands. The silverswords underwent a rapid radiation into many morphological forms but retained the ability to hybridize (Carr and Kyhos 1986). In both cases, older crossability data can be combined with more recent sequence data.

Here we have compiled pollen sterility and sequence data from artificial crosses between *Helianthus* and Madiinae species. We use these data to ask two questions: (i) Does reproductive isolation accrue in a clocklike manner? and (ii) Do annuals gain hybrid sterility faster than perennials? Additionally, we discuss possible causes of the differences in the rate of sterility evolution.

2.2 Methods

2.2.1 Data collection

Information on pollen sterility between *Helianthus* and Madiinae species was taken from the literature. *Helianthus* data included only crosses between sunflower species, while the Madiinae data included crosses between multiple genera of tarweeds. Artificial and natural hybrids were distinguished and only artificial crosses were used in our analysis. Direction of crosses was not distinguished, as this information was not available for all crosses.

Ten Madiinae crosses involved second-generation hybrids, e.g. *Dubautia knudsenii* X *D. laxa* crossed to *D. latifolia.* In these cases, the genetic distance used was the mean of the genetic distance from the first two species to the third species. These crosses were included in the phylogenetically corrected dataset only when the first two parental species were more closely related to each other than to the third species, i.e., when there was an unambiguous internal node. For the analysis of life history, these crosses were included because in each case all three parents were perennial, making assignment unambiguous. Life history was recorded as annual or perennial for each species. Thus crosses were annual-annual, perennial-perennial or annual-perennial.

Genetic distance was calculated from sequences of the external transcribed spacer (ETS) and the internal transcribed spacer (ITS) of 18S-26S nuclear ribosomal DNA for *Helianthus* and Madiinae, respectively. All sequences were obtained from Genbank

(Appendix A.3). Sequences were aligned using ClustalW (Larkin et al. 2007) and pairwise distance was calculated using MEGA5 (Tamura et al. 2011). Modeltest was used to determine the correct model of sequence evolution and only sites with ≥ 95% coverage were used (Posada and Crandall 1998).

2.2.2 Phylogenetic independence

Due to the nature of our dataset, the information provided by each individual cross was not phylogenetically independent. To alleviate this issue, we created a 'phylogenetically corrected' dataset (Coyne and Orr 1997). This collapsed all pairwise comparisons across a single internal node into a single data point. While this method does not provide complete phylogenetic independence, it is commonly used and ensures that any two data points do not share more than 50% of their phylogenetic history (Price and Bouvier 2002; Moyle et al. 2004; Larkin et al. 2007; Malone and Fontenot 2008).

Phylogenies for both datasets were taken from previously published work. For the *Helianthus* dataset the phylogeny was based on the same ETS sequences used to estimate genetic distance (Timme et al. 2007). For the Madiinae, no single published phylogeny covered our entire dataset of species so a consensus of multiple phylogenies was used. These phylogenies are based on ITS sequences (*Layia*, (Baldwin 2003); *Argyroxiphium*, *Dubautia*, *Wilkesia*, (Baldwin and Sanderson 1998)), both ETS and ITS (*Calycadenia*, (Baldwin and Markos 1998); *Deinandra*, (Baldwin 2007)), ETS,

ITS, and the *trnK* intron of chloroplast DNA (Madiinae, (Baldwin 2003)). Phylogenetic trees with nodes labeled are presented in Appendix A.1 and Appendix A.2.

To assess the effect of life history on the evolution of hybrid sterility, the dataset was first divided according to life cycle and then phylogenetically collapsed into independent nodes. The data were then brought back together into a single data set with independent data points of either type. Thus a single node on a tree may be represented in two separate categories, e.g. contain both an annual-annual and perennial-perennial comparison. The shared evolutionary history for these data points may obscure any differences in rate, but overall makes our test conservative in its conclusions.

Our method of assessing the effect of life history is simpler than the method used by Archibald et al. (2005), who assessed reproductive isolation in relation to annual or perennial *branch length,* but does not suffer from phylogenetic independence issues. Our test is likely less powerful but more conservative and does not rely upon the ability of the relatively short markers used to accurately reconstruct the phylogenetic relationships among the focal species.

2.2.3 Statistical analysis

We used genetic distance as a proxy for divergence times in our analysis. This relationship may be complicated by uneven rates of evolution or ongoing gene flow between species (but see discussion). As both pollen sterility and genetic distance were not normally distributed, both variables were arcsin transformed. We compared

genetic distance and pollen sterility between Madiinae crosses that were first and second generation hybrids (hybrid-hybrids) using a Kruskal-Wallis test (Kruskal and Wallis 1952). Transformed data were used to test for a correlation between pollen sterility and genetic distance using a non-parametric Spearman rank correlation to account for any residual non-normality.

To determine if life history affects the rate of reproductive isolation acquisition, we used an analysis of variance (ANOVA). We fit a linear model testing the effect of genetic distance, life history and their interaction on pollen sterility using the statistical programs in R (Ihaka and Gentleman 1996).

2.2.4 Testing evolutionary rate

Evolutionary rate was measured by comparing genetic distance between monophyletic groups of perennial or annual species with an outgroup that was equally related to all groups. Groups are indicated in supplementary figures 1 and 2. Genetic distance was measured with MEGA5 using Jukes-Cantor model with gamma parameter = 1 and complete deletion for missing positions.

2.3 Results

2.3.1 Data set

In *Helianthus* and Madiinae, we compiled data for 114 and 87 crosses representing 43 and 47 species, respectively. This included both within genera and between genera crosses as well as crosses where one or both of the parents were

themselves an F1 hybrid. These second generation hybrids were not different from the rest of the dataset in genetic distance or pollen sterility (d.f. = 1, $p = 0.594$; $p = 0.739$)

After collapsing the data to only phylogenetically independent nodes, 20 and 30 data points remained (shown in Appendix A.1 and Appendix A.2). The low number of independent nodes in the *Helianthus* dataset is for two reasons. First, the genus is divided into perennial and annual clades so all crosses between these clades (43 separate hybrids) are reduced to three nodes. Second, the perennial species are poorly resolved and many are not monophyletic. We were conservative in our use of these data so several species' relationships were reduced to single polytomies.

2.3.2 The relationship between pollen sterility and genetic distance.

There was a clear positive relationship between pollen sterility and genetic distance before phylogenetic correction for both Madiinae (rho=0.50, $p <$ E-6) and *Helianthus* (rho = 0.44, $p <$ E-6) datasets. In the phylogenetically independent datasets, this relationship is maintained for Madiinae (rho = 0.61, $p < 0.001$) but for *Helianthus* it is no longer significant (rho = 0.39, $p = 0.09$) (Table 2-1).

Table 2-1: Correlations between genetic distance and pollen viability for all comparisons and for the phylogenetically corrected dataset.

	N species	N Crosses $_{original}$	Spearman's rho$_{original}$		N Crosses$_{corrected}$	Spearman's rho$_{corrected}$	
Helianthus	43	114	rho=0.44	**p < E-06**	20	rho=0.39	p = 0.09
Madiinae	47	87	rho=0.50	**p < E-06**	30	rho=0.61	**p < 0.001**

2.3.3 Life history differences

Life history had a large effect in both data sets. Annual-annual crosses were much more strongly isolated than perennial-perennial crosses in terms of hybrid pollen viability (Figure 2-1). In both cases, when accounting for genetic distance, life history explained a significant portion of the variance in sterility (Table 2-2).

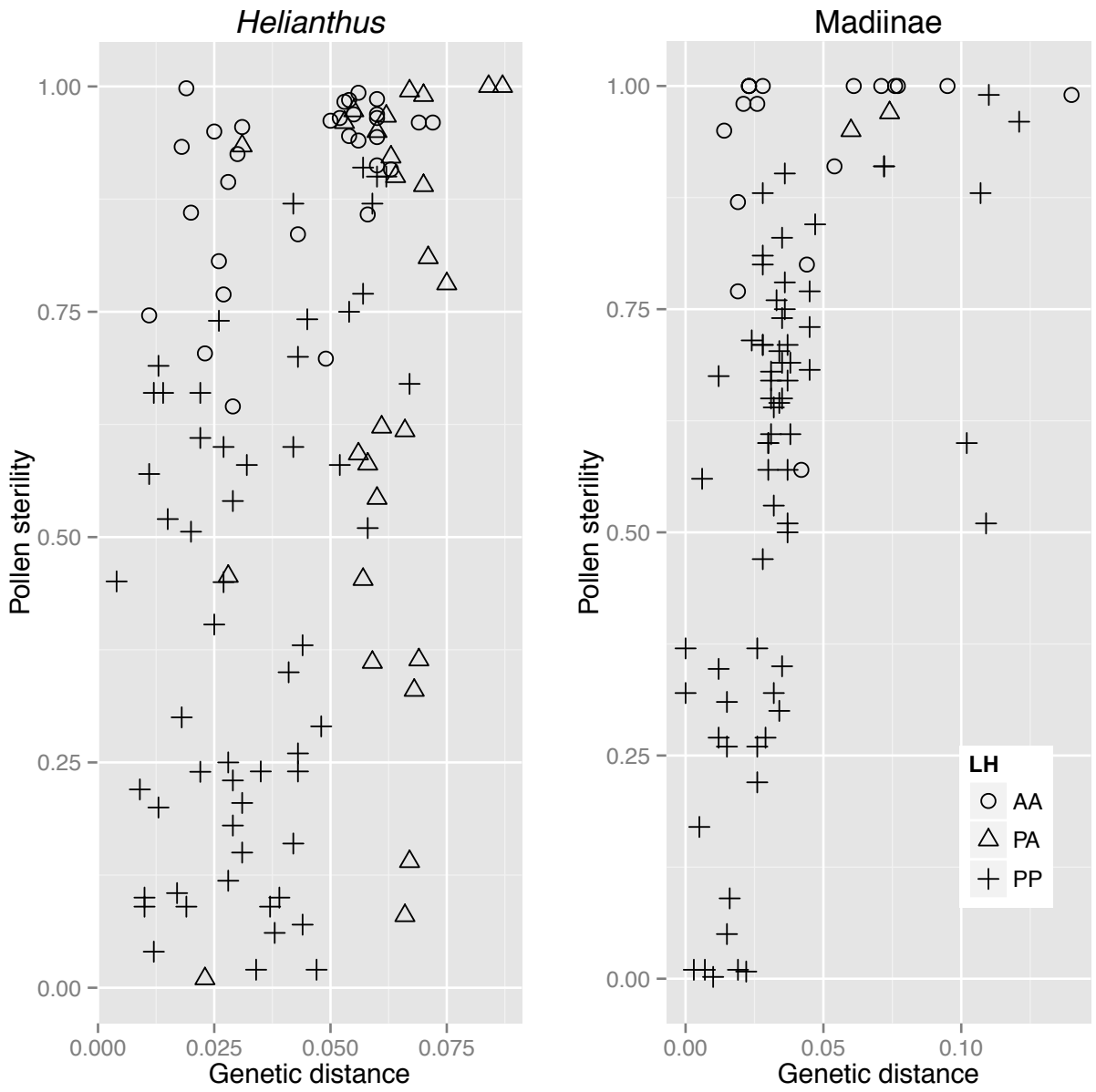


Figure 2-1: Pollen sterility and genetic distance for *Helianthus* and Madiinae data sets. Individual points are not phylogenetically corrected and are coded by life history combination. A is annual, P is perennial. Genetic distance was measured using ITS (Madiinae) or ETS (*Helianthus*).

Table 2-2: Results of analysis of variance for all variables tested using phylogenetically corrected datasets.
Genetic distance is arcsine transformed in all cases.

	Variable	Df	Sum Sq	Mean Sq	F value	p
Helianthus	Genetic Distance	1	0.7142	0.7142	14.0678	**0.001259**
	Life History	2	0.90428	0.45214	8.9059	**0.001714**
	Genetic Distance X Life History	2	0.06629	0.03315	0.6529	0.531289
	Residuals	20	1.01537	0.05077		
Madiinae	Genetic Distance	1	2.01713	2.01713	30.471	**9.77E-06**
	Life History	2	1.96053	0.98026	14.808	**5.73E-05**
	Genetic Distance X Life History	1	0.20736	0.20736	3.1324	0.08895
	Residuals	25	1.65496	0.0662		

2.3.4 Comparisons of rates of sequence evolution

For *Helianthus* data, perennial groups had mean genetic distances of 0.054, and 0.057, and the annuals had a mean distance of 0.064. For Madiinae, two paired perennial and annual clades had mean genetic distances of 0.098 versus 0.104 and 0.075 versus 0.084, respectively. In both cases, annual clades exhibited greater genetic distance.

2.4 Discussion

2.4.1 Hybrid sterility increases with genetic distance

It is intuitively obvious that reproductive isolation is correlated with genetic distance. Before populations diverge they should have little or no reproductive isolation and no genetic distance. Conversely, distantly related species have total reproductive isolation and high genetic distance. Positive correlation between genetic distance and sterility has been found repeatedly in animals, including Drosophila (Coyne and Orr 1997), frogs (Sasa et al. 1998), toads (Malone and Fontenot 2008), fish

(Russell 2003), birds (Price and Bouvier 2002) and butterflies (Presgraves 2002). Despite this, evidence for this pattern has been relatively scarce in plants; it was found in *Silene* and *Coreopsis* but missing in *Glycine, Streptanthus,* and *Frageria* (Moyle et al. 2004; Archibald et al. 2005; Nosrati et al. 2011). Here we show strong evidence for this relationship in both *Helianthus* (sunflowers) and Madiinae (tarweeds).

The positive correlation between reproductive isolation and genetic distance suggests that reproductive isolation is acquired in a relaxed clock-like manner. This occurs despite evidence that chromosomal rearrangements play a significant role in generating sterility (see below).

2.4.2 Life history

Our analysis clearly shows that annual species develop F1 hybrid sterility at a faster rate than perennials. Annual-annual crosses have mean pollen sterility of 90% (*Helianthus*) and 93% (Madiinae) versus 41% and 55% for perennial-perennial crosses. In fact, there are no annual-annual crosses with less than 57% sterility despite the inclusion of crosses between sister species.

It is interesting to note that although hybrids between perennial sunflowers are highly fertile, there seems to be a strong barrier to hybrid seed production (Heiser et al. 1969). Artificial crosses between perennial species require huge amounts of effort to obtain a few viable seeds; indeed, modern crosses involving perennial sunflowers often use embryo rescue (Kräuter et al. 1991).

2.4.3 Evolutionary rate

Our study uses genetic distance as a proxy for divergence time. This is not a perfect measure as rates of sequence evolution vary between lineages and, most relevantly, between life history strategies (Gaut et al. 2011). Several studies have shown that molecular evolutionary rates are faster in annuals than in perennials (Andreasen and Baldwin 2001; Kay et al. 2006; Soria-Hernanz et al. 2008); when taken into account with our results, this actually accentuates the pattern we find. If annuals evolve unusually fast in terms of nucleotide sequence, then annual-annual comparisons have lower divergence *times* and are *younger* than expected based on sequence divergence. Conversely, perennial-perennial pairs are *older* than what our sequence divergence suggests. Consider a scenario where there was no effect of life history and reproductive isolation evolved in a rate purely proportional to divergence time. Two pairs of species, one annual-annual and one perennial-perennial, that have been diverging for equal amounts of time would have equal reproductive isolation, but the annual-annual pair would have higher sequence divergence and, consequently, according to our measure, a slower rate of reproductive isolation gain. This is the opposite of the pattern we observe in the data; therefore differences in the rate of sequence evolution are not driving the patterns we see.

To confirm the differences in sequence divergence rate, we examined evolutionary rate in our dataset by comparing mean genetic distance between annual and perennial groups to outgroups (Appendix A.1 and Appendix A.2). In all cases

annual clades had greater genetic distance, suggesting faster sequence evolution. The variation between Madiinae pairs may represent long-term differences in rates of sequence divergence as these comparisons are between different genera. In each case, annual groups evolved faster in terms of nucleotide sequence than perennial groups. Thus, the more rapid evolution of hybrid sterility barriers in annuals does not appear to be a consequence of misestimating divergence times. Rather, differences in rates of sequence evolution appear to be causing the trend to be underestimated.

It is also possible that the low levels of hybrid sterility found between perennial species may permit significant interspecific gene flow, thereby reducing genetic divergence. However, this seems unlikely for perennial sunflowers, which appear to be reproductively isolated by strong prezygotic reproductive barriers. Also, this scenario does not explain why annuals developed high levels of reproductive isolation and perennials did not.

2.4.4 Causes of sterility

Hybrid sterility can be caused by epistatic interactions (including Dobzhansky-Muller incompatibilities) or chromosomal rearrangements. DM incompatibilities are negative epistatic interactions in hybrids originating from genes that evolved independently in the parental species. Chromosomal rearrangements, on the other hand, cause sterility through the production of chromosomally unbalanced gametes (Coyne and Orr 2004). While both cause sterility, there are distinct effects. DM incompatibilities typically are recessive and may therefore be masked in the F1 and

only appear in the F2 generation, leading to increased sterility in second-generation hybrids. Chromosomal rearrangements, on the other hand, are underdominant and would thus have the greatest effect in the F1, where all polymorphic loci are heterozygous. In the F2 generation heterozygosity is reduced and so sterility from chromosomal rearrangements will stay constant or be reduced. Additionally, in the absence of sex chromosomes, chromosomal rearrangements are symmetrical in their effect on sterility; it does not matter which species is the mother. DM incompatibilities can be bidirectional, like chromosomal rearrangements, or unidirectional and cause asymmetric sterility (Turelli and Moyle 2007). Lastly, artificial genome doubling using colchicine creates hybrids with perfectly paired chromosomes, alleviating the effect of chromosomal rearrangements but not DM incompatibilities (Stebbins 1958).

Based on these features, we have several reasons to believe that in these systems hybrid sterility is largely caused by chromosomal changes. Pollen sterility has been mapped to chromosomal rearrangements in *Helianthus* (Quillet et al. 1995; Lai et al. 2005), although epistatic interactions between sterility QTLs suggest DM incompatibilities contribute as well. Furthermore, among F1 *Helianthus* hybrids, pollen sterility was correlated with number of chromosomal translocations, although insignificantly (Chandler et al. 1986; Levin 2002). Similarly, in Hawaiian silverswords (subtribe Madiinae) the number of translocations between parental species is strongly correlated with pollen sterility in hybrids (Carr and Kyhos 1981; 1986; Levin 2002).

Chromosomal rearrangements have been extensively noted in both studied groups (Chandler et al. 1986; Carr and Kyhos 1986).

Asymmetry of sterility and the relative sterility of F1 versus F2 generations are not commonly reported or tested in our dataset so we cannot formally test them, but we examine the available data here. Cross sterility symmetry was not reported for Madiinae crosses, but for *Helianthus* crosses are generally found to be symmetrical (Long 1955; Lai et al. 2005), suggesting little contribution from unidirectional DM incompatibilities. In hybrids between the annual sunflowers *H. anunus* and *H. petiolaris*, pollen viability significantly increases from the F1 generation (5.6 ± 2.2 %, n=20) to the F2 (31.6 ± 12.4 %, n=20) (t-test, $p<0.0001$) (Rieseberg 2000). Contrary to this, in hybrids between the perennial sunflowers *H. decapetalus* and *H. laevigatus*, viability decreased from the F1 (80%) to the F2 (66%) generation (Heiser and Smith 1964). Lastly, colchicine-induced chromosome doubling, which helps alleviate chromosomal mispairing, has increased pollen fertility in several sunflower hybrids (Heiser and Smith 1964; Jan and Chandler 1989).

We believe this evidence is consistent with the idea that chromosome rearrangements are important in the hybrid sterility we measured, although almost certainly not the only cause. If we accept the importance of rearrangements, why are these rearrangements occurring more frequently or being fixed more often in annuals than perennials? More specifically, we would suggest that there are more karyotypic changes per nucleotide substitution in annuals than perennials. This could be because

chromosomal rearrangements occur more frequently or because demographic or selective factors cause them to fix at a greater rate. There are biological features that promote both of these options.

It is generally believed that chromosomal rearrangements primarily occur during meiosis mediated by the double strand breaks used in homologous recombination (Shaffer and Lupski 2000). By regenerating from seed every year, annuals may undergo more frequent meiosis events than perennials and accrue more chromosomal rearrangements as a consequence.

The increased chromosomal evolution may also be due to a difference in fixation rather than mutation rate. When faster sterility acquisition in annuals was first described by Stebbins (1958), he suggested that intense population fluctuations allow annuals to fix underdominant genic or chromosomal changes faster than perennials, which have more stable population sizes (Stebbins 1958). This intuitive explanation was later formalized by mathematical models demonstrating that chromosomal rearrangements could only be established in very small or inbred populations (Walsh 1982). Counter to this, in our dataset annual sunflowers, which have extremely high rates of chromosomal evolution (Burke et al. 2004), also have very high effective population size (Strasburg et al. 2011) indicating few species-wide bottlenecks. Within Madiinae, a majority of perennial crosses, which have relatively low sterility, involve silverswords, a group that speciated within the Hawaiian Islands and underwent repeated population bottlenecks (Witter and Carr 1988). Grant (1981) later suggested

that higher levels of selfing in annuals also contributed to higher rates of karyotypic evolution (Grant 1981). While selfing annuals may have high rates of chromosomal evolution, this does not explain the results reported here. In our datasets all species, including the annuals, are self-incompatible (with the exception of *H. agrestis,* which has 100% pollen sterility in both available crosses). Thus, differences in the fixation rate of karyoptypic changes due to variation in effective population size or mating system cannot account for the pattern in our dataset.

Chapter 3: Revisiting a classic case of introgression: Hybridization and gene flow in Californian sunflowers.

3.1 Introduction

In Verne Grant's seminal work "Plant Speciation", he lists four examples of introgression, one of which involves the sunflower *Helianthus bolanderi* (Grant 1981). Both morphology and habitat suggested that this largely ruderal species was a product of introgression between the smaller native serpentine endemic *H. exilis* and a larger recent weedy invader *H. annuus* (Heiser 1949). Work using early genetic markers failed to find evidence for a hybrid origin of *H. bolanderi* but the hybridization between *H. bolanderi* and *H. annuus* is ongoing as *H. annuus* invades California (Rieseberg et al. 1988; Carney et al. 2000). Here we re-investigate this classic example with high-resolution genomic data to ask if *H. bolanderi* is a product of introgression and also whether the direction of introgression, if any, is consistent with current theory.

During invasion, hybridization between the invader and native species can occur and is recognized as a major issue in species conservation (Rhymer and Simberloff 1996; Levin and Ortega 1996; Vilà et al. 2000; Allendorf et al. 2001). Although contamination of the native gene pool and "genome extinction" are the primary conservation issue, current models suggest that it is the invader that should be subject to the most introgression (Grant 1981; Currat et al. 2008). This is because hybrids will more often backcross with the invading species rather than the declining native

species. As the invasion spreads, these backcrossed individuals will advance with the wavefront. Therefore as the invasion continues, introgression should continue to increase until counteracted by selection. This pattern has been seen in many empirical studies (Heiser 1949; Martinsen et al. 2001; Donnelly et al. 2004; Secondi et al. 2006), but not all (Rieseberg et al. 1988; Goodman et al. 1999; Carney et al. 2000; Takayama et al. 2006) and is often attributed to the effects of selection or sex biased dispersal (Kulikova et al. 2004; Melo-Ferreira et al. 2005).

In Californian sunflowers, contemporary hybridization with *H. annuus* appears to be limited to *H. bolanderi* and not its sister species *H. exilis*. *Helianthus annuus* is native to central USA and has invaded California from south to north, up the Central Valley over the last several thousand years (Heiser 1949). Currently, it is found primarily south of Sacramento (38.5° N) and has replaced *H. bolanderi* populations in the Central Valley over the last 100 years (Carney et al. 2000). Hybridization is expected to be rarer with *H. exilis* because it occurs almost exclusively on serpentine soil, an extreme soil type characterized by a high Mg/Ca ratio and high levels of heavy metals, including Ni, Cr and Cd (Brooks 1987). Serpentine soil is deadly to non-adapted plant species but is home to a wide variety of endemic species (Safford et al. 2005; Brady et al. 2005). *Helianthus bolanderi* also occurs on serpentine soil, but not exclusively, while *H. annuus* has not been reported from serpentine soils. *Helianthus exilis* is morphologically differentiated from *H. bolanderi* by having lance-linear leaves, entire leaf margins and smaller flower heads and fruit.

We used genotyping-by-sequencing (GBS), a popular restriction enzyme-based method for reducing genome complexity, to interrogate the genomes of these three species. We ask the following three questions. (1) Is *H. bolanderi* of hybrid origin as hypothesized by Heiser (1949) and Grant (1981)? (2) Is there introgression between *H. bolanderi* and *H. annuus*? (3) Is introgression biased into the invader, *H. annuus*, as predicted by models? Our results provide the final resolution of a classic case study of the role of hybridization in plant evolution, and a test of contemporary theory regarding patterns of introgression during biological invasions.

3.2 Methods

3.2.1 Data preparation

3.2.1.1 Sampling

We collected *H. exilis* and *H. bolanderi* seeds from 10 sites across the known species ranges in August 2011 (Table 3-1). Additionally, we used seeds from the United States Department of Agriculture National Plant Germplasm System (USDA NPGS) (11 populations) and one population from Jake Schweitzer to supplement our collection. As there is controversy in the literature about the species' delimitation between *H. exilis* and *H. bolanderi*, we took an agnostic approach to collecting (Jain et al. 1992). Populations spanning the combined species ranges, including populations that had previously been identified as either species, were sampled. Similarly, all available samples of both species from the USDA NPGS were genotyped. Up to ten seeds were sampled per population. For personally collected populations, each seed came from a

separate maternal parent; for USDA NGRP seed, pooled parental seed was used. For samples from throughout the range of *H. annuus* as well as for several perennial sunflower outgroup species (specifically *H. divaricatus*, *H. giganteus*, *H. grosseserratus*, *H. maximiliani* and *H. nuttallii*), we employed GBS data previously generated in the Rieseberg lab using the same GBS protocol employed here (Baute 2015). These data are currently on the NCBI Sequence Read Archive (SRA) (Appendix B.2). Altogether we used 322 samples: 190 *H. bolanderi-exilis*, 102 *H. annuus* and 30 perennial sunflowers.

Table 3-1: Sample information by population.
Non-*H. bolanderi-exilis* samples are from a range of locations specified individually in Appendix B.2. Sample size information is post-sample quality filtering.

Population	Species	Sample size	Latitude	Longitude	Area	Serpentine?	Mg/Ca Ratio
G100	*H. bolanderi-exilis*	10	39.40117	-122.61349	Coast Mountains	yes	4.26
G101	*H. bolanderi-exilis*	3	39.26759	-122.48275	Coast Mountains	no	0.48
G102	*H. bolanderi-exilis*	10	39.12638	-122.43213	Coast Mountains	yes	3.38
G103	*H. bolanderi-exilis*	10	38.7804	-122.57185	Coast Mountains	yes	2.41
G108	*H. bolanderi-exilis*	11	38.87585	-120.8205	Sierra Nevada Mountains	yes	2.66
G109	*H. bolanderi-exilis*	10	39.17832	-121.75977	Central Valley	no	0.16
G110	*H. bolanderi-exilis*	6	39.25156	-121.88924	Central Valley	no	0.30
G111	*H. bolanderi-exilis*	10	39.34395	-121.44869	Central Valley	no	0.14
G114	*H. bolanderi-exilis*	11	41.28199	-122.85186	North Mountains	yes	4.53
G115	*H. bolanderi-exilis*	7	41.64306	-122.74711	North Mountains	yes	13.02
G116	*H. bolanderi-exilis*	5	39.066322	-122.478403	Coast Mountains	yes	NA
G118	*H. bolanderi-exilis*	9	39.2627	-122.51157	Coast Mountains	yes	1.89
G119	*H. bolanderi-exilis*	9	39.48584	-121.31271	Sierra Nevada Mountains	no	0.26
G120	*H. bolanderi-exilis*	8	38.543	-121.7383	Central Valley	no	NA
G121	*H. bolanderi-exilis*	10	38.82395	-122.33725	Coast Mountains	yes	NA
G122	*H. bolanderi-exilis*	8	38.73309	-122.52462	Coast Mountains	yes	2.78
G123	*H. bolanderi-exilis*	10	39.83434	-121.58227	Sierra Nevada Mountains	yes	6.25
G124	*H. bolanderi-exilis*	10	38.84119	-120.87647	Sierra Nevada Mountains	yes	2.50
G127	*H. bolanderi-exilis*	10	37.84557	-120.46388	Sierra Nevada Mountains	yes	1.82
G128	*H. bolanderi-exilis*	4	41.03086	-122.42451	North Mountains	yes	1.85
G129	*H. bolanderi-exilis*	6	39.88756	-122.63451	Coast Mountains	no	0.84
G130	*H. bolanderi-exilis*	10	41.29794	-122.72187	North Mountains	yes	2.56
cal_ann	*H. annuus*	24	NA	NA	California	NA	NA
cen_ann	*H. annuus*	76	NA	NA	Central USA	NA	NA
div	*H. divaricatus*	5	NA	NA	Central USA	NA	NA
gig	*H. giganteus*	5	NA	NA	Central USA	NA	NA
gro	*H. grosseserratus*	6	NA	NA	Central USA	NA	NA
max	*H. maximiliani*	10	NA	NA	Central USA	NA	NA
nut	*H. nuttallii*	3	NA	NA	Central USA	NA	NA

3.2.1.2 Soil sampling

For each site from which we collected seeds, we also collected soil for composition analysis. Soil was collected six inches below the surface in five randomly selected locations spanning the collection area and pooled. Soil was analyzed at A&L Western Labs and measured for organic matter, phosphorous, potassium, magnesium, calcium, sulphur, pH and hydrogen. Additionally, DTPA-Sorbitol extraction was used to measure the heavy metals nickel, chromium and cobalt.

For a subset of the USDA NGRP samples, calcium and magnesium concentrations in the soil were measured (Gulya and Seiler 2002). The remaining three sites had no soil measurements but two were from areas described as serpentine (G116, G121) and one from an area with no nearby serpentine (G120).

3.2.1.3 Genotyping-by-sequencing

Seeds were germinated and grown to seedling stage. DNA was extracted from young leaves using Qiagen DNeasy plant kit (Qiagen, Valencia, CA, USA), with RNase A. DNA quantity was assessed using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

GBS Library construction was done using the standard protocol of Elshire et al., (2011) except for the addition of a gel-isolation step to eliminate dimers generated by the polymerase chain reaction (PCR) (Elshire et al. 2011). Two libraries of 95 samples each were prepared.

3.2.1.4 Sequencing and data preparation

Both GBS libraries were paired end sequenced on an Illumina HiSeq2000 at the UBC Biodiversity Research Center, a single lane each. Individual data were demultiplexed from within read barcodes using a custom Perl script that also removed barcode sequence. Fastq files were then trimmed for low quality reads and Illumina adapters using Trimmomatic (Bolger et al. 2014). Raw demultiplexed data were uploaded to the SRA (SRP062491).

3.2.1.5 SNP calling

Data were aligned to the *H. annuus* reference genome (HA412.v1.1.bronze) using BWA (version 0.7.9a) and Stampy (version 1.0.23) using default parameters (Li and Durbin 2010; Lunter and Goodson 2011). Because we were aligning sequence data to a diverged species reference, we used Stampy to increase alignment quality. BAM files were cleaned, sorted and had their read group information added using Picard tools (1.114) (http://broadinstitute.github.io/picard/). We used the Genome Analysis ToolKit (version 3.3) to identify possible alignment issues and realign those areas using '*RealignerTargetCreator*' and '*IndelRealigner*' (Van der Auwera et al. 2002). BAM files were processed using the GATK '*HaplotypeCaller*' program and SNPs were ultimately called all together using '*GenotypeGVCFs*'. SNPs were converted to a flat table format using a custom Perl script which removed indels, required sites to have QUAL > 20 and MQ > 20, and required individual genotypes to have depth between 5 and 100,000 and

GT_QUAL > 20. Samples with below ~25,000 reads were removed because they did not have enough data to be informative.

After initial SNP calling, the data were divided into three datasets: only *H. bolanderi* and *H. exilis* (dataset 'BE'), *H. bolanderi*, *H. exilis* and *H. annuus* (dataset 'BE+A'), and all samples including the outgroup perennials (dataset 'BE+A+P'). These sets were filtered to remove sites with sample coverage < 60%, minor allele frequency < 1% and observed heterozygosity > 60% using a custom perl script. These are referred to as the 'filtered' datasets. For population structure analysis, linkage between markers can cause issues, so we subsequently thinned each filtered set so that each SNP is at least 1000bp from its nearest neighbor, effectively picking one SNP per GBS tag. These are referred to as the 'thinned' datasets.

3.2.2 Evaluating the genetic structure of *H. bolanderi* and *H. exilis*

3.2.2.1 Population structure and admixture

To detect admixture and population structure in *H. bolanderi-exilis*, we ran fastStructure using the 'BE' filtered dataset with K=1-10 (Raj et al. 2014), and repeated 100 times. The optimal K was found using the "chooseK" script bundled with fastStructure. *Admixture* was run from K=1-20, using the default parameters (Alexander et al. 2009). Cross-validation scores were used to determine the best K value. To control for linkage effects, this was repeated with the 'thinned' dataset that has neighbouring SNPs removed. Principal component analysis (PCA) was run using the "FactoMineR" packaged in R, using the command "PCA". Missing data were imputed

using the package "missMDA". These analyses were repeated using the same parameters with the 'BE+A' dataset.

Overall sample relatedness was visualized with an unrooted phylogenetic network using SplitsTree4 on the 'BE' filtered dataset (Huson 1998). Uncorrected P-distance was used and heterozygous sites were ignored (as per defaults). This was also run using the 'BE+A+P' filtered dataset.

We calculated F_{ST} between all pairs of populations using the Weir and Cockerham method (Weir and Cockerham 1984), and F_{IS} for each population. Both were calculated using custom Perl scripts.

3.2.2.2 Introgression with *H. annuus*

To determine if *H. bolanderi* is uniquely introgressed from *H. annuus*, we calculated Patterson's D statistic (Kulathinal et al. 2009; Green et al. 2010; Durand et al. 2011), which is commonly known as the ABBA-BABA test. It requires sequence data from four groups (either individual samples or allele frequencies). P1 and P2 are geographically separated populations of one species, P3 is a separate species in sympatry with P2, and P4 is an outgroup species. The test counts the number of ABBAs (where P2 and P3 share a derived allele) and BABAs (where P1 and P3 share a derived allele). Under incomplete lineage sorting, we would expect an equal number of ABBAs and BABAs, but if there is gene flow between P2 and P3, there would be excess ABBAs and D would be positive.

Since we had many samples of each group, we used allele frequencies instead of instance counts of single samples (Martin et al. 2015). The four groups used were all central *H. annuus* (i.e., all *H. annuus* not in California), all California *H. annuus*, an *H. bolanderi-exilis* population and all perennial sunflowers. Perennial sunflowers included *H. maximiliani*, *H. nuttallii*, *H. divaricatus*, *H. giganteus* and *H. grosseserratus.* This monophyletic group of species is an outgroup to the annual sunflowers that include *H. annuus* and *H. bolanderi-exilis*. Only biallelic sites for which all perennial samples were fixed for a single allele were used, because these sites gave the most confidence in determining the ancestral allele. We also calculated f_d, a measure of the amount of the genome involved in introgression (Martin et al. 2015). For each statistic, we calculated standard deviation, Z-score and p-value using a block jackknife approach with 10MB size blocks (Green et al. 2010). This test was run on each individual *H. bolanderi-exilis* population as well as all *H. bolanderi-exilis* samples together.

For this test, a positive D score indicates that ABBA > BABA, and California *H. annuus* and *H. bolanderi-exilis* share more derived alleles. A negative D score indicates that BABA > ABBA and central *H. annuus* and *H. bolanderi-exilis* share more derived alleles. The neutral expectation under no gene flow is ABBA = BABA and D = 0.
To evaluate hypotheses about introgression, we examined D and f_d in 10 Mb windows across the genome. We also used the *H. annuus* genetic map to compare recombination rate and introgression in 10 Mb windows using a type III ANOVA (Renaut et al. 2013).

A positive D statistic using allele frequencies from all samples may be driven by a subset of samples if introgression is not uniform among California *H. annuus* and *H. bolanderi-exilis* samples. It could also be caused by unmeasured introgression into central *H. annuus* by a third species (e.g. *H. petiolaris*, which is known to hybridize and is largely sympatric across the central USA range of *H. annuus* (Yatabe et al. 2007)). To account for this we used a subsampling strategy that isolates each sample individually (while retaining all samples for other groups) and calculates a D score. For example, one test would include one central *H. annuus* sample, all Californian *H. annuus*, all *H. exilis-bolanderi*, and all perennial samples. Thus, for each sample we get a D score reflecting its effect on the overall D score. Significance was calculated using a block jackknife approach (as above).

We use these single sample D scores to assess the hybrid origin of *H. bolanderi*. If *H. bolanderi* was a hybrid species, we would expect all *H. bolanderi-exilis* samples to have to fall into two distinct sets; one with high D scores (representing the hybrid *H. bolanderi)* and one with lower, but possibly still positive, D scores (representing non-introgressed *H. exilis*). A non-introgressed *H. exilis* may still produce a positive D score because of introgression in *H. annuus*, but a hybrid species should be distinctly higher. To evaluate the amount of introgression in each sample or population, we plotted individual sample D scores versus latitude (for *H. bolanderi-exilis* and *H. annuus*) and versus collection date (for *H. annuus*) (Wickham 2009). We used a type III ANOVA,

using the R package "car", to determine if each of these factors affected D or f_d (Fox and Weisberg 2010; R Core Team 2008).

3.2.3 Testing the directionality of gene flow with *H. annuus*

3.2.3.1 The partition D test

A positive D score indicates gene flow, but does not specify if the gene flow is into *H. bolanderi-exilis,* into *H. annuus,* or is bidirectional. To answer this question, we used the partitioned D statistic (Eaton and Ree 2013). This extension of the ABBA-BABA test uses five taxa instead of four and can determine directionality of introgression using a set of three different tests. The main difference between the partitioned D statistic and Patterson's D statistic is that the partitioned version divides the P3 clade (i.e., *H. bolanderi-exilis* in our analysis) into two lineages, $P3_1$ and $P3_2$, which are assumed not to be exchanging genes. The three partitioned D statistic tests then ask if the enrichment of shared derived alleles shown by the positive classic D statistic are from the first, second or both P3 lineages. Specifically, D_1 compares counts of ABBAA and BABAA looking for enriched shared derived alleles specifically in $P3_1$, D_2 compares counts of ABABA and BAABA looking for enriched shared derived alleles specifically in $P3_2$, and D_{12} compares counts of ABBBA and BABBA looking for enriched shared derived alleles in both $P3_1$ and $P3_2$.

Comparing the results of the three tests can be used to determine the directionality of gene flow. Consider the scenario where D_{12} is positive. This either suggests gene flow from P2 into the ancestor of $P3_1$ and $P3_2$, gene flow from P2 into

both $P3_1$ and $P3_2$, or gene flow from $P3_x$ into P2. If the first two scenarios can be ruled out by other tests or outside information, then gene flow in one direction is supported. In this scenario, the lineage of P3 that is donating genes is determined by the D_1 and D_2 tests. This in itself only indicates that gene flow is going in at least one direction, not that it is unidirectional, but by rotating the positions in the phylogeny (i.e., P1->$P3_2$, P2->$P3_1$, $P3_1$->P2, $P3_2$->P1), and repeating the tests we can make a case for the overall directionality of gene flow. For example, if in the rotated phylogeny scenario the D_{12} test is zero, then there is a lack of evidence for gene flow in the opposite direction and unidirectional gene flow is supported overall. With this framework in mind, we used two phylogenetic scenarios (i.e., the same phylogeny rotated differently) to get at the directionality of gene flow.

The first scenario uses the five groups in the following order: P1 = all central *H. annuus*, P2 = all California *H. annuus*, $P3_1$ = a southern *H. bolanderi-exilis* population, $P3_2$ a northern *H. bolanderi-exilis* population (G115), and P4 = perennial outgroup. In this case, we are treating G115 as non-introgressed due to its geographic isolation from any *H. annuus* population and the strong population structure, indicating little within species gene flow.

With our groupings in mind, the three tests from the partitioned D have different implications in this scenario. D_{12} asks if derived alleles found in both *H. bolanderi-exilis* populations are more often found in California *H. annuus*, than central *H. annuus*. A positive score suggests gene flow from any *H. bolanderi-exilis* into *H. annuus* because

otherwise the derived allele would not be present in both *H. bolanderi-exilis* populations. D_1 asks if derived alleles, not found in northern *H. bolanderi-exilis,* are present in California *H. annuus*. A positive score suggests that there is gene flow between the southern *H. bolanderi-exilis* and California *H. annuus*, or that there is gene flow between California *H. annuus* and a population of *H. bolanderi-exilis* more closely related to the southern *H. bolanderi-exilis* population tested. D_2 asks the same as D_1 but with northern and southern *H. bolanderi-exilis* populations reversed (i.e., this may suggest gene flow with northern *H. bolanderi-exilis* or close relative).

The test was repeated using each *H. bolanderi-exilis* population in $P3_1$, except G115, which is always in $P3_2$. This means that we did each test 21 times and our main reported result is how many of these tests were significantly positive. The number of positive tests is indicative of how consistent the signal is across the range of *H. exilis-bolanderi*. Since we tested every population, some tests involve two *H. bolanderi-exilis* populations that are both in the northern clade.

The second scenario involves a rotated phylogeny. The five groups are: P1 = a northern *H. bolanderi-exilis* (G115), P2 = a southern *H. bolanderi-exilis*, $P3_1$ = California *H. annuus*, $P3_2$ = central *H. annuus* and P4 = perennial outgroup. In this scenario, D_{12} asks if derived alleles found in all *H. annuus*, are present in the southern *H. bolanderi-exilis* and not the northern. A positive score indicates gene flow into *H. bolanderi-exilis*. Tests D_1 and D_2 ask if there are an excess of derived alleles from California *H. annuus* or central *H. annuus* respectively in southern *H. bolanderi-exilis*. Similarly in

this scenario we also repeat each test using a different southern *H. bolanderi-exilis* population and report the number of significantly positive tests.

For these tests we used allele frequencies instead of individual genomes and only included sites where all perennial samples were fixed for a single allele. Significance was tested using block jackknife bootstrapping, as before, and $p < 0.05$ was used as the p-value cut off. All tests were repeated using another population (G114) as the northern non-introgressed *H. bolanderi-exilis* population.

3.2.3.2 Demographic modeling

To explore the amount and direction of gene flow, we simulated the demographic history using **δaδi** (Gutenkunst et al. 2009). **δaδi** simulates the site frequency spectrum of demographic scenarios and uses diffusion approximation to explore the parameter space. In our model we use three populations (*H. bolanderi-exilis*, central *H. annuus*, and California *H. annuus*) and seven parameters; three effective population sizes, N_{BE}, N_{CenA}, and N_{CalA}, two times, T_1 and T_2, and two migration rates, $m_{CalA->BE}$ and $m_{BE->CalA}$. At time T_1, central *H. annuus* and *H. bolanderi-exilis* diverge, and at time T_2, *H. annuus* invades California and exchanges genes with *H. bolanderi-exilis* until present (Figure 3-1). We also ran the model with the migration events removed in all combinations.

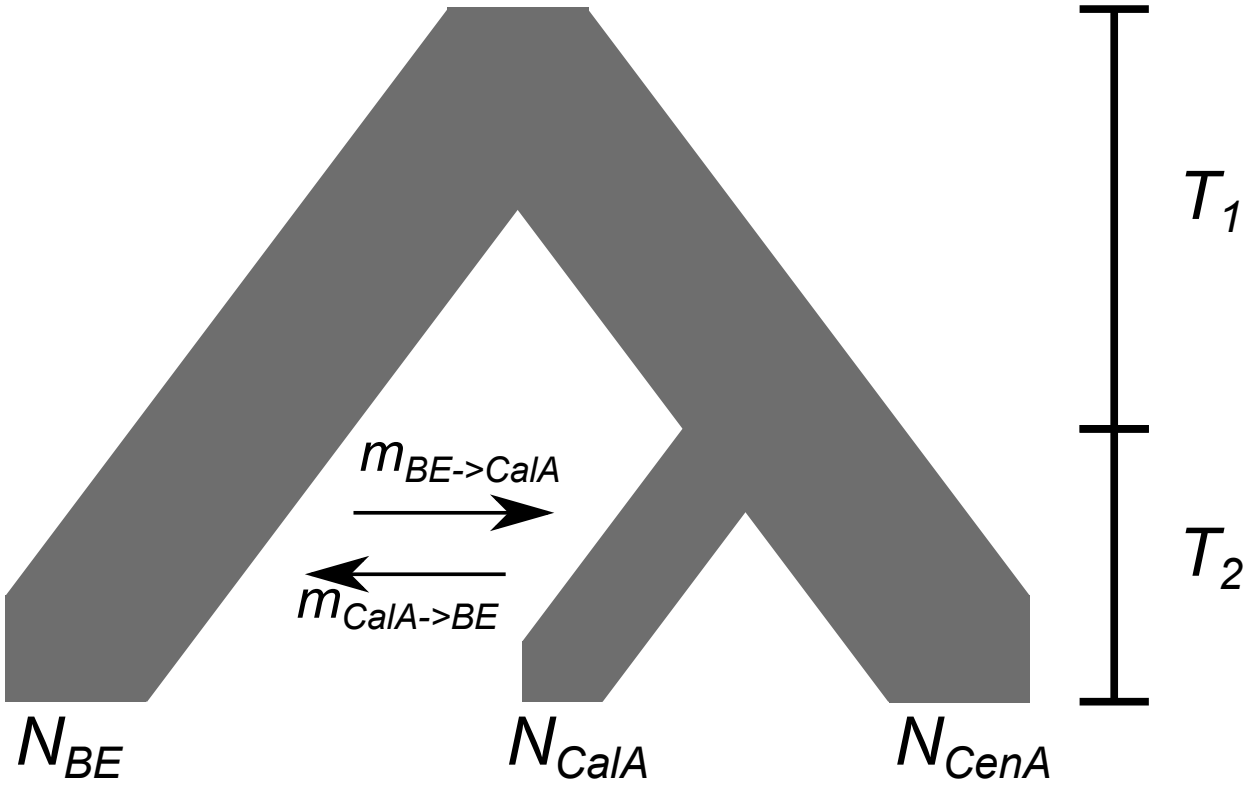


Figure 3-1: Demographic scenario modeled in δaδi including all modeled parameters. Including effective population size (N) for *H. bolanderi-exilis* ($_{BE}$), California *H. annuus* ($_{CalA}$), and central *H. annuus* ($_{CenA}$), migration rates (m) and time (T).

We used the BFGS optimization method to fit parameters for each model. Searches were started from 10 randomly perturbed starting positions with up to five iterations each. The best-fit parameters were used for a further optimization for up to 20 iterations. Samples were extrapolated to grid size of [175,75,25] to maximize the number of usable SNPs. Three hundred bootstrap site frequency spectra were generated using 1Mb block bootstrapping. This was used to calculate confidence intervals for all parameters. Parameters were corrected using the mutation rate of 6.1 * 10^{-9} substitutions/site/generation (Sambatti et al. 2012). Effective sequenced length

was estimated by measuring the number of sites with >5 reads in 88 *H. bolanderi-exilis*, 38 central *H. annuus* and 13 California *H. annuus* samples, including invariant sites. These numbers were chosen to reflect the extrapolation grid size.

3.3 Results:

3.3.1 Sample and SNP information

3.3.1.1 Sample sizes

We removed three *H. bolanderi-exilis* and two *H. annuus* samples for having < 25,000 reads. One perennial sample (GB148) was removed because it grouped with *H. annuus* samples in the splits network analysis. After removing samples, we had sequence data for 187 *H. bolanderi-exilis* samples, 100 *H. annuus* samples and 29 perennial sunflower samples (Appendix B.1).

3.3.1.2 Soil analysis

Serpentine sites are primarily characterized by Mg/Ca ratio > 1 (Kruckeberg 1985). All sites identified by plant composition and soil maps as serpentine were confirmed with soil measurements (Appendix B.1).

3.3.1.3 SNP calling

All demultiplexed data was uploaded to the SRA (SRP062491). Number of reads per sample and percent aligned reads are listed in Appendix B.2. After initial filtering for quality and depth, we found 131,150 SNPs total (Table 3-2). Subsequent filtering for coverage (> 60%), minor allele frequency (> 1%) and observed heterozygosity (< 60%) reduced that to 9,593 SNPs.

Table 3-2: Number of SNPs found for each dataset.
The filtered dataset removed sites where sample coverage < 60%, observed heterozygosity > 60% or minor allele frequency < 1%. The thinned dataset reduced the filtered dataset down to one SNP per 1000 bp.

Dataset	Total variant sites	Filtered	Thinned
Only *H. bolanderi-exilis* 'BE'	57,926	7,514	1,183
H. bolanderi-exilis and *H. annuus* 'BE+A'	103,318	8,915	1,095
All samples 'BE+A+P'	131,150	9,593	1,062

3.3.2 Population structure and introgression with *H. annuus*

3.3.2.1 Population structure approaches

ADMIXTURE and fastStructure suggest a fractal pattern of divergence in *H. bolanderi-exilis* based on geography rather than soil type. At K = 2, east and west populations are separated, at K = 3 northern populations become their own group, and at K = 4 southwest populations separate. At higher K values, individual populations become their own group and intermediate or admixed individuals are rare. Both ADMIXTURE and fastStructure generally agree on cluster assignment for lower K values (2 to 4) but above that there is inconsistency between runs and methods. Substantial admixture between *H. annuus* and *H. bolanderi-exilis* was not seen in either ADMIXTURE or fastStructure results (Figure 3-2). At K = 2, *H. annuus* and *H. bolanderi-exilis* are separate groups with the possible exception of the *H. bolanderi-exilis* population G128. ADMIXTURE showed G128 to have 1-2% ancestry from the *H. annuus* group. In fastStructure, this population had slightly elevated *H. annuus* ancestry but was of a lower magnitude (~0.5% admixed ancestry).

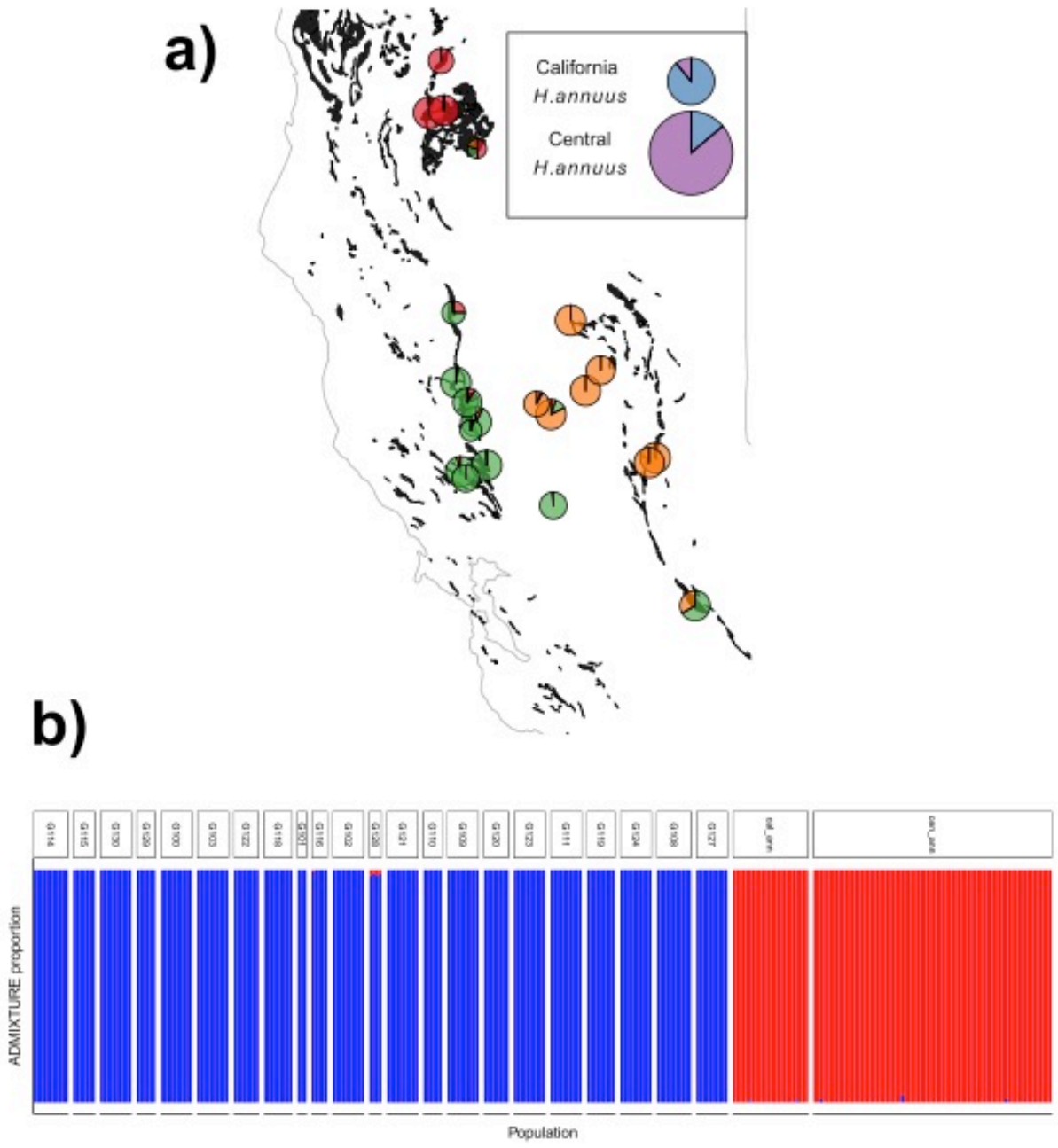


Figure 3-2: Admixture proportions at K=2 and K=5 for BE+A dataset.
a) A map of *H. bolanderi-exilis* locations, with ADMIXTURE proportions (based on the filtered BE+A dataset at K=5) indicated by color pie charts. Admixture group 1 (purple) and group 2 (blue) are only found in *H. annuus* samples. Groups 3 to 5 (red, green and orange) correspond to north, west and east regions respectively. Serpentine locations are highlighted in black on the map. b) ADMIXTURE proportion for K=2 for the filtered BE+A dataset. *Helianthus bolanderi-exilis* populations are ordered by latitude. Group 1 (red) corresponds to *H. annuus* samples and group 2 (blue) to *H. bolanderi-exilis* samples.

Splitstree and PCA recapitulated the results seen in ADMIXTURE and fastStructure (Figure 3-3 & Figure 3-4). For the splits network *H. bolanderi-exilis*, *H. annuus* and the perennial species form monophyletic groups without admixture. In the PCA, the first principal component separated *H. annuus* and *H. bolanderi-exilis,* and the second separated the east and west *H. bolanderi-exilis* populations.

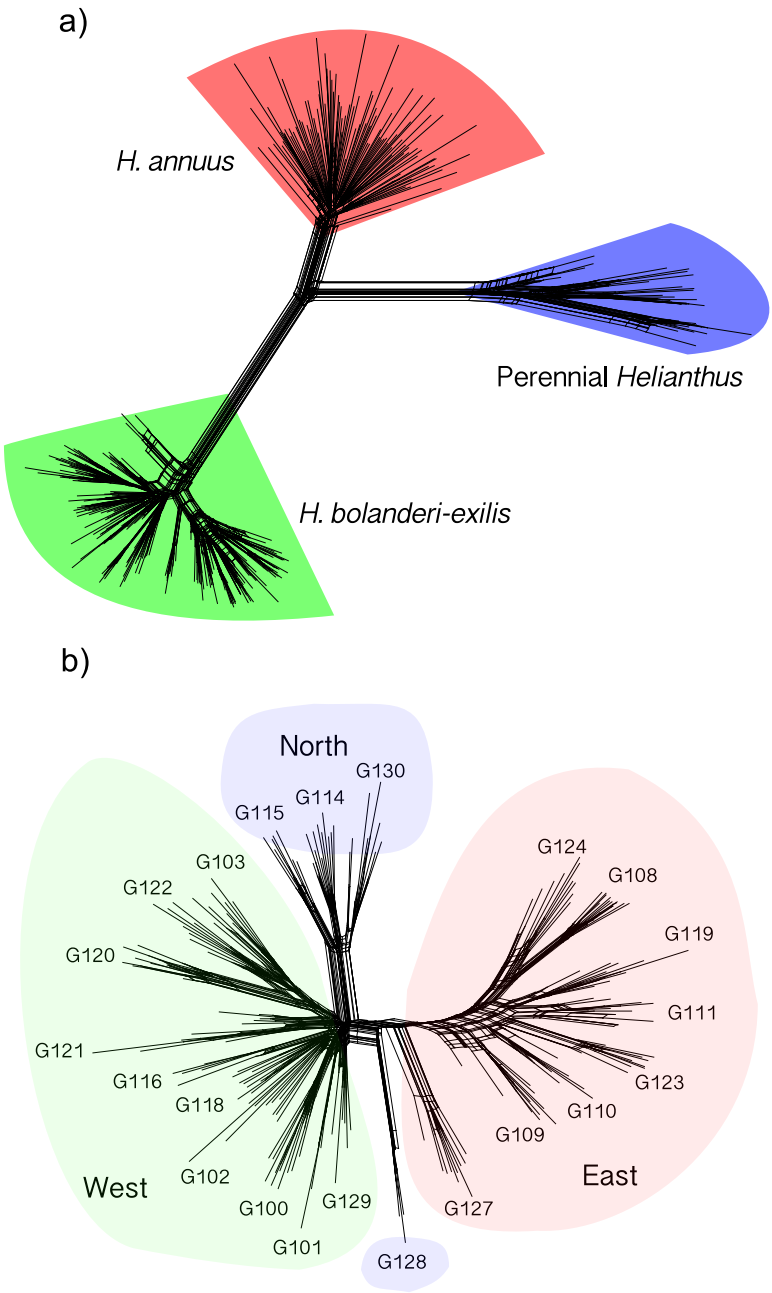


Figure 3-3: Splits network analysis of (a) the filtered BE+A+P dataset and (b) the filtered BE dataset. Network was made using Splitstree4 with uncorrected P-distance.

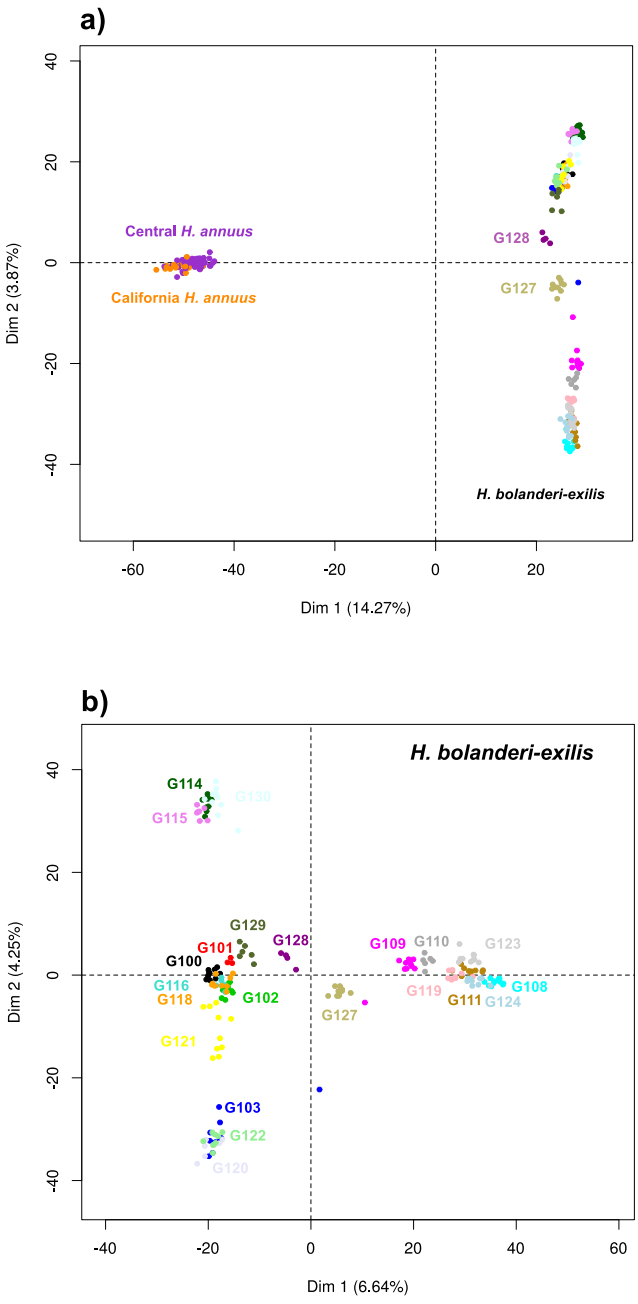


Figure 3-4: Principal component analysis of (a) the filtered BE+A dataset and (b) the filtered BE dataset.
In (a) populations G127 and G128 are labeled because they occupy the most intermediate position in the *H. bolanderi-exilis* cluster.

ADMIXTURE cross-validation testing found K = 8 for BE and K = 6 for BE+A to have the lowest error, although scores were relatively flat from K=5-10. For fastStructure, marginal likelihood was universally maximized at K=2 for BE and K=3 for BE+A. The K value that best explained population structure depended on the run and dataset: BE filtered = 3-5, BE thinned = 3-7, BE+A filtered = 3-4, BE+A thinned = 3-6. We do not further evaluate the best K value beyond the fact that *H. bolanderi-exilis* and *H. annuus* are never placed in the same group and that there is some level of geographic structure in *H. bolanderi-exilis*. The exact best K value to explain the geographic structure is not relevant to our hypotheses.

F_{ST} values between populations of *H. bolanderi-exilis* were high (0.041-0.509, mean=0.331), implying minimal gene flow between geographically distant populations, or population bottlenecks (Table 3-3). Between *H. bolanderi-exilis* and *H. annuus,* F_{ST} was also very high (mean F_{ST} = 0.508 and 0.472 for Californian and central *H. annuus* respectively).

Table 3-3: Weir and Cockerham F_{ST} between all pairs of populations of *H. bolanderi-exilis* and *H. annuus*.

F_{ST}	G100	G101	G102	G103	G108	G109	G110	G111	G114	G115	G116	G118	G119
G100	NA												
G101	0.248	NA											
G102	0.132	0.132	NA										
G103	0.249	0.244	0.143	NA									
G108	0.439	0.475	0.375	0.389	NA								
G109	0.323	0.339	0.249	0.276	0.338	NA							
G110	0.365	0.387	0.279	0.308	0.346	0.178	NA						
G111	0.400	0.418	0.325	0.347	0.327	0.229	0.208	NA					
G114	0.318	0.334	0.242	0.320	0.483	0.370	0.400	0.442	NA				
G115	0.307	0.338	0.225	0.308	0.488	0.369	0.396	0.441	0.214	NA			
G116	0.271	0.292	0.151	0.250	0.465	0.352	0.391	0.416	0.357	0.341	NA		
G118	0.170	0.168	0.041	0.177	0.401	0.276	0.308	0.350	0.261	0.239	0.195	NA	
G119	0.404	0.431	0.331	0.347	0.336	0.247	0.239	0.220	0.446	0.447	0.424	0.355	NA
G120	0.368	0.411	0.271	0.269	0.497	0.394	0.425	0.457	0.420	0.423	0.391	0.296	0.449
G121	0.224	0.226	0.123	0.192	0.413	0.295	0.327	0.372	0.291	0.278	0.229	0.141	0.357
G122	0.296	0.303	0.197	0.138	0.447	0.340	0.369	0.408	0.355	0.354	0.295	0.215	0.404
G123	0.456	0.496	0.383	0.402	0.398	0.279	0.279	0.294	0.482	0.496	0.478	0.413	0.293
G124	0.406	0.421	0.327	0.349	0.251	0.279	0.282	0.267	0.449	0.451	0.424	0.357	0.278
G127	0.335	0.357	0.257	0.297	0.388	0.326	0.341	0.366	0.390	0.394	0.358	0.283	0.365
G128	0.348	0.372	0.255	0.308	0.468	0.343	0.381	0.412	0.395	0.409	0.376	0.287	0.414
G129	0.196	0.177	0.101	0.212	0.414	0.271	0.306	0.353	0.260	0.258	0.236	0.140	0.356
G130	0.302	0.312	0.224	0.297	0.468	0.343	0.377	0.414	0.188	0.214	0.328	0.254	0.422
CalAnn	0.511	0.469	0.481	0.486	0.541	0.508	0.499	0.529	0.539	0.519	0.494	0.490	0.526
CenAnn	0.474	0.443	0.447	0.452	0.499	0.474	0.471	0.490	0.497	0.482	0.461	0.455	0.488

Fst	G120	G121	G122	G123	G124	G127	G128	G129	G130	CalAnn	CenAnn
G100											
G101											
G102											
G103											
G108											
G109											
G110											
G111											
G114											
G115											
G116											
G118											
G119											
G120	NA										
G121	0.295	NA									
G122	0.306	0.227	NA								
G123	0.509	0.420	0.458	NA							
G124	0.461	0.367	0.407	0.354	NA						
G127	0.402	0.297	0.344	0.418	0.340	NA					
G128	0.442	0.313	0.360	0.475	0.421	0.374	NA				
G129	0.333	0.183	0.270	0.414	0.362	0.280	0.258	NA			
G130	0.402	0.275	0.342	0.462	0.428	0.372	0.379	0.254	NA		
CalAnn	0.527	0.494	0.501	0.548	0.523	0.511	0.478	0.475	0.524	NA	
CenAnn	0.488	0.458	0.465	0.505	0.484	0.472	0.445	0.443	0.486	0.067	NA

F_{IS} showed no evidence of inbreeding in *H. bolanderi-exilis* populations, consistent with their self-incompatibility (Appendix B.2). Moderate inbreeding was observed in *H. annuus* and several perennial species, likely because samples from multiple populations were pooled and any population structure will result in increased F_{IS} (Wahlund 1928).

3.3.2.2 ABBA-BABA tests

We found a significant positive D score (suggesting Californian *H. annuus* – *H. bolanderi-exilis* gene flow) for the full dataset (0.123 ± 0.033, p= 1.6e-4) and for all individual *H. bolanderi-exilis* populations (Figure 3-5a). The fraction of the genome shared through introgression was overall 5-8% (f_d = 0.065 ± 0.017). When visualized across the genome, the amount of introgression was variable. In particular, chromosome Ha1 had high amounts of introgression, while introgression was low on Ha2, Ha11, Ha12 and Ha15. When D or f_d is compared with recombination rate in *H. annuus*, there is no association (p > 0.1).

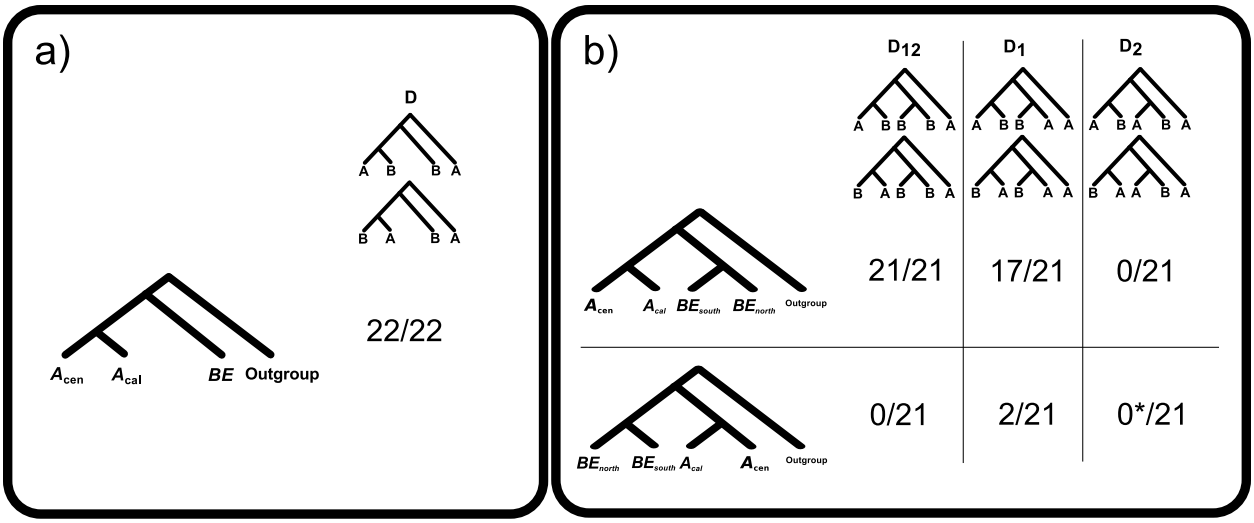


Figure 3-5: Number of significantly positive tests using (a) the Patterson's D statistic and (b) the partitioned D statistic.
(a) Each test uses a separate *H. bolanderi-exilis* population. (b) Each test uses a different *H. bolanderi-exilis* population in the BE_{south} position but keeps BE_{north} constant as G115. Phylogenetic scenarios being compared are included in each test diagram.

When looking at the effect of individual samples, we find positive D scores with 70/76 central *H. annuus* samples, 21/24 California *H. annuus*, and 187/187 *H. bolanderi-exilis* samples (Figure 3-6). Population G128, which exhibited slight evidence of admixture in the ADMIXTURE analysis, showed slightly below average D scores. We find no relationship between collection date or latitude and D or f_d for the California *H. annuus* samples (all $p > 0.12$), but latitude does correlate with D and f_d in *H. bolanderi-exilis* (D: $F_{1,183}=24.0$, $p < e\text{-}5$; f_d: $F_{1,183}=17.3$, $p < e\text{-}4$).

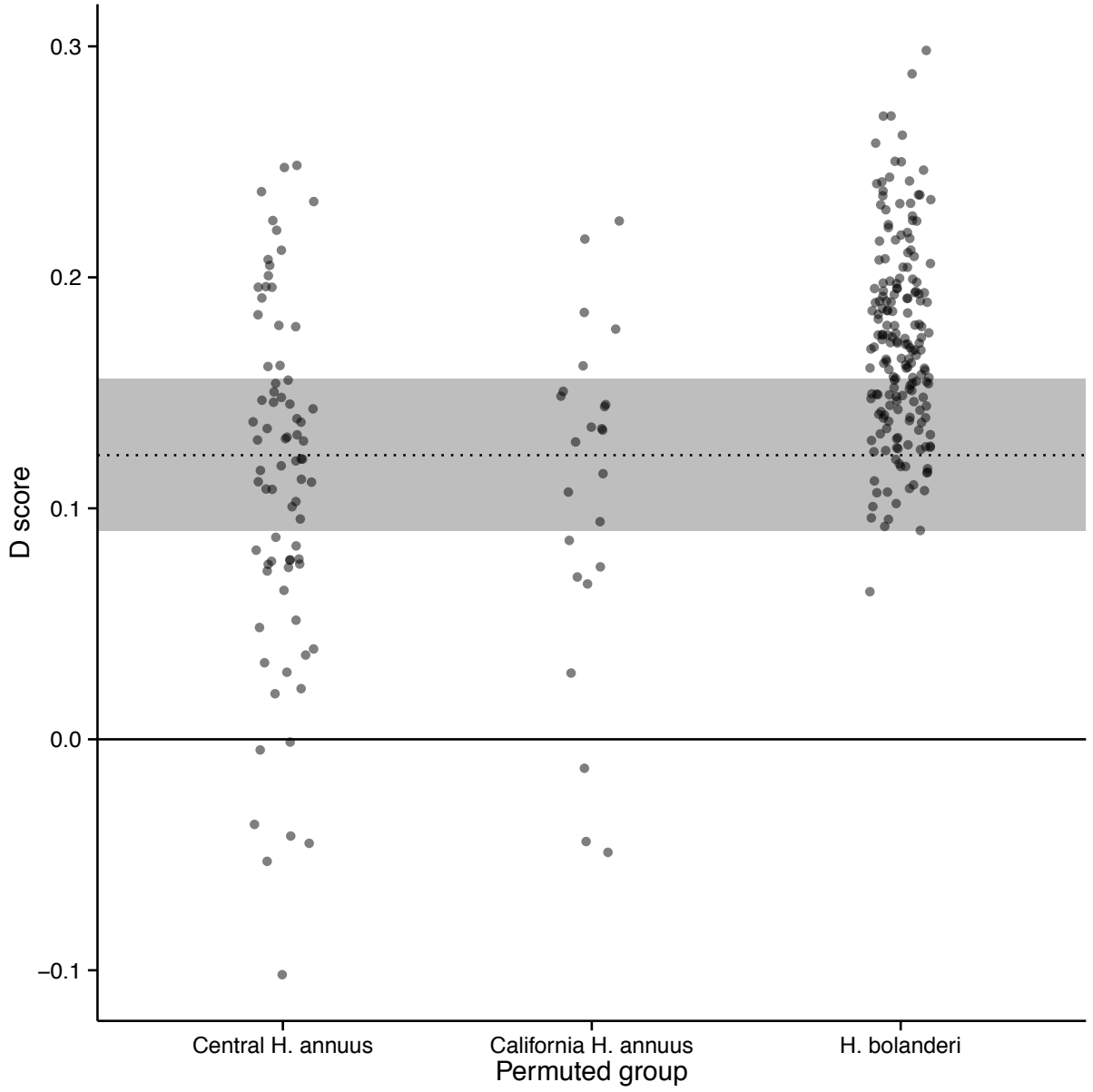


Figure 3-6: Patterson's D scores for subsampled results.
The dotted line represents the D score using all samples ± 1 standard error. The solid line represents the null expectation. Dots represent D scores when testing a single sample from that group.

3.3.3 Directionality of gene flow with *H. annuus*

3.3.3.1 Partitioned D tests

The partitioned D statistic using scenario one produced D_{12}, D_1 and D_2 tests that were significantly positive for 21/21, 17/21 and 0/21 populations respectively. For scenario two, the number of significantly positive populations was 0/21, 2/21 and 0/21 respectively (Figure 3-5b). In scenario two, test D_2, three populations produced significantly negative values. Using G114 as the reference northern population produced similar results.

3.3.3.2 Demographic modeling

Demographic modeling found the most likely model included bidirectional gene flow (Table 3-4). Both the unidirectional gene flow models were better than no migration (into California *H. annnus*: p = 0.0012; into *H. bolanderi-exilis*: p = 0.0059). Bidirectional gene flow was better supported than either unidirectional model (into California *H. annnus*: p = 0.0055; into *H. bolanderi-exilis*: p = 0.0046).

In the best-supported model, effective population size of central *H. annuus* effective is ~880,000, of California *H. annuus* is ~95,000 and of *H. bolanderi-exilis* is ~490,000. The model estimated ~410,000 years ago for the *H. annuus – H. bolanderi-exilis* split and 18,000 years ago for when *H. annuus* invaded California. Migration rates were below 1 migrant per generation (between 0.08 and 0.5).

Table 3-4: Parameters for all δaδi models.
Confidence intervals based on block bootstrapping. Migration is scaled to the number of migrants per generation in the receiving population.

	No migration		Into BE migration		Into CalA migration		Bidirectional migration	
	ML	95% CI	ML	95% CI	ML	95% CI	ML	95% CI
LL	-7494.10	-	-6605.07	-	-7262.47	-	-6464.80	-
Theta	469.66	-	321.84	-	321.85	-	313.91	-
N_{BE} (x 10^5)	5.70	5.65-5.75	4.96	4.85-5.07	4.05	4-4.09	4.94	4.83-5.05
N_{CenA} (x 10^5)	8.46	8.26-8.65	8.77	8.55-8.99	6.07	5.93-6.22	8.80	8.58-9.02
N_{CalA} (x 10^5)	0.97	0.87-1.07	1.21	1.21-1.21	0.49	0.48-0.5	0.95	0.94-0.95
T_1 (x 10^5)	3.15	3.12-3.18	3.97	3.88-4.06	2.36	2.34-2.39	4.14	4.07-4.22
T_2 (x 10^5)	0.19	0.17-0.21	0.22	0.22-0.22	0.10	0.1-0.1	0.18	0.18-0.18
$m_{CalA->BE}$	-	-	0.45	0.44-0.46	-	-	0.48	0.47-0.5
$m_{BE->CalA}$	-	-	-	-	0.11	0.06-0.17	0.08	0.05-0.11

3.4 Discussion:

3.4.1 The non-hybrid origin of *H. bolanderi*

Using our high-resolution genomic data, we can definitively rule out the putative hybrid origin theory of *H. bolanderi*, confirming early work by Rieseberg et al. (1988). Principal component, population structure and phylogenetic network analysis all fail to find evidence for admixture between a subset of *H. bolanderi-exilis* and *H. annuus.* If *H. bolanderi* were of hybrid origin, we would expect some of our sampled populations (particularly those in the eastern part of the range where *H. exilis* is not present) to be genetically closer to *H. annuus*, but we do not see this. This does not mean that there is no gene flow with *H. annuus* and, indeed, our ABBA-BABA testing shows that there is.

As a secondary hypothesis, we evaluated the possibility that *H. bolanderi* had undergone greater introgression with *H. annuus* than did *H. exilis*. The phenotypic intermediacy that motivated the hybrid origin hypothesis might be caused by small amounts of introgression, less than what is typically envisioned for a hybrid-species, and this may not be detected by the coarser population structure or clustering analyses. However, using the ABBA-BABA test, we failed to find support for this possibility as well. All *H. bolanderi-exilis* populations show positive D scores - there is no bimodality that can be attributed to two species, one of which hybridizes (although northern populations show some reduction in D, discussed below). In fact, our results do not support *H. exilis* and *H. bolanderi* as separate species, but are more consistent with a single species with population structure associated with geographic location.

The division between *H. exilis* and *H. bolanderi* has been a point of contention in the literature. Originally (and currently) designated as different species, they have also been classified as two subspecies, and two species plus one ecotype (Grey 1865; Heiser 1949; Jain et al. 1992). Further complicating this, the currently recognized morphological differences between the species, leaf shape, flower head size and seed size, can be confounded by phenotypic plasticity and the stunting effect of serpentine soil making *in situ* species identification difficult. Herbarium records for both species suggest that *H. exilis* is found in the North Coast and Klamath Ranges of California while *H. bolanderi* entirely encompasses that range and extends south and east into the northern Central Valley and Sierra Nevada Foothills. Our genetic data tell a different story.

At the highest level, populations are divided into east and west clades. Although this roughly corresponds to the ranges of *H. bolanderi* and *H. exilis* respectively, both clades are not present in the western range as expected based on current descriptions of species' ranges. Furthermore, the next level of population structure separates the northern populations from the rest, again inconsistent with two overlapping species. F_{ST} between populations is quite high, even for populations relatively close together and all individuals within a population cluster closely within the splits network analysis.

Taken together, this suggests a single species with many isolated populations. Future work should assess phenotypic variation in a common garden and hybrid

sterility for crosses between samples in the eastern, western and northern clades to determine if they are reproductively isolated. It could also establish whether the phenotypic differences purported between *H. exilis* and *H. bolanderi* follow the genetic divides we show here. We tentatively call the combined species, *H. bolanderi*. Both species names were published in the same issue by Asa Grey in 1865, but *H. bolanderi* was listed first and was considered to be the more widespread species (Grey 1865).

3.4.2 Gene flow with *H. annuus*

The genetic data we present here shows evidence for introgression between *H. annuus* and *H. bolanderi-exilis.* Although both population structure and clustering analyses do not show signs of admixture, the Patterson's D statistic is clear that introgression has occurred in California. When testing the effect of individual samples we found the vast majority produced positive D scores (Figure 3-6). This shows that the signal we are seeing is not from ghost introgression in a minority of samples (i.e., the effect of *H. petiolaris* introgression in central *H. annuus*). What the overall D statistic does not tell us is which way gene flow is occurring (e.g. *H. bolanderi-exilis* into *H. annuus*, *H. annuus* into *H. bolanderi-exilis* or bidirectional). To get at the direction of introgression we used the partitioned D statistic with two phylogenetic scenarios (Eaton and Ree 2013). In both of these, we treat the most northern *H. bolanderi-exilis* population as non-introgressed. We make this assumption for two reasons: (i) *H. annuus* is largely limited to the southern half of California and excluded from serpentine regions. The most northern *H. bolanderi-exilis* population (G115) is deep in a

Klamath Mountains, far from the range of *H. annuus* and on a serpentine patch. (ii) The high population structure and isolated nature of populations in *H. bolanderi-exilis* means that gene flow is low between populations and unlikely to have spread introgressed alleles that far in the relatively short period of time that *H. annuus* has been in California.

The partitioned D statistics show that gene flow is largely from *H. bolanderi-exilis* into *H. annuus.* This is seen critically in test D_{12} in both scenarios (Figure 3-5). Scenario one, D_{12} shows that derived alleles present in both *H. bolanderi-exilis* populations are enriched in the California *H. annuus* samples. This must be because of gene flow into *H. annuus* from *H. bolanderi-exilis* because the reverse could not spread the alleles to both populations. One alternative scenario is that gene flow occurred before the *H. bolanderi-exilis* populations diverged, but considering the high F_{ST} between populations of *H. bolanderi-exilis* and recent invasion of California by *H. annuus*, it is highly improbable that *H. annuus* was in California before *H. bolanderi-exilis* spread to its current range. For scenario two, D_{12} is never significant. This shows that the southern populations are not enriched for derived alleles present in all *H. annuus* populations, as would be expected if gene flow was bidirectional. Together these results suggest unidirectional gene flow from *H. bolanderi-exilis* into *H. annuus*. Demographic modeling supports bidirectional gene flow in California (Table 3-4). This is in partial conflict with the partitioned D statistic results. These methods use different ways of detecting gene flow; $\delta a \delta i$ models demographic scenarios that produce similar

site frequency spectra to the empirical data while the partitioned D statistic looks for imbalances in inheritance scenarios within a phylogeny. **δaδi** would not actually use information about shared derived alleles that is driving the partitioned D statistic signal. It is also possible that demographic modeling is affected by the population structure within the *H. bolanderi-exilis* samples. On the other hand, the partitioned D statistic may be under-powered for some scenarios and gene flow may be bidirectional, but unequal (i.e., there is gene flow into *H. bolanderi-exilis* but not enough to detect). Thus we have conclusive evidence of gene flow into California *H. annuus* and ambiguous signals of the reverse; therefore gene flow appears to be stronger into California *H. annuus*.

Theory by Currat et al. (2008) predicts that in this scenario the invader should have more introgressed alleles than the native species. Our results provide support for this theory - introgression does appear to be stronger into the invader *H. annuus*. Although we might expect introgression to be greater in more northern *H. annuus* populations (since they are in greater contact with *H. bolanderi-exilis*) or in populations collected at a later year (if introgression is ongoing), D scores for individual samples are not correlated with latitude or collection date. This is also counter to theory that predicts greater introgression in populations on the range edge (i.e., northern samples). This counter-intuitive result may be because the spread of *H. annuus* across California was not a simple expanding wave and hybridization occurred haphazardly or that hybridization occurred late in expansion and only some lineages were affected.

Furthermore, the model used by Currat et al. does not include reproductive isolation between the species and there is a significant sterility barrier between *H. bolanderi-exilis* and *H.* annuus (Chandler et al. 1986).

The Patterson's D statistic is positive in all *H. bolanderi-exilis* populations, but has regional variation. Specifically, the four northern populations have lower D statistics than the rest (mean 0.126 versus 0.187, students t-test $p < e\text{-}13$). This may be due to introgression in southern and central populations or, more likely, that introgressed alleles in *H. annuus* came from more southerly populations. The amount of introgression is not evenly spread across the genome; several chromosomes do not show evidence of introgression, in particular Ha2, Ha11, Ha12 and Ha15. Previous work has shown associations between low recombination rate and reduced introgression, but we do not see that in our data (Barton 1979; Machado et al. 2007; Yatabe et al. 2007). This may be because we do not have a genetic map of *H. bolanderi-exilis*, so our estimates of recombination rate are missing the major effects of chromosomal rearrangements. Chromosomal rearrangements are known to reduce introgression in sunflowers and other species (White 1978; Rieseberg 2001; Giménez et al. 2013; Barb et al. 2014) and, indeed, pollen sterility and meiotic abnormalities indicate there are several between *H. annuus and H. bolanderi-exilis* (Chandler et al. 1986). Particularly high values of introgression are seen in Ha1, perhaps from positive selection on loci or more neutrally from allele surfing (Hallatschek and Nelson 2008). Alternatively, simulation studies have shown that localized high D values may be due to the reduced

D_{xy} in the absence of gene flow so variation in D may be a side effect of this and not reflect true gene flow variation (Martin et al. 2015).

3.4.3 Edaphic quality and introgression.

The toxicity of serpentine soil excludes *H. annuus* migrants. Consequently, we would expect to see greater introgression in non-serpentine populations of *H. bolanderi-exilis* because both species can co-exist off serpentine sites. In our data this is not the case, Patterson's D scores of non-serpentine samples are not significantly lower than serpentine samples (student's t-test, p = 0.1097). This is consistent with our hypothesis that the samples we sequenced of *H. bolanderi-exilis* are not actually introgressed. Despite this, the hybridization between *H. bolanderi-exilis* and *H. annuus* most likely occurred on non-serpentine soil in California's Central Valley. Populations within the southern extent of this area collected in the 1950s are no longer present possibly due to genetic swamping by *H. annuus*. Extant non-serpentine samples appear to be in danger of a similar fate as *H. annuus* spreads north.

Chapter 4: The genomic composition of sunflower homoploid hybrid species

4.1 Introduction

Hybrid speciation is an extreme example of the constructive effects of hybridization (Mallet 2007). In homoploid hybrid speciation, hybridization without genome doubling brings together the genomes of two species to produce a third lineage that is reproductively isolated from both parental species. The parameter space allowing hybrid speciation and the resulting genomic composition has been modeled but, despite its emblematic importance for hybridization's role in speciation, the actual genomic consequences of hybrid speciation are largely unknown (McCarthy et al. 1995; Buerkle et al. 2000; Duenez-Guzman et al. 2009; Schumer et al. 2015). Homoploid hybrid speciation is much rarer than allopolyploidization (hybrid speciation with genome doubling), although in recent years more examples of the former have been discovered in both plants and animals (Schumer et al. 2014). One reason why hybrid speciation is thought to be rare is that it both requires and is constrained by hybridization (Buerkle et al. 2000). Initial hybridization is required to combine the parental genomes but it must cease for the new hybrid lineage to achieve reproductive isolation from its parents. There are three non-exclusive theories on how this can occur. The recombinational theory and the sorting hybrid incompatibility theory suggest that novel combinations of preexisting chromosomal rearrangements or

hybrid incompatibilities create a lineage that is intrinsically reproductively isolated from both parents (Grant 1958; Schumer et al. 2015). The 'segregation of a new type isolated by external barriers' theory extends this to extrinsic isolation and proposes that novel combinations of alleles allow the hybrid species to expand to a new niche that is geographically or ecologically isolated from the parents, or provides an assortative mating barrier (Grant 1981). In all cases, during hybrid species formation there should be genomic regions under selection that fix rapidly due to fertility (intrinsic) or ecological (extrinsic) selection.

Beyond the effect of selection during hybrid speciation, several other basic questions about hybrid species remain unexplored. For example, we do not have good estimates of genomic composition. This can range from ~2% admixed as is seen in Heliconius butterflies to 50% if parental contributions are equal (Heliconius Genome Consortium 2012). Similarly, estimates of rate at which hybrid genomes settle, or if they are even completely settled, have not been examined using modern genomic techniques.

The first step to answering these questions is identifying parentage blocks in a hybrid species genome, but this is difficult for several reasons. For one, the allele frequencies of the parents when the hybrids were formed will be different from the allele frequencies measured from contemporary populations. This is due to evolution in the parents, as well as the limits of sampling. It is likely that a hybrid species will form from contributions of a subpopulation of the total parental species. If those

subpopulations are not known, or not sampled, and the parental species have population structure then there will be differences. Additionally, hybrid species are independent evolutionary lineages so evolution since the hybridization event will shift allele frequencies and introduce novel mutations. Programs designed to detect admixed ancestry often make explicit assumptions about Hardy-Weinberg equilibrium (HWE) allele frequencies within groups (i.e., STRUCTURE). Thus if hybrid species are old, genetic drift and possibly selection will cause hybrid genome fragments to differ from HWE and potentially cause spurious results. To overcome this limitation, I designed a likelihood-based algorithm that does not make any population genetic assumptions. It simply uses parental allele frequencies and estimates the likelihood of different levels of admixture proportions of the two parents.

Here I apply this new method to three of the most well characterized cases of homoploid speciation: *H. anomalus*, *H. deserticola* and *H. paradoxus*. Each are hybrids between the common sunflower, *H. annuus* and the prairie sunflower, *H. petiolaris* (Rieseberg 1991). They each also occur on extreme habitats not normally inhabited by their parents. *H. anomalus* grows on sand dunes, *H. deserticola* grows on sand sheets and *H. paradoxus* grows on salt marshes (Heiser et al. 1969). It is thought that through transgressive adaptation to these extreme habitats, the hybrid species each separated from their parents both geographically and adaptively (Schwarzbach et al. 2001; Welch and Rieseberg 2002a; Gross et al. 2004).

To explore these issues, I use transcriptomic data from a range of annual sunflower species to ask a diverse array of questions about the origin(s), genomic composition, and ages of the hybrid lineages. I first ask whether *H. annuus* and *H. petiolaris* have been correctly identified as the parents of each hybrid species, and, if so, what is the proportional parentage in each hybrid species? The answers to these questions allows me to address the general hypothesis that hybrid species' genomes should resemble the more ecologically and morphologically similar parent.

I then explore how parental blocks are distributed across the genomes of the hybrid species. This information allows me to test the expectation that parental blocks should be non-randomly distributed across the genome because of strong fertility and ecological selection during the early stages of hybrid speciation. Likewise, I can assess whether hybrid genomes are more highly recombined than suggested by previous low resolution genome scans and associated simulation studies (Ungerer et al. 1998; Buerkle and Lexer 2008) and whether the hybrid genomes are completely stabilized potentially resolving a conflict between the relatively large effective population sizes reported for the hybrid species (Strasburg et al. 2011) and expectations from simulations.

Lastly, I determined the relative age(s) of the hybrid lineages and the overall similarity of their genomes with respect to parental chromosomal block distributions. This information offers a means for testing Schemske's (2000) proposition that most hybrid lineages, including the sunflower hybrids targeted by this paper, are recent

products of human disturbance. I also can assess the repeatability of hybrid speciation, thereby expanding our understanding of the predictability of evolution.

4.2 Methods

4.2.1 SNP preparation

I analyzed sequence variation in 101 transcriptomes from 9 annual *Helianthus* species (Table 4-1). Transcriptome sequencing of the wild species has been previously described (Lai et al. 2012; Renaut et al. 2013; 2014). RNA extractions, library preparation, and sequencing using the Illumina platform were carried out following Lai et al. 2012. Reads were trimmed using Trimmomatic (Bolger et al. 2014) using the sliding window option and final minimum read length of 36bp. Orphaned reads, those whose pair was entirely removed, were not included in analysis. Reads were aligned against a *H. annuus* reference genome (HA412.v1.1.bronze.20141015), using the Burrows-Wheeler Aligner (BWA) (Version:0.7.9a) using the 'aln' and 'sampe' command (Li and Durbin 2010). Alignments were refined using the command *subjunc* in the subread program to account for alignment issues derived from splicing (Liao et al. 2013). Alignments were converted to binary format using SAMtools (Version: 0.1.19) (Li et al. 2009). Read group information and PCR duplicate marking was completed using Picard (Version: 1.114) (http://broadinstitute.github.io/picard). Genotyping was performed using the 'HaplotypeCaller' and 'GenotypeGVCFs' commands in GATK (Version: 3.3) (Van der Auwera et al. 2002).

For all analyses, SNP data were converted from vcf format to tab separated using custom perl scripts. Only bialleleic sites were kept. Sites were discarded if either 'MQ' or 'Qual' were < 20 and individual genotypes were discarded if they had <= 5 or > 100,000 reads.

4.2.2 Sample diagnostics

I used SAMtools (Version: 0.1.19) to quantify the percent of reads aligned and custom scripts to count the number of bases genotyped in each sample. To visualize the phylogenetic relationships between samples I filtered the dataset for coverage (> 95%), minor allele frequency (> 2%) and observed heterozygosity (< 60%) and used SplitsTree4 (Huson 1998). For heterozygous sites, a single random allele was chosen. Samples that did not cluster with their predicted species, in this or previous phylogenetic networks were removed.

4.2.3 Parent determination

It is accepted that the parents of each *Helianthus* hybrid species are *H. annuus* and *H. petiolaris* (Rieseberg 1991). This is based on species distributions (both species have large ranges that overlap with the hybrid species' ranges) and early genetic markers, but this has not been formally tested with modern data. It is possible that the parents of the hybrids may be a close relative of either purported parent (assuming substantial historic range shifts) or the ancestor of multiple species (if the hybrid speciation event is older than the most recent speciation event). To evaluate this hypothesis I calculated pairwise genetic distance between each hybrid individual and

each individual of the potential parent species. These included *H. annuus* and its two closest relatives, *H. argophyllus* and *H. bolanderi*, and *H. petiolaris* and its two closest relative *H. debilis*, *H. praecox*. All sites with data were used. A permutation test (n=10,000) was used to compare the presumptive parents (*H. annuus* and *H. petiolaris*) with other possible parents to determine which had lower mean genetic distance.

For this and other analyses, I used a subset of transcriptomes available for *H. annuus*. The full dataset includes elite, landrace, wild, weedy and *texanus H. annuus* samples. I did not use elite or landrace samples because the domestication process modifies allele frequencies and does not represent the true species wide diversity. Additionally, interspecies gene flow is known to have occurred during improvement (Baute et al. 2015). Samples from Texas, identified as *H. annuus-texanus* were also not used because this subspecies is known to have introgression from *H. debilis* (Rieseberg et al. 1990b).

4.2.4 Parentage proportions

Once the parents of the hybrid species were confirmed to be *H. annuus* and *H. petiolaris* (see Results), I then asked what proportion of the genome for a hybrid individual came from each parent. To do this I selected sites with fixed differences in the parents and asked which parent the allele in the hybrid individual came from. Biases may be introduced from uneven sampling of the parents, so I implemented a dynamic subsampling procedure.

At each site, I counted the number of genotyped samples for each parental species. I then took the lower number and randomly selected that number of genotyped samples from each parental species. This ensures that the sample size is balanced. Since coverage is not equal across the genome in each sample, using this method allows for more sites to be kept than if I had just removed samples from the overrepresented parent from the start. Furthermore, it also removes the chance of sample selection bias, since all samples are still represented in the dataset. This subsampling procedure was also used in the hybrid genome composition analysis (see below).

4.2.5 Parental window assignment

I assigned parentage to genomic regions in individual hybrid samples using a maximum likelihood approach in a sliding window. The analysis was run twice, once with a non-overlapping window size of 1 Mb and once with a window size of one gene. At each site I required at least five samples of each parental species to be genotyped or the site was skipped. Parental samples were also dynamically subsampled (see above). I then calculated allele frequencies for both *H. annuus* (p^1) and *H. petiolaris* (p^2). If an allele was not present in one parental species, I assigned it a frequency of 0.01 to represent the possibility of missed alleles and to facilitate the likelihood approach. For admixture values, x, from 0 to 1 (representing 100% *H. annuus* to 100% *H. petiolaris*) in increments of 0.01, the log likelihood was calculated using the following formulae:

$$lnL(x) = \sum_{i=1}^{n} \ln\left(AA_i \cdot HWE(AA)_i + Aa_i \cdot HWE(Aa)_i + aa_i \cdot HWE(aa)_i\right)$$

$$HWE(AA)_i = ((p_i^1 \cdot x) + \left(p_i^2(1-x)\right))^2$$

$$HWE(Aa)_i = 2(((p_i^1 \cdot x) + \left(p_i^2(1-x)\right)) \cdot (\left((1-p_i^1)x\right) + \left((1-p_i^2)(1-x)\right)) \quad 4.1$$

$$HWE(aa)_i = (((1-p_i^1) \cdot x) + \left((1-p_i^2)(1-x)\right))^2$$

LnL, the log likelihood, is summed over the *n* sites, where *AA*, *Aa*, and *aa* represent the number of homozygous major allele, heterozygous and homozygous minor allele in individuals, respectively, and p^1 and p^2 are the major allele frequencies for *H. annuus* and *H. petiolaris*, parents respectively. Ultimately this produces a likelihood curve of *x* for each sample in each genomic window (Figure 4-1).

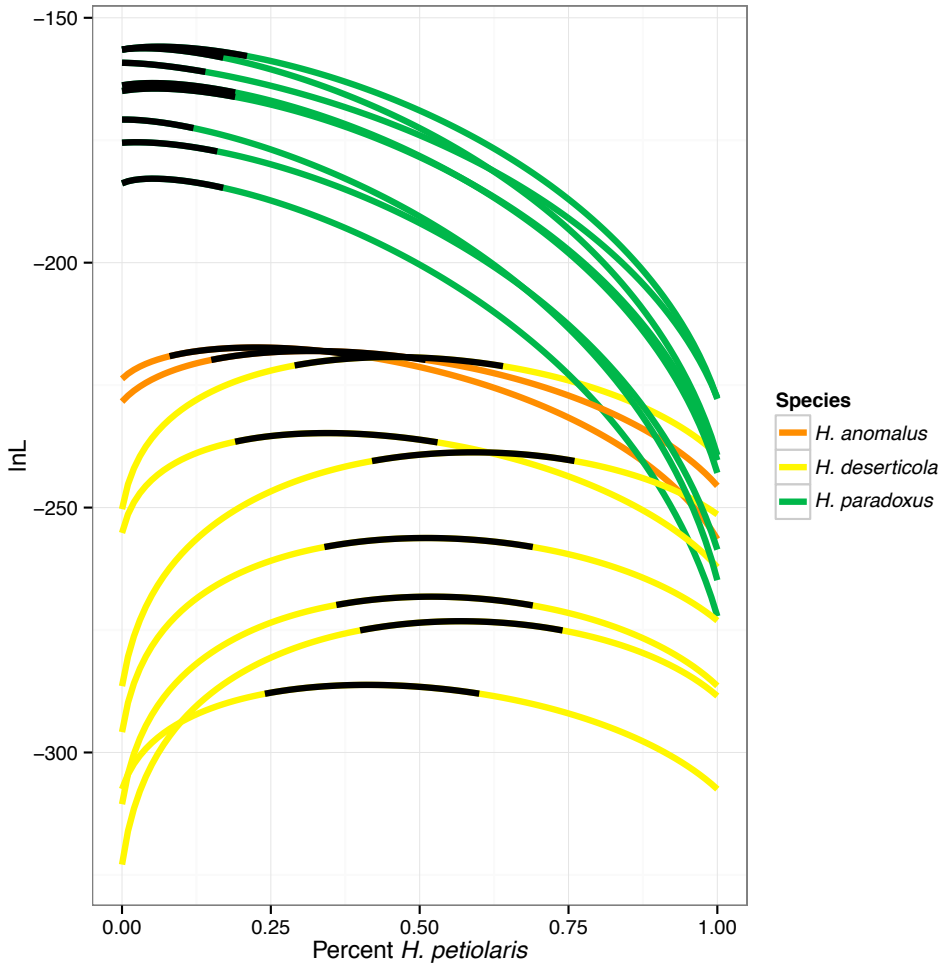


Figure 4-1: An example likelihood curve for one genomic window.
Helianthus paradoxus **(Par) samples are more likely to be from H. annuus, while *H. anomalus* (Ano) and *H. deserticola* (Des) are most likely to be admixed. Black area represents the chi-squared confidence interval.**

The maximum likelihood admixture value was found for each window and a 95% confidence interval was measured using a chi-squared test (df = 1, α = 0.05). This same analysis was repeated on a per gene basis, instead of a sliding window.

After confidence intervals were calculated, they were used to categorically divide windows in types. Windows where the confidence interval was wider than 0.5 (i.e., it

covered greater than half of the possible admixture values) were classified as "unknown". Windows where the confidence interval entirely fell below 0.5 were classified as "*H. annuus*", windows entirely above 0.5 were classified as "*H. petiolaris*" and windows that spanned 0.5 were classified as "admixed". Genetic map positions of chromosomal rearrangements between *H. annuus* and *H. petiolaris* were compared with admixture values (Kate Ostevik, unpublished).

To determine the approximate size of parental blocks, I calculated the cM size of consecutive blocks of the same parentage. Each admixed window was treated as its own block because it may represent multiple smaller parental blocks. Blocks were extended across "unknown" windows as these may be the result of a lack of data and not admixture.

4.2.6 Age of hybrid speciation

In the phylogenetic network analysis, all hybrid species had long branch lengths. This suggests that the hybrid species may be older than previously estimated. To roughly estimate the age of hybrid speciation, I calculated average genetic distance between all species at all genes. For a site to be included, it must have been genotyped in two individuals per species. Since intraspecific variation may contribute disproportionately to genetic distances, I subtracted the average intraspecific variation (i.e., $\boldsymbol{\pi}$) of the two species, from each genetic distance measure, effectively calculating the net nucleotide distance (Arbogast et al. 2002).

Hybrid species have genes from both parents, and comparing the genetic distance of a hybrid species to the wrong parent (i.e., the parent that did not contribute the allele) would incorrectly increase genetic distance, therefore I selected genes for which the parentage was confident and consistent among all samples of that hybrid species. Thus for each hybrid species I made a list of *H. annuus* and *H. petiolaris* genes. I took the net nucleotide distance for each gene against its purported parent and normalized it against the net nucleotide distance for that gene between *H. bolanderi* and *H. praecox*. I did not use the net nucleotide distance between *H. annuus* and *H. petiolaris* because ongoing gene flow has reduced overall divergence (Strasburg and Rieseberg 2008). *Helianthus bolanderi* and *H. praecox* are close relatives to *H. annuus* and *H. petiolaris* respectively so they diverged at the same time as *H. annuus – H. petiolaris*, and they are entirely allopatric so do not exchange genes.

4.2.7 Intraspecific genomic composition similarity

If each hybrid species originated only once, we would expect that genomic composition would be highly similar among individuals of the same species. Alternatively, if a hybrid species originated multiple times we expect similarity to be reduced or non-existent, although subsequent gene flow or parallel selection may influence this (see discussion). To determine whether the three hybrid species had multiple origins, I calculated pairwise correlation coefficients for the maximum likelihood admixture proportions between all samples within a given hybrid species. It is possible for correlations to be artificially increased or decreased due to missing

data; therefore I only included windows where in both samples the confidence interval spanned less than half the total possible range (i.e., < 0.5). This limits the comparison to windows where there is reasonable confidence in the admixture proportion. Both higher (< 0.3) and lower (< 0.7) stringencies were also tried.

4.2.8 Interspecific genomic composition similarity

To measure the similarity of genomic composition between species, I used the same measure as within species, pairwise correlation coefficients for the maximum likelihood admixture proportions. It's possible that there is a baseline correlation coefficient inherent to the analysis based on biases within the parental genomes. For example, a genomic region may be biased towards admixed values if there is very little differentiation between the parents. To control for this bias, I created a baseline correlation coefficient using simulated hybrid species genomes. The simulation modeled recombination events in a genetic map the same size as the *H. annuus* genetic map. Mating was random and the number of recombination events was drawn from a poisson distribution ($\lambda = 1$). For simplicity sake, the population size was set to 100 and was run for 400 generations. After this many generations, interspecific heterozygosity equaled ~0.01. For a single random simulated individual, parental genome fragments were translated into SNPs by drawing random alleles based on the parental allele frequencies for the appropriate parent. This simulated individual was then run through the same sliding window maximum likelihood script. This was

repeated 100 times and then the pairwise correlation coefficients of the maximum likelihood admixture proportions were calculated.

4.2.9 Shared origin of *H. anomalus* and *H. deserticola*

Interspecific consistency comparisons showed surprisingly high similarity in genomic composition between *H. anomalus* and *H. deserticola* samples. To assess whether this represents shared origin versus parallel genotypic evolution, I selected sites in the hybrid species with non-parental alleles (i.e., alleles not found in the parents) and asked whether the non-parental allele was found in more than one hybrid species. I only included sites where > 1 sample was genotyped in each hybrid species. Since *H. anomalus* had only two samples, all hybrid species were randomly subsampled to a sample size of two.

4.2.10 Genomic stabilization

During hybrid speciation, interspecific heterozygosity will decline due to drift and selection. Interspecific heterozgosity begins at 100% in the F1 and is expected to decline to minimal levels within hundreds or thousands of generations depending on effective population size (Buerkle and Rieseberg 2008). To measure the current levels of observed interspecific heterozygosity I selected all sites in the genome where *H. annuus* and *H. petiolaris* were fixed for different alleles. This included the subsampling procedure to balance sample sizes, so some sites used were not actually fixed differences in the entire dataset but were in the subsampled set. At each site I asked if

the hybrid species samples were heterozygous at this site or not and calculated the percent heterozygosity.

4.3 Results

4.3.1 Data quality

The number of reads used and percent of reads aligned for each sample are reported in Table 4-1. SNP calling produced genotype calls for 97,119,366 sites, after removing indels and filtering for genotype quality. This includes 6,240,995 bi-allelic and 438,363 tri-allelic sites. Splits network analysis confirmed species identity in almost all cases. Three samples were removed because they were putative contemporary hybrids ("Sample-Goblinvalley", "btm30-4" and "PET2343") (Figure 4-2).

Table 4-1: Names and read information for samples used in hybrid species analysis. Samples with * were removed from further analyses.

Name	Taxa	Number of reads	Number of reads aligned	Percent of reads aligned
Academy2	*annuus*	16017498	13770972	85.97
Academy7	*annuus*	21442334	18365606	85.65
ALB	*annuus*	20944799	15705557	74.99
Canal2	*annuus*	25474069	21916196	86.03
Canal5	*annuus*	30718522	26905742	87.59
Manteca4	*annuus*	25172855	22048289	87.59
Manteca8	*annuus*	23934604	20978593	87.65
SAW3	*annuus*	9852450	7507049	76.19
LEW1	*annuus*	24414987	19999903	81.92
NEW	*annuus*	22598438	18036863	79.81
TEW	*annuus*	20139861	15677004	77.84
Ano1495	*anomalus*	28492388	23426683	82.22
Sample-Ano1506	*anomalus*	43249420	35401342	81.85
Sample-des1486	*anomalus*	32392091	26570690	82.03
Sample-Goblinvalley*	*anomalus*	61130685	50046806	81.87
arg11B-11	*argophyllus*	20247227	17275312	85.32
arg14B-7	*argophyllus*	26240237	22289440	84.94
ARG1805	*argophyllus*	29055900	24516470	84.38
ARG1820	*argophyllus*	39802820	34013515	85.46
ARG1834	*argophyllus*	26969428	22912998	84.96
arg2B-4	*argophyllus*	21627940	17977941	83.12
arg4B-8	*argophyllus*	34276969	29250330	85.34
arg6B-1	*argophyllus*	32612250	27432294	84.12
btm10-5	*argophyllus*	18725992	15827749	84.52
btm13-4	*argophyllus*	28395684	24005375	84.54
btm17-4	*argophyllus*	26247563	22422066	85.43
btm19-1	*argophyllus*	30529452	26030359	85.26
btm20-8	*argophyllus*	24910558	20646712	82.88
btm21-4	*argophyllus*	27600852	23348716	84.59
btm22-8	*argophyllus*	25425154	21352036	83.98
btm25-2	*argophyllus*	30524207	25655371	84.05
btm26-4	*argophyllus*	19549044	16688985	85.37
btm30-6	*argophyllus*	20892621	17490824	83.72
btm27-3	*argophyllus*	24401569	20418717	83.68
BOL1037	*bolanderi-exilis*	26167182	18062305	69.03
BOL775	*bolanderi-exilis*	31223106	21457040	68.72
G109-13	*bolanderi-exilis*	42985566	36014216	83.78
G109-15	*bolanderi-exilis*	70944350	59454574	83.80
G110-2	*bolanderi-exilis*	44223718	37205650	84.13

Name	Taxa	Number of reads	Number of reads aligned	Percent of reads aligned
G110-3	*bolanderi-exilis*	66483881	55926052	84.12
G111-12	*bolanderi-exilis*	87774745	73253352	83.46
G111-14	*bolanderi-exilis*	107421789	90711457	84.44
Ames7109	*bolanderi-exilis*	26167182	18062305	69.03
EXI2348	*bolanderi-exilis*	44572676	31383970	70.41
EXI2356	*bolanderi-exilis*	39567225	25972904	65.64
EXI2359	*bolanderi-exilis*	29071969	19746853	67.92
EXI2360	*bolanderi-exilis*	29667258	19322507	65.13
EXI2363	*bolanderi-exilis*	33376559	22619177	67.77
EXI2368	*bolanderi-exilis*	20336788	13603176	66.89
EXI2370	*bolanderi-exilis*	29301945	20181556	68.87
EXI2371	*bolanderi-exilis*	25989621	12729594	48.98
EXI2373	*bolanderi-exilis*	21987899	14680951	66.77
EXI2375	*bolanderi-exilis*	30733966	20412868	66.42
RAR43	*debilis*	52013181	41790184	80.35
RAR46	*debilis*	50158666	41128480	82.00
RAR50	*debilis*	40491523	32819939	81.05
RAR55	*debilis*	35884259	27509497	76.66
RAR57	*debilis*	41991840	33969621	80.90
arg4B-14	*debilis-cucumerifolius*	45056394	36991069	82.10
btm33-4	*debilis-cucumerifolius*	27292831	21934794	80.37
btm30-4*	*debilis-cucumerifolius*	43834935	35629423	81.28
Des1484	*deserticola*	26497252	21512525	81.19
des2458	*deserticola*	43794883	34810632	79.49
Sample-Des2463	*deserticola*	35730029	28767695	80.51
Sample-desA2	*deserticola*	37114075	29778015	80.23
Sample-desc	*deserticola*	43723828	35081963	80.24
Sample-des1486	*deserticola*	32392091	26570690	82.03
Sample-DES1476	*deserticola*	39649439	31937316	80.55
Sample-king159B	*paradoxus*	40625152	33529189	82.53
Sample-king1443	*paradoxus*	53965251	44121249	81.76
king141B	*paradoxus*	32300999	26765443	82.86
king145B	*paradoxus*	16060683	13366646	83.23
king147A	*paradoxus*	37575827	30985837	82.46
King151	*paradoxus*	27694489	22691534	81.94
king152	*paradoxus*	23934831	19993734	83.53
King156B	*paradoxus*	36809637	30544925	82.98
GSD1439	*petiolaris*	29987906	20930366	69.80
GSD975	*petiolaris*	20657550	13585241	65.76
ISS19	*petiolaris*	14008016	9535625	68.07
KSG54	*petiolaris*	19264783	13960237	72.47
pet2119	*petiolaris*	47100253	38756757	82.29

Name	Taxa	Number of reads	Number of reads aligned	Percent of reads aligned
Pet2152	*petiolaris*	14130669	9907500	70.11
PET2341	*petiolaris*	41803114	34590157	82.75
PET2342	*petiolaris*	36629070	30033054	81.99
PET2343*	*petiolaris*	35743351	29215050	81.74
PET2344	*petiolaris*	38713386	31503910	81.38
PET-2	*petiolaris*	29564853	24452966	82.71
PET-3	*petiolaris*	27410418	22824081	83.27
pet489	*petiolaris*	42288917	35300159	83.47
Pi468805	*petiolaris*	8326375	5822403	69.93
PI468812	*petiolaris*	31765630	25627245	80.68
PI468815	*petiolaris*	12584477	8791918	69.86
PI503232	*petiolaris*	37859430	30866806	81.53
PI531058	*petiolaris*	28992050	23744474	81.90
PI547210	*petiolaris*	27495956	22292003	81.07
PI586932b	*petiolaris*	10269351	5591196	54.45
PI613767	*petiolaris*	43118173	35359396	82.01
PI649907	*petiolaris*	38150864	31009091	81.28
PL109	*petiolaris*	2578744	1811827	70.26
btm13-6	*praecox-runyonii*	43638715	35662815	81.72
btm14-4	*praecox-runyonii*	31370699	25672601	81.84
btm16-2	*praecox-runyonii*	39048713	32172596	82.39

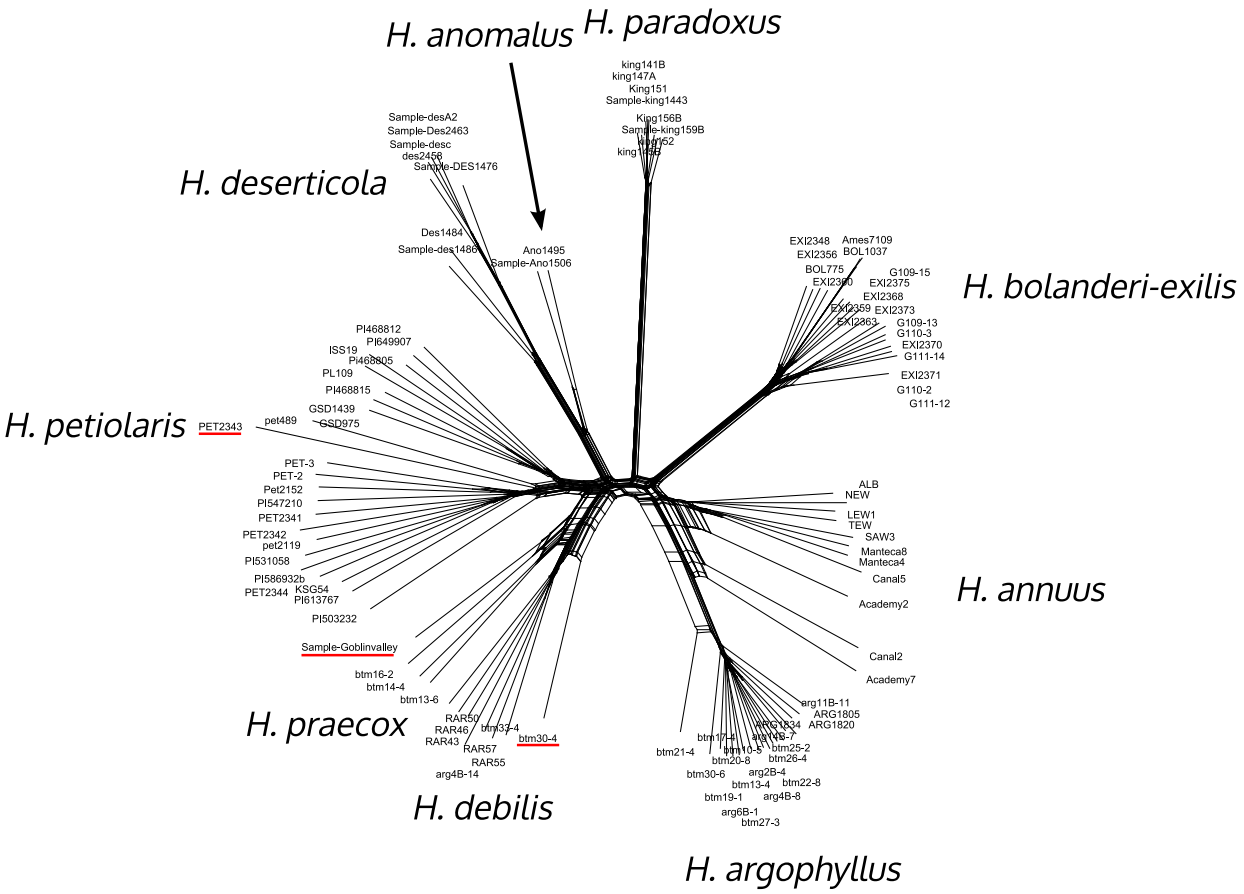


Figure 4-2: Splits network analysis of all EST samples.
Putative hybrids removed from future analyses are highlighted in red.

4.3.2 Parent identification

Based on raw genetic distance, the parents of each of the hybrid species are indeed *H. annuus* and *H. petiolaris* (Figure 4-3). I found that *H. annuus* is significantly genetically closer to each of the hybrid species than *H. bolanderi* and *H. argophyllus* (Table 4-2). On the other side, *H. petiolaris* is significantly genetically closer to each of the hybrid species than both of its closer relative, *H. debilis* and *H. praecox*. Within the

hybrid species, genetic distance was notably lower when comparing *H. anomalus* with *H. deserticola* then in any comparison with *H. paradoxus*.

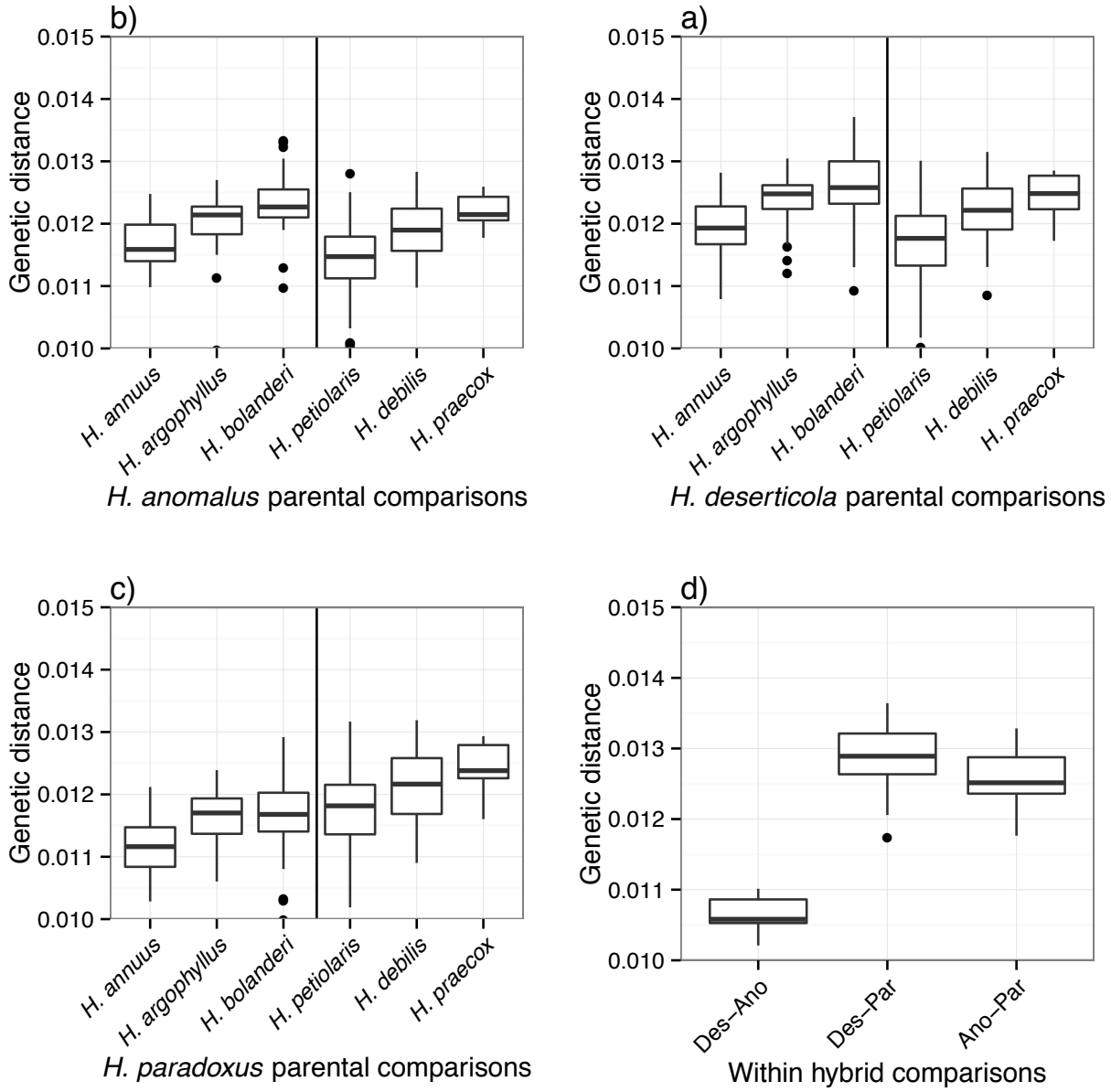


Figure 4-3: Average genetic distance between hybrid species and their potential parents. Genetic distance was calculated using all sites. Includes comparisons with a) *H. anomalus*, b) *H. deserticola*, c) *H. paradoxus* and d) inter-hybrid species comparisons.

Table 4-2: Results for permutation test comparing proposed hybrid parents with possible alternatives. P values < 0.05 are bolded.

	Ann vs Arg	Ann vs Bol	Pet vs Deb	Pet vs Pra
H. anomalus	**p = 0.004**	**p < E^{-4}**	**p = 0.007**	**p = 0.001**
H. deserticola	**p < E^{-4}**	**p < E^{-4}**	**p < E^{-4}**	**p < E^{-4}**
H. paradoxus	**p < E^{-4}**	**p < E^{-4}**	**p < E^{-4}**	**p < E^{-4}**

4.3.3 Genome average parental contribution

The parental contribution was biased toward *H. annuus* in *H. paradoxus* (58-59% *H. annuus*), and for *H. anomalus* and *H. deserticola* the genome was biased towards *H. petiolaris* (62-65% *H. petiolaris*) (Figure 4-4). Novel alleles, sites where the hybrid had neither of the parental alleles, were only present at about 1% of sites.

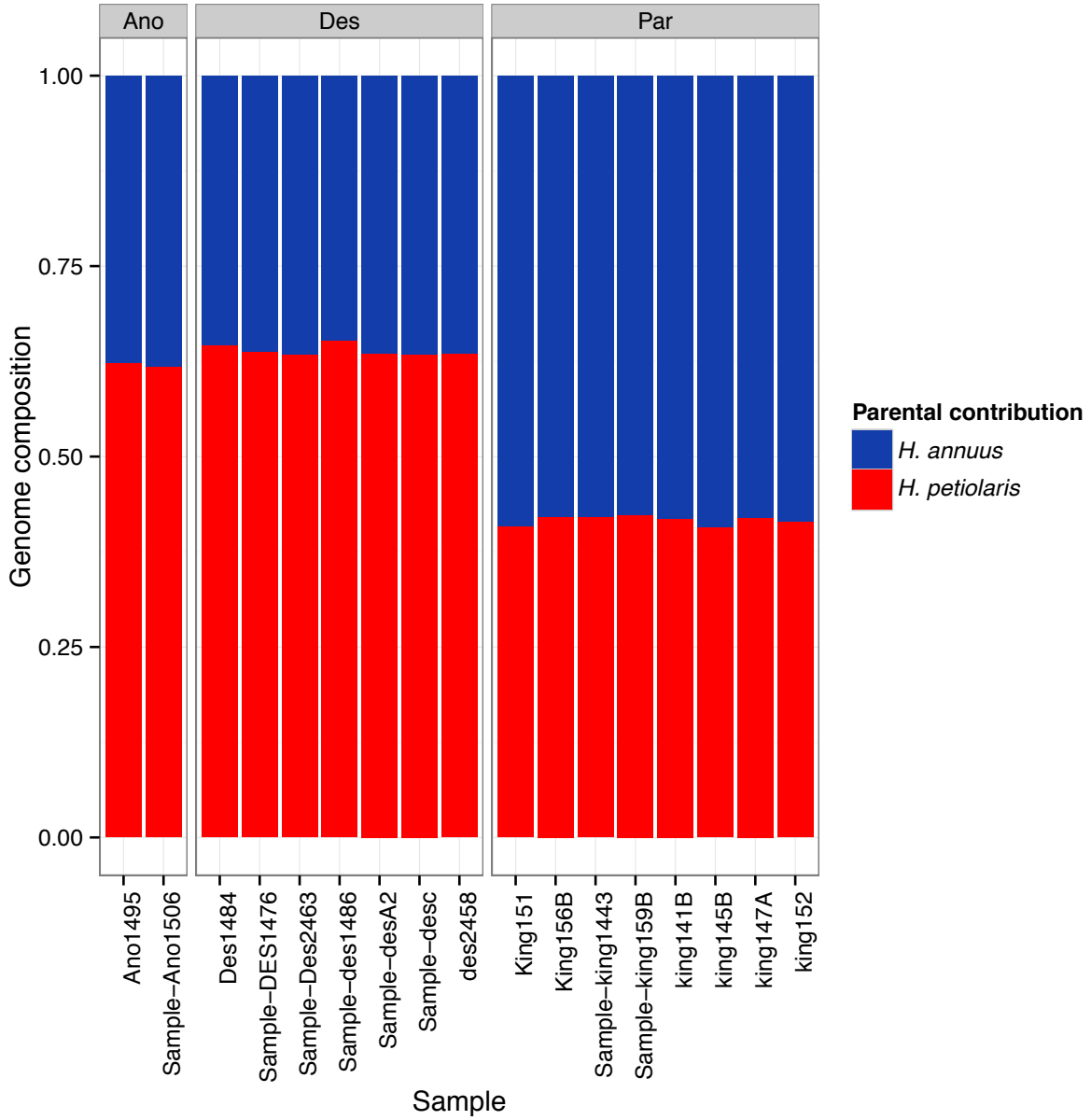


Figure 4-4: Genomic composition of hybrid species.
Calculated using loci with fixed differences between the parental species.

4.3.4 Genomic window parental contribution

I used the maximum likelihood admixture algorithm to assign a confidence interval range of admixture proportions for each genomic window in each sample. This

is summarized in Figure 4-5 which overlays confidence intervals of all samples by species. Genome windows are scaled by cM, and genetic map differences between the parental species are indicated. Values for individual samples are presented in Appendix C. The size of the confidence interval varied (Figure 4-6), but 80% of genomic windows had confidence intervals <= 0.5 admixture value wide (Figure 4-7). Parental blocks were generally very small, most under 1 cM (median size ~0.12 cM), although larger blocks were present in small numbers (Figure 4-8).

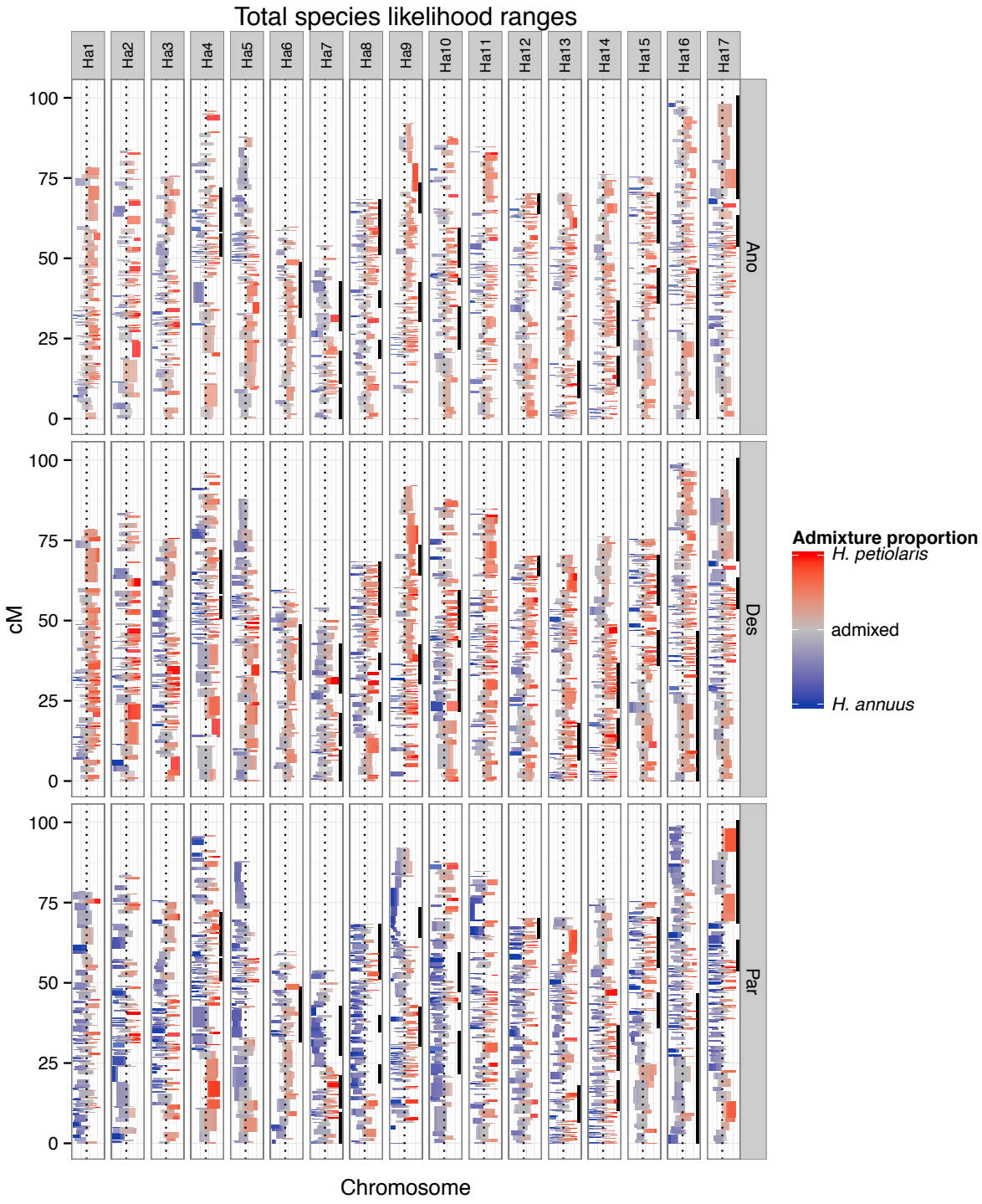


Figure 4-5: Admixture proportion confidence intervals overlaid for each hybrid species.
The width of the bars represents the width of the confidence interval in admixture proportion. The color of the bars represents the maximum likelihood admixture proportion. All samples of a

species are overlaid to represent the average value. Genomic windows are scaled by cM. Genetic map differences between *H. annuus* and *H. petiolaris* are highlighted with black bars.

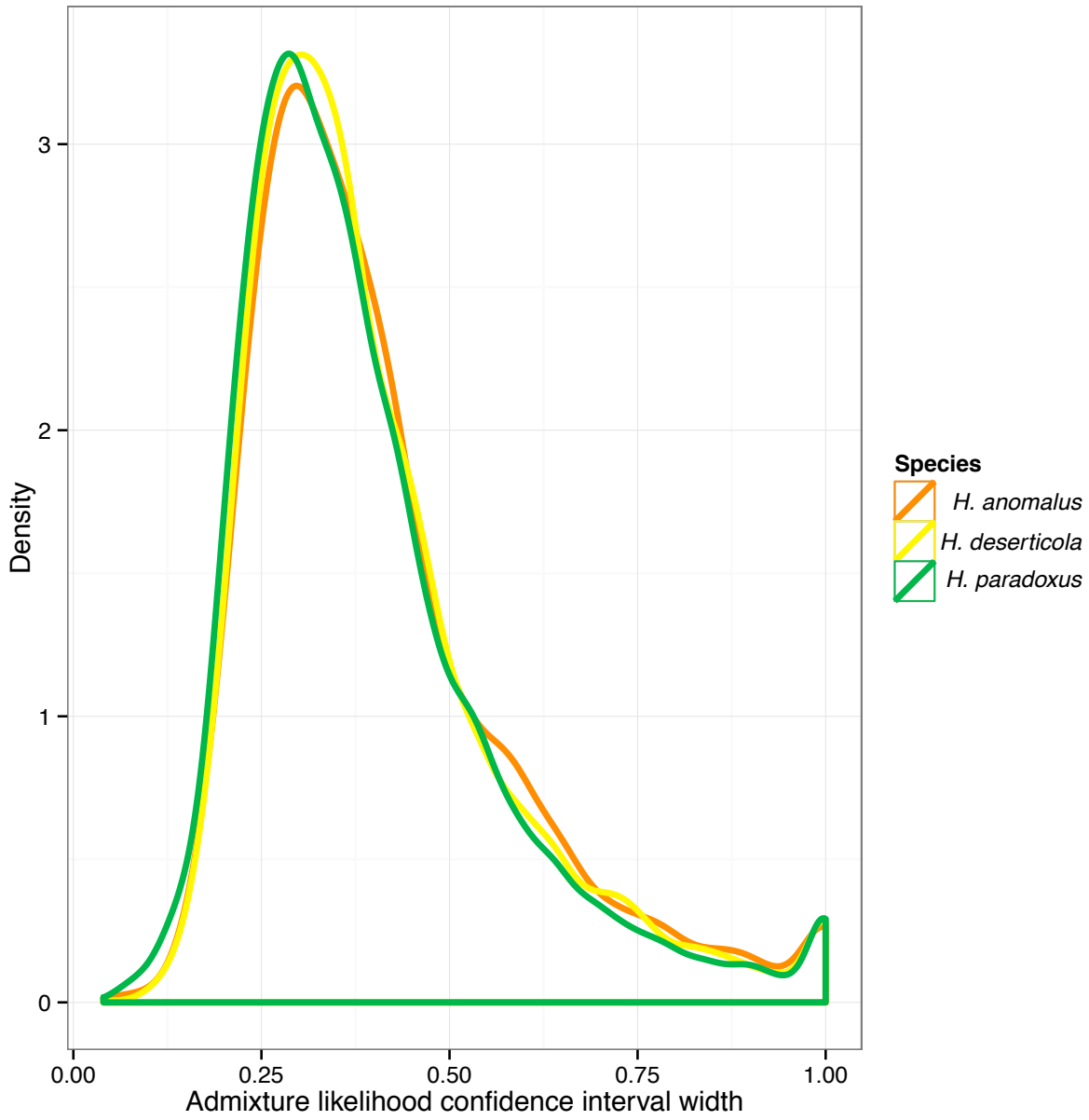


Figure 4-6: The distribution of admixture proportion confidence interval widths by species.

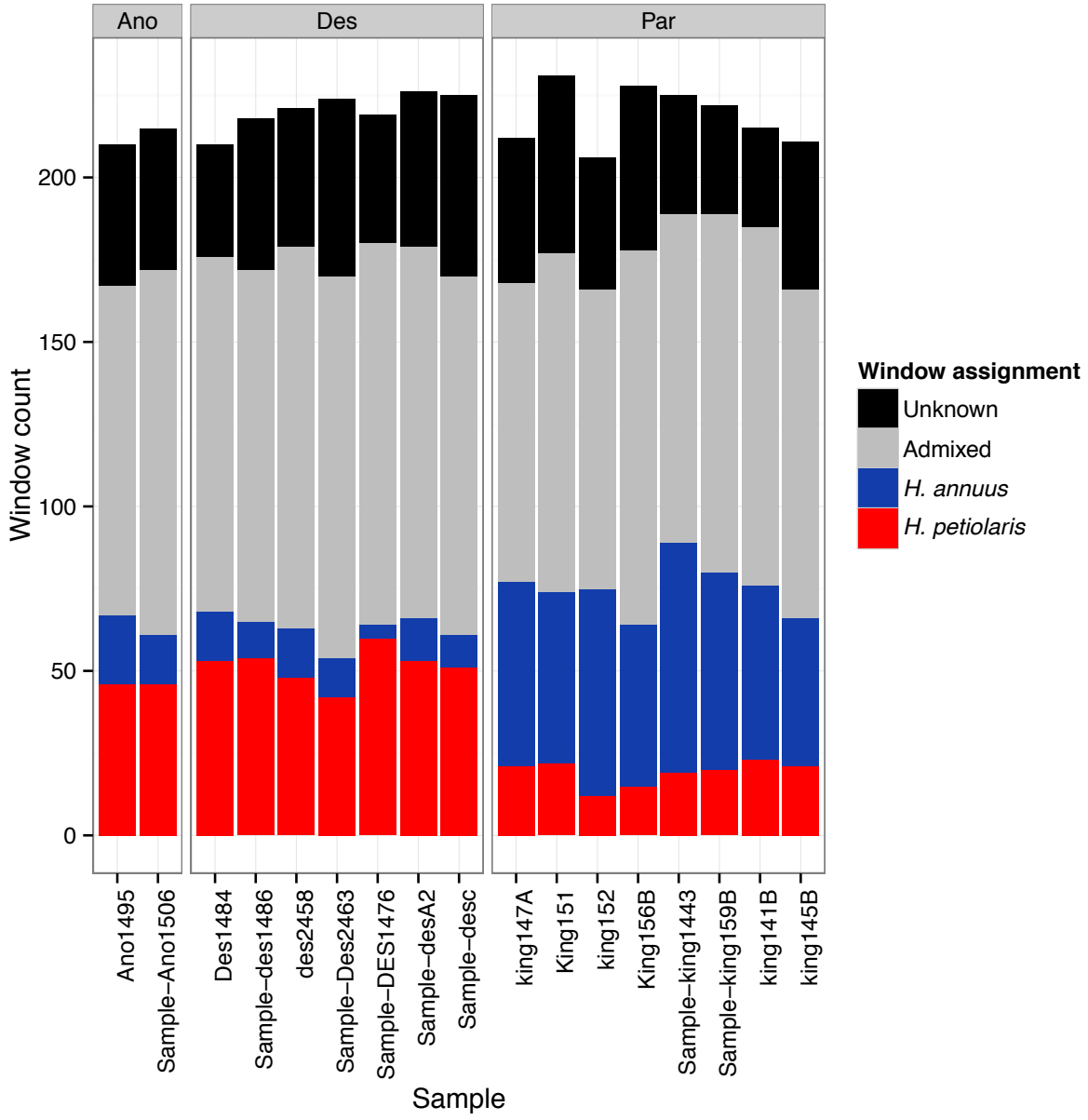


Figure 4-7: Counts of genomic windows in each category.
Unknown: confidence range > 0.5 wide. Admixed: confidence range overlapped 0.5. *H. annuus*: confidence range entirely < 0.5. *H. petiolaris*: confidence range entirely > 0.5.

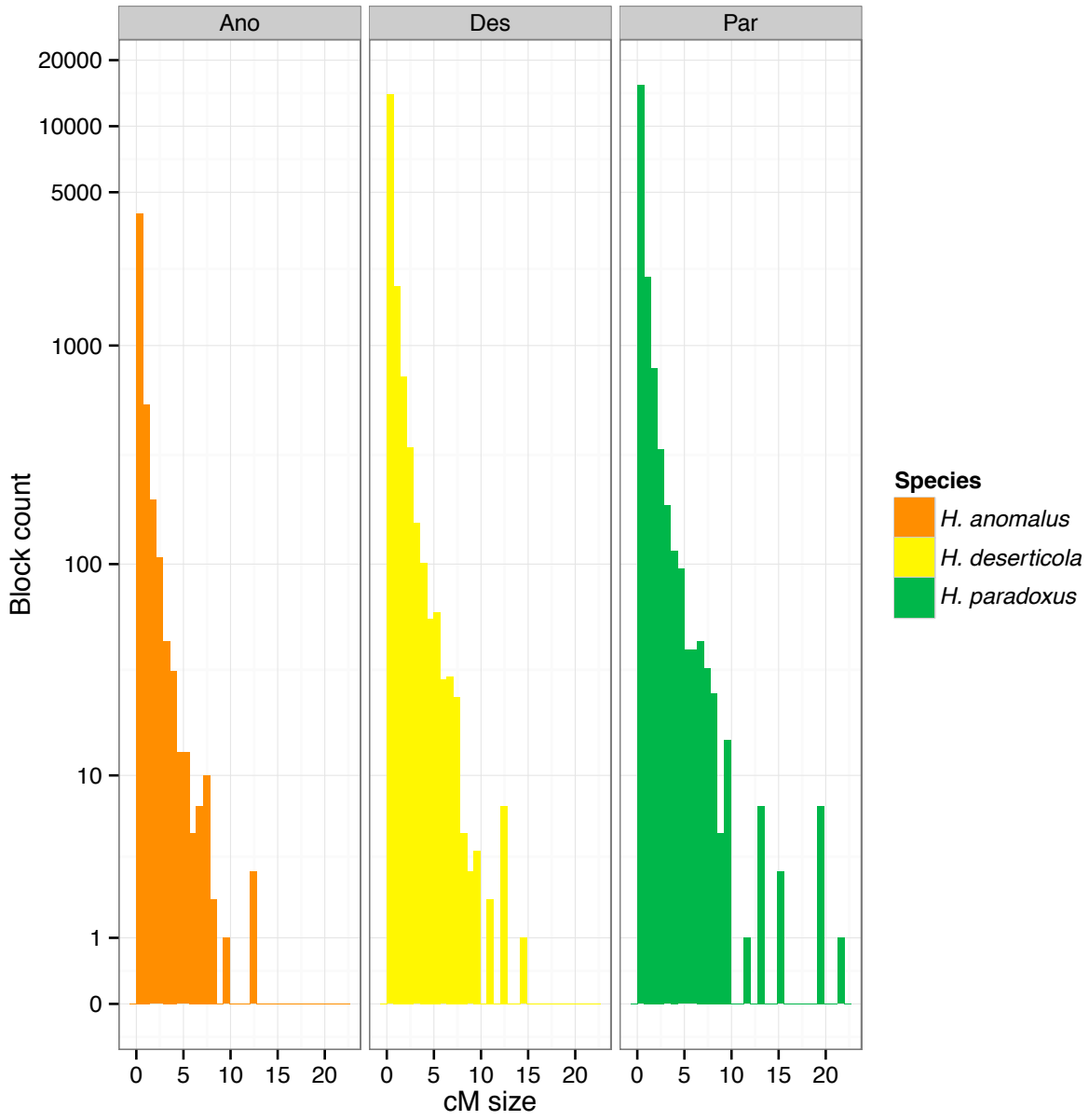


Figure 4-8: Parental block size in hybrid species.
Block size was measured using consecutive 1 Mb windows with the same parentage. Unknown windows were not considered. Admixed windows were their own individual blocks.

4.3.5 Age of hybridization

The net nucleotide distance between the hybrid species and their parents is surprisingly high (Figure 4-9). There is considerable variation by gene, but highest density values suggest the genetic distance is roughly ~0.35 to 0.65 times the genetic distance of *H. bolanderi – H. praecox* (Table 4-3).

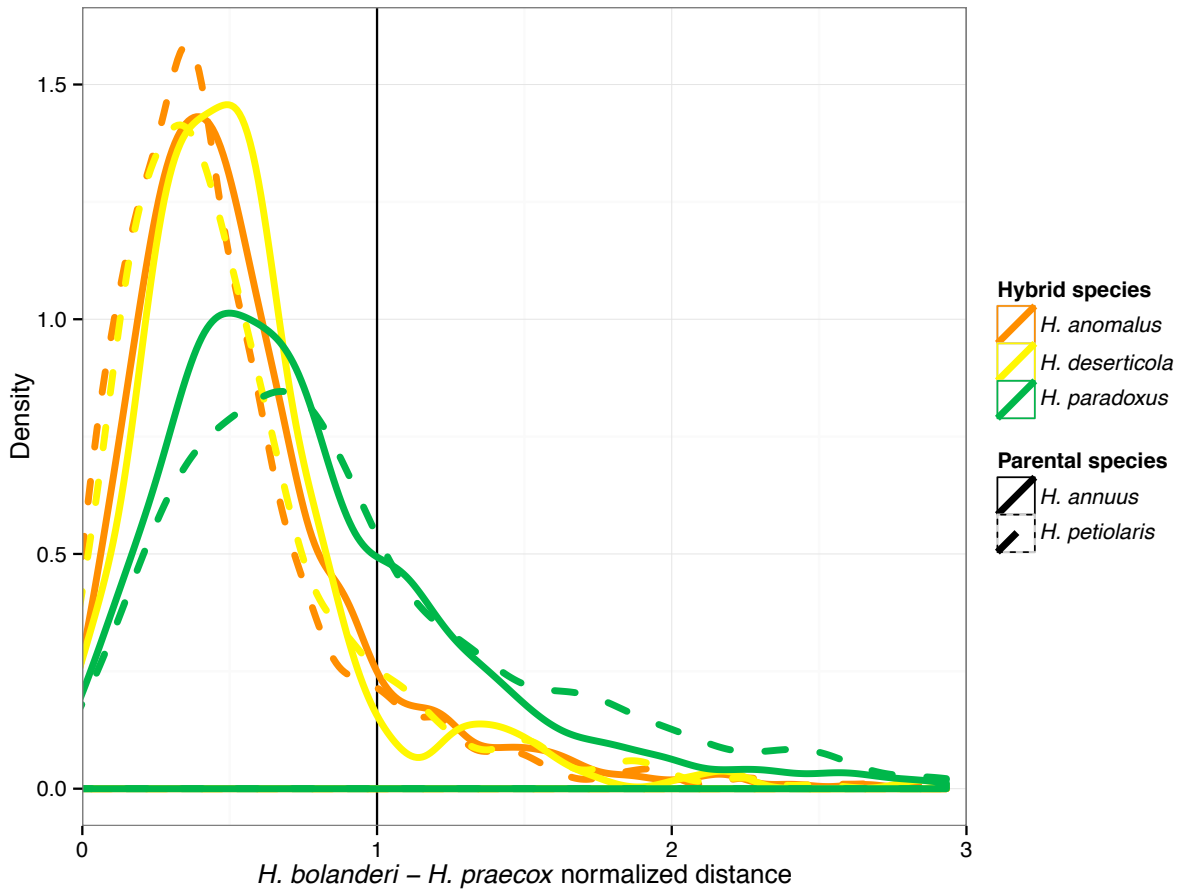


Figure 4-9: Normalized net nucleotide distance between hybrid species and their parents. Net nucleotide distance was normalized by *H. bolanderi – H. praecox* net nucleotide distance.

Table 4-3: Normalized net nucleotide distance between hybrid species and their parents. Net nucleotide distance was normalized by *H. bolanderi* – *H. praecox* net nucleotide distance. Numbers presented are max density values.

	H. anomalus	*H. deserticola*	*H. paradoxus*
H. annuus genes	0.391	0.330	0.508
H. petiolaris genes	0.355	0.489	0.660

4.3.6 Genomic similarity

Genomic composition was highly correlated when comparing samples within a species (mean Pearson's correlation coefficient: *H. anomalus* 0.748, *H. deserticola* 0.851 ± 0.061, *H. paradoxus* 0.924 ± 0.024) (Figure 4-10). Between samples from different species, *H. anomalus* and *H. deserticola* were the most correlated (0.659 ± 0.015) and either compared to *H. paradoxus* resulted in much lower correlations (0.315 ± 0.16 and 0.303 ± 0.019 respectively) (Figure 4-11). Simulated hybrid species resulted in minimal correlations (0.0015 ± 0.081). Increasing the stringency of the window filtering (i.e., only using windows with narrow confidence intervals) universally increased correlation coefficients. Inversely, decreasing the stringency decreased correlation coefficients.

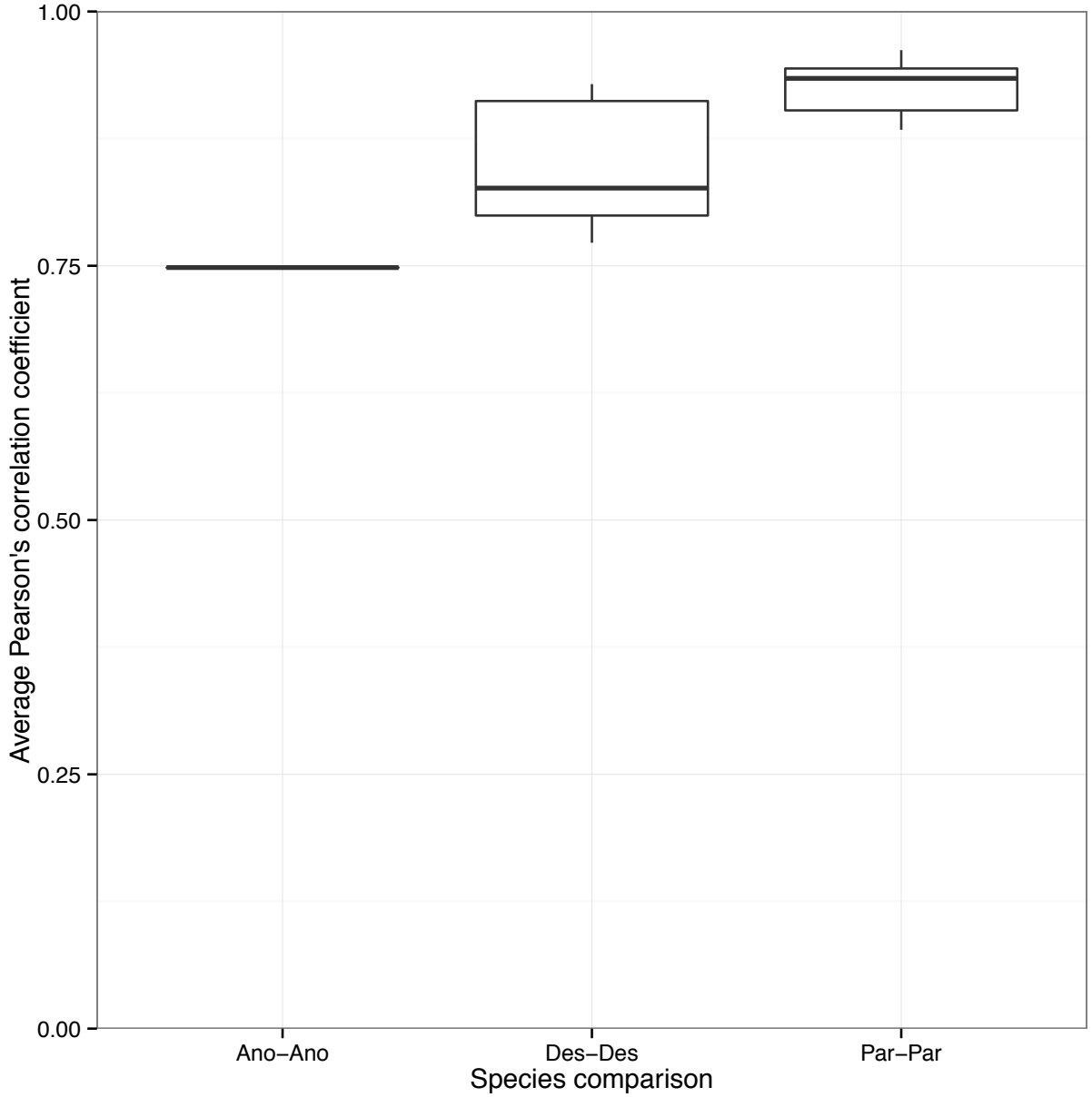


Figure 4-10: Average intraspecies composition correlation.
Values were calculated using the Pearson's correlation coefficient of windowed maximum likelihood admixture proportions. Admixture is measured in 1 Mb windows and only includes windows where both samples confidence intervals spanned less than 50% of the total range individually.

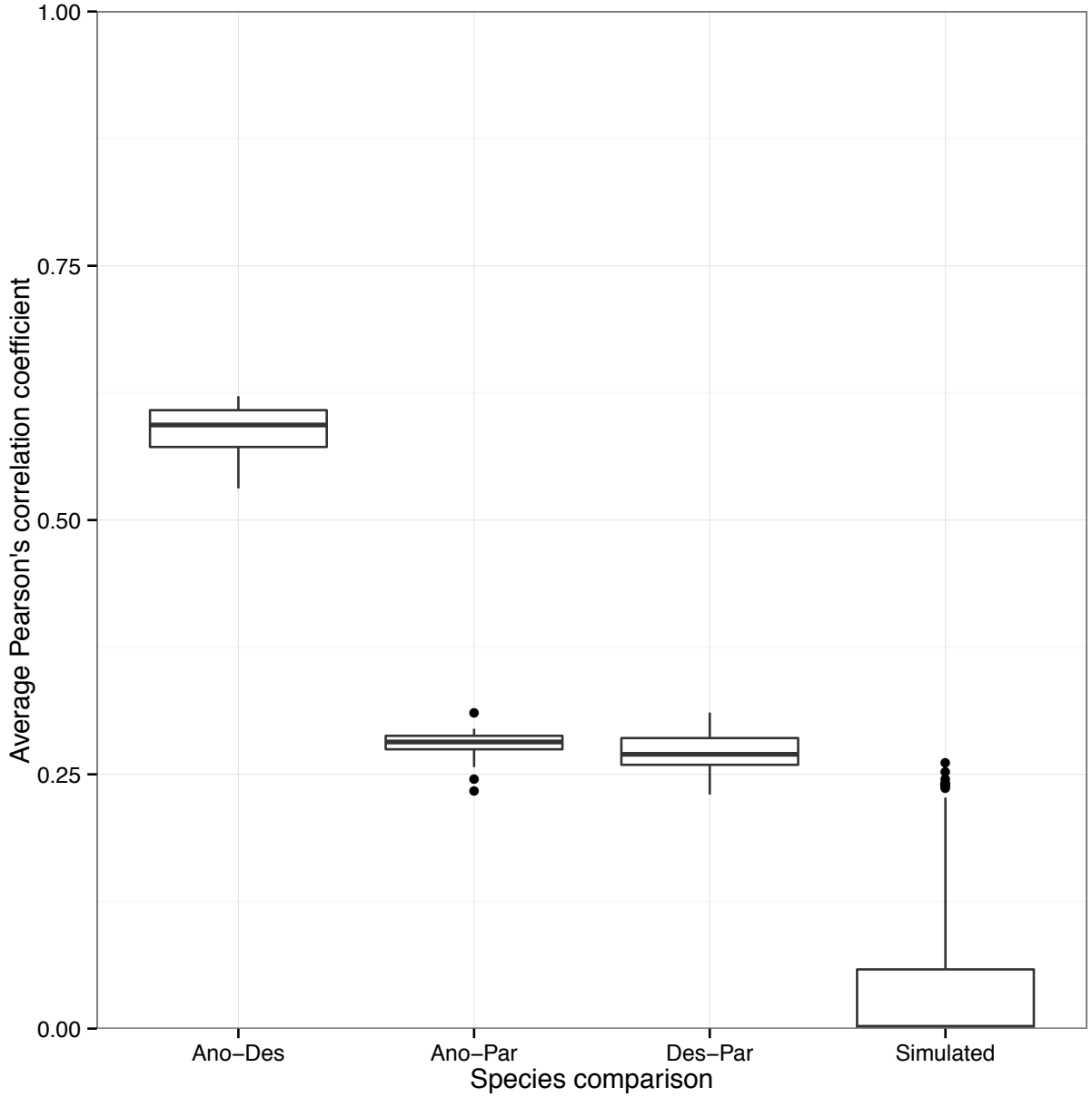


Figure 4-11: Average interspecies correlation coefficient including simulation.
Values were calculated using the Pearson's correlation coefficient of windowed maximum likelihood admixture proportions. Admixture is measured in 1 Mb windows and only includes windows where both samples confidence intervals spanned less than 50% of the total range individually. Simulated hybrid species were created using a population size of 100 and were run for 400 generations.

4.3.7 Shared origin of *H. anomalus* and *H. deserticola*

I identified non-parental alleles that were found in more than one hybrid species. I found that *H. anomalous* and *H. deserticola* shared roughly ten times more non-parental alleles than either did with *H. paradoxus* (Figure 4-12).

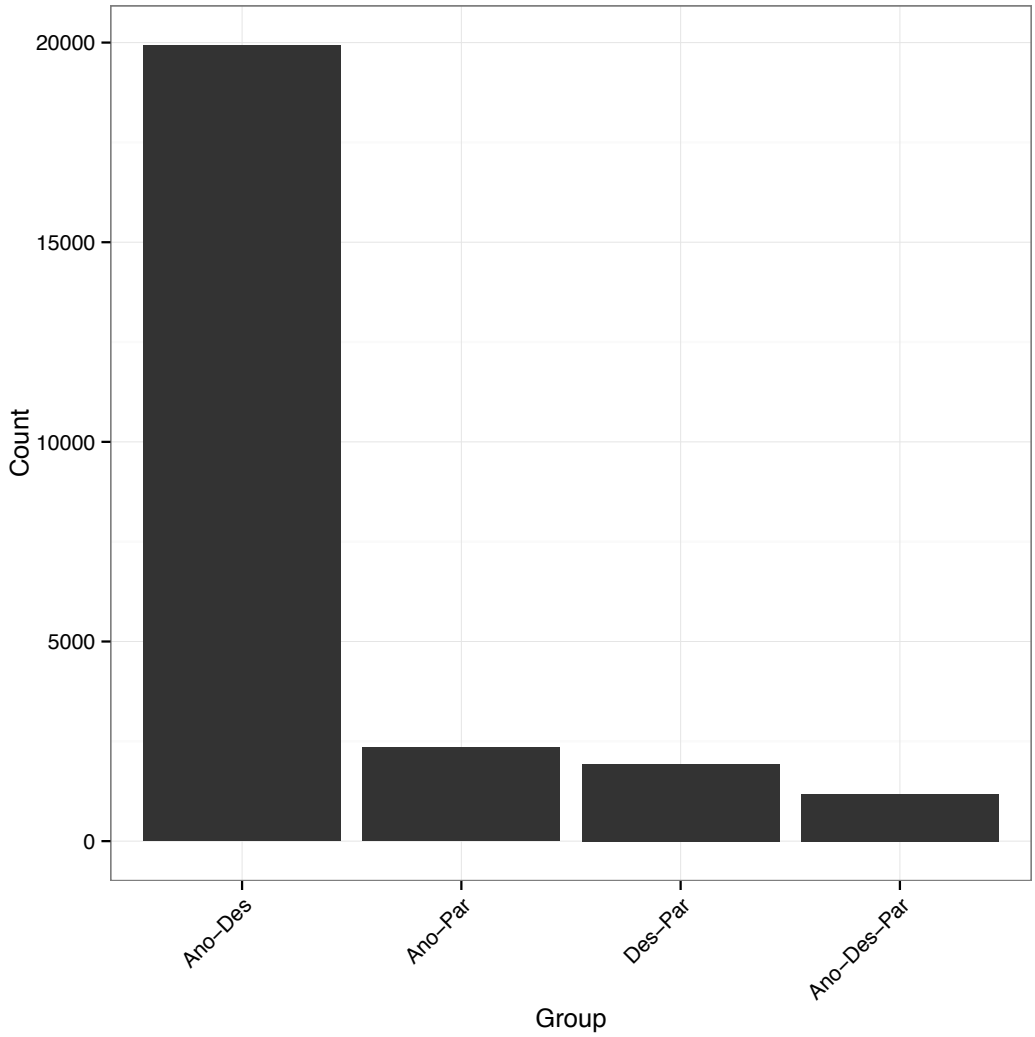


Figure 4-12: Counts of non-parental alleles shared by more than one hybrid species. Ano, Des and Par are *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively.

4.3.8 Genome stabilization

Interspecific heterozygosity was lowest in *H. paradoxus* (mean = 0.0079) and slightly higher in *H. deserticola* (mean = 0.035) and *H. anomalus* (mean = 0.054) (Figure 4-13). For each species this heterozygosity is very low, but appreciably higher than zero.

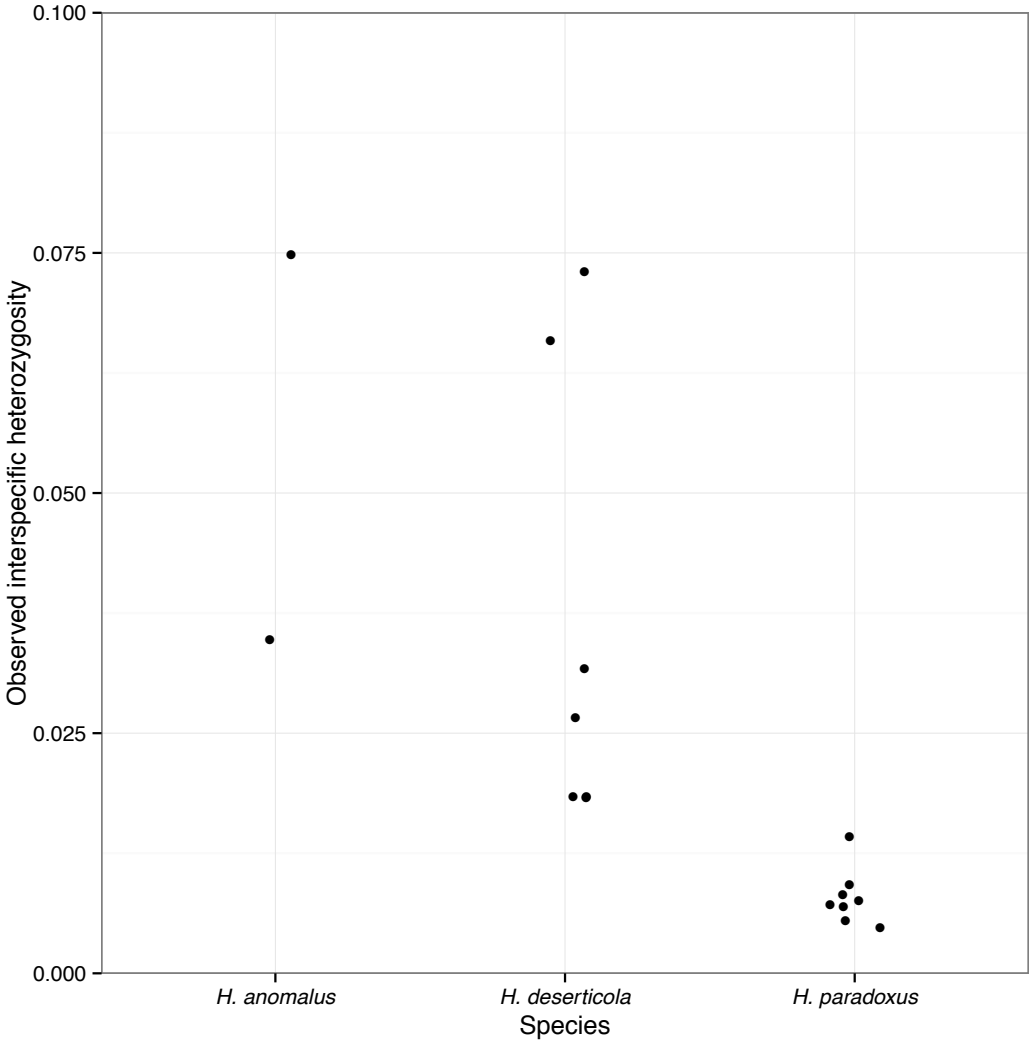


Figure 4-13: Observed interspecific heterozygosity in hybrid species.
Interspecific heterozygosity was calculated from fixed differences between the parental species.

4.4 Discussion

4.4.1 *Helianthus annuus* and *H. petiolaris* are the parental species

Before detailed work is done identifying genomic composition of the hybrids, it is important to confirm that the parental species are correctly identified. Original parentage identification was done using species ranges, morphology and restriction site data (Rieseberg 1991). The proposed parents *H. petiolaris* and *H. annuus* both have large ranges that overlap with all hybrid species, while the other potential parents are regional endemics that do not overlap with the hybrid species ranges. Alternate parents to *H. annuus* include *H. bolanderi*, a native of California and *H. argophyllus*, which is native to Texas. Alternates to *H. petiolaris* include *H. praecox* (native to Texas) and *H. debilis* (native to Texas, Mississippi and Florida). Despite this, it's possible that ranges have significantly changed over the last million years and species range overlaps were different when hybrid speciation occurred. Additionally, hybrid speciation may have occurred before other speciation events (i.e., the parental species may not be *H. annuus* but the ancestor of *H. annuus* and *H. argophyllus*).

The genetic distance scores show that of the petiolaris clade, *H. petiolaris* is the closest relative to each hybrid species (Table 4-2). Similarly, for the annuus clade it is the predicted parent *H. annuus* that is the closest relative (Table 4-2). This suggests that the hybrid speciation events occurred after *H. annuus* and *H. petiolaris* speciated from their nearest relatives and confirms previous hypotheses. Despite this result, it is important to note that these patterns could be driven by gene flow post-hybrid

speciation. Additionally, it is also possible that the hybrid speciation events occurred before the most recent speciation events and *H. annuus/H. petiolaris* are the closest genetic relatives because they have the largest effective population size and undergo the least drift.

Now that I have determined the parental species, I can measure the relative parental contributions to each hybrid species. I find that for both *H. anomalus* and *H. deserticola*, *H. petiolaris* is the dominant parent (62% and 63-65% from *H. petiolaris* respectively), while *H. paradoxus* has slightly more contribution from *H. annuus* (58-59% from *H. annuus*) (Figure 4-4). This is partially consistent with morphological data; both *H. anomalus* and *H. deserticola* are more similar to *H. petiolaris* than *H. annuus*, while *H. paradoxus*, which is slightly more evenly admixed, is roughly intermediate between the two (Rosenthal et al. 2002).

4.4.2 The hybrid genomes are highly recombined

At the beginning of hybrid speciation, parental genome fragments are very large. As the genome stabilizes, genomic regions harboring incompatibilities will tend to fix for one parental version. Before this happens, recombination will break up and intermix parental haplotypes. Thus the speed at which the genome settles will determine the size of parental fragments remaining after the genome has stabilized (Fisher 1954; Stam 1980; Chapman and Thompson 2002). This rate is not necessarily equal across the genome. If there is selection against interspecific heterozygosity, for

example from hybrid incompatibilities, then that genomic region will settle faster and with larger parentage blocks (Buerkle and Rieseberg 2008).

From previous work using sparse genetic maps of each hybrid species, I expected parental fragments to be large (Rieseberg 2003). However, high-density genomic data indicates that the hybrid genomes are highly recombined. A majority of windows show evidence for admixture (Figure 4-7), suggesting that they are not parentally pure across their entire 1 Mb size. I attempted to quantify the size of parental blocks by measuring consecutive blocks of single parentage (Figure 4-8). This distribution almost certainly overestimates the actual block sizes because the minimum block size is determined by window size and therefore the recombination rate. Considering the high number of admixed windows, which are treated as blocks of their own, it is likely that there are numerous actual parental blocks smaller than the minimum size.

Furthermore, the use of the *H. annuus* reference genome raises several possible problems. Reads in hybrid samples from the *H. annuus* parental regions may correctly align at a higher frequency and cause a bias towards *H. annuus* ancestry, consequently causing longer blocks than reality. Conversely, both small and large chromosomal rearrangements in the hybrid species genomes compared to *H. annuus* could cause block sizes to be under-estimated.

With these caveats in mind, there are genomic regions in each species where parental origin is consistent for > 5 cMs. These regions may harbor ecologically

important loci and/or hybrid incompatibilities between the parental species and were under selection during hybrid speciation.

The sorting of chromosomal rearrangements has been implicated as a major source of hybrid incompatibility in each hybrid species (Lai et al. 2005). In contrast, I find little correlation between rearranged regions and patterns of admixture (Figure 4-6). This may reflect variation within in the parental species, either geographically or temporally (i.e., that the actual parents of the hybrid species had different chromosomal structure than the populations used in contemporary genetic maps), karyotypic changes after hybrid speciation, or that selection against heterozygous chromosomal forms was weaker than currently thought.

4.4.3 The hybrid species are old

Because the hybrid sunflowers are arguable the best known examples of homoploid hybrid species, their ages have important implications for hybridization's role in speciation. Human activities are thought to have contributed to a recent expansion in the geographic range of *H. annuus*, leading some to suggest that the hybrid species are the direct result of human disturbance and are consequently very young (Schemske 2000). In contrast, estimates based on microsatellites, suggest they predate human involvement: *H. anomalus* 116,000 – 144,000, *H. deserticola* 63,000 – 170,000, *H. paradoxus* 75,000 – 208,000 ybp (Schwarzbach and Rieseberg 2002; Welch and Rieseberg 2002b; Gross et al. 2003). By comparing the genetic distance between hybrids and their parents and the genetic distance between the allopatric

species *H. bolanderi* and *H. praecox*, I find the hybrid species to be much older than earlier claim. Using the *H. bolanderi* and *H. praecox* divergence, which has been estimated to be 1.8 mya (Sambatti et al. 2012), as a baseline, then the hybrid speciation events are 0.6 to 0.8 mya (*H. anomalus* and *H. deserticola*) or 0.9 to 1.2 mya (*H. paradoxus*). These date estimates should be viewed with caution because normalized net nucleotide distance is a crude method of dating a divergence time. It is possible that these estimates are actually biased down slightly because I only used genes that had confidently assigned parentage. Genes that evolved quickly in the hybrid species may not be assigned to a parent because of mutations away from both parental haplotypes. Alternatively, the estimate may be biased upwards because it measures genetic distance to the entire parental species, and not the actual subpopulation involved in hybrid speciation.

We would expect that the divergence estimates would be the same for genes from both parents of a single hybrid species. Indeed, the divergence estimates largely overlap, although are not identical (Figure 4-9). *Helianthus annuus* genes appear to be slightly younger in *H. deserticola* and *H. paradoxus*, while the reverse is true in *H. anomalus*. Considering the large amount of variation in genetic distance, this should be interpreted with caution, but it raises the possibility of subsequent gene flow with the parental species, which could cause such a pattern. Future work should use an explicitly reticulate phylogenetic approach or estimate a phylogeny on each gene independently.

4.4.4 The hybrid genomes are not fully stabilized

Despite the fact that sequence divergence suggests ancient hybrid origins, observed interspecific heterozygosity remains. In simulations based on early genetic maps of the sunflower hybrid species interspecific heterozygosity declined rapidly to zero and genome stabilization occurred in hundreds to thousands of generations (Buerkle and Rieseberg 2008). Using transcriptome data, I find that the hybrid species are much older than previously appreciated but, despite this, may not be completely stabilized yet.

My measure of interspecific heterozygosity is based on fixed differences and is likely to overestimate interspecific heterozygosity due to limited parental sampling. Some of the sites, which were called as fixed differences, are not actually fixed differences in the larger gene pool. Thus some cases of *interspecific* heterozygosity could be *intraspecific* heterozygosity from a single parent. I expect this effect to be largely equivalent in each hybrid species, since they rely on the same parental sampling, so this does not explain the significantly higher observed interspecific heterozygosity in *H. deserticola* and *H. anomalus*.

Gene flow between the hybrid species or with their parents could also inflate observed interspecific heterozygosity. Indeed, *H. deserticola* and *H. anomalus* are known to hybridize and each of them has noticeably higher heterozygosity.
It is also possible that the high observed interspecific heterozygosity reflects a genome that is not completely stabilized. It's important to note that genome stabilization

occurs in two stages. First selection on underdominant or epistatic loci rapidly reduces interspecific heterozygosity immediately after hybridization. After these regions are fixed in the nascent hybrid species and the genome has achieved an adaptive form, the remaining interspecific heterozygosity is removed slower and is depend on effective population size (Equation 4-2).

$$t = \frac{log \frac{H_t}{H_0}}{\log (1 - \frac{1}{2N})} \quad \text{4-2}$$

In this equation, t = number generations, H is heterozygosity at time zero (i.e., 1 for the hybrid species) and at time t, and N is population size. *Helianthus paradoxus* has an estimated effective population size of circa 120,000 and, in my study, observed heterozygosity of 0.0079 (Strasburg et al. 2011). If we ignore the initial selective phase of stabilization, with these parameters t = ~1.1 million years and is in the upper range of my estimate age values. Early strong selection against heterozygosity would effectively reduce the H_0 value and consequently reduce the time required to stabilization, although exactly how much heterozygosity was lost through selection compared to drift is unknown.

Helianthus anomalus and *H. deserticola* haven't had their effective population sizes formally tested but they do have greater genetic diversity, and indeed we see they have greater interspecific heterozygosity (Schwarzbach and Rieseberg 2002; Welch and Rieseberg 2002b; Gross et al. 2003). Although each of the hybrid species is

relatively rare, they have multiple populations in different states (Heiser et al. 1969). If genome stabilization was incomplete before the original range expansion, then population structure and ongoing gene flow could maintain the observed interspecific heterozygosity. Thus it is theoretically plausible that the observed interspecific heterozygosity is a result of high effective population sizes for the hybrid species, in concert with the features discussed above.

If the genome were not fully stabilized we would also expect different parental variants to be sorting within the species. I find little evidence for this; only 0.1 to 0.4% of genes in a single species have samples called for each parent (i.e., in gene A, *H. anomalus* sample 1 is from *H. annuus* and *H. anomalus* sample 2 is from *H. petiolaris*). This is a much lower estimate than interspecific heterozygosity at the SNP level, but may reflect methodological limitations. In particular, individuals heterozygous for different parental alleles would necessarily be classified as "admixed", and not be counted. Also, if parental alleles were sorting within the species since hybrid speciation, it is likely that recombination will break up parental haplotypes and create alleles that would be classified as "admixed".

4.4.5 The hybrid species do not have evidence for multiple origins

During hybrid speciation, the genome is expected to stabilize into a single form as selection and drift removes interspecific heterozygosity. This means that stabilized genomes should have similar parental fragments across their genome. If parental fragments are not similar, this suggests that either the genome is not completely

stabilized yet, or that the species had multiple origins. Although the genomes of each hybrid species are not completely stabilized with regard to interspecific heterozygosity, based on genetic distance they are old and so I expect high similarity in genomic composition among samples of a single hybrid species if there is a single origin of each species.

Previous work has suggested that both *H. anomalus* and *H. deserticola* might have multiple origins (Schwarzbach and Rieseberg 2002; Gross et al. 2003). This is based off of cpDNA haplotypes, microsatellites and interfertility experiments. I find that the correlation coefficients for parental admixture are consistently high in both of these species supporting a single origin, although they are slightly lower than the values in *H. paradoxus*. For *H. anomalus* this conclusion is much weaker because I am only using two *H. anomalus* samples, it is possible that the two samples, although from geographically separated locations, are from only one of multiple origins. One putatively *H. anomalus* sample was removed early due to evidence that it was a hybrid with *H. petiolaris*, but it is possible that it represents a true lineage of *H. anomalus* with a unique composition.

There is also surprisingly high consistency of ancestry among different hybrid species. The highest is between *H. anomalus* and *H. deserticola*, which will be discussed in the next section, but the correlation coefficients are not zero between *H. paradoxus* and either other hybrid species. This correlation is far higher than the simulations that attempt to control for the genomic features of the parents, including

variable divergence. This is consistent with the action of fertility selection during hybrid speciation (Rieseberg et al. 1996). Under this scenario selection may favor particular combinations of parental alleles during the critical early hybrid generations when fertility must be restored. Indeed, artificial hybrids produced had more similar genomic compositions to the actual hybrid species than expected by chance (Rieseberg et al. 1996). They also had increased interfertility with the natural hybrid species, suggesting selection tended to favor the same combination of hybrid incompatibilities (Rieseberg 2000).

This speculation must be tempered by the fact that the simulations may be under estimating the neutral degree of correlation for several reasons. One is that the simulations were run under small effective population sizes, which produce large parental fragments. If a larger effective population size is used, parental fragments will decrease in size. Eventually if parental fragments become much smaller than the genomic window size, then all windows will look admixed and the correlation coefficients will increase. In actual hybrid genomes, both large and small parental fragments are seen so neither scenario is accurate.

4.4.6 *Helianthus anomalus* and *H. deserticola* may share a single origin

When Heiser first described *H. deserticola*, he posited that that *H. deserticola* and *H. anomalus* were close relatives (Heiser 1960). Although later molecular analysis attributed that morphological similarity to their parallel origin, using genomic data we must once again face the possibility that Heiser was right all along.

The first evidence for a shared origin was the finding that *H. anomalus* and *H. deserticola* have remarkable similarity in parental composition across their genome. Similarity in genomic composition may be driven by features of the parental genomes, for example if *H. annuus* and *H. petiolaris* are not differentiated in a genomic window, then hybrid species composition may be biased towards admixed values due to the relatively flat likelihood curve. To account for this, I simulated hybrid species with independent origins and large block sizes, and measured their correlation coefficients. These simulations showed very little correlation, marginally above zero, suggesting that these parental genome factors played little role in high correlation coefficients seen.

High correlation coefficients may also reflect selection during the hybrid speciation process. The genomic location of hybrid incompatibilities between the parental species were likely similar for all hybrid speciation events. We expect there to be strong fertility or ecological selection in early generations (Rieseberg et al. 1996) and these may force hybrid species into similar compositions as was seen in artificial hybrids (Rieseberg 2000).

To test whether the similarity is driven by shared origin, or parallel selection, I examined non-parental alleles (i.e., alleles not found in either parent). These alleles are either new mutations that have accumulated since the hybrid speciation events or are low frequency parental variants not picked up in our parental dataset. If *H. anomalus* and *H. deserticola* share a single origin, then I would expect them to share more of

these non-parental alleles due to their shared evolutionary history. Alternatively, if *H. anomalus* and *H. deserticola* are similar due to shared selective pressure, I do not expect them to share an elevated amount of non-parental alleles because selection is unlikely to select the same rare variants during hybrid speciation. As a reference, I also measured the number of non-parental alleles shared with *H. paradoxus*, which is not expected to have a shared origin.

I found that *H. anomalus* and *H. deserticola* share roughly ten times more non-parental alleles than either does with *H. paradoxus* (Figure 4-12). This strongly supports the shared origin hypothesis although it is also concordant with gene flow between *H. anomalus* and *H. deserticola*. Although these species are known to hybridize, they are currently strongly reproductively isolated, including several chromosomal rearrangements (although interesting *H. anomalus* and *H. deserticola* have the least reproductive isolation of the three hybrid species) (Lai et al. 2005). This suggests that gene flow in the recent past is unlikely (but see (Yatabe et al. 2007)). Thus our current results support a shared origin but cannot rule out independent origins followed by gene flow.

Chapter 5: Conclusion

Over the course of three data chapters I have explored hybridization from three angles. Together these chapters tell us something more broadly about hybridization's role in evolution and how to study it. First, reproductive isolation does not accrue uniformly in all taxa and therefore the prevalence and importance of hybridization is also going to vary (Price and Bouvier 2002; Moyle et al. 2004; Bolnick and Near 2005). If we had a better idea of the traits that affect this rate, we might better be able to predict which species are prone to hybridization. These species have access to a larger pool of potentially adaptive genetic variants but may also be more susceptible to hybridization-mediated extinction (Rhymer and Simberloff 1996; Todesco et al. 2016).

Second, detecting introgression between species is challenging and multiple methods should be used. STRUCTURE-type programs are often used as the main evidence for or against introgression (e.g. (Sato et al. 2010; Mucci et al. 2012; Zhang et al. 2014)), but in chapter 3 I found that FastStructure is unable to detect gene flow found using more explicit approaches. Variations on the ABBA-BABA test are in current development and allow for tests of a variety of gene flow scenarios, although their relative nature (i.e., they detect if gene flow is greater in one species, not if it exists at all) means that results should be interpreted with caution (Eaton and Ree 2013; Pease and Hahn 2015).

Lastly, the genome of a hybrid species does not stabilize evenly. My estimates of parental block sizes in the *Helianthus* hybrid species found that although there are regions of extended parental blocks, most of the genome is highly recombined. Furthermore, I find evidence that the genome is not actually entirely stabilized, despite the comparatively ancient origin. This suggests that during genome stabilization loci under selection, either fertility or ecological, may not be dense enough to fully stabilize the genome. Consequently, hybrid species may have more diversity than previously appreciated, since they can retain alleles from both parents indefinitely in some genomic regions. This may explain why *H. paradoxus* has a larger population size than expected for a species that has undergone a hybridization bottleneck (Strasburg et al. 2011).

5.1 Strengths and weaknesses

Studying reproductive isolation among many taxa is challenging because each measure is time consuming and potentially challenging. In chapter 2, I overcome this challenge by using a rich literature of experimental measures. This lets me examine the question in two separate lineages, *Helianthus* and Madiinae.

There are several limitations to this study. For one, I only looked at hybrid pollen sterility. This ignores both prezygotic isolation factors (e.g. habitat selection, ovule abortion) and other postzygotic isolation factors (e.g. ovule sterility, hybrid inviability). These other factors almost certainly play a role in maintaining species boundaries and may covary with pollen sterility. Second, in the annual species measured, pollen

sterility was already quite high for the most closely related pairs. This suggests that pollen sterility may begin to evolve in annuals before populations have diverged enough to be classified as different species. In this way, I missed the critical period for understanding the rate of reproductive isolation gain. Lastly, measures of genetic distance were based on ITS sequence, which has limited resolution and the sequences were not from the same populations tested experimentally. Thus genetic distance measures may be inaccurate, especially for close relatives.

To rule out the hybrid origin of *H. bolanderi*, my study improved on previous work by not only using better technology, but also much more thorough sampling of both the focal species *H. bolanderi-exilis* and its sister *H. annuus* (Rieseberg et al. 1988). This allowed me to make conclusions about the directionality of gene flow that I otherwise would not have been able. One shortcoming to my study is the unsatisfying picture of where introgression occcurs, both geographically and genomically. I have good evidence that California *H. annuus* as a whole harbors introgressions, but there seems to be no geographic pattern to where introgressed individuals live. Similarly, I failed to determine where in the genome introgression is occurring because of data limitations. Solutions to both of these questions require more detailed sampling in California *H. annuus* and higher resolution genomic data.

Homoploid hybrid species have not been examined with modern genomic techniques (although see Heliconius Genome Consortium 2012), so my study is a forerunner in this field. Consequently, this chapter is the first to explicitly assign

parentage to a substantial portion of genes in the genome of a homoploid hybrid species. Another strength of my analysis is the replication in species; I examine what were previously believed to be three separate examples of homoploid hybrid speciation. This allows me to look for common patterns that may be generalizable across other hybrid taxa. It will be interesting to see if similar patterns emerge in the much younger hybrid species *Senecio squalidus* (James and Abbott 2005). On the other hand, my study is limited by sample size, particularly for *H. anomalus*. Although genomic patterns were consistent within a species, the inconsistent sampling design limited my ability to ask questions about within species variation in a geographic context. Using transcriptome data gave me access to far more SNPs that any previous analyses, but ultimately it represents only a small fraction of the total genome. Lastly, the timing of divergence is challenging considering the hybrid genome and more detailed phylogenetic analyses are required to date the hybrid speciation events more precisely.

5.2 Future directions

In chapter 2 I showed that the rate of reproductive isolation acquisition varied by life history. This was shown in two clades within the Compositae family and has previously been shown within the genus Coreopsis, also in Compositae (Archibald et al. 2005). The relatively close relation between all groups tested raises the question of whether this is a family specific or more general pattern, which can be answered by further studies. The patterns presented also raise the question of whether alternate

mechanisms of reproductive isolation are more commonly used in perennials (e.g. prezygotic mechanisms like habitat or gametic isolation) or if perennial species tolerate higher levels of hybridization. Anecdotal evidence suggests perennial sunflower species have stricter gametic isolation than annual species, but this remains to be rigorously tested (Heiser et al. 1969).

In chapter 2 I showed that there was gene flow between *H. annuus* and *H. bolanderi-exilis*, but the question remains if the introgression is adaptive, as is seen in Texas (Whitney et al. 2010). Using whole genome shotgun data, genomic regions in the *H. annuus* genome harboring introgression can be identified and tested for signs of recent selection. For *H. bolanderi-exilis*, crossing studies can determine whether the two genetic clades are reproductively isolated enough to be named separate species. This has potential to be important for both agriculture and conservation. For example, *Helianthus exilis* was considered an endangered species in the past but my work suggests that the eastern clade of *H. bolanderi-exilis*, which has a more limited range, may be a more important conservation concern. Similarly, breeding effort to incorporate *H. bolanderi-exilis* genes into domestic *H. annuus* should focus on using collections from both genetic clades to incorporate the most diversity.

The *Helianthus* homoploid hybrid species still remain an enigma, despite the genomic analysis in chapter 4. For one, further work needs to be done to determine the age of hybridization. Genetic distance suggests the hybrid speciation events may

be quite old, but this could be confounded by gene flow among different branches in the clade.

To answer them, a comprehensive approach will need to be used that leverages multiple genome sequences from each hybrid species, each parent species and their close relatives and incorporates gene flow at multiple locations in the tree. This has been done in several systems although not with homoploid hybrid species (Marcussen et al. 2014; Liu et al. 2015; Wen et al. 2016).

Another feature of the homoploid hybrid species that needs exploring is the proliferation of transposable elements. We know that the genome size has enlarged in each hybrid species and that this is due to transposable elements, but artificial hybridization experiments have failed to find support for immediate hybridization induced proliferation (Kawakami et al. 2011). Long read genomic sequences could accurately reconstruct TE sequence and determine their age based on sequence divergence. From that, we could determine if the proliferation occurred immediately after hybridization, suggesting it was caused by the hybridization itself, or if it was a more gradual process.

Lastly, I find regions in the genome with unusually long parental block sizes. This may be the work of selection during hybrid speciation or could be the rare outcome of neutral processes. Future analyses could model hybrid speciation with and without loci under selection to determine if neutral processes alone could produce the patterns seen.

5.3 Conclusion

The role of hybridization in evolution is better appreciated now in the era of genomic data. Genome wide information has revealed a hidden side of hybridization in two of my chapters. It showed that introgression in Californian sunflowers occurred in the opposite direction from previous hypotheses, as well as an unexpected picture of homoploid hybrid species. I expect genomic data will continue to expose hidden hybridization in the evolutionary past, such that we will have to routinely consider the spread of adaptive alleles not just within a single species, but across an entire genus.

Bibliography

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

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664.

Allendorf, F. W., R. F. Leary, P. Spruell, and J. K. Wenburg. 2001. The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16:613–622.

Anderson, E. 1948. Hybridization of the Habitat. Evolution 2:1–9.

Anderson, E. 1949. Introgressive hybridization. Biol Rev 28:280–307.

Andreasen, K., and B. G. Baldwin. 2001. Unequal evolutionary rates between annual and perennial lineages of checker mallows (Sidalcea, Malvaceae): evidence from 18S-26S rDNA internal and external transcribed spacers. Mol Biol Evol 18:936–944.

Arbogast, B. S., S. V. Edwards, J. Wakeley, and P. Beerli. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu Rev Ecol Syst 33:707–740.

Archibald, J. K., M. E. Mort, D. J. Crawford, and J. K. Kelly. 2005. Life history affects the evolution of reproductive isolation among species of *Coreopsis* (Asteraceae). Evolution 59:2362–2369.

Arnold, M. L. 1996. Natural hybridization and evolution. Oxford University Press, USA.

Baack, E. J., K. D. Whitney, and L. H. Rieseberg. 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. New Phytol. 167:623–630.

Baldwin, B. G. 2003. A phylogenetic perspective on the origin and evolution of Madiinae. *in* S. Carlquist, B. G. Baldwin, and G. D. Carr, eds. Tarweeds and Silverswords: Evolution of the Madiinae (Asteraceae).

Baldwin, B. G. 2007. Adaptive radiation of shrubby tarweeds (*Deinandra*) in the

California Islands parallels diversification of the Hawaiian silversword alliance (Compositae-Madiinae). Am J Bot 94:237–248.

Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). P Natl Acad Sci Usa 95:9402–9406.

Baldwin, B. G., and S. Markos. 1998. Phylogenetic utility of the External Transcribed Spacer (ETS) of 18S–26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol Phylogenet Evol 10:449–463.

Barb, J. G., J. E. Bowers, S. Renaut, J. I. Rey, S. J. Knapp, L. H. Rieseberg, and J. M. Burke. 2014. Chromosomal evolution and patterns of introgression in *Helianthus*. Genetics 197:969–979.

Barker, M. S., N. Arrigo, A. E. Baniaga, Z. Li, and D. A. Levin. 2015. On the relative abundance of autopolyploids and allopolyploids. New Phytol., doi: 10.1111/nph.13698.

Barton, N. H. 2013. Does hybridization influence speciation? J Evolution Biol 26:267–269.

Barton, N. H. 1979. Gene flow past a cline. Heredity 43:333–339.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu Rev Ecol Syst 16:113–148.

Bateson, W. 1909. Heredity and variation in modern lights. Pp. 85–101 *in* A. C. Seward, ed. Darwin and Modern Science.

Baute, G. J. 2015. Genomics of sunflower improvement from wild relatives to a global oil seed.

Baute, G. J., N. C. Kane, C. J. Grassa, Z. Lai, and L. H. Rieseberg. 2015. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. New Phytol. 206:830–838.

Blair, W. F. 1955. Mating call and stage of speciation in the *Microhyla olivacea-M. carolinensis* complex. Evolution 9:469–480.

Bock, D. G., N. C. Kane, D. P. Ebert, and L. H. Rieseberg. 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytol 201:1021–1030.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Bomblies, K., and D. Weigel. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat Rev Genet 8:382–393.

Brady, K. U., A. R. Kruckeberg, and H. Bradshaw Jr. 2005. Evolutionary ecology of plant adaptation to serpentine soils. Annu Rev Ecol Evol Syst 243–266.

Brennan, A. C., G. Woodward, O. Seehausen, V. Muñoz-Fuentes, C. Moritz, A. Guelmami, R. J. Abbott, and P. Edelaar. 2015. Hybridization due to changing species distributions: adding problems or solutions to conservation of biodiversity during global change? Evol Ecol Res 16:475–491.

Brooks, R. R. 1987. Serpentine and its vegetation: a multidisciplinary approach. Dioscorides Press.

Buerkle, C. A., and C. Lexer. 2008. Admixture as the basis for genetic mapping. Trends Ecol Evol 23:686–694.

Buerkle, C. A., and L. H. Rieseberg. 2008. The rate of genome stabilization in homoploid hybrid species. Evolution 62:266–275.

Buerkle, C. A., R. J. Morris, M. A. Asmussen, and L. H. Rieseberg. 2000. The likelihood of homoploid hybrid speciation. Heredity 84 (Pt 4):441–451.

Burke, J. M., Z. Lai, M. Salmaso, T. Nakazato, S. Tang, A. Heesacker, S. J. Knapp, and L. H. Rieseberg. 2004. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. Genetics 167:449–457.

Butlin, R. 1987. Speciation by reinforcement. Trends Ecol Evol 2:8–13.

Carney, S., K. Gardner, and L. Rieseberg. 2000. Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (*Helianthus*). Evolution 54:462–474.

Carr, G. D., and D. W. Kyhos. 1981. Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae). I. Cytogenetics of spontaneous hybrids. Evolution 35:543–556.

Carr, G. D., and D. W. Kyhos. 1986. Adaptive radiation in the Hawaiian silversword

alliance (Compositae-Madiinae). II. Cytogenetics of artificial and natural hybrids. Evolution 959–976.

Chandler, J. M., C. C. Jan, and B. H. Beard. 1986. Chromosomal differentiation among the annual *Helianthus* species. Syst Bot 11:354–371.

Chapman, N. H., and E. A. Thompson. 2002. The effect of population history on the lengths of ancestral chromosome segments. Genetics 162:449–458.

Chen, Z. J. 2013. Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet 14:471–482.

Chunco, A. J. 2014. Hybridization in a warmer world. Ecol Evol 4:2019–2031.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc.

Coyne, J. A., and H. A. Orr. 1997. " Patterns of speciation in *Drosophila*" revisited. Evolution 51:295–303.

Currat, M., M. Ruedi, R. J. Petit, and L. Excoffier. 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–1920.

Darwin, C. 1859. On the origin of species.

Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. Am Nat 74:312–321.

Dobzhansky, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21:113–135.

Donnelly, M. J., J. Pinto, R. Girod, N. J. Besansky, and T. Lehmann. 2004. Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. Heredity 92:61–68.

Duenez-Guzman, E. A., J. Mavárez, M. D. Vose, and S. Gavrilets. 2009. Case studies and mathematical models of ecological speciation. 4. Hybrid speciation in butterflies in a jungle. Evolution 63:2611–2626.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239–2252.

Eaton, D. A. R., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst

Biol 62:689–706.

Edmands, S. 2002. Does parental divergence predict reproductive compatibility? Trends Ecol Evol 17:520–527.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Fisher, R. A. 1954. A fuller theory of "junctions" in inbreeding. Heredity 8:187–197.

Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, E. Kakani, S. N. Mitchell, Y.-C. Wu, H. A. Smith, R. R. Love, M. K. Lawniczak, M. A. Slotman, S. J. Emrich, M. W. Hahn, and N. J. Besansky. 2015. Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347:1258524.

Fox, J., and S. Weisberg. 2010. An R companion to applied regression.

Frankham, R., J. D. Ballou, M. D. B. Eldridge, R. C. Lacy, K. Ralls, M. R. Dudash, and C. B. Fenster. 2011. Predicting the probability of outbreeding depression. Conserv. Biol. 25:465–475.

Gaut, B., L. Yang, S. Takuno, and L. E. Eguiarte. 2011. The patterns and causes of variation in plant nucleotide substitution rates. Annu Rev Ecol Evol Syst 42:245–266.

Giménez, M. D., T. A. White, H. C. Hauffe, T. Panithanarak, and J. B. Searle. 2013. Understanding the basis of diminished gene flow between hybridizing chromosome races of the house mouse. Evolution 67:1446–1462.

Good, T. P., J. C. Ellis, C. A. Annett, and R. Pierotti. 2000. Bounded hybrid superiority in an avian hybrid zone: effects of mate, diet, and habitat choice. Evolution 54:1774–1783.

Goodman, S. J., N. H. Barton, G. Swanson, K. Abernethy, and J. M. Pemberton. 1999. Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus *Cervus*) in Argyll, Scotland. Genetics 152:355–371.

Grant, P. R., and B. R. Grant. 2011. How and why species multiply: the radiation of Darwin's finches. Princeton University Press.

Grant, V. 1981. Plant speciation. Columbia University Press.

Grant, V. 1958. The regulation of recombination in plants. Cold Spring Harb Symp

Quant Biol 23:337–363.

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

Grey, A. 1865. *Helianthus bolanderi*. Proc Amer Acad Arts 6:544–545.

Gross, B. L., A. E. Schwarzbach, and L. H. Rieseberg. 2003. Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). Am J Bot 90:1708–1719.

Gross, B. L., and L. H. Rieseberg. 2005. The ecological genetics of homoploid hybrid speciation. J. Hered. 96:241–252.

Gross, B. L., N. C. Kane, C. Lexer, F. Ludwig, D. M. Rosenthal, L. A. Donovan, and L. H. Rieseberg. 2004. Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. Am Nat 164:145–156.

Gulya, T. J., and G. J. Seiler. 2002. Plant Exploration Report.

Hallatschek, O., and D. R. Nelson. 2008. Gene surfing in expanding populations. Theor Popul Biol 73:158–170.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Oxford Surv Evol Biol 7:69–128.

Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol 22:4606–4618.

Heiser, C. B. 1947. Hybridization between the sunflower species *Helianthus annuus* and *H. petiolaris*. Evolution 1:249–262.

Heiser, C. B. 1951a. Hybridization in the annual sunflowers: *Helianthus annuus* X *H. argophyllus*. Am Nat 85:65–72.

Heiser, C. B. 1951b. Hybridization in the annual sunflowers: *Helianthus annuus* X *H.*

debilis var. cucumerifolius. Evolution 5:42–51.

Heiser, C. B. 1965. Species Crosses in *Helianthus*: III. Delimitation of "Sections." Annals of the Missouri Botanical Garden 52:364–370.

Heiser, C. B. 1949. Study in the evolution of the sunflower species *Helianthus annuus* and *H. bolanderi*. Univ. Calif. Publ. Bot.

Heiser, C. B., and D. Smith. 1964. Species crosses in *Helianthus*: II. Polyploid species. Rhodora 344–358.

Heiser, C. B., D. M. Smith, S. B. Clevenger, and W. C. Martin. 1969. The North American sunflowers (*Helianthus*). Mem Torrey Bot Club 22:1–218.

Heiser, C. B., Jr. 1960. A new annual sunflower, *Helianthus deserticolus*, from southwestern United States. Proc Ind Acad Sci 70:209–212.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

Hopkins, R., and M. D. Rausher. 2012. Pollinator-mediated selection on flower color allele drives reinforcement. Science 335:1090–1092.

Hoskin, C. J., M. Higgie, K. R. McDonald, and C. Moritz. 2005. Reinforcement drives rapid allopatric speciation. Nature 437:1353–1356.

Husband, B. C. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. P R Soc B 267:217–223.

Huson, D. H. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.

Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. J Comput Graph Stat 5:299–314.

Jackson, R. C., and A. T. Guard. 1956. Analysis of some natural and artificial interspecific hybrids in *Helianthus*. Proc Ind Acad Sci 66.

Jain, S. K., R. Kesseli, and A. Olivieri. 1992. Biosystematic status of the serpentine sunflower, *Helianthus exilis* Gray. Pp. 391–408 *in* J. Proctor, A. J. M. Baker, and R. D. Reeves, eds. The vegetation of ultramafic (serpentine) soils.

James, J. K., and R. J. Abbott. 2005. Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on

Mount Etna, Sicily. Evolution 59:2533–2547.

Jan, C. C. 1997. Cytology and interspecific hybridization. Sunflower Tech Prod 35:497–558.

Jan, C. C., and J. M. Chandler. 1989. Sunflower interspecific hybrids and amphiploids of *Helianthus annuus* × *H. bolanderi*. Crop Science 29:643.

Jónsson, H., M. Schubert, A. Seguin-Orlando, A. Ginolhac, L. Petersen, M. Fumagalli, A. Albrechtsen, B. Petersen, T. S. Korneliussen, J. T. Vilstrup, T. Lear, J. L. Myka, J. Lundquist, D. C. Miller, A. H. Alfarhan, S. A. Alquraishi, K. A. S. Al-Rasheid, J. Stagegaard, G. Strauss, M. F. Bertelsen, T. Sicheritz-Ponten, D. F. Antczak, E. Bailey, R. Nielsen, E. Willerslev, and L. Orlando. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. P Natl Acad Sci Usa 111:18655–18660.

Karrenberg, S., C. Lexer, and L. H. Rieseberg. 2007. Reconstructing the history of selection during homoploid hybrid speciation. Am Nat 169:725–737.

Kawakami, T., P. Dhakal, A. N. Katterhenry, C. A. Heatherington, and M. C. Ungerer. 2011. Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons. Genome Biol Evol 3:156–167.

Kay, K. M., J. B. Whittall, and S. A. Hodges. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol Biol 6:36.

Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz, A. Sivasundar, S. Wittwer, and O. Seehausen. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Mol Ecol 22:2848–2863.

Kräuter, R., A. Steinmetz, and W. Friedt. 1991. Efficient interspecific hybridization in the genus *Helianthus* via embryo rescue and characterization of the hybrids. Theor Appl Genet 82:521–525.

Kruckeberg, A. R. 1985. California serpentines: flora, vegetation, geology, soils, and management problems. University of California Press.

Kruskal, W. H., and W. A. Wallis. 1952. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621.

Kulathinal, R. J., L. S. Stevison, and M. A. F. Noor. 2009. The genomics of speciation in

Drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genet 5:e1000550.

Kulikova, I. V., Y. N. Zhuravlev, and K. G. McCracken. 2004. Asymmetric hybridization and sex-biased gene flow between eastern spot-billed ducks (*Anaz zonorhyncha*) and mallards (*A. platyrhynchos*) in the Russian far east. The Auk 121:930–949.

Lai, Z., N. C. Kane, A. Kozik, K. A. Hodgins, K. M. Dlugosch, M. S. Barker, M. Matvienko, Q. Yu, K. G. Turner, S. A. Pearl, G. D. M. Bell, Y. Zou, C. Grassa, A. Guggisberg, K. L. Adams, J. V. Anderson, D. P. Horvath, R. V. Kesseli, J. M. BURKE, R. W. Michelmore, and L. H. Rieseberg. 2012. Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression. Am J Bot 99:209–218.

Lai, Z., T. Nakazato, M. Salmaso, J. M. Burke, S. Tang, S. J. Knapp, and L. H. Rieseberg. 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. Genetics 171:291–303.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Leclercq, P. 1969. Une sterilite male cytoplasmique chez le tournesol. Ann Amelior Plant 19:99–106.

Levin, D. A. 1979. Hybridization: an evolutionary perspective. Dowden, Hutchinson & Ross, Inc.

Levin, D. A. 1975. Minority cytotype exclusion in local plant populations. Taxon 24:35–43.

Levin, D. A. 2002. The role of chromosomal change in plant evolution. Oxford University Press.

Levin, D. A., and J. F. Ortega. 1996. Hybridization and the extinction of rare plant species. Conserv Biol 10:10–16.

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.

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

Liao, Y., G. K. Smyth, and W. Shi. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 41:e108.

Lijtmaer, D., B. Mahler, and P. Tubaro. 2003. Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution 57:1411–1418.

Liu, K. J., E. Steinberg, A. Yozzo, Y. Song, M. H. Kohn, and L. Nakhleh. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. P Natl Acad Sci Usa 112:196–201.

Long, R. W. 1955. Hybridization in perennial sunflowers. Am J Bot 42:769–777.

Lunter, G., and M. Goodson. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21:936–939.

Machado, C. A., T. S. Haselkorn, and M. A. F. Noor. 2007. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics 175:1289–1306.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol Evol 20:229–237.

Malone, J. H., and B. E. Fontenot. 2008. Patterns of reproductive isolation in toads. PLoS ONE 3:e3900.

Marcussen, T., S. R. Sandve, L. Heier, M. Spannagl, M. Pfeifer, K. S. Jakobsen, B. B. H. Wulff, B. Steuernagel, K. F. X. Mayer, and O.-A. Olsen. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol Biol Evol 32:244–257.

Martinsen, G. D., T. G. Whitham, R. J. Turek, and P. Keim. 2001. Hybrid populations selectively filter gene introgression between species. Evolution 55:1325–1335.

Mavarez, J., and M. Linares. 2008. Homoploid hybrid speciation in animals. Mol Ecol 4181–4185.

Mayr, E. 1963. Animal species and evolution. Harvard University Press.

McCarthy, E. M., M. A. Asmussen, and W. W. Anderson. 1995. A theoretical assessment

of recombinational speciation. Heredity 74:502–509.

Melo-Ferreira, J., P. Boursot, F. Suchentrunk, N. Ferrand, and P. C. Alves. 2005. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. Mol Ecol 14:2459–2464.

Moore, W. S. 1977. An evaluation of narrow hybrid zones in vertebrates. Q Rev Biol 52:263–277.

Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. Evolution 58:1195–1208.

Mucci, N., F. Mattucci, and E. Randi. 2012. Conservation of threatened local gene pools: landscape genetics of the Italian roe deer (*Capreolus c. italicus*) populations. Evol Ecol Res 14:897–920.

Muller, H. J. 1942. Isolating mechanisms, evolution and temperature. Biol Symp 6:71–125.

Nosil, P., and B. J. Crespi. 2006. Ecological divergence promotes the evolution of cryptic reproductive isolation. P R Soc B 273:991–997.

Nosrati, H., A. H. Price, and C. C. Wilcock. 2011. Relationship between genetic distances and postzygotic reproductive isolation in diploid *Fragaria* (Rosaceae). Biol J Linn Soc 104:510–526.

Panero, J. L., and V. A. Funk. 2002. Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). P Biol Soc Wash 115:909–922.

Pardo-Diaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-Ready, M. Joron, W. O. McMillan, and C. D. Jiggins. 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. PLoS Genet 8:e1002752.

Pease, J. B., and M. W. Hahn. 2015. Detection and polarization of introgression in a five-taxon phylogeny. Syst Biol 64:651–662.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Presgraves, D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution 56:1168–1183.

Price, T. D., and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083–2089.

Quillet, M. C., N. Madjidian, Y. Griveau, and H. Serieys. 1995. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. Theor Appl Genet 91:1195–1202.

R Core Team. 2008. R: A Language and Environment for Statistical Computing.

Raj, A., M. Stephens, and J. K. PRITCHARD. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573–589.

Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, J. M. Burke, and L. H. Rieseberg. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nature Comm 4:1827.

Renaut, S., H. C. Rowe, M. C. Ungerer, and L. H. Rieseberg. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. Philos Trans R Soc Lond B Biol Sci 369.

Rhymer, J. M., and D. Simberloff. 1996. Exinction by hybridization and introgression. Annu Rev Ecol Syst 27:83–109.

Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. Trends Ecol Evol 16:351–358.

Rieseberg, L. H. 2000. Crossing relationships among ancient and experimental sunflower hybrid lineages. Evolution 54:859–865.

Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. Am J Bot 78:1218–1237.

Rieseberg, L. H. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216.

Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. Science 317:910–914.

Rieseberg, L. H., C. Van Fossen, and A. M. Desrochers. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375:313–316.

Rieseberg, L. H., D. E. Soltis, and J. D. Palmer. 1988. A molecular reexamination of introgression between *Helianthus annuus* and *H. bolanderi* (Compositae). Evolution 42:227–238.

Rieseberg, L. H., H. Choi, R. Chan, and C. Spore. 1993. Genomic map of a diploid hybrid species. Heredity 70:285–293.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity 83:363–372.

Rieseberg, L. H., R. Carter, and S. Zona. 1990a. Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). Evolution 44:1498–1511.

Rieseberg, L. H., S. Beckstrom-Sternberg, and K. Doan. 1990b. *Helianthus annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. P Natl Acad Sci Usa 87:593–597.

Rieseberg, L., B. Sinervo, C. Linder, M. Ungerer, and D. Arias. 1996. Role of gene interactions in hybrid apeciation: evidence from ancient and experimental hybrids. Science 272:741–745.

Rosenthal, D. M., A. E. Schwarzbach, L. A. Donovan, O. Raymond, and L. H. Rieseberg. 2002. Phenotypic differentiation between three ancient hybrid taxa and their parental species. Int. J Plant Sci. 163:387–398.

Rundle, H. D., and M. C. Whitlock. 2001. A genetic interpretation of ecologically dependent isolation. Evolution 55:198–201.

Russell, S. T. 2003. Evolution of intrinsic post-zygotic reproductive isolation in fish. Ann Zool Fennici 40:321–329.

Safford, H., J. Viers, and H. SP. 2005. Serpentine endemism in the California flora: a database of serpentine affinity. Madroño 54:222–257.

Saino, N., and S. Villa. 1992. Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. The Auk 109:543–555.

Sambatti, J. B. M., J. L. Strasburg, D. Ortiz-Barrientos, E. J. Baack, and L. H. Rieseberg. 2012. Reconciling extremely strong barriers with high levels of gene exchange in annual sunflowers. Evolution 66:1459–1473.

Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52:1811–1820.

Sato, T., T. Demise, H. Kubota, M. Nagoshi, and K. Watanabe. 2010. Hybridization, isolation, and low genetic diversity of kirikuchi char, the southernmost populations of the genus *Salvelinus*. T Am Fish Soc 139:1758–1774.

Schemske, D. W. 2000. Understanding the origin of species. Evolution 54:1069–1073.

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

Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014. How common is homoploid hybrid speciation? Evolution 68:1553–1560.

Schumer, M., R. Cui, G. G. Rosenthal, and P. Andolfatto. 2015. Reproductive isolation of hybrid populations driven by genetic incompatibilities. PLoS Genet 11:e1005041.

Schwarzbach, A. E., and L. H. Rieseberg. 2002. Likely multiple origins of a diploid hybrid sunflower species. Mol Ecol 11:1703–1715.

Schwarzbach, A. E., L. A. Donovan, and L. H. Rieseberg. 2001. Transgressive character expression in a hybrid sunflower species. Am J Bot 88:270–277.

Secondi, J., B. Faivre, and S. Bensch. 2006. Spreading introgression in the wake of a moving contact zone. Mol Ecol 15:2463–2475.

Servedio, M. R., J. Hermisson, and G. S. van Doorn. 2013. Hybridization may rarely promote speciation. J Evolution Biol 26:282–285.

Shaffer, L. G., and J. R. Lupski. 2000. Molecular mechanisms for constitutional chromosomal rearrangements in humans. Annu Rev Genet 34:297–329.

Soltis, D. E., P. S. Soltis, and J. A. Tate. 2004. Advances in the study of polyploidy since Plant speciation. New Phytol. 161:173–191.

Song, Y., S. Endepols, N. Klemann, D. Richter, F.-R. Matuschka, C.-H. Shih, M. W. Nachman, and M. H. Kohn. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr Biol 21:1296–1301.

Soria-Hernanz, D. F., J. M. Braverman, and M. B. Hamilton. 2008. Parallel rate heterogeneity in chloroplast and mitochondrial genomes of Brazil nut trees (Lecythidaceae) is consistent with lineage effects. Mol Biol Evol 25:1282–1296.

Stam, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet Res 35:131–155.

Staton, S. E., M. C. Ungerer, and R. C. Moore. 2009. The genomic organization of Ty3/gypsy-like retrotransposons in *Helianthus* (Asteraceae) homoploid hybrid species. Am J Bot 96:1646–1655.

Stebbins, G. L. 1958. The inviability, weakness, and sterility of interspecific hybrids. Adv

Genet 9:147–215.

Stebbins, G. L. 1959. The role of hybridization in evolution. P Am Philos Soc 103:231–251.

Strasburg, J. L., and L. H. Rieseberg. 2008. Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris* - large effective population sizes and rates of long-term gene flow. Evolution 62:1936–1950.

Strasburg, J. L., N. C. Kane, A. R. Raduski, A. Bonin, R. Michelmore, and L. H. Rieseberg. 2011. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. Mol Biol Evol 28:1569–1580.

Takayama, K., T. Kajita, J. Murata, and Y. Tateishi. 2006. Phylogeography and genetic structure of Hibiscus tiliaceus--speciation of a pantropical plant with sea-drifted seeds. Mol Ecol 15:2871–2881.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739.

Templeton, A. R. 1981. Mechanisms of speciation--a population genetic approach. Annu Rev Ecol Syst 12:23–48.

Timme, R. E., B. B. Simpson, and C. R. Linder. 2007. High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. Am J Bot 94:1837–1852.

Todesco, M., M. A. Pascual, G. L. Owens, K. L. Ostevik, B. T. Moyers, S. Hubner, S. M. Heredia, M. A. Hahn, C. Caseys, D. G. Bock, and L. H. Rieseberg. 2016. Hybridization and Extinction. Evol Appl.

Turelli, M., and L. C. Moyle. 2007. Asymmetric postmating isolation: Darwin"s corollary to Haldane"s rule. Genetics 176:1059–1088.

Turesson, G. 2010. Zur Natur und Begrenzung der Arteinheiten. Hereditas 12:323–334.

Ungerer, M. C., S. J. Baird, J. Pan, and L. H. Rieseberg. 1998. Rapid hybrid speciation in wild sunflowers. P Natl Acad Sci Usa 95:11757–11762.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2002. From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. Curr Protoc

Bioinform 43:11.10.1–11.10.33.

Vilà, M., E. Weber, and C. M. D. Antonio. 2000. Conservation implications of invasion by plant hybridization. Biol Invasions 2:207–217.

Wahlund, S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 10:65–106.

Walsh, J. B. 1982. Rate of accumulation of reproductive isolation by chromosome rearrangements. Am Nat 120:510–532.

Wang, H., E. D. McArthur, S. C. Sanderson, J. H. Graham, and D. C. Freeman. 1997. Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. Evolution 51:95–102.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Welch, M. E., and L. H. Rieseberg. 2002a. Habitat divergence between a homoploid hybrid sunflower species, *Helianthus paradoxus* (Asteraceae), and its progenitors. Am J Bot 89:472–478.

Welch, M. E., and L. H. Rieseberg. 2002b. Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. Evolution 56:2126–2137.

Wen, D., Y. Yu, M. W. Hahn, and L. Nakhleh. 2016. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. Mol Ecol, doi: 10.1111/mec.13544.

White, M. 1978. Modes of speciation. W. H. Freeman and Company, San Francisco.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytol. 187:230–239.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer New York.

Witter, M. S., and G. D. Carr. 1988. Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). Evolution 42:1278–1287.

Wolf, D. E., N. Takebayashi, and L. H. Rieseberg. 2001. Predicting the risk of extinction through hybridization. Conserv Biol 15:1039–1053.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H.

Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. P Natl Acad Sci Usa 106:13875–13879.

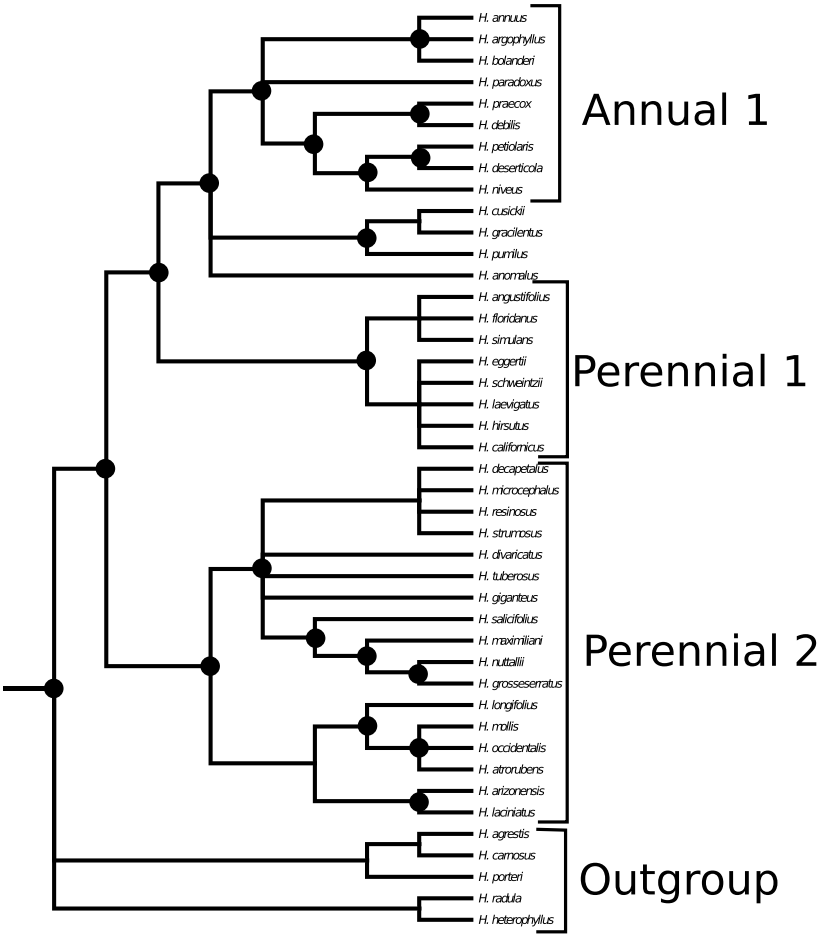
Yatabe, Y., N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 175:1883–1893.

Zhang, D., T. Xia, M. Yan, X. Dai, J. Xu, S. Li, and T. Yin. 2014. Genetic introgression and species boundary of two geographically overlapping pine species revealed by molecular markers. PLoS ONE 9:e101106–e101106.

Zirkle, C. 1935. The beginnings of plant hybridization. University of Pennsylvania Press.

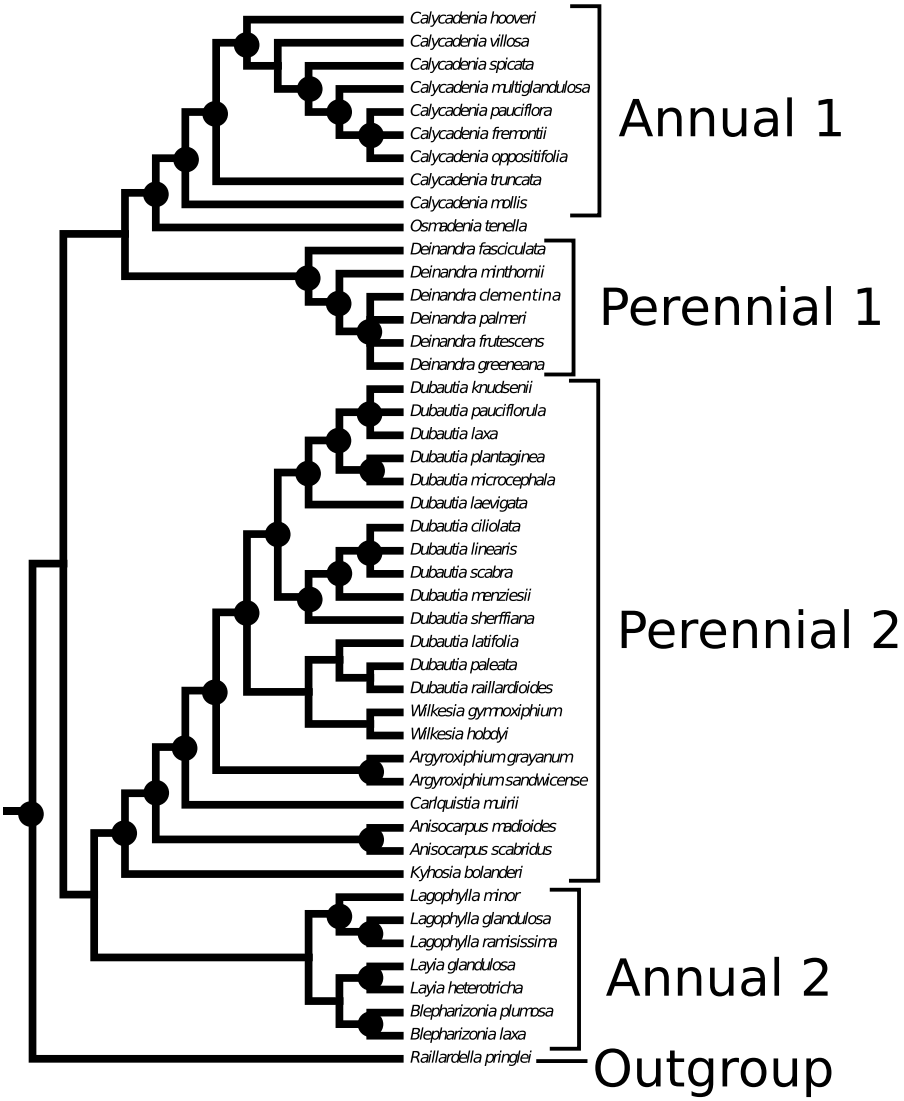
Appendices

Appendix A Supplementary information for Chapter 2



A.1 Phylogeny of Helianthus used for creating the phylogenetically corrected dataset.

Circles are nodes where multiple crosses were averaged together to construct a single independent data point. Phylogenetic relationships were inferred from Timme et al. (2007). Labeled clades were used to test differences in evolutionary rate (see text).



A.2 Phylogeny of Madiinae used for creating the phylogenetically corrected dataset.

Circles are nodes where multiple crosses were averaged together to construct a single independent data point. See text for references used to construct phylogeny. Labeled clades were used to test differences in evolutionary rate (see text).

A.3 Accession numbers for molecular sequence used in chapter 2.

Name	Life History	Accession Number
Anisocarpus madioides	P	AF061914.1
Anisocarpus scabridus	P	M93799.1
Argyroxiphium grayanum	P	AF061886.1
Argyroxiphium sandwicense	P	EU341969.1
Blepharizonia laxa	A	AF283548.1
Blepharizonia plumosa	A	AF229323.1
Calycadenia fremontii	A	U04249.1
Calycadenia hooveri	A	U04251.1
Calycadenia mollis	A	U04253.1
Calycadenia multiglandulosa	A	U04254.1
Calycadenia oppositifolia	A	U04257.1
Calycadenia pauciflora	A	U04259.1
Calycadenia spicata	A	U04260.1
Calycadenia truncata	A	U04262.1
Calycadenia villosa	A	U04264.1
Carlquistia muirii	P	M93798.1
Deinandra clementina	P	EF059624.1
Deinandra fasciculata	A	EF059605.1
Deinandra frutescens	P	EF059667.1
Deinandra greeneana subsp. greeneana	P	EF059649.1
Deinandra greeneana subsp. peninsularis	P	EF059683.1
Deinandra minthornii	P	EF059613.1
Deinandra palmeri	P	EF059659.1
Dubautia ciliolata	P	EU341946.1
Dubautia knudsenii	P	AF061903.1
Dubautia laevigata	P	AF061899.1
Dubautia latifolia	P	AF061900.1
Dubautia laxa	P	EU341947.1
Dubautia linearis	P	AF061910.1
Dubautia menziesii	P	M93791.1
Dubautia microcephala	P	AF061902.1
Dubautia paleata	P	AF061888.1
Dubautia pauciflorula	P	AF061896.1
Dubautia plantaginea	P	AF061891.1
Dubautia raillardioides	P	AF061897.1
Dubautia scabra	P	AF061906.1
Dubautia sherffiana	P	AF061907.1
Kyhosia bolanderi	P	M93794.1
Lagophylla glandulosa	A	DQ188074.1
Lagophylla minor	A	AF229311.1
Lagophylla ramosissima	A	AF229310.1
Layia glandulosa	A	DQ188043.1
Layia heterotricha	A	DQ188075.1

Name	Life History	Accession Number
Osmadenia tenella	A	U04266.1
Raillardella pringlei	P	M93797.1
Wilkesia gymnoxiphium	P	M93800.1
Wilkesia hobdyi	P	AF061882.1
Helianthus agrestis	A	DQ486530.1
Helianthus angustifolius	P	DQ486532.1
Helianthus annuus	A	DQ486533.1
Helianthus anomalus	A	DQ486535.1
Helianthus argophyllus	A	DQ486537.1
Helianthus arizonensis	P	DQ486540.1
Helianthus atrorubens	P	DQ486544.1
Helianthus bolanderi	A	DQ486545.1
Helianthus californicus	P	DQ486546.1
Helianthus carnosus	P	DQ486548.1
Helianthus cusickii	P	DQ486551.1
Helianthus debilis	A	DQ486554.1
Helianthus decapetalus	P	DQ486557.1
Helianthus deserticola	A	DQ486561.1
Helianthus divaricatus	P	DQ486563.1
Helianthus floridanus	P	DQ486571.1
Helianthus giganteus	P	DQ486572.1
Helianthus gracilentus	P	DQ486578.1
Helianthus grosserratus	P	DQ486581.1
Helianthus heterophyllus	P	DQ486582.1
Helianthus hirsutus	P	DQ486584.1
Helianthus laciniatus	P	DQ486585.1
Helianthus laevigatus	P	DQ486587.1
Helianthus longifolius	P	DQ486589.1
Helianthus maximiliani	P	DQ486590.1
Helianthus microcephalus	P	DQ486592.1
Helianthus mollis	P	DQ486595.1
Helianthus niveus	P	DQ486596.1
Helianthus nuttallii	P	DQ486598.1
Helianthus occidentalis	P	DQ486601.1
Helianthus paradoxus	A	DQ486606.1
Helianthus petiolaris	A	DQ486611.1
Helianthus praecox	A	DQ486613.1
Helianthus pumilius	P	DQ486614.1
Helianthus radula	P	DQ486615.1
Helianthus resinosus	P	DQ486617.1
Helianthus salicifolius	P	DQ486619.1
Helianthus schweinitzii	P	DQ486622.1
Helianthus simulans	P	DQ486625.1
Helianthus strumosus	P	DQ486627.1
Helianthus porteri	P	DQ486612.1
Helianthus eggertii	P	DQ486570.1

Name	Life History	Accession Number
Helianthus tuberosus	P	DQ486630.1

Appendix B Supplementary information for chapter 3

B.1 Sample information by population for chapter 3. Including soil measurements for *H. bolanderi-exilis* samples, F_{IS}, sample location, and seed accession.

Population	Species	Sample size	F_{IS}	Latitude	Longitude	Area	Elevation (Feet)
G100	*H. bolanderi-exilis*	10	0.092294135	39.40117	-122.61349	Coast Mountains	2382
G101	*H. bolanderi-exilis*	3	-0.116186363	39.26759	-122.48275	Coast Mountains	1312
G102	*H. bolanderi-exilis*	10	0.166771372	39.12638	-122.43213	Coast Mountains	1270
G103	*H. bolanderi-exilis*	10	-0.003609194	38.7804	-122.57185	Coast Mountains	1030
G108	*H. bolanderi-exilis*	11	-0.038774216	38.87585	-120.8205	Sierra Nevada Mountains	2304
G109	*H. bolanderi-exilis*	10	-0.03104927	39.17832	-121.75977	Central Valley	113
G110	*H. bolanderi-exilis*	6	-0.069590884	39.25156	-121.88924	Central Valley	37
G111	*H. bolanderi-exilis*	10	0.041052216	39.34395	-121.44869	Central Valley	298
G114	*H. bolanderi-exilis*	11	0.09922861	41.28199	-122.85186	North Mountains	3670
G115	*H. bolanderi-exilis*	7	0.026255805	41.64306	-122.74711	North Mountains	3140
G116	*H. bolanderi-exilis*	5	0.013262151	39.066322	-122.478403	Coast Mountains	756
G118	*H. bolanderi-exilis*	9	0.040231136	39.2627	-122.51157	Coast Mountains	1309
G119	*H. bolanderi-exilis*	9	0.072372213	39.48584	-121.31271	Sierra Nevada Mountains	2497
G120	*H. bolanderi-exilis*	8	0.011322698	38.543	-121.7383	Central Valley	NA
G121	*H. bolanderi-exilis*	10	0.06427413	38.82395	-122.33725	Coast Mountains	1300
G122	*H. bolanderi-exilis*	8	0.08619789	38.73309	-122.52462	Coast Mountains	976
G123	*H. bolanderi-exilis*	10	-0.007381898	39.83434	-121.58227	Sierra Nevada Mountains	2320
G124	*H. bolanderi-exilis*	10	0.048286179	38.84119	-120.87647	Sierra Nevada Mountains	1600
G127	*H. bolanderi-exilis*	10	0.037019361	37.84557	-120.46388	Sierra Nevada Mountains	1100
G128	*H. bolanderi-exilis*	4	-0.172127528	41.03086	-122.42451	North Mountains	2682
G129	*H. bolanderi-exilis*	6	0.028531963	39.88756	-122.63451	Coast Mountains	2240
G130	*H. bolanderi-exilis*	10	0.093057773	41.29794	-122.72187	North Mountains	4530
cal_ann	*H. annuus*	24	0.339046504	NA	NA	California	NA
cen_ann	*H. annuus*	76	0.393825339	NA	NA	Central USA	NA
div	*H. divaricatus*	5	-0.026229967	NA	NA	Central USA	NA
gig	*H. giganteus*	5	0.347009762	NA	NA	Central USA	NA
gro	*H. grosseserratus*	6	0.299273251	NA	NA	Central USA	NA
max	*H. maximiliani*	10	0.262673156	NA	NA	Central USA	NA
nut	*H. nuttallii*	3	-0.368190528	NA	NA	Central USA	NA

Population	Serpentine?	Mg/Ca Ratio	Organic Matter %	P (ppm)	K (ppm)	Mg (ppm)	Ca (ppm)	Na (ppm)	pH	Cation Exchange Capacity	S (ppm)	Cr (mg/kg)	Co (mg/kg)	Ni (mg/kg)
G100	yes	4.26	4.1	28	149	2807	659	13	7.9	26.8	4	0	0	5.2
G101	no	0.48	3	14	142	1486	3080	82	7.5	28.3	9	0	0	1
G102	yes	3.38	6.5	16	139	2934	868	41	7.6	29	21	0	0	12.9
G103	yes	2.41	4.9	31	165	2043	848	21	6.9	21.9	6	0	0	15.7
G108	yes	2.66	3.9	61	80	1475	554	11	7.3	15.1	3	0	0	26.5
G109	no	0.16	2.1	78	235	293	1865	6	6.3	13.8	2	0	0	1.3
G110	no	0.30	1	38*	105	238	786	13	5.5	8.4	5	0	0.1	0.9
G111	no	0.14	2.8	65	87	267	1914	4	6.7	12.6	7	0	0.1	0.3
G114	yes	4.53	3	43	104	2707	598	1	7.7	25.5	1	0	0.2	10.9
G115	yes	13.02	3.7	66	105	3229	248	42	7.8	28.2	2	0	0	7.3
G116	yes	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G118	yes	1.89	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G119	no	0.26	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G120	no	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G121	yes	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G122	yes	2.78	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G123	yes	6.25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G124	yes	2.50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G127	yes	1.82	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G128	yes	1.85	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G129	no	0.84	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G130	yes	2.56	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
cal_ann	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
cen_ann	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
div	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
gig	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
gro	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
max	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nut	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Population	Collected by	USDA-GRIN collection
G100	Gregory Owens	*PI 649893
G101	Gregory Owens	NA
G102	Gregory Owens	NA
G103	Gregory Owens	*PI 649888
G108	Gregory Owens	*PI 664632
G109	Gregory Owens	NA
G110	Gregory Owens	NA
G111	Gregory Owens	NA
G114	Gregory Owens	*PI 649896
G115	Gregory Owens	*PI 649895
G116	Jake Schweitzer	NA
G118	NA	Ames 27232
G119	NA	PI 649899
G120	NA	PI 435644
G121	NA	PI 468662
G122	NA	PI 649889
G123	NA	PI 649898
G124	NA	PI 649900
G127	NA	PI 649901
G128	NA	PI 649897
G129	NA	PI 664630
G130	NA	PI 649894
cal_ann	NA	NA
cen_ann	NA	NA
div	NA	NA
gig	NA	NA
gro	NA	NA
max	NA	NA
nut	NA	NA

B.2 Sample information by individual for chapter 3, including read number, percent reads aligned, sample location, SRA accession and seed accession.

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
DB114	DB24	gro	1027221	971482	0.95	*Helianthus grosseserratus*
DB118	DB23	gro	709482	668205	0.94	*Helianthus grosseserratus*
DB124	DB25	gro	808156	772022	0.96	*Helianthus grosseserratus*
DB129	DB22	gro	1615518	1505050	0.93	*Helianthus grosseserratus*
DB209	DB05	gro	468094	447231	0.96	*Helianthus grosseserratus*
DB291	DB02	gig	954540	900749	0.94	*Helianthus giganteus*
DB295	DB03	gig	2083312	1986001	0.95	*Helianthus giganteus*
DB297	DB04	gig	1655969	1574442	0.95	*Helianthus giganteus*
DB302	DB06	gro	641355	608036	0.95	*Helianthus grosseserratus*
DB320	DB07	div	3279458	2980108	0.91	*Helianthus divaricatus*
DB322	DB10	div	1688213	1582053	0.94	*Helianthus divaricatus*
DB324	DB19	div	3370246	1914645	0.57	*Helianthus divaricatus*
DB325	DB08	div	1781305	1456847	0.82	*Helianthus divaricatus*
DB329	DB09	div	3245652	2951454	0.91	*Helianthus divaricatus*
DB38	DB21	gig	358842	331481	0.92	*Helianthus giganteus*
DB94	DB20	gig	580214	549996	0.95	*Helianthus giganteus*
G100.12	G100.12	G100	1367798	1175175	0.86	*Helianthus bolanderi-exilis*
G100.13	G100.13	G100	2865447	2731907	0.95	*Helianthus bolanderi-exilis*
G100.14	G100.14	G100	2355890	2252860	0.96	*Helianthus bolanderi-exilis*
G100.2	G100.2	G100	1785164	1336832	0.75	*Helianthus bolanderi-exilis*
G100.20	G100.20	G100	3617864	3036259	0.84	*Helianthus bolanderi-exilis*
G100.21	G100.21	G100	6968997	2066593	0.30	*Helianthus bolanderi-exilis*
G100.22	G100.22	G100	2210704	1665056	0.75	*Helianthus bolanderi-exilis*
G100.4	G100.4	G100	1735667	1106484	0.64	*Helianthus bolanderi-exilis*
G100.5	G100.5	G100	7915323	3575732	0.45	*Helianthus bolanderi-exilis*
G100.6	G100.6	G100	5996926	5594429	0.93	*Helianthus bolanderi-exilis*
G101.3	G101.3	G101	1396666	1239402	0.89	*Helianthus bolanderi-exilis*
G101.4	G101.4	G101	7241017	5949508	0.82	*Helianthus bolanderi-exilis*
G101.5	G101.5	G101	3445544	2900088	0.84	*Helianthus bolanderi-exilis*
G102.1	G102.1	G102	1859898	1803985	0.97	*Helianthus bolanderi-exilis*
G102.12	G102.12	G102	7175304	1058980	0.15	*Helianthus bolanderi-exilis*
G102.13	G102.13	G102	5022218	4813288	0.96	*Helianthus bolanderi-exilis*
G102.2	G102.2	G102	2581442	2326584	0.90	*Helianthus bolanderi-exilis*
G102.23	G102.23	G102	4170116	2072163	0.50	*Helianthus bolanderi-exilis*
G102.3	G102.3	G102	13535912	4132358	0.31	*Helianthus bolanderi-exilis*
G102.4	G102.4	G102	3910313	2164469	0.55	*Helianthus bolanderi-exilis*
G102.7	G102.7	G102	8458282	1912682	0.23	*Helianthus bolanderi-exilis*
G102.8	G102.8	G102	5825942	5296334	0.91	*Helianthus bolanderi-exilis*
G102.9	G102.9	G102	4036120	2079960	0.52	*Helianthus bolanderi-exilis*
G103.1	G103.1	G103	1996322	1844417	0.92	*Helianthus bolanderi-exilis*
G103.12	G103.12	G103	5676614	5349161	0.94	*Helianthus bolanderi-exilis*
G103.2	G103.2	G103	4039934	3884843	0.96	*Helianthus bolanderi-exilis*
G103.3	G103.3	G103	4767335	4392495	0.92	*Helianthus bolanderi-exilis*
G103.4	G103.4	G103	4527145	3896977	0.86	*Helianthus bolanderi-exilis*
G103.5	G103.5	G103	2846455	2679088	0.94	*Helianthus bolanderi-exilis*
G103.6	G103.6	G103	1307524	1145043	0.88	*Helianthus bolanderi-exilis*
G103.7	G103.7	G103	2799123	2650635	0.95	*Helianthus bolanderi-exilis*

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
G103.8	G103.8	G103	6096422	5497322	0.90	*Helianthus bolanderi-exilis*
G103.9	G103.9	G103	4471996	4099077	0.92	*Helianthus bolanderi-exilis*
G108.13	G108.13	G108	3721161	3598988	0.97	*Helianthus bolanderi-exilis*
G108.17	G108.17	G108	1987803	1837015	0.92	*Helianthus bolanderi-exilis*
G108.2	G108.2	G108	4325736	4058601	0.94	*Helianthus bolanderi-exilis*
G108.20	G108.20	G108	1586021	1465115	0.92	*Helianthus bolanderi-exilis*
G108.3	G108.3	G108	1329141	1281663	0.96	*Helianthus bolanderi-exilis*
G108.4	G108.4	G108	1944987	1885006	0.97	*Helianthus bolanderi-exilis*
G108.5	G108.5	G108	4557272	4117701	0.90	*Helianthus bolanderi-exilis*
G108.6	G108.6	G108	1044158	938720	0.90	*Helianthus bolanderi-exilis*
G108.7	G108.7	G108	2667881	2564686	0.96	*Helianthus bolanderi-exilis*
G108.8	G108.8	G108	1932091	1826532	0.95	*Helianthus bolanderi-exilis*
G108.9	G108.9	G108	5482199	5179544	0.94	*Helianthus bolanderi-exilis*
G109.1	G109.1	G109	3770793	2914480	0.77	*Helianthus bolanderi-exilis*
G109.10	G109.10	G109	1235215	1156610	0.94	*Helianthus bolanderi-exilis*
G109.2	G109.2	G109	2625719	2075495	0.79	*Helianthus bolanderi-exilis*
G109.3	G109.3	G109	2826289	2703145	0.96	*Helianthus bolanderi-exilis*
G109.4	G109.4	G109	1392252	1307409	0.94	*Helianthus bolanderi-exilis*
G109.5	G109.5	G109	3739033	3591434	0.96	*Helianthus bolanderi-exilis*
G109.6	G109.6	G109	967898	922753	0.95	*Helianthus bolanderi-exilis*
G109.7	G109.7	G109	3222241	2872665	0.89	*Helianthus bolanderi-exilis*
G109.8	G109.8	G109	224816	214692	0.95	*Helianthus bolanderi-exilis*
G109.9	G109.9	G109	5186039	4933732	0.95	*Helianthus bolanderi-exilis*
G110.1	G110.1	G110	2741812	2665060	0.97	*Helianthus bolanderi-exilis*
G110.11	G110.11	G110	3681024	3586153	0.97	*Helianthus bolanderi-exilis*
G110.12	G110.12	G110	1981722	1887655	0.95	*Helianthus bolanderi-exilis*
G110.3	G110.3	G110	1001666	864694	0.86	*Helianthus bolanderi-exilis*
G110.6	G110.6	G110	1180872	1110079	0.94	*Helianthus bolanderi-exilis*
G110.9	G110.9	G110	1985916	1931398	0.97	*Helianthus bolanderi-exilis*
G111.1	G111.1	G111	2878353	2719995	0.94	*Helianthus bolanderi-exilis*
G111.10	G111.10	G111	4962233	4768535	0.96	*Helianthus bolanderi-exilis*
G111.11	G111.11	G111	2150224	1834459	0.85	*Helianthus bolanderi-exilis*
G111.3	G111.3	G111	2129274	1437117	0.67	*Helianthus bolanderi-exilis*
G111.4	G111.4	G111	1451146	1391552	0.96	*Helianthus bolanderi-exilis*
G111.5	G111.5	G111	1582412	1525287	0.96	*Helianthus bolanderi-exilis*
G111.6	G111.6	G111	1687187	1580404	0.94	*Helianthus bolanderi-exilis*
G111.7	G111.7	G111	690425	668331	0.97	*Helianthus bolanderi-exilis*
G111.8	G111.8	G111	1263823	1218031	0.96	*Helianthus bolanderi-exilis*
G111.9	G111.9	G111	7113246	6778430	0.95	*Helianthus bolanderi-exilis*
G114.10	G114.10	G114	2224966	1207279	0.54	*Helianthus bolanderi-exilis*
G114.13	G114.13	G114	2171035	2074520	0.96	*Helianthus bolanderi-exilis*
G114.14	G114.14	G114	3090909	2945715	0.95	*Helianthus bolanderi-exilis*
G114.15	G114.15	G114	1130094	1082354	0.96	*Helianthus bolanderi-exilis*
G114.18	G114.18	G114	2635268	2555707	0.97	*Helianthus bolanderi-exilis*
G114.19	G114.19	G114	767313	732752	0.95	*Helianthus bolanderi-exilis*
G114.20	G114.20	G114	3524906	3361591	0.95	*Helianthus bolanderi-exilis*
G114.21	G114.21	G114	3887691	3760768	0.97	*Helianthus bolanderi-exilis*
G114.24	G114.24	G114	1209387	1163926	0.96	*Helianthus bolanderi-exilis*
G114.25	G114.25	G114	1514703	1467049	0.97	*Helianthus bolanderi-exilis*
G114.29	G114.29	G114	2653363	2506420	0.94	*Helianthus bolanderi-exilis*
G115.10	G115.10	G115	5684206	3520157	0.62	*Helianthus bolanderi-exilis*
G115.11	G115.11	G115	3872470	3742778	0.97	*Helianthus bolanderi-exilis*
G115.12	G115.12	G115	2487729	2126326	0.85	*Helianthus bolanderi-exilis*

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
G115.3	G115.3	G115	7803789	5610356	0.72	*Helianthus bolanderi-exilis*
G115.4	G115.4	G115	1420499	1377938	0.97	*Helianthus bolanderi-exilis*
G115.7	G115.7	G115	28782566	28071051	0.98	*Helianthus bolanderi-exilis*
G115.9	G115.9	G115	1982604	1917170	0.97	*Helianthus bolanderi-exilis*
G116.13	G116.13	G116	4576583	4427395	0.97	*Helianthus bolanderi-exilis*
G116.14	G116.14	G116	1062011	1017201	0.96	*Helianthus bolanderi-exilis*
G116.15	G116.15	G116	2266291	2093633	0.92	*Helianthus bolanderi-exilis*
G116.4	G116.4	G116	3928123	3776587	0.96	*Helianthus bolanderi-exilis*
G116.6	G116.6	G116	4757866	4636599	0.97	*Helianthus bolanderi-exilis*
G118.11	G118.11	G118	2617281	2461465	0.94	*Helianthus bolanderi-exilis*
G118.12	G118.12	G118	1898691	1803272	0.95	*Helianthus bolanderi-exilis*
G118.2	G118.2	G118	4470480	4282012	0.96	*Helianthus bolanderi-exilis*
G118.3	G118.3	G118	4177652	4024286	0.96	*Helianthus bolanderi-exilis*
G118.5	G118.5	G118	1487042	1425830	0.96	*Helianthus bolanderi-exilis*
G118.6	G118.6	G118	3381717	3276542	0.97	*Helianthus bolanderi-exilis*
G118.7	G118.7	G118	3274741	3146063	0.96	*Helianthus bolanderi-exilis*
G118.8	G118.8	G118	1418224	1361215	0.96	*Helianthus bolanderi-exilis*
G118.9	G118.9	G118	259394	249641	0.96	*Helianthus bolanderi-exilis*
G119.1	G119.1	G119	3025761	2923838	0.97	*Helianthus bolanderi-exilis*
G119.2	G119.2	G119	2345468	2278077	0.97	*Helianthus bolanderi-exilis*
G119.3	G119.3	G119	560398	544415	0.97	*Helianthus bolanderi-exilis*
G119.4	G119.4	G119	1369987	1259387	0.92	*Helianthus bolanderi-exilis*
G119.5	G119.5	G119	1766036	1708644	0.97	*Helianthus bolanderi-exilis*
G119.6	G119.6	G119	1090753	1065048	0.98	*Helianthus bolanderi-exilis*
G119.7	G119.7	G119	4227755	4120688	0.97	*Helianthus bolanderi-exilis*
G119.8	G119.8	G119	7344003	6993569	0.95	*Helianthus bolanderi-exilis*
G119.9	G119.9	G119	4083811	3544592	0.87	*Helianthus bolanderi-exilis*
G120.10	G120.10	G120	6173064	5615232	0.91	*Helianthus bolanderi-exilis*
G120.11	G120.11	G120	3892757	3662080	0.94	*Helianthus bolanderi-exilis*
G120.12	G120.12	G120	1623804	1555019	0.96	*Helianthus bolanderi-exilis*
G120.15	G120.15	G120	2701160	2584554	0.96	*Helianthus bolanderi-exilis*
G120.17	G120.17	G120	2968903	1708356	0.58	*Helianthus bolanderi-exilis*
G120.2	G120.2	G120	2918393	2797076	0.96	*Helianthus bolanderi-exilis*
G120.7	G120.7	G120	3647002	3482010	0.95	*Helianthus bolanderi-exilis*
G120.8	G120.8	G120	4686330	4379243	0.93	*Helianthus bolanderi-exilis*
G121.1	G121.1	G121	1861826	1812230	0.97	*Helianthus bolanderi-exilis*
G121.10	G121.10	G121	4198496	4098312	0.98	*Helianthus bolanderi-exilis*
G121.2	G121.2	G121	2275170	2215126	0.97	*Helianthus bolanderi-exilis*
G121.3	G121.3	G121	2366926	2305774	0.97	*Helianthus bolanderi-exilis*
G121.4	G121.4	G121	7830599	7613137	0.97	*Helianthus bolanderi-exilis*
G121.5	G121.5	G121	12170859	11762574	0.97	*Helianthus bolanderi-exilis*
G121.6	G121.6	G121	1938738	1864830	0.96	*Helianthus bolanderi-exilis*
G121.7	G121.7	G121	7375027	7205726	0.98	*Helianthus bolanderi-exilis*
G121.8	G121.8	G121	1176261	814042	0.69	*Helianthus bolanderi-exilis*
G121.9	G121.9	G121	1617356	1579693	0.98	*Helianthus bolanderi-exilis*
G122.1	G122.1	G122	2774955	2618256	0.94	*Helianthus bolanderi-exilis*
G122.11	G122.11	G122	8925364	8763464	0.98	*Helianthus bolanderi-exilis*
G122.2	G122.2	G122	1736236	1662927	0.96	*Helianthus bolanderi-exilis*
G122.3	G122.3	G122	2771564	2693549	0.97	*Helianthus bolanderi-exilis*
G122.5	G122.5	G122	641828	613477	0.96	*Helianthus bolanderi-exilis*
G122.6	G122.6	G122	912889	876707	0.96	*Helianthus bolanderi-exilis*
G122.7	G122.7	G122	3001265	2605349	0.87	*Helianthus bolanderi-exilis*
G122.8	G122.8	G122	1211834	1176106	0.97	*Helianthus bolanderi-exilis*

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
G123.12	G123.12	G123	6138261	5818225	0.95	*Helianthus bolanderi-exilis*
G123.13	G123.13	G123	3175055	2827261	0.89	*Helianthus bolanderi-exilis*
G123.15	G123.15	G123	5887046	5602697	0.95	*Helianthus bolanderi-exilis*
G123.17	G123.17	G123	2870118	1866121	0.65	*Helianthus bolanderi-exilis*
G123.2	G123.2	G123	5335592	5079734	0.95	*Helianthus bolanderi-exilis*
G123.4	G123.4	G123	897336	860134	0.96	*Helianthus bolanderi-exilis*
G123.5	G123.5	G123	2301583	2199156	0.96	*Helianthus bolanderi-exilis*
G123.6	G123.6	G123	4101844	3936776	0.96	*Helianthus bolanderi-exilis*
G123.7	G123.7	G123	3354079	3206396	0.96	*Helianthus bolanderi-exilis*
G123.8	G123.8	G123	7740995	7508855	0.97	*Helianthus bolanderi-exilis*
G124.1	G124.1	G124	4336703	4198072	0.97	*Helianthus bolanderi-exilis*
G124.10	G124.10	G124	304324	291560	0.96	*Helianthus bolanderi-exilis*
G124.11	G124.11	G124	3306566	3195749	0.97	*Helianthus bolanderi-exilis*
G124.12	G124.12	G124	637464	604358	0.95	*Helianthus bolanderi-exilis*
G124.2	G124.2	G124	1820010	1732364	0.95	*Helianthus bolanderi-exilis*
G124.3	G124.3	G124	5310484	5042455	0.95	*Helianthus bolanderi-exilis*
G124.5	G124.5	G124	2723584	2594604	0.95	*Helianthus bolanderi-exilis*
G124.6	G124.6	G124	2593317	2464618	0.95	*Helianthus bolanderi-exilis*
G124.7	G124.7	G124	1105565	1047874	0.95	*Helianthus bolanderi-exilis*
G124.9	G124.9	G124	3555760	3379088	0.95	*Helianthus bolanderi-exilis*
G127.1	G127.1	G127	2181245	2126830	0.98	*Helianthus bolanderi-exilis*
G127.10	G127.10	G127	3104007	3012153	0.97	*Helianthus bolanderi-exilis*
G127.2	G127.2	G127	3632501	3369583	0.93	*Helianthus bolanderi-exilis*
G127.3	G127.3	G127	5834284	5650853	0.97	*Helianthus bolanderi-exilis*
G127.4	G127.4	G127	4688246	4467911	0.95	*Helianthus bolanderi-exilis*
G127.5	G127.5	G127	1834236	1752407	0.96	*Helianthus bolanderi-exilis*
G127.6	G127.6	G127	8447723	7746341	0.92	*Helianthus bolanderi-exilis*
G127.7	G127.7	G127	3658743	3527701	0.96	*Helianthus bolanderi-exilis*
G127.8	G127.8	G127	1327017	1259099	0.95	*Helianthus bolanderi-exilis*
G127.9	G127.9	G127	1858320	1779876	0.96	*Helianthus bolanderi-exilis*
G128.1	G128.1	G128	3396585	3270459	0.96	*Helianthus bolanderi-exilis*
G128.2	G128.2	G128	8583870	8181670	0.95	*Helianthus bolanderi-exilis*
G128.3	G128.3	G128	7061432	6769607	0.96	*Helianthus bolanderi-exilis*
G128.4	G128.4	G128	2564790	2372173	0.92	*Helianthus bolanderi-exilis*
G129.11	G129.11	G129	4898678	4749838	0.97	*Helianthus bolanderi-exilis*
G129.4	G129.4	G129	4903122	3751479	0.77	*Helianthus bolanderi-exilis*
G129.5	G129.5	G129	1126189	1075596	0.96	*Helianthus bolanderi-exilis*
G129.6	G129.6	G129	2809633	2740814	0.98	*Helianthus bolanderi-exilis*
G129.8	G129.8	G129	2983976	2270035	0.76	*Helianthus bolanderi-exilis*
G129.9	G129.9	G129	4531209	4236061	0.93	*Helianthus bolanderi-exilis*
G130.1	G130.1	G130	2554400	2461415	0.96	*Helianthus bolanderi-exilis*
G130.10	G130.10	G130	1790668	1725658	0.96	*Helianthus bolanderi-exilis*
G130.2	G130.2	G130	2738894	2617300	0.96	*Helianthus bolanderi-exilis*
G130.3	G130.3	G130	2507265	2253653	0.90	*Helianthus bolanderi-exilis*
G130.4	G130.4	G130	2246092	2169184	0.97	*Helianthus bolanderi-exilis*
G130.5	G130.5	G130	1467241	1357915	0.93	*Helianthus bolanderi-exilis*
G130.6	G130.6	G130	3956354	3799289	0.96	*Helianthus bolanderi-exilis*
G130.7	G130.7	G130	5057877	4797240	0.95	*Helianthus bolanderi-exilis*
G130.8	G130.8	G130	885118	850496	0.96	*Helianthus bolanderi-exilis*
G130.9	G130.9	G130	311646	301957	0.97	*Helianthus bolanderi-exilis*
GB001	nut01	nut	4705239	4397353	0.93	*Helianthus nutallii*
GB002	nut02	nut	1661075	1566437	0.94	*Helianthus nutallii*
GB003	nut03	nut	1775383	1665927	0.94	*Helianthus nutallii*

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
GB011	ann01	cen_ann	6638468	5303152	0.80	*Helianthus annuus*
GB013	ann02	cen_ann	401261	372389	0.93	*Helianthus annuus*
GB014	ann93	cal_ann	2807149	2719411	0.97	*Helianthus annuus*
GB015	ann03	cen_ann	2867641	2642730	0.92	*Helianthus annuus*
GB016	ann04	cen_ann	1251824	1163305	0.93	*Helianthus annuus*
GB020	ann05	cen_ann	6870322	6594239	0.96	*Helianthus annuus*
GB025	ann06	cen_ann	4361425	4274154	0.98	*Helianthus annuus*
GB026	ann94	cal_ann	6052798	5825744	0.96	*Helianthus annuus*
GB027	ann95	cal_ann	7119034	5479942	0.77	*Helianthus annuus*
GB028	ann96	cal_ann	4389235	4090087	0.93	*Helianthus annuus*
GB029	ann07	cen_ann	3686192	3535486	0.96	*Helianthus annuus*
GB031	ann08	cen_ann	3788363	3253346	0.86	*Helianthus annuus*
GB032	ann09	cen_ann	5034343	4606045	0.91	*Helianthus annuus*
GB034	ann10	cen_ann	4381006	4206143	0.96	*Helianthus annuus*
GB035	ann11	cen_ann	3469247	3359348	0.97	*Helianthus annuus*
GB036	ann12	cal_ann	7143395	6865045	0.96	*Helianthus annuus*
GB037	ann13	cen_ann	2716148	2055151	0.76	*Helianthus annuus*
GB041	ann14	cen_ann	7225304	5793539	0.80	*Helianthus annuus*
GB042	ann15	cal_ann	6344334	5187280	0.82	*Helianthus annuus*
GB043	ann16	cen_ann	3147274	2934371	0.93	*Helianthus annuus*
GB044	ann17	cen_ann	3163584	3049466	0.96	*Helianthus annuus*
GB047	ann18	cen_ann	5177316	4893072	0.95	*Helianthus annuus*
GB048	ann19	cal_ann	10395611	10171813	0.98	*Helianthus annuus*
GB049	ann20	cen_ann	3870164	3769410	0.97	*Helianthus annuus*
GB050	ann21	cen_ann	3948545	3530145	0.89	*Helianthus annuus*
GB051	ann22	cen_ann	6411931	5978221	0.93	*Helianthus annuus*
GB052	ann23	cen_ann	1776605	1742374	0.98	*Helianthus annuus*
GB053	ann24	cen_ann	1980619	1799313	0.91	*Helianthus annuus*
GB054	ann25	cen_ann	3445156	3144586	0.91	*Helianthus annuus*
GB062	max01	max	2947500	2779966	0.94	*Helianthus maximilliani*
GB063	max02	max	1026646	909298	0.89	*Helianthus maximilliani*
GB064	max03	max	1632594	1528253	0.94	*Helianthus maximilliani*
GB065	max04	max	1753720	1556985	0.89	*Helianthus maximilliani*
GB098	ann26	cen_ann	2155028	2021514	0.94	*Helianthus annuus*
GB099	ann27	cen_ann	2295913	2213407	0.96	*Helianthus annuus*
GB100	ann28	cen_ann	974662	936649	0.96	*Helianthus annuus*
GB101	ann29	cen_ann	713558	671912	0.94	*Helianthus annuus*
GB102	ann30	cen_ann	346552	323041	0.93	*Helianthus annuus*
GB103	ann31	cen_ann	3733300	3575257	0.96	*Helianthus annuus*
GB104	ann32	cen_ann	2426577	2366797	0.98	*Helianthus annuus*
GB105	ann33	cen_ann	3315898	3213037	0.97	*Helianthus annuus*
GB106	ann34	cen_ann	2237353	2180567	0.97	*Helianthus annuus*
GB107	ann35	cen_ann	5136470	4995951	0.97	*Helianthus annuus*
GB110	ann36	cen_ann	3109624	2946418	0.95	*Helianthus annuus*
GB111	ann37	cen_ann	4086839	3894501	0.95	*Helianthus annuus*
GB113	ann38	cen_ann	646208	595144	0.92	*Helianthus annuus*
GB114	ann39	cen_ann	3601263	3439961	0.96	*Helianthus annuus*
GB115	ann40	cen_ann	2126371	2057999	0.97	*Helianthus annuus*
GB116	ann41	cen_ann	2769993	2543446	0.92	*Helianthus annuus*
GB117	ann42	cen_ann	2372985	2209441	0.93	*Helianthus annuus*
GB118	ann43	cen_ann	1273891	1195111	0.94	*Helianthus annuus*
GB119	ann44	cen_ann	3568814	3508489	0.98	*Helianthus annuus*
GB120	ann45	cal_ann	6704884	6286021	0.94	*Helianthus annuus*

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
GB121	ann46	cal_ann	5555605	5404404	0.97	*Helianthus annuus*
GB122	ann47	cen_ann	5174409	4799240	0.93	*Helianthus annuus*
GB123	ann48	cal_ann	3845140	3654905	0.95	*Helianthus annuus*
GB124	ann49	cal_ann	3322099	3121764	0.94	*Helianthus annuus*
GB125	ann50	cal_ann	4784296	4461408	0.93	*Helianthus annuus*
GB126	ann51	cen_ann	330291	283664	0.86	*Helianthus annuus*
GB127	ann52	cal_ann	3560541	3280046	0.92	*Helianthus annuus*
GB128	ann53	cen_ann	1996477	1933746	0.97	*Helianthus annuus*
GB129	ann54	cen_ann	3814164	3720359	0.98	*Helianthus annuus*
GB130	ann55	cen_ann	2263824	2163222	0.96	*Helianthus annuus*
GB131	ann56	cen_ann	5364128	5075297	0.95	*Helianthus annuus*
GB132	ann57	cen_ann	2679263	2526449	0.94	*Helianthus annuus*
GB133	ann58	cen_ann	387672	358166	0.92	*Helianthus annuus*
GB134	ann59	cen_ann	2759810	2657997	0.96	*Helianthus annuus*
GB135	ann60	cen_ann	2909910	2801782	0.96	*Helianthus annuus*
GB142	max05	max	4078604	3822172	0.94	*Helianthus maximilliani*
GB143	max06	max	2084452	1953842	0.94	*Helianthus maximilliani*
GB146	max07	max	1910067	1792293	0.94	*Helianthus maximilliani*
GB169	ann61	cen_ann	762182	727093	0.95	*Helianthus annuus*
GB170	ann62	cen_ann	2563939	2481044	0.97	*Helianthus annuus*
GB171	ann63	cen_ann	822431	785935	0.96	*Helianthus annuus*
GB172	ann64	cen_ann	1180574	1135998	0.96	*Helianthus annuus*
GB173	ann65	cen_ann	3585037	3490127	0.97	*Helianthus annuus*
GB174	ann66	cen_ann	1790801	1721807	0.96	*Helianthus annuus*
GB175	ann67	cen_ann	1455714	1404169	0.96	*Helianthus annuus*
GB176	ann68	cen_ann	2470821	2386255	0.97	*Helianthus annuus*
GB177	ann69	cen_ann	3419934	3316802	0.97	*Helianthus annuus*
GB178	ann70	cen_ann	984149	950801	0.97	*Helianthus annuus*
GB182	ann71	cen_ann	1619217	1559379	0.96	*Helianthus annuus*
GB183	ann72	cen_ann	1374906	1282946	0.93	*Helianthus annuus*
GB184	ann73	cal_ann	5435889	5281876	0.97	*Helianthus annuus*
GB185	ann74	cen_ann	1615152	1343802	0.83	*Helianthus annuus*
GB186	ann75	cen_ann	1049978	998197	0.95	*Helianthus annuus*
GB187	ann76	cen_ann	1418974	1369138	0.96	*Helianthus annuus*
GB188	ann77	cen_ann	2318174	2253418	0.97	*Helianthus annuus*
GB189	ann78	cal_ann	900482	877954	0.97	*Helianthus annuus*
GB190	ann79	cal_ann	5444450	5320730	0.98	*Helianthus annuus*
GB191	ann80	cal_ann	1633092	1583482	0.97	*Helianthus annuus*
GB192	ann81	cal_ann	2940602	2848306	0.97	*Helianthus annuus*
GB193	ann82	cal_ann	937211	900700	0.96	*Helianthus annuus*
GB194	ann83	cal_ann	3907213	3822224	0.98	*Helianthus annuus*
GB195	ann84	cal_ann	4573313	4472428	0.98	*Helianthus annuus*
GB198	ann85	cen_ann	6086721	5961171	0.98	*Helianthus annuus*
GB199	ann86	cen_ann	649744	639017	0.98	*Helianthus annuus*
GB200	ann87	cen_ann	3942287	3869602	0.98	*Helianthus annuus*
GB201	ann88	cen_ann	967292	910198	0.94	*Helianthus annuus*
GB202	ann89	cen_ann	2854041	2793784	0.98	*Helianthus annuus*
GB204	ann204	cen_ann	4376267	4293663	0.98	*Helianthus annuus*
GB205	ann91	cen_ann	8639669	8494239	0.98	*Helianthus annuus*
GB206	ann92	cen_ann	4045263	3966800	0.98	*Helianthus annuus*
GB225	ann225	cal_ann	912353	889670	0.98	*Helianthus annuus*
GB249	ann97	cal_ann	3737638	3678289	0.98	*Helianthus annuus*
GB250	bol250	cal_ann	3561392	3491292	0.98	*Helianthus annuus*

Sample name	Alternate name	Population	# Reads	Mapped reads	% mapped	Species
GB255	ann255	cen_ann	5602936	5502888	0.98	*Helianthus annuus*
GB277	max277	max	391768	369393	0.94	*Helianthus maximilliani*
GB278	max08	max	7021568	6647214	0.95	*Helianthus maximilliani*
GB282	max09	max	11941077	11328615	0.95	*Helianthus maximilliani*

Sample name	Seed accession	Latitude	Longitude	SRA number
DB114	PI 547195	45.2	-85.16666667	SRR2169752
DB118	PI 586890	42.16666667	-100.3833333	SRR2169753
DB124	PI 547192	40.73333333	-88.76666667	SRR2169754
DB129	PI 547202	41.68333333	-93.13333333	SRR2169755
DB209	PI 468726	33.46666667	-89.71666667	SRR2169756
DB291	PI 664647	41.59083333	-83.76194444	SRR2169747
DB295	PI 664710	35.81166667	-82.1972222	SRR2169748
DB297	PI 468719	36.3	-78.58333333	SRR2169749
DB302	PI 468725	34.91666667	-95.3	SRR2169757
DB320	PI 503218	40	-77	SRR2169731
DB322	PI 664604	43.06666667	-89.43333333	SRR2169732
DB324	PI 503209	37	-80	SRR2169733
DB325	PI 664645	38.81083333	-83.53027778	SRR2169734
DB329	PI 547174	39.18333333	-88.8	SRR2169735
DB38	PI 503223	36	-77	SRR2169750
DB94	*PI 649893	45.25	-88.6	SRR2169751
G100.12	*PI 649893	39.40117	-122.61349	SRR2169854
G100.13	*PI 649893	39.40117	-122.61349	SRR2169855
G100.14	*PI 649893	39.40117	-122.61349	SRR2169856
G100.2	*PI 649893	39.40117	-122.61349	SRR2169857
G100.20	*PI 649893	39.40117	-122.61349	SRR2169858
G100.21	*PI 649893	39.40117	-122.61349	SRR2169859
G100.22	*PI 649893	39.40117	-122.61349	SRR2169860
G100.4	*PI 649893	39.40117	-122.61349	SRR2169861
G100.5	*PI 649893	39.40117	-122.61349	SRR2169862
G100.6	*PI 649893	39.40117	-122.61349	SRR2169863
G101.3	NA	39.26759	-122.48275	SRR2169864
G101.4	NA	39.26759	-122.48275	SRR2169865
G101.5	NA	39.26759	-122.48275	SRR2169866
G102.1	NA	39.12638	-122.43213	SRR2169867
G102.12	NA	39.12638	-122.43213	SRR2169868
G102.13	NA	39.12638	-122.43213	SRR2169869
G102.2	NA	39.12638	-122.43213	SRR2169870
G102.23	NA	39.12638	-122.43213	SRR2169871
G102.3	NA	39.12638	-122.43213	SRR2169872
G102.4	NA	39.12638	-122.43213	SRR2169873
G102.7	NA	39.12638	-122.43213	SRR2169874
G102.8	NA	39.12638	-122.43213	SRR2169875
G102.9	NA	39.12638	-122.43213	SRR2169876
G103.1	*PI 649888	38.7804	-122.57185	SRR2169877
G103.12	*PI 649888	38.7804	-122.57185	SRR2169878
G103.2	*PI 649888	38.7804	-122.57185	SRR2169879
G103.3	*PI 649888	38.7804	-122.57185	SRR2169880
G103.4	*PI 649888	38.7804	-122.57185	SRR2169881

Sample name	Seed accession	Latitude	Longitude	SRA number
G103.5	*PI 649888	38.7804	-122.57185	SRR2169882
G103.6	*PI 649888	38.7804	-122.57185	SRR2169883
G103.7	*PI 649888	38.7804	-122.57185	SRR2169884
G103.8	*PI 649888	38.7804	-122.57185	SRR2169885
G103.9	*PI 649888	38.7804	-122.57185	SRR2169886
G108.13	*PI 664632	38.87585	-120.8205	SRR2169887
G108.17	*PI 664632	38.87585	-120.8205	SRR2169888
G108.2	*PI 664632	38.87585	-120.8205	SRR2169889
G108.20	*PI 664632	38.87585	-120.8205	SRR2169910
G108.3	*PI 664632	38.87585	-120.8205	SRR2169890
G108.4	*PI 664632	38.87585	-120.8205	SRR2169891
G108.5	*PI 664632	38.87585	-120.8205	SRR2169892
G108.6	*PI 664632	38.87585	-120.8205	SRR2169893
G108.7	*PI 664632	38.87585	-120.8205	SRR2169894
G108.8	*PI 664632	38.87585	-120.8205	SRR2169895
G108.9	*PI 664632	38.87585	-120.8205	SRR2169896
G109.1	NA	39.17832	-121.75977	SRR2169897
G109.10	NA	39.17832	-121.75977	SRR2169898
G109.2	NA	39.17832	-121.75977	SRR2169899
G109.3	NA	39.17832	-121.75977	SRR2169900
G109.4	NA	39.17832	-121.75977	SRR2169901
G109.5	NA	39.17832	-121.75977	SRR2169902
G109.6	NA	39.17832	-121.75977	SRR2169903
G109.7	NA	39.17832	-121.75977	SRR2169904
G109.8	NA	39.17832	-121.75977	SRR2169905
G109.9	NA	39.17832	-121.75977	SRR2169906
G110.1	NA	39.25156	-121.88924	SRR2169907
G110.11	NA	39.25156	-121.88924	SRR2169908
G110.12	NA	39.25156	-121.88924	SRR2169909
G110.3	NA	39.25156	-121.88924	SRR2169911
G110.6	NA	39.25156	-121.88924	SRR2169912
G110.9	NA	39.25156	-121.88924	SRR2169913
G111.1	NA	39.34395	-121.44869	SRR2169914
G111.10	NA	39.34395	-121.44869	SRR2169915
G111.11	NA	39.34395	-121.44869	SRR2169916
G111.3	NA	39.34395	-121.44869	SRR2169917
G111.4	NA	39.34395	-121.44869	SRR2169918
G111.5	NA	39.34395	-121.44869	SRR2169919
G111.6	NA	39.34395	-121.44869	SRR2169920
G111.7	NA	39.34395	-121.44869	SRR2169921
G111.8	NA	39.34395	-121.44869	SRR2169922
G111.9	NA	39.34395	-121.44869	SRR2169923
G114.10	*PI 649896	41.28199	-122.85186	SRR2169924
G114.13	*PI 649896	41.28199	-122.85186	SRR2169925
G114.14	*PI 649896	41.28199	-122.85186	SRR2169926
G114.15	*PI 649896	41.28199	-122.85186	SRR2169927
G114.18	*PI 649896	41.28199	-122.85186	SRR2169928
G114.19	*PI 649896	41.28199	-122.85186	SRR2169929
G114.20	*PI 649896	41.28199	-122.85186	SRR2169930
G114.21	*PI 649896	41.28199	-122.85186	SRR2169931
G114.24	*PI 649896	41.28199	-122.85186	SRR2169932
G114.25	*PI 649896	41.28199	-122.85186	SRR2169933
G114.29	*PI 649896	41.28199	-122.85186	SRR2169934

Sample name	Seed accession	Latitude	Longitude	SRA number
G115.10	*PI 649895	41.64306	-122.74711	SRR2169935
G115.11	*PI 649895	41.64306	-122.74711	SRR2169936
G115.12	*PI 649895	41.64306	-122.74711	SRR2169937
G115.3	*PI 649895	41.64306	-122.74711	SRR2169938
G115.4	*PI 649895	41.64306	-122.74711	SRR2169939
G115.7	*PI 649895	41.64306	-122.74711	SRR2169940
G115.9	*PI 649895	41.64306	-122.74711	SRR2169941
G116.13	NA	39.066322	-122.478403	SRR2169942
G116.14	NA	39.066322	-122.478403	SRR2169943
G116.15	NA	39.066322	-122.478403	SRR2169944
G116.4	NA	39.066322	-122.478403	SRR2169945
G116.6	NA	39.066322	-122.478403	SRR2169946
G118.11	Ames 27232	39.2627	-122.51157	SRR2169947
G118.12	Ames 27232	39.2627	-122.51157	SRR2169948
G118.2	Ames 27232	39.2627	-122.51157	SRR2169949
G118.3	Ames 27232	39.2627	-122.51157	SRR2169950
G118.5	Ames 27232	39.2627	-122.51157	SRR2169951
G118.6	Ames 27232	39.2627	-122.51157	SRR2169952
G118.7	Ames 27232	39.2627	-122.51157	SRR2169953
G118.8	Ames 27232	39.2627	-122.51157	SRR2169954
G118.9	Ames 27232	39.2627	-122.51157	SRR2169955
G119.1	PI 649899	39.48584	-121.31271	SRR2169956
G119.2	PI 649899	39.48584	-121.31271	SRR2169957
G119.3	PI 649899	39.48584	-121.31271	SRR2169958
G119.4	PI 649899	39.48584	-121.31271	SRR2169959
G119.5	PI 649899	39.48584	-121.31271	SRR2169960
G119.6	PI 649899	39.48584	-121.31271	SRR2169961
G119.7	PI 649899	39.48584	-121.31271	SRR2169962
G119.8	PI 649899	39.48584	-121.31271	SRR2169963
G119.9	PI 649899	39.48584	-121.31271	SRR2169964
G120.10	PI 435644	38.543	-121.7383	SRR2169965
G120.11	PI 435644	38.543	-121.7383	SRR2169966
G120.12	PI 435644	38.543	-121.7383	SRR2169967
G120.15	PI 435644	38.543	-121.7383	SRR2169968
G120.17	PI 435644	38.543	-121.7383	SRR2169969
G120.2	PI 435644	38.543	-121.7383	SRR2169970
G120.7	PI 435644	38.543	-121.7383	SRR2169971
G120.8	PI 435644	38.543	-121.7383	SRR2169972
G121.1	PI 468662	38.82395	-122.33725	SRR2169973
G121.10	PI 468662	38.82395	-122.33725	SRR2169974
G121.2	PI 468662	38.82395	-122.33725	SRR2169975
G121.3	PI 468662	38.82395	-122.33725	SRR2169976
G121.4	PI 468662	38.82395	-122.33725	SRR2169977
G121.5	PI 468662	38.82395	-122.33725	SRR2169978
G121.6	PI 468662	38.82395	-122.33725	SRR2169979
G121.7	PI 468662	38.82395	-122.33725	SRR2169980
G121.8	PI 468662	38.82395	-122.33725	SRR2169981
G121.9	PI 468662	38.82395	-122.33725	SRR2169982
G122.1	PI 649889	38.73309	-122.52462	SRR2169983
G122.11	PI 649889	38.73309	-122.52462	SRR2169984
G122.2	PI 649889	38.73309	-122.52462	SRR2169985
G122.3	PI 649889	38.73309	-122.52462	SRR2169986
G122.5	PI 649889	38.73309	-122.52462	SRR2169987

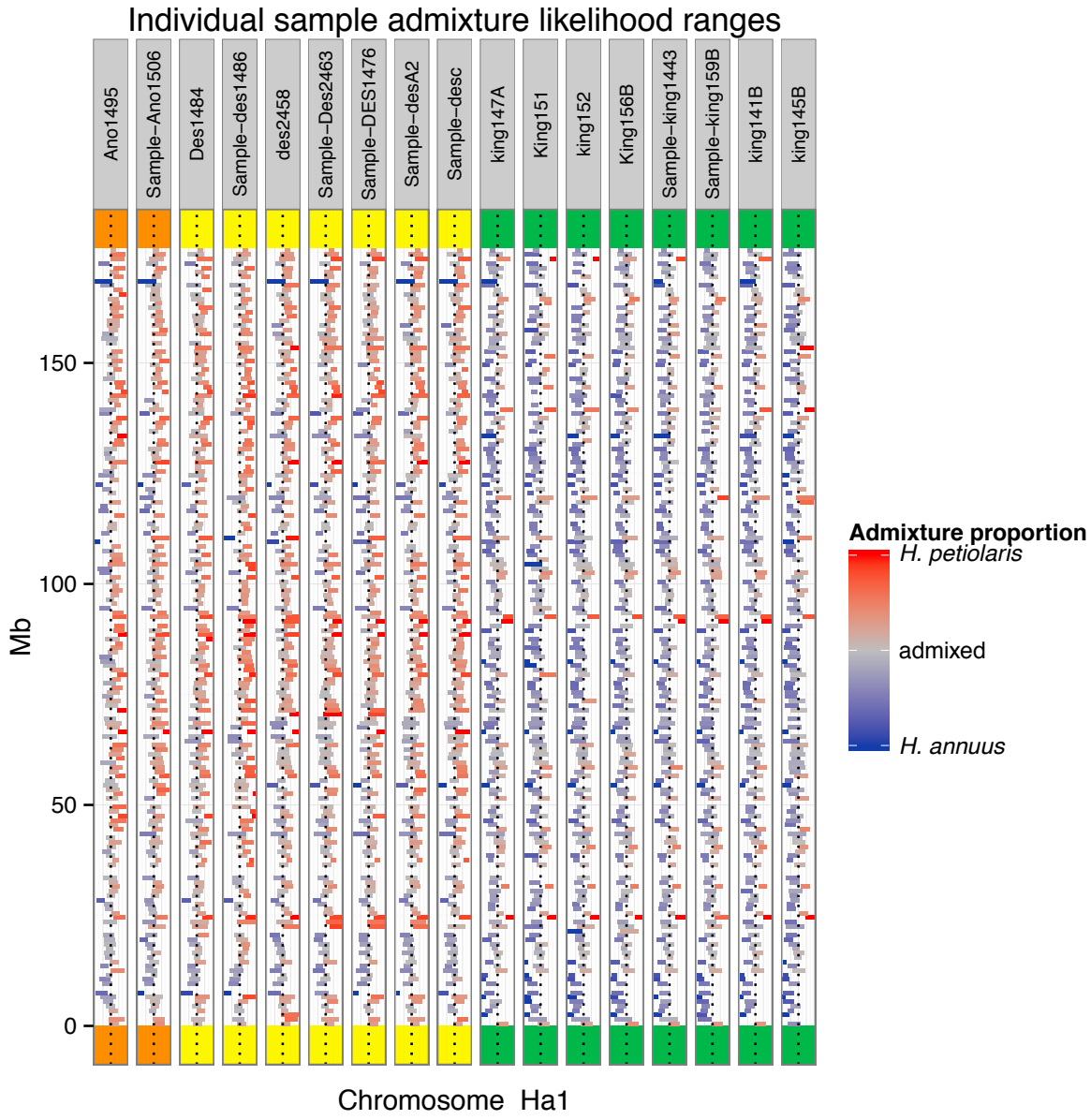
Sample name	Seed accession	Latitude	Longitude	SRA number
G122.6	PI 649889	38.73309	-122.52462	SRR2169988
G122.7	PI 649889	38.73309	-122.52462	SRR2169989
G122.8	PI 649889	38.73309	-122.52462	SRR2169990
G123.12	PI 649898	39.83434	-121.58227	SRR2169991
G123.13	PI 649898	39.83434	-121.58227	SRR2169992
G123.15	PI 649898	39.83434	-121.58227	SRR2169993
G123.17	PI 649898	39.83434	-121.58227	SRR2169994
G123.2	PI 649898	39.83434	-121.58227	SRR2169995
G123.4	PI 649898	39.83434	-121.58227	SRR2169996
G123.5	PI 649898	39.83434	-121.58227	SRR2169997
G123.6	PI 649898	39.83434	-121.58227	SRR2169998
G123.7	PI 649898	39.83434	-121.58227	SRR2169999
G123.8	PI 649898	39.83434	-121.58227	SRR2170000
G124.1	PI 649900	38.84119	-120.87647	SRR2170001
G124.10	PI 649900	38.84119	-120.87647	SRR2170002
G124.11	PI 649900	38.84119	-120.87647	SRR2170003
G124.12	PI 649900	38.84119	-120.87647	SRR2170004
G124.2	PI 649900	38.84119	-120.87647	SRR2170005
G124.3	PI 649900	38.84119	-120.87647	SRR2170006
G124.5	PI 649900	38.84119	-120.87647	SRR2170007
G124.6	PI 649900	38.84119	-120.87647	SRR2170008
G124.7	PI 649900	38.84119	-120.87647	SRR2170009
G124.9	PI 649900	38.84119	-120.87647	SRR2170010
G127.1	PI 649901	37.84557	-120.46388	SRR2170011
G127.10	PI 649901	37.84557	-120.46388	SRR2170012
G127.2	PI 649901	37.84557	-120.46388	SRR2170013
G127.3	PI 649901	37.84557	-120.46388	SRR2170014
G127.4	PI 649901	37.84557	-120.46388	SRR2170015
G127.5	PI 649901	37.84557	-120.46388	SRR2170016
G127.6	PI 649901	37.84557	-120.46388	SRR2170017
G127.7	PI 649901	37.84557	-120.46388	SRR2170018
G127.8	PI 649901	37.84557	-120.46388	SRR2170019
G127.9	PI 649901	37.84557	-120.46388	SRR2170020
G128.1	PI 649897	41.03086	-122.42451	SRR2170021
G128.2	PI 649897	41.03086	-122.42451	SRR2170022
G128.3	PI 649897	41.03086	-122.42451	SRR2170023
G128.4	PI 649897	41.03086	-122.42451	SRR2170024
G129.11	PI 664630	39.88756	-122.63451	SRR2170025
G129.4	PI 664630	39.88756	-122.63451	SRR2170026
G129.5	PI 664630	39.88756	-122.63451	SRR2170027
G129.6	PI 664630	39.88756	-122.63451	SRR2170028
G129.8	PI 664630	39.88756	-122.63451	SRR2170029
G129.9	PI 664630	39.88756	-122.63451	SRR2170030
G130.1	PI 649894	41.29794	-122.72187	SRR2170031
G130.10	PI 649894	41.29794	-122.72187	SRR2170032
G130.2	PI 649894	41.29794	-122.72187	SRR2170033
G130.3	PI 649894	41.29794	-122.72187	SRR2170034
G130.4	PI 649894	41.29794	-122.72187	SRR2170035
G130.5	PI 649894	41.29794	-122.72187	SRR2170036
G130.6	PI 649894	41.29794	-122.72187	SRR2170037
G130.7	PI 649894	41.29794	-122.72187	SRR2170038
G130.8	PI 649894	41.29794	-122.72187	SRR2170039
G130.9	PI 649894	41.29794	-122.72187	SRR2170040

Sample name	Seed accession	Latitude	Longitude	SRA number
GB001	King 140-38	NA	NA	SRR2169810
GB002	King 140-32	NA	NA	SRR2169811
GB003	King 140-32	NA	NA	SRR2169812
GB011	PI 613783	41.352778	-94.092222	SRR2169560
GB013	IAF 54-46	NA	NA	SRR2169561
GB014	PI 649867	36.331667	-118.353333	SRR2169657
GB015	PI 592317	50.355	-104.466389	SRR2169562
GB016	PI 613727	36.401111	-92.262222	SRR2169563
GB020	PI 468556	33.511389	-104.535556	SRR2169564
GB025	PI 413021	41.786111	-103.735833	SRR2169565
GB026	PI 649869	36.453889	-118.364722	SRR2169658
GB027	PI 649868	36.301667	-118.231667	SRR2169659
GB028	PI 649867	36.331667	-118.353333	SRR2169660
GB029	PI 586809	47.471111	-99.363333	SRR2169566
GB031	PI 613752	35.960556	-82.079167	SRR2169567
GB032	PI 468580	33.039722	-114.374444	SRR2169568
GB034	PI 547167	39.816667	-88.35	SRR2169569
GB035	PI 435612	35.733056	-80.658611	SRR2169570
GB036	PI 413130	34.678611	-120.227778	SRR2169571
GB037	PI 592318	50.163056	-104.558611	SRR2169573
GB041	PI 435368	34.256389	-98.483611	SRR2169574
GB042	PI 613737	36.300833	-118.218056	SRR2169575
GB043	PI 435406	36.866667	-99.133333	SRR2169576
GB044	PI 435410	33.995	-97.175	SRR2169577
GB047	PI 468615	36.213333	-107.290833	SRR2169578
GB048	PI 435589	38.525	-120.030278	SRR2169579
GB049	PI 468571	33.138611	-109.875556	SRR2169580
GB050	PI 468545	34.535833	-102.909722	SRR2169582
GB051	PI 435531	33.605278	-100.208333	SRR2169583
GB052	PI 468476	31.272778	-101.307778	SRR2169585
GB053	PI 468463	29.808333	-100.441667	SRR2169586
GB054	PI 413157	32.187222	-107.666667	SRR2169587
GB062	PI 468747	NA	NA	SRR2169762
GB063	PI 592333	49.709167	-98.037778	SRR2169763
GB064	PI 650010	46.65	-96.766667	SRR2169764
GB065	PI 613794	42.451111	-95.805833	SRR2169765
GB098	PI 649814	33.493611	-111.063611	SRR2169590
GB099	PI 435471	36.342778	-103.1	SRR2169591
GB100	PI 592312	49.961667	-106.243056	SRR2169592
GB101	PI 586887	45.5	-97.883333	SRR2169593
GB102	PI 435414	33.6025	-94.571944	SRR2169594
GB103	PI 435850	27.586111	-96.546944	SRR2169595
GB104	PI 468562	33.332778	-107.934722	SRR2169596
GB105	PI 435598	36.111111	-110.766111	SRR2169597
GB106	PI 435557	38.4825	-99.093333	SRR2169598
GB107	PI 586864	39.85	-94.166667	SRR2169599
GB110	PI 468613	35.793333	-109.495556	SRR2169600
GB111	PI 468616	36.785556	-107.313611	SRR2169601
GB113	PI 649854	43.066667	-95.50	SRR2169602
GB114	PI 413173	42.928333	-99.248056	SRR2169603
GB115	PI 613749	45.011667	-98.044722	SRR2169604
GB116	PI 435359	32.448611	-98.267222	SRR2169605
GB117	PI 653547	33.901944	-105.131389	SRR2169606

Sample name	Seed accession	Latitude	Longitude	SRA number
GB118	PI 586879	42.916667	-99.80	SRR2169607
GB119	PI 413097	35.0525	-117.826944	SRR2169608
GB120	PI 413103	38.286944	-120.3225	SRR2169609
GB121	PI 413131	34.678611	-120.227778	SRR2169610
GB122	PI 413155	32.252778	-108.168611	SRR2169611
GB123	PI 413080	32.815	-114.627222	SRR2169612
GB124	PI 413079	32.815	-114.627222	SRR2169613
GB125	PI 413095	35.002222	-116.3525	SRR2169614
GB126	PI 468542	35.078889	-101.6	SRR2169615
GB127	PI 413120	37.957778	-120.710278	SRR2169616
GB128	PI 435456	35.288056	-101.938611	SRR2169617
GB129	PI 586853	38.466667	-99.516667	SRR2169618
GB130	PI 586860	39.166667	-94.983333	SRR2169619
GB131	PI 586818	46.333333	-104.166667	SRR2169620
GB132	PI 586819	45.833333	-104.333333	SRR2169621
GB133	PI 613787	40.81	-94.196111	SRR2169622
GB134	PI 435442	29.702778	-100.65	SRR2169623
GB135	PI 435448	30.863333	-101.126944	SRR2169624
GB142	PI 613757	51.533611	-99.991944	SRR2169766
GB143	PI 531041	47.00	-107.65	SRR2169767
GB146	PI 531041	47.00	-107.65	SRR2169768
GB169	PI 435457	35.370278	-101.906111	SRR2169625
GB170	PI 468494	33.785	-96.266667	SRR2169626
GB171	PI 468456	27.635833	-98.516667	SRR2169627
GB172	PI 468512	28.455	-95.112222	SRR2169628
GB173	PI 597901	43.083333	-95.816667	SRR2169629
GB174	PI 435841	35.8125	-109.805556	SRR2169630
GB175	PI 468457	28.033333	-98.65	SRR2169631
GB176	PI 597890	43.05	-96.50	SRR2169632
GB177	PI 468596	39.608056	-118.749167	SRR2169633
GB178	PI 435534	31.845556	-101.632778	SRR2169634
GB182	PI 435397	NA	NA	SRR2169635
GB183	PI 468548	32.856111	-102.237778	SRR2169636
GB184	PI 468583	33.733333	-116.833333	SRR2169637
GB185	PI 468536	31.40	-101.129167	SRR2169638
GB186	PI 432524	35.138611	-106.621667	SRR2169639
GB187	PI 649806	41.417778	-103.902222	SRR2169640
GB188	PI 435598	36.111111	-110.766111	SRR2169641
GB189	PI 413088	36.815278	-118.008056	SRR2169642
GB190	PI 413088	36.815278	-118.008056	SRR2169643
GB191	PI 413088	36.815278	-118.008056	SRR2169644
GB192	PI 413079	32.815	-114.627222	SRR2169645
GB193	PI 413079	32.815	-114.627222	SRR2169646
GB194	PI 413088	36.815278	-118.008056	SRR2169647
GB195	PI 413088	36.815278	-118.008056	SRR2169648
GB198	PI 435442	29.702778	-100.65	SRR2169649
GB199	PI 586853	38.466667	-99.516667	SRR2169650
GB200	PI 586853	38.466667	-99.516667	SRR2169651
GB201	PI 435442	29.702778	-100.65	SRR2169652
GB202	PI 435442	29.702778	-100.65	SRR2169653
GB204	PI 468542	35.078889	-101.6	SRR2169581
GB205	PI 468580	33.039722	-114.374444	SRR2169655
GB206	PI 468580	33.039722	-114.374444	SRR2169656

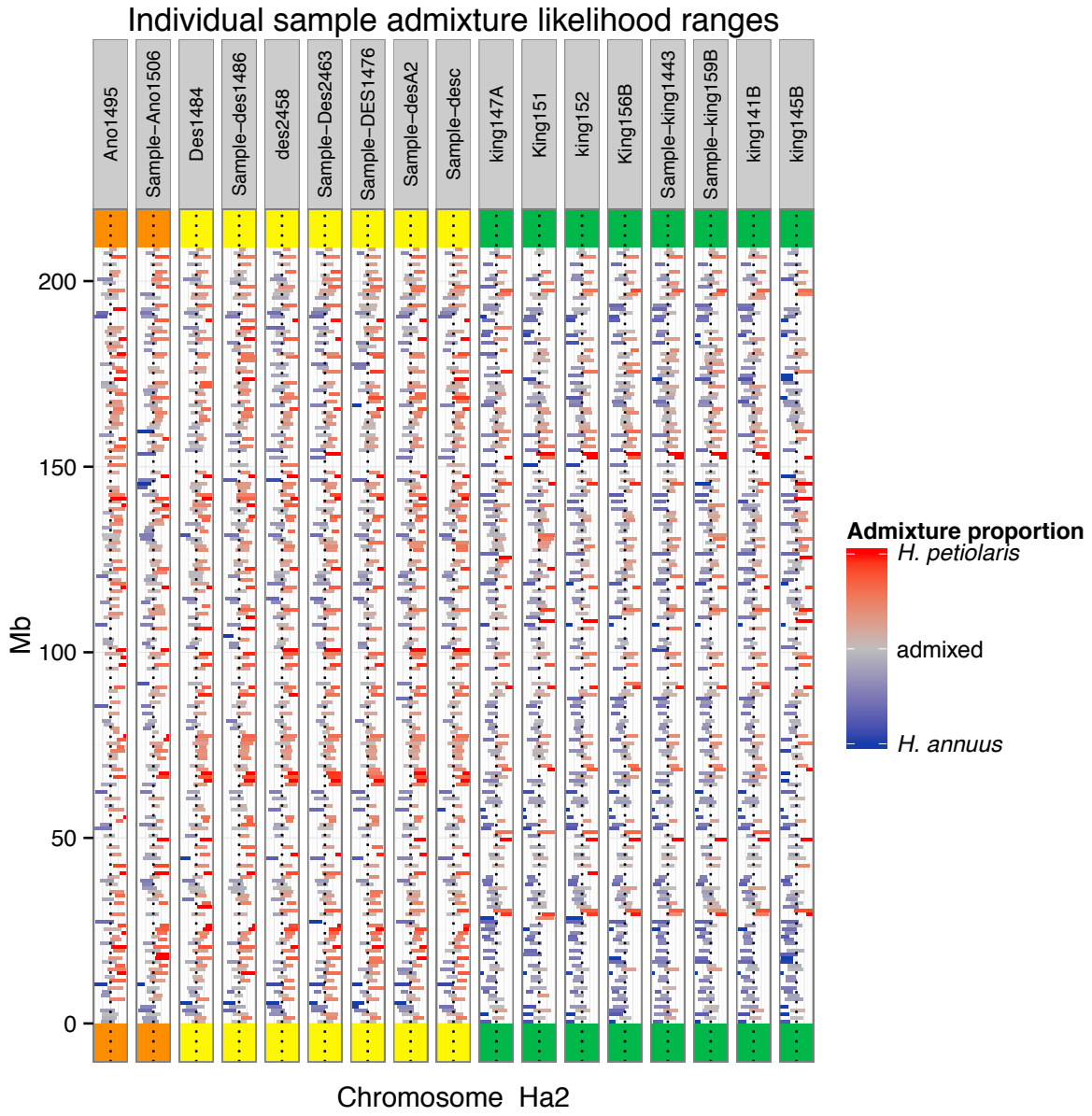
Sample name	Seed accession	Latitude	Longitude	SRA number
GB225	PI 435400	38.678611	-120.227778	SRR2169584
GB249	PI 649869	36.453889	-118.364722	SRR2169661
GB250	PI 649869	36.453889	-118.364722	SRR2169588
GB255	PI 435483	35.209444	-101.80	SRR2169589
GB277	PI 531041	47.00	-107.65	SRR2169772
GB278	PI 531041	47.00	-107.65	SRR2169769
GB282	PI 531041	47.00	-107.65	SRR2169770

Appendix C Supplementary information for chapter 4



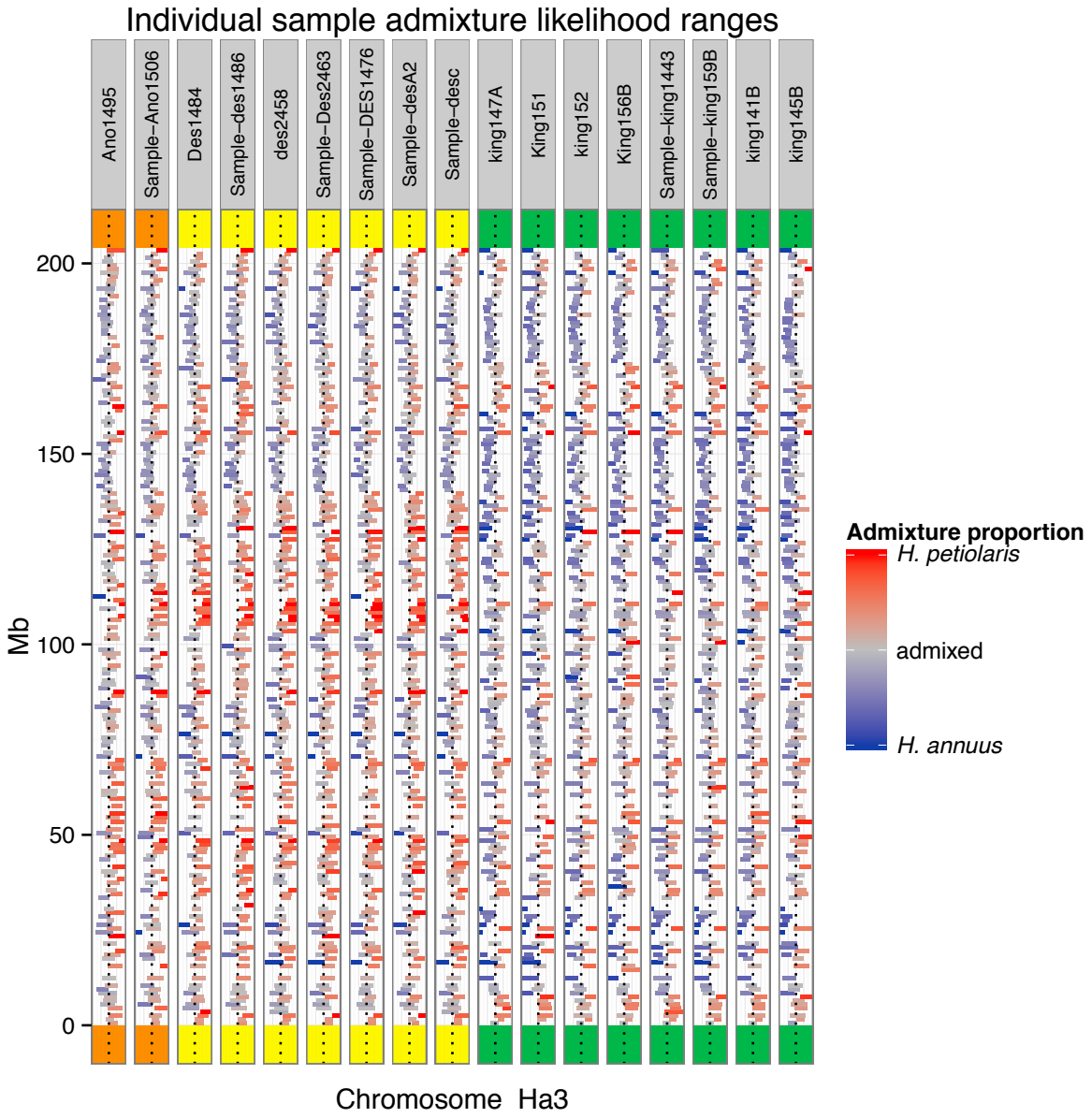
C.1 Genomic composition for individual samples (Ha1).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



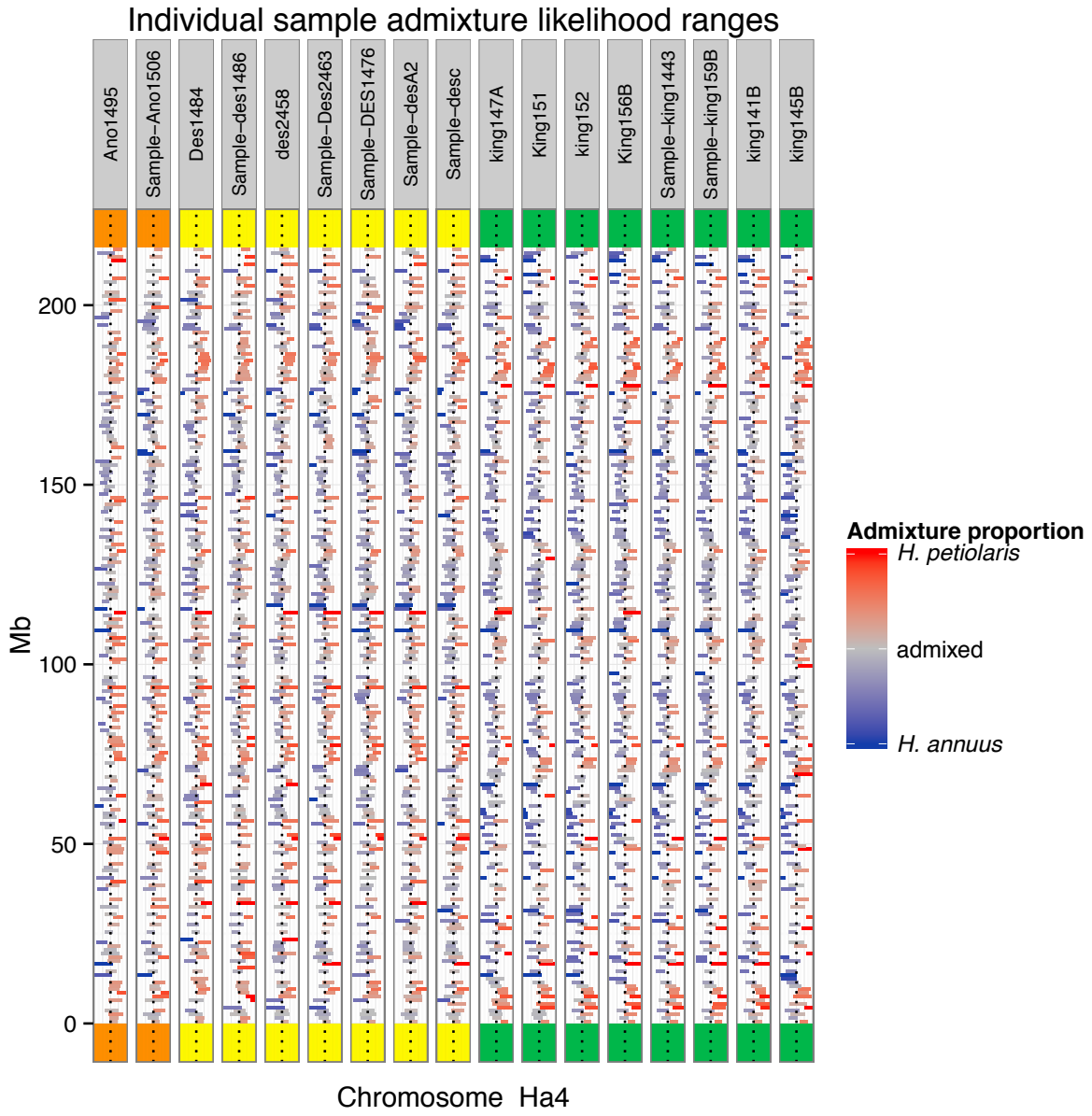
C.2 Genomic composition for individual samples (Ha2).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



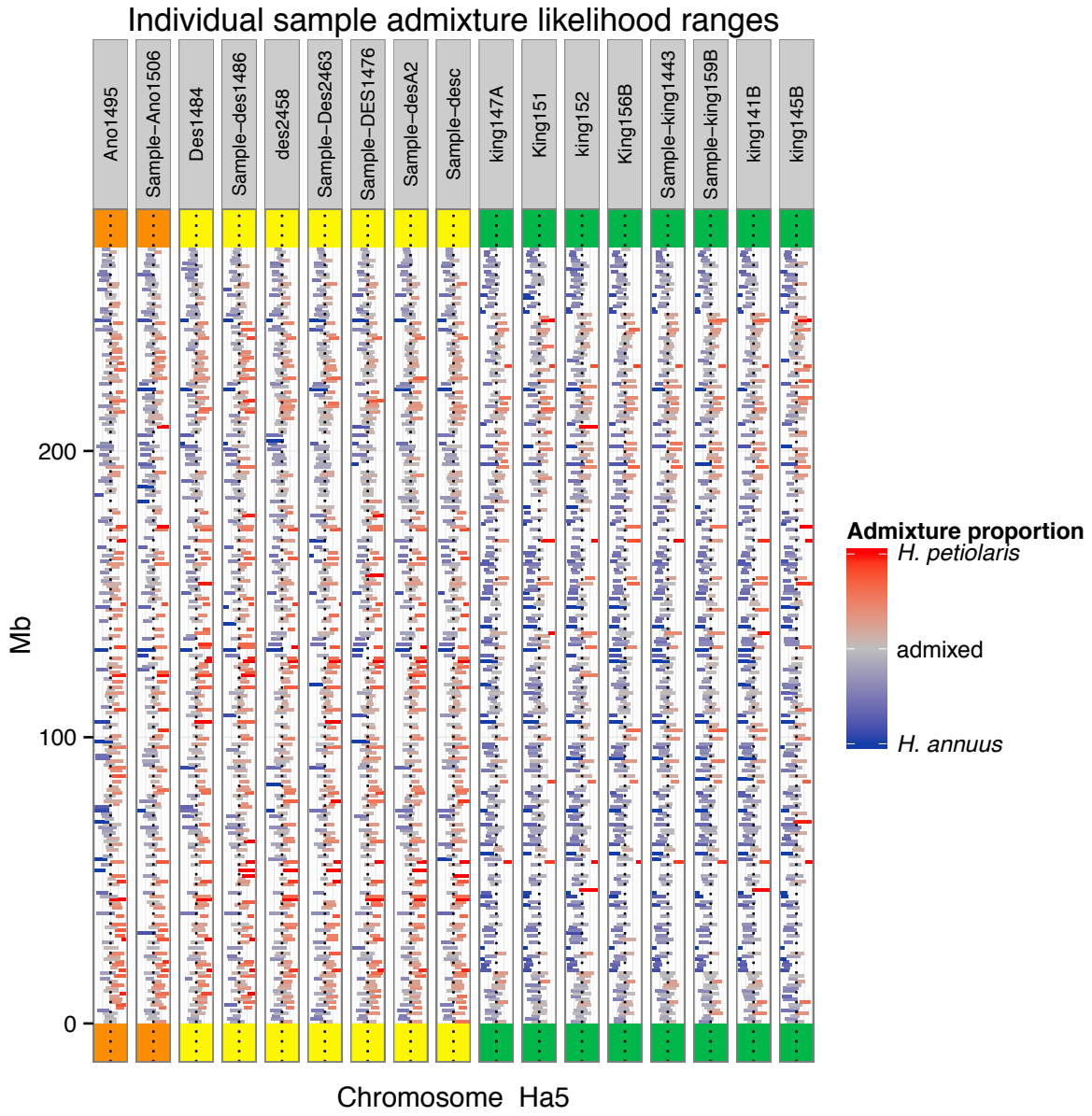
C.3 Genomic composition for individual samples (Ha3).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



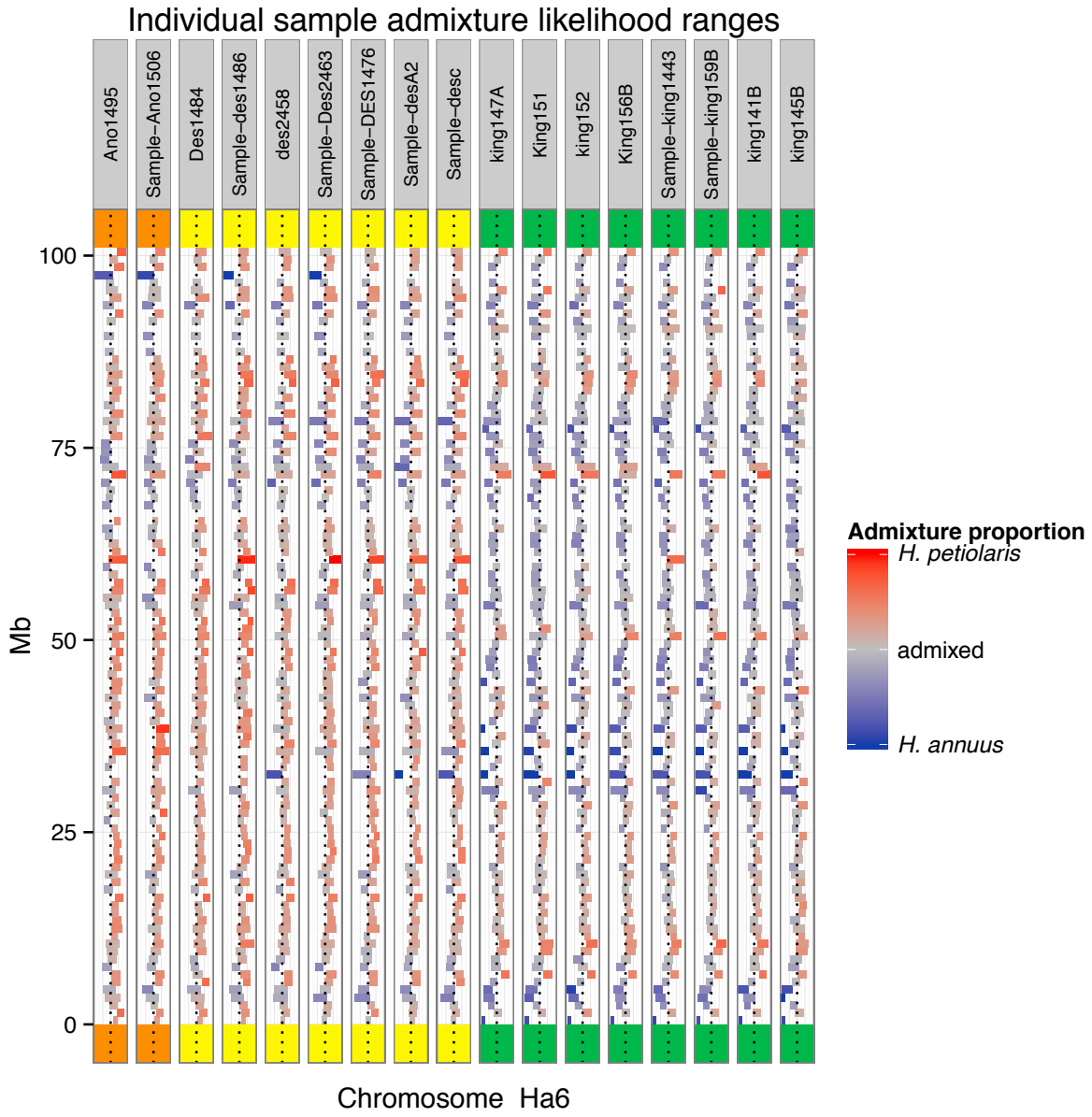
C.4 Genomic composition for individual samples (Ha4).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



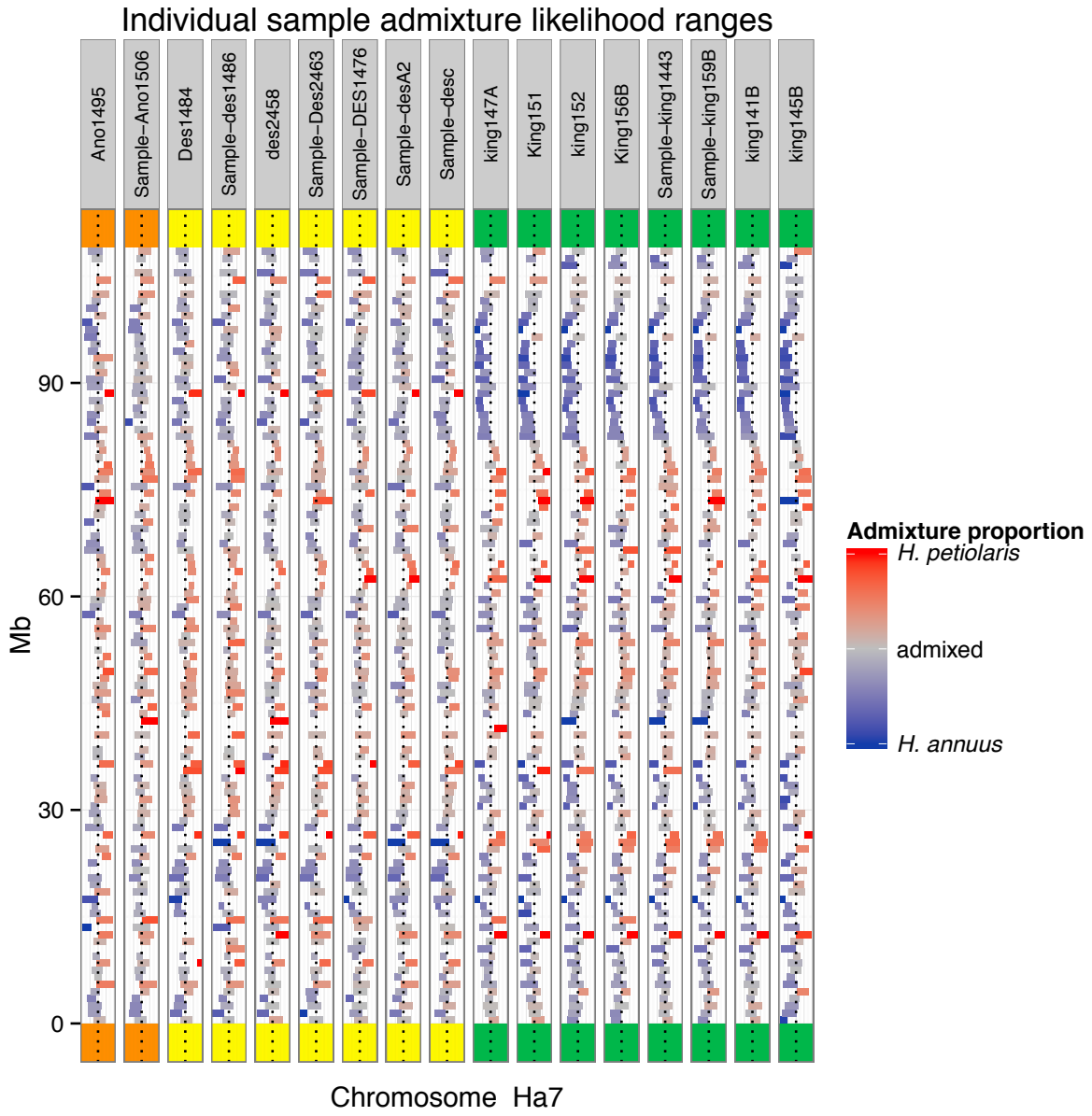
C.5 Genomic composition for individual samples (Ha5).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



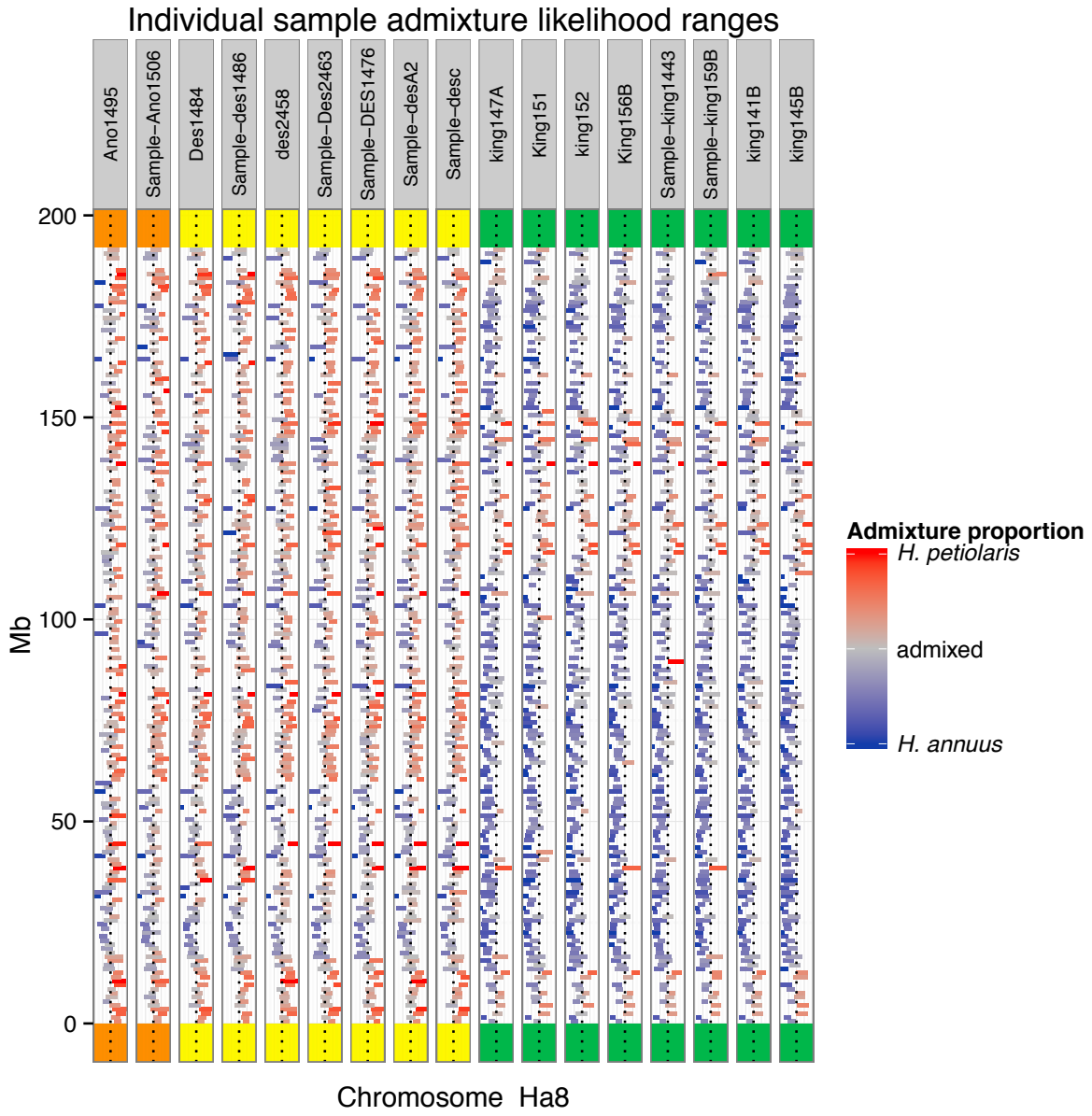
C.6 Genomic composition for individual samples (Ha6).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



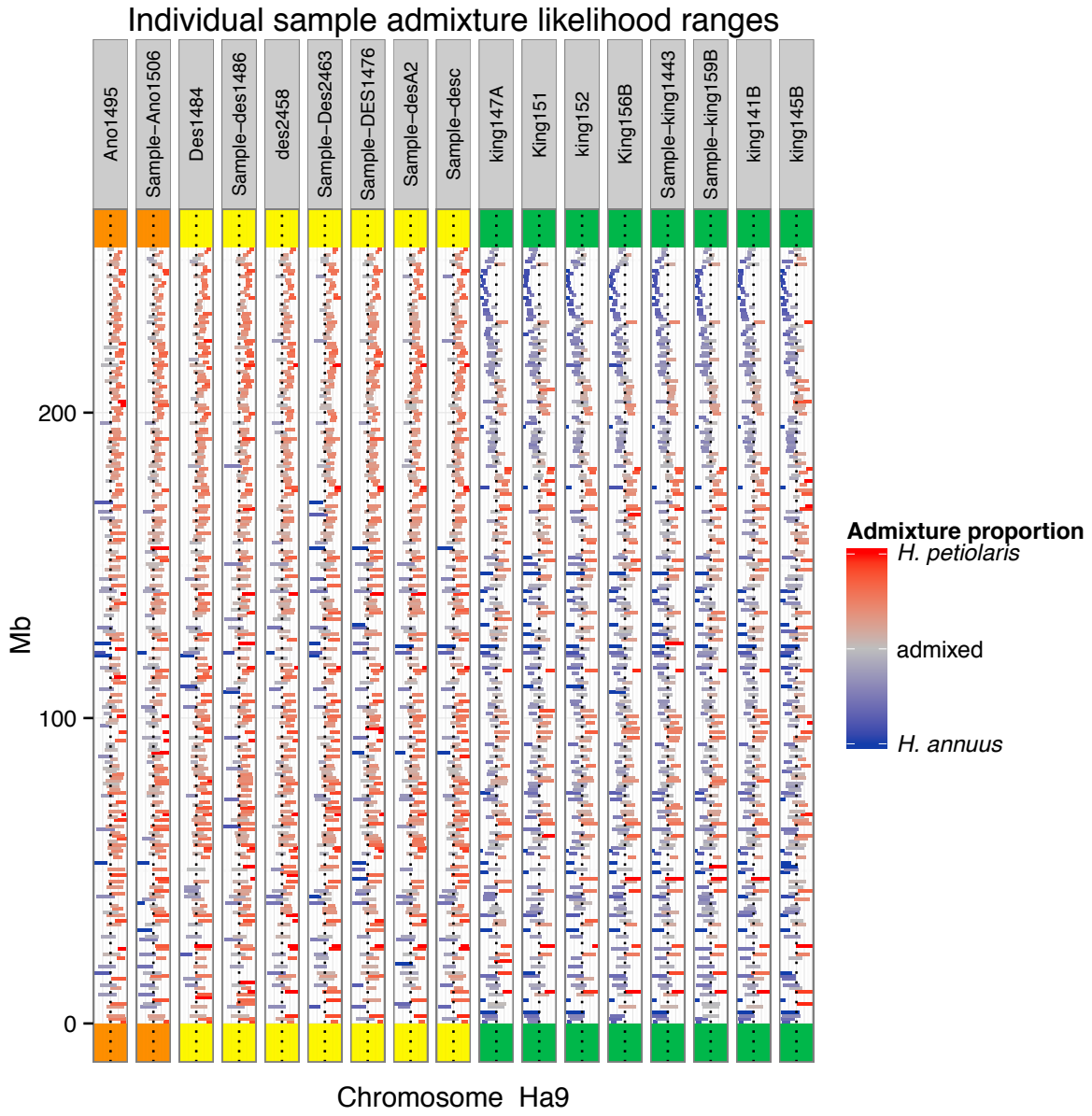
C.7 Genomic composition for individual samples (Ha7).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



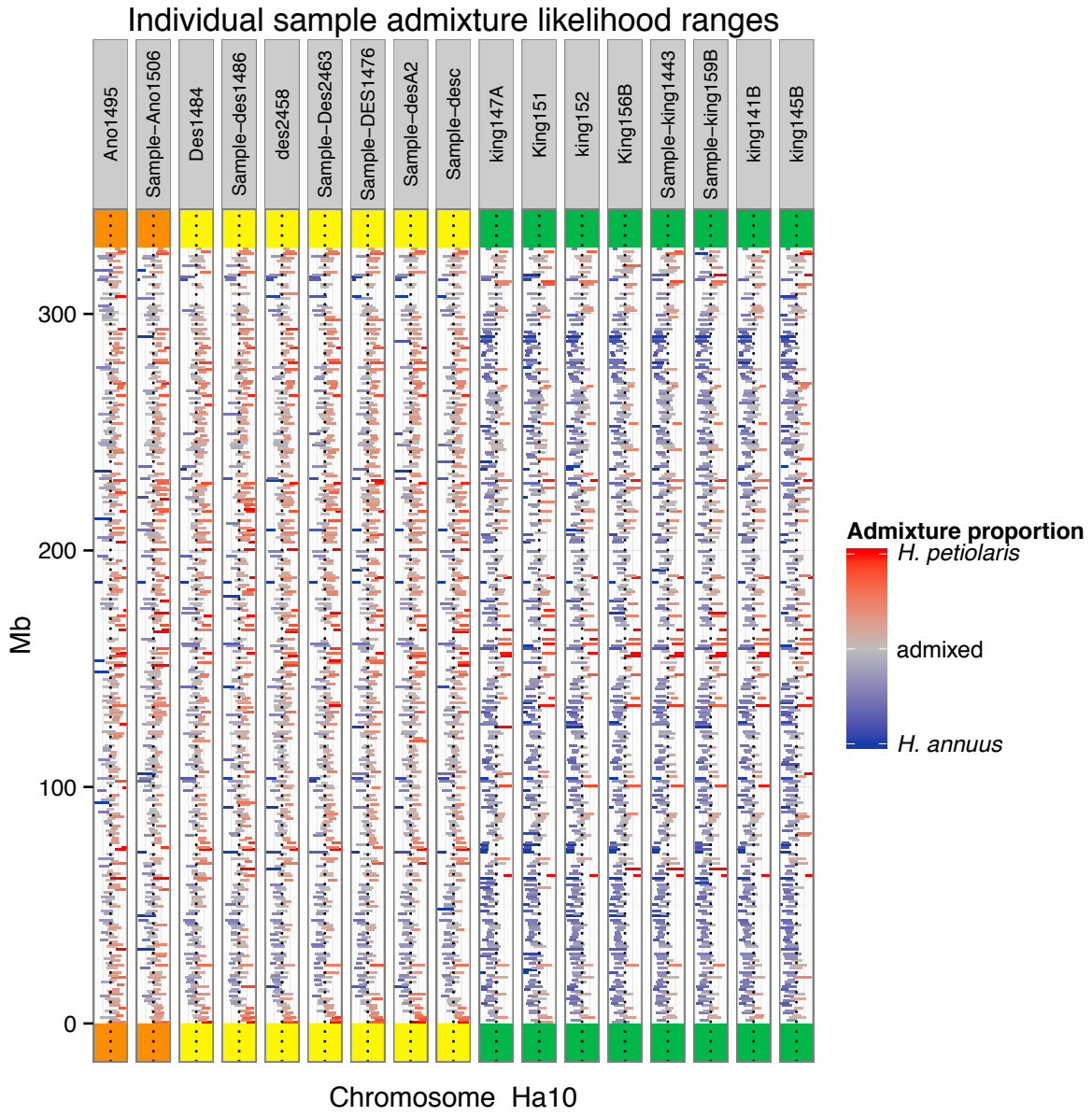
C.8 Genomic composition for individual samples (Ha8).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



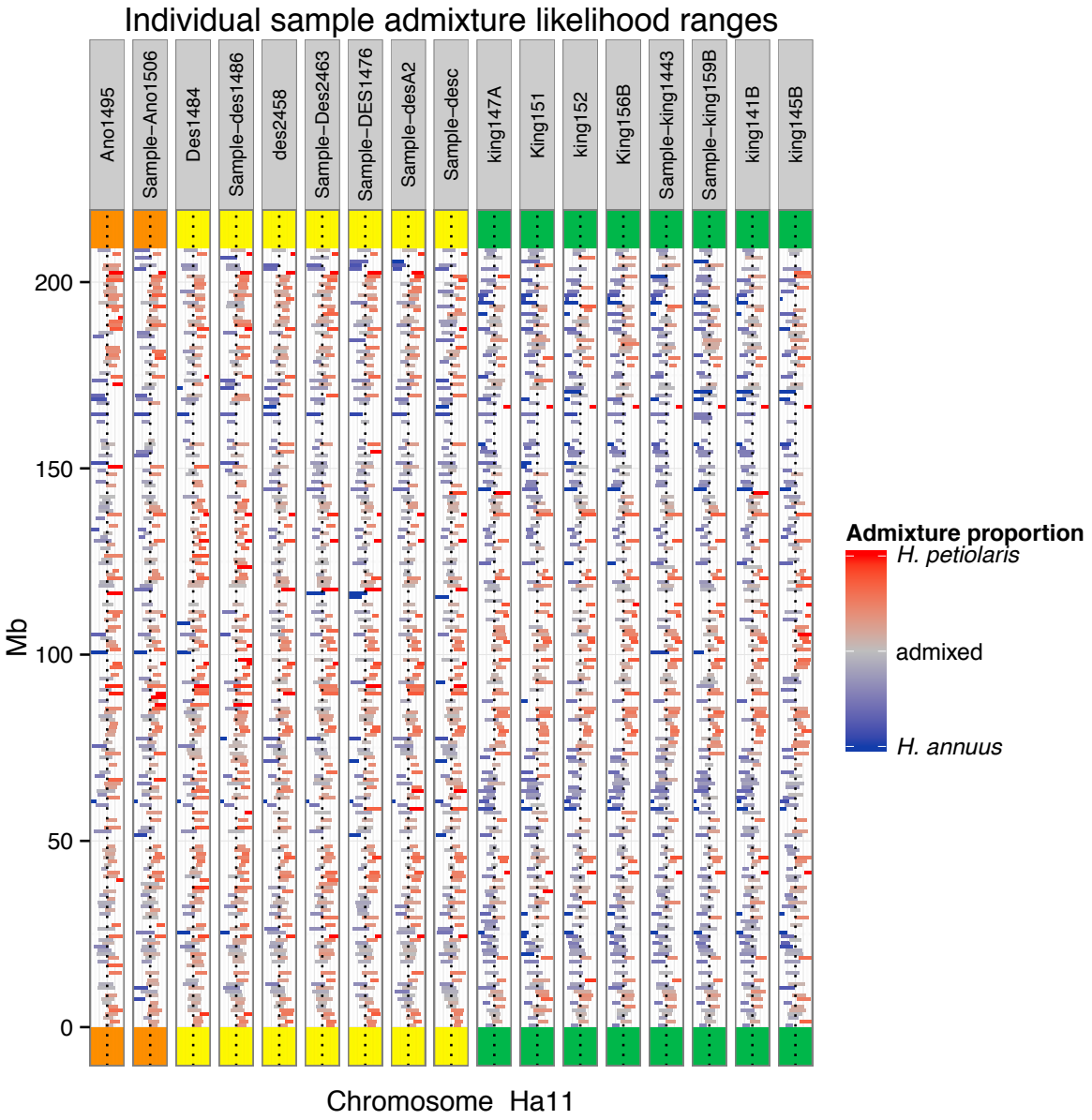
C.9 Genomic composition for individual samples (Ha9).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



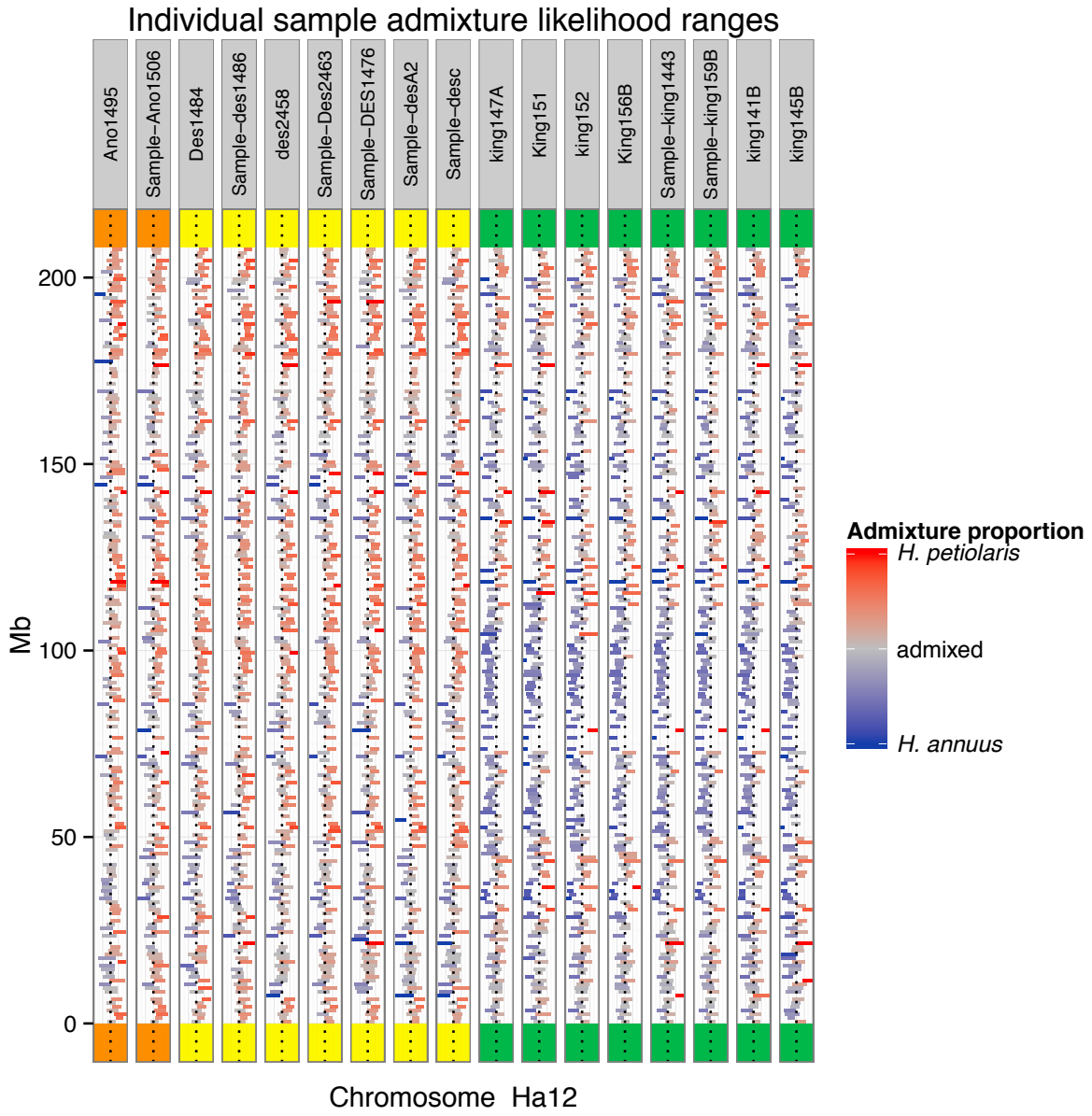
C.10 Genomic composition for individual samples (Ha10).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



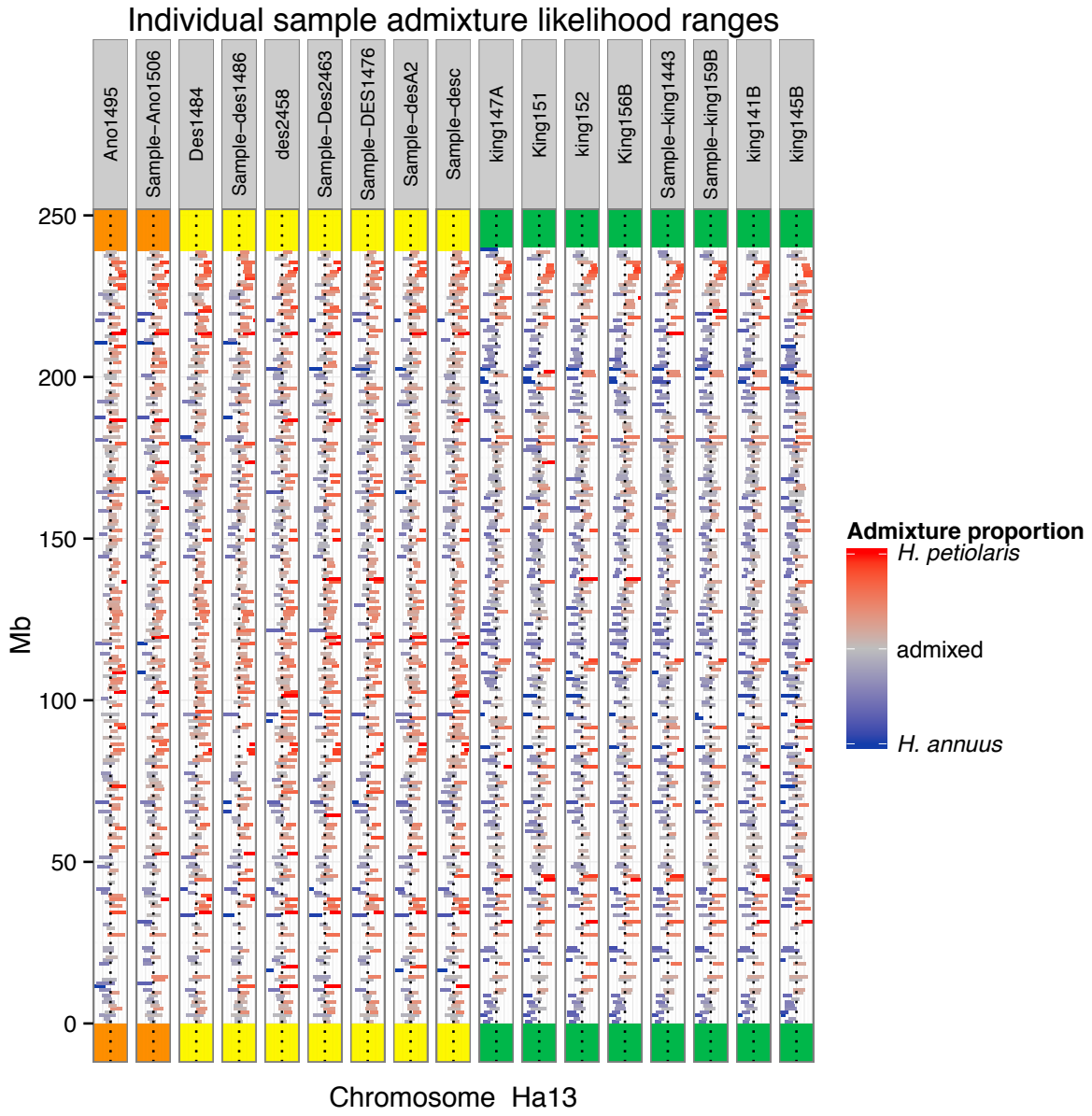
C.11 Genomic composition for individual samples (Ha11).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



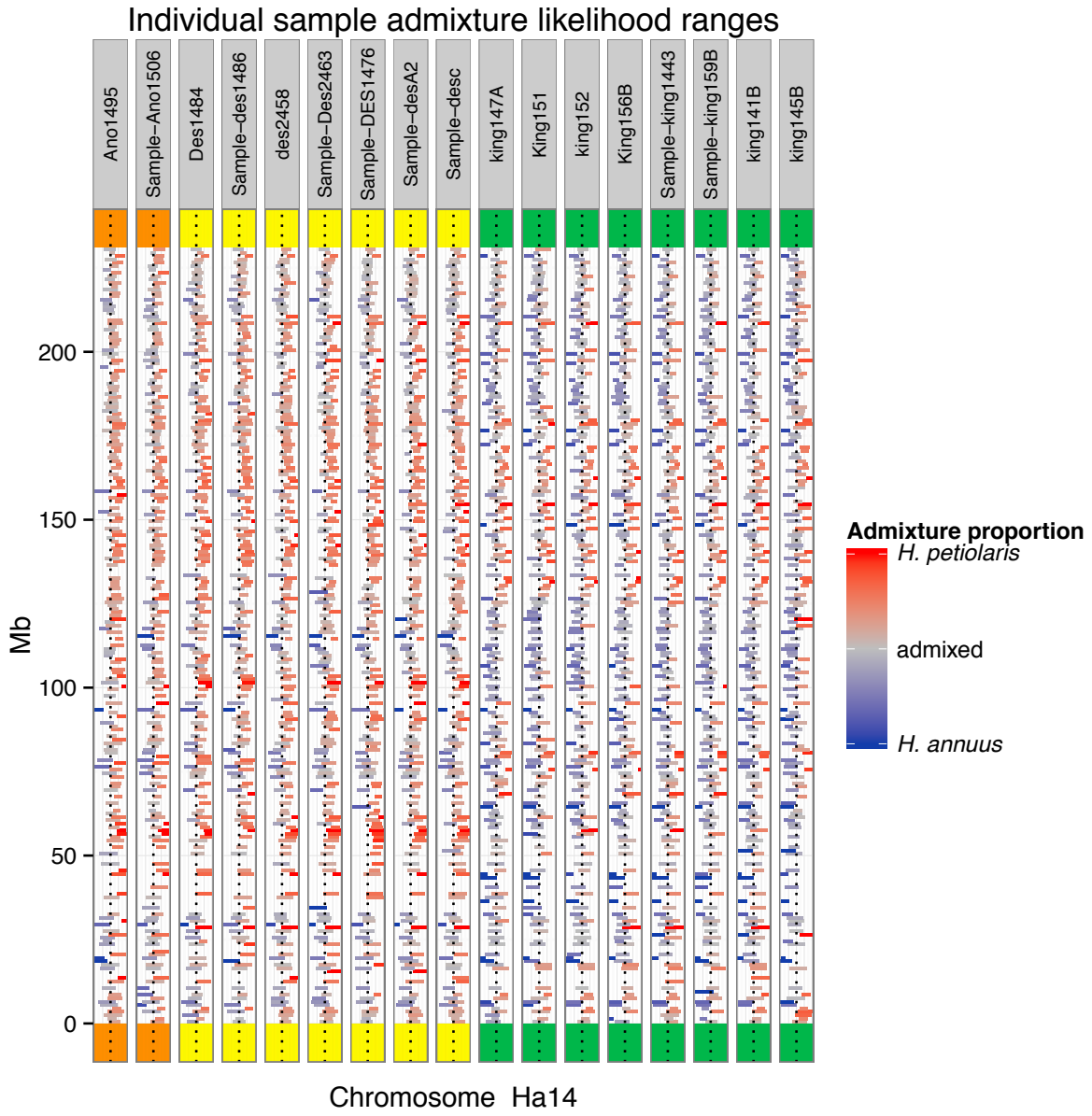
C.12 Genomic composition for individual samples (Ha12).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



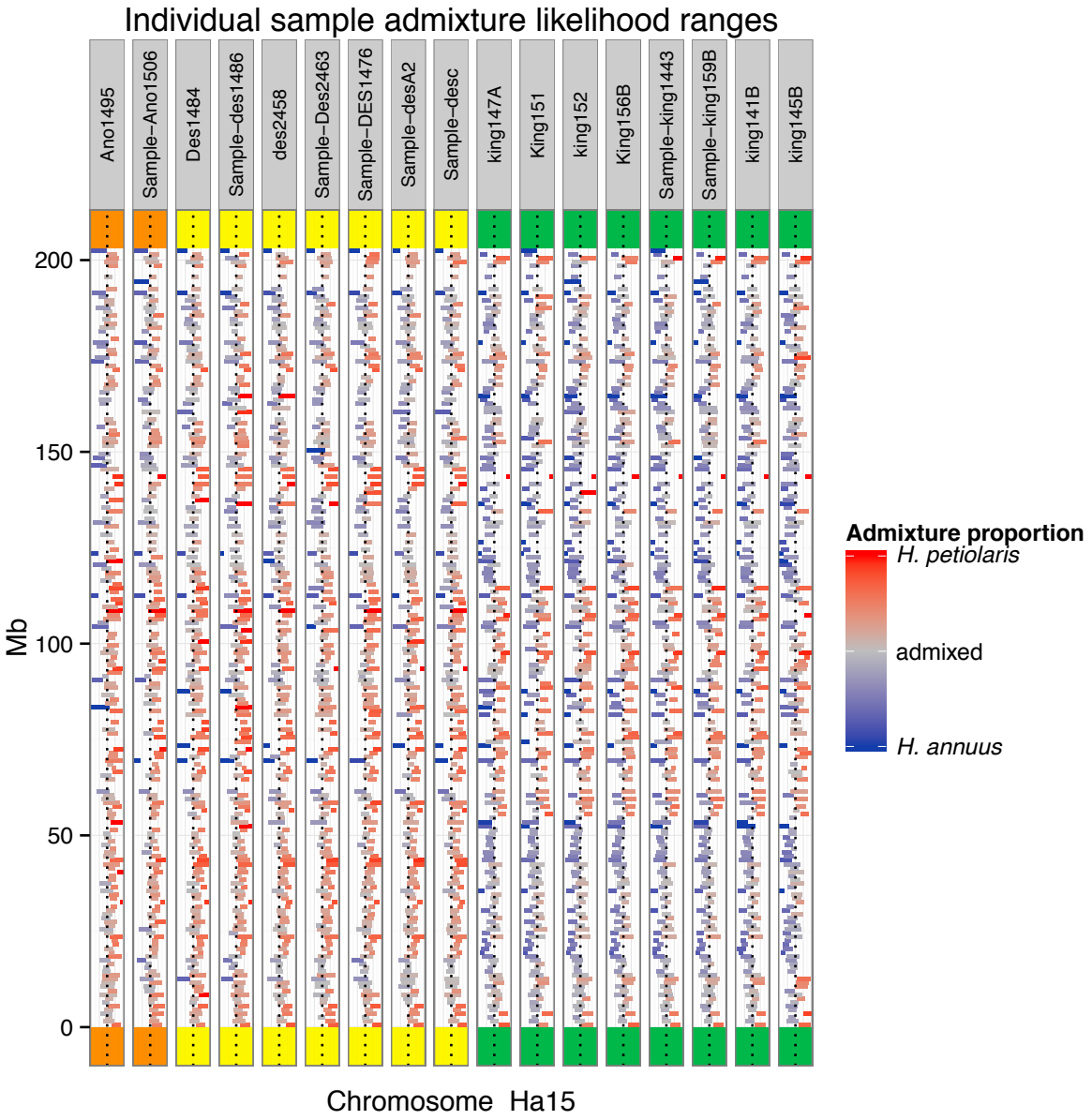
C.13 Genomic composition for individual samples (Ha13).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



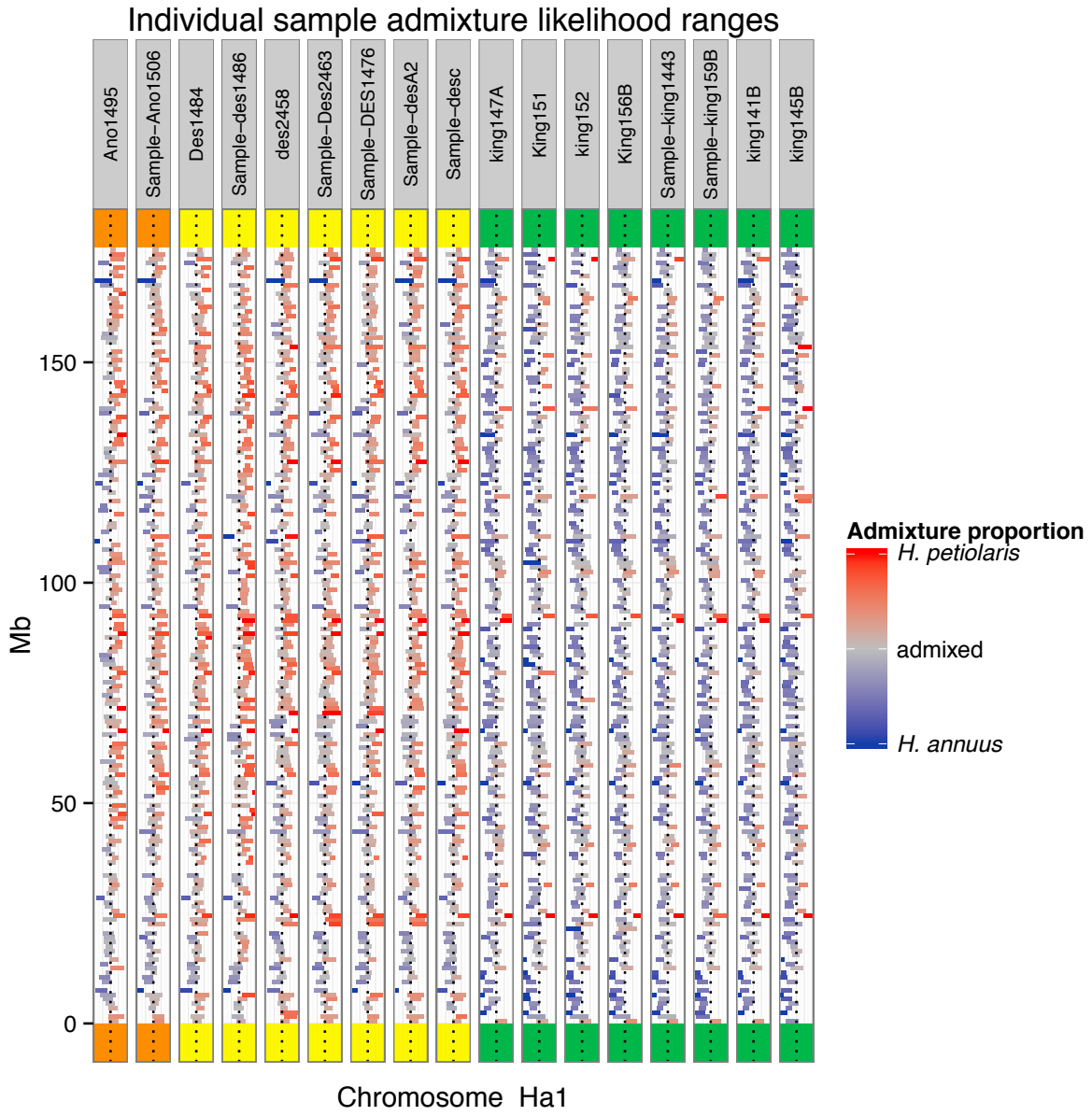
C.14 Genomic composition for individual samples (Ha14).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



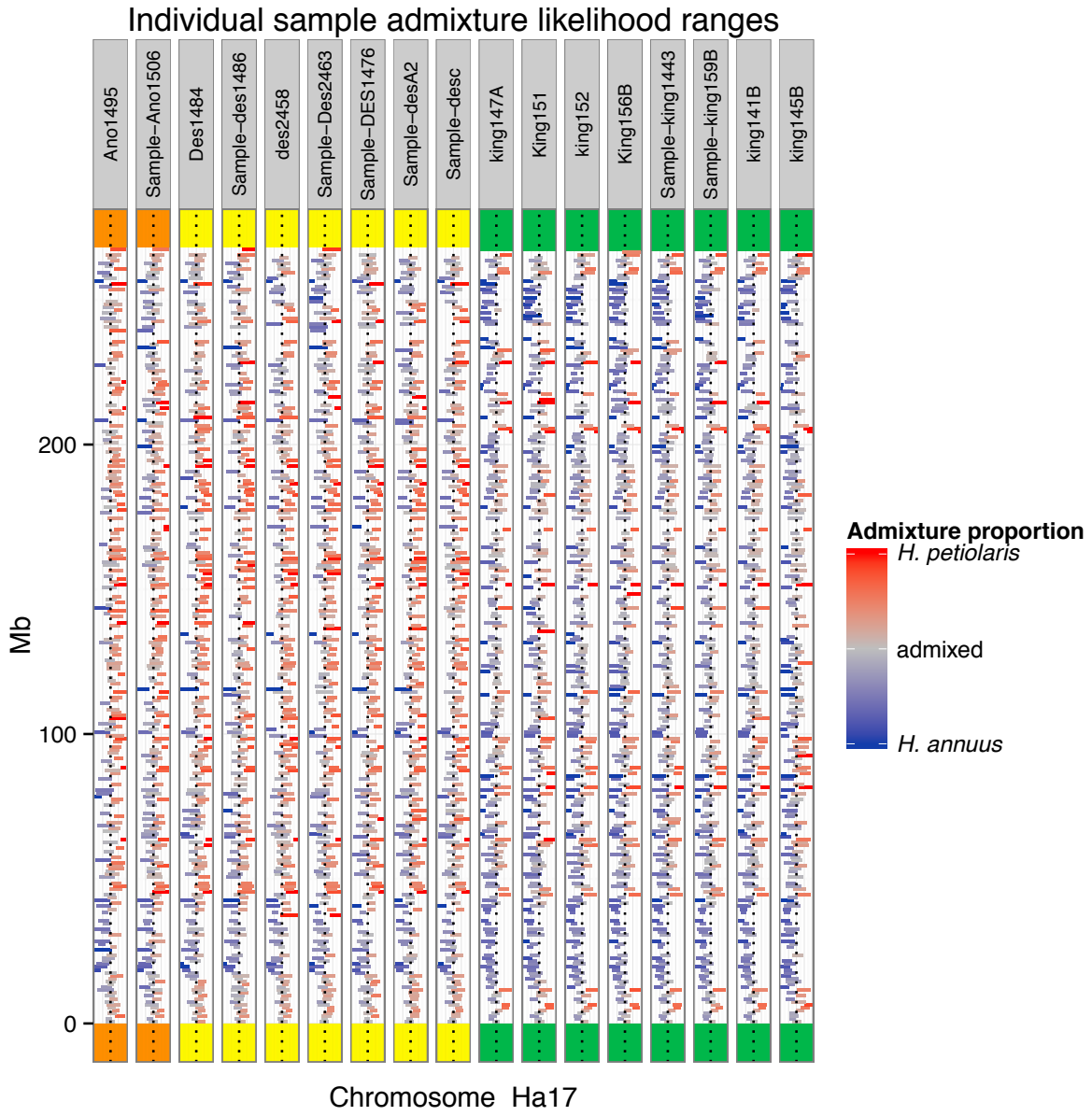
C.15 Genomic composition for individual samples (Ha15).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



C.16 Genomic composition for individual samples (Ha16).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.



C.17 Genomic composition for individual samples (Ha17).

Orange, yellow and green highlights signify *H. anomalus*, *H. deserticola* and *H. paradoxus* respectively. The bar represents the total confidence interval and the color indicates the maximum likelihood value.