
# Hybridization and Wild Tomato

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
**Thomas Ian Beddows**
aus Phoenix, AZ

Düsseldorf, August 2017

aus dem Institut für Populationsgenetik
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Laura Rose

2. Prof. Dr. Martin Lercher

Tag der mündlichen Prüfung:

**Selbständigkeitserklärung**

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation habe ich in dieser oder ähnlicher Form noch bei keiner anderen Institution vorgeleget. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

**Statement of authorship**

I declare under oath that I have compiled my dissertation independently and without any undue assistance by third parties under consideration of the ‘Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf’”. No other person’s work has been used without due acknowledgement. The dissertation has not been submitted in the same or similar format to any other institution. I have not previously failed a doctoral examination procedure.

Düsseldorf, August 2017 _____

Thomas Ian Beddows

Decisive influence is wielded not by the orthodox majority, but by a significant, progressive minority

G. L. Stebbins, 1959

# Zusammenfassung

Die vorliegende Arbeit umfasst zwei Erstautormanuskripte und eine Einleitung in das Thema der Pflanzenhybride. Hybridisierung kann eine die Evolution bestimmende Kraft, bis hin zur Entstehung neuer Arten, sein. Ihr Einfluss auf die evolutionäre Geschichte der Pflanzen wird allerdings von ihrer Häufigkeit in der Natur bestimmt. Des Weiteren kann sich die Hybridisierungsfrequenz zwischen phylogenetischen Gruppen unterscheiden und außerdem in Abhängigkeit zu den ökophysilogischen Merkmalen des Organismus stehen.

Das erste Manuskript, "*Factors determining hybridization rate in plants",* untersucht und schätzt die Häufigkeit von Hybridisierungen in Gefäßpflanzen im Bundestaat Michigan, U.S.A. ab. Diese Abschätzung bildet die Grundlage einer neuen Methode, um die hybridisierungsraten in Pflanzen in Abhängigkeit von ökologischen und physiologische Faktoren zu ermitteln. Insgesamt konnten 17,5% der untersuchten Arten hybridisieren und es gibt 0,127 einzigartige Hybride pro nicht-hybrider Art. Dabei sind die Hybridisierungsraten signifikant unterschiedlich zwischen Lebensforme und -geschichte: 97% der hybridisierenden Arten sind mehrjährig. Langlebige Arten, wie Bäume, Sträucher und "fern allies" haben die höchsten Hybridisierungsraten. Die Störung eines Habitats ist mit der Hybridisierungsrate assoziiert, das bedeutet, dass Arten die auf ungestörte Pflanzengemeinschaften begrenzt sind, weniger hybridisieren als Arten in gestörten. Dies zeigt, dass die Störung des Habitats eine kritische Komponente für die Entstehung und Etablierung von Hybriden ist. Die Daten zeigen auch, dass Hybridisierung auf bestimmte taxonomische Gruppen konzentriert ist, trotzdem hat die Phylogenie insgesamt keine Auswirkungen auf die Hybridisierungsrate. Diese Studie ist die erste, die ökologische und physiologische Faktoren nutzt, um die Hybridisierungsraten in Pflanzen zu ermitteln. Die Ergebnisse haben weitreichende Auswirkungen auf Verständnis von Hybridisierung auf die Evolution und könnte helfen taxonomischen Entscheidungen zu treffen.

Das zweite Manuskript, "*Population genomics reveals a new wild tomato species with a history of hybridization*", ist eine phylogenetische Fallstudie von

Hybridisierung in Wildtomaten. Die untersuchte Tomatenklade umfasst 12 Wildarten, deren Lebensraum das westliche Südamerika ist. Transkriptome von Individuen aus 38 unterschiedlichen Populationen wurden sequenziert und zusammen mit bereits publizierten Daten in einem umfassenden genomischen Datensatz der gesamten Klade analysiert. Als ein erstes Ergebnis kann diese Arbeit die phylogenetischen Verwandtschaftsverhältnisse der Wildtomaten auflösen und verbessern. Das Hauptaugenmerk des Manuskripts liegt aber auf der Populationsgenetik zweier hochvariabler Arten: *Solanum peruvianum* und *S. chilense*. Für die Art mit den meisten Polymorphismen, *S. peruvianum*, wird gezeigt, dass sie, genetisch wie geographisch, aus zwei Subpopulationen besteht. *Solanum chilense* hat ausgeprägte südliche, küstennahe Populationen, die eine reduzierte Heterozygosität besitzen. Dies weist auf eine jüngere Erweiterung des Lebensraums nach Süden und der Abspaltung von *S. peruvianum* vor ca. 1,25 Millionen Jahren hin. Die umfassende Sammlung von Individuen zeigt, dass unabhängige Populationen, welche momentan als *S. chilense* beschrieben werden, genetisch zwischen *S. chilense* und *S. peruvianum* stehen. Basierend auf molekularen, morphologischen und Kreuzungsdaten konnten wir die Hypothese, dass diese unabhängigen '*S. chilense*' Populationen ein Beispiel für Artbildung durch Hybridisierung sind, überprüfen. Artbildung durch homoploide Hybridisierung ist kaum bekannt, daher diskutieren wir die Schwierigkeiten bei der Identifizierung und Unterscheidung zwischen den alternativen demographischen Szenarien. Insgesamt bietet diese Entdeckung neue Möglichkeiten die Resultate von Hybridisierungen in Pflanzen zu verstehen und zeigt eine neue adaptive genetische Struktur in *S. peruvianum* und *S. chilense*.

Zusammenfassend betrachtet, tragen diese Arbeiten zu einem besseren Verständnis von Hybridbildung in Pflanzen bei und zeigen eine detaillierte Analyse der Ergebnisse von Hybridisierung in der Klade der Tomaten.

# Summary

This thesis consists of two first-author research manuscripts and an introduction to the topic of plant hybridization. Hybridization can be a significant force in evolution, including in the origin of new species and the transfer of genetic material between species. However, the importance of hybridization to the evolutionary history of plants is tied to its frequency in the wild. Furthermore, hybridization frequency may differ between phylogenetic groups and could also depend on an organisms ecophysiological traits.

The first manuscript, "*Factors determining hybridization rate in plants*," estimates the frequency of hybridization in the vascular plants of Michigan. This estimation is followed by a novel approach to test the association of hybridization rate with various ecological and physiological factors. In total 17.5% of species were found to hybridize and there were 0.127 unique hybrids per non-hybrid species. Hybridization rates were significantly different between life forms and life histories: 97% of hybridizing species are perennial, and long-lived life forms such as trees, shrubs, and fern allies have the highest rates of hybridization. Habitat disturbance was also found to be associated with hybridization rate: species restricted to more undisturbed plant communities hybridize less whereas species in disturbed habitats hybridize more. This indicates that habitat disturbance is critical to the formation and establishment of hybrids. The data also indicated that hybridization is concentrated in particular taxonomic groups but that there is no broad effect of phylogeny on hybridization rate. This is the first study to test any ecological or physiological factors on the hybridization rate in plants, and the intriguing results have important implications for the evolutionary interpretation of hybridization and could also help guide taxonomic decision making.

The second manuscript, "*Population genomics reveals a new wild tomato species with a history of hybridization*," is a phylogenomic case study of hybridization in wild tomatoes. The tomato clade includes 12 wild species distributed in western South America. Transcriptomes from individuals of 38 different populations were sequenced and combined with published data to build a comprehensive genomic dataset for the entire clade. The first accomplishment of

this study was to resolve phylogenetic relationships and clarify taxonomy in wild tomatoes. The main body of the work, however, focuses on the population genetics of two highly variable species: *Solanum peruvianum* and *S. chilense*. The most polymorphic species, *S. peruvianum*, is shown to have two geographical subpopulations. *Solanum chilense* is found to have distinct southern coastal populations with reduced heterozygosity, and this indicates a recent expansion south following speciation from *S. peruvianum* ca. 1.25 million years ago. The comprehensive sampling revealed that discontinuous populations currently described as *S. chilense* are genetically intermediate between *S. chilense* and *S. peruvianum*. Based upon molecular, morphological, and crossing data, we test the hypothesis that these discontinuous *'S. chilense'* populations are an example of hybrid speciation. Homoploid hybrid speciation is rarely reported, and we discuss the difficulties in identifying it and differentiating between alternative demographic scenarios. Overall, this discovery presents a new opportunity to understand the genomic outcomes of hybridization in plants, and identifies putatively adaptive genetic structure in both *S. peruvianum* and *S. chilense*.

Taken together, the manuscripts in this thesis provide both broad insight into hybridization in plants and a detailed analysis of the outcomes of hybridization in the tomato clade.

# Acknowledgements

Big thanks go to Laura for being a great supervisor over the past four years. It has been a period of extensive growth and development for me, and I thank you for being supportive, encouraging, and interested throughout. Thanks also to you Martin for your support, encouragement, and feedback. Many thanks to Robin and the people in her lab for their support and mentoring during my stay at MSU. Extra thanks to Grant for your shared interest in plant diversity and for giving me a copy of the Field Manual! Thank you to all of the people who have been in and out of the lab the last four years. Habib, I wish you the best finishing up your thesis. Same goes for Janina and Sarah. Sophie, best of luck in your future scientific career. Piet, thanks for the apples and all the money I won from you playing basketball. Thorsten, thanks for teaching me to program and the extensive use of your BBQ, letting me store stuff at your house, taxes etc. – very much appreciated! Best of luck to both of you on winning the lottery. Melanie, it was nice working with you, I'm sure your hard work will pay off soon! Amir, come to America and I will show you a bigly good time, thanks for the last year and a half! Thanks to Kevin for helping with the tomato crosses. Lampros, I had an awesome time in Greece! Best of luck to all the iGRAD members finishing your theses! Thanks to Dominik and Marc for all the sport. Special thanks to Sigrun and Petra, you have both helped tremendously. Thanks also to Susanne for introducing me to HHU (and Rob Last for bringing me to MSU). Special thanks to Roger Chetelat for the invitation to come to UC Davis – that meant a lot. Thanks also to the greenhouse staff for keeping my tomato plants alive. Antimo, thanks for the good times here and in Naples! I also want to thank my good friends Brandon, Pat, and Dylan for helping me get into botany. Thanks Alan Rebertus for being my undergraduate supervisor – your field courses were awesome! Finally, I want to thank my whole family for their support. Special thanks also to Emilie. Your support has been first-rate and essential to completing this thesis.

# Table of Contents

**Zusammenfassung** ..... **i**

**Summary** ..... **iii**

**Acknowledgements** ..... **v**

**Table of Contents** ..... **vi**

**1. Introduction** ..... **10**

*1.1. Background on Hybridization* ..... *10*

Figure 1 Examples of hybrids ..... 12

Figure 2 Examples of hybridizing species ..... 13

*1.2. Differentiation of Populations* ..... *14*

1.2.1. Gene flow and population structure ..... 14

1.2.2. Forces structuring plant populations ..... 14

1.2.2.1. Ecology ..... 14

1.2.2.2. Breeding behavior ..... 15

1.2.2.3. Polyploidy ..... 16

*1.3. Species Integrity and Hybridization* ..... *18*

1.3.1. The Biological Species Concept (BSC) ..... 18

1.3.2. Hybridization and the BSC ..... 19

1.3.3. Reproductive barriers between species ..... 20

1.3.3.1. Premating ..... 21

1.3.3.2. Postzygotic ..... 21

*1.4. Features of Interspecific Hybrids* ..... *23*

1.4.1. Rarity ..... 23

1.4.2. Reduced fertility ..... 23

1.4.3. Intermediate phenotypes ..... 24

1.4.4. Occurrence in disturbed habitat ..... 24

1.4.5. Occurrence in hybrid zones ..... 25

*1.5. Evolutionary Outcomes of Hybridization* ..... *26*

1.5.1. Introgression ..... 26

1.5.2. Speciation ..... 27

1.5.2.1. Homoploid hybrid speciation ..... 27

1.5.2.2. Polyploid speciation ..... 30

*1.6. Examples of Wild Plant Hybridization* ..... *32*

1.6.1. Ecology of hybridization in Iris ..... 32

1.6.2. Transfer of adaptations between species in *Senecio* ..... 32

1.6.3. Biogeography of a hybrid zone ..... 34

1.6.4. Species integrity *Quercus* ..... 35

1.6.5. Homoploid hybrid speciation in *Pinus* and *Argyanthemum* ..... 36

1.6.6. Allopolyploid hybrid speciation in *Tragopogon* ..... 38

*1.7. Frequency of Natural Hybridization* ..... *40*

1.7.1. Motivations of my research on this topic ..... 40

1.7.2. Outcomes of my research on this topic ..... 40

*1.8. Case Study of Hybridization in Wild Tomato* ..... *42*

1.8.1. Wild tomato taxonomy ..... 42

1.8.2. Modern methods and tools for studying hybridization ..... 43

1.8.3. Motivations and outcomes of my research....45
*1.9. Summary of the Scientific Contributions of this Thesis....46*
*1.10. References....47*

**2. Factors Determining Hybridization Rate in Plants....52**
*2.1. Overview....52*
*2.2. Abstract....53*
*2.3. Introduction....54*
*2.4. Materials and Methods....56*
2.4.1. The data....56
2.4.2. Defining mating system....57
2.4.3. Determining the extent of hybridization....57
2.4.4. Quantifying hybridization....58
2.4.5. Statistical tests of significance....59
2.4.6. Data availability....59
*2.5. Results....60*
2.5.1. Frequency of interspecific hybridization....60
2.5.2. Comparison of hybridization metrics....60
2.5.3. Hybrid species....61
2.5.4. Phylogenetic distribution of hybridizing species....61
2.5.5. Factors associated with hybridization frequency....62
*2.6. Discussion....64*
2.6.1. Metrics of hybridization and recommendations for future studies....64
2.6.2. Estimating the true hybridization frequency....65
2.6.3. Ecological correlates of hybridization: habitat disturbance as a critical factor....66
2.6.4. Phylogenetic distribution of hybridization....66
2.6.5. Life history and hybridization propensity....67
2.6.6. High rates of hybridization in woody plants....68
2.6.7. Interesting examples of hybridization in Michigan....69
*2.7. Conclusions....71*
*2.8. Acknowledgements....71*
*2.9. References....72*
*2.10. Figures....75*
Figure 1....75
Figure 2....76
Figure 3....77
Figure 4....78
Figure 5....79
*2.11. Tables....80*
Table 1....80
Table 2....81
Table 3....82
Table 4....83
Table 5....83
*2.12. Supplemental Figures....84*
Supplemental Figure 1....84
Supplemental Figure 2....85
Supplemental Figure 3....86
Supplemental Figure 4....87
Supplemental Figure 5....88

Supplemental Figure 6 ....89
*2.13. Supplemental Tables* ....*90*
Supplemental Table 1 ....90
Supplemental Table 2 ....92
Supplemental Table 3 ....107
Supplemental Table 4 ....107
2.13.1. References for Supplemental Table 2 and Supplemental Table 4 ....109
**3. Case Study of Hybridization in Wild Tomato** ....**112**
*3.1. Overview* ....*112*
*3.2. Abstract* ....*113*
*3.3. Introduction* ....*114*
*3.4. Materials and Methods* ....*116*
3.4.1. Transcriptome Data ....116
3.4.2. Additional Data ....117
3.4.3. Read mapping to the reference genome and SNP calling ....117
3.4.4. Interspecific relationships and genetic groups within wild tomato ....118
3.4.5. Within-species nucleotide diversity and individual heterozygosity ....119
3.4.6. Modeling the joint demography of *S.chi* and *S.per* ....119
3.4.7. Test crosses ....120
*3.5. Results* ....*121*
3.5.1. Distribution of genetic variation across groups ....121
3.5.2. Recent speciation of *S.chi* and *S.per* ....122
3.5.3. Evidence of natural *S.chi* × *S.per* hybrid populations in southern Peru ....123
3.5.4. Broader phylogenetic implications ....125
*3.6. Discussion* ....*128*
3.6.1. Demography and speciation of *S.chi* and *S.per* ....128
3.6.2. *S. corneliomulleri* and *S. huaylasense* ....130
*3.6.3. Solanum arcanum* ....131
3.6.4. Evidence for widespread cryptic hybrid populations ....132
3.6.5. A putative example of recombinational speciation ....133
*3.7. References* ....*137*
*3.8. Figures* ....*141*
Figure 1 ....141
Figure 2 ....142
Figure 3 ....143
Figure 4 ....144
Figure 5 ....145
Figure 6 ....146
*3.9. Tables* ....*147*
Table 1 ....147
*3.10. Supplemental Methods* ....*148*
3.10.1. Ploidy (C-value) determination by flow cytometry ....148
3.10.2. ∂a∂i Analysis ....148
*3.11. Supplemental Figures* ....*150*
Supplemental Figure 1 ....150
Supplemental Figure 2 ....151
Supplemental Figure 3 ....152
Supplemental Figure 4 ....153
Supplemental Figure 5 ....154

Supplemental Figure 6 ..... 155
Supplemental Figure 7 ..... 156
Supplemental Figure 8 ..... 157
Supplemental Figure 9 ..... 158
Supplemental Figure 10 ..... 159
Supplemental Figure 11 ..... 160
Supplemental Figure 12 ..... 161
Supplemental Figure 13 ..... 164
Supplemental Figure 14 ..... 165
Supplemental Figure 15 ..... 166
Supplemental Figure 16 ..... 167
Supplemental Figure 17 ..... 168
*3.12. Supplemental Tables ..... 169*
Supplemental Table 1 ..... 169
Supplemental Table 2 ..... 171
Supplemental Table 3 ..... 172
Supplemental Table 4 ..... 174
Supplemental Table 5 ..... 175

**4. Appendix 1: The Hybrids Database Website ..... 181**
*4.1. Motivation for the website ..... 181*
*4.2. Example pages from the website ..... 182*
4.2.1. Home page ..... 182
4.2.2. About page ..... 182
4.2.3. Search for data ..... 183
4.2.4. Search results ..... 183
4.2.5. Individual data record for a hybridizing species ..... 184
4.2.6. Individual data record for a hybrid species ..... 184
4.2.7. Individual data record for a unique hybrid ..... 184

**5. Appendix 2: Additional Tomato Data ..... 185**
*5.1. Additional figures explaining the RNA-seq data ..... 185*
*5.2. Model of S. chilense and S. peruvianum speciation ..... 186*
*5.3. Haplotype lengths in the hybrid populations ..... 187*
*5.4. Fixed differences between S. chilense and S. peruvianum ..... 188*
*5.5. Seed germination rates for all crosses ..... 189*
*5.6. Hybrid viability ..... 190*
*5.7. Additional tomato crossing data ..... 191*

# 1. Introduction

## 1.1. Background on Hybridization

The term hybrid can describe the product of any two heterogeneous things. The botanist G. L. Stebbins (1959, p. 231) proposed an evolutionary definition of hybridization as “crossing between individuals belonging to separate populations which have different adaptive norms.” Hybrids between species are termed interspecific hybrids and were historically thought to be sterile misfits. However, interspecific hybridization is now recognized as an important and creative force in evolution, including in the origin of new species.

Documented experiments in hybridization date back to the Babylonians and Assyrians who realized the need to cross fertilize dioecious date palms in order to get fruits (Roberts, 1929). Linnaeus and others also did experimental crosses in many plant genera, including *Veronica*, *Verbascum*, and *Tragopogon* (Figure 1). Wilhelm Focke was the first to systematically study hybridization in his book “Die Pflanzen-Mischlinge” (Focke, 1881). Focke made the important observation that hybrids were widespread and that some hybrid seeds were fertile after he repeated the experimental crosses of Linnaeus in *Tragopogon* (Focke, 1890). Following the rediscovery of Mendel’s hybridization work, Lotsy (1916) was the first to recognize hybridization as an important force in evolution (Stebbins, 1959). Winge (1917) was apparently the first to implicate polyploidy and hybridization as potential forces for speciation (Stebbins, 1959).

The proceeding 100 years after Winge have been full of research and insights into the topic. Today, hybridization is critical to the breeding and production of most important crops. This includes all cereal crops that depend on hybrid vigor for high yields. Other crops such as seedless watermelon, seedless grapes, and seedless banana are the sterile triploid products of hybridization. Furthermore, disease and drought resistance alleles have been reintroduced via interspecific hybridization from wild relatives into the breeding populations of many crops for their improvement, including tomato (Lin et al., 2014). Many wild crop ancestors have a recent history of natural hybridization including the allopolyploids maize, tobacco, potato, cotton, strawberry, peanut, and coffee. Other iconic

species like the coastal redwoods or the cultivated sycamore also ultimately originate from hybridization events – even some populations of the lowly dandelion are triploid hybrids! Hybridization is therefore both critical to agriculture and common in wild plant species (Figure 2).

Indeed, one of the most exciting discoveries of genetics has been the realization that natural hybridization is widespread in plants (Rieseberg, 1995). It is now established that hybridization is a mechanism of both speciation and the transfer of advantageous adaptations between species. In this introduction, the forces that distribute genetic variation in plant populations are briefly reviewed. Next, the idea and integrity of plant species is examined, followed by an overview of plant reproductive barriers. Then, a few features that are common to many interspecific hybrids are discussed. This is followed by an examination of the main outcomes of hybridization: introgression and speciation. Finally, the features, outcomes, and idiosyncrasies of hybrids are explored using case studies.

In section 1.7, the frequency of natural hybridization is briefly reviewed. This topic forms the basis for the original research presented in chapter two, “Factors determining hybridization rate in plants”. Finally, in section 1.8, we build upon the outcomes of hybridization by introducing wild tomato and a putative example of hybrid speciation in this clade. This topic forms the basis of the original research of chapter three, “Case study of hybridization in wild tomato.”

**Figure 1** **Examples of hybrids**. (**a**) *Platanus* ×*acerifolium* is a vigorous and commonly planted man-made hybrid between the North American and European sycamores. (**b**) Peppermint (*Mentha* ×*piperita*) is the hybrid of watermint and spearmint. (**c**) *Nicotiana tabacum* (tobacco) is an allotetraploid hybrid of *N. sylvestris* and *N. tomentosiformis*. (**d**) *Capsella bursa-pastoris* is a 100,000-300,000 year old allotetraploid hybrid species. (**e**) The genus *Cardamine* includes the hybrid species *C. maxima.* There are several hybrid species in *Tragopogon* and studies of hybridization in this genus date back to the 1700s.

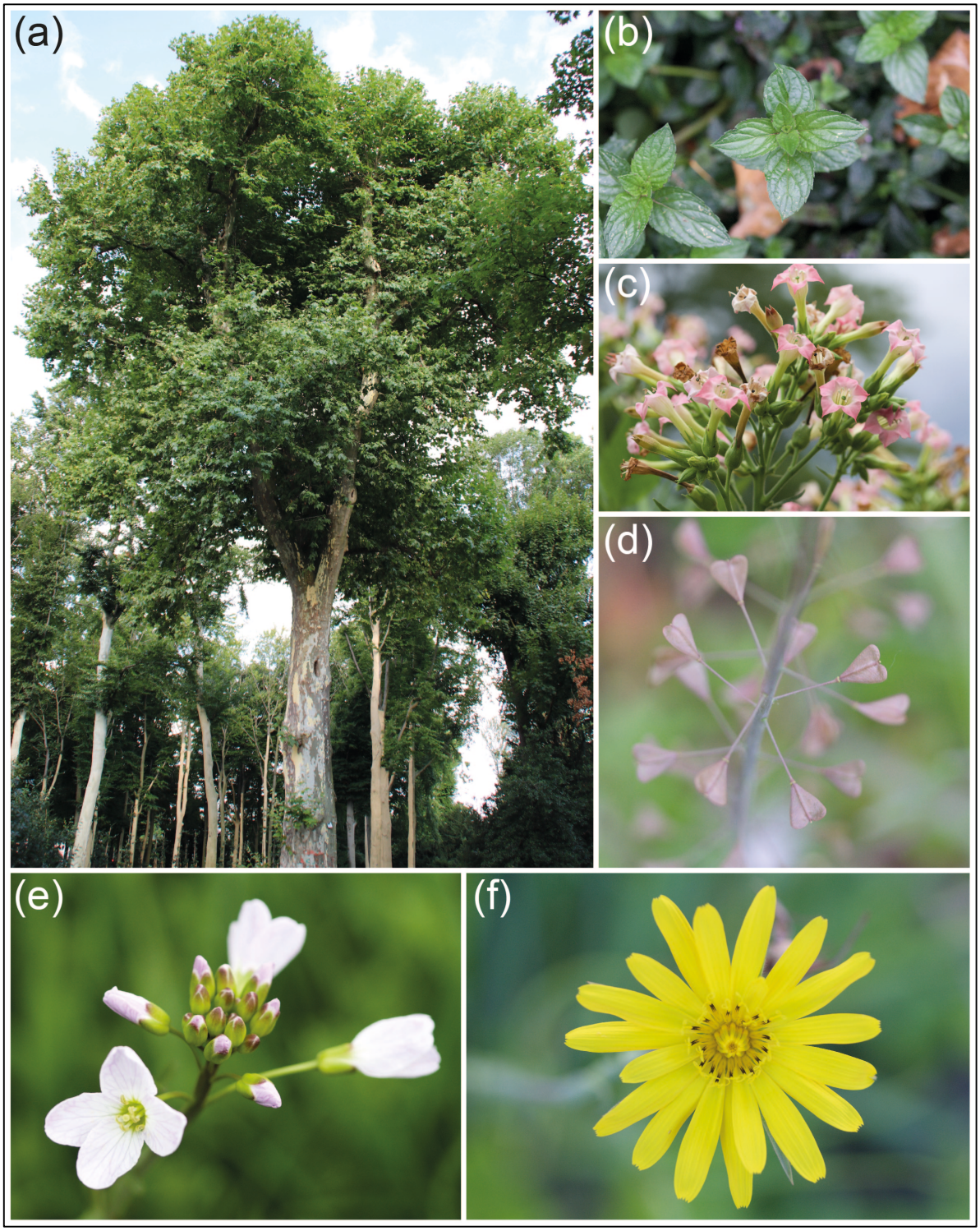
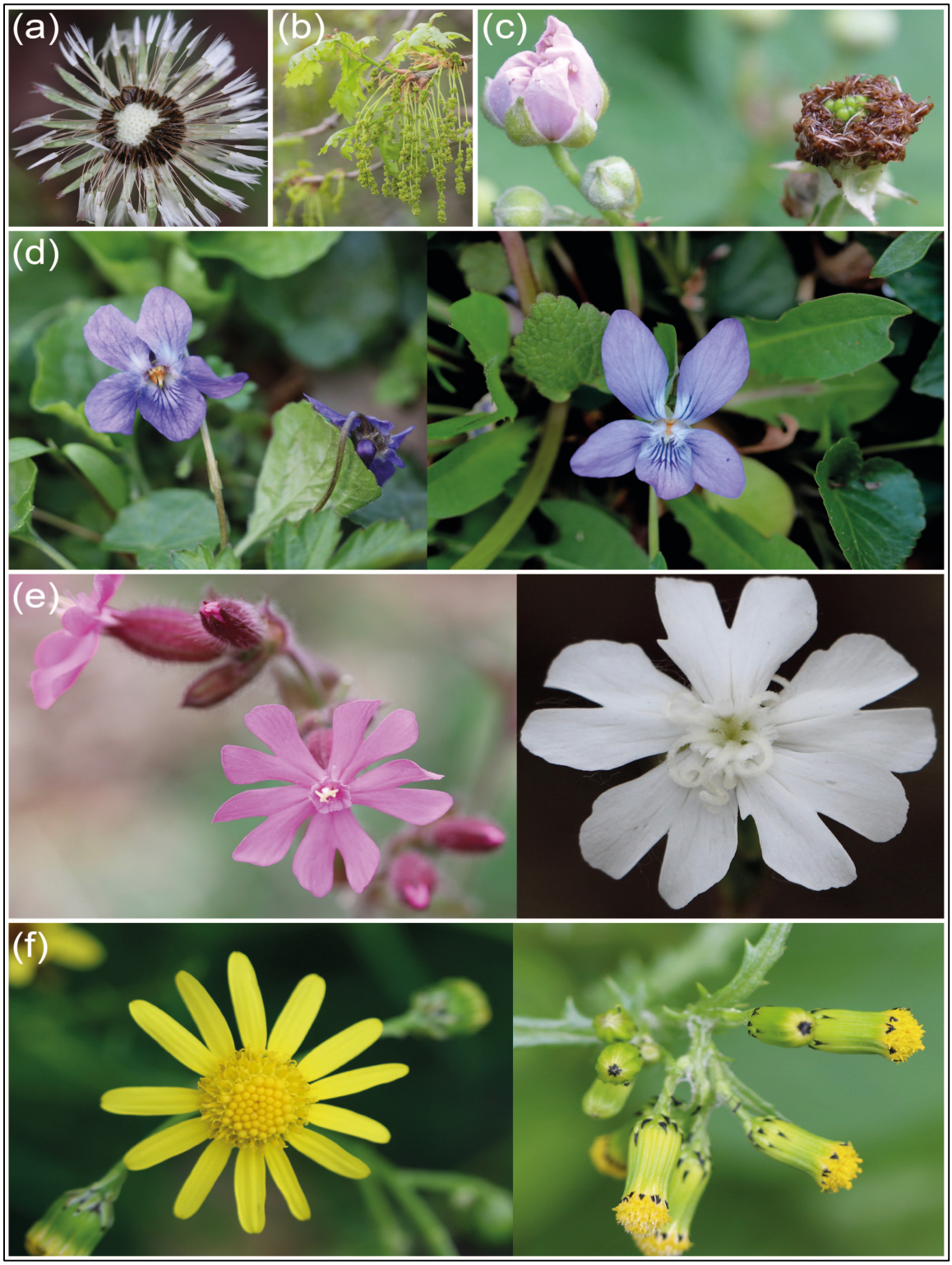


**Figure 2** **Examples of hybridizing species**. (**a**) Asexual seeds of dandelion (*Taracacum officinale*). (**b**) Male and female flowers of *Quercus*. (**c**) *Rubus spp.* flowers and developing fruit. The genera *Taraxacum*, *Quercus*, and *Rubus* are well-known for hybridizing (**d**) Two *Viola* specimens from Düsseldorf (left) and Karlsruhe (right) showing variation within a species. (**e**) *Silene dioica* and *S. latifolia* are known to hybridize. (**f**) A radiate *Senecio sp.* (left) and the non-radiate *S. vulgaris* (right). The radiate phenotype has been transferred between species in *Senecio* via introgressive hybridization.



## 1.2. Differentiation of Populations

### 1.2.1. Gene flow and population structure

The Hardy-Weinberg equilibrium (HWE) describes the effect of random mating on allele and genotype frequencies within a population. However, most plant species are not distributed uniformly, do not mate at random, and are not in HWE (Hedrick, 2011). They instead consist of many subpopulations. The movement between subpopulations (migration) that results in genetic exchange between them is termed gene flow. All subpopulations have reciprocal gene flow at rates dependent on several dynamic ecological and physiological factors. Often there is a hierarchy of increasingly subdivided units. For example, alpine plants occupying different mountains also have local populations on each mountain.

The subdivision of genetic variation is important because allele frequencies can change independently in different subpopulations. These changes can be due to drift or local adaptation. Eventually change in allele frequencies (evolution) at many loci can result in reproductive barriers between the populations (i.e. speciation). Thus, the forces structuring variation in populations are fundamental to the origin of species and therefore fundamental to understanding interspecific hybridization.

### 1.2.2. Forces structuring plant populations

Plant populations are particularly dynamic because of their diverse life histories, reproductive modes, and life forms. Here, the main forces structuring genetic variation in populations are divided into: ecology, breeding behavior, and polyploidy.

#### 1.2.2.1. Ecology

Habitat may be the most important force structuring variation in populations. Most species are adapted to one or a few ecological niches, and they are therefore restricted by the availability of habitat. Discontinuous habitats result in disjunct populations. Given that most seeds/gametes disperse near the parent (in the 10 to 100s of meters with a low frequency of long-distance dispersals), most matings will occur between neighboring individuals. With a clumped distribution due to habitat, wide outcrosses between inbred neighborhoods will be

comparatively rare even in obligate outcrossing species. This creates a population substructure of many discontinuous races which Grant (1981) termed colonial because the species consist of many interbreeding “colonies” which are local and do not breed with one another. Populations of both alpine species occupying different mountaintops and wild tomatoes in different river valleys of the Andes are examples of species expected to be colonial. The clumped distribution of differentiated populations due to habitat is akin to those due to autogamy and asexual reproduction.

More widespread species (e.g. long-lived boreal trees) often form large, homogenously distributed populations. Clumped inbred colonies are rare for this type of species. However, unbroken geographic differentiation, “geographic races,” do develop. This is common, for example, in arctic-adapted circumpolar species (e.g. *Arctostaphylos uva-ursi*). Geographic races at the ends of a species’ distribution can have different “adaptive norms” and may even be incompatible. This is the idea behind ring species: an array of interbreeding, negligibly differentiated populations around a geographic barrier (e.g. a plateau). When the terminal races of the ring species meet they are divergent enough to have genomic incompatibilities (i.e. may be termed species despite unbroken interbreeding around the ring).

#### 1.2.2.2. Breeding behavior

Reproductive mode is an important factor determining the migration rate between populations. Reproductive events can be divided into whether zygotes are formed from outcrossing, autogamy, or apomixis. Most plants are outcrossing and many are obligate outcrossers due to genetic self-incompatibility, dioecy, dichogamy, and mechanical barriers to inbreeding. However, many species are also facultative autogamous. Self-fertilization (autogamy) reduces heterozygosity by half per generation and results in inbreeding depression, but it is always more fit to have some reproduction even if there is inbreeding depression. Overall, there are dozens of ecological and genetic factors that dictate the rate of outcrossing, and more than one fourth of seed plants are predominately autogamous (data from Barrett and Eckert (1990), autogamous defined as an outcrossing rate <0.2; they analyzed outcrossing rate in 139 species of seed plants).

Different populations of the same species can also have different rates of autogamy. For example, the wild tomato species *S. pimpinellifolium* has variable rates of outcrossing (0-40%) within one population (Rick et al., 1978). Woody perennials have higher rates of outcrossing than herbaceous perennials, and annuals have the highest rates of autogamy (Barrett and Eckert, 1990). Because an annual life history is evolutionarily risky (i.e. only one opportunity at reproduction), autogamy (i.e. reproductive assurance) is expected to evolve more frequently.

Other species forego mating altogether (apomixis). These species can form seeds asexually using different mechanisms like the development of unreduced eggs in the embryo (e.g. *Taraxacum*). This is akin to clonal reproduction but via seeds and is termed agamospermy. Vegetative clonal reproduction is also very common in perennial species, such as the extremely long-lived clones of *Populus spp.* or the short-lived forb *Trientalis borealis*. Both of these species reproduce asexually by forming ramets (akin to individuals) at the ends of rhizomes and/or stolons; all ramets are genetically identical to the mother plant. In this way, large colonies, termed a genet, are formed. Both autogamy and asexual reproduction result in many relatively homogenous local populations that are differentiated from one another and reproduce almost exclusively inter se. Sometimes these are termed microspecies.

#### 1.2.2.3. Polyploidy

Polyploidy is also important in structuring genetic variation in plants. Polyploidy is often cryptic, but one estimate has 13% of species with multiple cytotypes (Soltis et al., 2007). These cytotypes are often intersterile (see 1.3.3.2). One of the advantages of polyploidy is fixed heterozygosity. This is the case in arctic plants, for example, which are highly autogamous and highly polyploid (up to 18x!) (Brochmann et al., 2004). Interestingly, most of the heterozygosity within these species is also within any one individual, and this is a reversal of normal population subdivision that actually results in a deficiency of heterozygotes (Hedrick, 2011). Polyploidy is especially important for hybrid speciation (1.5.2). Whether to recognize intersterile cytotypes as species is also of fundamental importance to the frequency of interspecific hybridization (1.7).

In conclusion, there are many interdependent and dynamic forces that structure variation within species, and the interplay of forces is different for each species. Genetic structure within a species is often recognized by the formal nomenclatural ranks of subspecies, variety, and form (infraspecific ranks). Other terms like semispecies, race, and deme are not formal, but are frequently employed to describe infraspecific variation. Formal recognition at the rank of a species usually depends on reproductive isolation between differentiated populations (1.3.3). However, many of the forces that restrict breeding between populations within one species also reproductively isolate species. Hybridizations between species are therefore *not fundamentally different* than crosses between infraspecific ranks, but there are a few defining features of interspecific hybrids that make them particularly interesting (1.4).

## 1.3. Species Integrity and Hybridization

*I view the term species as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other.*

C. Darwin (1859, p. 52)

*A species is a group of individuals fully fertile inter se, but barred from interbreeding with other similar groups by its physiological properties (producing incompatibilities of parents, or sterility of hybrids, or both).*

T. Dobzhansky (1935, p. 353)

### 1.3.1. The Biological Species Concept (BSC)

Species are the basic unit in taxonomy, systematics, ecology, and evolutionary biology; but defining them is difficult. For most of history, species have been defined from gaps in morphological variation between related populations; this is the taxonomic and/or morphological species definition. However, plant populations are highly variable and it is not obvious which morphological differences are evolutionarily important or meaningful. Today, the biological species concept (BSC) is the most widely used species definition (Coyne and Orr, 2004). It defines species as “groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups” (Mayr 1942, p. 120). Reproductive barriers between species are therefore of fundamental importance for the BSC (1.3.3).

Many influential botanists have accepted the BSC (Grant, 1981; Rieseberg et al., 2006; Schemske, 2000; Soltis and Soltis, 2009; Stebbins, 1950), but others, for a variety of reasons, have not (Donoghue, 1985; Ehrlich and Raven, 1969; Levin, 2000). First, all microspecies of autogamous and apomictic taxa would need to be ranked as species under the BSC. There are, in fact, over 2,000 named species of the agamospermous *Taraxacum officinale*; all of these species are defined from morphology, *but with the understanding* that they do not interbreed (Kirschner and Štěpánek, 1994). Many populations have morphologically cryptic cytotypes, and because cytotypes can have strong reproductive barriers, these also represent biological species (Ramsey and Schemske, 1998; Soltis et al., 2007). Third, from a theoretical systematic

perspective, intercrossability is a symplesiomorphy (a shared, ancestral character) and therefore not really appropriate for defining clades (Soltis and Soltis, 2009). Fourth, some have argued that the population – not the species – is the unit of evolution and species are therefore invalid (Levin, 1979; Levin, 2000). This idea may be true for some populations/semispecies, but appears to be mostly unfounded (Morjan and Rieseberg, 2004). Finally, the most practical problem with the BSC is that, for most taxa, it is unrealistic to do the test crosses needed to define a species. Thus, in practice nearly all plant species are defined by their morphology (taxonomic species) not their intercrossing relationships.

Because of these issues (and hybridization, 1.3.2), many other species definitions were formulated, including those focusing on morphological, evolutionary, ecological, or phylogenetic criteria. These are extensively reviewed in Grant (1981), Templeton (1989), Baum and Donoghue (1995), Coyne and Orr (2004), Rieseberg and Willis (2007), de Queiroz (2007) and Richards (2010). All of these definitions also have inherent issues. For example, multiple origin polyploids interbreed but are not monophyletic; species with high carrying capacities that take many generations to get reciprocal monophyly for all alleles; and gaps in morphological variation are not always evolutionarily important. Thus, many (if not most) botanists think in terms of the BSC today (Rieseberg and Willis, 2007; Schemske, 2000).

### 1.3.2. Hybridization and the BSC

There are many taxonomic groups where there are few breeding barriers between species and hybridization is frequent; one example is the oak syngameon (1.6.4). Hybridization between species is, at first thought, irreconcilable with the BSC. However, many species retain their reproductive isolation despite hybridization: their species integrity is not destroyed by hybridization. The fact that *hybridization does not equate to extensive gene flow* is the reason that occasional hybridization is allowed in most interpretations of the BSC, including here (Coyne and Orr, 2004; Rieseberg et al., 2006).

Furthermore, as Mayr (1992, p. 228) remarks, species are "*the property of populations, not of individuals.* A population does not lose its species status when an individual belonging to it makes a mistake." Grant (1981, p. 92) also provides the insight that "all populations will *not* be grouped into discrete biological species

at any moment in history. The fact that *some* biological species exist is the essential point." There is indeed no definite cutoff for when variation between populations becomes variation between species, and because species do not instantiate immediately, there will be murky cases (aka semispecies).

In any case, species are hypotheses of relationships to be tested on a case-by-case basis using a variety of methods and all available data. They should represent meaningful and evolutionarily in- and inter-dependent units of evolution. Judd et al. (p. 151) conclude that, "it is difficult to generalize about plant species because each one has a unique diversification." This is indeed the case, diversity is idiosyncratic. That said, the majority of species are evolutionarily meaningful units whether defined from morphology, interbreeding, ecology, reciprocal monophyly or any other criteria (Coyne and Orr, 2004; Rieseberg et al., 2006). Finally, the fact that most species do not hybridize makes the rare instances of interspecific hybridization very interesting.

### 1.3.3. Reproductive barriers between species

*Even if there is evidence of backcrossing but the intergrading types remain relatively uncommon in comparison with sharply distinct parent types, it may be presumed that there is so much selection against the hybrids that they do not destroy the integrity of the two species.*

S. Wright (1978, p. 5)

Species integrity is dependent on functioning reproductive barriers. Levin (2000) divides reproductive barriers into ecological (prezygotic) and genomic (postzygotic). One could also divide them into hybrid formation (prezygotic) and hybrid establishment (postzygotic). Here, I divide barriers into premating (1.3.3.1) and postzygotic (1.3.3.2). There are also postmating prezygotic barriers, but these are not reviewed here. The relative importance of the different barriers is dependent on the species, and can also be dynamic depending, for example, on the maturity and composition of an ecological community.

Only prezygotic barriers need to be overcome for the formation of a hybrid. However, for the hybrids to exert an influence on their parental species or become established as species themselves, postzygotic barriers also need to be overcome (reinforcement is a potential outcome of hybridization that does not depend on

hybrids overcoming postzygotic barriers). Reproductive barriers are not only important to species integrity but also dictate the features of hybrids, including if and where they will form (1.4).

#### 1.3.3.1. Premating

On p. 14, habitat and geography were shown to be important in structuring heterozygosity within a species. Both also function as reproductive barriers between species. In fact, Mayr (1982, p. 273) modified the BSC in recognition of the importance of ecology and habitat: “a species is a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature.”

Most closely related species occupy different ecological and/or geographic niches. If species are allopatric, then their biogeography is obviously a strong reproductive barrier. When species are parapatric, hybrids are normally formed/established in hybrid zones (1.4.5). In geographically sympatric species, habitat differences can also be effective breeding barriers. An informative example of ecological isolation is the adaptive radiation of the silversword alliance on Hawaii. None of these species have post reproductive barriers but all occupy vastly different habitats (Seehausen, 2004). Therefore, these closely related species neither interact nor interbreed. Habitat breeding barriers are more dynamic than geographic or genomic ones because they easily can easily break down following habitat disturbance (1.4.4).

Besides geography and habitat, reproductive mode can isolate species. First, in autogamous or facultative autogamous species, the frequency of hybridization obviously decreases with an increasing frequency of autogamy. For sympatric outcrossing species, different flowering times can inhibit interspecific matings (e.g. *Salix*). Other outcrossing species can be isolated by their pollinators or due to mechanical barriers (e.g. *Asclepias*).

#### 1.3.3.2. Postzygotic

Postzygotic barriers are those that reduce fitness of the hybrids following zygote formation. While it is generally thought that premating barriers are more important (Lowry et al., 2008; Rieseberg and Willis, 2007; Stebbins, 1950),

postzygotic barriers are critical to species integrity when there are high interspecific mating rates (e.g. in oaks, 1.6.4).

The first postzygotic barrier is again ecological: in the absence of suitable habitat, hybrid individuals will have decreased fecundity. Like the prezygotic habitat barrier, this type of barrier can breakdown following habitat disturbance. Other postzygotic barriers are governed by genetic interactions.

The Bateson-Dobzhansky-Muller (BDM) model involves divergent alleles at one or more loci. For any BDM locus, there are two alleles that function with the ancestral genotype, but are incompatible with one another. Thus, the BDM alleles interact negatively in hybrids, either resulting in developmental difficulties or other detrimental fitness outcomes. Relatively few BDM incompatibilities between wild tomato species have been shown to reduce the fertility of interspecific hybrids (Guerrero et al., 2017; Moyle and Graham, 2005). Disease resistance loci can also function as BDM incompatibilities. One example is the triggering of autoimmunity in *Arabidopsis thaliana* hybrids (Bomblies et al., 2007). BDM incompatibilities are also important in models of homoploid hybrid speciation (1.5.2.1).

Hybrid sterility is a feature of many interspecific hybrids, (e.g. wood avens and *Populus*). And even if hybrid plants are vigorous, their inability (or reduced ability) to produce viable gametes is an effective reproductive barrier. Sterility can be due to structural mispairings during meiosis and the resulting aneuploid gametes. Another model of hybrid sterility involves a reciprocal translocation between two chromosomes. One reciprocal translocation results in first generation hybrids producing 50% unbalanced (and inviable) gametes. Interestingly, one mechanism to overcome sterility due to chromosomal structural differences (including translocations) is polyploidy. However, polyploidy cannot overcome incompatibilities due to BDM incompatibilities.

Moreover, polyploidy can itself be a strong postzygotic barrier between cytotypes. The triploid block is a well-known example: triploid individuals produce a majority of aneuploid gametes. A low percentage of gametes from triploids will be viable, but most are not (Ramsey and Schemske, 1998). Moreover, 3x, 5x, 7x etc. polyploids are sterile, but hybrids between e.g. a diploid and a hexaploid are 4x and are expected to be fertile. Thus, polyploidy can be a strong reproductive barrier but not in all cases.

## 1.4. Features of Interspecific Hybrids

There are a few features of offspring of crosses between individuals at the rank of species (interspecific hybrids) that distinguish them from crosses between individuals from within one variable species. These are *features not rules*.

### 1.4.1. Rarity

Many hybrids have only been collected once and most hybrids are rare. In fact, Focke (1881) recognized that hybrids are more common when one of the parents is rare. Rarity is due to multiple barriers that inhibit hybridization, and if hybrids were both frequent and widespread (they may be frequent in local hybrid zones), then the parents would probably intergrade and this is evidence for hybridization between geographic races not species. In my own experience, hybrids are extremely rare in any given habitat. Stebbins (1959) argued that occasional hybridization is the rule in angiosperms, but this is quite a different idea than hybrid *individuals* being the rule (1.7). Often hybridization is rare and cryptic enough that it is realized only from footprints of hybridization left in the genome (1.8.2).

### 1.4.2. Reduced fertility

For most of history, hybrids have been associated with sterility; they were seen as freaks, misfits, and monsters. Sterility of interspecific plant hybrids is indeed a common feature. Of course, not all hybrids are fully sterile, and full fertility is often restored in later generations (Arnold et al., 1999; Rieseberg, 1995). Moreover, most studies that have quantified fertility by examining only one or a few individuals. In the wild, even those hybrids deemed inviable can occasionally reproduce. In the case of *Helianthus annuus* x *H. bolanderi*, for example, first-generation hybrids are highly sterile in the lab, but there some races of *H. bolanderi* that have many introgressed individuals (Carney et al., 2000; Heiser, 1949). This indicates that wild hybrids must be somewhat fertile. Nevertheless, sterility (or a quantifiable reduction in fertility) is a frequent feature of interspecific hybrids.

### 1.4.3. Intermediate phenotypes

First-generation hybrids are by definition genetically intermediate between the parental taxa. They are also expected to be intermediate in morphology. This is true for quantitative phenotypes (i.e. those involving many loci), but phenotypes governed by only one or a few genes are often not intermediate (Rieseberg, 1995). In the case of a dominant allele, for example, all hybrids will resemble the dominant parent.

Hybrids often exhibit extreme phenotypes that are outside of the variation found in the parental species (aka transgressive phenotypes). While hybrid vigor is common in first generation man-made hybrids between inbred lines, transgressive phenotypes are more common in later generation wild hybrids. This is because later generation hybrids have unique genotypes (and phenotypes) due to Mendel's first law. The $F_2$ generation usually has extremely high variability and can be expected to fill the entire phenotypic gap between the parental species (e.g. Brochmann et al. 2000).

These unique phenotypes are often important in hybrid speciation because they render the ability to colonize new ecological niches and habitats not occupied by the parental species; adaptation that results in an ecological barrier to backcrossing and allows reproduction inter se – criteria critical for the formation of a new species.

### 1.4.4. Occurrence in disturbed habitat

*When man 'hybridizes the habitat' as in burning, ditching, pasturing, tilling the soil, he produces new, and frequently relatively unoccupied, niches in which introgressants are at a selective advantage.*

E. Anderson (1953 p. 289)

For two species occupying adjacent but very different habitats (e.g. a forest and a swamp), their hybrids are not expected to be adapted to either habitat. However, the hybrids could be adapted to an intermediate habitat (e.g. forest/swamp edges). This is fundamentally different than a geographic hybrid zone because habitat barriers can break down randomly and locally following disturbance. They are therefore not really bounded by geography. Habitat

disturbance frequently results in local hybrid swarms (i.e. populations of the parental species and multiple generations of hybrids/backcrosses).

A good example of the interaction between habitat and hybridization are two species of shrub, *Amelanchier bartraminana* and *A. laevis*, which are both found in northeastern North America. *Amelanchier bartraminana* inhabits undisturbed forests, bogs, and wet conifer swamps (Reznicek and Voss, 2012); *A. laevis* is found more often in early successional habitats like roadsides, forest edges, and recently burned areas. The disturbance of *A. bartraminana* habitat creates an opportunity for *A. laevis* immigrants to establish. This brings the opportunity for hybridization (Judd et al., 2016; Weber and Campbell, 1989). Their hybrid, *A.* ×*neglecta* is intermediate between the parents in morphological characters and frequently forms following disturbances of *A. bartraminana* habitat (Weber and Campbell, 1989). Without habitat disturbance immigrants of *A. laevis* are unfit and hybridization does not occur. Thus, the different niche preferences of these species are a strong ecological reproductive barrier, but it is a non-genetic barrier that can break down locally following habitat disturbance. This is also the case for a well-known example of hybridization in *Iris* (1.6.1).

### 1.4.5. Occurrence in hybrid zones

Anderson (1953) made the observation that most introgressions into Eastern North American species were coming from the Rocky Mountains, the Ozark Plateau, the more boreal north, or the Appalachian Mountains. These are all examples of geographic hybrids zones. In plants, hybrid zones are broad and not as geographically well-defined as in birds or other species (Harrison, 1993).

One good example of a bounded hybrid zone is in two Northern Flicker species (*Colaptis*, Picidae). Their hybrid zone follows the Rocky Mountains from Texas to Canada. The hybrids are not fundamentally unfit, but the differential adaptations in the parents maintain the hybrid zone. In turn, hybrid fitness and mean offspring dispersal distance determine its width. Hybrid zones can be stable over long periods of time. This is the case for *Helianthus annuus* and *H. petiolaris* which have many hybrid swarms, but the species nevertheless maintain their identities (1.3.2). The hybrid zone in *Tragopogon* is another example of a stable hybrid zone (1.6.6).

## 1.5. Evolutionary Outcomes of Hybridization

*The role of hybridization in evolution depends not on the frequency with which hybrids occur in nature, but on the effects that the hybrids which occur may have on genetic variability in natural populations*

G. Stebbins (1959, p. 235)

Many hybrids are infrequently formed and fail to reproduce or establish. For those hybrids that do make it, there are many proposed evolutionary outcomes. The two most important are briefly discussed here: introgression and hybrid speciation. Additional examples of these outcomes are given in 1.6.2 and 1.6.5, respectively.

### 1.5.1. Introgression

A species may plunder the genetic heritage of its relatives via introgressive hybridization. The basic idea is that alleles from one species are incorporated into the genomic background of another species. This occurs if a hybrid and its offspring repeatedly backcrosses to one of the parental species for many generations. Hybridization is not always equivalent to introgression, but it is widely thought to be the most frequent outcome of hybridization (Grant, 1981). One important and immediate result of introgression is increased genetic diversity in the introgressed species.

In some cases, introgression can have dramatic phenotypic outcomes: this has been demonstrated in plants by Kim et al. (2008) (1.6.2). However, most introgressions remain within the local hybridizing population (i.e. populations outside of the hybrid zone are not affected by hybridization). This can be due to natural selection against alleles in habitats for which they are maladaptive. Introgressions beyond a hybrid zone (aka dispersed introgressions) are rare, but do occur.

In plants, cytoplasmic DNA is introgressed more frequently than nuclear DNA (Rieseberg and Soltis, 1991). In fact, the discordance between chloroplast and nuclear phylogenies was some of the initial genetic evidence that hybridization was more widespread than previously thought (Rieseberg, 1995). A model of why cytoplasmic DNA can be more easily introgressed involves

immigration of one individual into an interspecific population. In this case, the immigrant will be fertilized exclusively by the neighboring individuals of the other species. However, the immigrant will not fertilize others (because of conspecific pollen preference). Thus, all admixed offspring will have 50:50 nuclear DNA and be 100% immigrant cytoplasmic DNA. The hybrids will also mate unequally, and eventually only the cytoplasmic DNA could be left with the nuclear genome of the established species (and perhaps a few introgressions).

There are many remarkable examples of introgressive hybridization and the transfer of adaptive alleles between species, including in humans. For example, all non-African humans have Neanderthal introgressions that account for, depending on the population, 1-2% of their DNA. It has been suggested that these introgressions were adaptive and helped humans transition to the cold and diseases of Europe and Asia (Racimo et al., 2015). There is also evidence that introgressions from a Denisovan-like ancestor helped Tibetan populations adapt to high-altitudes (Huerta-Sanchez et al., 2014). Introgressions from hardy and high-altitude adapted yaks are also used to improve cattle (Medugorac et al., 2017). Cases of introgression have also been documented between our closest relatives: chimpanzee and bonobo (de Manuel et al., 2016).

### 1.5.2. Speciation

First generation hybrids are usually morphologically intermediate between their parents, but later generation hybrids are highly variable and frequently have novel phenotypes. If these new types develop a reproductive barrier to their parents, then they can form new species. The origin of species via hybridization is a break from traditional thinking of two populations diverging in allopatry and is a creative outcome of hybridization. Hybrid (aka reticulate) speciation is possible without a change in ploidy (1.5.2.1), but these cases are few in number. More often hybrid speciation involves polyploidy which introduces an immediate breeding barrier between hybrid and parents (1.3.4.2). In either case, hybrid speciation is an example of the rapid origin of new species.

#### 1.5.2.1. Homoploid hybrid speciation

Homoploid hybrid speciation is often called recombinational speciation because of Grant's original model (Grant, 1981). His model depends on alleles of

two BDM incompatibility loci being recombined in the hybrids so that the hybrid population is incompatible with both parents. There are few convincing examples of recombinational speciation, but the case studies in *Helianthus*, *Iris*, and *Senecio* demonstrate the main ideas. The second manuscript of this thesis presents a putative example of homoploid hybrid speciation in wild tomato. However, we are very careful in our interpretation because there are inherent difficulties in distinguishing ancient hybrid speciation from other scenarios like introgressions and substructure using molecular data (1.8.2). These difficulties are discussed in a critical overview of two putative cases of recombinational speciation in *Argyranthemum* and *Pinus* (1.6.5).

There are three sunflower (*Helianthus*, Asteraceae) species of hybrid origin: *H. anomalous*, *H. deserticola*, and *H. paradoxus*. All three species are diploid and derived from two self-incompatible short-lived perennial species common throughout central North America: *H. annuus* (Common Sunflower) and *H. petiolaris* (Prairie Sunflower) (Rieseberg, 1991). *Helianthus annuus* prefers heavier and more mesic soils than *H. petiolaris* which is more often found in drier and sandier soils, but both are occasionally found together (Gross and Rieseberg, 2005). Where they occur together, extensive hybrid zones form and hybridization rates (% interspecific matings) have been estimated at 4-15% (Rieseberg et al., 1998). Hybrids are semisterile due to >10 chromosomal rearrangements that distinguish the parental species, but later generation recombinant hybrids are often fit and frequently found in the wild (Rieseberg, 1995).

All of the hybrid species occupy extreme habitats relative to the parents: *H. anomalus* is found on sand dunes in northern Arizona and Utah, *H. deserticola* in deserts of Nevada, Utah, and northern Arizona, and *H. paradoxus* in desert salt marshes of New Mexico and Western Texas (Gross and Rieseberg, 2005). These are ancient hybrids that have formed within the last 100,000-160,000 years. Apparently, each formed multiple times (Gross and Rieseberg, 2005). Initial estimates by Ungerer et al. (1998) had *H. anomalous* forming within 60 generations, but Buerkle and Rieseberg (2008) found that it took 100s of generations. In either case, this is rapid speciation relative to homoploid divergence.

Rieseberg et al. (2003) were able to recreate the extreme phenotypes of the hybrids by crossing *H. annuus* and *H. petiolaris* followed by further

generations of intercrossing and backcrossing. The wild hybrids are composed of parental blocks derived from *H. annuus* and *H. petiolaris*. (this is also the case for the novel example of hybrid speciation in wild tomato given in this thesis). Interestingly, the hybrids created by Rieseberg et al. (2003) all appear to have recombinant genotypes similar to the wild hybrid species. These were also compatible with the wild hybrid species *H.* anomalous and had some postzygotic reproductive barriers to the parental species. Together this is evidence that reproductive barriers – which are critical to new hybrid species needing to avoid gene swamping from their parents – can be formed quickly via recombination of incompatibilities (like Grant's model), but that only a few recombinant types will be viable. It also demonstrates that novel phenotypes (and adaptations) can be created via recombining parental traits; the resulting ecological differences are interpreted as an additional reproductive barrier.

The genus *Iris* (Iridaceae) is well known for hybridization (1.6.1), including an example of homoploid hybrid speciation which is interesting because it involves three parental species: *Iris nelsonii* (2n=42) is the hybrid derivative of *I. fulva* (2n=42), *I. hexagona* (2n=44), and *I. brevicaulis* (2n=44). *Iris nelsonii* co-occurs geographically with the parental species, but all of the parents occupy different niches (1.6.1). In particular, *I. nelsonii* occupies deeper waters than any of the other species. There have been morphological (Randolph, 1966), cytological (Randolph and Mitra, 1961), and several molecular studies (Arnold et al., 1990) of *I. nelsonii*. In the study of Arnold et al. (1990), *I. nelsonii* had alleles from both *I. fulva* and *I. hexagona*. Alleles from *I. brevicaulis* were detected later (Arnold, 1993). This case of hybrid speciation has been criticized by Coyne & Orr (2004) because there is no knowledge of reproductive barriers between *I. nelsonii* and the other species.

Hybrid speciation in groundsel (*Senecio*, Asteraceae) demonstrates both opportunism and the importance of an unoccupied habitat for new homoploid hybrid species (Abbott et al., 2010; Brennan et al., 2012). Two short-lived perennial species, *S. aethnensis* and *S. chrysanthemifolius*, have a hybrid zone on the Mount Etna volcano. *Senecio aethnensis* inhabits recent lava flows above 2000 m whereas *S. chrysanthemifolius* occupies habitats below 1000 m. Both species are self-incompatible and highly interfertile. They form occasional hybrids,

but these are unfit at the intermediate altitudes where they occur (Brennan et al., 2012).

Following the introduction of hybrid individuals to Britain, however, they began reproducing inter se and occupying disturbed habitats like railroads, roadsides, etc. These new populations are evolutionarily independent of their parents (now allopatric) and have the rank of species: *S. squalides*. There is reduced allelic diversity in *S. squalides*, indicating a founder event; and this species is distinct from all Mt. Etna hybrids examined, indicating extensive drift and/or adaptive evolution in about 300 generations. Drift may actually be an important feature in the founding of all hybrid species (Templeton, 2008). Importantly, several quantitative traits in *S. squalides* are transgressive relative to both parental species. This apparently includes novel adaptations for tolerating drought and temperature stress (Allan and Pannell, 2009). Overall, the evidence that hybridization in *Senecio* has led to a new and uniquely adapted species is quite good. Hybridization between *S. squalides* and another *Senecio* species is also the best example in plants of adaptive traits being transferred between species (1.6.2)

Overall hybrid speciation without a change in ploidy is much less common than allopolyploid speciation (Schumer et al., 2014; Soltis and Soltis, 2009). There are only 15-20 documented cases, and few of these have molecular data (Rieseberg and Willis, 2007). This is probably because this type of speciation is underestimated (i.e. there is no cytological evidence and drift/mutation obscure genetic evidence quickly when there is a founder event) and because it is theoretically more difficult to form a homoploid hybrid species (i.e. no immediate reproductive barriers).

#### 1.5.2.2. Polyploid speciation

Polyploidy is widespread in plants with about 34.5% of species polyploidy relative to their generic base and 13% of species harboring multiple cytotypes (Soltis et al., 2007; Wood et al., 2009). Allopolyploid speciation is also common in plants. Otto and Whitton (2000) estimated that a minimum of 2-4% of speciation events in angiosperms and 7% in ferns were due to polyploidy. Wood et al. (2009) estimate 15% of speciation events in angiosperms and 31% in ferns are due to polyploidy. All of the cases of hybrid speciation identified in the first manuscript of

this thesis involve polyploidy (p. 81). There are two important features of polyploid speciation: it is rapid and most polyploidy species have multiple origins (Soltis and Soltis, 2009). Allopolyploid speciation thought to be more common than autopolyploid speciation, and rates of interspecific hybridization are therefore fundamental to understanding rates of polyploid speciation.

Meiotic nonreduction is the most common mechanism for polyploid formation (Coyne & Orr 2004, p. 325), but there are other mechanisms, including the triploid bridge. In most cases, the homeologous genomes are divergent enough to have diploid behavior during meiosis (these are termed amphidiploids). It appears that greater genome divergence may even drive polyploid formation following hybridization (Paun et al., 2009).

Whole genome duplication (WGD) has extensive evolutionary implications, including epigenetic changes and the resulting process of diploidization. Functional redundancy from a WGD also results in neofunctionalization and gene loss. This is akin to processes following individual gene duplication but for the entire genome. Polyploid species also undergo a strong bottleneck during their formation and frequently transition to autogamy (Douglas et al., 2015).

One well-documented example of allopolyploid speciation is in *Spartina* (Poaceae). The two parental species have repeatedly formed independent homoploid hybrids. In one instance, these homoploid hybrids have undergone genome doubling and begun reproducing inter se. This new species is now vigorous in salt marshes and also spreads clonally. The parental species are themselves hexaploid, so the new allopolyploid is a dodecaploid. This species underwent a strong bottleneck accompanied by gene loss within the last 150 years (Salmon et al., 2005). Section 1.6.6 examines the well-known instance of polyploidy speciation in *Tragopogon*.

## 1.6. Examples of Wild Plant Hybridization

*The guiding forces which have determined the direction of evolutionary trends in higher plants are more likely to be clarified by careful detailed explorations.*

G. L. Stebbins (1967 p. 104)

The following case studies of hybridization demonstrate the features (1.4), outcomes (1.5), and idiosyncrasies of wild plant hybridization.

### 1.6.1. Ecology of hybridization in Iris

One of the best-studied examples of hybridization is in Louisiana *Iris.* It is interesting that, following the recognition of hybridization in this clade, the 80 named species were reduced to four species and their hybrids (Judd et al., 2016). Two of these species, *Iris fulva* and *I. hexagona*, are very easily distinguished. They form fertile hybrids in the Mississippi delta, but hybrid establishment is highly dependent on habitat disturbance.

The facts are as follows. *Iris fulva* inhabits natural upland levees while *I. hexagona* inhabits submerged waters in neighboring marshes. Both species are highly clonal. They form extensive hybrid swarms, but only in mad-made pastures and cleared woodlands bordering the natural habitat of the parental species (Anderson, 1949). Both species also hybridize with *I. brevicaulis* (Hamlin and Arnold, 2014).

Hybrid zygotes between *I. fulva* and *I. hexagona* must always form at some rate, but habitat is a very strong postzygotic barrier that inhibits hybrid establishment. When the barrier is removed (i.e. formation of an open, “hybridized” habitat), hybrids do establish. This influential case demonstrates the importance of habitat disturbance in hybridization. It also demonstrates that hybridization does not necessarily equate to a breakdown of reproductive barriers or disintegration of species (Arnold et al., 1999). Finally, it demonstrates that unrecognized hybridization can confuse taxonomists using morphology.

### 1.6.2. Transfer of adaptations between species in *Senecio*

It is now relatively easy to detect introgression using genomic data (1.8.2), but the demonstration that introgression results in adaptive changes is much more

difficult. Members of the Compositae are characterized by capitulate inflorescences composed of many individual flowers (florets). In radiate Compositae, the outermost florets have bilateral symmetry (Figure 2f, p. 13). *Senecio vulgaris* (Groundsel, Asteraceae) is a non-radiate species, but some British populations have partially radiate phenotypes. These radiate morphs have a higher outcrossing rate making the trait of evolutionary importance (Marshall and Abbott, 1984). *Senecio vulgaris* is an allotetraploid and occasionally hybridizes with *Senecio squalides*, a diploid radiate species (itself of hybrid origin, 1.5.2.1). Their triploid hybrids have low fertility (<0.2%) (Lowe and Abbott, 2000), but Ingram et al. (1980) nevertheless hypothesized that radiate *S. vulgaris* populations gained the phenotype via introgressions from *S. squalides.*

The radiate phenotype is ancestral and controlled by the *RAY* locus (Kim et al. 2009). The two *RAY* locus genes (*RAY1* and *RAY2*) are regulatory and expressed only in the outermost florets of radiate species (R/R). The non-radiate alleles are derived and semi-dominant.

Kim et al. 2008 sequenced the *RAY1* and *RAY2* loci from wild populations of *S. vulgaris* and *S. squalides*. They found two *RAY* haplotypes in *S. squalides* populations. One of these was identical to the *RAY* haplotype in radiate *S. vulgaris*. They interpret this as evidence that the radiate *S. vulgaris RAY* haplotype is the result of introgression. Furthermore, they are able to recreate partially radiate *S. vulgaris* in the lab by crossing non-radiate populations to *S. squalides*.

However, the allele found in both *S. vulgaris* and *S. squalides* could be an ancestral polymorphism in *S. vulgaris* (it was also in other *Senecio* species). The situation is further confused by having four alleles in tetraploid *S. vulgaris*. Nevertheless, the case of radiate *S. vulgaris* is the best evidence to date in plants that “key morphological and ecological traits controlled by regulatory genes may be gained, lost, and regained during evolution” (Kim et al. 2000, p. 1116).

### 1.6.3. Biogeography of a hybrid zone

North American *Juniper* is a good example of a geographic hybrid zone. Creeping Juniper (*Juniperus horizontalis*), Red-Cedar (*J. virginiana*), and Rocky Mountain Juniper (*J. scopulorum*) are three widespread and closely related species common throughout North America (Figure 3). Most populations of these species are allopatric, but taken together the species are parapatric. The species have distinct forms and plant habits: *J. virginiana* and *J. scopulorum* are both upright shrubs/small trees whereas *J. horizontalis* is a creeping subshrub. The more boreal *J. horizontalis* grows on rock outcrops and sandy/rocky shores while the other species prefer more open forest habitats.

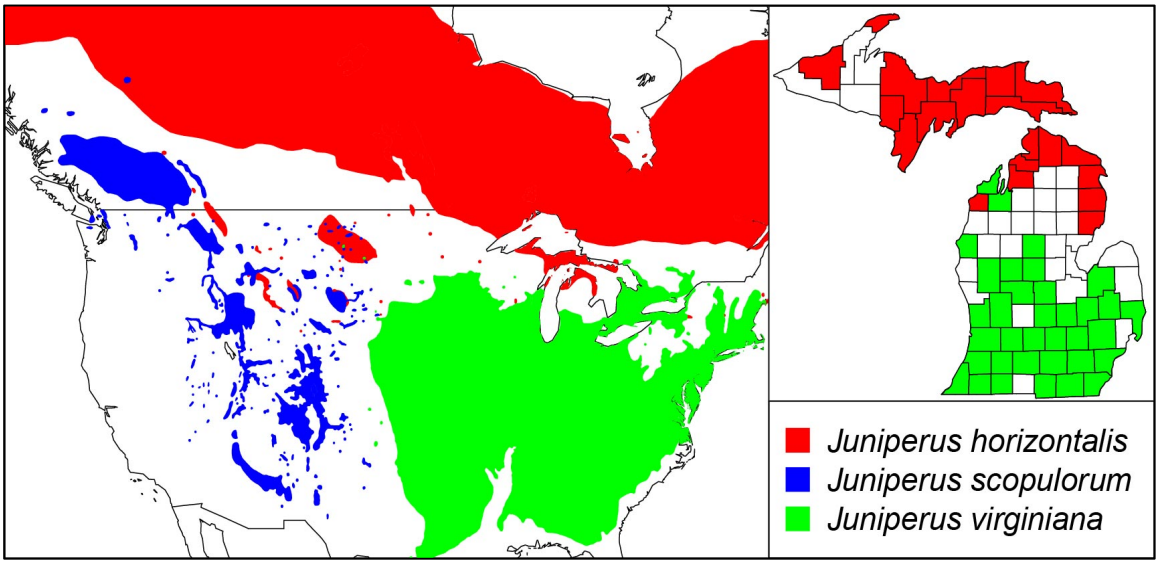



**Figure 3** The distribution of three parapatric *Juniperus* species in North America, including a more detailed view of the distribution of *J. horizontalis* and *J. virginiana* in Michigan.

All three species hybridize with one another, but their biogeographical differences inhibit hybridization for nearly all of their populations. Hybrids between *J. virginiana* and *J. horizontalis* form a hybrid zone in Wisconsin (Palmaotal et al., 1983). *Juniperus virginiana* is also said (from morphology) to have introgressions from *J. scopulorum* from Western North America (Grant 1981 p. 217). *Juniperus horizontalis* and *J. scopulorum* also hybridize. The hybrids only occur in hybrid zones and geography is a fundamental prezygotic barrier maintaining species integrity. Where populations are geographically sympatric, their habitat differences may further inhibit hybridization. Hybrids are not known, for example, in Michigan despite both species occurring together in the upper lower peninsula. Overall,

interspecific hybridization is probably relatively unimportant for allele frequencies in most *Juniperus* populations.

It is also interesting that northern populations of *J. virginiana* commonly grow in old fields and dry uplands while southern populations prefer sand dunes and coastal river sandbanks. There are no obvious morphological differences between these populations, but their habitat preferences indicate different adaptive norms and the formation of geographic races (Judd et al. 2016, p. 128). This intraspecific variation (i.e. geographic races) further muddles hybridization in the group.

### 1.6.4. Species integrity *Quercus*

Grant (1981) uses North American *Quercus* (Fagaceae) as an example of a syngameon. He defines a syngameon as "the most inclusive unit of interbreeding in a hybridizing species group" (Grant 1981, p. 234). The idea being that the individual units are identifiable morphological species, but together the syngameon behaves as a biological species. All of the widespread *Quercus* species are diploid, outcrossing, and wind-pollinated, and 14/16 North American oaks hybridize with other sympatric species. However, it is now evident that oaks maintain their integrity as individual species and are therefore both definable under the BSC and may not actually be a 'worst-case scenario' as termed by Coyne & Orr 2004, p. 43. Furthermore, as I show in Chapter 2, the typical hybridizing species has only one interspecific partner, and this indicates that syngameons as defined by Grant are exceedingly rare in the wild.

Whittemore and Schaal (1991) found high allele sharing in sympatric oak using cpDNA markers. They found no allele sharing between allopatric populations, and this is therefore good evidence of hybridization followed by local introgression. Moran et al. (2012) report that >20% of seeds are of hybrid origin in other sympatric oak species. Other studies have estimated the hybridization rate at 2-30% in *Quercus* (% zygotes formed from interspecific matings) (Cavender-Bares and Pahlich, 2009; Curtu et al., 2007; Lepais et al., 2009). Thus, oak species do hybridize at a fairly high rate.

However, postzygotic barriers appear to be maintaining species integrity in *Quercus*. First, Abadie et al. (2012) and others have made the observation that many first-generation oak hybrids have low fitness. This idea is demonstrated in

the field by Lagache et al. (2014) in two species of European oak that form extensive hybrids: *Q. petraea* and *Q. robur*. They find that the observed rate of hybridization is different for different life stages: rates are higher if seeds are measured and lower if adults are measured. This indicates a decrease in the hybrid cohort as the trees mature. Most interestingly, Lagache et al. (2014) explore the idea that community composition affects hybridization rate. *Quercus robur* is a smaller and early successional tree whereas *Q. petraea* is found in mature forests. Thus, as the forest matures, the frequency of the trees changes along with the frequency of con-and heterospecific mates. This, in turn, results in differential rates of hybridization.

Another dynamic example of hybridization can be found in *Populus*. In a hybrid zone between *P. alba* and *P. tremula*, Lindtke et al. (2014) report many hybrid seeds and mature first-generation hybrids. However, they find few adult backcrossed individuals. This again indicates that, like in oaks, postzygotic barriers maintain species integrity. Furthermore, Muir et al. (2000) found that *Q. robur* and *Q. petraea* were discrete at many loci. Indeed, much of the observed 'hybridization' may be due to other evolutionary processes like ancestral polymorphisms and large population sizes (Muir and Schlotterer, 2005).

In chapter two, my results indicate that trees (and shrubs) have high rates of hybridization. Trees define their ecological communities, and are known to have slow rates of molecular evolution (Kay et al., 2006; Smith and Donoghue, 2008); thus, their inability to form hybrids (which is clock-like, Mallet 2005) may take longer to evolve. This has important implications for the phylogenetic distribution and rate of hybridization. These ideas are discussed in more detail in 2.6.5 and 2.6.6. In conclusion, the total evidence indicates postzygotic reproductive barriers are strong enough to maintain species integrity in oaks even with a fairly high rate of hybridization and local introgressions. The situation is probably similar in other well-known syngameon trees like *Populus* and *Salix*.

### 1.6.5. Homoploid hybrid speciation in *Pinus* and *Argyanthemum*

The two cases briefly overviewed here are often referenced as good examples of homoploid hybrid speciation (Coyne and Orr, 2004; Gross and Rieseberg, 2005; Paun et al., 2009). However, both fail to meet many of the criteria to be a definitive case outlined by Schumer et al. (2014). These criteria are

reproductive isolation, genetic evidence of hybridization, and demonstrating that hybridization resulted in genetic isolation.

The first example involves *Argyanthemum* (Asteraceae) from the Canary Islands. Two closely related self-incompatible species occupy coastal xerophytic (*A. frutescens*) or upland forest (*A. broussonetii*) habitats on the same islands (Brochmann et al., 2000). Hybrid swarms between these two species were observed in a recently deforested area and there are few postzygotic barriers between them ($F_1$ hybrids have do have somewhat reduced fertility). There are two known populations in two different valleys of a third species, *A. sundingii*, which is hypothesized on the basis of intermediate morphology to be of hybrid origin (hybrid of *A. frutescens* and *A. broussonetii*).

Brochmann et al. (2000) quantified the morphology of two populations of *A. sundingii* using multivariate analysis, and determined them to be intermediate between the putative parents. One of these populations had the *A. frutescens* cpDNA and one had *A. broussonetti*. From this evidence Brochmann et al. (2000) conclude the two populations have independent origins but represent the same hybrid species. However, they did not test the compatibility of the two populations with one another nor *A. sundingii*'s compatibility to its putative parents. They did 'recreate' *A. sundingii* in that hybrid derivates from crosses between the parents resembled the two *A. sundingii* populations (some of the plants from the hybrid swarm are also said to resemble *A. sundingii* but this was not quantified). Finally, from the low morphological variation within the *A. sundingii* populations, they conclude that the hybrids are genetically stabilized. In conclusion, *A. sungingii* does appear to be of hybrid origin, but without any evidence of reproductive isolation it is premature to conclude that it is a hybrid species.

The second putative case of homoploid hybrid speciation involves *Pinus*. It was suggested on the basis of morphology that *Pinus densata* arose via hybridization from two other mostly allopatric species: *P. tabuliformis* and *P. yunnanensis* (Wang et al., 2001; Wu, 1956;  ing et al., 2014). All three species occur in the mountains of the Tibetan plateau, but *Pinus densata* is found at higher elevations than the other species and is more adapted to the cold (Ma et al., 2010). These species are all diploid (2n=2x=24), wind-pollinated, long-lived trees. Wang et al. (2001) found that 9 of 12 loci analyzed from 14 populations of the three species were polymorphic. Interestingly, populations of the putative

hybrid have the highest heterozygosity. They hypothesize from their data that *P. densata* is both multiply derived and millions of years old. However, they also report that the *P. densata* population sympatric with *P. yunnanensis* (Pd-7) has the most allele sharing with *P. yunnanensis*: this indicates on-going gene flow and introgression, and their data is not sufficient to support a new hybrid species.

Ma et al. (2006) sequenced 7 haplotype loci from female gametophytes and found that many of the haplotypes were shared among all species (akin to the finding of Wang et al. 2001). Wang et al. (2011) found unique alleles in *P. densata* and genetic subdivision in this species by sequencing cytoplasmic DNA in 54 populations of *P. densata* and its putative parents. Gao et al. (2012) hypothesize an origin about 6 million years ago for *P. densata* using coalescent modeling on the data of Wang et al. (2011). Taken together, despite the repeated claim that "genetic analyses indicate that *Pinus densata* is a natural homoploid hybrid originating from *Pinus tabuliformis* and *Pinus yunnanensis*" ( ing et al. 2014 p. 1890) and others, there is not really any evidence to differentiate between ancient homoploid hybridization, ongoing introgression, or a history of homoploid divergence without hybridization. Indeed, the allele sharing in sympatric populations is good evidence for current introgressions. This may well be a true case of hybrid speciation, but the repeated failure to consider and test alternative demographic histories makes this, in my opinion, a doubtful case of homoploid hybridization.

### 1.6.6. Allopolyploid hybrid speciation in *Tragopogon*

Goat's Beards (*Tragopogon*, Asteraceae) are a good example of a multiple-origin polyploid species. Three diploid (2n=2x=12) species were introduced to North America around 1900: *T. dubius*, *T. porrifolius*, and *T. pratensis*. Hybridizations between *T. dubius* and the other two species in Washington have resulted in two allotetraploid species: *T. miscellus* (*T. dubius* × *T. pratensis*) and *T. mirus* (*T. dubius* × *T. porrifolius*). These new species are independent because backcrosses to their parents are triploid and mostly sterile. This case has been extensively characterized using morphology, genetic markers and cytological studies (reviewed in Symonds et al. 2010). Both allopolyploid species have formed multiple times (11 times for *T. muris* and 21 times for *T. miscellus*) (Soltis et al., 2004). The maternal parent of *T. muris* is *T. porrifolius* in all of the known origins

(i.e. cytoplasmic DNA always from *T. porrifolius*). While there is no directionality in the other case, the hybrids do have different phenotypes depending on the maternal species.

This case of hybrid speciation is especially interesting in the context of hybridization between *T. pratensis* and *T. porrifolius* in Europe. Their wild hybrid, *T.* ×*mirabilis*, has been documented in Britain and elsewhere in Europe for over 250 years (Clausen, 1966; Matthews et al., 2015). These species have not, however, formed any new species of hybrid origin. Matthews et al. (2015) sampled populations of these species and their putative hybrids in England and used nuclear and cytoplasmic markers to identify *T. pratensis, T. porrifolius,* $F_1$ hybrid, and backcrossed individuals. In total, they sequenced 15 *T. porrifolius*, 14 *T. pretense,* 6 hybrids, and one backcross using both nuclear and cytoplasmic markers. They demonstrate that hybrids have 90% aborted seeds (25% and 8% aborted seeds in the parents) and that *T. pratensis* is the maternal parent in all cases. Importantly, these two species are more closely related to one another than to *T. dubius*, and allopolyploids are known to form more often between distantly related species while homoploid hybrid species result from more closely related species (Paun et al., 2009). In the absence of tetraploidy, *T. pratensis* × *T. porrifolius* hybrids have no intrinsic reproductive barrier to their parents and, being less fertile and with no ecological difference, have failed to become independent.

## 1.7. Frequency of Natural Hybridization

### 1.7.1. Motivations of my research on this topic

The outcomes of hybridization and the individual case studies establish that hybridization can be an important evolutionary force with diverse outcomes. However, the fraction of plant species that hybridize in the wild remains a relatively open question. This is an important question because hybridization rate determines its importance as an evolutionary force and can help answer related questions like the rate of allopolyploid speciation, the rate of divergent speciation, and the reliability of the BSC. The first manuscript of this thesis, "*Factors determining hybridization rate in plants*," estimates the frequency of natural hybridization of vascular plants. This estimation is followed by a novel approach to test the effect of various ecological and physiological factors on the rate of hybridization.

As discussed in 1.2, many factors can influence the genetic structure and hybridization rate between populations and species. Several ecological factors (e.g. natural disturbances) are said to be important and repeatedly occur in the case studies, but their importance has never been systematically quantified. Other physiological factors (e.g. being a tree) are also frequent in the case studies, but the statistical association of physiological factors and hybridization propensity has never been directly tested. To my knowledge, this is the *first systematic study to directly test any ecological or physiological factors on hybridization rate*.

### 1.7.2. Outcomes of my research on this topic

In this study, we first quantify all hybridizing species in the "Field Manual of Michigan Flora" and their unique hybrids (Reznicek & Voss 2012). We then statistically test if life form, life history, reproductive mode, phylogeny, and habitat disturbance are significant associated with hybridization rate. The manuscript has the following key findings:

- 17.5% of vascular plant species hybridize in Michigan
- 0.127 unique hybrids per non-hybrid species
- There are significantly more herbarium collections for hybridizing species
- Dioecy is not correlated with hybridization

- Hybridization is concentrated in certain taxonomic groups, but no strong correlations between phylogeny and hybridization rate are detected
- Trees, shrubs, and fern-allies have the highest rates of hybridization
- 97% of hybridizing species are perennial
- Hybridization rates are lower in natural remnant communities and higher in disturbed habitats

Many of these findings are novel and of interest. For example, no study has analyzed potential bias in hybridization rates by investigating the number of herbarium collections for hybridizing vs. non-hybridizing species. Furthermore, this study tests and compares multiple metrics to estimate hybridization frequency (a recommendation of Whitney et al. 2010, p. 176). The results are followed by a careful discussion of the frequency of natural hybridization and the factors affecting it including rates of evolution and the interaction of ecology and physiology. Finally, the extensive documentation of all hybridizing taxa by the authors will facilitate future research on hybrids.

## 1.8. Case Study of Hybridization in Wild Tomato

The second publication included in this thesis, 'Population genomics reveals a new wild tomato species with a history of hybridization', is a case study of hybridization, taxonomy, and population genetics in wild tomato. The cultivated tomato (*Solanum lycopersicum*) is of significant economic importance, but like all crops has undergone an extreme bottleneck following domestication and improvement (Bevan et al., 2017; Lin et al., 2014). The wild species are a reservoir of allelic diversity that can been used for improvement via introgressions into tomato breeding populations. For example, two disease resistance alleles, *Ty-1* and *Tm-2*, have been introgressed from *S. chilense* and *S. peruvianum* respectively (Lin et al., 2014). Most of these breeding efforts rely extensively on the work of C. Rick and the wild germplasm available to researchers from the Tomato Genetics Resource Center at UC Davis. One idea is to identify adaptive alleles and potentially adaptive population substructure in this germplasm that can inform and guide breeding efforts. However, this depends on having a good taxonomic and population genetic understanding of the wild tomato species.

### 1.8.1. Wild tomato taxonomy

There are 12 wild species of tomato native to the western coast of South America. Relationships within the wild tomato clade (*Solanum* sect. Lycopersicon; Solanaceae) have been tested using nearly all technologies as they developed (Aflitos et al., 2014; Alvarez et al., 2001; Breto et al., 1993; Labate et al., 2014; Marshall et al., 2001; McClean and Hanson, 1986; Miller and Tanksley, 1990; Palmer and Zamir, 1982; Pease et al., 2016; Peralta and Spooner, 2005; Spooner et al., 1993; Zuriaga et al., 2009). At the same

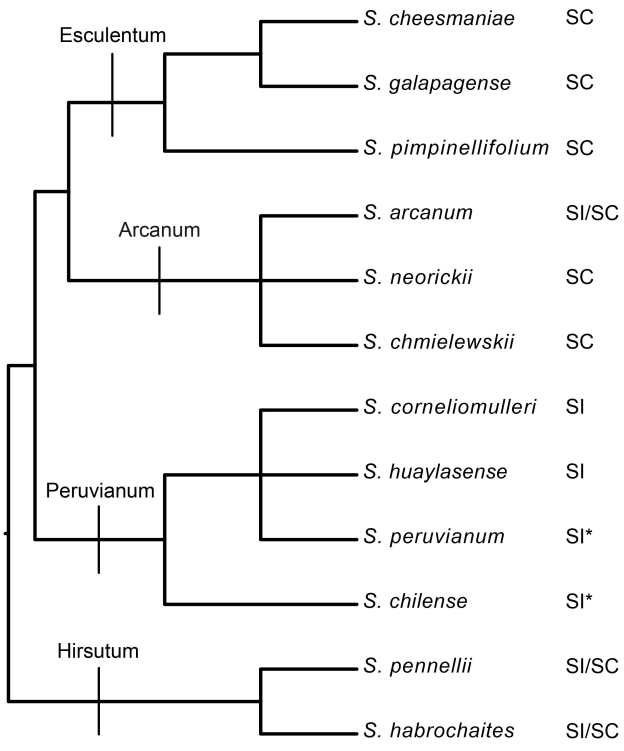



**Figure 4** Current relationships in the wild tomato clade. SC, self-compatible; SI, self-incompatible. An * indicates that SC population of these species were identified in this study.

time, biosystematics studies attempted to delineate species based on the BSC (Rick, 1963, 1979, 1986; Rick and Lamm, 1955). Other studies focused on identifying species and relationships from morphological differences (Darwin et al., 2002; Peralta et al., 2005; Peralta and Spooner, 2005).

These studies have returned a relatively good idea of the relationships within the clade with four informal species groups well-supported: Esculentum, Arcanum, Peruvianum, and Hirsutum (Figure 4). There are, however, a few unresolved relationships, including two new species that were recently circumscribed out of the highly variable *S. peruvianum* (see 3.6.2 and 3.6.3).

### 1.8.2. Modern methods and tools for studying hybridization

For most of history, hybrids have been described and quantified from morphology. In fact, R. A. Fisher developed multivariate analysis for analyzing hybridization in *Iris* (Judd et al. 2016). Morphological descriptions of hybrids remain common today, but the first thing a hybrid normally does is backcross to one of its parents. This makes identification difficult if, for example, there are only backcrossed individuals (which resemble the parents after only a few generations of backcrossing) or impossible if no hybrids exist at the time of observation/collection. That said, for those hybrids that are identified, the parentage is usually confirmed with molecular data (Cronn and Wendel, 2004; Matthews et al., 2015) (but see Rieseberg et al. (1988) for a counterexample).

Genomic data has greatly expanded the repertoire of tools available to study hybridization. Moreover, footprints of past hybridization can be found in the genome of modern species. That said, recombination and mutation obscure evidence of hybridization in the genome relatively quickly (Moody and Rieseberg, 2012). There are a few methods to infer hybridization from the genetic data, even if only one individual genome per species is available. First, discordant phylogenies between markers in the nuclear and cytoplasmic genomes remain good evidence of hybridization. Second, the ABBA-BABA test (aka the D-statistic) developed by Patterson et al. (2012) detects introgressions in a population using two outgroups and two ingroup species (reviewed in Racimo et al. 2015).

There are more methods available to infer hybridization when population genetic data is available. These methods usually infer hybridization by analyzing allele frequencies between the hybridizing populations. First, the model-based

clustering algorithms of STRUCTURE/fineSTRUCTURE/ADMI TURE are widely used. These find genotype cluster from population data and assign individuals to those clusters. Admixed assignment is often interpreted as hybridization (e.g. Hamlin et al. 2014 in *Iris*), but this is not definitive because population subdivision and other processes also create 'admixed' individuals (Falush et al., 2016). The $F_3$ statistic was developed by Patterson et al. (2012), and basically detects allele frequencies intermediate between the putative parental populations in the putative hybrid population. Hybrids are expected to have intermediate frequencies over many loci. Given that most new hybrid species have a strong founder effect (Templeton 2008), drift can quickly obscure this type of evidence. Thus, while the $F_3$ is statistical evidence of hybridization, it is only good for detecting recent hybridization events and has been mostly used in humans.

Other interesting approaches exploit inferences about the behavior of neutral alleles within a hybridizing population. One fact is that the probability of an allele being introduced via introgression is dependent on its frequency in the donor population. Thus, derived alleles of high frequency are expected to exhibit greater sharing (de Manuel et al., 2016). Furthermore, donated segments of the genome should have low divergence to the donor species and relatively high heterozygosity in the recipient population. Together these criteria can be used to identify introgressions, for example bonobo haplotypes in chimpanzee (de Manuel et al., 2016).

Haplotype-based methods try to identify parental haplotypes in introgressed populations. This is, however, different than the D-statistic because multiple individuals are needed (e.g. HAPMI ). Using haplotypes, it is also possible to estimate the age of hybridization from haplotype lengths with genomic data (because recombination breaks up haplotypes in a clock-like manner). In plants, this was first done with a few molecular markers by Ungerer et al. (1998) and Buerkle and Rieseberg (2008). However, the genomic data now available allows more accurate inference (e.g. ARGweaver) (Rasmussen et al., 2014).

In conclusion, there are many new and powerful methods to explore hybridization and introgression using genomic data. Many of these methods have been applied in the study of wild tomato hybridization.

### 1.8.3. Motivations and outcomes of my research

This is the first comprehensive genetic analysis of all wild tomato species to include a detailed sampling of the most polymorphic species *S. peruvianum* and *S. chilense*. This work has allowed several questions about the evolutionary history of this clade to be definitively answered. These answers are relevant and necessary for the hundreds of studies that work on these wild species every year.

This work first resolves relationships in the tomato clade; dates the speciation event of *S. chilense* and *S. peruvianum* to circa 1.25 million years ago; estimates the current effective population sizes and historical migration rates of these species; documents extensive population structure in *S. peruvianum* that probably arose from ecological adaptation; provides evidence that two taxa currently recognized at the rank of species are neither monophyletic nor good biological species; and re-annotates one very northern accession as *S. chilense*.

Finally, we present extensive evidence for a new species of putative hybrid origin. This is an exciting discovery because there are only a few examples where natural hybridization generates a new entity. This discovery was followed up with extensive genomic analyses and experimental crossing studies testing compatibility of the putative hybrid populations to their parental species. It also revealed how cryptic hybrids have misled previous work in this clade regarding the speciation of *S. chilense* and *S. peruvianum*. The extensive analyses and data in this work can be built upon to both identify adaptive changes in the subpopulations of *S. peruvianum* and to explore the history of the putative hybrid species.

## 1.9. Summary of the Scientific Contributions of this Thesis

Hybridization is a force of evolution: it can introduce reticulate events into phylogenies, transfer potentially adaptive alleles between species, and result in the origin of new species. This thesis makes two significant contributions to the field of plant hybridization. First, the frequency of natural plant hybridization is estimated and a number of physiological and ecological factors are found to be significantly associated with hybridization rate. Second, using an extensive genomic dataset, relationships within the tomato clade are resolved, including extensive population structure in the polymorphic species *S. peruvianum* and the discovery of a putative species of hybrid origin.

## 1.10. References


Abadie, P., Roussel, G., et al., 2012. Strength, diversity and plasticity of postmating reproductive barriers between two hybridizing oak species (Quercus robur L. and Quercus petraea (Matt) Liebl.). J. Evol. Biol. 25, 157-173.

Abbott, R.J., Hegarty, M.J., et al., 2010. Homoploid hybrid speciation in action. Taxon 59, 1375-1386.

Aflitos, S., Schijlen, E., et al., 2014. Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. Plant J. 80, 136-148.

Allan, E., Pannell, J.R., 2009. Rapid divergence in physiological and life-history traits between northern and southern populations of the British introduced neo-species, Senecio squalidus. Oikos 118, 1053-1061.

Alvarez, A.E., van de Wiel, C.C.M., et al., 2001. Use of microsatellites to evaluate genetic diversity and species relationships in the genus Lycopersicon. Theor. Appl. Genet. 103, 1283-1292.

Anderson, E., 1949. Introgressive hybridization. John Wiley and Sons, NY.

Anderson, E., 1953. Introgressive hybridization. Biol. Rev. Camb. Philos. Soc. 28, 280-307.

Arnold, M.L., 1993. Iris nelsonii (Iridaceae): origin and genetic composition of a homoploid hybrid species. Am. J. Bot., 577-583.

Arnold, M.L., Bennett, B.D., et al., 1990. Natural hybridization between Iris fulva and Iris hexagona: pattern of ribosomal DNA variation. Evolution 44, 1512-1521.

Arnold, M.L., Bulger, M.R., et al., 1999. Natural hybridization: how low can you go and still be important? Ecology 80, 371-381.

Barrett, S., Eckert, 1990. Variation and evolution of mating systems in seed plants. Biological Approaches and Evolutionary Trends in Plants, 229-254.

Baum, D.A., Donoghue, M.J., 1995. Choosing among alternative phylogenetic species concepts. Syst. Bot., 560-573.

Bevan, M.W., Uauy, C., et al., 2017. Genomic innovation for crop improvement. Nature 543, 346-354.

Bomblies, K., Lempe, J., et al., 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biol. 5, 1962-1972.

Brennan, A.C., Barker, D., et al., 2012. Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of Senecio squalidus (Asteraceae). Heredity 108, 87-95.

Breto, M.P., Asins, M.J., et al., 1993. Genetic variability in Lycopersicon species and their genetic relationships. Theor. Appl. Genet. 86, 113-120.

Brochmann, C., Borgen, L., et al., 2000. Multiple diploid hybrid speciation of the Canary Island endemic Argyranthemum sundingii (Asteraceae). Plant Syst. Evol. 220, 77-92.

Brochmann, C., Brysting, A.K., et al., 2004. Polyploidy in arctic plants. Biol. J. Linn. Soc. 82, 521-536.

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

Carney, S.E., Gardner, K.A., et al., 2000. Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (Helianthus). Evolution 54, 462-474.

Cavender-Bares, J., Pahlich, A., 2009. Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, Quercus virginiana and Q. geminata (Fagaceae). Am. J. Bot. 96, 1690-1702.

Clausen, J., 1966. Stability of genetic characters in Tragopogon species through 200 years, Transactions of the Botanical Society of Edinburgh. Taylor & Francis, pp. 148-158.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland, MA.

Cronn, R., Wendel, J.F., 2004. Cryptic trysts, genomic mergers, and plant speciation. New Phytol. 161, 133-142.

Curtu, A.L., Gailing, O., et al., 2007. Evidence for hybridization and introgression within a species-rich oak (Quercus spp.) community. BMC Evol. Biol. 7, 218.

Darwin, C., 1859. On the origin of the species by natural selection.

Darwin, S.C., Knapp, S., et al., 2002. Taxonomy of tomatoes in the Galápagos Islands: native and introduced species of Solanum section Lycopersicon (Solanaceae). Syst. Biodivers. 1, 29-53.

de Manuel, M., Kuhlwilm, M., et al., 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science 354, 477-481.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879-886.

Donoghue, M.J., 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist, 172-181.
Douglas, G.M., Gos, G., et al., 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid Capsella bursa-pastoris. Proc. Natl. Acad. Sci. USA 112, 2806-2811.
Ehrlich, P.R., Raven, P.H., 1969. Differentiation of populations. Science 165, 1228-1232.
Falush, D., van Dorp, L., et al., 2016. A tutorial on how (not) to over-interpret STRUCTURE/ADMI TURE bar plots. bioRxiv, 066431.
Focke, W., 1890. Versuche und Beobachtungen über Kreuzung und Fruchtensatz bei Blutenpflanzen. Abhandlungen herausgegeben vom Naturwissenschaftlichen Verein zu Bremen 11, 413-421.
Focke, W.O., 1881. Die Pflanzen-Mischlinge: ein Beitrag zur Biologie der Gewächse. Gebr. Borntraeger.
Gao, J., Wang, B.S., et al., 2012. Demography and speciation history of the homoploid hybrid pine Pinus densata on the Tibetan Plateau. Mol. Ecol. 21, 4811-4827.
Gillespie, J.H., 2004. Population genetics: a concise guide, Johns Hopkins University Press.
Grant, V., 1981. Plant speciation. New York: Columbia University Press.
Gross, B.L., Rieseberg, L.H., 2005. The ecological genetics of homoploid hybrid speciation. J. Hered. 96, 241-252.
Guerrero, R.F., Muir, C.D., et al., 2017. Pervasive antagonistic interactions among hybrid incompatibility loci. PLoS Genet. 13.
Hamlin, J.A.P., Arnold, M.L., 2014. Determining population structure and hybridization for two iris species. Ecology and Evolution 4, 743-755.
Harrison, R.G., 1993. Hybrid zones and the evolutionary process. Oxford University Press on Demand.
Hedrick, P.W., 2011. Genetics of populations. Jones & Bartlett Learning.
Heiser, C.B., 1949. Study in the evolution of the sunflower species Helianthus annuus and H. bolanderi. Univ. Calif. Publ. Bot 23, 157-208.
Huerta-Sanchez, E., Jin, ., et al., 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512, 194-+.
Ingram, R., Weir, J., et al., 1980. New evidence concerning the origin of inland radiate groundsel, S. vulgaris L. var. hibernicus Syme. New Phytol. 84, 543-546.
Judd, W.S., Campbell, C.S., et al., 2016. Plant Systematics: A Phylogenetic Approach, Fourth ed. Sinauer Associates.
Kay, K.M., Whittall, J.B., et al., 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.
Kirschner, J., Štěpánek, J., 1994. Clonality as a part of the evolution process in Taraxacum. Folia Geobotanica 29, 265-275.
Labate, J.A., Robertson, L.D., et al., 2014. Genetic structure of the four wild tomato species in the Solanum peruvianum s.l. species complex. Genome 57, 169-180.
Lagache, L., Klein, E.K., et al., 2014. Distinct male reproductive strategies in two closely related oak species. Mol. Ecol. 23, 4331-4343.
Lepais, O., Petit, R., et al., 2009. Species relative abundance and direction of introgression in oaks. Mol. Ecol. 18, 2228-2242.
Levin, D.A., 1979. The nature of plant species. Science 204, 381-384.
Levin, D.A., 2000. The origin, expansion, and demise of plant species. Oxford University Press on Demand.
Lin, T., Zhu, G., et al., 2014. Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. 46, 1220-1226.
Lindtke, D., Gompert, Z., et al., 2014. Unexpected ancestry of Populus seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. Mol. Ecol. 23, 4316-4330.
Lotsy, J.P., 1916. Evolution by means of hybridization. Martinus Nijhoff, The Hague.
Lowe, A.J., Abbott, R.J., 2000. Routes of origin of two recently evolved hybrid taxa: Senecio vulgaris var. hibernicus and York radiate groundsel (Asteraceae). Am. J. Bot. 87, 1159-1167.
Lowry, D.B., Modliszewski, J.L., et al., 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. Philosophical Transactions of the Royal Society B-Biological Sciences 363, 3009-3021.

Ma, F., Zhao, C.M., et al., 2010. Enhanced drought-tolerance in the homoploid hybrid species Pinus densata: implication for its habitat divergence from two progenitors. New Phytol. 185, 204-216.
Ma, .F., Szmidt, A.E., et al., 2006. Genetic structure and evolutionary history of a diploid hybrid pine Pinus densata inferred from the nucleotide variation at seven gene loci. Mol. Biol. Evol. 23, 807-816.
Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229-237.
Marshall, D., Abbott, R., 1984. Polymorphism for outcrossing frequency at the ray floret locus in Senecio vulgaris L. II. Confirmation. Heredity 52, 331-336.
Marshall, J.A., Knapp, S., et al., 2001. Molecular systematics of Solanum section Lycopersicum (Lycopersicon) using the nuclear ITS rDNA region. Theor. Appl. Genet. 103, 1216-1222.
Matthews, A., Emelianova, K., et al., 2015. 250 years of hybridization between two biennial herb species without speciation. Aob Plants 7.
Mayr, E., 1942. Systematics and the origin of species. Columbia University Press, New York, NY, USA.
Mayr, E., 1982. The growth of biological thought: diversity, evolution, and inheritance. Harvard University Press.
Mayr, E., 1992. A local flora and the biological species concept Am. J. Bot. 79, 222-238.
McClean, P.E., Hanson, M.R., 1986. Mitochondrial DNA sequence divergence among Lycopersicon and related Solanum species. Genetics 112, 649-667.
Medugorac, I., Graf, A., et al., 2017. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nat. Genet.
Miller, J.C., Tanksley, S.D., 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theor. Appl. Genet. 80, 437-448.
Moody, M.L., Rieseberg, L.H., 2012. Sorting through the chaff, nDNA gene trees for phylogenetic inference and hybrid identification of annual sunflowers (Helianthus sect. Helianthus). Mol. Phylogen. Evol. 64, 145-155.
Moran, E.V., Willis, J., et al., 2012. Genetic evidence for hybridization in red oaks (Quercus sect. Lobatae, Fagaceae). Am. J. Bot. 99, 92-100.
Morjan, C.L., Rieseberg, L.H., 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. 13, 1341-1356.
Moyle, L.C., Graham, E.B., 2005. Genetics of hybrid incompatibility between Lycopersicon esculentum and L. hirsutum. Genetics 169, 355-373.
Muir, G., Fleming, C.C., et al., 2000. Taxonomy: species status of hybridizing oaks. Nature 405, 1016.
Muir, G., Schlotterer, C., 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (Quercus spp.). Mol. Ecol. 14, 549-561.
Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401-437.
Palmaotal, M., Moore, W.S., et al., 1983. Morphological, chemical, and biogeographical analyses of a hybrid zone involving Juniperus virginiana and J. horizontalis in Wisconsin. Canadian Journal of Botany 61, 2733-2746.
Palmer, J.D., Zamir, D., 1982. Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. Proc Natl Acad Sci U S A 79, 5006-5010.
Patterson, N., Moorjani, P., et al., 2012. Ancient admixture in human history. Genetics 192, 1065-1093.
Paun, O., Forest, F., et al., 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytol. 182, 507-518.
Pease, J.B., Haak, D.C., et al., 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 14, e1002379.
Peralta, I.E., Knapp, S.K., et al., 2005. New species of wild tomatoes (Solanum section Lycopersicon: Solanaceae) from northern Peru. Syst. Bot. 30, 424-434.
Peralta, I.E., Spooner, D.M., 2005. Morphological characterization and relationships of wild tomatoes (Solanum L. Sect. Lycopersicon). Monographs In Systematic Botany 104, 227-257.
Racimo, F., Sankararaman, S., et al., 2015. Evidence for archaic adaptive introgression in humans. Nature Reviews Genetics 16, 359-371.
Ramsey, J., Schemske, D.W., 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. 29, 467-501.

Randolph, L., 1966. Iris nelsonii, a new species of Louisiana iris of hybrid origin. Baileya 14, 143-169.
Randolph, L., Mitra, J., 1961. Karyotypes of Iris species indigenous to the USSR. Am. J. Bot., 862-870.
Rasmussen, M.D., Hubisz, M.J., et al., 2014. Genome-Wide Inference of Ancestral Recombination Graphs. PLoS Genet. 10.
Reznicek, A.A., Voss, E.G., 2012. Field manual of Michigan flora. University of Michigan Press.
Richards, R.A., 2010. Species problem – a philosophical analysis. Wiley Online Library.
Rick, C.M., 1963. Barriers to interbreeding in Lycopersicon peruvianum. Evolution 17, 216-232.
Rick, C.M., 1979. Biosystematic studies in Lycopersicon and closely related species of Solanum, in: Hawkes, J.G., Lester, R.N., Skelding, A.D. (Eds.), The Biology and Taxonomy of the Solanaceae. Academic Press, London, pp. 667-678.
Rick, C.M., 1986. Reproductive isolation in the *Lycopersicon peruvianum* complex, in: D'Arcy, W.G. (Ed.), Solanaceae Biology and Systematics. Columbia University Press, New York, pp. 477-495.
Rick, C.M., Holle, M., et al., 1978. Rates of cross-pollination in Lycopersicon pimpinellifolium: impact of genetic variation in floral characters. Plant Syst. Evol. 129, 31-44.
Rick, C.M., Lamm, R., 1955. Biosystematic studies on the status of Lycopersicon chilense. Am. J. Bot. 42, 663-675.
Rieseberg, L.H., 1991. Homoploid reticulate evolution in Helianthus (Asteraceae) - evidence from ribosomal genes. Am. J. Bot. 78, 1218-1237.
Rieseberg, L.H., 1995. The role of hybridization in evolution - old wine in new skins. Am. J. Bot. 82, 944-953.
Rieseberg, L.H., Baird, S.J., et al., 1998. Patterns of mating in wild sunflower hybrid zones. Evolution 52, 713-726.
Rieseberg, L.H., Raymond, O., et al., 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301, 1211-1216.
Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5, 65-84.
Rieseberg, L.H., Soltis, D.E., et al., 1988. A molecular reexamination of introgression between Helianthus annuus and H. bolanderi (Compositae). Evolution 42, 227-238.
Rieseberg, L.H., Willis, J.H., 2007. Plant speciation. Science 317, 910-914.
Rieseberg, L.H., Wood, T.E., et al., 2006. The nature of plant species. Nature 440, 524-527.
Roberts, H.F., 1929. Plant hybridisation before Mendel. Princeton University Press.
Salmon, A., Ainouche, M.L., et al., 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Mol. Ecol. 14, 1163-1175.
Schemske, D.W., 2000. Understanding the origin of species. Evolution 54, 1069-1073.
Schumer, M., Rosenthal, G.G., et al., 2014. How common is homoploid hybrid speciation? Evolution 68, 1553-1560.
Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198-207.
Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322, 86-89.
Soltis, D.E., Soltis, P.S., et al., 2004. Recent and recurrent polyploidy in Tragopogon (Asteraceae): cytogenetic, genomic and genetic comparisons. Biol. J. Linn. Soc. 82, 485-501.
Soltis, D.E., Soltis, P.S., et al., 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56, 13-30.
Soltis, P.S., Soltis, D.E., 2009. The role of hybridization in plant speciation. Annu. Rev. Plant Biol. 60, 561-588.
Spooner, D.M., Anderson, G.J., et al., 1993. Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). Am. J. Bot. 80, 676-688.
Stebbins, G., 1967. Adaptive radiation and trends of evolution in higher plants. Evolutionary biology 1, 101-142.
Stebbins, G.L., 1950. Variation and Evolution in Plants, New York.
Stebbins, G.L., 1959. The role of hybridization in evolution. Proceedings of the American Philosophical Society 103, 231-251.
Symonds, V.V., Soltis, P.S., et al., 2010. Dynamics of polyploid formation in Tragopogon (Asteraceae): recurrent formation, gene flow, and population structure. Evolution 64, 1984-2003.
Templeton, A.R., 1989. The meaning of species and speciation: a genetic perspective. The units of evolution: essays on the nature of species, 159-183.

Templeton, A.R., 2008. The reality and importance of founder speciation in evolution. Bioessays 30, 470-479.
Ungerer, M.C., Baird, S.J.E., et al., 1998. Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. USA 95, 11757-11762.
Wang, B.S., Mao, J.F., et al., 2011. Colonization of the Tibetan Plateau by the homoploid hybrid pine Pinus densata. Mol. Ecol. 20, 3796-3811.
Wang, .R., Szmidt, A.E., et al., 2001. Genetic composition and diploid hybrid speciation of a high mountain pine, Pinus densata, native to the Tibetan plateau. Genetics 159, 337-346.
Weber, J.E., Campbell, C.S., 1989. Breeding system of a hybrid between a sexual and an apomictic species of Amelanchier, shadbush (Rosaceae, Maloideae). Am. J. Bot., 341-347.
Whitney, K.D., Ahern, J.R., et al., 2010. Patterns of hybridization in plants. Perspectives in Plant Ecology Evolution and Systematics 12, 175-182.
Whittemore, A.T., Schaal, B.A., 1991. Interspecific gene flow in sympatric oaks. Proceedings of the National Academy of Sciences 88, 2540-2544.
Winge, Ö., 1917. The chromosome. Their numbers and general importance. Compt. Rend. Trav. Lab. Carlsberg. 13, 131-175.
Wood, T.E., Takebayashi, N., et al., 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA 106, 13875-13879.
Wright, S., 1978. Evolution and the genetics of populations, vol. 4. Variability within and among populations. University of Chicago Press, Chicago.
Wu, C.-L., 1956. The taxonomic revision and phytogeographical study of Chinese pines. Acta Phytotaxonomica Sinica 5, 131-163.
ing, F.Q., Mao, J.F., et al., 2014. Needle morphological evidence of the homoploid hybrid origin of Pinus densata based on analysis of artificial hybrids and the putative parents, Pinus tabuliformis and Pinus yunnanensis. Ecology and Evolution 4, 1890-1902.
Zuriaga, E., Blanca, J., et al., 2009. Classification and phylogenetic relationships in Solanum section Lycopersicon based on AFLP and two nuclear gene sequences. Genet. Resour. Crop Evol. 56, 663-678.

# 2. Factors Determining Hybridization Rate in Plants

## 2.1. Overview

*Title:* Factors determining hybridization rate in plants

*Authors:* Ian Beddows[12] and Laura Rose[123]

[1]Institute of Population Genetics,
[2]International Graduate School in Plant Sciences (iGRAD-Plant),
[3]Cluster of Excellence on Plant Sciences (CEPLAS),
Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

*Author for correspondence*:
Ian Beddows
*Tel*: +49 15115852902
*Email*: Ian.Beddows@hhu.de

*Published*: This manuscript is currently under peer review. The manuscript has been reformatted for this thesis.

*Word count*: Abstract 190, Introduction 871, Materials and Methods 1203, Results 1250, Discussion 2661, Conclusions 118, Acknowledgements/Author Contributions 58, **Total 6351**.
*Number of figures*: 5
*Number of tables*: 5
*Number of Supporting Information files*: 1 (Supplemental Figure 1-6; Supplemental Tables 1, 2 and 4). Supplemental Table 3 online only.

*Contribution:*

- Contributed the idea for the study
- Performed all analyses
- Made all figures
- Interpreted data
- Wrote the manuscript

## 2.2. Abstract

Hybridization has long been recognized as an important process in plant evolution, but the factors that determine hybridization propensity are not known. The recent publication of the “Field Manual of Michigan Flora” has made it possible to analyze the extent of hybridization in Michigan’s vascular plants, and – by utilizing information on life history, physiognomy, phylogeny, mating system, and habitat – to test if these factors are associated with hybridization rate. We find that 17.5% of species produce natural hybrids with one or more other species and the ratio of unique hybrids to non-hybrid species was 0.127. We demonstrate that this is an underestimate for the true number of Michigan species which hybridize. Our data did not show a phylogenetic signal for hybridization rate. Nor was there an association between mating system and hybridization rate. However, life history, life form, and habitat disturbance are significantly associated with the hybridization rate. Nearly 97% of hybridizing taxa are perennial. Fern allies (but not ferns) and woody groups (shrubs and trees) are found to have the highest hybridization frequencies. Hybridization is more frequent in disturbed habitats with hybridization rare in high-quality natural plant communities.

*Keywords*: tree, shrub, woody, introgression, ecology

## 2.3. Introduction

Hybridization is generally defined as a cross between two species, but broader definitions also include crosses between genetically differentiated populations of one species (Grant, 1981; Stebbins, 1959). Hybrids were originally associated with sterility and have been derided as misfits, mistakes, and even hopeful monsters (Mallet, 2005; Mayr, 1992; Otto and Whitton, 2000). However, hybrids are often vigorous, fertile, and are of considerable importance to humans and in wild populations. Intraspecific hybridization is critical, for example, to the breeding and production of all important cereal crops (Bevan et al., 2017). Other crops such as seedless watermelon, grapes, and banana are the triploid products of hybridization between a diploid and a tetraploid. In addition, disease and drought resistance alleles have been reintroduced into the breeding populations of tomato, wine grapes, and many others via interspecific hybridization with wild relatives (Di Gaspero et al., 2012; Lin et al., 2014). Finally, many wild crop relatives themselves have a relatively recent history of natural hybridization (e.g. maize, tobacco, potato, cotton, *Citrus*, strawberry, peanut, sweet potato, blueberry and coffee) (Bevan et al., 2017; Cronn and Wendel, 2004; Judd et al., 2016; Maurin et al., 2007; Shulaev et al., 2011; Wu et al., 2014).

Besides crop relatives, hybridization is an important feature in the history of *most* wild plant species. Foremost, all angiosperms share a whole genome duplication event which probably resulted from an ancient interspecific hybridization (Cui et al., 2006). And nearly 4-15% of speciation events in angiosperms and 7-31% in ferns are associated with allopolyploidy (Otto and Whitton, 2000; Wood et al., 2009). Hybridization can both transfer adaptations between species and result in new adaptations through allelic recombination (Kim et al., 2008; Lewontin and Birch, 1966; Rieseberg and Wendel, 1993). Hybridization is also thought to have a creative role in evolution by producing novel genetic entities, including new species (Rieseberg et al., 2003; Rieseberg and Willis, 2007). Furthermore, sophisticated tests of introgression using genetic data have recently revealed cryptic hybridization to be more widespread than was recognized from morphological analyses alone (Cronn and Wendel, 2004; Pennisi, 2016; Twyford and Ennos, 2012). Hybridization is so widespread in flowering plants that some have argued occasional hybridization is the rule not the

exception (Stebbins, 1959). However, despite its apparent frequency and importance the factors that contribute to the likelihood of hybridization have not been systematically tested.

The few existing systematic studies of hybridization frequency in plants have made important insights. Ellstrand et al. (1996) and more recently Whitney et al. (2010) both examined multiple floras to estimate hybridization frequency. They respectively estimate 0.109 and 0.093 hybrids per non-hybrid species. These two studies have three floras in common – The British Isles, American Intermountain, and the American Great Plains. For these floras, Whitney et al. (2010) report higher estimates; this may reflect more hybrids having been described in the intervening time or could also result from the more conservative approach of Ellstrand et al. (1996) in quantifying hybrids in taxonomically challenging groups. In both studies, the highest frequency of hybrids is in the British Isles at c. 0.25 hybrids per non-hybrid species. This reflects a long history of hybrid study in Britain (Matthews et al., 2015; Stace, 1975), and indicates that more hybrids are likely to be identified elsewhere.

Ecological disturbance has long been thought to be an important determinant of hybridization (Anderson, 1949). But previous studies have not explicitly tested for ecological factors affecting hybridization propensity. Ellstrand et al. (1996) note that most of the hybridizing species were perennial and outcrossing, and both studies find that the distribution of hybrids is non-random and concentrated in particular genera and families, but the significance of these observations was not tested. However, Whitney and colleagues (2010) do find phylogeny to have a significant effect. Insights from case studies also indicate that woody groups hybridize the most, but it is not clear to what extent these qualitative observations hold up in quantitative analyses.

Michigan is a unique test case for a detailed look at hybridization due to its natural history and extensive herbarium collections. Michigan is comprised of two large peninsulas which include many different local habitats such as islands, bogs, fens, marshes, lakes, dunes, coastlines, swamps, forests, railroads, roads, and ditches. Michigan was fully glaciated during the last glacial maximum 18-20 thousand years ago; re-colonization came from species inhabiting both glacial refugia and non-glaciated North America, and, due to the glaciation, Michigan has relatively few plant species. Today, both temperate and boreal ecosystems and

their transition zone occur in Michigan. Transitional ecosystems have the potential for hosting a high frequency of hybrids when closely related species with different adaptive norms establish secondary contact (Remington, 1968).

Here, we build upon previous work by quantifying the number of hybridizing species and the number of unique hybrids per non-hybrid species in the flora of Michigan. Then, we explicitly test factors that might be associated with hybridization, including physiognomy (life form and life history), habitat, mating system, and phylogeny. We conclude by discussing the frequency of hybridization, the factors associated with it, and interesting but understudied cases of hybridization in Michigan that are ready for a genetic approach; data which could clarify (or disprove) their history of hybridization and provide new insights about hybridization and introgression.

## 2.4. Materials and Methods

### 2.4.1. The data

The flora of Michigan includes 2,878 species of vascular plants (Reznicek and Voss, 2012; Voss, 1996). From these, 49 species are reported as extirpated and were excluded from all analyses. Our analysis includes ferns and other spore-bearing vascular plants which are not published in Reznicek and Voss (2012) but are included in the Michigan Flora Online website (Reznicek et al., 2011). Reznicek et al. (2011) also includes data on the number of herbarium collection per species. Online data was accessed on 29 November 2016.

In addition to information published in Reznicek and Voss (2012), the online flora (http://michiganflora.net) includes information on the number of collections for all species and incorporates data from the Michigan Floristic Quality Assessment (FQA) (Herman, 2001). FQAs measure an area's ecological integrity and are based on the plant communities' coefficients of conservatism (*C*). *C* values are between 0 and 10 and measure a species faithfulness to natural remnant communities. A $C < 3$ indicates that the species is found in disturbed habitats and a $C > 7$ indicates that it is found almost exclusively in high-quality natural plant communities. For example, the ragweed (*Ambrosia artemisiifolia*) is common on roadsides, ditches etc. and has a *C* of 0 whereas the fringed orchid (*Platanthera ciliaris*) is found only in sphagnum bogs and has a *C* of 10.

The FQA also assigns all species a wetness coefficient. This coefficient indicates the wetness of a species' preferred habitat and is divided into five groups: obligate wetland (OBL), facultative wetland (FACW), facultative (FAC), facultative upland (FACU), and obligate upland (UPL).

Finally, the FQA includes all species' physiognomy. We divide this into life history and life form. Life history has three categories: annual, biennial, or perennial. If a plant is capable of being perennial then it was assigned as perennial (http://michiganflora.net/home.aspx). The life form has nine groups that describe the form (and to some extent the ecological niche and phylogeny) of a species: fern, fern-ally, forb (non-woody), grass, sedge, shrub, tree, vine (non-woody), and woody-vine. If a plant *could* be woody, then it was assigned to a woody group even if it is not always woody.

While these data do not fully describe the life history or habitat of any species, they together provide quantitative information about a species' ecology; information that can be used to statistically test hypotheses about factors affecting hybridization likelihood.

### 2.4.2. Defining mating system

Plants have a number of mating systems from asexual to hermaphroditic self-compatible to fully dioecious. The idea behind testing mating system is that forced outcrossing (i.e. dioecy) potentially affects hybridization (from the perspective that hybridization is really a wide outcross). For all species in the flora, we either assigned them as cosexual or dioecious. If a species could be cosexual, then it was assigned as cosexual even if some individuals are unisexual. The information on mating system was compiled from Judd et al. (2016), The Flora of North America, and the keys of Reznicek and Voss (2012). Ferns and their allies have a sexual gametophyte and were excluded from the analyses testing the effect of mating system on hybridization.

### 2.4.3. Determining the extent of hybridization

All species descriptions were read to determine if the species was reported to hybridize with one or more other species. The totals of both the unique hybrid crosses and the parental species involved in hybridization were determined. Two species can only form one unique hybrid. The total number of non-hybrid species

excluded species of recent hybrid origin. While backcrossing is expected from an early-generation hybrid, hybrid species were only counted as hybridizing with their parents if the flora indicated it. For example, the large toothwort (*Cardamine maxima*) is thought to be of recent hybrid origin (Sweeney and Price, 2000), but is not treated as a hybridizing taxon in our study because no hybrids with *C. maxima* and its parental species are described. In contrast, the rattlesnake plantain (*Goodyera tesselata*) is the tetraploid hybrid of two diploid parental species. In this case, triploids have been collected, indicating backcrossing. We did not include as hybridizing those hybrid species that have escaped from cultivation unless they hybridized with other species in the wild (e.g. *Pyrus calleryana* × *P. communis*).

Some genera are characterized by agamospermy (e.g. *Rubus*) etc. and are extremely difficult for taxonomists. In a few of these cases, hybridization was reported to be widespread, but unique hybrids were not indicated. For these genera, the species which are said to hybridize are included as hybridizing but no unique hybrids with them as parents are included in the analysis.

In several genera, the taxonomy at the species level is somewhat muddled. As an example, some populations of perennial ryegrass (*Lolium perenne*) are occasionally recognized as a second species, *L. multiflorum.* However, despite clear morphological differences, the taxa hybridize to such an extent that there are too many intermediate individuals to recognize either individual species (Reznicek and Voss, 2012). We recognized only those species included in the flora and only quantified hybridization between species. Some results may therefore change depending on taxonomic revisions, but these cases are few in number.

### 2.4.4. Quantifying hybridization

Hybridization was quantified in three ways: hybridization frequency, hybrid ratio, and hybridization propensity. The hybridization frequency is the number of species that hybridize with one or more other species divided by the number of non-hybrid species. The hybrid ratio is the number of unique hybrids per non-hybrid species. Two species can only have one unique hybrid even if they hybridize multiple times. The hybridization propensity is taken from Whitney et al. (2010) and is the realized proportion of possible hybrids per genera. For example, in a genus of n species, there are (n(n-1)/2) possible unique hybrids and the hybridization propensity is then 100 × (number of unique hybrids in the genus /

possible unique hybrids). The idea behind the hybridization propensity is to normalize for the *opportunity* to hybridize, i.e. species in large genera have more potential interspecific hybridizing partners.

The hybridization frequency and hybrid ratio were calculated for all species, per genus, and per family. But the per family and per order hybridization propensities are the weighted means of their genera's propensities (Whitney et al., 2010). Genera with only one species were not included in the analysis of hybridization propensity because their propensity is undefined. Unique intergeneric hybrids were assigned 50-50 to the parental genera.

### 2.4.5. Statistical tests of significance

One-sample Chi-Square tests for goodness of fit were done in R to test if the hybridization was non-random with respect to the coefficient of wetness, coefficient of conservatism, life history, life form, and sexual system. Linear regressions were calculated in R using the lm function.

A phylogeny for the 41 orders that have one or more genera with one or more species was built using Angiosperm Phylogeny Group IV (APG) relationships (Byng et al., 2016). The per order hybridization frequencies, hybrid ratios, and hybridization propensities were then mapped onto the phylogeny and ancestral nodes were inferred using phytools v0.6-00 (Revell, 2012). Trees were then tested for a phylogenetic signal (λ) with regards to all hybridization measures using the ML model of BayesTraits v3 (Meade and Pagel, 2016).

### 2.4.6. Data availability

The data used in this study is available in the online version of this article.

## 2.5. Results

In total, 373 species were identified as hybridizing with one or more species in the wild. There was a large discrepancy between the number of hybrids in native and adventive species (Table 1). This is probably due to collection bias and/or rarity in adventive taxa, and does not reflect the natural hybridization frequency. The mean number of herbarium collections for native species (103) was significantly more than for adventive species (31) (Wilcox test $P < 2.2e\text{-}16$). We therefore exclude adventive species and their unique hybrids from the analyses and results. The remaining 1,759 native species belong to 572 genera, 151 families, and 51 orders.

### 2.5.1. Frequency of interspecific hybridization

A total of 306 native species were identified as hybridizing. Twelve species were identified as being of hybrid origin, but only six of these were hybridizing with other species. The hybridization frequency is therefore 17.5% (hybridizing species=306 / total non-hybrid species=1747).

There are 221 unique hybrids belonging to 81 genera, including 78 named hybrids that are not included in the flora as taxa (Supplemental Table 1). The overall hybrid ratio was 0.127 (221 unique hybrids / 1747 non-hybrid species). Out of 572 total genera, 282 are monotypic in the flora, and if these genera are excluded, then the hybridization frequency and hybrid ratio increase to 20.8% and 0.151, respectively. All hybridizing genera, species, and unique hybrids are described further in Supplemental Table 2. The final data used for analyses is available in Supplemental Table 3 online.

### 2.5.2. Comparison of hybridization metrics

All three metrics for quantifying the extent of hybridization are significantly correlated when compared by family and genus (Supplemental Figure 1). The hybridization frequency is higher than the hybrid ratio except for a few cases. This is because most hybridizing species have only a single interspecific partner (Supplemental Figure 2), and the median number of species per genus is two; this results in genera with a hybridization frequency of 1 but hybrid ratio of 0.5. Hybridization propensity is always lower than hybridization frequency, and, with

few exceptions, lower than the hybrid ratio (Supplemental Figure 1). The hybridization propensity relies on the observation that nearly all hybridization is intrageneric, but ignores phylogenetic structure below the genus level that can affect the opportunity to hybridize. Furthermore, because the number of possible unique interspecific hybrids increases exponentially with the number of species, hybridization propensity is high in small genera and remains low in large genera (Figure 1).

For hybridizing species, the mean number of interspecific partners was 1.47 ± 0.82 (mean of 0.25 partners per species for all species). A total of 32 species (1.8% of all species) had more than two interspecific partners and only 11 species in four genera had four or more interspecific partners (Supplemental Figure 2). Thus, most hybridization involves only two species, with few cases of free hybridization within a genus.

### 2.5.3. Hybrid species

There are 12 hybrid species in the flora. A few more instances were identified as potential examples of hybrid speciation due to partial fertility in the hybrids, allopatric hybrid populations, and/or potential postzygotic barriers to backcrossing between the hybrid and both its parents (i.e. an allotetraploid with two diploid parents). These 12 hybrid species and the evidence for hybrid speciation are given in Table 2. The other potential instances can be found in Supplementary Table 2.

### 2.5.4. Phylogenetic distribution of hybridizing species

For both families and genera, the number of recorded hybrids increases significantly with the number of taxa (family = Adjusted R−Sqr 0.78, P=6.32e−51; genus = Adjusted R−Sqr 0.47, P=3.69e−80) (Supplemental Figure 3). Hybrids were non-randomly distributed across genera and families (Figure 2). Overall 29.8% of families had one or more hybridizing taxa, but more than 50% of hybridizing species belonged to just six families (3.4% of the total # of families). For genera, 11.1% had one or more hybridizing taxa, but 20 genera (2.2% of the total # of genera) included more than half of hybridizing species. The hybridization frequency, hybrid ratio, and hybrid propensity for the 20 families with the most hybridizing taxa are given in Table 3. The equivalent data for genera is in Table 4.

Neither hybridizing frequency, hybrid ratio, nor the hybridization propensity had a significant phylogenetic signal using BayesTraits ($\lambda < 0.001$) (Supplemental Figure 4).

### 2.5.5. Factors associated with hybridization frequency

All of the tested habitat and physiognomy variables were significantly associated with hybridization frequency with the exception of mating system for which we tested only the effect of being cosexual vs. dioecious. Ninety seven percent of hybridizing species were perennial, and this is significantly more than other life histories (Chi-sqr P = 1.68e-08) (Table 5). To test if this result was confounded by life form (because fern allies, trees, and shrubs are always perennial and all have high hybridization frequencies), we tested the significance of life history on hybridization exclusively in forbs (N=959) which include all life histories. While 79.7% of the total forbs are perennial, 94.9% of hybridizing forbs were perennial. Our data therefore indicates a significant interaction between life history and hybridization frequency (Chi-sqr P = 3.90e-05).

The frequency of hybridization was significantly different between wetness coefficients (Chi-sqr P = 0.010). Obligate and facultative wet species had the highest hybridization percentages at 21.9% and 18.5% respectively. In contrast, only 11.8% of obligate upland species hybridized (Supplemental Figure 5).

Hybridization frequency was significantly different between life forms (Chi-sqr P=4.46e-7) (Figure 3). The fern allies had the highest frequency of hybridizing species at 37.5% (N=32). Fern allies in Michigan include one order each of clubmoss (Lycopodiales), quillwort (Isoetales), spikemoss (Selaginellales), and horsetail (Equisetales). Note that horsetails are assigned as fern allies by Reznicek et al. (2011), but included within the monilophytes by Byng et al. (2016). The next highest hybridizing percentages were found in shrubs (29.5%) and trees (27.4%). Interestingly, with the exception of non-woody vines (N=30) which had no hybridizing taxa, ferns (N=72) had the lowest hybridization frequency at 10%.

The coefficient of conservatism was significantly correlated with the number of hybridizing species if species with *C*=0 were removed from the analysis. When *C*=0 species were removed, there was a strong negative correlation between *C* and hybridization frequency: species restricted to more undisturbed plant communities hybridize less and species in disturbed habitats hybridize more

(Adjusted R-sqr = 0.81, P < 0.001) (Figure 4a). The *C*=0 species include significantly more annual species which have low rates of hybridization (Figure 4b). There also appears to be a strong relationship between *C* and the number of herbarium collections per individual: there are fewer collections for species with higher *C* values except the *C*=0 species, which have relatively fewer collections (Figure 4c).

In fact, the number of herbarium collections is correlated with hybridization frequency. Hybridizing species have significantly more collections (144 ± 95) than non-hybridizing species (98 ± 86, Wilcox test P=3.09e-17) (Figure 5). The hybridization frequency also increases significantly with the mean number of collections per genera (P<0.001), but does not explain much of the data (Adjusted R-sqr = 0.034); hybridization frequency and mean number of collection per species are not significantly correlated at the family level.

Furthermore, shrubs have significantly more (Wilcoxon text P<0.01) herbarium collections than all other life forms except woody vines; besides shrubs there are no further significant differences between life forms (Supplemental Figure 6a). Trees, forbs, and sedges have significantly more (Wilcoxon text P<0.01) collections for hybridizing species than non-hybridizing species; but there is no significant difference between the number of collections for hybridizing and non-hybridizing species in the other life forms, including shrubs (Supplemental Figure 6b).

## 2.6. Discussion

This study quantifies the frequency of hybridization in Michigan's vascular plants and finds that nearly one in five species hybridize in the wild. We show that perennials hybridize at a significantly higher rate than species with other life histories and that different life forms hybridize at different rates, with fern-allies, shrubs, and trees having the highest proportion of hybridizing taxa in Michigan. We further show that species inhabiting locally disturbed habitats hybridize at a higher rate than species restricted to natural remnant plant communities. We find no effect of mating system, but do find that hybridizing species are non-randomly distributed in certain taxonomic groups; however, phylogenetic relationships do not appear to be a significantly associated with hybridization rate.

One interesting outcome of this study is that there are significantly fewer herbarium specimens for non-hybridizing species and fewer collections for species restricted to natural remnant communities. This needs to be interpreted cautiously. First, species with a higher *C* value are – by definition – restricted in habitat and rarer than other species. They will therefore be collected less. Moreover, "some botanists have a tendency to collect along roadsides" (Coyne & Orr 2004 p. 45). Second, hybridizing species are often difficult and peculiar and therefore collected more frequently (Wiegand, 1935). However, as a counter example, there are over 500 collections of *Trillium grandiflorum*, and additional collections will most certainly not produce evidence of hybridization. Thus, it is premature to argue that hybridization is biased by collection number, but the observation is an intriguing one.

### 2.6.1. Metrics of hybridization and recommendations for future studies

There are three systematic studies of hybridization in plants (excluding Mayr (1992) which has been strongly criticized by Whittemore (1993) and others). Ellstrand et al. (1996) and Whitney et al. (2010) report 642 and 770 hybrids in the British Isles, respectively. However, both quantified hybrids from the New Flora of the British Isles, 2nd ed. The discrepancy could be due to differences in quantifying hybrids in 'nightmare' taxonomic groups. Whitney et al. (2010) took half of possible unique hybrids. For example, in *Rubus* (N=12 species in Michigan) there are $\frac{12\times11}{2} = 132$ possible unique hybrids; however, we report only 9 hybridizing

taxa and only 5 unique hybrids. Thus, this approach may be disproportionately influenced by taxonomic nightmares and overestimate hybridization.

Both Ellstrand et al. (1996) and Whitney et al. (2010) quantified the number of unique hybrids (equivalent to our hybrid ratio) but not the hybridization frequency. In contrast, Mallet (2005) quantified the equivalent of our hybridization frequency. Both of these metrics are useful and provide different insights. The final metric, hybridization propensity from Whitney et al. (2010), is more problematic. It appears biased against large genera and ignores phylogenetic structure below the genus level. As a recommendation, future systematic studies should quantify both hybridizing taxa, their mating partners, and all unique hybrids, but avoid inferring hybrids not directly reported in a flora and be cautious of using hybridization propensity.

### 2.6.2. Estimating the true hybridization frequency

This and other studies probably underestimate the true number of hybridizing species, despite arguments to the contrary (Coyne & Orr p. 41). First, cryptic hybrids may not be readily identifiable based upon morphology, including cryptic polyploid species which are widespread in plants (Soltis et al., 2007). Neither cytologic nor molecular data exist for most species, but effort to collect this type of data is increasing, and could reveal novel cryptic species and hybrids.

The second argument for underestimating the true number of hybridizers is that hybridization is restricted by geography. Indeed, many of Michigan's species are known to hybridize elsewhere but not within Michigan. Consider the gymnosperms (N=13) which have no hybrids in Michigan but are known to hybridize elsewhere. For example, two species of spruce, *Picea* glauca and *P. mariana*, are found in Michigan and known to hybridize in Minnesota (Little and Pauley, 1958); jack pine (*Pinus banksiana*) forms a hybrid zone with lodgepole pine (*P. cordata*) in western North America (Wagner et al., 1987); lodgepole pine is not found in Michigan. *Juniperus virginiana* has a hybrid zone with *J. horizontalis* in Wisconsin (Palmaotal et al., 1983). Both of these species are in Michigan. *Abies balsamea* is the only fir species in Michigan, but hybridizes with both *A. fraseri* in Eastern North America and has a hybrid zone with *A. lasiocarpa* in Western North America (Cinget et al., 2015; Clark et al., 2000). No evidence for hybridization in the other seven taxa was found, but these examples would

increase the hybridization frequency of gymnosperms found in Michigan to nearly 50%. Species not said to be hybridizing in Michigan that are known to hybridize elsewhere are not restricted to gymnosperms. We identified at least 19 additional examples in other groups (Supplementary Table 4).

The conclusion from this exercise is *not* that more than 17.5% of species hybridize in Michigan. Nor is the conclusion that the true hybridization frequency is greater than 17.5%. First, many of the potential interspecific mates are not found in Michigan (i.e. hybridization could not happen in Michigan because the populations are allopatric). Second, even if both species are found in Michigan, they may not hybridize there; the right habitat (discussed below) is critical to hybrid formation, and may not exist in Michigan. However, the gymnosperm example demonstrates that more than 17.5% of the species found in Michigan do hybridize in the wild elsewhere.

Previous systematic studies that quantified hybridization in multiple floras did not quantify the overlap in species or hybrids. As more floras are analyzed, the total number of hybridizing species could *only* increase. However, the hybridization frequency and hybrid ratio could either increase or decrease depending on if there is an increasing or diminishing return for the number of hybrids per new species. For example, if Michigan and Wisconsin share 90% of unique hybrids but 50% of taxa, then the hybridization frequency and hybrid ratio will both decrease. As more floras are analyzed in conjunction, we will arrive at a more accurate estimate of natural hybridization frequency.

### 2.6.3. Ecological correlates of hybridization: habitat disturbance as a critical factor

The role of ecology in hybridization was already recognized by (Focke, 1881). One of the best examples of the ecology of hybridization remains that of *Iris fulva* and *I. hexagona* whose hybrids occur only in disturbed habitats (Anderson, 1949). This is the first study to show a systematic relationship between a species' preference for local habitat disturbance and hybridization.

### 2.6.4. Phylogenetic distribution of hybridization

As in previous studies, we find that hybridization is concentrated in certain genera and families (Ellstrand et al., 1996; Mallet, 2005; Whitney et al., 2010).

However, these studies also note that different groups have been shown to have different hybridization rates depending on the flora examined. We have demonstrated that this is the case for the gymnosperms of Michigan, and we also find that ferns, which have high rates of hybridization elsewhere, have an unexpectedly low hybridization frequency in Michigan. One reason may be because, following Reznicek et al. (2011), we include the Equisetales (hybridization frequency = 0.33, N= 9) as fern allies not as ferns.

In contrast to Whitney et al. (2010), this study does not find a significant effect of phylogeny on hybridization propensity (or other measures of hybridization). However, Whitney et al. (2010) examined more taxa and therefore may have had more power to detect an effect. That said, the extremely high rates of hybridization in the lycophytes and monilophytes (which are lumped together while the rest of their phylogeny has orders as tips) might bias their result. On the other hand, they could also be detecting the correlation of life history and taxonomy. In our opinion, physiological properties are more proximal to hybridization frequency than phylogeny.

### 2.6.5. Life history and hybridization propensity

We found that 97% of hybrids are perennial and that they hybridize at a significantly higher rate than species with other life histories. Stace (1975) observed that more hybrids were perennial than annuals and attributed this to autogamy being more frequent in annuals. Another explanation is vegetative reproduction which is widespread in perennial plants (Grant, 1981). For example, most species of oaks have rhizomes which produce suckers capable of developing into new plants, and *Populus* species form extremely long-lived colonies via vegetative spread. Both of these genera have high rates of hybridization. Paun et al. (2009) make the distinction between hybrid formation and establishment, and it may be that hybrid establishment is easier in long-lived species. This trait could be important because many hybrids show reductions in fertility despite being otherwise vigorous, and vegetative reproduction may allow them to hold on and produce some viable offspring. Stebbins (1959, p. 237) computes the reproductive output of a long-lived, but nearly sterile *Elymus* hybrid with 0.001% fertility, concluding that this individual could produce 200 viable seeds in its lifetime.

### 2.6.6. High rates of hybridization in woody plants

Woody groups were already recognized by Stebbins (1959) as frequently hybridizing, and some of the most prominent case studies of hybridization come from trees (Lagache et al., 2014; Lindtke et al., 2014). However, Petit and Hampe (2006) note that “few tree-rich floras have been examined for the frequency of hybrids.” This study includes 105 tree and 146 shrub species, and we find a significant enrichment of hybridization in these two life forms. Thus, the observation of (Stebbins, 1959) and others does not appear to be disproportionately influenced by the inclusion of some ‘nightmare’ tree taxa like the oak syngameon. The respective hybridization frequencies of 0% and 19% for non-woody and woody vines also substantiates this observation.

One of the features of trees (and shrubs and woody vines) that may contribute to their high rates of hybridization is that trees are predominately outcrossing and no tree species are exclusively selfing (Petit and Hampe, 2006). Many mechanisms ensure outcrossing including dioecy, and dioecy is more frequent in woody plants than forbs (Vamosi and Vamosi, 2004); it occurs in 8.5% of Michigan trees and 10.3% of shrubs. However, we did not find any significant difference in hybridization frequency between cosexual and dioecious species. One explanation is that dioecy is only one of many mechanisms to ensure outcrossing (others include dichogamy, genetic self-incompatibility). Thus, outcrossing rate could still be an important factor even if being dioecious is not.

Mallet (2005) concludes that “the loss of a tendency to hybridize is reasonably, but relatively coarsely clock-like” (with variable rates). A second explanation for high hybridization in woody groups is that they have a slower molecular clock. Evolutionary rates are on average about 2-2.5 times faster in herbs than shrubs/trees (Kay et al., 2006; Smith and Donoghue, 2008), but this result may reflect generation time not ‘woodiness’ per se because perennials also have slower rates of molecular evolution (Andreasen and Baldwin, 2001) and high rates of hybridization (this study). Trees typically have both long generation times and large population sizes (Petit and Hampe, 2006). Both factors lead to a slower rate of molecular evolution.

It is interesting to consider four species that are the only representatives of their genera in Michigan: *Liriodendron tulipifera* (Tulip-Tree), *Platanus occidentalis*

(Sycamore), *Sassafras* albidum, and *Morus rubra* (Mulberry). All of these species are long-lived trees with no opportunity for natural hybridization in Michigan, and they all also have closely related species in Eastern Asia or Europe. All produce to some degree fertile hybrids in test crosses with their closely related cousins (Judd et al., 2016; Nie et al., 2007; Parks and Wendel, 1990). Both their cross compatibility and their morphological similarity indicate low rates of genomic divergence between these species. Thus, without natural selection for breeding barriers, neutral incompatibility due to drift takes a long time to arise between large populations of long-lived species.

The ability to form fertile hybrids with relatives from different continents is not exclusive to woody groups, and it would be very interesting to compare population sizes and generation times of hybridizing vs. non-hybridizing species in perennial forbs. One other possibility is the reproductive strategy of long-lived perennials relative to annuals. Besides annuals being more autogamous (Barrett and Eckert, 1990), long-lived perennials produce high numbers of offspring and their fecundity greatly exceeds carrying capacity. Thus, even if there are incompatibilities between species, occasional crosses might overcome these incompatibilities and the resulting offspring could be vigorous, grow to maturity, and become well-established populations.

### 2.6.7. Interesting examples of hybridization in Michigan

There are many examples of hybridization and possible hybrid speciation in Michigan, but few are supported by population-level molecular data. Two interesting cases that would benefit from molecular data are described in some detail here: *Fraxinus* and *Viola*.

Two polyploid species have recently been described from *F. americana* L.: hexaploid *F. biltmoreana* Beadle and tetraploid *F. smallii* Britton (Nesom, 2010). Currently, only *F. americana* is recognized by Reznicek and Voss (2012). However, tetraploid *F. americana* (= *F. smallii*) is known from Michigan and there could be hexaploid individuals as well (Nesom, 2010). Wallander (2008), using nuclear and cytoplasmic markers, found that one *F. biltmoreana* was distinct from two other *F. americana* individuals; no individuals of *F. smallii* were included. Because triploid and pentaploid individuals are rare (and would presumably be sterile), *F. smallii* is thought to be reproductively isolated from both *F. americana*

and *F. biltmoreana*; however, tetraploids from a diploid by hexaploid cross would be fertile (i.e. *F. americana* × *F. biltmoreana* would result in fertile tetraploids with the same cytotype as *F. smallii*). Nesom (2010) reasons that because *F. smallii* is found further west than *F. biltmoreana*, these tetraploids cannot be hybrids of *F. americana* and *F. biltmoreana*. However, there is no genetic data for any tetraploid (= *F. smallii*) populations, and their origin is therefore not known. Nor is the origin of *F. biltmoreana* clear. Were these species formed once or do the tetraploid and hexaploid populations have multiple origins? Are the parental species *F. americana* and the closely related *F. pennsylvanica* Marsh. or were they formed from crosses between distinct *F. americana* populations? This interesting case exemplifies the difficulties in taxonomy and hybridization, and the opportunity for the application of molecular data.

The violets (*Viola*) of Michigan were described by (Ballard, 1994) as a "pack of incorrigible wolves." One reason is that parental crosses in the lab result in hybrids that immediately produce well-developed cleistogamous seeds, and the hybrids fill morphological and ecological gaps between parental species (Ballard, 1994). There are 22 native *Viola* taxa; however, Reznicek and Voss (2012) try not to over differentiate the genus (e.g. *V. pubescens* is considered one species). Nevertheless, depending on one's taxonomic disposition and molecular data, the number of species and hybrids could easily change.

Three potential hybrid species in *Viola* are of particular interest: *V. primulifolia*, *V.* ×*subsinuata*, and *V.* ×*palmata*. *Viola primulifolia*, a fertile species known from the Atlantic coastal plain, occurs in small colonies with many ramets interconnected by stolons (2n=24). The Michigan hybrid *V. macloskeyi* × *V. lanceolata* is morphologically similar to *V. primulifolia*. Most of these populations are sterile, but there are fertile populations known from Michigan and northern Illinois. Later generation hybrids are said to breed true and more closely resemble one of the parental taxa which may indicate some backcrossing. The open question is if the Michigan fertile populations are equivalent to more Eastern ones or were independently formed. A larger question might be whether the leaf morphology differences used to differentiate *V. primulifolia* and *V. lanceolata* is of evolutionary significance and if these are even 'good' species.

The other two potential hybrid species, *V.* ×*subsinuata*, and *V.* ×*palmata* are understudied. They are found throughout the Eastern United States, have

clesitogamous seeds, and apparently are from populations recently (and currently being) formed via hybridization. The last treatment of Michigan's violets was by Ballard in 1994 and he relied exclusively on morphology. Because there are more idiosyncrasies than rules for plant hybrids, investigations into both *Fraxinus* and *Viola* could result in both novel insights and taxonomic clarity.

## 2.7. Conclusions

In conclusion, we demonstrate that nearly one in five of Michigan's species hybridize and that there are 0.127 unique hybrids per non-hybrid species. We show that these are underestimates of the number of hybridizing species, but that the true frequency of hybridizing taxa could be under or overestimated. Finally, we demonstrate that both life form and habitat disturbance are significant factors associated with hybridization propensity, but that these factors may be confounded/biased by rare species that are infrequently collected. Perennials, and especially fern allies, shrubs and trees have the highest rates of hybridization. Finally, this study strongly advocates for the continued funding of herbaria which are of fundamental importance to future studies in plant evolution (Deng, 2015).

## 2.8. Acknowledgements

I.B. thanks Grant Godden for introduction to the Field Manual of Michigan Flora. Thanks to E. Voss and A. Reznicek for their many years of hard work studying the plant species of Michigan and for making the data available online. I.B. was supported by a fellowship from the International Graduate School for Plant Science program (DFG GRK1525).

## 2.9. References

Anderson, E., 1949. Introgressive hybridization. John Wiley and Sons, NY.
Andreasen, K., Baldwin, B.G., 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.
Ballard, H., 1994. Violets of Michigan. Mich. Bot 33, 131-199.
Barnes, B.V., Dancik, B.P., 1985. Characteristics and origin of a new birch species, Betula murrayana, from southeastern Michigan. Canadian Journal of Botany 63, 223-226.
Barrett, S., Eckert, 1990. Variation and evolution of mating systems in seed plants. Biological Approaches and Evolutionary Trends in Plants, 229-254.
Bevan, M.W., Uauy, C., et al., 2017. Genomic innovation for crop improvement. Nature 543, 346-354.
Byng, J.W., Chase, M.W., et al., 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181, 1-20.
Carlsen, T., Bleeker, W., et al., 2009. Biogeography and phylogeny of cardamine (Brassicaceae). Annals of the Missouri Botanical Garden 96, 215-236.
Cinget, B., de Lafontaine, G., et al., 2015. Integrating phylogeography and paleoecology to investigate the origin and dynamics of hybrid zones: insights from two widespread North American firs. Mol. Ecol. 24, 2856-2870.
Clark, C.M., Wentworth, T.R., et al., 2000. Genetic discontinuity revealed by chloroplast microsatellites in eastern North American Abies (Pinaceae). Am. J. Bot. 87, 774-782.
Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland, MA.
Cronn, R., Wendel, J.F., 2004. Cryptic trysts, genomic mergers, and plant speciation. New Phytol. 161, 133-142.
Crowder, A.A., Pearson, M.C., et al., 1990. Drosera L. J. Ecol. 78, 233-267.
Cui, L.Y., Wall, P.K., et al., 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16, 738-749.
Deng, B., 2015. Plant collections left in the cold by cuts. Nature 523, 16-16.
Di Gaspero, G., Copetti, D., et al., 2012. Selective sweep at the Rpv3 locus during grapevine breeding for downy mildew resistance. Theor. Appl. Genet. 124, 277-286.
Dobes, C.H., Mitchell-Olds, T., et al., 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American Arabis drummondii, A. x divaricarpa, and A. holboellii (Brassicaceae). Mol. Ecol. 13, 349-370.
Ellstrand, N.C., Whitkus, R., et al., 1996. Distribution of spontaneous plant hybrids. Proc. Natl. Acad. Sci. USA 93, 5090-5093.
Focke, W.O., 1881. Die Pflanzen-Mischlinge: ein Beitrag zur Biologie der Gewächse. Gebr. Borntraeger.
Grant, V., 1981. Plant speciation. New York: Columbia University Press.
Herman, K.D., 2001. Floristic quality assessment with wetland categories and examples of computer applications for the state of Michigan. Michigan Dept. of Natural Resources, Wildlife Division, Natural Heritage Program.
Judd, W.S., Campbell, C.S., et al., 2016. Plant Systematics: A Phylogenetic Approach, Fourth ed. Sinauer Associates.
Juslen, A., Vare, H., et al., 2011. Relationships and evolutionary origins of polyploid Dryopteris (Dryopteridaceae) from Europe inferred using nuclear pgiC and plastid trnL-F sequence data. Taxon 60, 1284-1294.
Kallunki, J.A., 1976. Population studies in Goodyera (Orchidaceae) with emphasis on the hybrid origin of G. tesselata. Brittonia 28, 53-75.
Kay, K.M., Whittall, J.B., et al., 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.
Kim, M., Cui, M.-L., et al., 2008. Regulatory genes control a key morphological and ecological trait transferred between species. Science 322, 1116-1119.

Koch, M.A., Dobes, C., et al., 2003. Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in north American Arabis divaricarpa (Brassicaceae). Mol. Biol. Evol. 20, 338-350.
Kowal, R.R., Judziewicz, E.J., et al., 2011. Packera insulae-regalis (Asteraceae, Senecioneae), a new species endemic to Isle Royale, Michigan, USA. Brittonia 63, 343-354.
Lagache, L., Klein, E.K., et al., 2014. Distinct male reproductive strategies in two closely related oak species. Mol. Ecol. 23, 4331-4343.
Laureto, P.J., Barkman, T.J., 2011. Nuclear and Chloroplast DNA Suggest a Complex Single Origin for the Threatened Allopolyploid Solidago houghtonii (Asteraceae) Involving Reticulate Evolution and Introgression. Syst. Bot. 36, 209-226.
Laureto, P.J., Pringle, J.S., 2010. Solidago vossii (Asteraceae), A new species of Goldenrod from Northern Michigan. Mich. Bot, 105-117.
Lewontin, R.C., Birch, L.C., 1966. Hybridization as a source of variation for adaptation to new environments. Evolution 20, 315-336.
Lin, T., Zhu, G., et al., 2014. Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. 46, 1220-1226.
Lindtke, D., Gompert, Z., et al., 2014. Unexpected ancestry of Populus seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. Mol. Ecol. 23, 4316-4330.
Little, E.L., Pauley, S.S., 1958. A natural hybrid between black and white spruce in Minnesota. The American Midland Naturalist 60, 202-211.
Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229-237.
Marcks, B.G., 1974. Preliminary reports on the flora of Wisconsin: no. 66. Cyperaceae: 2. - Sedge family: 2. The genus Cyperus - the umbrella sedges. Trans. Wisconsin Acad. Sci. 62, 261-284.
Matthews, A., Emelianova, K., et al., 2015. 250 years of hybridization between two biennial herb species without speciation. Aob Plants 7.
Maurin, O., Davis, A.P., et al., 2007. Towards a phylogeny for Coffea (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. Ann. Bot. 100, 1565-1583.
Mayr, E., 1992. A local flora and the biological species concept Am. J. Bot. 79, 222-238.
Meade, A., Pagel, M., 2016. BayesTraits V3 <http://www.evolution.rdg.ac.uk/>.
Nesom, G., 2010. Fraxinus biltmoreana and Fraxinus smallii (Oleaceae), forest trees of the eastern United States. Phytoneuron 51, 1-30.
Nie, Z.L., Wen, J., et al., 2007. Phylogeny and biogeography of Sassafras (Lauraceae) disjunct between eastern Asia and eastern North America. Plant Syst. Evol. 267, 191-203.
Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401-437.
Palmaotal, M., Moore, W.S., et al., 1983. Morphological, chemical, and biogeographical analyses of a hybrid zone involving Juniperus virginiana and J. horizontalis in Wisconsin. Canadian Journal of Botany 61, 2733-2746.
Parks, C.R., Wendel, J.F., 1990. Molecular divergence between asian and north-american species of Liriodendron (Magnoliaceae) with implications for interpretation of fossil floras. Am. J. Bot. 77, 1243-1256.
Paun, O., Forest, F., et al., 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytol. 182, 507-518.
Pennisi, E., 2016. Shaking up the Tree of Life. American Association for the Advancement of Science.
Petit, R.J., Hampe, A., 2006. Some evolutionary consequences of being a tree. Annual Review of Ecology Evolution and Systematics 37, 187-214.
Pryer, K.M., Haufler, C.H., 1993. Isozymic and chromosomal evidence for the allotetraploid origin of Gymnocarpium dryopteris (Dryopteridaceae). Syst. Bot. 18, 150-172.
Remington, C.L., 1968. Suture-zones of hybrid interaction between recently joined biotas, Evolutionary biology. Springer, pp. 321-428.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217-223.
Reznicek, A., Voss, E.G., et al., 2011. Michigan Flora Online <http://michiganflora.net/home.aspx>. University Of Michigan.
Reznicek, A.A., Voss, E.G., 2012. Field manual of Michigan flora. University of Michigan Press.
Rieseberg, L.H., Raymond, O., et al., 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301, 1211-1216.
Rieseberg, L.H., Wendel, J.F., 1993. Introgression and its consequences in plants. Hybrid zones and the evolutionary process, 70-109.
Rieseberg, L.H., Willis, J.H., 2007. Plant speciation. Science 317, 910-914.
Runk, K., Zobel, M., et al., 2012. Biological flora of the British Isles: Dryopteris carthusiana, D. dilatata and D. expansa. J. Ecol. 100, 1039-1063.
Shulaev, V., Sargent, D.J., et al., 2011. The genome of woodland strawberry (Fragaria vesca). Nat. Genet. 43, 109-116.
Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322, 86-89.
Soltis, D.E., Soltis, P.S., et al., 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56, 13-30.
Stace, C.A., 1975. Hybridization and the flora of the British Isles. Academic Press, London.
Stebbins, G.L., 1959. The role of hybridization in evolution. Proceedings of the American Philosophical Society 103, 231-251.
Sweeney, P.W., Price, R.A., 2000. Polyphyly of the genus Dentaria (Brassicaceae): Evidence from trnL intron and ndhF sequence data. Syst. Bot. 25, 468-478.
Twyford, A.D., Ennos, R.A., 2012. Next-generation hybridization and introgression. Heredity 108, 179-189.
Vamosi, J.C., Vamosi, S.M., 2004. The role of diversification in causing the correlates of dioecy. Evolution 58, 723-731.
Voss, E., 1996. Michigan flora, parts I–III. Cranbrook Institute of Science. Ann Arbor, Michigan, USA.
Wagner, D.B., Furnier, G.R., et al., 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. Sci. USA 84, 2097-2100.
Wallace, L.E., 2003. Molecular evidence for allopolyploid speciation and recurrent origins in Platanthera huronensis (Orchidaceae). Int. J. Plant Sci. 164, 907-916.
Wallander, E., 2008. Systematics of Fraxinus (Oleaceae) and evolution of dioecy. Plant Syst. Evol. 273, 25-49.
Whitney, K.D., Ahern, J.R., et al., 2010. Patterns of hybridization in plants. Perspectives in Plant Ecology Evolution and Systematics 12, 175-182.
Whittemore, A.T., 1993. Species concepts: a reply to Ernst Mayr Taxon 42, 573-583.
Wiegand, K.M., 1935. A taxonomist's experience with hybrids in the wild. Science 81, 161-166.
Wood, C.E., 1955. Evidence for the hybrid origin of Drosera anglica. Rhodora 57, 105-130.
Wood, T.E., Takebayashi, N., et al., 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA 106, 13875-13879.
Wu, G.A., Prochnik, S., et al., 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. 32, 656-662.

## 2.10. Figures

**Figure 1** Different measures of hybridization. Each bar represents one of the 92 genera that have one or more hybridizing species. (**a**) Number of species in each genus. (**b**) Total number of species that are hybridizing in each genus. (**c**) Hybrid ratio, (**d**) Hybridization frequency, and (**e**) Hybridization propensity in each genus. The hybridization propensity appears biased against large genera relative to hybrid ratio and hybridization frequency.

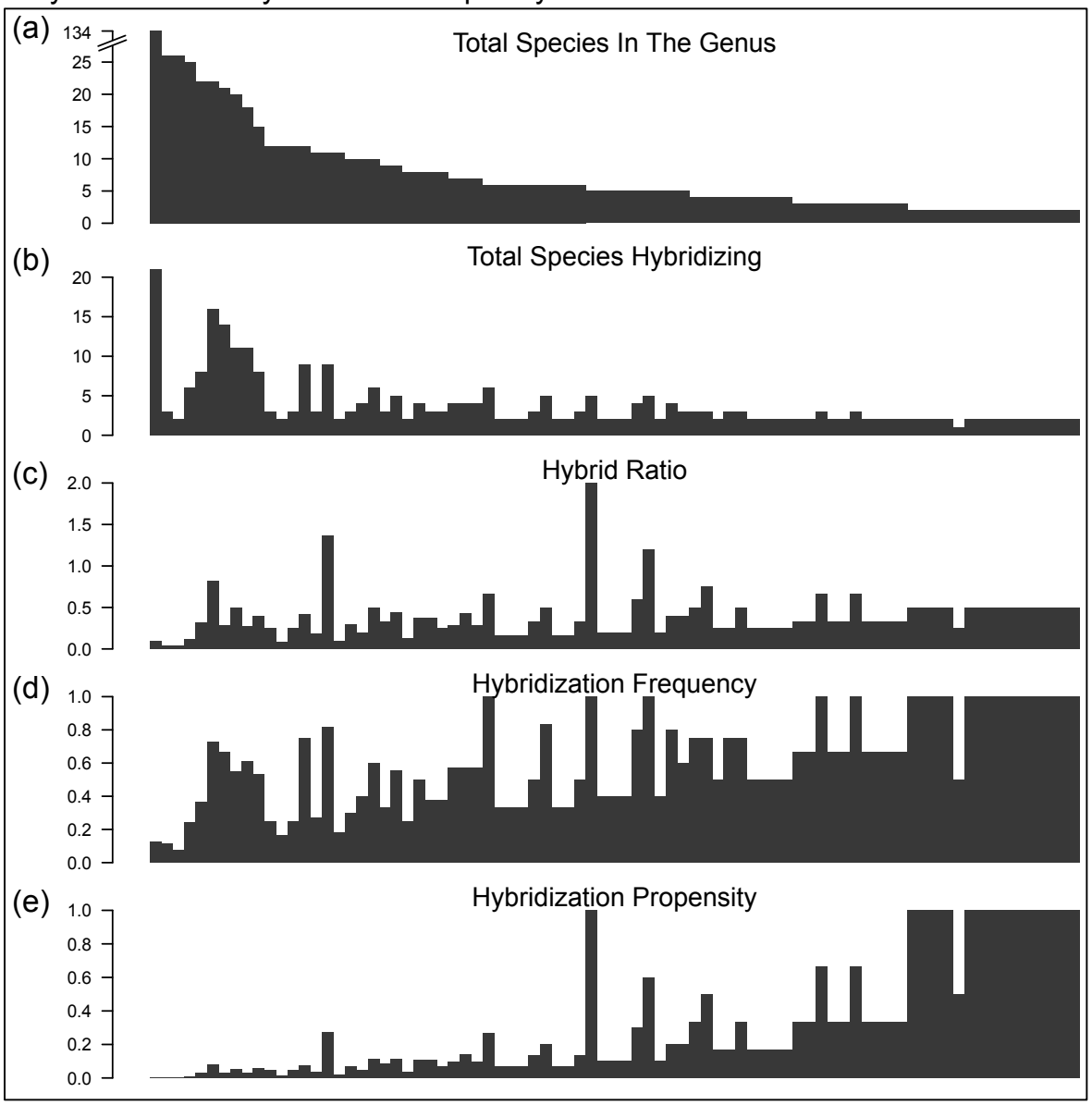


**Figure 2** Cumulative frequency distribution of species and hybrids by taxonomy. The cumulative frequency of all species by family (**a**) and genus (**b**). Cumulative frequency of all hybrids by family (**c**) and genus (**d**). The total number of hybridizing species are concentrated in relatively few families and genera.

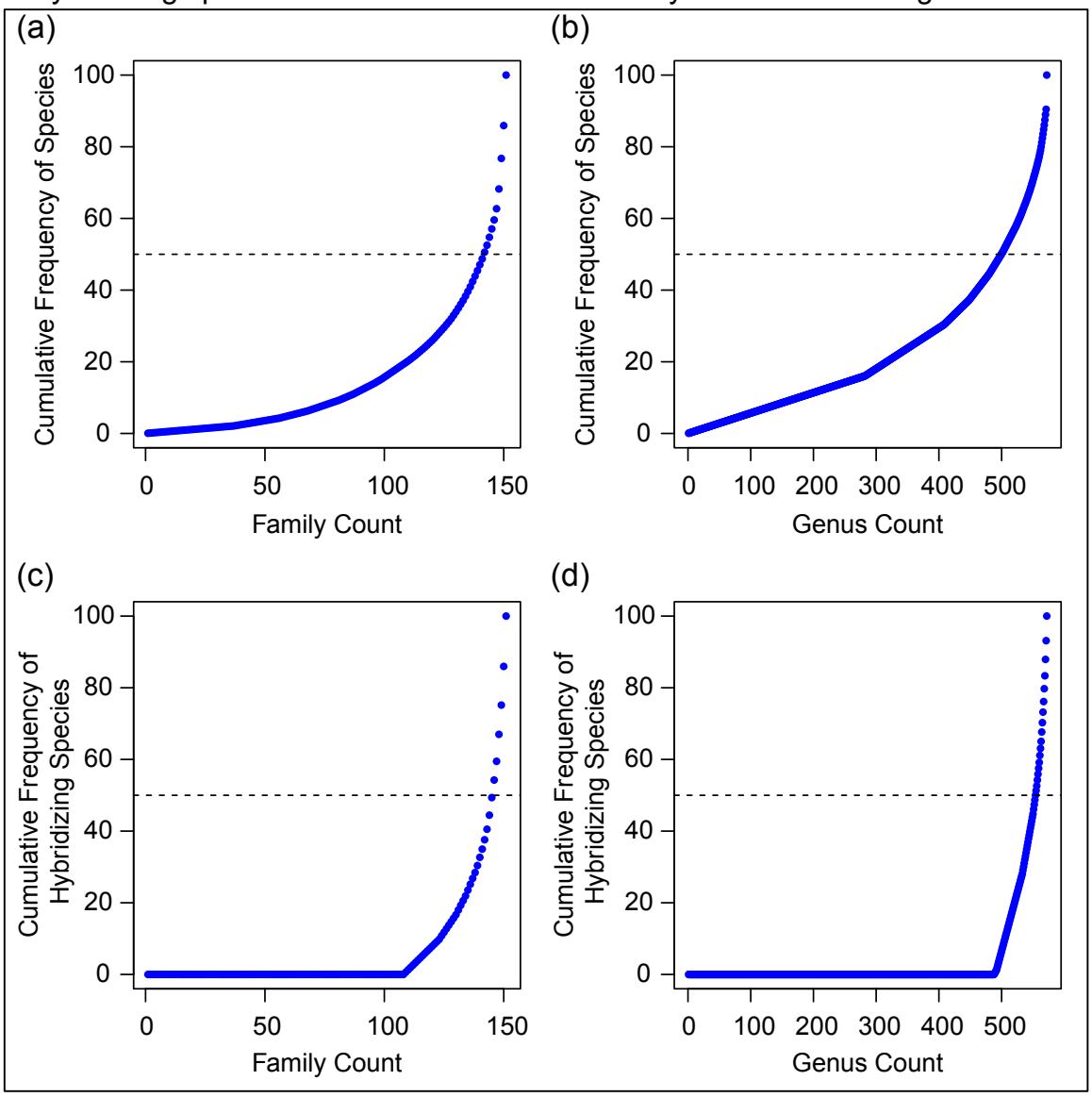


**Figure 3** Life form vs hybridization frequency. Life forms differ significantly for hybridization frequency (Chi-sqr P = 4.46e-7). Fern allies, shrubs, and trees have the highest rates of hybridization.

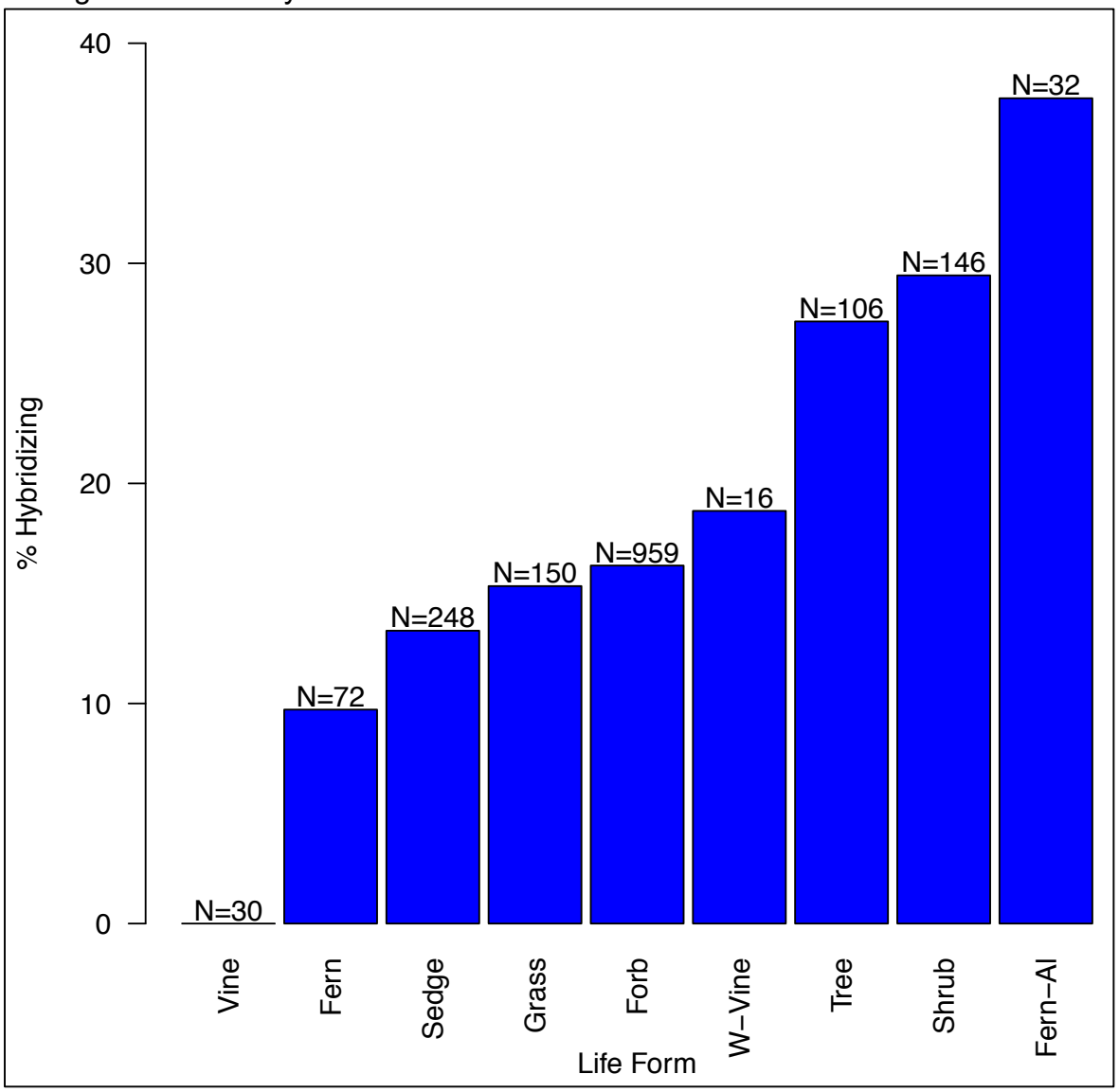


**Figure 4** Association of coefficient of conservatism (*C*) and hybridization rate. (**a**) Percent of species hybridizing by *C*. (**b**) The proportion of species which are annual (A), biennial (B), or perennial (P) for each coefficient of conservatism class. The greater proportion of annual species could explain the low hybridization rate for species with a *C* of 0. (**c**) The number of herbarium collections by *C*.

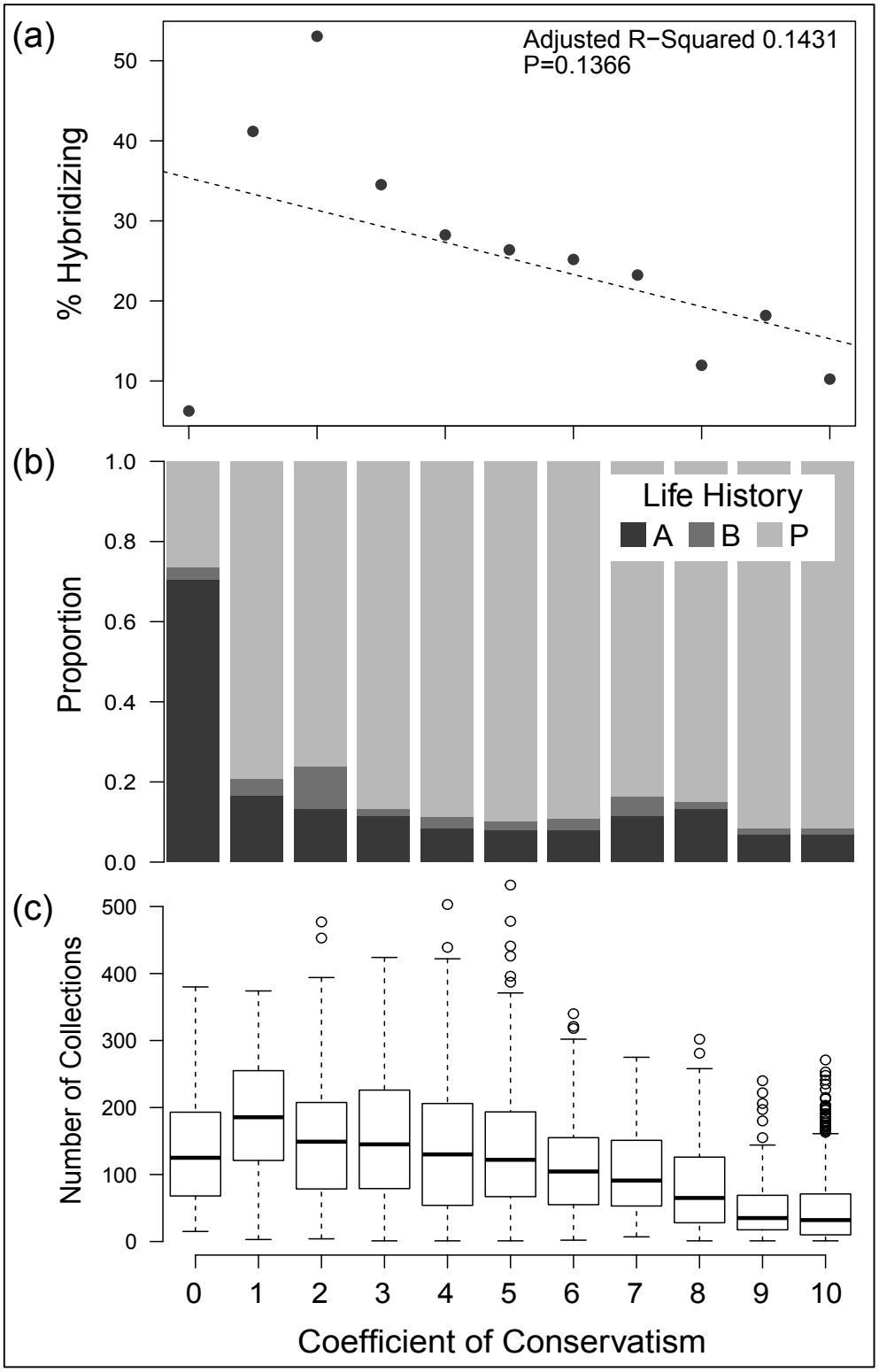
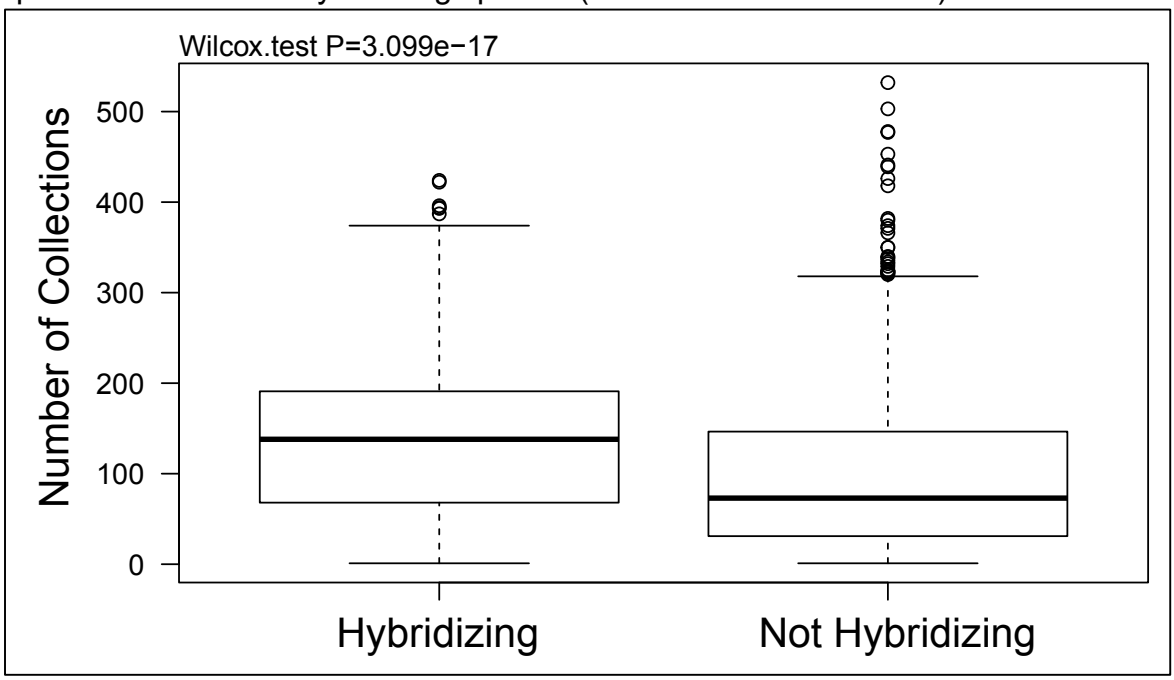


**Figure 5** Association of herbarium collection number and hybridization rate. The number of collections per species is significantly greater for hybridizing species than for non-hybridizing species (Wilcoxon test P=3.1e-17).



## 2.11. Tables

**Table 1**. Summary of Michigan's hybridizing species and unique hybrids.

| | *Families* | | | *Genera* | | | Species hybridizing | Non-hybrid species number | % of species hybridizing | *Unique hybrids* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total number | Number with hybrids | % with hybrids | Total number | Number with hybrids | % with hybrids | | | | Number | Hybrid ratio[†] |
| Native | 151 | 47 | 31.1% | 572 | 92 | 16.1% | 306 | 1747 | 17.5% | 221 | 0.127 |
| Adventive | 116 | 18 | 15.5% | 536 | 29 | 5.4% | 67 | 1070 | 3.9% | 24 | 0.022 |
| Total | 174 | 52 | 29.8% | 922 | 102 | 11.1% | 373 | 2832 | 13.2% | 261 | 0.092 |

[†]Excluding species of hybrid origin

**Table 2**. Examples of hybrid species in Michigan (N=12) and the supporting evidence.

| *Hybrid Species* | *Parents* | *Evid-ence*[a] | *Citation* |
|---|---|---|---|
| *Betula murrayana* (2n=8x=112) | Unreduced B. ×*purpussii* (2n=5x=70) × *B. alleghaniensis*. This is a fertile backcross because B. ×*purpussii* = *B. pumilla* (2n=4x=56) × *B. alleghaniensis* (2n=6x=84). | M,C | (Barnes and Dancik, 1985) |
| *Boechera grahamii* (2n=3x/4x) | *B. collinsii* (2x) × *B. stricta* (2x/3x). (*B. grahamii* = Arabis ×*divaricarpa*, *B. collinsii* = *A. holboellii*, and *B. stricta* = *A. drummondii*) | M,P,N | (Dobes et al., 2004; Koch et al., 2003) |
| *Cardamine maxima* (2n=?) | *C. concatenata* (2n=32x=256) × *C. diphylla* (2n=32x=256). *C. maxima* (Large toothwort) has not been found with mature fruits or seeds and thought to be sterile. | M,P[b] | (Carlsen et al., 2009; Sweeney and Price, 2000) |
| *Cyperus houghtonii* (2n=?) | *C. lupulinus* (2n=?) × *C. schweinitzii* (2n=?). C. houghtonii found only in jack pine barrens. | M,E | (Marcks, 1974) |
| *Drosera anglica* (2n=4x=40) | *D. linearis* (2n=2x=20) × *D. rotundifolia* (2n=2x=20). Diploid hybrids frequent but sterile. | M,C | (Crowder et al., 1990; Wood, 1955) |
| *Dryopteris carthusiana* (2n=164) | *D. intermedia* (2n=82) × *D. ludoviciana* (2n=82). | M,C,P,N | (Juslen et al., 2011; Runk et al., 2012) |
| *Goodyera tesselata* (2n=4x=60) | *G. repens* (2n=2x=30) × *G. oblongifolia* (2n=2x=30). | M,C | (Kallunki, 1976) |
| *Gymnocarpium dryopteris* (2n=160) | *G. disjuncta* (2n=80) × *G. appalachianum* (2n=80). | M,C,N | (Pryer and Haufler, 1993) |
| *Packera insulae-regalis* (2n=6x=66) | *P. paupercula* (2n=4x=44) × *P. indecora (*2n=8x=88). Local to Isle Royal. | M,C | (Kowal et al., 2011) |
| *Platanthera huronensis* (4n=84) | *P. aquilonis* (2n=42) × *P. dilatata* (2n=42). Multiple origins. | M,C,P,N | (Wallace, 2003) |
| *Solidago houghtonii* (2n=6x=54) | Three parents: *S. ptarmicoides* × *S. riddellii* × *S. ohioensis* (all 2n=8x=18). Upper Great Lakes only. Exact history not specified. | M,C,P,N | (Laureto and Barkman, 2011) |
| *Solidago vossii* (2n=8x=72) | *S. houghtonii* (2n=2x=18) × *S. ptarmicoides* (2n=2x=18). | M,C,E | (Laureto and Pringle, 2010) |

[a]Evidence for hybrid speciation. M = Morphology, C = Cytology, P = Chloroplast DNA Markers, N = Nuclear DNA Markers, E = Ecology/Habitat differences

[b]Sweeney and Price (2000) showed *C. maxima* was distinct, not that it is of hybrid origin; Carlsen et al. (2009) did not include *C. maxima*.

**Table 3**. The 20 families with the highest number of species hybridizing.

| *Family* | *Number Species Hybridizing* | *Number Species Total* | *Hybrid Ratio* | *Hybridization Frequency* | *Hybridization Propensity* |
|---|---|---|---|---|---|
| Asteraceae | 46 | 137 | 0.29 | 34% | 9% |
| Cyperaceae | 33 | 241 | 0.10 | 14% | 1% |
| Rosaceae | 26 | 87 | 0.20 | 30% | 7% |
| Poaceae | 23 | 119 | 0.10 | 19% | 4% |
| Violaceae | 16 | 22 | 0.82 | 73% | 8% |
| Orchidaceae | 15 | 44 | 0.25 | 34% | 10% |
| Salicaceae | 15 | 23 | 0.39 | 65% | 7% |
| Lamiaceae | 12 | 35 | 0.23 | 34% | 11% |
| Fagaceae | 9 | 11 | 1.36 | 82% | 27% |
| Potamogetonaceae | 8 | 27 | 0.15 | 30% | 8% |
| Fabaceae | 7 | 32 | 0.22 | 22% | 10% |
| Lycopodiaceae | 7 | 17 | 0.24 | 41% | 24% |
| Gentianaceae | 6 | 9 | 0.44 | 67% | 39% |
| Ranunculaceae | 5 | 9 | 0.09 | 15% | 8% |
| Hypericaceae | 5 | 13 | 0.29 | 36% | 18% |
| Apocynaceae | 5 | 14 | 0.31 | 38% | 18% |
| Betulaceae | 5 | 33 | 1.11 | 56% | 56% |
| Onagraceae | 4 | 8 | 0.10 | 20% | 13% |
| Cornaceae | 4 | 20 | 0.38 | 50% | 11% |
| Anacardiaceae | 4 | 7 | 0.29 | 57% | 24% |

**Table 4.** The 20 genera with the highest number of hybridizing species.

| *Genus* | *Number Species Hybridizing* | *Number Species Total* | *Hybrid Ratio* | *Hybridization Frequency* | *Hybridization Propensity* |
|---|---|---|---|---|---|
| *Carex* | 21 | 168 | 0.09 | 13% | 0% |
| *Viola* | 16 | 22 | 0.82 | 73% | 8% |
| *Dichanthelium* | 14 | 21 | 0.29 | 67% | 3% |
| *Symphyotrichum* | 11 | 20 | 0.5 | 55% | 5% |
| *Salix* | 11 | 18 | 0.28 | 61% | 3% |
| *Quercus* | 9 | 11 | 1.36 | 82% | 27% |
| *Rubus* | 9 | 12 | 0.42 | 75% | 8% |
| *Solidago* | 8 | 22 | 0.32 | 36% | 3% |
| *Platanthera* | 8 | 15 | 0.4 | 53% | 6% |
| *Helianthus* | 6 | 10 | 0.5 | 60% | 11% |
| *Amelanchier* | 6 | 6 | 0.67 | 100% | 27% |
| *Potamogeton* | 6 | 25 | 0.12 | 24% | 1% |
| *Betula* | 5 | 5 | 2 | 100% | 100% |
| *Lespedeza* | 5 | 5 | 1.2 | 100% | 60% |
| *Scirpus* | 5 | 9 | 0.44 | 56% | 11% |
| *Rosa* | 5 | 6 | 0.5 | 83% | 20% |
| *Cornus* | 4 | 8 | 0.38 | 50% | 11% |
| *Gentiana* | 4 | 5 | 0.6 | 80% | 30% |
| *Hieracium* | 4 | 7 | 0.43 | 57% | 14% |
| *Acer* | 4 | 7 | 0.29 | 57% | 10% |

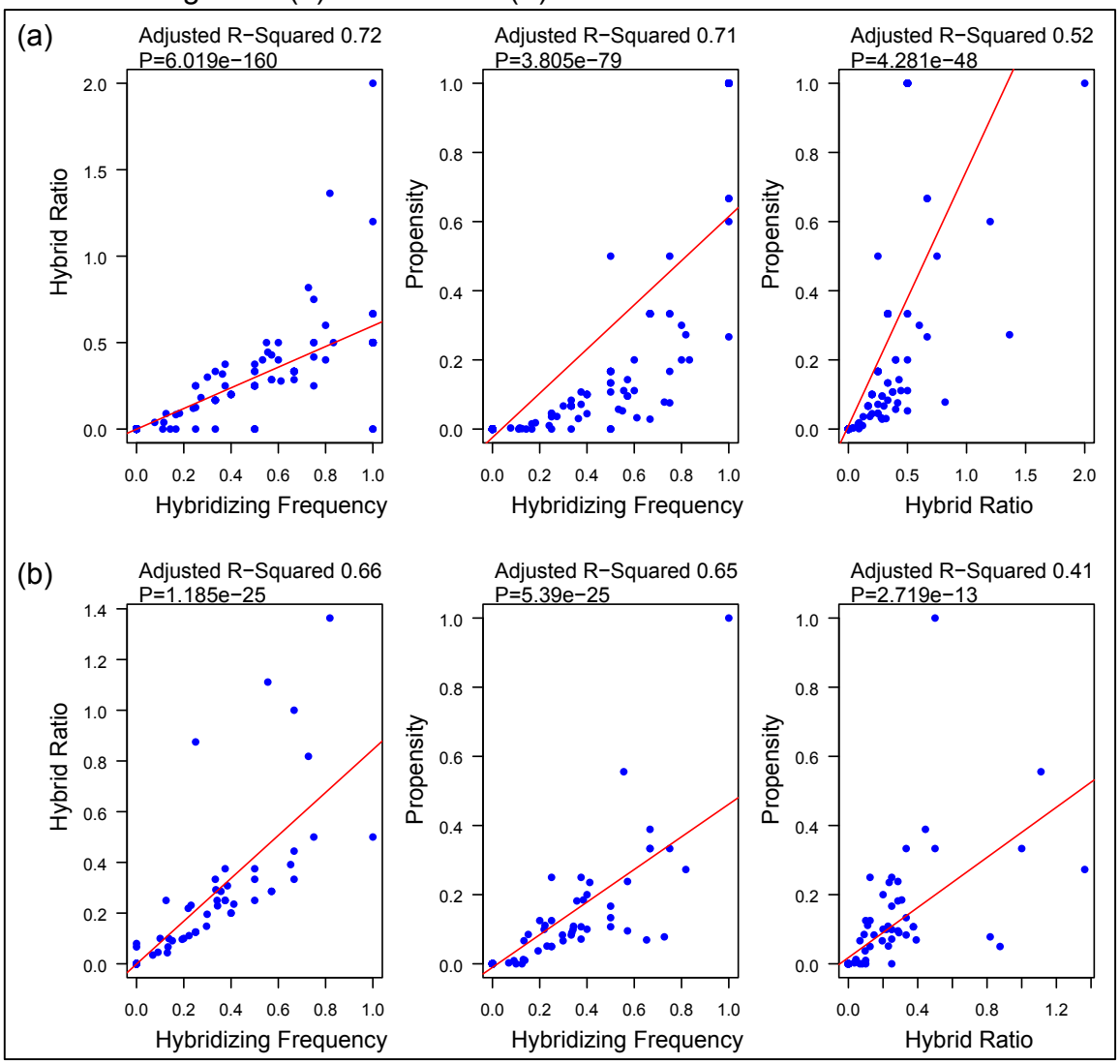
**Table 5.** Significant association between life history and hybridization frequency.

| *Life History* | *Hybridizing*[†] | *Not Hybridizing*[†] | *Hybridization Frequency* | *% of all hybrids* | *Average # of collections* |
|---|---|---|---|---|---|
| *Annual* | 5 | 181 | 2.68% | 1.66% | 70 ± 66 |
| *Biennial* | 3 | 43 | 6.52% | 1.00% | 94 ± 64 |
| *Perennial* | 292 | 1223 | 19.28% | 97.33% | 111 ± 91 |

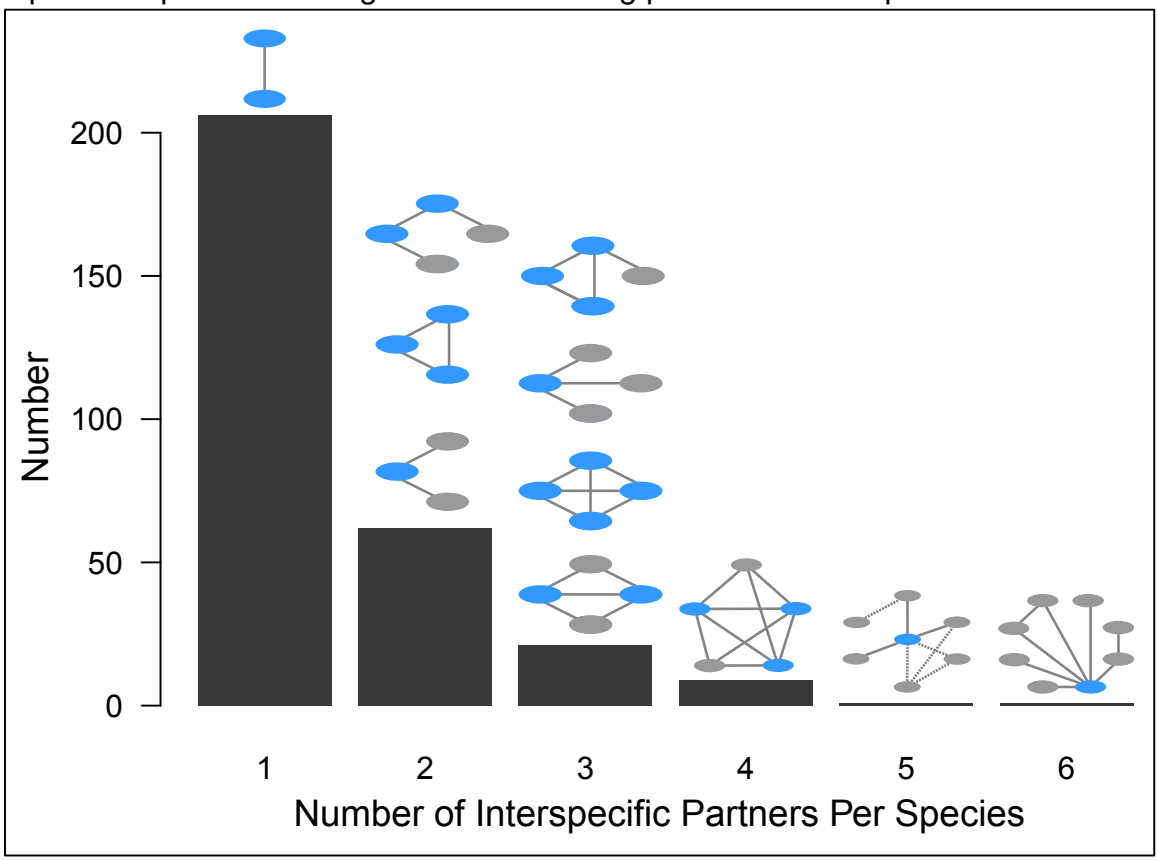
[†]Excluding species of hybrid origin

## 2.12. Supplemental Figures

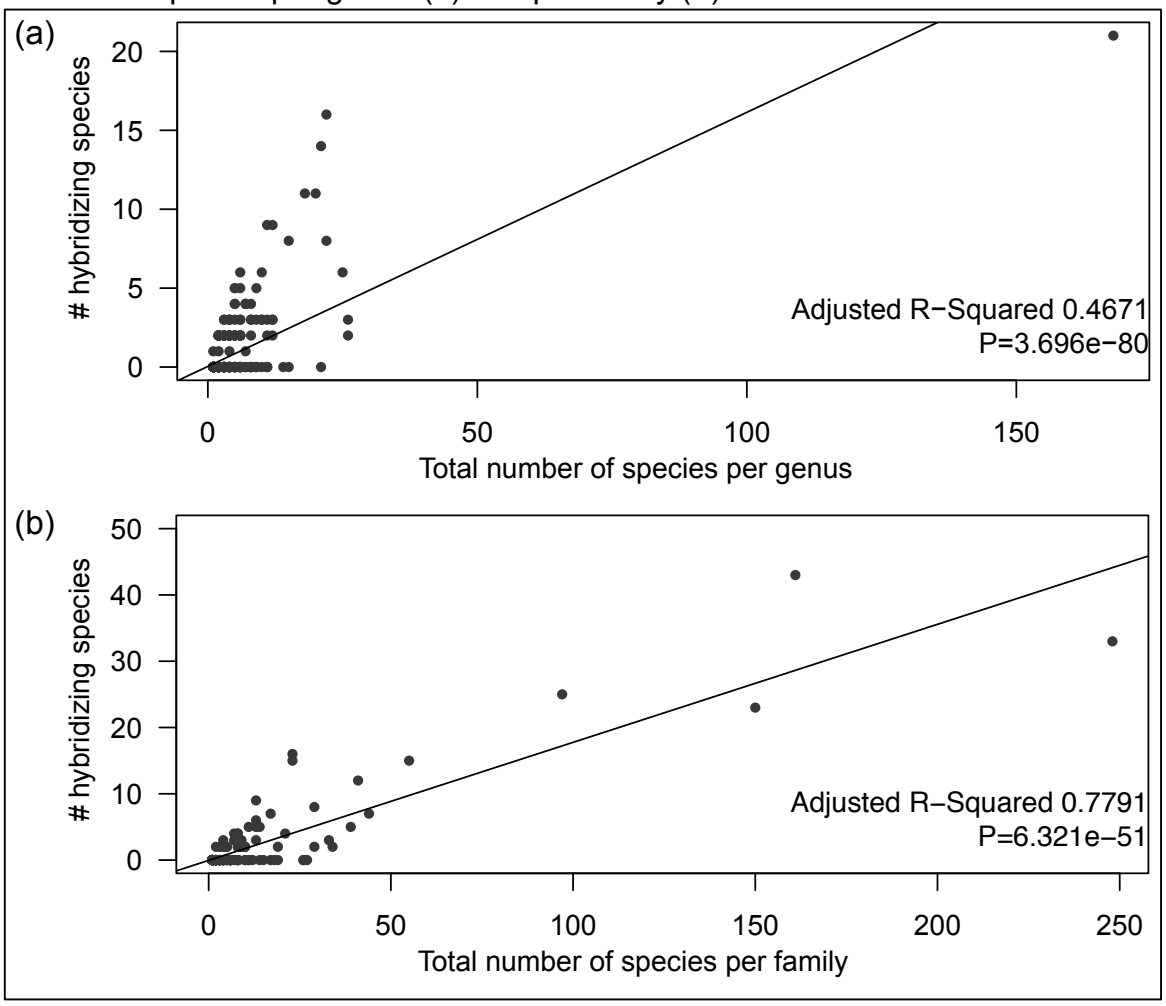
**Supplemental Figure 1** Relationships between different metrics of hybridization for individual genera (**a**) and families (**b**).



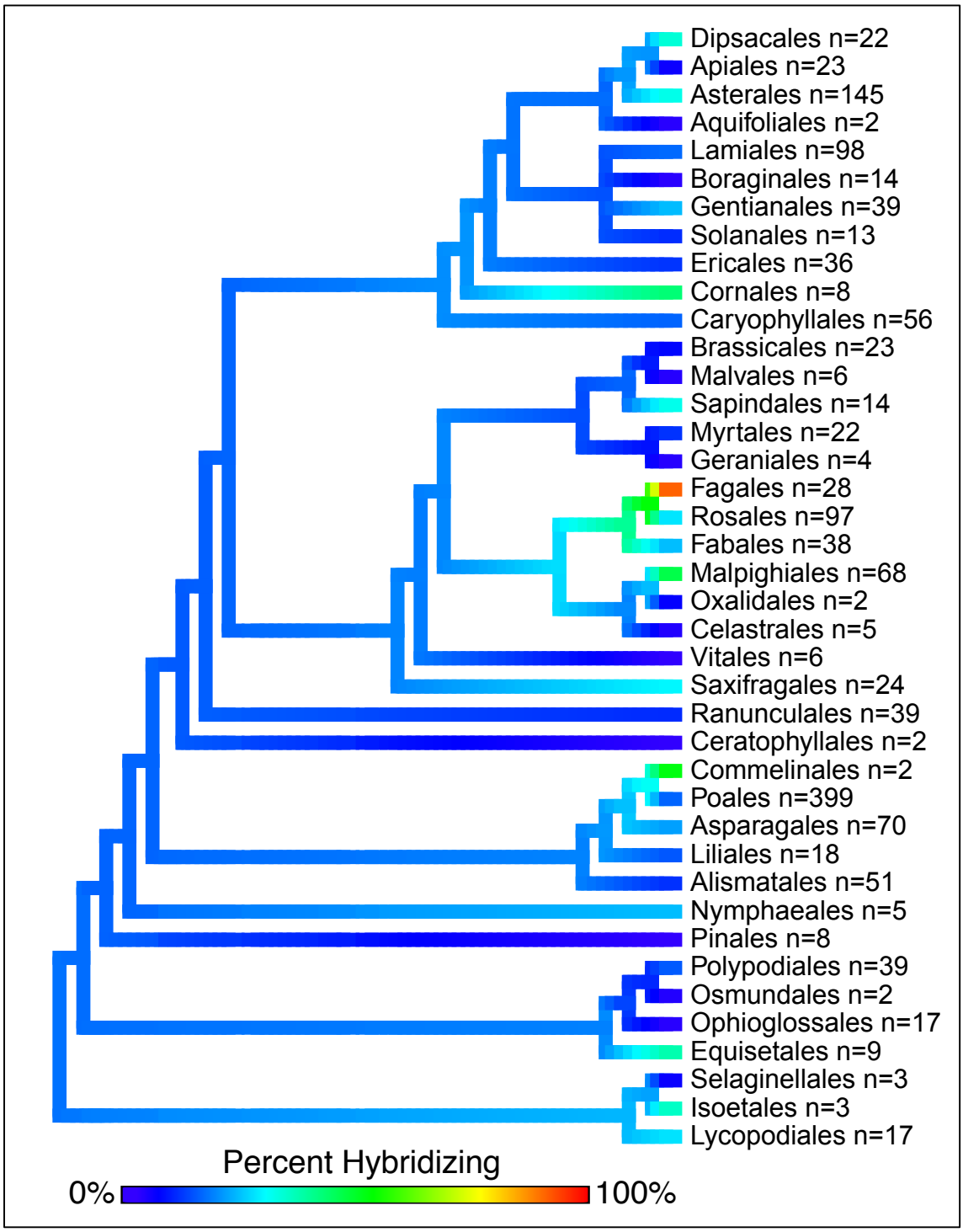
**Supplemental Figure 2** Histogram of the number interspecific mating partners per species. The median number of partners per species is one. A total of 32 species (1.8% of all species) had more than two partners and only 11 species in four genera had four or more interspecific partners. The most commonly observed networks for each number of interspecific matings are also depicted; nodes represent species and edges connect mating partners. Focal species are in blue.



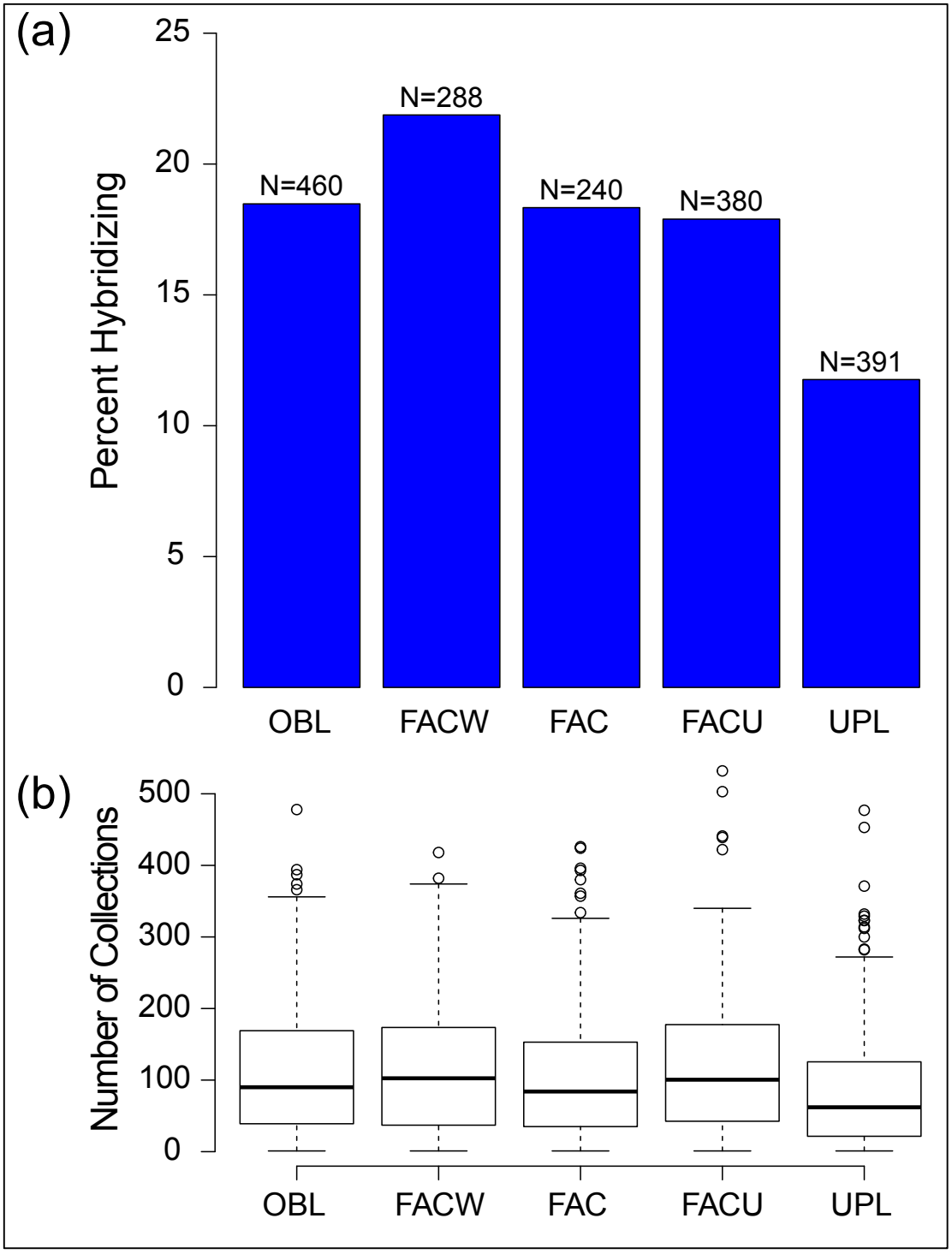
**Supplemental Figure 3** Relationship between total species and total hybrids. The y-axis shows the number of hybridizing species and the x-axis shows the total number of species per genus (**a**) and per family (**b**).



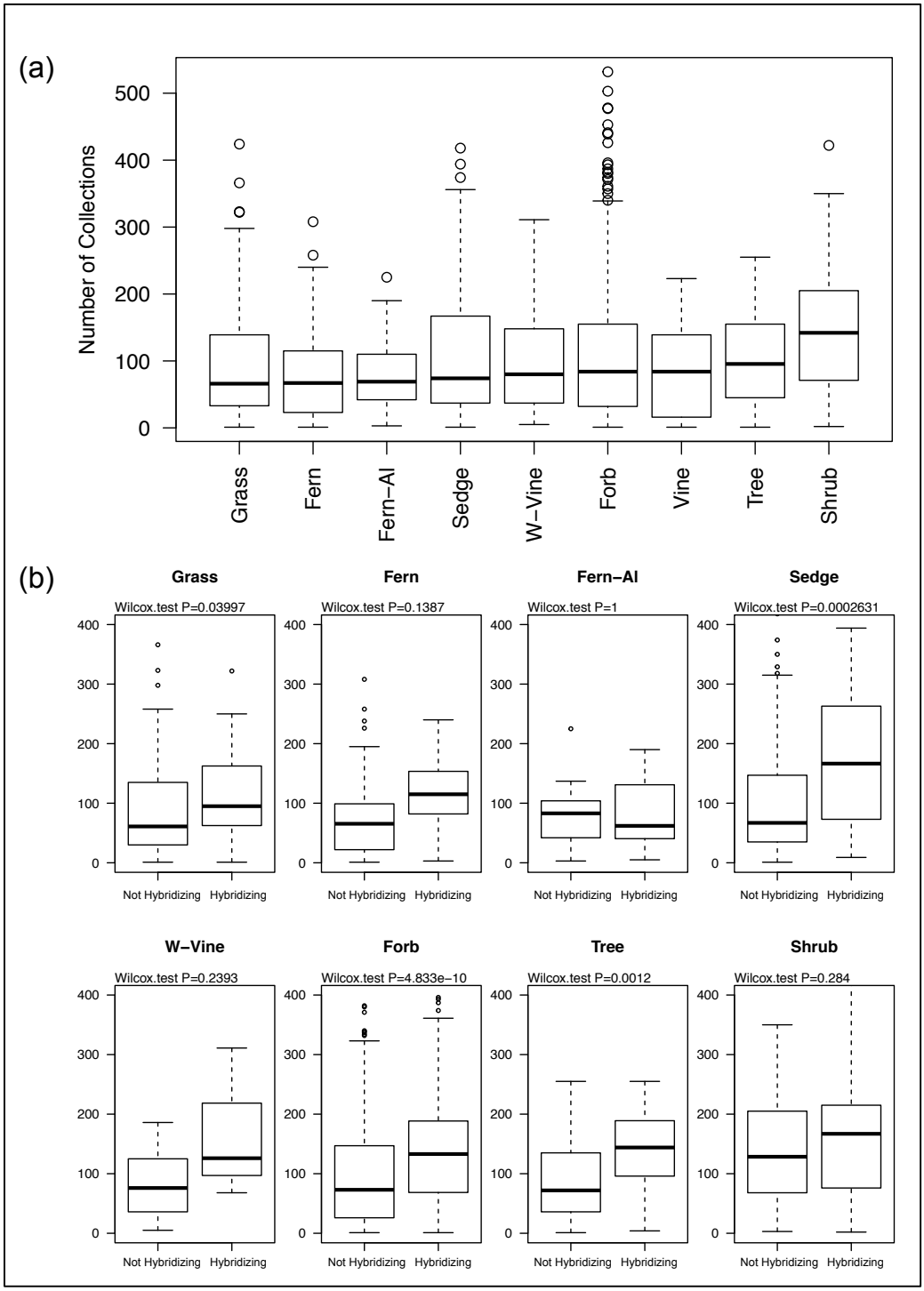
**Supplemental Figure 4** Hybridization frequency mapped to APG IV phylogeny. The hybridization frequency for all taxonomic orders (following the Angiosperm Phylogeny Group IV (APG IV) system) is the average hybridization frequency for all genera in the order. The per order hybridization frequencies were mapped onto the phylogeny and ancestral nodes were inferred using phytools v0.6-00. The tip label gives the taxonomic order and number of species in that taxonomic order. No phylogenetic signal for hybridization frequency (nor hybrid ratio nor hybridization propensity) was found using BayesTraits v3 under the maximum likelihood model.



**Supplemental Figure 5** Correlation between habitat wetness and hybridization frequency. (**a**) The frequency of hybridization was significantly different between wetness coefficients (Chi-sqr P = 0.010). (**b**) The number of collections per species by wetness coefficient. The number of upland collections was significantly less (Wilcoxon test P<0.01) than all other wetness coefficients; there were no other significant differences. OBL, obligate wetland; FACW, facultative wetland; FAC, facultative; FACU, facultative upland; UPL, obligate upland.



**Supplemental Figure 6** Number of herbarium collections per by life form. (**a**) Boxplots for the number of collections for all life forms. Shrubs are significantly different (Wilcoxon test P<0.01) from all other life forms except woody vines; there are no other significant differences. (**b**) Boxplots for the number of collections per life form for hybridizing and non-hybridizing species. The Wilcoxon rank sum test p-value was significant (<0.01) for sedges, forbs, and trees only; in these cases, hybridizing species have significantly more collections.



## 2.13. Supplemental Tables

**Supplemental Table 1.** There are 78 named hybrids in 43 genera that are not included in the flora of Michigan.

| *Genus* | *Named Hybrid* | *Parent 1* | *Parent 2* |
|---|---|---|---|
| *Actaea* | *×ludovici* | *A. rubra* | *A. pachypoda* |
| *Ambrosia* | *×helenae* | *A. artemisiifolia* | *A. trifida* |
| *Apocynum* | *×floribundum* | *A. cannabinum* | *A. androsaemifolium* |
| *Betula* | *×purpusii* | *B. pumilla* | *B. alleghaniensis* |
| *Betula* | *×sandbergii* | *B. pumilla* | *B. papyrifera* |
| *Carex* | *×connectens* | *C. limosa* | *C. magellanica* |
| *Carex* | *×knieskernii* | *C. arctata* | *C. castanea* |
| *Carex* | *×subimpressa* | *C. pellita* | *C. hyalinolepsis* |
| *Cornus* | *×friedlanderi* | *C. foemina* | *C. rugosa* |
| *Cyperidium* | *×andrewsii* | *C. candidum* | *C. parviflorum* |
| *Drosera* | *×beleziana* | *D. intermedia* | *D. rotundifolia* |
| *Dryopteris* | *×boottii* | *D. cristata* | *D. intermedia* |
| *Dryopteris* | *×triploidea* | *D. carthusiana* | *D. intermedia* |
| *Epilobium* | *×wisconsinense* | *E. ciliatum* | *E. coloratum* |
| *Equisetum* | *×ferrissii* | *E. hyemale* | *E. laevigatum* |
| *Equisetum* | *×mackayi* | *E. variegatum* | *E. hyemale* |
| *Equisetum* | *×nelsonii* | *E. laevigatum* | *E. variegatum* |
| *Gentiana* | *×pallidocyanea* | *G. andrewsii* | *G. alba* |
| *Gentiana* | *×billingtonii* | *G. andrewsii* | *G. puberulenta* |
| *Gentiana* | *×grandilacustris* | *G. andrewsii* | *G. rubricaulis* |
| *Gymnocarpium* | *×intermedium* | *G. dryopteris* | *G. jessoense* |
| *Helianthus* | *×ambiguus* | *H. divaricatus* | *H. giganteus* |
| *Helianthus* | *×divariserratus* | *H. divaricatus* | *H. grosseserratus* |
| *Helianthus* | *×laetiflorus* | *H. pauciflorus* | *H. tuberosus* |
| *Helianthus* | *×luxurians* | *H. grosseserratus* | *H. giganteus* |
| *Huperzia* | *×buttersii* | *H. selago* | *H. lucidula* |
| *Huperzia* | *×josephbeitelii* | *H. selago* | *H. appressa* |
| *Iris* | *×robusta* | *I. versicolor* | *I. virginiana* |
| *Isoetes* | *×hickeyi* | *I. echinospora* | *I. lacustris* |
| *Lespedeza* | *×nuttallii* | *L. violaceae* | *L. hirta* |
| *Liatris* | *×gladewitzii* | *L. cylindracea* | *L. aspera* |
| *Lobelia* | *×speciosa* | *L. cardinalis* | *L. syphilitica* |
| *Lycopus* | *×sherardii* | *L. uniflorus* | *L. virginicus* |
| *Lysimachia* | *×commixta* | *L. terrestris* | *L. thyrsifolia* |
| *Lysimachia* | *×producta* | *L. terrestris* | *L. quadrifolia* |
| *Monarda* | *×media* | *M. didyma* | *M. fistulosa* |
| *Neottia* | *×veltmanii* | *N. auriculata* | *N. convallarioides* |
| *Nuphar* | *×rubrodisca* | *N. variegata* | *N. microphylla* |
| *Petasites* | *×vitifolius* | *P. sagittatus* | *P. frigidus* |
| *Platanthera* | *×andrewsii* | *P. lacera* | *P. psycodes* |
| *Platanthera* | *×bicolor* | *P. blephariglottis* | *P. ciliaris* |
| *Platanthera* | *×hollandiae* | *P. leucophaeae* | *P. lacera* |
| *Platanthera* | *×reznicekii* | *P. leucophaeae* | *P. psycodes* |
| *Platanthera* | *×vossii* | *P. blephariglottis* | *P. clavellata* |
| *Populus* | *×jackii* | *P. balsamifera* | *P. deltoides* |

| | | | |
|---|---|---|---|
| *Populus* | ×*smithii* | *P. tremuloides* | *P. grandidentata* |
| *Potamogeton* | ×*haynesii* | *P. strictifolius* | *P. zosteriformis* |
| *Proserpinaca* | ×*intermedia* | *P. pectinata* | *P. palustris* |
| *Quercus* | ×*bebbiana* | *Q. alba* | *Q. macrocarpa* |
| *Quercus* | ×*beckyae* | *Q. macrocarpa* | *Q. prinoides* |
| *Quercus* | ×*deamii* | *Q. macrocarpa* | *Q. muehlenbergii* |
| *Quercus* | ×*faxonii* | *Q. alba* | *Q. prinoides* |
| *Quercus* | ×*hawkinsiae* | *Q. rubra* | *Q. velutina* |
| *Quercus* | ×*jackiana* | *Q. alba* | *Q. bicolor* |
| *Quercus* | ×*leana* | *Q. imbricata* | *Q. velutina* |
| *Quercus* | ×*palaeolithicola* | *Q. ellipsoidalis* | *Q. velutina* |
| *Quercus* | ×*runcinata* | *Q. imbricata* | *Q. rubra* |
| *Quercus* | ×*schuettei* | *Q. bicolor* | *Q. macrocarpa* |
| *Quercus* | ×*wagneri* | *Q. bicolor* | *Q. prinoides* |
| *Rhus* | ×*pulvinata* | *R. typhina* | *R. glabra* |
| *Rosa* | ×*palustriformis* | *R. blanda* | *R. palustris* |
| *Rubus* | ×*fraseri* | *R. parviflorus* | *R. odoratus* |
| *Rubus* | ×*neglectus* | *R. strigosus* | *R. occidentalis* |
| *Rubus* | ×*paracaulis* | *R. acaulis* | *R. pubescens* |
| *Salix* | ×*glatfelteri* | *S. nigra* | *S. amygdaloides* |
| *Schoenoplectus* | ×*contortus* | *S. americanus* | *S. pungens* |
| *Scutellaria* | ×*churchilliana* | *S. laterifolia* | *S. galericulata* |
| *Solidago* | ×*krotkovii* | *S. ptarmicoides* | *S. ohioensis* |
| *Stuckenia* | ×*suecica* | *S. filliformis* | *S. pectinata* |
| *Symphyotrichum* | ×*amethystinum* | *S. novae-angliae* | *S. ericoides* |
| *Verbena* | ×*engelmannii* | *V. urticifolia* | *V. hastata* |
| *Viola* | ×*braunie* | *V. rostrata* | *V. striata* |
| *Viola* | ×*subsinuata* | *V. pedatifida* | *V. sororia* |
| *Viola* | ×*palmata* | *V. sororia* | *V. sagittata* |
| *Viola* | ×*primulifolia* | *V. lanceolata* | *V. macloskeyi* |
| *Viola* | ×*eclipes* | *V. labradorica* | *V. striata* |
| *Viola* | ×*malteana* | *V. labradorica* | *V. rostrata* |
| *Woodsia* | ×*abbeae* | *W. ilvensis* | *W. oregano* |

**Supplemental Table 2**. Information on all unique hybrids in the Field Manual of Michigan Flora.

| *Genus / Unique Hybrid* | *Parental Species 1* | *Parental Species 2* | *Information* | *Citation* | *Total # of Unique Hybrids* |
|---|---|---|---|---|---|
| ×*Calammophila don-hensonii* | *Ammophila breviligulata* | *Calamagrostis canadensis* | The hybrid ×*Calammophila don-hensonii* is the only intergeneric hybrid in Michigan described by (Reznicek and Judziewicz 1996). This hybrid is described from a collection on Grand Island, Michigan. An individual intermediate between the hybrid and C. canadensis is thought to be a backcross indicating that the hybrid is not fully sterile. Normal fruits were also observed on the hybrid. The parental species are common in Michigan. | (Reznicek and Judziewicz 1996) | 1 |
| *Acer* | Multiple | Multiple | From seven native species, two hybrid combinations are given: *Acer saccharinum* × *rubrum* and *A. nigrum* × *saccharum*. Note that black maple is only recently circumscribed from sugar maple and is sometimes considered *A. saccharum* subsp. *nigrum*. | | 2 |
| *Achillea* | *A. millefolium* | None | *Achillea millefolium* is the only native yarrow and is therefore not included as a hybrid. However, there are distinct subpopulations of this widespread species in Michigan, different ploidy levels, and this species is known to hybridize elsewhere. Therefore, there may be yarrow hybrids in Michigan. | | 0 |
| *Actaea* ×*ludovici* | *A. rubra* | *A. pachypoda* | The parental species are quite distinct in pedicel morphology (A. pachypoda pedicels being thicker) and (normally) fruit color. Hybrids uncommon. The third baneberry species in the flora, *A. racemosa*, is not said to hybridize. | | 1 |
| Ajuga ×hybrida | A. genevensis | A. reptans | *Ajuga* ×*hybrida* is the hybrid of the only two *Ajuga* (bugle) species in the flora. | | 1 |
| *Amaranthus* | Multiple | Multiple | There are two native and 10 adventive *Amaranthus* species. Four combinations involving five species are given: *A. hybridus* × *A. tuberculatus*, *A. tuberculatus* × *A. retroflexus, A. caudatus* × *cruentus*, and *A. caudatus* × *retroflexus* – all hybrids apparently sterile. | | 4 |

| | | | | | |
|---|---|---|---|---|---|
| *Ambrosia* | *A. artemisiifolia* | Multiple | *Ambrosia* ×*helenae* is the hybrid of the two native ragweeds, *A. artemisiifolia* (common ragweed) and *A. trifida* (giant ragweed). The third ragweed in the flora, *A. psilostachya* (western ragweed), is not native to Michigan but hybridizes readily with *A. artemisiifolia* producing *A.* ×*intergradiens*. This is a more common hybrid in Michigan than *A.* ×*helenae*. | | 2 |
| *Amelanchier* | Multiple | Multiple | All six of the flora's *Amelanchier* species are native and appear to freely cross. This has led to systematic problems in many floras; a situation greatly complicated by mostly asexual reproduction. There are three combinations, all involving *A. laevis*, given: *A. laevis* × *arborea, A. laevis* × *bartramiana, A. laevis* × *spicata*. *Amelanchier* interior is also thought be a hybrid swarm involving crosses between *A. laevis/arborea* and *A. spicata/sanguinea*. This is included as one unique hybrid, and *A. laevis* is considered both as hybridizing to be of hybrid origin. | (Nielsen 1939) | 4 |
| *Apocynum* | *A. cannabinum* | *A. androsaemifolium* | The two native *Apocynum* species are known to hybridize. The authors note that intermediate plants are hybrids but that some individuals have character combinations from both (e.g. flowers of one parent with habit of the second) probably resulting from introgressions into the parent species. This make herbarium specimen identification difficult. There are no introduced species in this genus. | | 1 |
| *Aronia* ×*prunifolia* | *A. melanocarpa* | *A. arbutifolia* | *A. melanocarpa* and *A. arbutifolia* are thought to cross and give *A.* ×*prunifolia*. Hybrid populations seem to form independent, self-reproducing populations. However, *A.* ×*prunifolia* is not really distinct from *A. melanocarpa*. Only *A. prunifolia* and not *A. arbutifolia* are included in the flora. *A. prunifolia* is quantified as hybridizing but not itself a hybrid. | | 1 |
| *Asclepias* | *A. syriaca* | Multiple | From 10 native species of milkweed, two are said to hybridize with the common milkweek, *Asclepias syriaca*: *A. syriaca* × *exaltata* and *A. syriaca* × *purpurascens*. The only adventive *Asclepias* species in the flora, *A. speciosa*, is native from western Minnesota westward and was collected in Mackinaw county where it was hybridizing with *A. syriaca*; the hybrid was said to be intermediate in phenotype. *A. tuberosa* is known to hybridize elsewhere (Wyatt and Antonovics 1981), but these may be subspecies. | | 3 |
| *Aureolaria* | *A. flava* | *A. pedicularia* | The hybrids are said to be uncommon. No hybrids are given for the third native species, *A. viriginica*. | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| *Berberis ×ottawensis* | *B. thunbergii* | *B. vulgaris* | There are no native and four introduced barberry species. | | 1 |
| *Betula* | Multiple | Multiple | Betula is very cross compatible genus often with morphological continuity between species. The situation is greatly complicated by having anywhere from diploid to dodecaploid taxa. In Michigan, there are four main taxa: white, yellow, bog, and gray birch; and all five species hybridize in all combinations. There are also two named hybrids and one hybrid species. First, *B. ×purpusii* (2n=5x=70, sterile) is *B. pumilla* (2n=4x=56) × *alleghaniensis* (2n=6x=84). The hybrid species is *B. murrayana* (2n=8x=112) which is the fertile octoploid backcross derived from an unreduced *B. ×purpusii* gamete and a reduced *B. alleghaniensis* one. x=14 in *Betula*. The second named hybrid is B. ×sandbergii (*B. pumila* × *papyrifera*). | (Barnes and Dancik 1985; Walters and Yawney 2004) | 10 |
| *Boechera grahamii* | *B. collinsia* | *B. stricta* | The rock cresses constitute a taxonomically difficult genus; all seven native species were formerly in *Arabis*. *Boechera grahamii* is said from morphological evidence to be the hybrid of *B. collinsia* and *B. stricta*. Note that *B. collinsii* is native to the west and north and not included in Michigan's flora. | (Flora of North America Editorial Committee 1993+-b) | 0 |
| *Bromus* | *B. pubescens* | *B. ciliatus* | There are six native and 11 adventitious *Bromus* species. Only one individual from a tamarack swamp is intermediate and suspected to be a hybrid of native *B. pubescens* and *B. ciliatus*. | | 1 |
| Calystegia | *C. silvatica* | *C. sepium* | There are three adventitious and one native *Caystegia* in the flora. Hybrids between *C. silvatica* and native *C. sepium* were identified from morphology. | | 1 |
| *Cardamine maxima* | *C. concatenata* | *C. diphylla* | There are eight native and three introduced toothworts. *Cardamine maxima* is morphologically intermediate between *C. concatenata* and *C. diphylla* and thought to be their hybrid. This species has been shown to be genetically distinct from the parents by (Sweeney and Price 2000), but there is no evidence other than morphology that it is of hybrid origin. | (Sweeney and Price 2000) | 1 |
| *Carex* | *Multiple* | *Multiple* | There are 171 native (and 13 non-native) species of *Carex* in the flora. Twenty-one of which are said to hybridize in 15 combinations. There are also three named hybrids: *C. arctata* × *castanea* (*C. ×knieskernii*) – sterile. *C. atherodes* × *trichocarpa. C. blanda* × *laxiflora* – sterile. *C. castanea* × *gracillima. C. comosa* × *hystericina* – sterile. *C. comosa* × *pseudocyperus. C. cryptolepis* × *viridula. C. hyalinolepis* × *lacustris, C. hyalinolepis* × *pellita* (*C. ×subimpressa*) – sterile. *C. hystericina* × *pseudocyperus* – sterile. *C. lacustris* × *pellita, C. limosa* × *magellanica* (*C. ×connectens*). *C. lupulina* × *lurida* – sterile. *C. lupulina* × *retrorsa* – sterile. *C. lupulina* × *vesicaria* - sterile. | | 15 |

| | | | | | |
|---|---|---|---|---|---|
| *Centaurea* | Multiple | Multiple | All 10 *Centaurea* species are non-native. Two named hybrid combinations are given: *C. diffusa* × *stoebe* (*C.* ×*psammogena*) and *C. jacea* × *nigra* (*C.* ×*moncktonii*). Both hybrids are common and fertile and seem to backcross with their parents: "the hybrid and *C. diffusa* grade insensibly into one another" and "there is full gradation toward the parents, especially *C. jacea*." Quote from Reznicek and Voss (2012). | | 2 |
| *Circaea* ×*sterilis* | *C. alpina* | *C. canadensis* | The only two *Circaea* species in the flora hybridize and as the name suggests, their hybrid is sterile and its ovaries abort. The hybrid is morphologically intermediate. | | 1 |
| *Coreopsis* | *C. grandiflora* | *C. lanceolata* | There are three native and two introduced *Coreopsis* taxa in the flora. Hybrids between *C. lanceolata* and native *C. grandiflora* are suspected from morphology. | | 1 |
| *Cornus* | Multiple | Multiple | There are eight native dogwoods. Four of these hybridize in three combinations: *C. foemina* × *rugosa* (*C.* ×*friedlanderi*), *C. drummondii* × *amomum*, and *C. amomum* × *foemina*. The *C. amomum* × *foemina* hybrids have mostly infertile pollen. | | 3 |
| *Crataegus* | Multiple | Multiple | There are 26 native hawthorns. The genus is difficult with characteristics similar to *Rubus* (i.e. cross-compatibility, agamospermy). The only native hybrid combination given is *C. punctata* × *crus-galli*. These hybrids were once classified as *C. disperma* which is now removed from the flora. *Crataegus nitidula* may also be a hybrid of *C. punctata* and another species, but is given as synonymous with *C. punctata* in the flora. Another questionable hybrid is *C. colea* which "may be an intersectional hybrid," but no more information is available and it is therefore not included as a hybrid. One combination involving only introduced taxa is given: *C. laevigata* × *monogyna* (*C.* ×*media*). The final combination is between *C. monogyna* and native *C. margaretta*. | | 3 |
| *Cyperus* | Multiple | Multiple | There are 13 native and two non-native species in this sedge genus. Three native species are said to hybridize: *C. houghtonii* is the stablizied hybrid of *C. lupulinus* and *C. schweinitzii*, and there is apparent backcrossing and introgression because some *C. schweinitzii* individuals approach *C. houghtonii* in culm and fruit characters. | (Marcks 1974) | 3 |
| *Cypripedium* ×*andrewsii* | *C. candidum* | *C. parviflorum* | There are five native lady slippers with the white (*C. candidum*) and yellow (*C. parviflorum*) hybridizing. The hybrid is said to and intermediate corolla color. | | 1 |
| *Desmodium* | *D. paniculatum* | *D. canadense* | This is the only hybrid combination from 12 native trefoil species. The leaf and fruit characters of the two parents are combined (not intermediate) in the hybrid. | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| *Dichanthelium* | Multiple | Multiple | There are 21 native species of *Dichanthelium*, a large and confusing genus. Six hybrid combinations from nine native taxa are given by (Reznicek and Voss 2012): *D. praecocius* × *implicatum. D. implicatum* × *lindheimeri. D. perlongum* × *depauperatum. D. perlongum* × *linearifolium. D. linearifolium* × *latifolium. D. xanthophysum* × *boreale* (*D.* ×*calliphyllum*, sterile triploid). An additional three crosses are recognized by the Utah State Herbarium which was consulted to help with this genus: *D. commutatum* × *dichotomum. D. clandestinum* × *dichotomum* (*D.* ×*recognitum*, sterile hybrid), *D. polyanthes* × *sphaerocarpon*. There may be more hybrid combinations in this genus (i.e. *D. lindheimeri*). | (Freckmann and Lelong) | 9 |
| *Diphasiastrum* ×*sabinifolium* | *D. sitchense* | *D. tristachyum* | This clubmoss is the only native hybrid taxa recognized in the flora (as denoted by an × before the species name). It is included in the flora because one of the parents, *D. sitchense*, is common on the north shore of Lake Superior, but not known in Michigan. The hybrid is known from one county only. | | 0 |
| *Drosera* | Multiple | Multiple | There are four native sundews including one hybrid species, D. *anglica*, which is the tetraploid derivative of diploids *D. linearis* × *rotundifolia*. Diploid F1 hybrids are sterile. There is also the named hybrid, *D.* ×*beleziana*, from the cross of *D. intermedia* and *D. rotundifolia*. The fourth species in the flora, native *D. linearis*, is not said to hybridize. *D. anglica* does hybridize to both of its parents, but this was discovered too late for inclusion in the analyses. | (Wood 1955; Crowder et al. 1990) | 2 |
| *Dryopteris* | Multiple | Multiple | Three of 10 native log ferns are said to hybridize. First, *D. carthusiana* is a stabilized tetraploid hybrid from diploid *D. intermedia* and an unknown *Dryopteris* species (probably *D. ludoviciana* from North America). The backcross of *D. carthusiana* to *D. intermedia* is a named sterile triploid *D.* ×*triploidea*. A second hybrid taxa said to be easily recognizable is *D.* ×*boottii* derived from *D. cristata* × *intermedia*. | (Runk et al. 2012) | 3 |
| *Elymus* | Multiple | Multiple | From 10 native species of wild rye, two are said to hybridize with *E. virginicus*: *E. riparius and E. hystrix*. Individuals deemed hybrids from their intermediate morphology and the individual taxa are highly distinctive. The only non-native species, *E. repens*, hybridizes with *E. trachycaulus*. | | 3 |
| *Epilobium* ×*wisconsinense* | *E. ciliatum* | *E. coloratum* | This genus is known for hybridization and well-studied by P. H. Raven and probably turned him against the biological species concept due its hybridization (Ehrlich and Raven 1969). Only two of the five native species and none of the three introduced species are said to hybridize. | | 1 |
| *Equisetum* | Multiple | Multiple | There are 10 native species of horsetail and three are said to hybridize with all combinations named: *E. variegatum* × *hyemale* (*E.* ×*mackayi*), *E. variegatum* × *laevigatum* (*E.* ×*nelsonii*), and *E. hyemale* × *variegatum* (*E.* ×*ferrissii*). | | 3 |

| | | | | | |
|---|---|---|---|---|---|
| *Eupatorium ×polyneuron* | *E. perfoliatum* | *E. serotinum* | There are two native and two introduced bonesets. Late boneset, *E. serotinum*, is adventive from farther south and hybridizes with native *E. serotinum*. This is a well-known hybrid and the description of shriveled achenes indicates that the hybrid is (mostly) sterile. | | 1 |
| *Euthamia* | *E. graminifolia* | *E. caroliniana* | There are two native *Euthamia* species, and intermediate individuals are said to be hybrids. This is a difficult taxonomic group closely related to *Solidago*. | | 1 |
| *Fallopia ×bohemica* | *F. japonica* | *F. sachalinensis* | This is the tetraploid or hexaploid (depending on *F. japonica* cytotype) hybrid of two introduced species. The hybrid is apparently fertile. | | 1 |
| *Fraxinus* | *F. pennsylvanica* | *F. americana* | Hybrids between *F. pennsylvanica* and *F. americana* are mentioned in the flora but said to be uncommon. Hybrids between these species are thought to have resulted in two hybrid species: i) hexaploid *F. biltmoreana* (from a tetraploid *F. americana* × diploid *F. pennsylvanica* - not occurring in Michigan) and ii) tetraploid *F. smallii* (apparently in Michigan but not included in the flora). The numerous ploidy levels and cross compatibility among the five native ashes make this an interesting group for studying hybridization. | (Nesom 2010) | 1 |
| *Galeopsis* | *G. tetrahit* | *G. bifida* | There are two introduced species of Hemp-Nettle in the flora: *G. tetrahit* and *G. ladanum*. *G. tetrahit* is said to hybridize with *G. bifida* which is circumscribed under *G. tetrahit* in the flora. Included in this table (but not counted as a unique hybrid) because the line between species, subspecies etc. is difficult to draw, and it is not recognized by Reznicek and Voss (2012). Molecular data may help in this case. | | 0 |
| *Gentiana* | *G. andrewsii* | *Multiple* | From six native species, there are three named hybrids; all involve *Gentiana andrewsii*: *G. andrewsii* × *alba* (*G. ×pallidocyanea*), *G. andrewsii* × *puberulenta* (*G. ×billingtonii*), and *G. andrewsii* × *rubricaulis* (*G. ×grandilacustris*). No information is given about the three species that hybridize with *G. andrewsii* in regards to their hybridizing with one another. There are two additional native species which are not said to hybridize at all. | | 3 |
| *Gentianopsis* | *G. virgata* | *G. crinita* | These two species “are not as clearly separable as one might wish (and are considered to hybridize in some regions) … [and] … floral differences often cited between this species and *G. crinita* do not hold up.” (Reznicek and Voss 2012). This is either good evidence of hybridization or of over-differentiation in the genus. | | 1 |
| *Geum ×catlingii* | *G. urbanum* | *G. canadense* | There are nine native and one introduced *Geum* species in the flora. Eurasian *G. urbanum* is said to hybridize with *G. canadense* and the hybrids are apparently intermediate. | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| *Goodyera tesselata* | *G. repens* | *G. oblongifolia* | There are four native species of rattlesnake plant. The two parental species are diploid (2n=30) and *G. tesselata* is tetraploid, morphologically intermediate, and thought to be of hybrid origin. Triploids were found in northern Michigan by (Kallunki 1976) indicating backcrossing. The fourth native *Goodyera* species is uncommon and not found with the others. *Goodyera oblongifolia* is mostly in western North America with the Michigan populations disjunct whereas *G. repens* is circumboreal. | (Kallunki 1976) | 3 |
| *Gymnocarpium* | Multiple | Multiple | There are three native oak ferns, but only *Gymnocarpum* dryopteris (tetraploid) and *G. jessoense* (diploid) hybridize giving *G. ×intermedium*, a sterile triploid. The third native species, *G. robertianum*, is not said to hybridize. The situation is, however, more complicated because *G. dryopteris* is itself an allotetraploid from *G. disjuncta* × *appalachianum* (both diploid, 2n=80). Note that *G. disjuncta* may well be native while *G. appalachianum* is not native. Neither species is included in the flora. The sterile triploid of backcross of *G. dryopteris* to *G. disjuncta* is *G. ×brittonianum* but is not included because *G. disjuncta* is not in the flora. | (Pryer and Haufler 1993) | 1 |
| *Helianthus* | Multiple | Multiple | From 11 native species, five hybrid combinations and four named hybrids are given: *H. pauciflorus* × *tuberosus* (*H. ×laetiflorus*), *H. mollis* × *giganteus, H. grosseserratus* × *giganteus* (*H. ×luxurians*), *H. giganteus* × *divaricatus* (*H. ×ambiguus*), *H. divaricatus* × *grosseserratus* (*H. ×divariserratus*). Two of the introduced species, *H. annuus* and *H. petiolaris*, are well-known for hybridizing and have spun off three distinct homoploid hybrid species in western North America. However, there is no indication of there being hybrids involving these or the other two introduced species in Michigan. | (Rieseberg et al. 2003) | 5 |
| *Heuchera* | *H. americana* | *H. richardsonii* | The two native alum roots are difficult to divide from morphology. While most *H. americana* have symmetrical flowers, the Michigan collections have somewhat bilateral flowers more characteristic of *H. richardsonii* – the authors interpret this as evidence of hybridization. | | 1 |
| *Hieracium* | Multiple | Multiple | There are 15 *Hieraceum* species in the flora with seven native. From the native hawkweeds, four are said to hybridize in three combinations: *H. scabrum* × *venosum*, *H. kalmii* × *scabrum,* and *H. kalmii* × *umbellatum*. There are also three combinations involving only introduced species: *H. aurantiacum* × *piloselloides* hybrids occur occasionally where both parents are found and can be recognized by the intermediate color of the ligules; *H. aurantiacum* × *caespitosum*; and *H. flagellare* × *caespitosum*. Native *H. kalmii* also hybridizes with *H. lachenalii* producing *H. ×grohii*. Reznicek and Voss (2012) write that "Morphological characters are not very helpful in confirming hybrids." | | 7 |

| | | | | |
|---|---|---|---|---|
| *Huperzia* | *H. selago* | Multiple | All three native Huperiza species hybridize: *H. appressa* × *selago* (*H. ×josephbeitelii*, mostly sterile) and *H. selago* × *lucidula* (*H. ×buttersii*). | 2 |
| *Hypericum* | Multiple | Multiple | Twelve native and one introduced species of St. John's wort. Three closely related species hybridize in all combinations: *H. majus, H. canadense*, and *H. mutilum*. | 3 |
| *Iris ×robusta* | *I. versicolor* | *I. virginica* | Hybrids between *I. versicolor* and *I. virginiana* are known from Mackinaw county. The third native species, *I. lacustris* (Dward Lake Iris), is not known to hybridize (it is also the state flower!). This genus is well-known for hybridization and hybrid speciation. | 1 |
| *Isoetes* | *I. echinospora* | *I. lacustris* | There are three native *Isoetes* species. *Isoetes ×hickeyi* is the named hybrid of *I. echinospora* × *lacustris*. *Isoetes lacutris* is a decaploid (2n=100) and *I. echinospora* is diploid (2n=22) and the hybrid is therefore likely sterile in backcrosses, but this is not mentioned. There are likely more hybrid species currently described under the third native species, *I. engelmannii*, which currently includes a number of different ploidy levels, but this species is not identified as hybridizing here. | 1 |
| *Juncus* | *J. alpinoarticulatus* | *J. articulatus* | Only two of 26 native *Juncus* species are said to hybridize producing more or less sterile intermediates. | 1 |
| *Lactuca* | *L. biennis* | *L. canadensis* | These two widespread native lettuce species are known to hybridize, but their hybrids are not common. | 1 |
| *Leersia* | *L. oryzoides* | *L. virginica* | Morphologically intermediate individuals suggest hybridization between the only two native *Leersia* species. | 1 |
| *Lespedeza* | Multiple | Multiple | One named hybrid and six more hybrid crosses involving five native species are given for *Lespedeza*: *L. virginica* × *frutescens, L. frutescens* × *violaceae, L. violaceae* × *hirta* (*L. ×nuttallii*), *L. capitata* × *violaceae, L. capitata* × *hirta*, and *L. capitata* × *virginica*. The final native species, *L. procumbens*, is not said to hybridize (extirpated), nor is hybridization said to occur for the three introduced *Lespedeza* species. | 6 |
| *Liatris* | *Multiple* | *Multiple* | There are six native (two of which are extirpated) and one introduced species. The named hybrid between native species *Liatris cylindracea* and *L. aspera* (*L. ×gladewitzii*) is said to be uncommon. A second hybrid combination between native *L. spicata* and introduced *L. pycnostachya* is given with the evidence of individuals with intermediate characters. | 2 |

| | | | | | |
|---|---|---|---|---|---|
| *Lobelia ×speciosa* | *L. cardinalis* | *L. siphilitica* | This hybrid has only been found once in Michigan. The four other native taxa do not hybridize. The hybrid is intermediate in phenotype. | | 1 |
| *Lonicera* | Multiple | Multiple | There are six native honeysuckles with some collections of *L. dioica* thought to actually be hybrids between *L. hirsuta* and a large-flowered species. From nine introduced species, there are six unique hybrids involving four species: *L. ×bella* is the hybrid of *L. morrowii* × *tatarica* and is said to backcross with both parents, *L. morrowii* × *ruprechtiana* (*L. ×muscaviensis*; note that *L. ruprechtiana* is not included in the flora but the hybrid has been collected in Michigan), *L. xylosteum* × *tatarica* (*L. ×xylosteoides*), *L. xylosteum* × *morrowii* (*L. ×minutiflora*). | | 5 |
| *Lycopodiella* | Multiple | Multiple | No specific hybrid combinations are given for the three native *Lycopodiella* clubmosses. However, *L. subappressa* × *margueritae* hybrids are fertile (both tetraploids). Further *L. margueritae* is known to hybridize with diploid *L. inundata*; these hybrids are sterile. | (Flora of North America Editorial Committee 1993+-c) | 2 |
| *Lycopus* | Multiple | Multiple | There are four native and two introduced species in this genus of mints. *Lycopus ×sherardii* is the named hybrid of *L. uniflorus* and *L. virginicus*. *Lycopus uniflorus* is also said to hybridize with *L. americanus*. The fourth native *Lycopus* species, *L. rubellus*, is not said to hybridize. Introduced *L. europaeus* hybridizes with native *L. americanus*. Individuals intermediate between *L. uniflorus* and adventive *L. asper* have been collected, but the authors do not say that there has been hybridization. | | 3 |
| *Lysimachia* | *L. terrestris* | Multiple | There are seven native loosestrifes, with two named hybrids from crosses involving three species: *L. terrestris* × *quadrifolia* (*L. ×producta*) and *L. terrestris* × *thyrsiflora* (*L. ×commixta*). There are no hybrids involving the four introduced *Lysimachia* species. | | 2 |
| *Monarda ×media* | *M. didyma* | *M. fistulosa* | There are four native *Mondarda* species and two are thought to hybridize giving *Monarda ×media*, but there is uncertainty. See Reznicek and Voss (2012) for more information. | | 1 |
| *Morus* | *M. rubra* | *M. alba* | The only native mulberry, *M. rubra*, hybridizes with the only introduced species, *M. alba*. Hybrids are vigorous. | | 1 |
| *Myriophyllum* | *M. spicatum* | *M. sibiricum* | There are two introduced and six native *Myriophyllum* species. *M. spicatum* is introduced and *M. sibricum* is native. Hybrids have an intermediate leaf morphology. | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| *Neottia ×veltmanii* | *N. auriculata* | *N. convallarioides* | The hybrid *N. ×veltmanii* is also known from New England. There is no indication that the third *Neottia* species in the flora, native *N. cordata*, hybridizes with the other native taxa. | (Haines et al. 2011) pg. 206 | 1 |
| *Nuphar ×rubrodisca* | *N. variegata* | *N. microphylla* | Hybrids are intermediate in leaf and flower morphology and mostly sterile. The third native *Nuphar* species, *N. advena*, is intermediate between *N. variegata* and *N. microphylla* and was considered by some authors as their hybrid, but this is disputed. Hybrids between *N. advena* and *N. variegata* are suspected (Padgett et al. 1998). It is worth noting that early botanists considered all North American *Nuphar* one species. | | 1 |
| *Packera insulae-regalis* | *P. paupercula* | *P. indecora* | This is an example of hybrid speciation in ragwort which is local to Isle Royal, Michigan. *Packera indecora* is an octaploid and *P. paupercula* is a tetraploid. Their hybrid, *P. insulae-regalis* is hexaploid and was not formerly distinguished from *P. paupercula*. It appears to be a new species. | (Kowal et al. 2011) | 1 |
| *Pascopyrum* | *P. smithii* | *Elymus repens* | Hybridization between the only *Pascopyrum* in the flora and *E. repens* is inferred from leaf morphology. Both of these species are introduced. | | 1 |
| *Petasites ×vitifolius* | *P. sagittatus* | *P. frigidus* | The authors elect to keep these species separate, but some authors consider them subspecies (for example, in the Flora of North America). These are the only two *Petasites* species in the flora. | | 1 |
| *Platanthera* | Multiple | Multiple | Eight of 15 species in *Platanthera* are noted to hybridize with three named hybrids and one hybrid species: *P. huronensis*. *Platanthera huronensis* (tetraploid) is morphologically intermediate between diploids *P. aquilonis* and *P. dilatata*. The hybrid relationship has been supported by molecular data (Wallace 2003). The remaining taxa hybridize in the following five combinations: *P. leucophaea* × *lacera (P. ×hollandia), P. leucophaea* × *psycodes (P. ×reznicekii), P. lacera* × *psycodes (P. ×andrewsii), P. blephariglottis* × *ciliaris, P. blephariglottis* × *clavellata*. | (Wallace 2003) | 6 |
| *Populus* | Multiple | Multiple | Two hybrid combinations are given for native species: *P. tremuloides* × *grandidentata* (*P. ×smithii*) and *P. balsamifera* × *P. deltoides* (*P. ×jackii*). The final native species, *P. heterophylla*, is uncommon and not said to hybridize. The introduced *P. alba* hybridizes with both *P. tremuloides* (*P. ×heimburgeri*) and *P. grandidentata* (*P. ×rouleauiana*). Two more hybrids involving Michigan species are known in Europe: *P. deltoides* × *trichocarpa* (*P. ×generosa*), *P. deltoides* × *P. nigra* (*P. ×canadensis*) (Seybold 2009). These are not included in any analysis but mentioned here. | (Seybold 2009) | 4 |

| | | | | | |
|---|---|---|---|---|---|
| *Potamogeton* | Multiple | Multiple | A large and difficult genus of aquatic plants. Of 25 native species, only *P. ×haynesii* (*P. strictifolius* × *zosteriformis*) is indicated as a hybrid. *P. foliosus* × *pusillus, P. berchtoldii* × *vaseyi* hybrids are known from the literature. | (Les et al. 2009) | 3 |
| *Proserpinaca ×intermedia* | *P. pectinata* | *P. palustris* | The hybrid *Proserpinaca ×intermedia* is intermediate between its parents. It is recognized as a species by some authors and may be reproductively independent. These are the only two native *Proserpinaca* species. | (Catling 1998) | 1 |
| *Pycnanthemum* | *P. pilosum* | Multiple | From five native mountain mints, three are known to hybridize in two combinations. First, *P. pilosum* × *verticillatum* (*P. pilosum* is sometimes considered synonymous with *P. verticillatum*). Second, *P. pilosum* × *muticum*. The one introduced species is not said to hybridize. | | 2 |
| *Pyrus* | *P. calleryana* | *P. communis* | Hybrids between these species of pear are suspected from a combination of leaf and fruit characters. All three of the flora's *Pyrus* species are non-native. | | 1 |
| *Quercus* | Multiple | Multiple | Oaks are the main example of a syngameon from (Grant 1981) – and are a true taxonomic nightmare. The subgenus Quercus has two sections: *Quercus* (white oaks) and *Lobatae* (red oaks). Interspecific hybridization, while common, remains within the section. Of 11 native species, all but *Q. shumardii* and *Q. palustris* are said to hybridize. The following 15 combinations are given in the flora (with 11 named hybrids): ***section Quercus:*** *Q. alba* × *macrocarpa (Q. ×bebbiana). Q. alba* × *bicolor (Q. ×jackiana). Q. alba* × *prinoides (Q. ×faxonii). Q. alba* × *muehlenbergii. Q. bicolor* × *macrocarpa (Q. ×schuettei). Q. bicolor* × *prinoides (Q. ×wagneri). Q. macrocarpa* × *muehlenbergii (Q. ×deamii). Q. macrocarpa* × *prinoides (Q. ×beckyae). Q. muehlenbergii* × *prinoides*. ***section Lobatae:*** *Q. ellipsoidalis* × *rubra. Q. ellipsoidalis* × *imbricaria. Q. ellipsoidalis* × *velutina (Q. ×palaeolithicola). Q. imbricaria* × *rubra (Q. ×runcinata). Q. imbricaria* × *veluntia (Q. ×leana). Q. rubra* × *velutina (Q. hawkinsiae*). The one introduced species, *Q. montana*, is not said to hybridize. | | 15 |
| *Rhamnus* | *R. utilis* | *R. cathartica* | There are no native but three introduced buckthorn species. Hybrids between these two have intermediate leaf morphology. | | 1 |
| *Rhus ×pulvinata* | *R. typhina* | *R. glabra* | Rhus glabra is well-known for hybridizing with R. typhina. The authors write that hybrids are more common than R. typhina itself. They also differentiate four groups of hybrids. The other two native Sumac species are not known to hybridize. | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| *Ribes* | Multiple | Multiple | No specific hybrid combinations are given for the eight native *Ribes*, but twice the reference to hybridization is made for the gooseberries of this genus. The three native gooseberries (but not the introduced *R. missouriense*) are given as hybridizing: *R. oxyacanthoides, R. hirtellum*, and *R. cynosbati*. | | 3 |
| *Rosa* | Multiple | Multiple | There are six native and 10 introduced roses in the flora. Five native species are known to hybridize in three combinations. *R. blanda* (diploid) × *palustris* (*R. ×palustriformis*), *R. carolina* × *R. arkansana*, and *R. blanda* × *R. acicularis* (hexaploid, resulting in partly sterile tetraploids). If partly sterile tetraploid hybrids backcrossed, then sterile triploids or pentaploids would be expected and this may therefore be a potential example of hybrid speciation given that they hybrids are not fully sterile. No hybrid combinations are given for the 10 introduced species. | (Lewis 2008) | 3 |
| *Rubus* | Multiple | Multiple | This genus is one of the 'taxonomic nightmares' that have led botanists to question the species concept. The difficulty is due to agamospermy, cross compatibility, and numerous ploidy levels. There are 12 native species including blackberries, dewberries, raspberries, and the thimbleberry. Five hybrid combinations involving nine species are given: *R. strigosus* × *occidentalis* (*R. ×neglectus*) – raspberry × raspberry. *R. acaulis* × *pubescens* (*R. ×paracaulis*) – raspberry × raspberry. *R. canadensis* × *allegheniensis* – dewberry × blackberry. *R. pensilvanicus* × *allegheniensis* – dewberry × blackberry. *R. parviflorus* × *odoratus* (*R. ×fraseri*) – thimbleberry × raspberry. No hybrid combinations are given for the five introduced species. | | 5 |
| *Rumex* | *R. crispus* | *R. obtusifolius* | There are no hybrids involving the six native docks, but from six introduced species, *R. crispus* and *R. obtusifolius*, are said to hybridize. | | 1 |
| *Salix* | Multiple | Multiple | There are 18 native species of cross-compatible, perennial, and wind-pollinated willows which all appear to be diploid (n=19). Nine native species are said to hybridize with the following five combinations: *S. discolor* × *humilis. S. myricoides* × *cordata. S. nigra* × *amygdaloides (S. ×glatfelteri). S. sericea* × *petiolaris, S. planifolia* × *discolor*. Two more native species, *S. bebbiana* and *S. candida*, are said to hybridize with many species but no specifics are given. From eight introduced species, two combinations involving only non-natives are given: *S. alba* × *euxina* (*S. ×fragilis*) and *S. pentandra* × *euxina* (*S. ×meyeriana*). | (Reznicek and Voss 2012) | 7 |
| *Schoenoplectus* | *Multiple* | *Multiple* | Two hybrid combinations involving four of seven native species are described: *S. tabernaemontani* × *acutus* and *S. americanus* × *pungens* (*S. ×contortus*). | | 2 |

| | | | | | |
|---|---|---|---|---|---|
| *Scirpus* | *Multiple* | *Multiple* | Five of nine native species in this genus of bulrush hybridize. First, *S. hattorianus* × *atrovirens* hybridize with their hybrids having abortive seeds. Second*, S. cyperinus, S. atrocinctus,* and *S. pedicellatus* are said to hybridize in all combinations. | | 4 |
| *Scutellaria* ×*churchilliana* | *S. lateriflora* | *S. galericulata* | There are seven native skullcaps with one hybrid, *S.* ×*churchilliana*, said to have intermediate leaves between the parents. | | 1 |
| Silene ×hampeana | *S. dioica* | *S. latifolia* | There are four native and 10 introduced *Silene* species. Hybrids (*S.* ×*hampeana*) from introduced *S. latifolia* and *S. dioica* have been found in Michigan. These two species are European natives with pink and white flowers, respectively. | | 1 |
| *Silphium* | *S. terebinthinaceum* | *S. laciniatum* | Two of four native species hybridize. The hybrid has intermediate leaves and involucral bracts. | | 1 |
| *Solidago* | Multiple | Multiple | There are 22 native Solidago species, including two hybrid species. First, *S. houghtonii* is a hybrid derived from S*. ptarmicoides* × *ohioensis* × *riddellii* (all diploids). Second, the cross of *S. houghtonii* (hexaploid) × *ptarmicoides* (diploid) resulted in the octaploid *S. vossii* (2n=8x=72). Finally, seven additional taxa hybride in five more combinations: *S. patula* × *uliginosa, S. uliginosa* × *gigantean, S. rugosa* × *uliginosa, S. ptarmicoides* × *ohioensis* (*S.* ×*krotkovii*), *S. gigantea* × *flexicaulis*. The only introduced species, *S. sempervirrens*, does not hybridize. | (Laureto and Pringle 2010) | 7 |
| *Stachys* | *S. hispida* | *S. tenuifolia* | Two of six native hedge-nettles are said to hybridize. Note that *Stachys hispida* was once included within *S. tenuifolia* (synonymous with *S. tenuifolia* var. *hispida*). Uncommon hybrids are morphologically intermediate in respect to leaf pubescence and petiole length between the parental taxa. The two introduced *Stachys* do not hybridize. | | 1 |
| *Stuckenia* ×*suecica* | *S. filiformis* | *S. pectinata* | These pondweeds are closely related to *Potamogeton* and *Stuckenia filliformis* × *pectinata* give the named hybrid *S.* ×*suecica.* | | 1 |
| *Symphyotrichum* | Multiple | Multiple | There are 20 native asters in the genus *Symphyotrichum*, 11 of which hybridize in 10 combinations: *S. novae-angliae* × *ericoides (S.* ×*amethystinum), S. lanceolatum* × *pilosum, S. lanceolatum* × *lateriflorum, S. lanceolatum* × *puniceum, S. lanceolatum* × *firmum, S. ciliolatum* × *laeve, S. ciliolatum* × *lanceolatum, S. boreale* × *lanceolatum, S. boreale* × *puniceum*. The 10$^{th}$ combination is *S. racemosum* × *dumosum; S. racemosum* is not in the flora but may be native (Brouillet et al. 2006) (therefore *S. dumosum* is included as hybridizing with a native species – *S. racemosum* is not included). | (Brouillet et al. 2006) | 10 |

| | | | | |
|---|---|---|---|---|
| *Thalictrum* | *T. dasycarpum* | *Multiple* | Of six native meadow rues, two are said to cross with *T. dasycarpum: T. dasycarpum* × *venulosum* and *T. dasycarpum* × *revoltum*. | 2 |
| *Toxicodendron* | *T. rydbergii* | *T. radicans* | These two species of poison ivy are suspected of hybridization. The third native *Toxicodendron* species, poison sumac, is quite distinct from the other two and does not hybridize with them. | 1 |
| *Tradescantia* | *T. virginiana* | *T. ohiensis* | There are diploid and tetraploid *Tradescantia ohiensis* in Michigan, and they apparently have different habitats, so this is a candidate example of autopolyploid speciation. It is not said if hybrids are diploid or triploid. The third native species, *T. bracteata*, is no longer found in Michigan and there are no introduced species. | 1 |
| *Triadenum* | *T. fraseri* | *T. virginicum* | Intermediate individuals are said to be found between the only two *Triadenum* species in the flora, and this is interpreted as hybridization. | 1 |
| *Trillium* | *T. erectum* | Multiple | There are eight native trilliums. *Trillium erectum* is said to hybridize with two species: *T. flexipes* (drooping trillium) and *T. cernuum* (nodding trillium). Hybrids are intermediate in corolla morphology. | 2 |
| *Typha* ×*glauca* | *T. latifolia* | *T. angustifolia* | There are two cattails in the flora: native *T. latifolia* and introduced *T. angustifolia*. Their hybrid, *T.* ×*glauca*, is said to be dominant in some marshes with a transgressive phenotype (greater height and longer inflorescences) relative to the parents. | 1 |
| *Vaccinium* | *V. angustifolium* | *V. corymbosum* | From Reznicek and Voss (2012), “Hybridization with lowbush taxa (producing half-high shrubs) is suspected.” There are 11 native species in the genus including bilberries, blueberries, and cranberries. | 1 |
| *Verbascum* | *V. thapsus* | Multiple | All five *Verbascum* species in the flora are introduced. Three species hybridize with *V. thapsus: V. thapsus* × *densiflorum (V.* ×*humnickii), V. thapsus* × *lychnitis,* and *V. thapsus* × *phlomoides* (sterile *V.* ×*kerneri*). All hybrids apparently intermediate in phenotype. | 3 |
| *Verbena* | Multiple | Multiple | There are three native and three introduced *Verbena* species and three named hybrid combinations involving four species: native *V. urticifolia* and *V. hastata* hybridize to produce *V.* ×*engelmannii* which is intermediate in corolla morphology, style length, and inflorescence density between the parental taxa; *V. hasta* × *stricta* (*V.* ×*rydbergii*); and *V. stricta* × *bracteata* (*V.* ×*deamii* which has a combination of parental characters). This family has a high hybrid propensity according to (Whitney et al. 2010). | 3 |

| | | | | | |
|---|---|---|---|---|---|
| *Viola* | Multiple | Multiple | There are a minimum of 22 native violets in Michigan and Ballard writes that these are a “pack of taxonomically incorrigible wolves.” Which is probably due to their habit of being cleistogamous and selfing. But their cross compatibility and morphological diversity also add to the difficulty. Seventeen species are said to hybridize in 18 combinations, not including the *V. canadensis* × *pubescens* cross which is likely. The hybrid crosses are divided into four groups: **Uncommon and mostly sterile hybrids**: V. adunca × labradorica (sterile), *V. affinis* × *cucullata*, *V. affinis* × *sororia, V. blanda* × *lanceolata* (sterile), *V. blanda* × *renifolia, V. cucullata* × *nephrophylla, V. cucullata* × *sororia, V. labradorica* × *rostrata* (*V.* ×*malteana*), *V. labradorica* × *striata* (*V.* ×*eclipes*), *V. palmata* × *pedatifida*. **Widespread hybrids and hybrid species**: *V. lanceolata* × *macloskeyi* (*V.* ×*primulifolia* – may be synonymous with *V. primulifolia*, see main text), *V. pedatifida* × *sororia* (*V.* ×*subsinuata*), *V. sororia* × *sagittata* (*V.* ×*palmata*). **From (Ballard 1994) but not in the flora***: V. rostrata* × *conspera* (sterile), *V. rostrata* × *striata* (*V.* ×*braunie*), *V macloskeyi* × *blanda* (sterile), *V. macloskeyi* × *renifolia*, *V. nephrophylla* × *sororia*. These are included as hybrids here. **Not confirmed but likely in Michigan**: V. adunca × rostrata, V. adunca × striata, V. canadensis × pubescens. These are not included as hybrid crosses here but might show up in Michigan. | (Ballard 1994) | 18 |
| *Woodsia* ×*abbeae* | *W. ilvensis* | *W. oregana* | *Woodsia* ×*abbeae* is a morphologically intermediate sterile triploid found in the western upper peninsula only. There are four native *Woodsia*, which are leptosporangiate fern species, but only one hybrid combination resulting in (probably sterile) *W.* ×*abbeae*; all Michigan *W. oregano* are tetraploid subsp. *cathcartiana* (2n = 152) and *W. ilvensis* is diploid (2n = 82). | (Flora of North America Editorial Committee 1993+-c) | 1 |

**Supplemental Table 3**. Available online only.

**Supplemental Table 4**. Examples of non-hybridizing Michigan species (N=19) known to hybridize elsewhere.

| *Michigan Taxa* | *Information* | *Evidence*[a] | *Citation* |
|---|---|---|---|
| *Aesculus glabra* | Hybridizes with *A. pavia* and *A. flava*. | M | (Depamphilis and Wyatt 1989) |
| *Asclepias tuberosa* | Said to hybridize with populations of *Ascelpias tuberosa interior*, currently a subspecies of *Asclepias tuberosa*. | M | (Wyatt and Antonovics 1981) |
| *Phlox divariata* and *P. pilosa* | These two hybridize and *P. pilosa* is also said to hybridize with *P. deameii*. | M | (Levin 1967) (Levin and Schaal 1972) |
| *Phlox maculate* | Said to hybridize with *P. glaberrima*. | M | (Hadley and Levin 1969). |
| *Phlox bifida*[c] | *Phlox bifida* (now extirpated in Michigan) is said to hybridize with *P. amoena*. | M | (Anderson and Gage 1952) |
| *Solidago rugose* | Said to hybridize with *S. semprivens*, an East Coast halophyte which now occurs in the Great Lakes region, including Michigan. | M | (Goodwin 1937) |
| *Oenothera biennis* and *O. parviflora* | These two species are said to hybridize in Britain. This cross resulted in *O. muricata.* All appear to be 2n=2x=14. | M | (Stace 1975 p. 16) |
| *Cercis Canadensis* | *Cercis canadensis* is said to form hybrid swarms with *C. reniformis*. | M* | (Anderson 1953) |
| *Phytolacca americana* | Thought to hybridize with *P. octandra*. | M* | (Anderson 1953) |
| *Carex rostrate* (or *C. utriculata*) | Carex rostrata hybridizes with *C. rotunda*. These species occupy different habitats with C. rostrata in lakes and *C. rotunda* in peat bogs. In Alaska, they hybridize and have formed fertile populations named *C. paludivagans* segregating out of hybrid swarm that occurs between their habitats. The '*C. rostrata'* parent appears to have been *C. utriculata (but this is also a Michigan species).* | M | (Grant 1981 p. 252; (Drury 1956; Ford et al. 1993) |
| Carex gynocrates | The hybrid between Carex gynocrates and C. maritime is named C. ×langeana Fernald. | M | (Flora of North America Editorial Committee 1993+-a) |
| *Pinus banksiana* | Hybridizes with *Pinus cordata*. Discussed in the main text. | M,P | (Wagner et al. 1987) |

| | | | |
|---|---|---|---|
| *Picea mariana* and *P. glauca* | Hybrid zone in Minnesota. Discussed in the main text. | M | (Little and Pauley 1958) |
| *Juniperus virginiana* and *J. horizontalis* | Hybrid zone in Wisconsin. Discussed in the main text. In addition, *J. virginiana* is said to have introgressions from *J. scopulorum* from Western North America (Grant 1981 p. 217). | M,E | (Flake et al. 1978; Palmaotal et al. 1983) |
| *Abies balsamea* | Hybridizes with *A. fraseri* and *A. lasiocarpa*. Discussed in the main text. | P | (Cinget et al. 2015) |

[a]M = Morphology, P = Chloroplast DNA Markers, E = Ecology/Habitat/Biogeography inferences.
*Qualitative not quantitative data.

### 2.13.1. References for Supplemental Table 2 and Supplemental Table 4.

Anderson E. 1953. Introgressive hybridization. *Biol Rev Camb Philos Soc* **28**: 280-307.

Anderson E, Gage A. 1952. Introgressive hybridization in Phlox bifida. *Am J Bot* **39**: 399-404.

Ballard H. 1994. Violets of Michigan. *Mich Bot* **33**: 131-199.

Barnes BV, Dancik BP. 1985. Characteristics and origin of a new birch species, Betula murrayana, from southeastern Michigan. *Canadian Journal of Botany* **63**: 223-226.

Brouillet L, Semple J, Allen G, Chambers K, Sundberg S. 2006. Symphyotrichum. *Flora of North America* **20**: 465-539.

Catling PM. 1998. A synopsis of the genus Proserpinaca in the southeastern United States. *Castanea*: 408-414.

Cinget B, de Lafontaine G, Gerardi S, Bousquet J. 2015. Integrating phylogeography and paleoecology to investigate the origin and dynamics of hybrid zones: insights from two widespread North American firs. *Mol Ecol* **24**: 2856-2870.

Crowder AA, Pearson MC, Grubb PJ, Langlois PH. 1990. Drosera L. *J Ecol* **78**: 233-267.

Depamphilis CW, Wyatt R. 1989. Hybridization and introgression in buckeyes (Aesculus, Hippocastanaceae) - a review of the evidence and a hypothesis to explain long-distance gene flow. *Syst Bot* **14**: 593-611.

Drury WH. 1956. The ecology of the natural origin of a species of Carex by hybridization. *Rhodora* **58**: 51-72.

Ehrlich PR, Raven PH. 1969. Differentiation of populations. *Science* **165**: 1228-1232.

Flake R, Urbatsch L, Turner B. 1978. Chemical documentation of allopatric introgression in Juniperus. *Syst Bot*: 129-144.

Flora of North America Editorial Committee. 1993+-a. *Flora of North America, Magnoliophyta: Commelinidae (in Part): Cyperaceae.*

Flora of North America Editorial Committee. 1993+-b. *Flora of North America, Magnoliophyta: Salicaceae to Brassicaceae*.

Flora of North America Editorial Committee. 1993+-c. *Flora of North America, Pteridophytes and gymnosperms*.

Ford BA, Ball PW, Ritland K. 1993. Genetic and macromorphologic evidence bearing on the evolution of members of Carex section Vesicariae (Cyperaceae) and their natural hybrids. *Canadian journal of botany* **71**: 486-500.

Freckmann RW, Lelong MG. Dichanthelium (Hitchc. & Chase) Gould. Vol 2017. Utah State Herbarium.

Goodwin RH. 1937. The cyto-genetics of two species of Solidago and its bearing on their polymorphy in nature. *Am J Bot*: 425-432.

Grant V. 1981. *Plant speciation*. New York: Columbia University Press.

Hadley EB, Levin DA. 1969. Physiological evidence of hybridization and reticulate evolution in Phlox maculata. *Am J Bot* **56**: 561-&.

Haines A, Farnsworth E, Morrison G. 2011. *New England Wild Flower Society's Flora Novae Angliae: a manual for the identification of native and naturalized higher vascular plants of New England*. Yale University Press.

Kallunki JA. 1976. Population studies in Goodyera (Orchidaceae) with emphasis on the hybrid origin of G. tesselata. *Brittonia* **28**: 53-75.
Kowal RR, Judziewicz EJ, Edwards J. 2011. Packera insulae-regalis (Asteraceae, Senecioneae), a new species endemic to Isle Royale, Michigan, USA. *Brittonia* **63**: 343-354.
Laureto PJ, Pringle JS. 2010. Solidago vossii (Asteraceae), A new species of Goldenrod from Northern Michigan. *Mich Bot*: 105-117.
Les DH, Murray NM, Tippery NP. 2009. Systematics of Two Imperiled Pondweeds (Potamogeton vaseyi, P. gemmiparus) and Taxonomic Ramifications for Subsection Pusilli (Potamogetonaceae). *Syst Bot* **34**: 643-651.
Levin DA. 1967. Variation in Phlox divaricata. *Am J Bot* **49**: 92-108.
Levin DA, Schaal BA. 1972. Seed protein polymorphism in Phlox pilosa (Polemoniaceae). *Brittonia* **24**: 46-&.
Lewis WH. 2008. Rosa carolina (Rosaceae) subspecies and hybrids in eastern and Midwestern United States, Canada, and Mexico. *Novon* **18**: 192-198.
Little EL, Pauley SS. 1958. A natural hybrid between black and white spruce in Minnesota. *The American Midland Naturalist* **60**: 202-211.
Marcks BG. 1974. Preliminary reports on the flora of Wisconsin: no. 66. Cyperaceae: 2. - Sedge family: 2. The genus Cyperus - the umbrella sedges. *Trans Wisconsin Acad Sci* **62**: 261-284.
Nesom G. 2010. Fraxinus biltmoreana and Fraxinus smallii (Oleaceae), forest trees of the eastern United States. *Phytoneuron* **51**: 1-30.
Nielsen EL. 1939. A taxonomic study of the genus Amelanchier in Minnesota. *The American Midland Naturalist* **22**: 160-206.
Padgett DJ, Les DH, Crow GE. 1998. Evidence for the hybrid origin of Nuphar x Rubrodisca (Nymphaeaceae). *Am J Bot* **85**: 1468-1476.
Palmaotal M, Moore WS, Adams RP, Joswiak GR. 1983. Morphological, chemical, and biogeographical analyses of a hybrid zone involving Juniperus virginiana and J. horizontalis in Wisconsin. *Canadian Journal of Botany* **61**: 2733-2746.
Pryer KM, Haufler CH. 1993. Isozymic and chromosomal evidence for the allotetraploid origin of Gymnocarpium dryopteris (Dryopteridaceae). *Syst Bot* **18**: 150-172.
Reznicek A, Judziewicz E. 1996. A new hybrid species, x Calammophila don-hensonii (Ammophila breviligulata x Calamagrostis canadensis, Poaceae), from Grand Island, Michigan. *Mich Bot* **35**: 35-40.
Reznicek AA, Voss EG. 2012. *Field manual of Michigan flora*. University of Michigan Press.
Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211-1216.
Runk K, Zobel M, Zobel K. 2012. Biological Flora of the British Isles: Dryopteris carthusiana, D. dilatata and D. expansa. *J Ecol* **100**: 1039-1063.
Seybold S. 2009. Schmeil-Fitschen—Die Flora Deutschlands und der angrenzenden Länder. *Quelle and Meyer, Heidelberg, Wiesbaden*.
Stace CA. 1975. *Hybridization and the flora of the British Isles*. Academic Press, London.
Sweeney PW, Price RA. 2000. Polyphyly of the genus Dentaria (Brassicaceae): Evidence from trnL intron and ndhF sequence data. *Syst Bot* **25**: 468-478.

Wagner DB, Furnier GR, Saghaimaroof MA, Williams SM, Dancik BP, Allard RW. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc Natl Acad Sci USA* **84**: 2097-2100.

Wallace LE. 2003. Molecular evidence for allopolyploid speciation and recurrent origins in Platanthera huronensis (Orchidaceae). *Int J Plant Sci* **164**: 907-916.

Walters R, Yawney H. 2004. Silvics Manual: Volume 2: Hardwoods. *United States Department of Agriculture, Forest Service, Washington, DC*.

Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS. 2010. Patterns of hybridization in plants. *Perspectives in Plant Ecology Evolution and Systematics* **12**: 175-182.

Wood CE. 1955. Evidence for the hybrid origin of Drosera anglica. *Rhodora* **57**: 105-130.

Wyatt R, Antonovics J. 1981. Butterflyweed re‐revisited: spatial and temporal patterns of leaf shape variation in Asclepias tuberosa. *Evolution* **35**: 529-542.

# 3. Case Study of Hybridization in Wild Tomato

## 3.1. Overview

*Title*: Population genomics reveals a new wild tomato species with a history of hybridization

*Authors*: Ian Beddows[1,2], Aparna Reddy[1], Thorsten Kloesges[1] and Laura E. Rose[1,2,3]

[1]Institute of Population Genetics,
[2]International Graduate School in Plant Sciences (iGRAD-Plant),
[3]Cluster of Excellence on Plant Sciences (CEPLAS),
Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

*Author for correspondence*:
Laura Rose
*Tel*: +49 2118113406
*Email*: Laura.Rose@hhu.de

*Published:* This manuscript is currently under peer review. The manuscript has been reformatted for this thesis.

*Word Count*: Abstract 196, Introduction 1024, Materials and Methods 1225, Results 2433, Discussion 2743, Acknowledgements/Author Contributions/Data Accessibility 97, **Total 7425**.
*Number of figures*: 6
*Number of tables*: 1
*Number of Supporting Information files*: 1 (Supplemental Figure 1-17; Supplemental Tables 1-5).

*Contribution*:

- Contributed ideas and data
- Performed nearly all analyses
- Made all figures
- Interpreted data
- Wrote the manuscript

## 3.2. Abstract


- Hybridization between closely related plant species is widespread, but the outcomes of hybridization are not fully understood. This study investigates phylogenetic relationships and the history of hybridization in the wild tomato clade (*Solanum* sect. *Lycopersicon*).
- We sequenced RNA from individuals of 38 different populations and, by combining this with published data, build a comprehensive genomic dataset for the entire clade.
- The data indicate that many taxa are not monophyletic and many individuals are admixed due to repeated hybridization. The most polymorphic species, *S. peruvianum*, has two genetic and geographical subpopulations, while its sister species, *S. chilense,* has distinct coastal populations and reduced heterozygosity indicating a recent expansion south following speciation from *S. peruvianum* c. 1.25 million years ago. Discontinuous populations west of 72° are currently described as *S. chilense*, but are genetically intermediate between *S. chilense* and *S. peruvianum*.
- Based upon molecular, morphological, and crossing data, we test the hypothesis that these discontinuous *'S. chilense'* populations are an example of recombinational speciation. Recombinational speciation is rarely reported, and we discuss the difficulties in identifying it and differentiating between alternative demographic scenarios. This discovery presents a new opportunity to understand the genomic outcomes of hybridization in plants.


*Key words*: admixture, gene flow, hybridization, Lycopersicon, population genetics

## 3.3. Introduction

By some estimates, greater than 25% of plant species hybridize in the wild, and the prevalence of hybridization has convinced some botanists to question if Mayr's Biological Species Concept (BSC) is appropriate for plants (Coyne and Orr 2004; Ehrlich and Raven 1969; Levin 1979; Mallet 2005; Mayr 1942). The BSC defines species as actual or potentially interbreeding populations which are reproductively isolated from other such groups. At first, the observation of interspecific hybridization does seem to be at odds with this definition, but two species can remain distinct, even if they occasionally hybridize. For example, hybrids may be produced very rarely, the hybrid embryos may abort, or the hybrids themselves may be unfit or sterile. Because of the distinction between hybridization and gene flow, most authors do not take a hard line on hybridization between otherwise ‘good’ species as defined by the BSC (Coyne and Orr 2004). Furthermore, as shown by Rieseberg, et al. (2006), most taxonomic plant species fit the rubric of the BSC, in that they represent reproductively independent groups.

Speciation is a long-term process and breeding barriers will necessarily be incomplete when the taxa in question are young. In this case, hybridization can happen naturally during the divergence process and is not necessarily unexpected (Grant 1981). This idea is supported by hybridization being widespread in many taxonomic groups, including between humans and our closest relatives (Racimo, et al. 2015). Hybridization is, however, more common in some groups than others. For example, nearly 1 in 10 bird species hybridize with at least one other species (Grant and Grant 1992). Hybridization is also relatively common in flowering plants with c. 0.09 hybrids per non-hybrid species (Whitney, et al. 2010).

One of the most important consequences of hybridization is introgression (Anderson 1949). If first generation hybrids are not completely sterile, then they can backcross to one or both of the parental taxa, and successive generations of recurrent backcrosses will lead to introgression – the incorporation of one species’ alleles into the background genome of another. Introgression can transfer advantageous alleles between species, as in the case of Neanderthal introgressions which constitute 1-2% of present-day non-African genomes or Denisovan introgressions into the ancestors of Tibetans (Racimo, et al. 2015). Introgressions are also common in plants. For example, introgressions from oxford

ragwort (*Senecio squalidus*) potentially increase the rate of outcrossing in the common groundsel (*S. vulgaris*) by reintroducing a phenotype (ray flowers) lost in this otherwise predominately selfing species (Kim, et al. 2008).

A second important outcome of hybridization is hybrid or 'reticulate' speciation which combines two parental species to create a new one (Rieseberg and Willis 2007). In plants, hybrid speciation is normally accomplished by a doubling of chromosome number in the hybrid compared to the parental taxa (Soltis and Soltis 2012). In this case, breeding barriers between the hybrid and its parents are immediate because backcrosses are sterile due to abnormal meiosis (Paun, et al. 2009). However, hybrid speciation without a change in chromosome number can also occur and is known as recombinational speciation (Grant 1981). The rate of recombinational speciation is not known, but there are several well-documented examples in plants (Coyne and Orr 2004; Paun, et al. 2009).

The tomato clade (Solanum sect. *Lycopersicon*) split from the nearest neighboring section (Juglandifolia) c. 5.8-8 million years ago (Mya), but likely diversified only c. 2.5 Mya (Pease, et al. 2016; Sarkinen, et al. 2013). There are 13 species in the clade. All are diploid (n=x=12) and share high levels of chromosomal synteny, but there are also detectable cytological differences between some species (Anderson, et al. 2010; Chetelat and Ji 2007; Peralta, et al. 2008). Many species, including relatively distant taxa within the clade, are compatible in test crosses. Others are incompatible, normally resulting in aborted embryos and hybrid breakdown when intercrossed (post-zygotic incompatibility).

Considerable intraspecific diversity in plant size, shape, habit, and other characters has made systematics following the morphological species concept difficult in wild tomato. Furthermore, incomplete lineage sorting (ILS) and introgressions continue to present a challenge for molecular taxonomic studies, even as new methods have been adopted (Breto, et al. 1993; Pease, et al. 2016; Zuriaga, et al. 2009). Thus, although wild tomato species have been the focus of numerous morphological, phylogenetic, and biosystematic studies, the ancestry and definition of specific taxa within the clade remain unresolved.

This study is focused on sister species, *Solanum chilense* Dunal (*S.chi*) and *S. peruvianum* L. (*S.per*) which are the most polymorphic in wild tomato. These species are both perennial, green-fruited, and self-incompatible. They are

sympatric in southern Peru, but display differences in morphology, particularly leaflet shape (Fig. 1a,b), and *S.chi* has several adaptations for more arid habitats including grayish pubescence and deep roots (Moyle 2008). Their most recent common ancestor (MRCA) has been dated to between 0.5 and 2 Mya making them quite young (Pease, et al. 2016; Staedler, et al. 2008). Hybrid seed failure is the predominant outcome when they are intercrossed in the lab, but several studies have found genetic evidence for allele sharing in the wild, including the suggestion of speciation under residual gene flow (Staedler, et al. 2008; Staedler, et al. 2005).

In this study, we sequenced the RNA transcriptomes from different populations of *S.chi* and *S.per*. We aimed to determine the divergence time of the two species, test the hypothesis of speciation under residual gene flow, assemble a dataset that can serve as a null model for evolutionary studies, and – by including comparable public data – tackle taxonomic problems in the entire tomato clade that have remained unresolved without a comprehensive sampling of *S.chi* and *S.per*.

By combining our data with comparable datasets, we recovered the main phylogenetic groups within the clade, but also discovered taxonomic conflicts not evident before, including evidence of hybridization in multiple taxa. Surprisingly, the genomic analyses reveal little allele sharing between *S.chi* and *S.per,* with the exception of populations described as *S.chi* near Arequipa, Peru. We tested the hypothesis that this group of populations represents a recent example of hybrid speciation, and discuss both how natural hybridization can generate new genetic entities within a clade and the difficulties in distinguishing hybrid speciation from alternate demographic scenarios.

## 3.4. Materials and Methods

### 3.4.1. Transcriptome Data

We sequenced 18 *S.chi* and 17 *S.per* covering the known distribution of these species (Fig. 1c; Table S1). Two outgroups and one *S. corneliomulleri* J. F. Macbr. (*S.cor*) were also sequenced. Seeds originating from natural populations in Peru and Chile were kindly provided by the Charles M. Rick Tomato Genetics Resource Center (TGRC), University of California, Davis (tgrc.ucdavis.edu). The

seeds were germinated following TGRC guidelines and grown in a glasshouse in Düsseldorf, Germany.

We chose mRNA sequencing due to the high level (>76%) of heterochromatic repeats that would constitute a majority of the reads if genomic DNA was sequenced (Peterson, et al. 1996). Leaf RNA was extracted from one individual per accession using the RNeasy Plant Mini Kit (Qiagen, Germany). The leaf mRNA was then prepared with the TruSeq RNA Library Preparation Kit v2 or the NEBNext Ultra Directional RNA Library Prep Kit for Illumina and sequenced with Illumina HiSeq2500 100-nt paired-end technology at the Max Planck Genome Center (Cologne, Germany). Final libraries had a minimum of 35 million reads with a median of 92.9 million 100-nt reads following quality control and adapter removal.

### 3.4.2. Additional Data

To systematically evaluate the wild tomato clade, 28 genomic (ENA PRJEB5235) and 14 transcriptomic (NCBI Bioproject PRJNA305880) Illumina libraries were downloaded (Aflitos, et al. 2014; Pease, et al. 2016) and co-analyzed. By including this data, 80 individuals from 13 accessions (including the two outgroups) are represented (Table S1). Additional genomic data from Lin, et al. (2014) was included for one ∂a∂i analysis (Table S2).

### 3.4.3. Read mapping to the reference genome and SNP calling

Read libraries were individually mapped to the *S. lycopersicum* Heinz 1706 reference genome release SL2.50 with BWA v0.7.10 (Li and Durbin 2009; Sato, et al. 2012). We allowed up to 5% divergence from the reference and disallowed insertions greater than 25 (-k 1 -l 25 -n 0.05 -e 15 -i 10). For the mRNA libraries only, reads not mapped by BWA were remapped in TopHat2 (Kim, et al. 2013). Alignment files were then sorted and indexed using SAMtools v0.1.19 (Li, et al. 2009). All non-uniquely aligned reads and reads with mapping quality <30 were removed.

To identify polymorphisms, we used the multiallelic caller of BCFtools v1.3.1. Indels were removed, and the resulting unphased files were processed in BEAGLE 4.1 to infer haplotypes (Browning and Browning 2016). Positions with

coverage less than 10 in any single individual were treated as missing data, and positions with >50% missing data across all individuals were excluded. Polymorphisms were categorized as 5'UTR, coding sequence (CDS), intron, 3'UTR, or intergenic using the reference GFF. Polymorphisms mapping to CDS were further characterized as synonymous, nonsynonymous, changing a start codon, changing a stop codon, or nonsense mutations.

### 3.4.4. Interspecific relationships and genetic groups within wild tomato

We inferred the relationships of all species by maximum-likelihood (ML) using 429,881 synonymous positions. We ran 100 bootstraps under the rapid bootstrap algorithm of RAxML v8.2.9 with a GTR-GAMMA model of nucleotide substitution and an ascertainment bias for invariable sites (Stamatakis 2014). The two allied *Solanum* species were used to root the trees which were visualized in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/).

A second phylogenetic analysis was implemented with SNAPP v1.2.5 which uses a coalescent model without the need to directly infer trees (Bryant, et al. 2012). To reduce computational time, we randomly selected 1,250 synonymous polymorphisms and excluded allied *Solanums* and the 'Hirsutum' group, which is the first diverging lineage in sect. *Lycopersicon*. XML input files were created with the default parameters of BEAUti v2.3.1 (Drummond, et al. 2012). Following 1 million MCMC iterations in BEAST v2.3.1 and examination of log files in Tracer v1.6.0 (Rambaut, et al. 2014) the burn-in was set to 100,000 iterations. Coalescent trees were visualized with Densitree v2.2.2 (Bouckaert, et al. 2014).

The model-based clustering software STRUCTURE v2.3.4 was used to determine the number of genetic clusters (K) in sect. *Lycopersicon* (Pritchard, et al. 2000). To find K, we first removed the 'Hirsutum' group and generated 10 independent sets of 10,000 randomly chosen synonymous polymorphisms from positions with ≥10 coverage and <10% missing data. We modeled K=1-7 with a burn-in period of 100,000 followed by 100,000 MCMC steps under the admixture model. The greedy algorithm from CLUMPP assigned average membership coefficients from 10 independent runs (Jakobsson and Rosenberg 2007). The program STRUCTUE HARVESTER v0.6.94 was used to calculate the ad hoc

statistic ΔK (Earl and Vonholdt 2012). Evanno, et al. (2005) showed that ΔK, which is derived from the second order change in the log-likelihood, is accurate at finding the true K at the maximum hierarchical level.

Due to evidence for hybridization between taxa (i.e. intermediate individuals) from STRUCTURE, we built a reticulate network using SplitsTree4 (Huson and Bryant 2006). Reticulate networks do not require a tree-like model which allows more complicated evolutionary histories to be represented. Similarly, a principle component analysis (PCA) on synonymous polymorphisms was calculated using the R package APE. The PCA was run using the prcomp function (Paradis, et al. 2004; R Core Team 2014).

### 3.4.5. Within-species nucleotide diversity and individual heterozygosity

To estimate pairwise nucleotide diversity (π) within species, we derived accession-specific genomes for all individuals using the reference genome and called SNPs. Sites were called for all positions with coverage ≥10 reads. Coding sequences based on the ITAG2.4 genome annotation were then extracted and π at synonymous ($\pi_{syn}$) and at nonsynonymous ($\pi_{nonsyn}$) sites were calculated following Nei and Gojobori (1986). Heterozygosity was calculated for all individuals by dividing the total number of heterozygous positions with ≥10 coverage by all positions with ≥10 coverage in that individual. $F_{ST}$ was calculated between different species and subgroups with VCFtools v0.1.13 (Danecek, et al. 2011).

### 3.4.6. Modeling the joint demography of *S.chi* and *S.per*

We estimated the joint demography of *S.chi* and *S.per* using *∂a∂i* (Gutenkunst, et al. 2009). Demographic inference in *∂a∂i* uses a diffusion-based approach to model the distribution of multi-population allele frequency spectra. The joint site frequency spectrum (JSFS) was derived from 289,563 synonymous polymorphisms that had a nonzero allele frequency in *S.chi* or *S.per*. Individuals of *S. huaylasense* Peralta (*S.hua*) and *S.cor* were included as *S.per* by default, but any individual with >10% mixed ancestry in the STRUCTURE analysis was excluded.

Demographic parameters were estimated in a simple model that had an ancestral speciation event at time tau (T1) followed by the potential for population size changes and migration between species (Supplemental Information). We did 100 independent 10,000-iteration runs from randomized starting parameter values to find the optimum global parameter values and 100 conventional bootstraps to determine confidence intervals. To normalize for sequence length, we divided theta by the total number of potentially synonymous positions in all individuals (Supplemental Information).

### 3.4.7. Test crosses

Test crosses were made to determine if reproductive barriers were present between two cryptic hybrid populations (see Results) and their parental species. Multiple individuals from the following accessions were grown: Hybrid-LA1930, Hybrid-LA1932, *S.chi*-LA2930, *S.chi*-LA1960, *S.per*-LA1954, *S.per*-LA2732, *S.per*-LA0153, *S.per*-LA2964, and *S.cor*-LA1274. Test crosses were done for all combinations with the exception of *S.chi*-LA1960 and Hybrid-LA1930, which failed to flower. From the remaining 7 accessions, 815 flowers were bagged and hand-pollinated. Following a minimum of 50 days, fruit diameter, the number of seeds per fruit, and the number of seed-like structures (SLS; i.e. ovules not completely developed into seeds) were counted. The germination of seeds and some larger SLSs were determined following TGRC germination protocols (tgrc.ucdavis.edu).

## 3.5. Results

We analyzed 80 individuals from 11 wild tomato species and two outgroups. On average 78% of the mRNA reads were uniquely aligned to the Heinz 1706 reference genome with a mapping quality ≥30 (Fig. S1a; Table S1). In contrast, likely due to heterochromatic repeats, only 56% of the reads from the genomic data (ENA PRJEB5235) aligned uniquely. A mean of 63 million positions per individual had ≥10 mapping coverage for mRNA data (mean of 285 million positions for genomic data), but only 8.04 million positions had a ≥10 coverage in *all* individuals (Fig. S1b; Fig. S2). By allowing an intermediate amount of missing data (50%) as recommended by Streicher, et al. (2016), the number of positions increases to 35.64 million. This represents 4.2% of the total genome, but contains more than 66% of coding positions. In total, 4,866,729 bi-allelic polymorphisms with ≥10 coverage and ≤50% missing data were identified. This included 1,385,292 synonymous, 1,243,177 nonsynonymous, 903 start codon change, 2,410 stop codon change, and 62,000 nonsense mutations. The remaining polymorphisms were intergenic, 5'UTR, intronic, or 3'UTR (Fig. S3).

### 3.5.1. Distribution of genetic variation across groups

Three major genetic groups were consistently identified (Fig. 2, Fig. 3, Fig. S4, Fig. S5). One group contained only individuals of *S.chi*, one group contained individuals of *S.per* sensu lato (including *S.cor* and *S.hua*) and the third group contained all of the mostly autogamous taxa in the Esculentum and Arcanum species groups (named here Arc+Esc, Fig. 2). Sequence polymorphism was the greatest for the *S.per* group ($\pi_{syn}$ = 1.69%) followed by *S.chi* (1.27%). The autogamous group (Arc+Esc) had the lowest diversity ($\pi_{syn}$ = 1.04%). The level of nonsynonymous polymorphism was similar, but considerably lower with $\pi_{non}$ = 0.22% (*S.per*), $\pi_{non}$ = 0.18% (*S.chi*), and $\pi_{non}$ = 0.16% (Arc+Esc). Average individual heterozygosity was greatest in the allogamous-SI *S.per* (0.51%) and *S.chi* (0.43%) (Fig. S6; Table S3).

*Solanum peruvianum* showed strong population subdivision, evident in nearly all analyses (Fig. 2; Fig. 3). One subpopulation contains low-elevation collections from the sandy coast and/or Lomas formations of the Peruvian desert (Fig. 4). These populations (LA1951, LA1954, LA1336, LA1333, LA1474, LA2964,

and LA3218) are located between 14° and 17° S, and are less than 5 km from the coast and collected at sites at less than 600 m in elevation (with the exception of LA1474 which is located at 1300 m, 14 km from the coast). The second subpopulation has collections distributed across many different watersheds, mainly in central Peru, but also includes three *S.per* from northern Chile, *S.cor*-LA0118, *S.cor*-LA1274, and all of *S.hua.* This non-coastal subpopulation has higher nucleotide diversity ($\pi_{syn}$ = 1.57%) than the coastal one ($\pi_{syn}$ = 1.07%). Individuals from the non-coastal subpopulation also have significantly more unique SNPs per individual (65,127 ± 15,531, SD) compared to the coastal ones (30,205 ± 7,025 SD, t-test $P < 10^{-5}$).

Mean $F_{ST}$ between *S.chi* and *S.per* was 0.070 which was less than the mean $F_{ST}$ between the two subpopulations of *S.per* ($F_{ST}$=0.074) (Fig. S7). The coastal subpopulation of *S.per* has a higher mean $F_{ST}$ to *S.chi* (0.14) and Esculentum (0.33) than the non-coastal deme does (0.09 and 0.13).

*Solanum chilense* shows relatively little population structure compared to *S.per.* Individual heterozygosity decreases in the more southern populations of *S.chi* ($r = 0.72$, $P < 1.1 \times 10^{-5}$, Fig. S8). Furthermore, three accessions (LA2750, LA2930 and LA0752) are always distinguishable as a clade (Fig. 2, Fig. 5). Two of these accessions, LA2750 and LA2930, are from low-elevation coastal regions of Chile, and LA0752 is a northern *S.chi* population described in the final results section.

### 3.5.2. Recent speciation of *S.chi* and *S.per*

We used *∂a∂i* to model the ancestry of the sister species *S.chi* and *S.per* because *∂a∂i* is appropriate for detecting recent demographic events (Gutenkunst, et al. 2009). We fit the synonymous JSFS to a simple model and this returned an estimate for the speciation time (tau) at 1.46-1.56 times the size of the MRCA (Table S4). If we assume a per site mutation rate of $5.1 \times 10^{-9}$ (Roselius, et al. 2005; Staedler, et al. 2008), then the speciation time is estimated to be between 1.2 and 1.4 million generations ago. We found evidence of population expansion in both species relative to their MRCA. *S.per* has an estimated population size of 1.54 – 1.70 million individuals which is nearly three times larger than the estimate for *S.chi* (0.52 – 0.58 million). We detect low levels of reciprocal gene flow

between the two species when compared to previous reports; gene flow from *S.chi* into *S.per* was 0.27 individuals per generation and the reciprocal was 0.12 individuals per generation. All maximum-likelihood parameter values are provided in Table S4 and the model fit is visualized in Fig. S9.

### 3.5.3. Evidence of natural *S.chi* × *S.per* hybrid populations in southern Peru

Three accessions of *S.chi* (LA1782, LA1930, and LA1932) were all collected in the Acari river drainage near Arequipa, in southern Peru. These populations were considered to be among the northernmost of *S.chi* (Fig 1b), but the following observations indicate that these populations are genetically not individuals of *S.chi*:

1. The genomes of all individuals show c. 35% corresponding to *S.per* and the remainder corresponding to *S.chi* in the STRUCTURE analysis (Fig. 3).
2. These populations are intermediate between the *S.chi* and *S.per* in the PCA and network analyses (Fig. 5, Fig. S4).
3. They form a monophyletic clade located between *S.chi* and *S.per* in all phylogenetic analyses (Fig. 2).

According to collection records from the TGRC (tgrc.ucdavis.edu), wild individuals from these populations are described as vigorous, stress-tolerant, and long-lived (>10 years old). To characterize them morphologically and to exclude the possibility of seed or sample contamination, we regrew individuals from accessions LA1930, LA1932, *S.chi*-LA2930, *S.chi*-LA0752, *S.per*-LA0153, *S.per*-LA1954, *S.per*-LA2732, and *S.cor*-LA1274. Individuals from LA1930 and LA1932 were fast-growing, large, and could be distinguished from both *S.chi* and *S.per* although they were similar in leaf shape to northern *S.chi* populations (Fig. 1; Fig. 6). Three-month old plants had significantly longer leaves, thicker stems, and reduced lateral branching in comparison to the tested *S.chi*, *S.per*, and *S.cor* (t-test, $P < 0.01$; Fig. S10). They also had a very high density of type I trichomes on the lower stem, although this character was not quantified (Fig. 6i). These populations flowered reluctantly and later than the other species in our hands. The corollas were frequently recurved and 2-3 cm in diameter. The style was straight

and extended c. 2 mm beyond the anther tube. This and an absence of fruits on unpollinated flowers are consistent with outcrossing. Pollen fertility was normal.

Due to their genotype, phenotype, and the fact that two tetraploid populations of *S.chi* have been reported from southern Peru (Chetelat and Ji 2007; Rick 1990), we considered the possibility that these populations represent allotetraploids. However, flow cytometry indicated that they have a diploid C content (Supporting Information; Fig. S11).

These individuals appeared to be hybrids between *S.chi* and *S.per*, but it is difficult to distinguish hybridization from ILS or population subdivision using methods such as STRUCTURE, PCA, and phylogenetic reconstruction. The three-population test of (Patterson, et al. 2012) is a formal test for admixture and results in a negative $f_3$ if the tested population is admixed between parental populations by essentially looking for intermediate allele frequencies in the tested population with respect to the parents. This test did not result in a negative $f_3$ and therefore does not provide evidence for or against a recent history of admixture.

A second method for potentially differentiating hybridization from other scenarios is to look for chromosomal blocks from the parental species in the hybrids (Ungerer, et al. 1998). This was done with the program HAPMIX (Price, et al. 2009)(Fig. S12). We ran HAPMIX with a uniform recombination rate using all *S.chi* and *S.per* individuals to identify parental haplotypes. The admixed individuals were indeed composed of *S.chi* and *S.per* haplotypes. The mean haplotype was 112 kb long. Consistent with our previous analyses, the hybrid individuals appeared more *S.chi*-like than *S.per*-like, and the mean *S.chi* haplotype (218 kb) was longer than the mean *S.per* haplotype (55 kb). This analysis indicated that these individuals did indeed have a history of hybridization.

Test crosses with Hybrid-LA1932 largely failed to produce viable seeds with the four tested *S.per* populations and with *S.chi*-LA2930 (Table 1). Fruits of these crosses instead contained a large number of seed-like structures (i.e. SLS). The total number of SLS for crosses to a hybrid parent were 1,892 and only a small number of seeds were recovered (total seeds across all crosses with hybrid parent = 47) (Fig. S13). In contrast, seed number from fruits of conspecific crosses always substantially outnumbered SLSs (total seeds of conspecific crosses = 2,950 seeds; total SLS of conspecific crosses = 520). Furthermore, the small

seeds from the LA1932 × *S.chi* cross all failed to germinate (N=12) (Table 1). These hybrid populations are known to set seed in test crosses with other *S.chi* populations, but in reduced numbers (R. Chetelat pers. comm.).

Crosses between the hybrid and two *S.per* accessions (LA0153 and LA2964) resulted in a total of 35 seeds (average of two seeds per fruit for these two crosses). These seeds had a 50% germination frequency, indicating some potential for backcrossing to *S.per*, but all of the interspecific $F_1$ individuals eventually died while conspecific seedlings did not. No other *S.per* × LA1932 crosses produced seeds.

Interspecific crosses between *S.chi* and *S.per* recapitulated the outcome of crosses between the hybrid and each of these species. Fruits had an excess of SLS and few viable seeds (*S.chi* × *S.per* crosses resulted in 55 fruits, 971 SLS and 45 seeds). These interspecific seeds germinated, but all individuals died before reaching maturity.

Some intraspecific reproductive incompatibility was detected within *S.per* between the coastal and non-coastal demes. In one case, 56% of the ovules were aborted in the 14 fruits from the *S.per*-LA1954 × *S.per*-LA0153 crosses (but not the reciprocal). In a second example, seeds from *S.per*-LA2964 × *S.per*-LA2732 and the reciprocal cross had only 25% germination rate despite normal seed set. Overall, crosses were between the two genetic groups identified within *S.per* had significantly lower number of seeds per fruit compared to within-deme crosses (Wilcoxon test, $P < 0.05$). Some incompatibility was also detected between *S.cor*-LA1274 and *S.per*-LA1954 (mean of 5 seeds and 12 SLS per fruit). However, the *S.cor*-LA1274 × *S.per*-LA1954 cross was not different than within-deme *S.per* crosses. All crossing data is available in Table S5.

### 3.5.4. Broader phylogenetic implications

This large genetic dataset allowed us to test and validate some earlier phylogenetic observations within the Lycopersicon clade. Three of four well-established species groups within sect. *Lycopersicon* were monophyletic independent of the phylogenetic method (Fig. 2). In contrast, the Peruvianum group was not monophyletic in either the ML or coalescent phylogenies. In the ML analysis, Arc+Esc was initially derived from within *S.per* making *S.per* paraphyletic

(Fig. S14). To test if this was due to the inclusion of potentially admixed individuals (eight individuals had mixed membership between *S.per* and Arc+Esc in the STRUCTURE analysis), accessions with >10% mixed membership were removed and the ML analysis was redone. This resulted in Arc+Esc as a relative outgroup to *S.chi* and *S.per* as expected and restored monophyly to the Peruvianum group (Fig. 2a). However, *S.per* itself remained paraphyletic due to inclusion of *S.hua* and *S.cor*, both of which themselves are polyphyletic.

In contrast to the ML analysis, Arc+Esc is not derived from within *S.per* in the coalescent analysis. The Peruvianum group was, however, polyphyletic due to *S.per*-LA1913 and *S.hua*-LA1983 which are in a clade with Arc+Esc. Furthermore, in contrast to the ML analysis excluding admixed accessions, Arc+Esc and *S.per* are sister taxa with *S.chi* as a relative outgroup (Fig. 2b). These differences appear to be dependent on the inclusion of certain admixed accessions, giving strong indication that reticulate events are influencing phylogenetic relationships within the clade.

The species *Solanum arcanum* Peralta (*S.arc*) was polyphyletic in the ML and coalescent phylogenies. This is due to *S.arc*-LA2157 which is always an outgroup to all other species of the Arcanum group (Fig. 2). Evidence for a close relationship of *S.arc*-LA2157 and *S.per* is present in the STRUCTURE analysis (Fig. 3) and in the reticulate network (Fig. 5). suggesting that the polyphyly *S.arc* may be the result of past hybridization between an individual of Arc+Esc and *S.per*.

One *S.hua* accession, *S.hua*-LA1358, is indistinguishable from *S.per* while the five others appear to have mixed ancestry between Arc+Lyc and *S.per*. One of these accessions, *S.hua*-LA1360 was shown to have introgressions from the Esculentum group by Pease, et al. (2016) using the ABBA-BABA test. This accession and *S.hua*-LA1983 are nearly intermediate between *S.per* and Arc+Lyc in the STRUCTURE and network analyses (Fig. 3, Fig. 5). Like *S.hua, S.cor* is also polyphyletic in all phylogenetic analyses. Two *S.cor* individuals are indistinguishable from *S.per* while *S.cor*-LA0118 has c. 1% and *S.cor*-LA1274 c. 10% Arc+Esc component. The *S.cor* individuals never group together in any phylogenetic analyses.

We detect a minor signature of *S.chi* component in *S.per*-LA3636, *S.per*-LA1616, *S.cor*-LA0444, *S.cor*-LA0107, and *S.hua*-LA1358. This is also seen in the association of these accessions with *S.chi* along the first principle component (Fig. S4). *S.per*-LA1913 was unique in having mixture from all three STRUCTURE groups. Interestingly, as the number of clusters inferred by STRUCTURE increases, *S.per*-LA1913 continues to have genetic material from all of them (Fig. 3).

*Solanum chilense* was always monophyletic, and there was no evidence of mixed ancestry in any *S.chi* accessions with the exception of three samples that appear nearly intermediate between *S.chi* and *S.per,* as described above.

One '*S.per*' accession, LA0752, was collected in central Peru, yet is closest genetically to the southern coastal populations of *S.chi*. This anomalous accession was described as *S.chi*-like when collected, and we confirmed this by growing multiple individuals (Fig. S15). Based on the genetic and phenotypic evidence, LA0752 has been re-annotated as *S.chi*. To our knowledge, this is the most northerly accession of *S.chi* described. Furthermore, *S.chi*-LA0752 has the lowest heterozygosity of any *S.chi* individual (Fig. S8), and plant growth was weak. Its disjunct location, weak growth and low heterozygosity may indicate a long-distance dispersal and subsequent founder effect in the history of this population. However, collection error (i.e. mislabeling) cannot be excluded.

Finally, in the independent datasets used to build phylogenies of sect. *Lycopersicon,* individuals of six accessions from four species were duplicated: *S.arc*-LA2172, *S.hua*-LA1364, *S.neo*-LA2133, *S.per*-LA1954, *S.per*-LA2744, and *S.per*-LA2964. These duplicates are always sister taxa in the ML, coalescent, and network analyses (Fig. 2, Fig. 5). This concordance demonstrates the feasibility and consistency of combining sequence data from many independent studies (and labs) to address problematic questions in evolutionary biology.

## 3.6. Discussion

Our comprehensive population genomic analysis provides an in-depth view of population and lineage divergence among a group of closely related plant species. While many of our analyses confirmed the existence of three previously well-established groups within the section including 1) the monophyletic Hirsutum group, 2) the Esculentum clade, and 3) the Arcanum group, the novelty of our study is the extensive sampling and analysis of the two most polymorphic species in the clade: *S.chi* and *S.per.* In fact, the focus on these taxa lead to the surprising discovery of a new species of hybrid origin.

### 3.6.1. Demography and speciation of *S.chi* and *S.per*

Prior estimates of the divergence time between *S.chi* and *S.per* indicate a very recent speciation time of 0.18 x $2N_e$, where $N_e$ is the estimated size of *S.chi* (Naduvilezhath, et al. 2011). Assuming one generation per year and a mutation rate of 5.1 x $10^{-9}$ site per year, this corresponds to a split 730,000 years ago. This estimate was based upon a modest dataset of seven nuclear genes containing 954 polymorphic positions. However, the sample unknowingly included up to seven Quicacha individuals that we now know to represent *S.chi* × *S.per* hybrids (discussed below). The inclusion of these individuals not only contributed to the signature of on-going gene flow between species, but likely caused the speciation time to be under-estimated. Grounded upon a much larger dataset and following the identification and exclusion of admixed individuals, our analysis indicates that the species split was 1.51 x $2N_e$ generations before present. Assuming the same generation time and mutation rate, this corresponds to approximately 1.25 Mya, which is consistent with a previous family-wide dated phylogeny (Sarkinen, et al. 2013). However, age estimates based on the molecular clock need to be approached cautiously given disagreement between molecular data and new fossil evidence (Wilf, et al. 2017). The recency is, however, consistent with the observed low $F_{ST}$ and few fixed differences between these species. In our analyses, migration rates of 0.12 and 0.27 individuals per generation were estimated – in contrast to higher estimates from data which included cryptic hybrid individuals (e.g. Staedler *et al*. 2005).

We detect a lower amount of synonymous nucleotide diversity in *S.per*

(1.7%) compared to previous estimates of 2.1%, 2.5%, and 3.1% (Arunyawat, et al. 2007; Staedler, et al. 2012; Staedler, et al. 2005). This could be in part due to method-specific differences (i.e. next generation versus Sanger sequencing). However, our numbers may reflect a more accurate estimate of nucleotide diversity since it is based upon orders of magnitude larger number of nucleotide positions, more populations from both species, and did not include interspecific hybrid individuals.

We detected two genetic groups within *S.per,* similar to those described by Rick (1963) based upon morphology and by Nakazato, et al. (2012) based upon AFLPs. These two groups appear to represent distinct geographic demes and/or subspecies occupying different ecological niches. One contained seven low-elevation populations restricted to the coast and/or lomas formations of southern Peru. The second deme contained non-coastal central Peruvian populations, including most individuals of *S.cor* and *S.hua*.

Rick (1963) described the coastal *S.per* group as having less variation in shape, size, and habit between populations, but greater variation within any single population. Conversely, the non-coastal Peru populations were described as more restricted and more idiosyncratic (Rick 1963). This morphological observation is reinforced by our genetic data showing higher diversity and more private polymorphism in the non-coastal Central populations. An observation which could be explained by limited dispersal between different Andean river drainages.

In contrast, there are few geographical barriers to inhibit gene flow between coastal populations. The coastal deme also has higher mean $F_{ST}$ with both *S.chi* and Esculentum than the non-coastal deme. We interpret this difference due to recent gene flow between the non-coastal deme and both *S.chi* and the Esculentum group*:* the *S.chi* × *S.per* hybrids have non-coastal *S.per* component and the admixed Peruvianum group accessions (including *S.cor* and *S.hua*) are from the non-coastal deme whereas the coastal deme shows no admixture. Alternatively, the greater mean $F_{ST}$ could reflect differentiation of the coastal deme, perhaps from ecological adaptation.

Climatic conditions differ between the two subpopulations: fog is abundant along the coast in the Lomas formation from May to October, while rainfall is abundant in the central river drainages November to May (Taylor 1986). These

climatic differences may influence flowering time resulting in a prezygotic barrier that could account for the population subdivision. The sub-populations also show reduced interfertility (i.e. reduced seed number in LA0153 x LA1954 crosses). Recognizing these geographic races/subspecies of *S.per* as distinct species is not warranted due the low amount of genetic differentiation and the absence of pronounced incompatibility. Overall, these geographic races are a good opportunity for the 'magnifying glass' approach to study speciation-in-action (Via 2009).

The five *S.per* populations showing a low amount (<10%) of genetic similarity to *S.chi* according to STRUCTURE are all from the non-coastal deme near Lima, Peru. We do not detect admixture between sympatric populations, consistent with previous crossing studies (Rick and Lamm 1955). Interestingly, the five *S.per* populations with *S.chi* admixture are physically near the *S. chi*-LA0752 population (Fig. 5). Thus, if *S.chi*-LA0752 is not a collection error and represents a long-distance dispersal event, this could explain how genes from *S.chi* could be introduced into non-coastal *S.per* populations.

### 3.6.2. *S. corneliomulleri* and *S. huaylasense*

Relationships in the Peruvianum group remain challenging, even in the face of such an extensive dataset. Although many different species concepts exist, there is general consensus that species form discrete, evolutionarily independent lineages (de Queiroz 2005). Our data show that neither *S.cor* nor *S.hua* form discrete genetic clades as currently circumscribed. C. H. Muller (1940) first described *S.cor* as *Lycopersicon glandulosum*, but Macarthur and Chiasson (1947) and later Rick (1963) demonstrated the compatibility of *L. glandulosum* with other *S.per* accessions. Therefore *L. glandulosum* was renamed *L. peruvianum* var. glandulosum and later designated as a race of *S.per* (Warnock 1988). In fact, Rick (1963) noted at least five additional races, some currently included within *S.cor*, that were equally distinct from *S.per*. Our data is in agreement with other studies reporting the lack of genetic or ecological differentiation between *S.cor* and *S.per* (Labate, et al. 2014; Nakazato, et al. 2010; Pease, et al. 2016; Rodriguez, et al. 2009; Zuriaga, et al. 2009).

*Solanum huaylasense* was delineated from *S.per* using morphologically by

Peralta, et al. (2005). Our data included six individuals of this species from five accessions. One accession, LA1958, is indistinguishable from *S.per*, but the remaining five show some admixture with Arc+Esc. For example, LA1364, described to be admixed by Labate, et al. (2014) and again by Pease, et al. (2016), has c. 7% mixed ancestry. Other *S.hua* accessions show even more admixture, and the species is consistently polyphyletic. Thus, as currently defined, *S.hua* does not appear to be a natural group and some, maybe most, of the individuals described as *S.hua* could be of hybrid origin.

Given the increasing evidence, recognizing *S.cor* and *S.hua* as distinct species seems untenable. Instead, we have identified two well-differentiated demes in S.per: the coastal and non-coastal demes. The currently recognized representatives of *S.cor* and *S.hua* belong to the non-coastal deme. This deme is highly idiosyncratic and there has undoubtedly also been admixture, which has further contributed to the taxonomic difficulties and confusion.

### *3.6.3. Solanum arcanum*

A different variety of *S.per* called humifusum was first described by Muller (1940) and largely corresponds to the species now recognized as *S.arc*. Morphologically *S.arc* can be distinguished from other taxa by its unbranched inflorescences, straight anther tubes, and short styles (Müller 1940; Rick 1986). Although Rick and colleagues detected a reduction of cross-compatibility between the typical *S.per* and variety humifusum (aka *S.arc*), Rick, following the BSC, did not feel justified in recognizing humifusum as a distinct species from *S.per* because gene flow was theoretically possible through several intermediate populations (Rick 1979; Rick and Lamm 1955), Instead, Rick (1986) delineated four more-or-less reproductively isolated assemblages based on extensive reciprocal test-crosses: Chotano-humifusum, Chamaya-Cuvita, Marañ n, and typical *S.per*. Today's *S.arc* is the sum of the first three assemblages (Peralta, et al. 2008). But gene flow is possible between these assemblages and the typical *S.per,* through e.g. Chotano-humifisum populations (Rick 1979). In our data, *S.arc*-LA2172 (Marañ n) shared a more recent ancestor with *S. neorickii* while *S.arc*-LA2157 (Chotano) is an outgroup to the Arcanum group species and appears to have *S.per* ancestry. Interestingly, these two *S.arc* accessions have

actually been shown to be incompatible with one another by Rick (1986). Thus, while evidence supports the Arcanum group as biologically meaningful, the number of species within the group and the species-level assignment of individual accessions (particularly of individuals of *S.arc*) deserves further study and clarification.

### 3.6.4. Evidence for widespread cryptic hybrid populations

Herbarium records and TGRC collection data indicate that there are several other collections of *S.chi* from Arequipa near the interspecific hybrids identified in this study, and that all of these collections are geographically discontinuous from the rest of *S.chi* (Fig. S16). This does not appear to be a sampling artifact because many other wild tomatoes have been collected from this area. The reported collections of *S.chi* west of 72° are LA0869, LA1782, LA1917, LA1930, LA1931, LA1932, LA1934, LA1938, LA1939, LA3780, LA3784, LA3785, and LA3786. Most of these were collected in 1979 or 1996 and field notes indicated that they included many "tall upright plants" with "long peduncles" and "very large growth and very heavy fruit set," resembling our morphological observations. The genetic evidence for hybridization in three of these populations, the distinct and consistent morphological differences of these populations, and their geographic discontinuity with other *S.chi* led to the hypothesis that *all* of these discontinuous northern *S.chi* populations are of hybrid origin.

This hypothesis is supported by the following observations from previously published work. First, LA1782 and LA4117A were chosen to represent *S.chi* in a study of wild tomato evolution by Pease, et al. (2016). LA1782 was collected independently and c. 9 km from LA1930 and LA1932 in 1977, and is genetically indistinguishable from these populations (Fig. 2, Fig. 3). A finding consistent with the relatively long coalescent of the two sampled *S.chi* individuals in figures 2a and 2b of Pease, et al. (2016).

Second, Boendel, et al. (2015) sequenced 30 genes from 23 *S.chi* populations, including Hybrid-LA1930 and another putative hybrid, LA3784. The similarity between these two accessions and the separation of these two accessions from other *S.chi* in their analyses is consistent with LA3784 being genetically comparable to Hybrid-LA1930. In fact, Boendel, et al. (2015) explored

the idea of hybridization in these populations in their discussion, but were not able to make conclusions because they had data only from *S.chi* and not from *S.per*.

Third, Staedler and colleagues collected plants and seeds from *S.chi* and *S.per* in Peru and Chile in 2004 (Fig. S16)(Roselius, et al. 2005). We examined the voucher specimens from this 2004 trip, including two *S.chi* collections from Arequipa, Peru near Acari: Quicacha (QUI) and Nazca (NAZ) (Fig. S17). The NAZ collection includes a specimen of *S.chi* that is phenotypically very similar to the *S.chi* QUI population. The leaf morphology of both QUI and NAZ specimens is more similar to our sampled hybrids than to typical *S.chi* or *S.per*, indicating that the QUI and NAZ *S.chi* samples are also hybrids. Furthermore, genetic studies using this material came to conclusions consistent with our phenotypic observations. Based on this material, Staedler and colleagues estimated a very recent split time between the two species (<0.55 Mya), found an absence of fixed differences, and concluded that speciation occurred under residual gene flow (Staedler, et al. 2008; Staedler, et al. 2005). Subsequent studies based on this material describe trans-specific allele sharing and selection (due to the presence of *S.per* alleles in the QUI population) (Mboup, et al. 2012; Xia, et al. 2010). While there is no reason to question the data, we argue that the allele sharing resulted from the inclusion of cryptic hybrids in their sample rather than natural selection as they hypothesize.

Together, these observations support the conclusion that *all* of the populations described as *S.chi* west of 72° are genetically equivalent and therefore of hybrid origin. Given that their genomic constitution is composed of *S.chi* and *S.per* haplotype blocks and the hybrid accessions are diploid, these data are consistent with this being an example of recombinational speciation in wild tomato. However, because they are not shown to be definitively admixed when formally tested, other hypotheses must also be considered, including, for example, recent introgressions of *S.per* haplotypes into distinct populations of *S.chi*. Note that the lack evidence for admixture according to the $f_3$ test could be due to the small number of hybrid individuals tested or drift within the hybrid populations following their creation.

### 3.6.5. A putative example of recombinational speciation

Recombinational speciation is the rapid formation of a new species resulting from a cross between two closely related species, without a change in chromosome number. It is rare, with only c. 20 examples in plants, and many of these examples are unconvincing (Rieseberg 1997; Rieseberg and Willis 2007; Stuessy, et al. 2014). Rarity may be due to poor documentation because there is no cytotaxonomic evidence and because hybrid species may be difficult to recognize and distinguish morphologically (Rieseberg 1997). However, this form of speciation may simply be less common because barriers to gene flow, such as those introduced by a change in ploidy, are not present at the outset. Without such barriers, hybrids inevitably backcross to the more abundant parental species, leading to their eventual disappearance (Baack, et al. 2005).

Theoretical studies on hybrid speciation have therefore emphasized the role of ecology and the necessity of an open habitat for the hybrids to separate them from their parental taxa (Anderson 1949; Buerkle, et al. 2000; Buerkle and Rieseberg 2008; Gross and Rieseberg 2005). Simulations also show that this type of rapid speciation is more likely in perennial species (reviewed in Stuessey et al. 2014), a condition met by *S.chi* and *S.per*. While recombinational speciation is theoretically more likely in self-compatible species, it can also occur in outcrossing taxa, and, interestingly, most of the convincing examples are outcrossers such as *S.chi* and *S.per* (McCarthy, et al. 1995; Rieseberg 1997).

The hybrid populations show strong reproductive barriers to all tested *S.per* populations. While the cross-compatibility of the hybrid populations to northern *S.chi* (R. Chetelat, pers. comm.) would allow backcrossing, their non-overlapping distribution would generally shield them from gene swamping by *S.chi*. However, their morphological similarity to the northern Chilean *S.chi* populations and their genomic composition seems to indicate historical backcrossing to *S.chi*. Alternatively, their similarity to *S.chi* could be accounted for by differential segregation in the $F_2$ or later generation hybrids, or these populations could be a distinct subpopulation of *S.chi* with introgressions from *S.per*. It is difficult to distinguish between these scenarios, but the consistent phenotype and genotype of the hybrids from independent collections from the 1970s to 2004 (and following seed expansion at TGRC) and the small haplotype size make it clear that they are stabilized derivatives and not first or early generation hybrids. Because they are

older, the different scenarios of hybridization, introgression, and population subdivision are especially difficult to distinguish.

Schumer, et al. (2014) give three criteria that need to be met in a definitive example of recombinational speciation. Most purported examples fail to meet all of these criteria. The first criterion is reproductive isolation of the hybrid species from its parents. The second is genetic evidence of hybridization. These two criteria are fulfilled here, but the third criterion – showing that hybridization resulted in reproductive barriers and speciation – is more challenging. This has only been demonstrated once in plants by Rieseberg, et al. (2003) who recreated the extreme phenotypes of hybrid sunflower species. These tomato populations are a good starting point for tests of reproductive barriers, and further mapping and cytological work employing them could narrow down the incompatibility loci as done for other species pairs in the clade (Moyle and Nakazato 2008). Such work is also ultimately needed to demonstrate recombinational speciation in wild tomato. Further studies can also help clarify the date of admixture and the exact compatibilities of these populations.

In conclusion, section *Lycopersicon* provides a window into the speciation continuum, from population subdivision to speciation, and includes one and possibly more hybrid taxa. Knowing the ancestry of these populations and species is fundamental for addressing future questions about the genomics of ecological adaptation and the development of breeding barriers in the clade.

**Acknowledgements**

The authors thank the TGRC for seed and Roger Chetelat for comments, Sandra Knapp for herbarium records, and Hajo Esser for herbarium specimens. This research was supported by the Research Training Group GRK1525 and DFG grants: Ro 2491/5-2 and Ro 2491/6-1.

**Author Contributions**

L.R. and I.B. conceived the study. I.B. and A.R. grew the samples and A.R. extracted RNA. I.B. and T.K. analyzed data. I.B. and L.R. wrote the manuscript. All authors approved the final version of the manuscript.

**Data Accessibility**

Nucleotide sequence data generated in this study have been deposited in NCBI under Bioproject PRJNA329478.

## 3.7. References

Aflitos S, et al. 2014. Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. Plant Journal 80: 136-148.

Anderson E. 1949. Introgressive hybridization: John Wiley and Sons, NY.

Anderson LK, Covey PA, Larsen LR, Bedinger P, Stack SM 2010. Structural Differences in Chromosomes Distinguish Species in the Tomato Clade. Cytogenetic and Genome Research 129: 24-34.

Arunyawat U, Stephan W, Stadler T 2007. Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. Molecular Biology and Evolution 24: 2310-2322.

Baack EJ, Whitney KD, Rieseberg LH 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in Helianthus homoploid hybrid species. New Phytologist 167: 623-630.

Boendel KB, et al. 2015. North-south colonization associated with local adaptation of the wild tomato species Solanum chilense. Molecular Biology and Evolution 32: 2932-2943.

Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10: e1003537.

Breto MP, Asins MJ, Carbonell EA 1993. Genetic variability in Lycopersicon species and their genetic relationships. Theoretical and Applied Genetics 86: 113-120.

Browning BL, Browning SR 2016. Genotype Imputation with Millions of Reference Samples. American Journal of Human Genetics 98: 116-126.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29: 1917-1932.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH 2000. The likelihood of homoploid hybrid speciation. Heredity 84: 441-451.

Buerkle CA, Rieseberg LH 2008. The rate of genome stabilization in homoploid hybrid species. Evolution 62: 266-275.

Chetelat RT, Ji YF. 2007. Cytogenetics and evolution. In: Razdan MK, Mattoo AK, editors. Genetic improvement of Solanaceous crops. Enfield, N.H.: Science Publishers. p. 77-112.

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

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

de Queiroz K. 2005. Ernst Mayr and the modern concept of species. In: Hey J, Fitch W, Ayala FJ, editors. Systematics and the origin of species. Washington D.C.: National Academy of Sciences. p. 243 - 263.

Drummond AJ, Suchard MA, Xie D, Rambaut A 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29: 1969-1973.

Earl DA, Vonholdt BM 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

Ehrlich PR, Raven PH 1969. Differentiation of populations. Science 165: 1228-1232.

Evanno G, Regnaut S, Goudet J 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

Grant PR, Grant BR 1992. Hybridization of bird species. Science 256: 193-197.

Grant V. 1981. Plant speciation: New York: Columbia University Press.

Gross BL, Rieseberg LH 2005. The ecological genetics of homoploid hybrid speciation. Journal of Heredity 96: 241-252.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics 5: e1000695.

Huson DH, Bryant D 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267.

Jakobsson M, Rosenberg NA 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.

Kim D, et al. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14: R36.
Kim M, et al. 2008. Regulatory genes control a key morphological and ecological trait transferred between species. Science 322: 1116-1119.
Labate JA, Robertson LD, Strickler SR, Mueller LA 2014. Genetic structure of the four wild tomato species in the Solanum peruvianum s.l. species complex. Genome 57: 169-180.
Levin DA 1979. The nature of plant species. Science 204: 381-384.
Li H, Durbin R 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.
Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
Lin T, et al. 2014. Genomic analyses provide insights into the history of tomato breeding. Nature Genetics 46: 1220-1226.
Macarthur JW, Chiasson LP 1947. Cytogenetic notes on tomato species and hybrids. Genetics 32: 165-177.
Mallet J 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20: 229-237.
Mayr E. 1942. Systematics and the origin of species. New York, NY, USA: Columbia University Press.
Mboup M, Fischer I, Lainer H, Stephan W 2012. Trans-species polymorphism and allele-specific expression in the CBF gene family of wild tomatoes. Molecular Biology and Evolution 29: 3641-3652.
McCarthy EM, Asmussen MA, Anderson WW 1995. A theoretical assessment of recombinational speciation. Heredity 74: 502-509.
Moyle LC 2008. Ecological and evolutionary genomics in the wild tomatoes (Solanum sect. Lycopersicon). Evolution 62: 2995-3013.
Moyle LC, Nakazato T 2008. Comparative genetics of hybrid incompatibility: Sterility in two Solanum species crosses. Genetics 179: 1437-1453.
Müller CH 1940. A revision of the genus Lycopersicon. U.S.D.A. Misc. Publ. 382: 1-28.
Naduvilezhath L, Rose LE, Metzler D 2011. Jaatha: a fast composite-likelihood approach to estimate demographic parameters. Molecular Ecology 20: 2709-2723.
Nakazato T, Franklin RA, Kirk BC, Housworth EA 2012. Population Structure, Demographic History, and Evolutionary Patterns of a Green-fruited Tomato, Solanum peruvianum (Solanaceae), Revealed by Spatial Genetics Analysis. American Journal of Botany 99: 1207-1216.
Nakazato T, Warren DL, Moyle LC 2010. Ecological and geographic modes of species divergence in wild tomatoes. American Journal of Botany 97: 680-693.
Nei M, Gojobori T 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418-426.
Paradis E, Claude J, Strimmer K 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.
Patterson N, et al. 2012. Ancient Admixture in Human History. Genetics 192: 1065-1093.
Paun O, Forest F, Fay MF, Chase MW 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytologist 182: 507-518.
Pease JB, Haak DC, Hahn MW, Moyle LC 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLoS Biology 14: e1002379.
Peralta IE, Knapp SK, Spooner DM 2005. New species of wild tomatoes (Solanum section Lycopersicon: Solanaceae) from northern Peru. Systematic Botany 30: 424-434.
Peralta IE, Spooner DM, Knapp S 2008. Taxonomy of wild tomatoes and their relatives. Systematic Botany Monographs 84.
Peterson DG, Price HJ, Johnston JS, Stack SM 1996. DNA content of heterochromatin and euchromatin in tomato (Lycopersicon esculentum) pachytene chromosomes. Genome 39: 77-82.
Price AL, et al. 2009. Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. PLoS Genetics 5.

Pritchard JK, Stephens M, Donnelly P 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.
R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. In: ISBN 3-900051-07-0.
Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E 2015. Evidence for archaic adaptive introgression in humans. Nature Reviews Genetics 16: 359-371.
Rambaut A, Suchard MA, Xie D, Drummond AJ 2014. Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.
Rick CM 1963. Barriers to interbreeding in Lycopersicon peruvianum. Evolution 17: 216-232.
Rick CM. 1979. Biosystematic studies in Lycopersicon and closely related species of Solanum. In: Hawkes JG, Lester RN, Skelding AD, editors. The Biology and Taxonomy of the Solanaceae. London: Academic Press. p. 667-678.
Rick CM 1990. New or otherwise noteworthy accessions of wild tomato species. Tomato Genet. Coop. Rep. 40.
Rick CM. 1986. Reproductive isolation in the *Lycopersicon peruvianum* complex. In: D'Arcy WG, editor. Solanaceae Biology and Systematics. New York: Columbia University Press. p. 477-495.
Rick CM, Lamm R 1955. Biosystematic studies on the status of Lycopersicon chilense. American Journal of Botany 42: 663-675.
Rieseberg LH 1997. Hybrid origins of plant species. Annual Review of Ecology and Systematics 28: 359-389.
Rieseberg LH, et al. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301: 1211-1216.
Rieseberg LH, Willis JH 2007. Plant speciation. Science 317: 910-914.
Rieseberg LH, Wood TE, Baack EJ 2006. The nature of plant species. Nature 440: 524-527.
Rodriguez F, Wu FN, Ane C, Tanksley S, Spooner DM 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? BMC Evolutionary Biology 9: 191-207.
Roselius K, Stephan W, Staedler T 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics 171: 753-763.
Sarkinen T, Bohs L, Olmstead RG, Knapp S 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evolutionary Biology 13: 214.
Sato S, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635-641.
Schumer M, Rosenthal GG, Andolfatto P 2014. How common is homoploid hybrid speciation? Evolution 68: 1553-1560.
Soltis PS, Soltis DE. 2012. Polyploidy and genome evolution: Springer.
Staedler T, Arunyawat U, Stephan W 2008. Population genetics of speciation in two closely related wild tomatoes (Solanum section Lycopersicon). Genetics 178: 339-350.
Staedler T, Florez-Rueda AM, Paris M 2012. Testing for "snowballing" hybrid incompatibilities in Solanum: impact of ancestral polymorphism and divergence estimates. Molecular Biology and Evolution 29: 31-34.
Staedler T, Roselius K, Stephan W 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. Evolution 59: 1268-1279.
Stamatakis A 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.
Streicher JW, Schulte JA, Wiens JJ 2016. How Should Genes and Taxa be Sampled for Phylogenomic Analyses with Missing Data? An Empirical Study in Iguanian Lizards. Systematic Biology 65: 128-145.
Stuessy TF, Crawford DJ, Soltis DE, Soltis PS. 2014. Plant systematics: the origin, interpretation, and ordering of plant biodiversity.: Koeltz Scientific Books.
Taylor IB. 1986. Biosystematics of the tomato. In: Atherton JG, Rudich J, editors. The tomato crop: a scientific basis for improvement. London: Chapman and Hall. p. 1–34.

Ungerer MC, Baird SJE, Pan J, Rieseberg LH 1998. Rapid hybrid speciation in wild sunflowers. Proceedings of the National Academy of Sciences of the United States of America 95: 11757-11762.
Via S 2009. Natural selection in action during speciation. Proceedings of the National Academy of Sciences of the United States of America 106: 9939-9946.
Warnock SJ 1988. A review of taxonomy and phylogeny of the genus Lycopersicon. HortScience 23: 669-673.
Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS 2010. Patterns of hybridization in plants. Perspectives in Plant Ecology Evolution and Systematics 12: 175-182.
Wilf P, Carvalho MR, Gandolfo MA, Ruben Cuneo N 2017. Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. Science 355: 71-74.
Xia H, Camus-Kulandaivelu L, Stephan W, Tellier A, Zhang ZW 2010. Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. Molecular Ecology 19: 4144-4154.
Zuriaga E, Blanca J, Nuez F 2009. Classification and phylogenetic relationships in Solanum section Lycopersicon based on AFLP and two nuclear gene sequences. Genetic Resources and Crop Evolution 56: 663-678.

## 3.8. Figures

**Figure 1** Diversity of *S. peruvianum* (*S.per*) and *Solanum chilense* (*S.chi*). (**a**) Differences in flower morphology between *S.per* populations including one *S. corneliomulleri* (*S.cor*). Note that the extended stigma and large yellow petals are indicative of outcrossing. (**b**) Differences in leaf morphology within and between *S.per*, *S.chi*, and *S.cor*. (**c**) Collection locations of populations sampled in this study. Horizontal and vertical lines indicate the distribution of *S.per* and *S.chi* respectively.

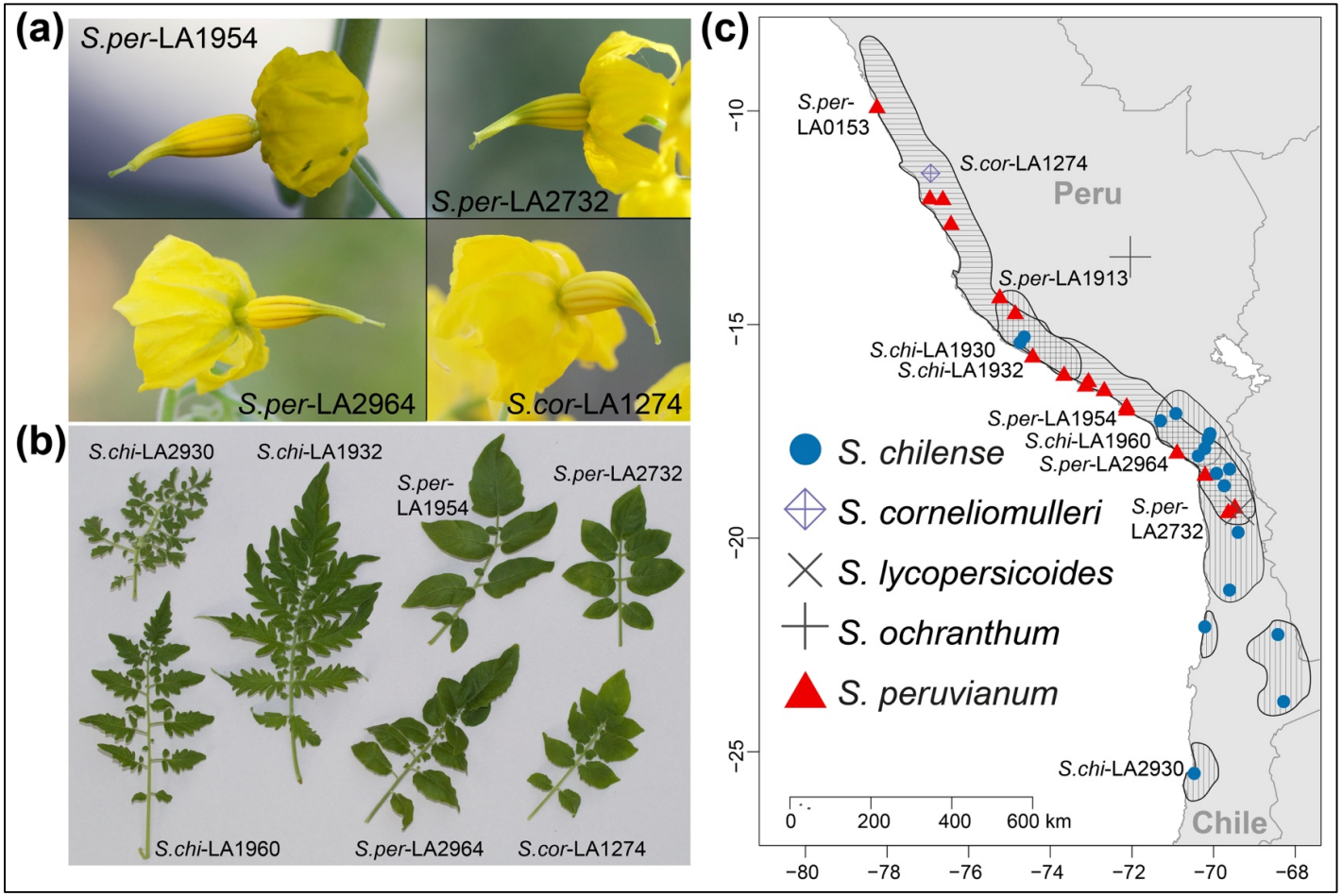


**Figure 2** Phylogeny of sect. *Lycopersicon*. (**a**) A maximum-likelihood phylogeny from all accessions excluding those with >10% admixture according to STRUCTURE. The species groups are delineated by black lines and labelled. Node labels give bootstrap support. (**b**) Coalescent phylogeny with all accessions excluding the early diverging 'Hirsutum' group. Taxon abbreviations: *S. arcanum* (*S.arc*), *S. chmielewskii* (*S.chm*), *S. corneliomulleri* (*S.cor*), *S. chilense* (*S.chi*), *S. galapagense* (*S.gal*), *S. huaylasense* (*S.hua*), *S. habrochaites* (*S.hab*), *S. lycopersicoides* (*S.lyc*), *S. neorickii* (*S.neo*), *S. ochranthum* (*S.och*), *S. pennellii* (*S.pen*), *S. peruvianum* (*S.per*), *S. pimpinellifolium* (*S.pim*).

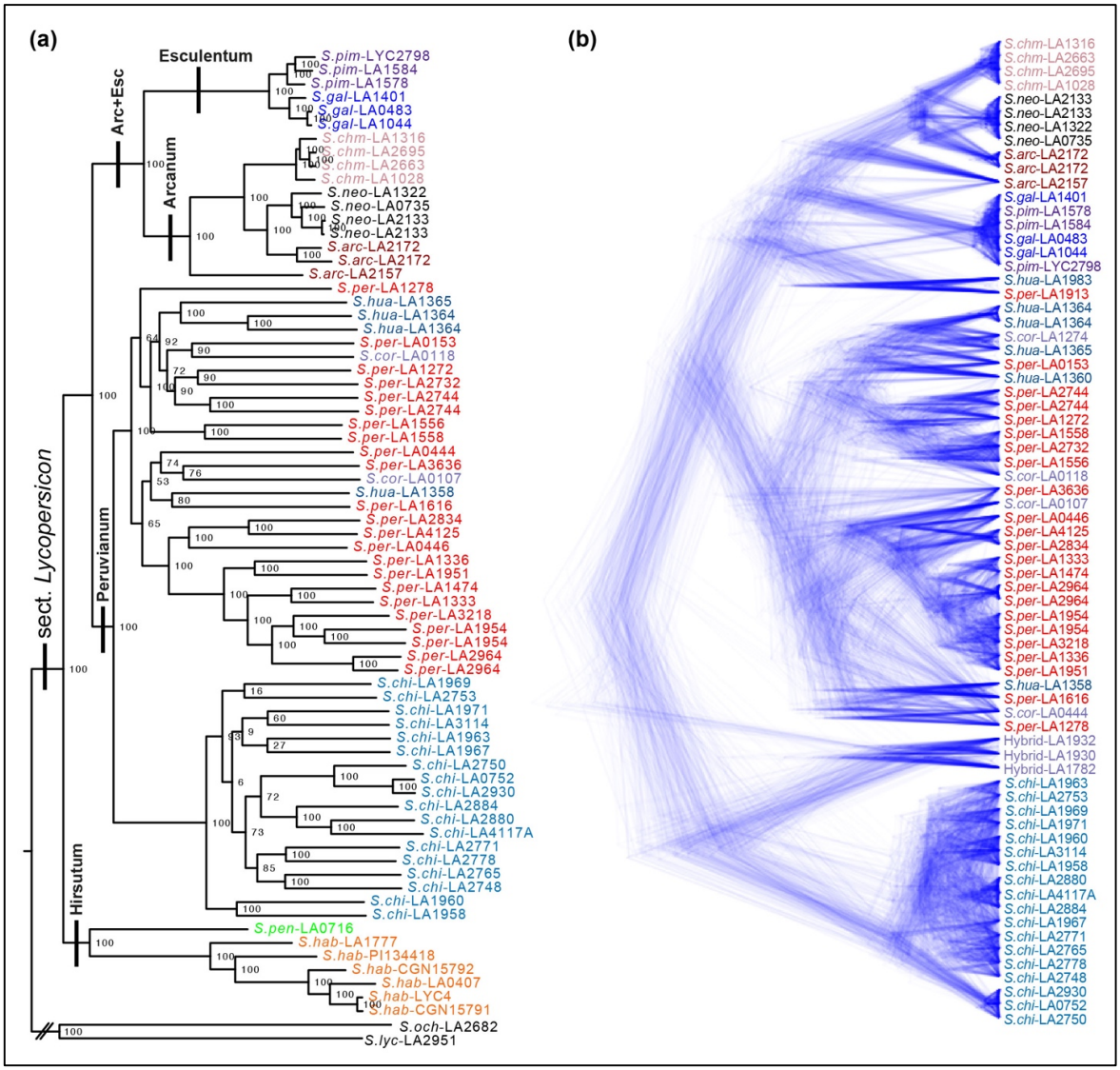


**Figure 3** STRUCTURE analysis of all sect. *Lycopersicon* accessions excluding the 'Hirsutum' group. The most likely number of clusters (K) was 3, but data for K=4 and K=5 is also shown. The subdivision of *S. peruvianum* is noticeable at K>4. Taxon abbreviations are given in Fig. 2.

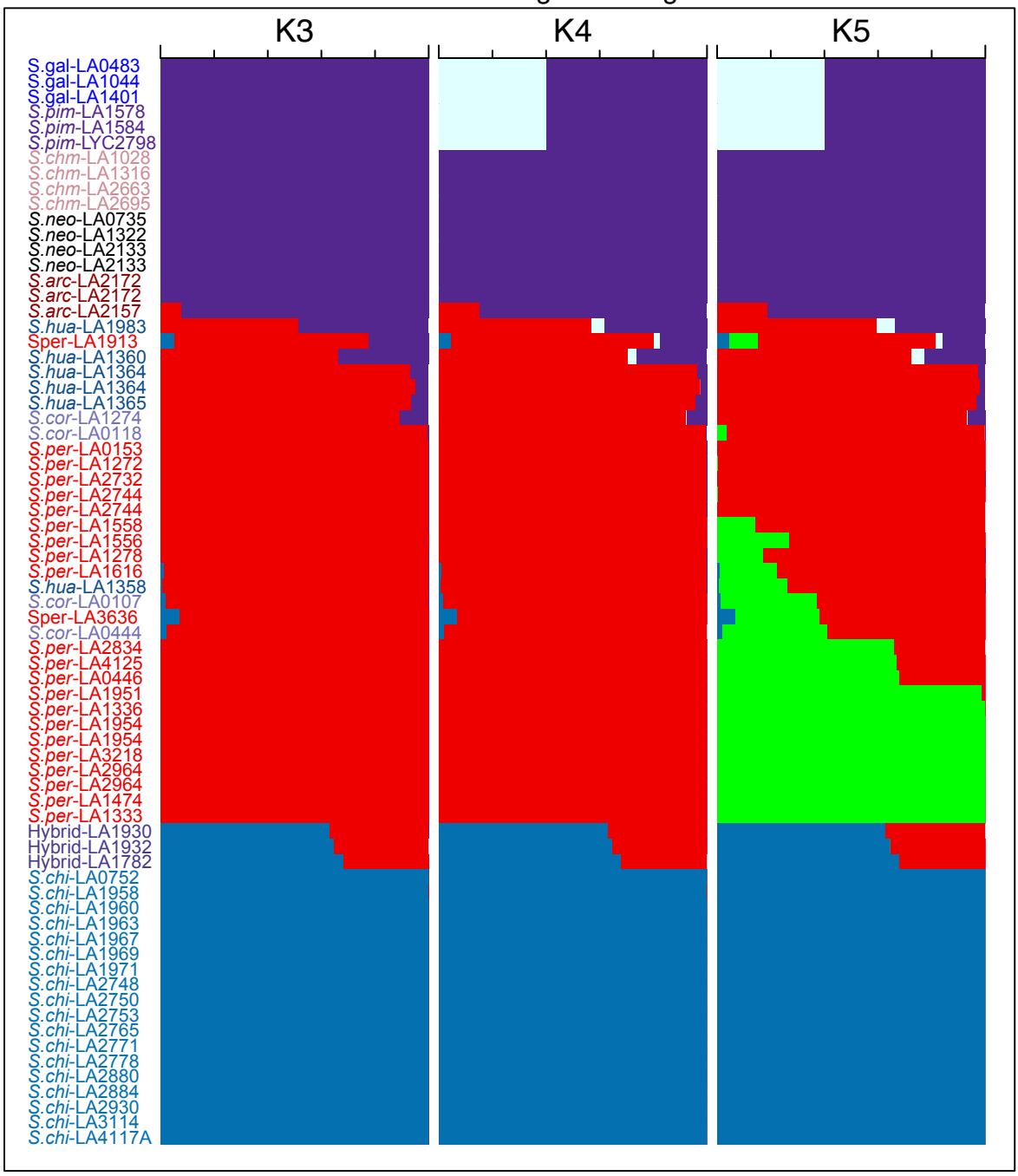


**Figure 4** A pie chart of STRUCTURE groups based on K=5 for all individuals shown in Fig. 3 at their collection location. The coastal subpopulation of *S. peruvianum* is circled. Two highly admixed accessions, *S.per*-LA1913 and *S. huaylasense* (*S.hua*) LA1983 as well as the anomalous *S. chilense* (*S.chi*) accession LA0752 and the cryptic hybrid populations LA1930 and LA1932 are marked.

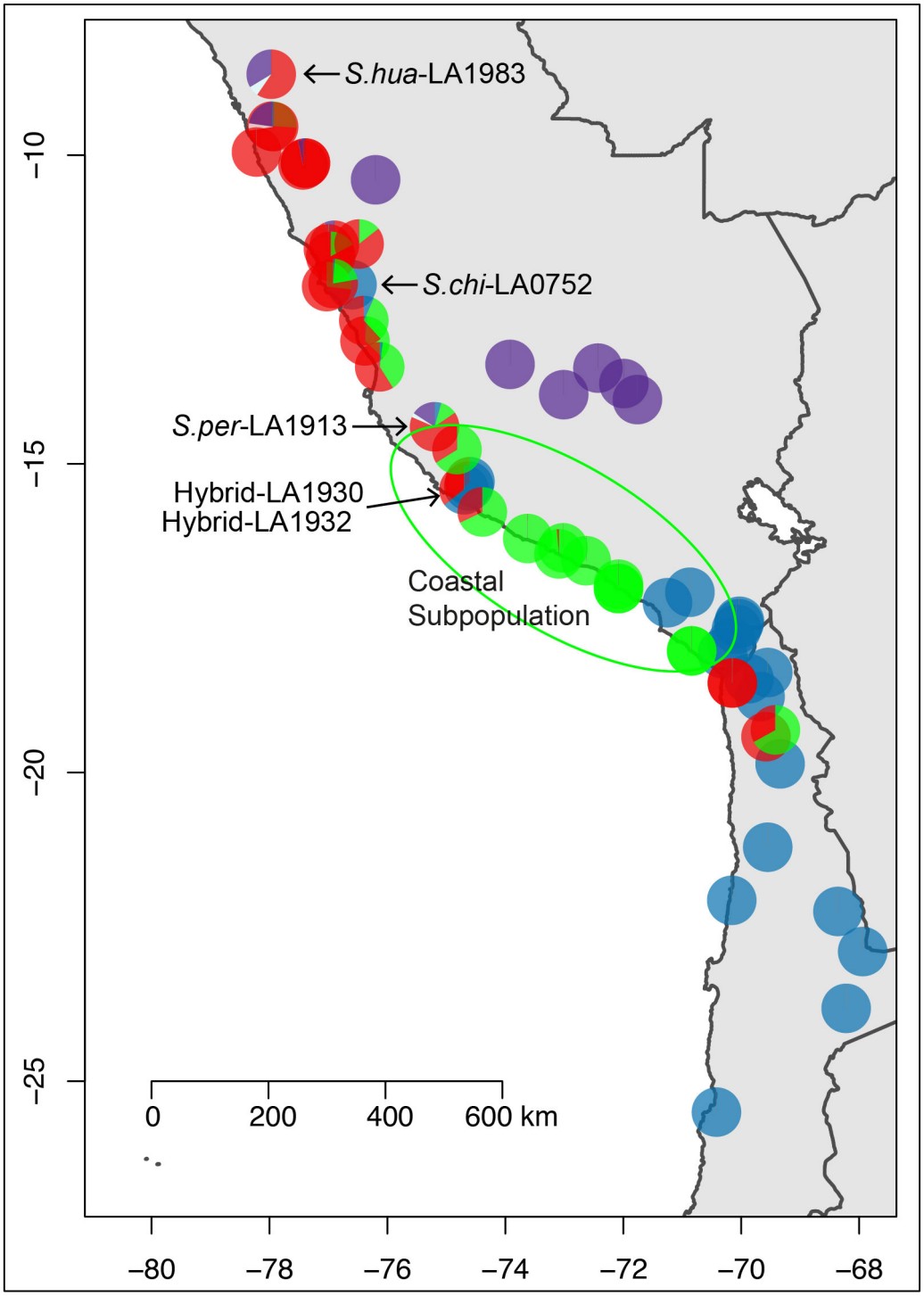


**Figure 5** Reticulate network based on SplitsTree4. The reticulate network shows reticulation within the 'Peruvianum' group species and between *S. peruvianum* and *S. chilense*, including reticulation of the hybrid populations. Species abbreviations are given in Fig. 2.

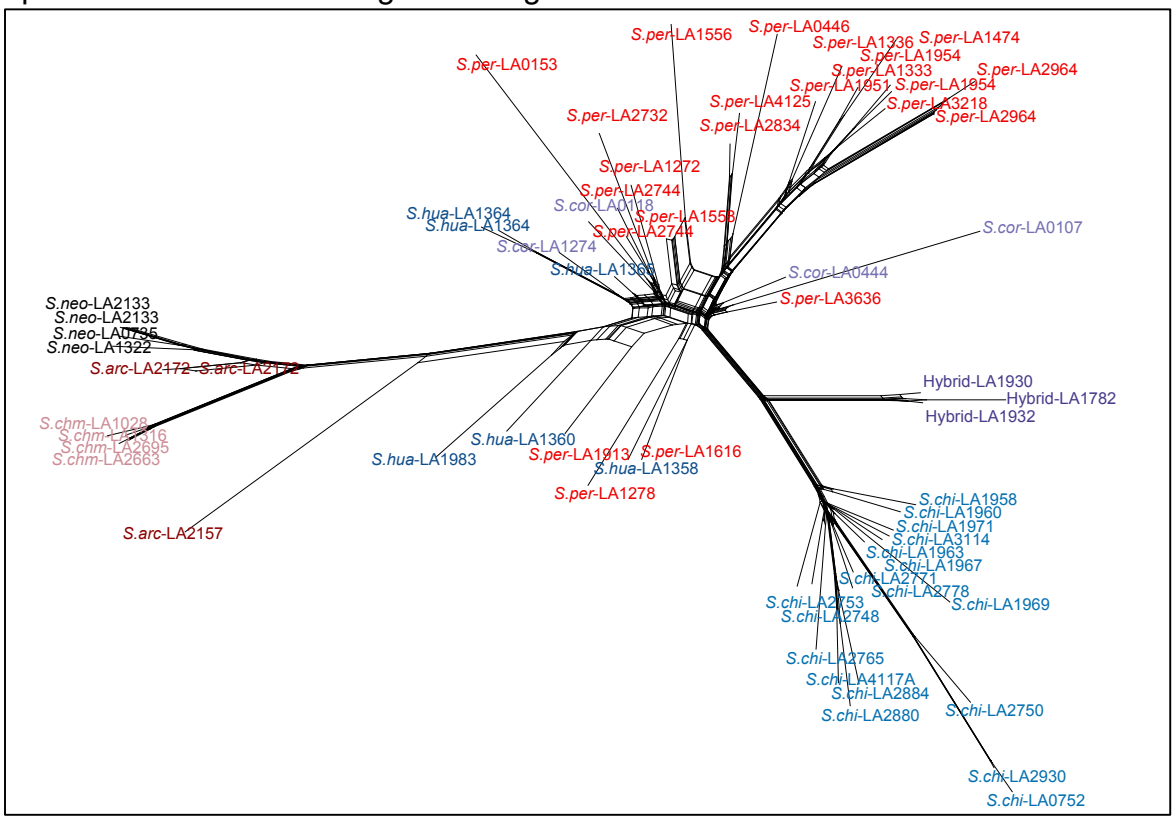
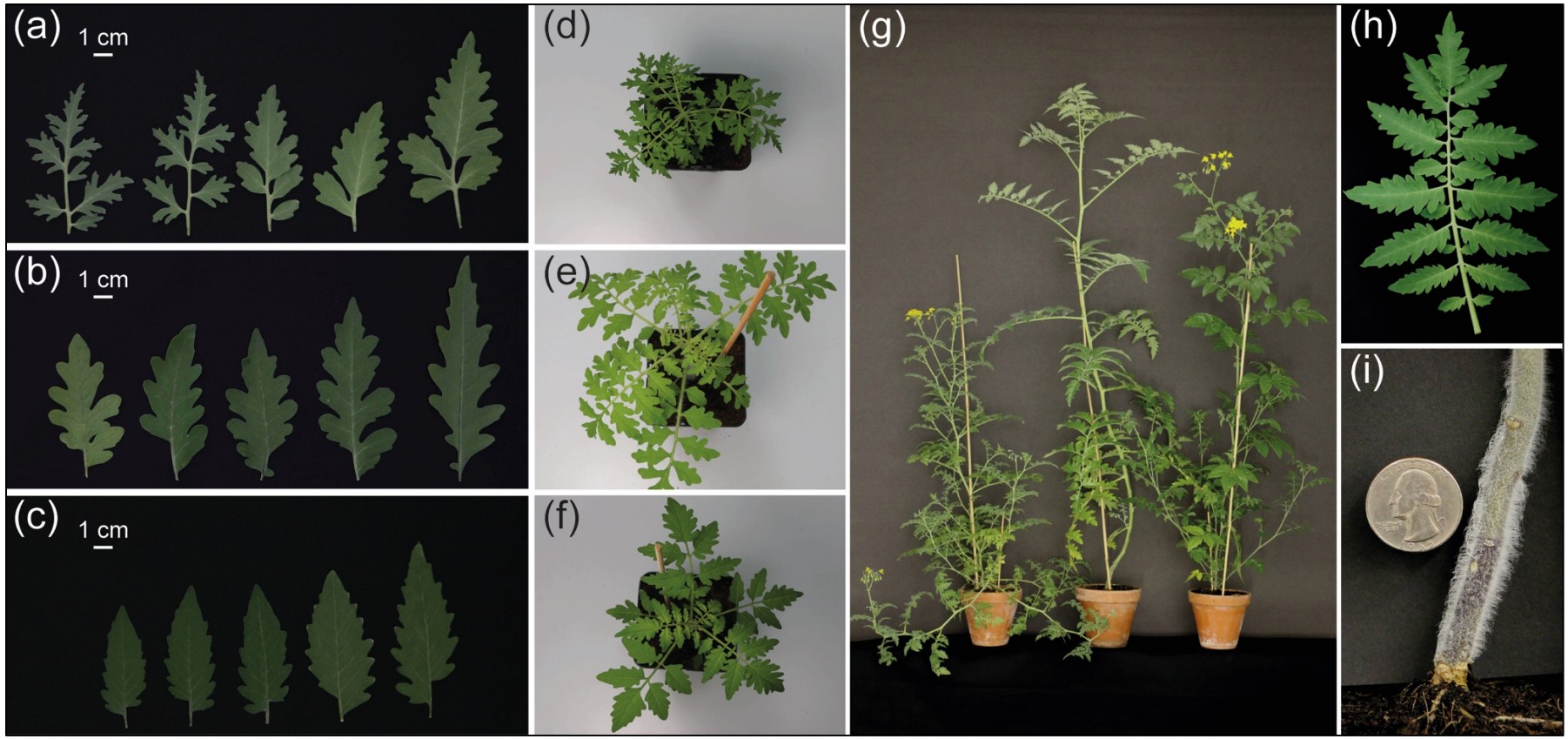


**Figure 6** Phenotype of the hybrid accessions. (**a-c**) Leaflet diversity across individuals from a single population of (a) *S. chilense* LA2930, (b) Hybrid-LA1930, and (c) *S. peruvianum* LA1954. (**d-f**) Typical 3-week-old (d) *S. chilense* LA2930, (e) Hybrid-LA1932, and (f) *S. peruvianum* LA1954. (**g**) 10-week *S. chilense* LA2930 (left), Hybrid-LA1932 (center), and *S. peruvianum* LA1954 (right). All plants were germinated and grown together. (**h**) Typical leaf of Hybrid-LA1932 at 10 weeks. (**i**) Typical Hybrid-LA1932 stem at 10 weeks showing the thickness and high density of type I trichomes. In all cases, the phenotype of Hybrid-LA1930 was indistinguishable from Hybrid-LA1932.



## 3.9. Tables

**Table 1.** The mean number of seeds and seed-like structures (SLS) per fruit for all cross combinations. Germination tests on seeds were carried out following TGRC guidelines (http://tgrc.ucdavis.edu/seed_germ.aspx).

| *Cross* | *Mean seed number per fruit* | *Mean SLS number per fruit* | *Germination rate* | *Offspring relative to within deme S.per × S.per crosses* |
|---|---|---|---|---|
| *S.per* × *S.per* - within deme | 44.6 | 7.5 | 72.4% | 100% |
| *S.per* × *S.per* - between deme | 33.2 | 2.9 | 74.6% | 77% |
| *S.cor* × *S.per* | 16.7 | 8 | 35% | 18% |
| *S.cor* × *S.chi* | 0 | 10.9 | n/a | 0% |
| *S.chi* × *S.per* | 0.82 | 17.7 | 100% | 3%[†] |
| *S.chi* × Hybrid | 0.92 | 17.2 | 0% | 0% |
| *S.per* × Hybrid | 0.49 | 23.5 | 50% | 1% |

[†]all of these individuals died before reaching maturity

## 3.10. Supplemental Methods

### 3.10.1. Ploidy (C-value) determination by flow cytometry

To determine if two cryptic hybrid individuals (LA1930, LA1932) were allopolyploids, they, two *S.per* (LA2732, LA1954), *S.chi*-LA2930, and two diploid *S. lycopersicum* controls were analyzed on the FACS Aria II Flow Cytometer. Leaf tissue from one individual from these accessions was chopped in 500 μL of 15 mM PBS buffer containing 2 mM EDTA, 15 mM 2-Mercaptoethanol, 0.1% Triton X-100, 0.5 mM Spermine, 100 μg/mL RNase, and 5 μg/mL propidium iodide. The nuclei suspension was then run through a 50 μm filter and the filtered nuclei were then immediately analyzed. A minimum of 10,000 nuclei from each accession were analyzed. Samples were normalized to the diploid *S. lycopersicum* controls. The flow cytometry data for all samples is shown in Fig. S12.

### 3.10.2. ∂a∂i Analysis

Our analysis using ∂a∂i highlighted two informative problems. First, the number of singleton SNPs appeared to be underestimated, perhaps due to low sequencing coverage of these rare variants (Nielsen et al. 2011). We addressed this by masking singletons in the JSFS before running ∂a∂i. Second, according to ∂a∂i, too many high-frequency shared derived alleles were detected. This could result from ancestral state misidentification, and a misidentification rate of 1.1 – 1.5% in the outgroup could account for these high-frequency shared derived alleles. To address this potential problem, an additional outgroup misidentification rate parameter was added to the model.

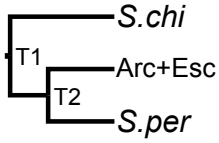
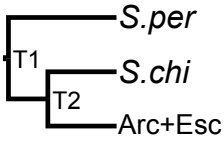
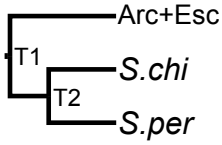
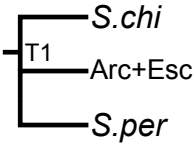
According to the branching order in the coalescent analysis and the ML analysis using all data, the divergence of *S. chilense* predates the divergence between *S. peruvianum* and the Arc+Esc lineage. This is surprising and goes against what is known from morphology and the literature. However, a large degree of uncertainty on the branching order of these lineages is evident (for example, if taxa

with >10% admixture are removed), and it is also difficult distinguish these patterns from a concurrent divergence of all three lineages (a true polytomy).

Therefore, we wanted to directly test four alternative evolutionary scenarios to help determine the branching order of *S. chilense, S. peruvianum,* and Arc+Esc. For this we used ∂a∂i because it is able to estimate demographic parameters from a three-dimensional site frequency spectrum (Gutenkunst et al. 2009). The four alternative hypotheses are: (Arc+Esc, (*S. chilense*, *S. peruvianum*)), (*S. chilense*, (Arc+Esc, *S. peruvianum*)), (*S. peruvianum*, (*S. chilense*, Arc+Esc)), or simultaneous divergence (polytomy). The model written for this purpose had two ancestral splits (T1 and T2) which were both followed by growth in all populations. The time interval of T1 was restricted to 0 to estimate a simultaneous speciation event.

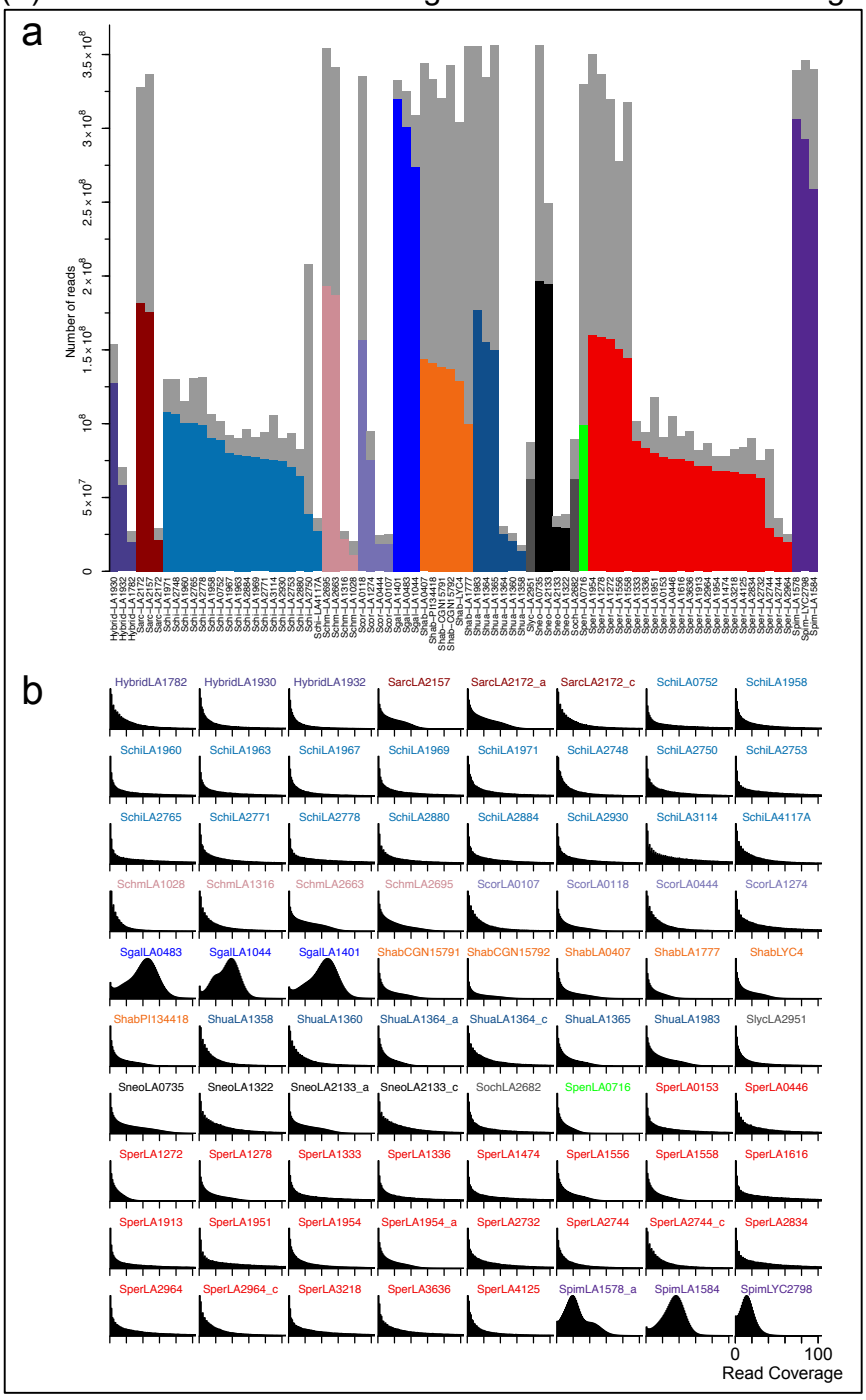
Unfortunately, the multispecies Arc+Esc lineage did not have an appropriate site frequency spectrum for demographic analysis in ∂a∂i. As an alternative we downloaded 17 *S. pimpinellifolium* ('Esculentum group) accessions reported by Lin et al. 2014 (NCBI SRA SRP045767). These are described in Supplemental Table S2. All of these *S. pimpinellifolium* accessions were then used to represent the Arc+Esc group. After excluding the admixed individuals identified in STRUCTURE, the three-dimensional JSFS was derived from 92,939 synonymous variants with a nonzero MAF. Ten independent runs of 10,000 iterations and 10 bootstraps were computed for each scenario and the scenario with the highest composite log-likelihood was selected. The model with the highest composite likelihood was (Arc+Esc), (*S. chilense*, *S. peruvianum*)), supporting *S. chilense* and *S. peruvianum* as sister species:

The four alternative demographic scenarios tested in ∂a∂i and the maximum log-likelihood returned from 10 independent runs from randomized starting parameters.

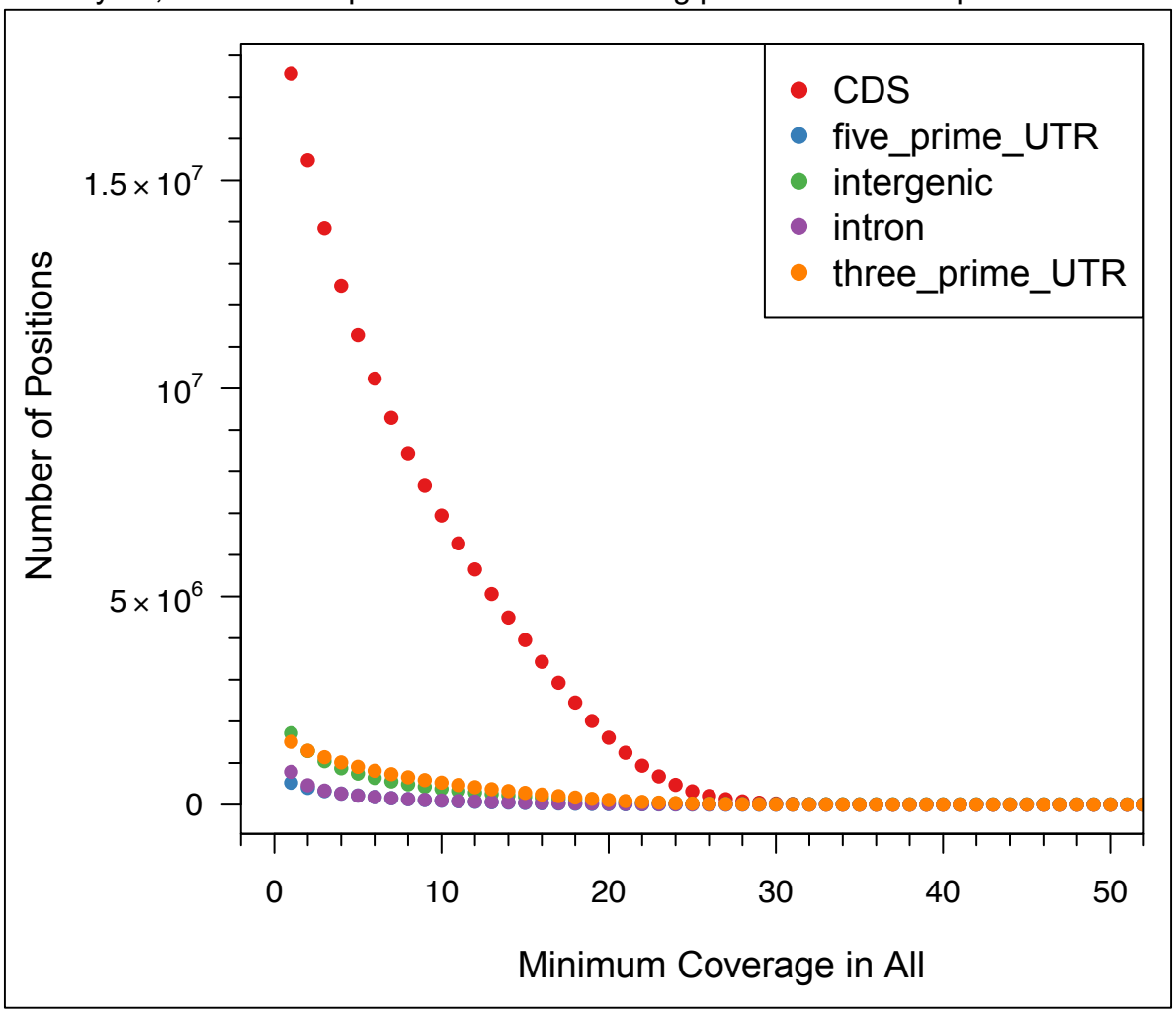
| Hypothesis | *S. chilense* diverges first | *S. peruvianum* diverges first | Arc+Esc diverges first | Simultaneous divergence |
|---|---|---|---|---|
| Demographic scenario |  |  |  |  |
| Maximum log-likelihood | -20,538 | -23,119 | **-20,059** | -22,026 |

## 3.11. Supplemental Figures

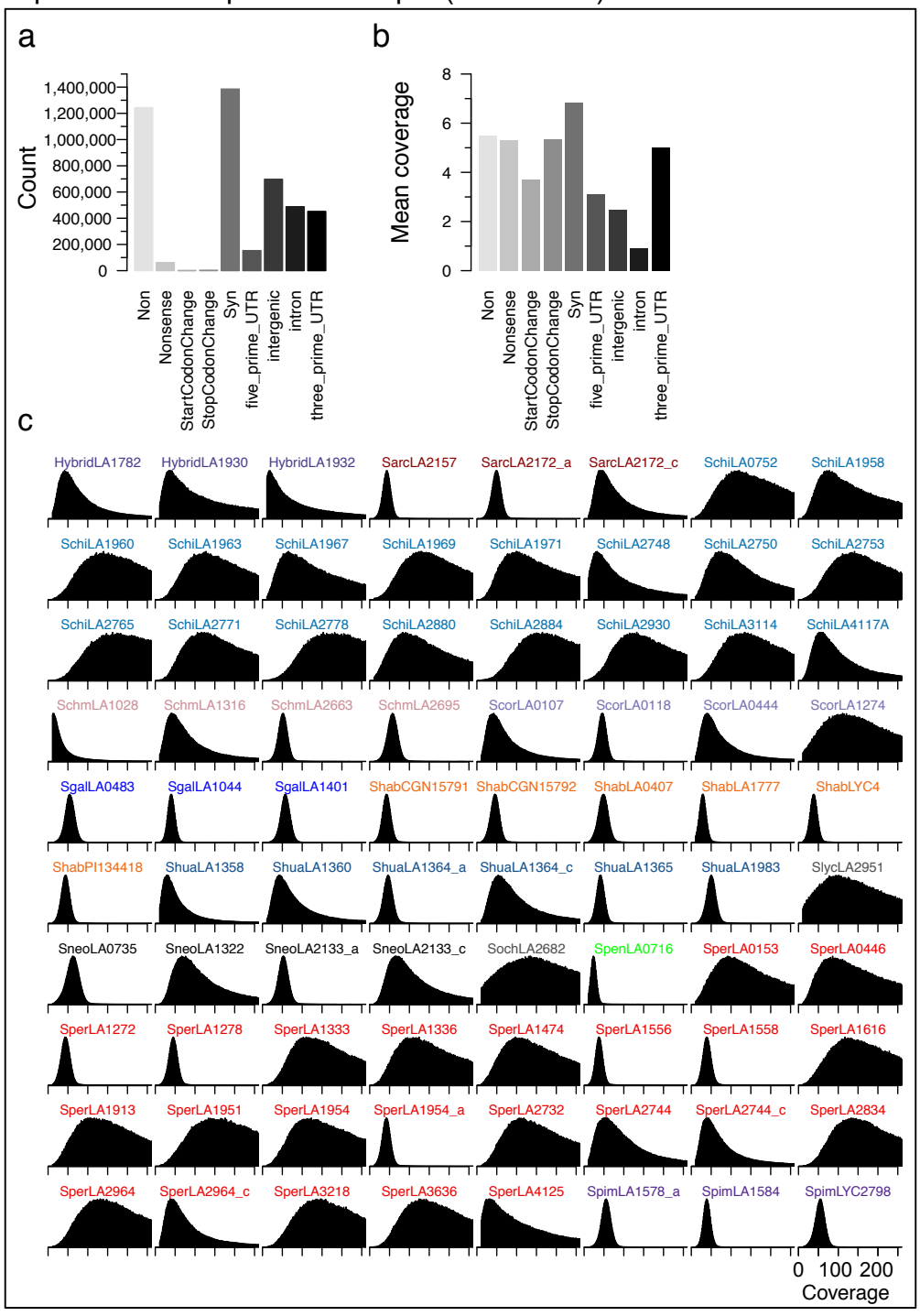
**Supplemental Figure 1** Read mapping data by individual. (**a**) Number of reads mapping to the reference genome sequence for the 80 wild tomato accessions. The colored segment represents the proportion of reads uniquely aligning to the genome and the gray segment the proportion of reads non-uniquely aligned or not aligned. Mapping data for these and the additional accessions is in Supplemental Table S1. (**b**) Distribution of read coverage in the 80 libraries following mapping.



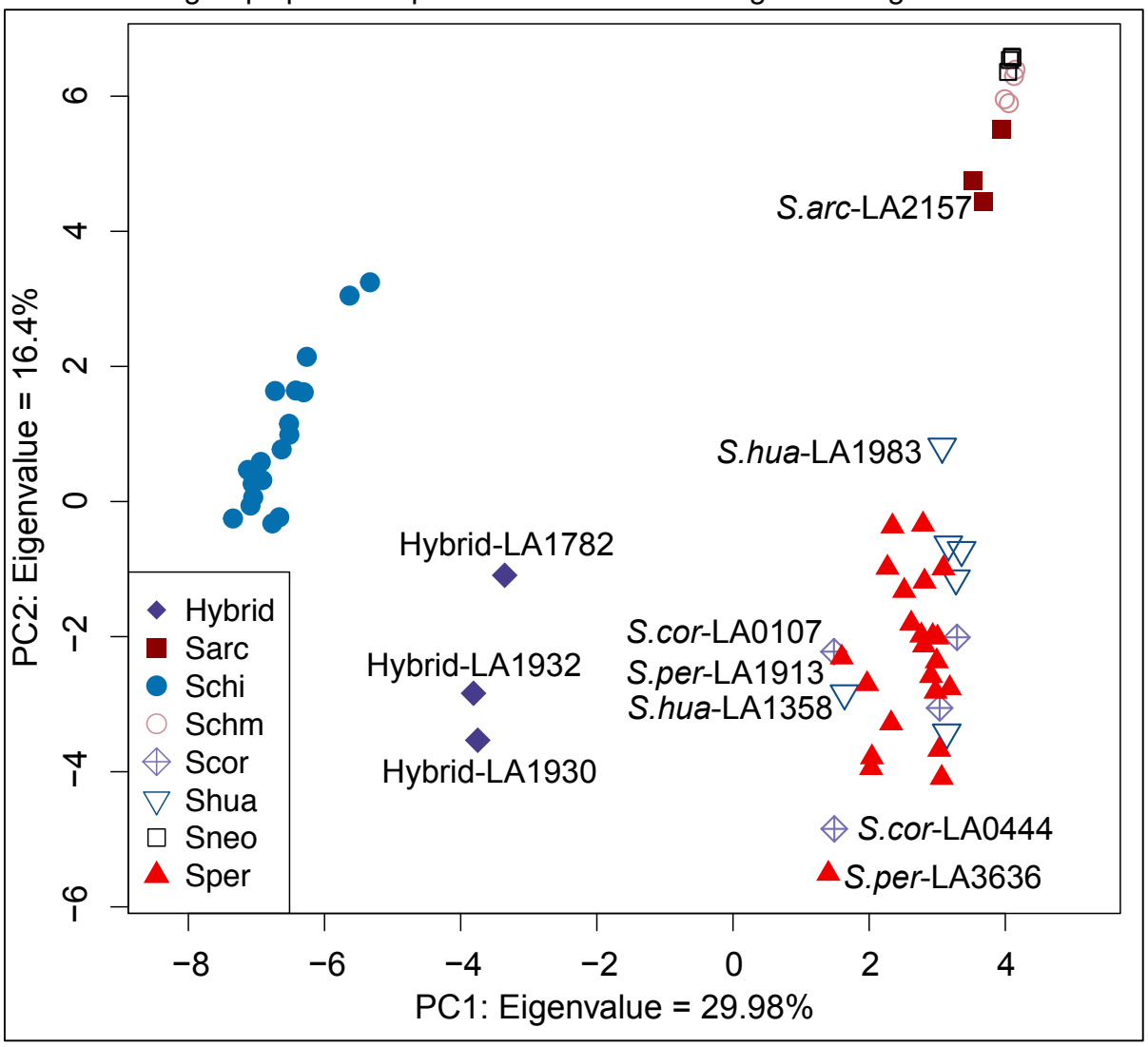
**Supplemental Figure 2** Number of positions (y-axis) sequenced to a minimum coverage (x-axis) in all 80 individuals. The number of positions is divided into categories following the reference GFF file. Note that the number of positions sequenced to a minimum depth also includes all positions sequenced to a greater depth of coverage. For example, 527,655 positions were sequenced to ≥10 coverage but only 56,433 had a depth of 10. The remaining positions had a depth >10.



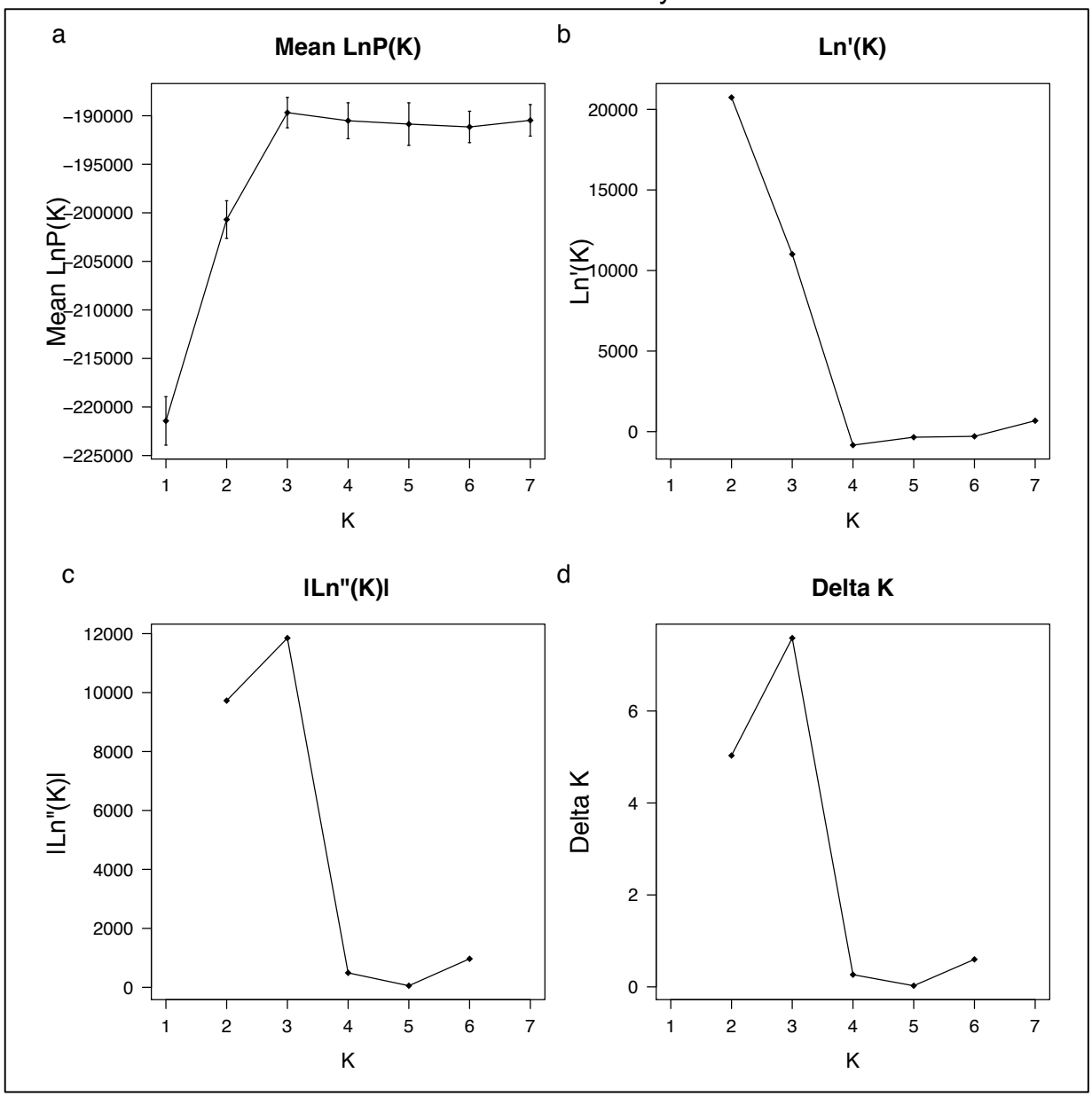
**Supplemental Figure 3** Annotation and coverage in the high-quality variant dataset. (**a**) Number of variants assigned as 5'UTR, CDS (synonymous + nonsynonymous), intron, 3'UTR, and intergenic. The number of positions which change a start or stop codon and nonsense mutations are also given (**b**) Median depth of coverage at variable sites for the classes in (a). (**c**) The distribution of read depth at variable positions. Depth (from 0-250) is on the x-axis.



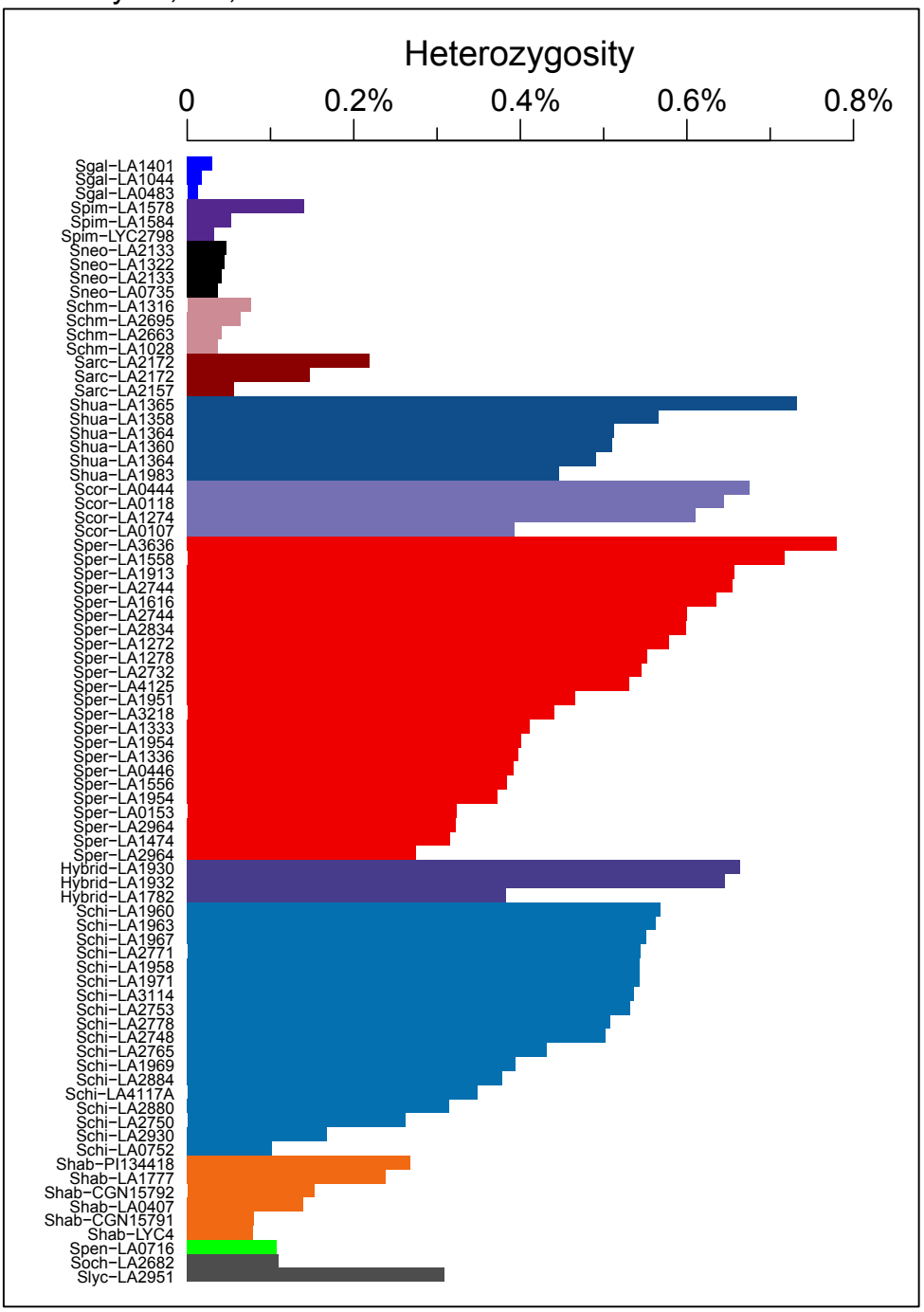
**Supplemental Figure 4** Principle component analysis of all accessions excluding the 'Hirsutum' and 'Lycopersicon' groups. The first principle component explained 29.98% of the variance and delinated *S. chilense* (*S.chi*) from the 'Arcanum' and 'Peruvianum' group species. Species abbreviations are given in Fig. 2.



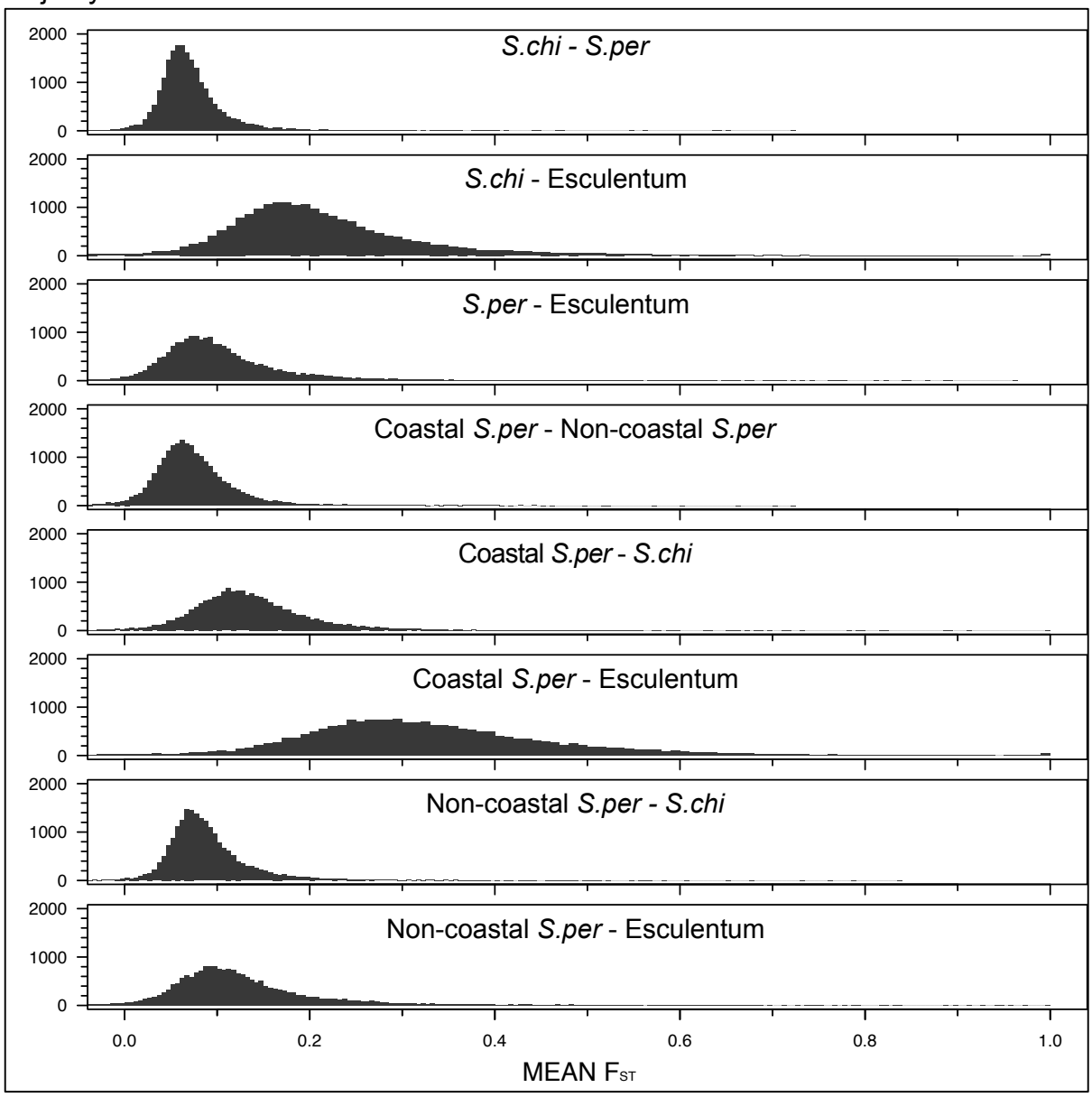
**Supplemental Figure 5** Log-likelihood of the data from STRUCTURE. (**a**) The average log likelihood ± s.d. over 10 runs (y-axis) modelled for K=1 – K=7 (x-axis). (**b**) The average difference between successive likelihood values of K, L'(K) = L(K) – L(K-1). (**c**) The second order rate change of L'(K), |L''(K)| = |L'(K+1) – L'(K)|. (**d**) The ad hoc ΔK statistic which is the average second order rate change (|L''(K)|) divided by the standard deviation of L(K) [see Evanno *et al*. (2005) for full details]. The log likelihood and ΔK both indicate that K=3 is most likely.



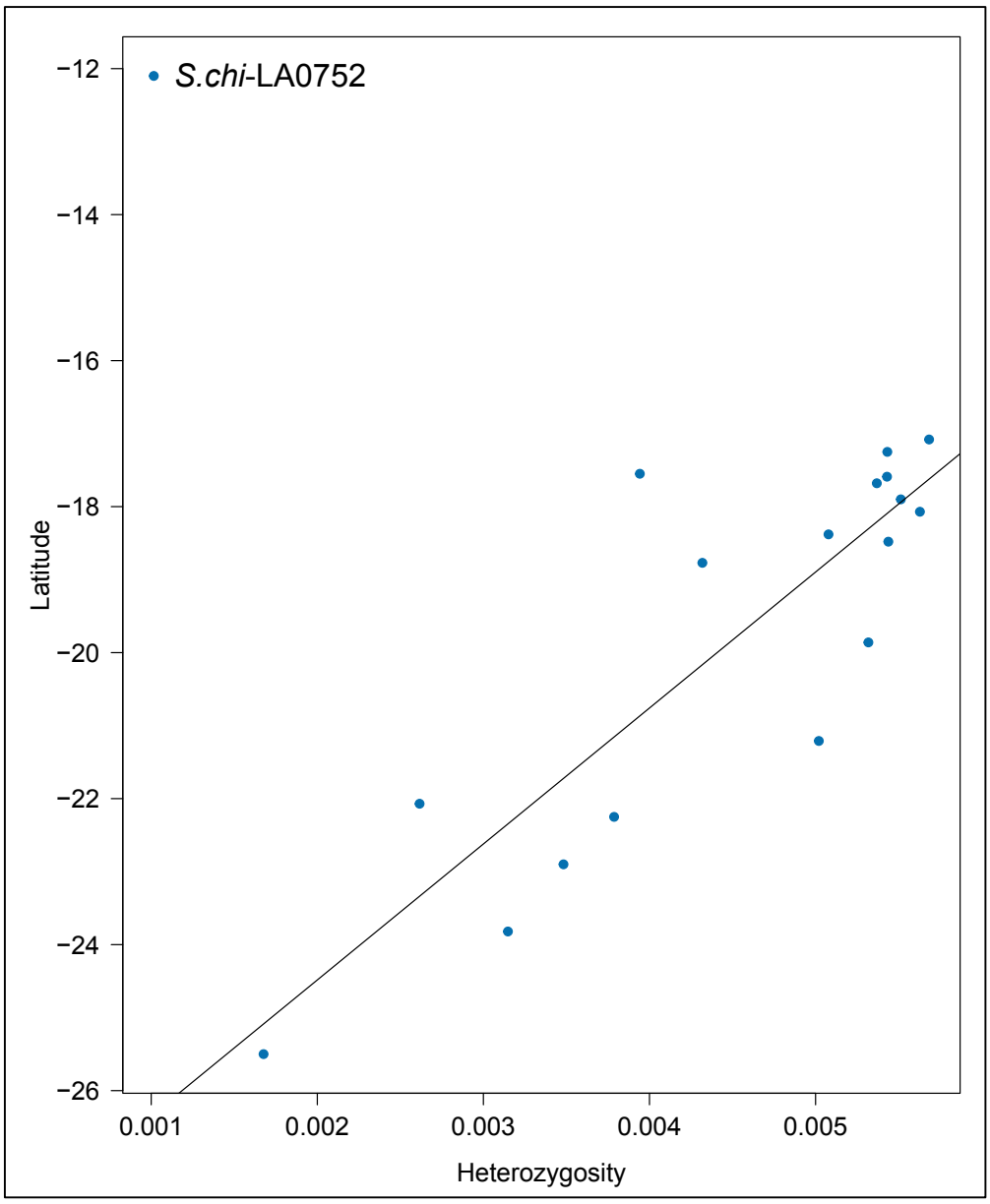
**Supplemental Figure 6** Heterozygosity per accession. Individual heterozygosity was calculated for each accession by dividing the number of heterozygous sites by the number of positions. For example, from the 4,866,729 bi-allelic polymorphisms 154,538 had ≥10 coverage and were heterozygous in *S.chi*-LA1958. These bi-alleleic polymorphisms were drawn from the 34,617,016 positions with ≥10 coverage with ≤50% missing data. However, only 28,448,837 of the 34,617,016 positions had ≥10 coverage in *S.chi*-LA1958. Thus, the heterozygosity for this accession is 154,538 divided by 28,448,837.



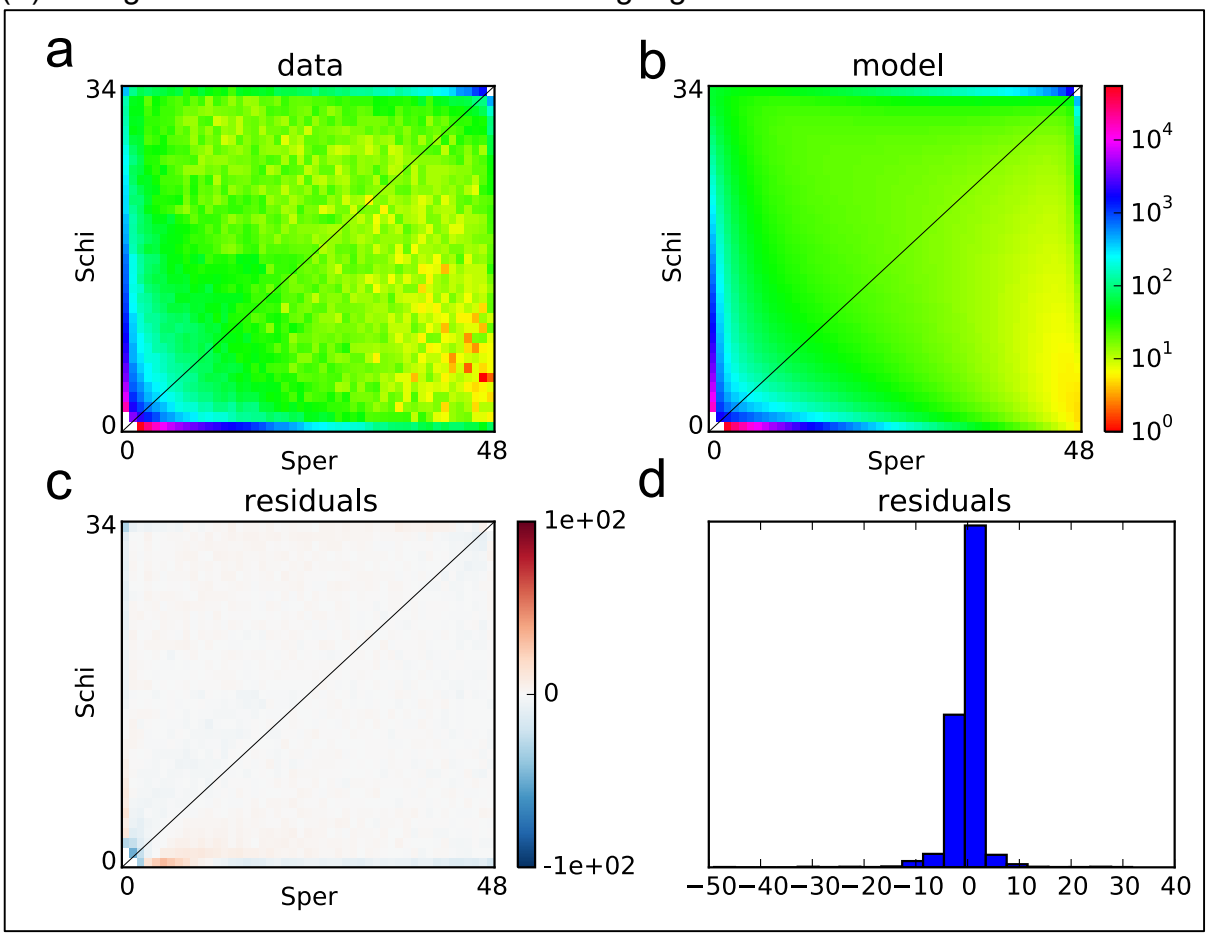
**Supplemental Figure 7** $F_{ST}$ between species and subpopulations of *S. peruvianum*. The mean $F_{ST}$ was calculated using VCFTools for all genes in pairwise comparisons between all populations shown. The species/accessions included in the Arcanum and Esculentum species groups is indicated in Fig. 2. For the *S.per* subpopulations were assigned based on the STRUCTURE result (Fig. 3) with a majority rule.



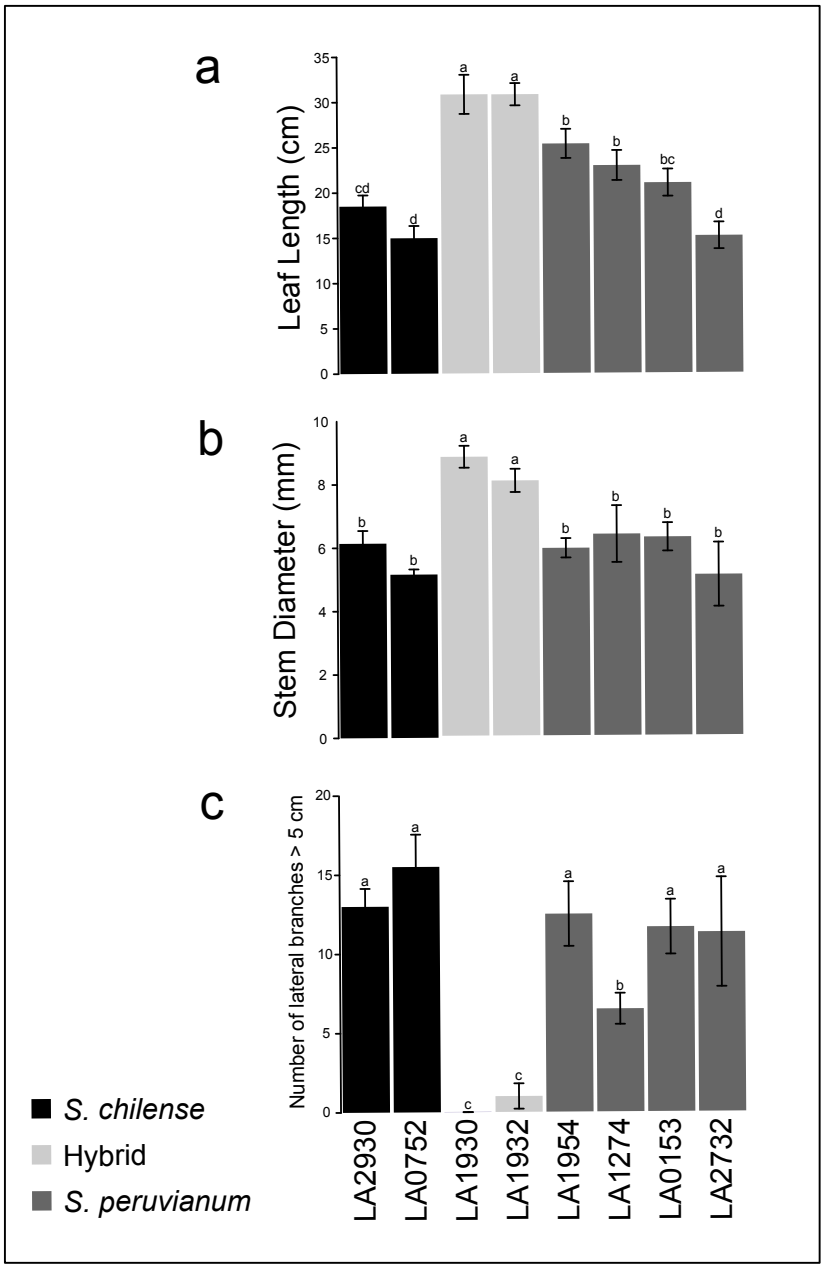
**Supplemental Figure 8** The relationships of heterozygosity (x-axis) and latitude (y-axis) of *S. chilense* accessions. The anomolous *S. chilense* accession LA0752 is labelled.



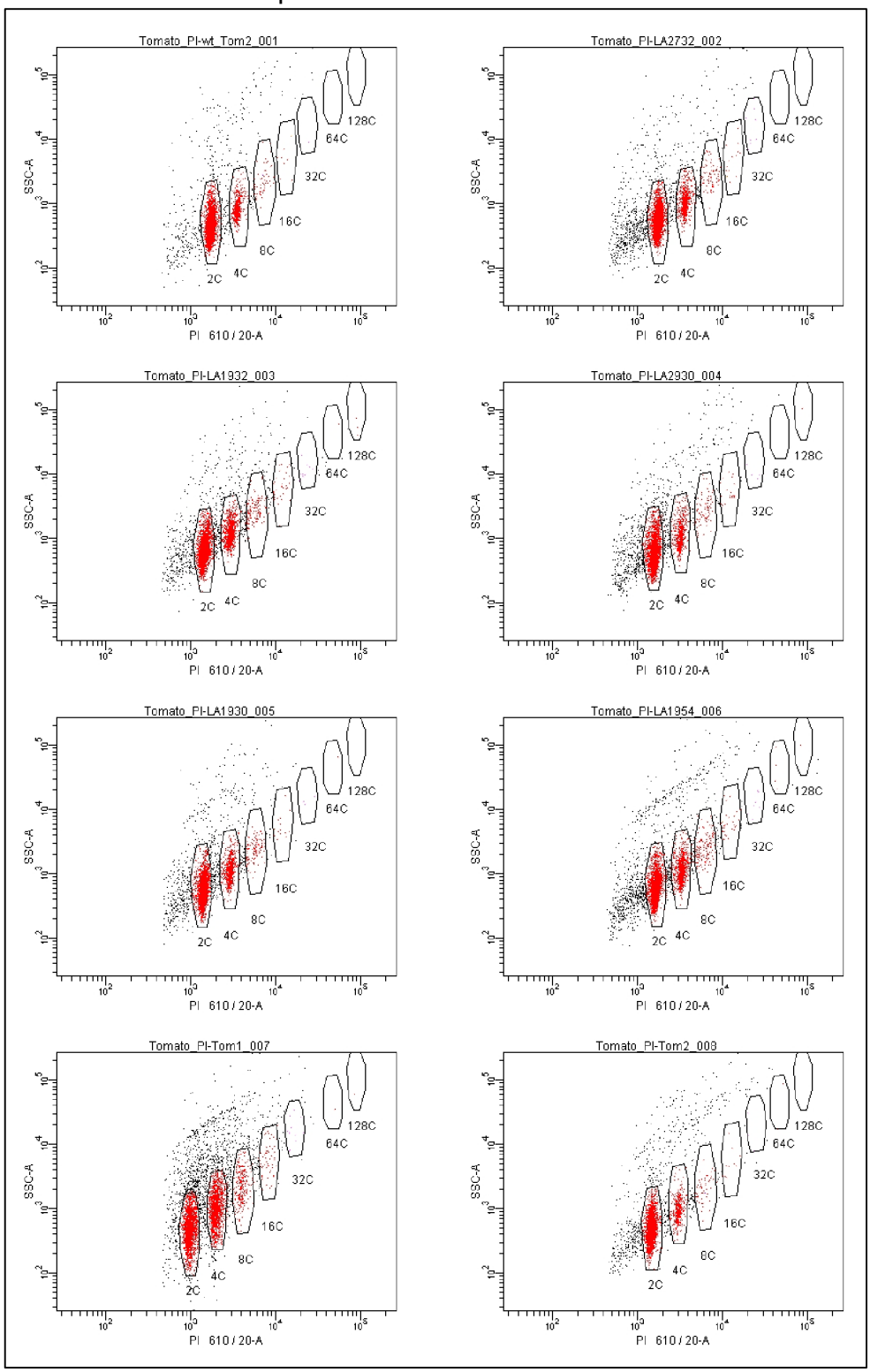
**Supplemental Figure 9** Fit of the *S. chilense* and *S. peruvianum* speciation event to the model in ∂a∂i. (**a**) The synonymous joint site frequency spectrum (JSFS) of *S. chilense* and *S. peruvianum*. (**b**) the modeled JSFS. (**c**) The residuals (data - model). (**d**) Histogram of the residual values showing a good model fit.



**Supplemental Figure 10** Distinct phenotypes of hybrid populations LA1930 and LA1932. Approximately four seeds from LA1930, LA1932, *S.chi*-LA2930, *S.chi*-LA0572, three *S.per*, and one *S.cor* accession were germinated and grown together in a greenhouse in Düsseldorf, Germany (29 plants total). Measurements were taken after three months for all individuals and grouped by accession. Bars with no shared letters indicates significant difference between groups (pairwise t-test, $P < 0.01$) and error bars show the s.e.m. Measurements for (**a**) Leaf length and (**b**) stem diameter were taken at nodes 1 through 10 for 3-5 individuals per accession. (**c**) The number of lateral branches >5 cm.

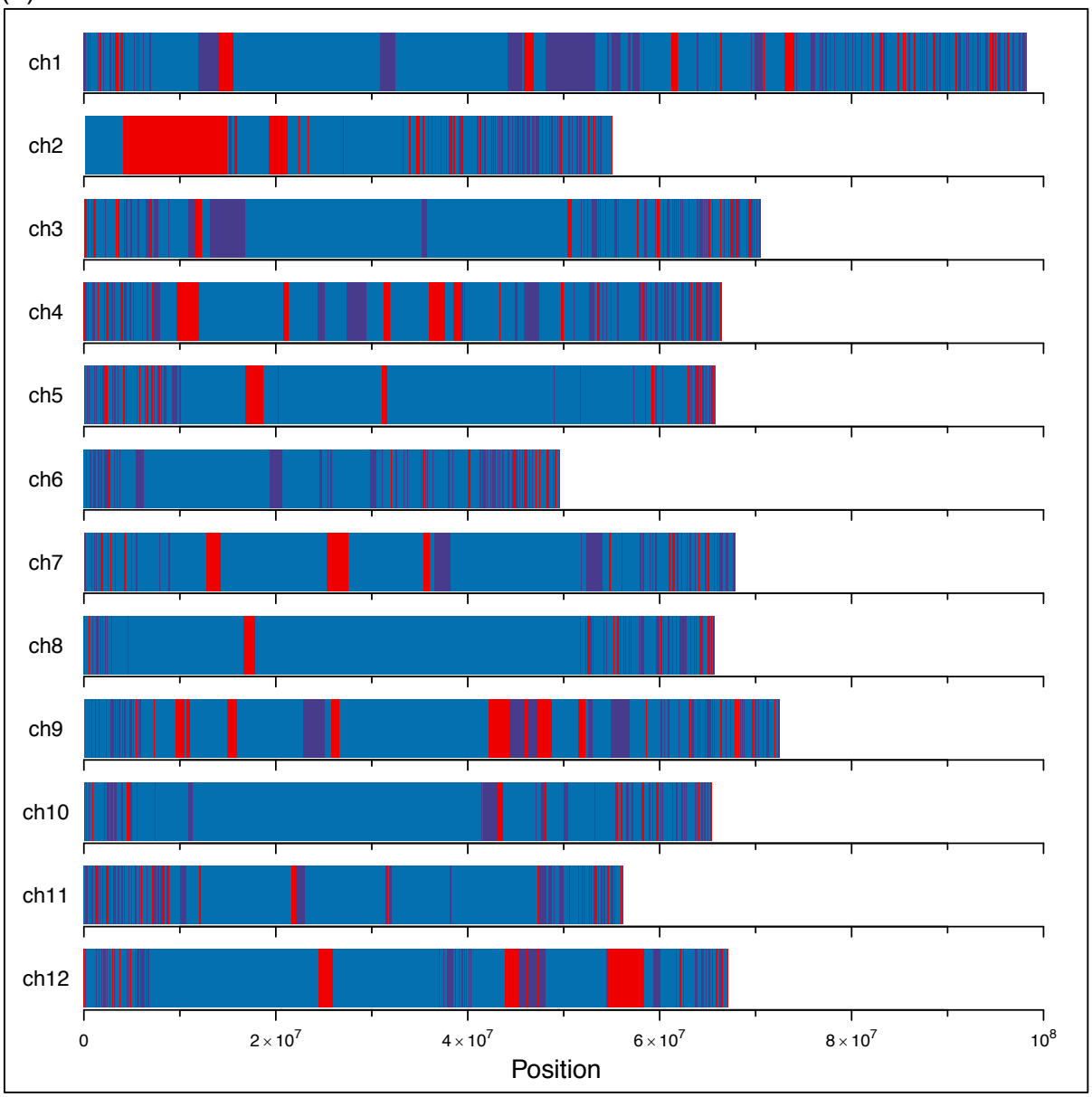


**Supplemental Figure 11** Flow cytometry results. The two-dimensional plots show side scattered light (y-axis) and propidium idodide intensity (x-axis). The ploidy level gates were determined from the first wt tomato control and confirmed again with the same and a second diploid cultivated control.

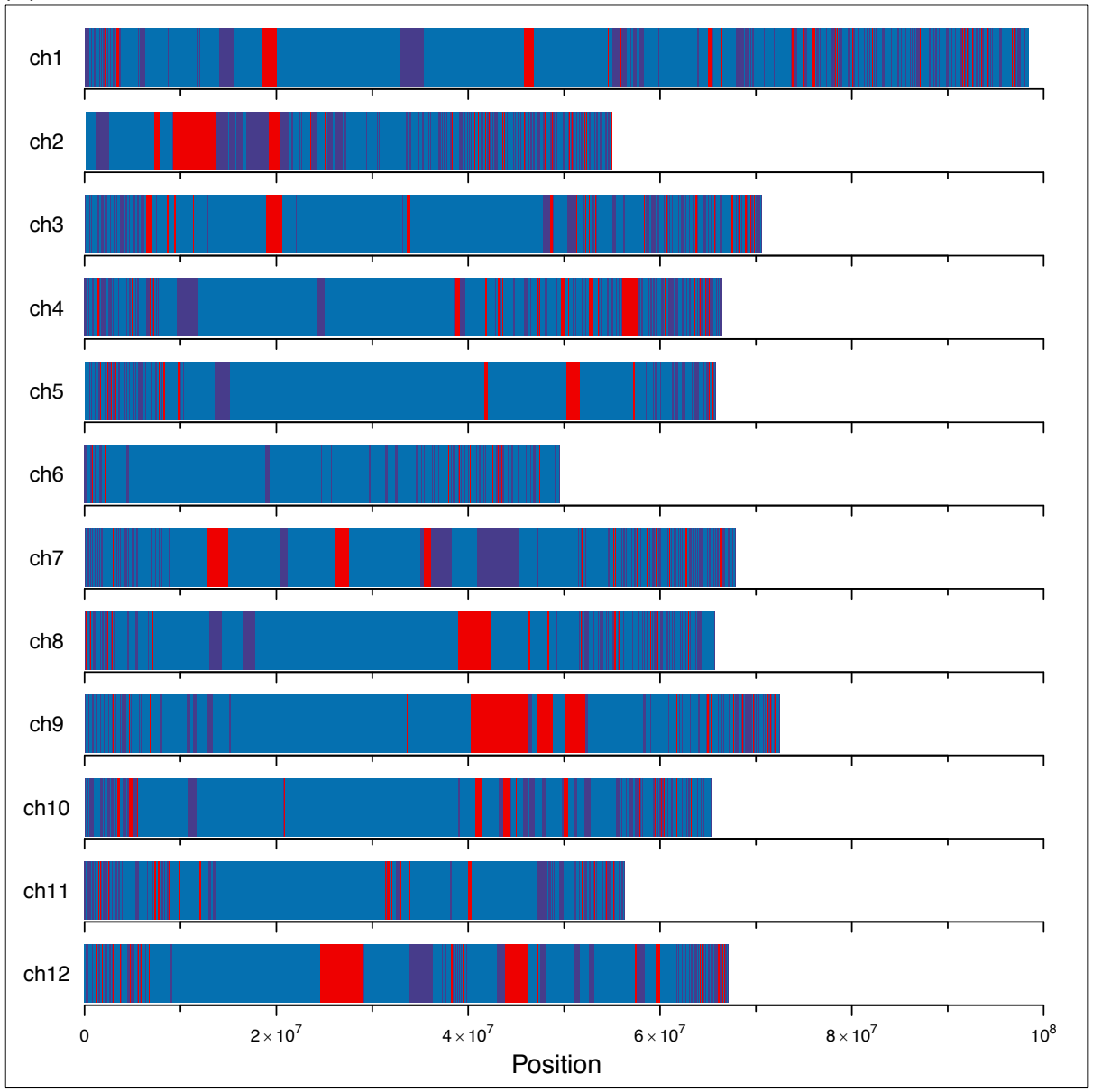


**Supplemental Figure 12** Haplotype blocks identified in three hybrid individuals from HAPMIX. Parental haplotype blocks were identified in the three individual accessions identified as hybrids in this study. Blocks were assigned using all accessions from the parental species *S.chi* and *S.per.* Blocks were then plotted to the genomic position and colored depending on the parental species (red = *S.per* and blue = *S.chi*). Note that there is little coverage in non-genic regions for these accessions because it is RNAseq data, so the assignment of large blocks to *S.chi* in non-genic regions might be misleading. Data is shown for (**a**) LA1782, (**b**) LA1930, and (**c**) LA1932.

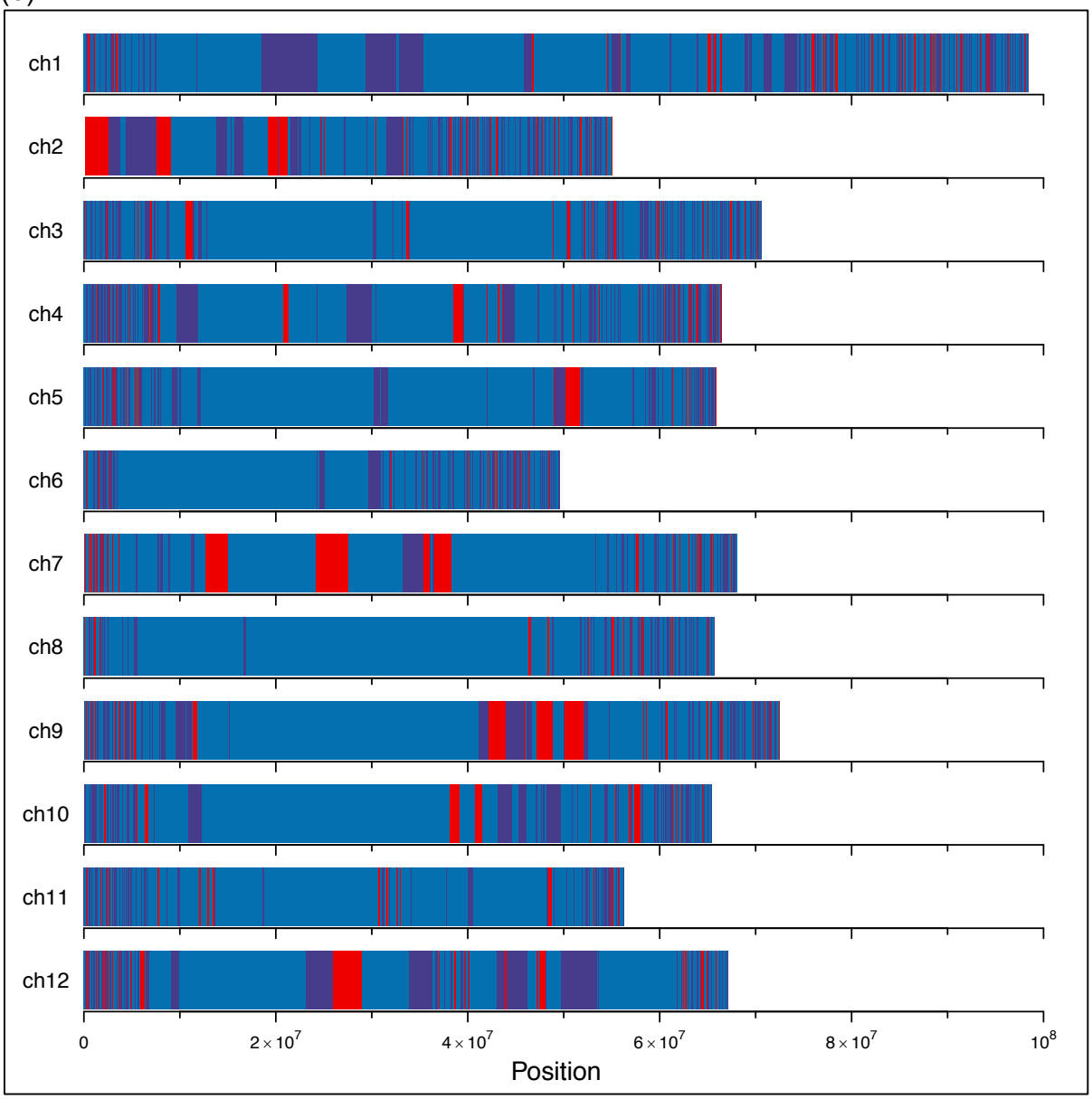
(**a**)



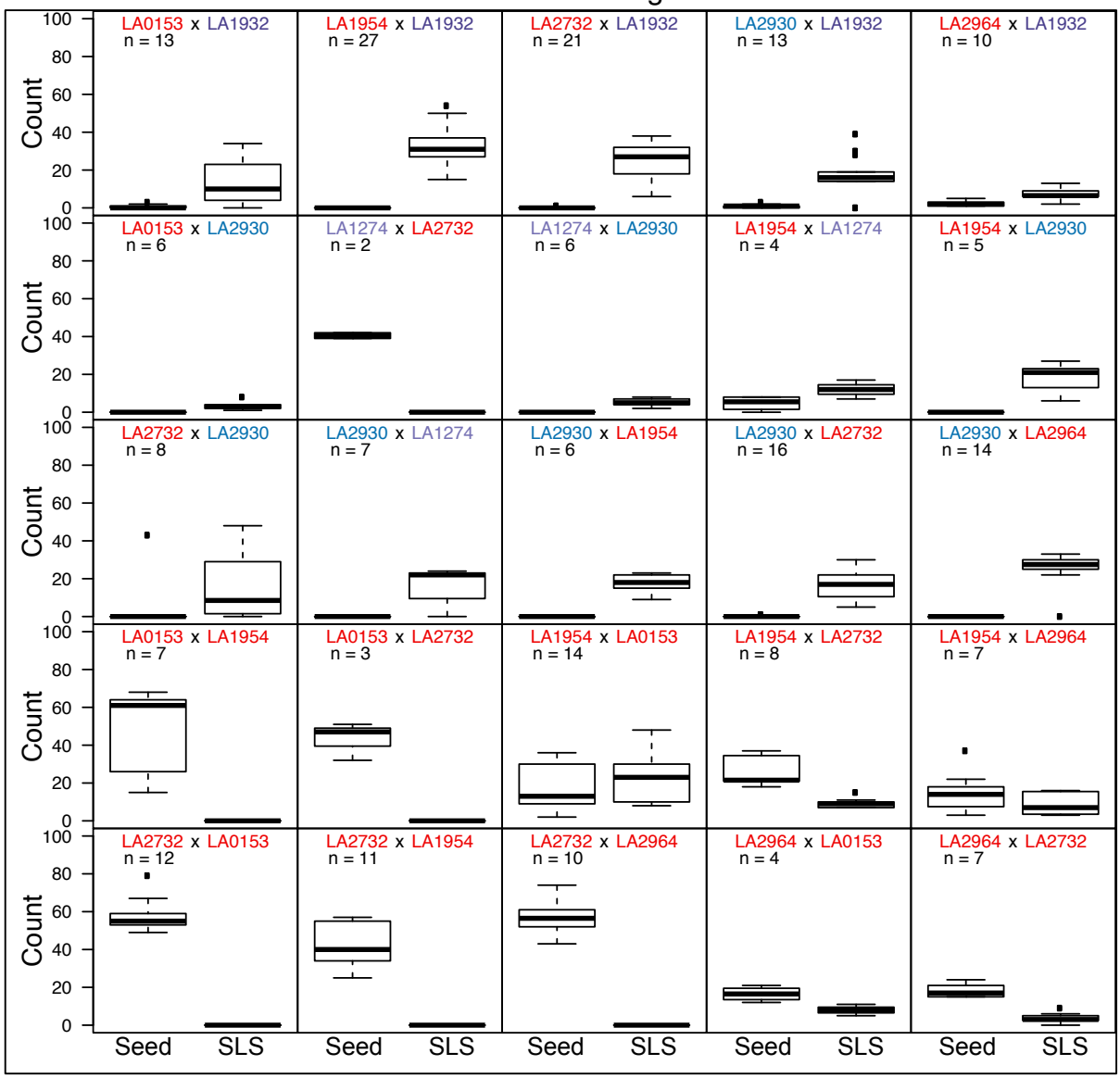
(**b**)


ch1
ch2
ch3
ch4
ch5
ch6
ch7
ch8
ch9
ch10
ch11
ch12
0
2 × 10^7
4 × 10^7
6 × 10^7
8 × 10^7
10^8
Position

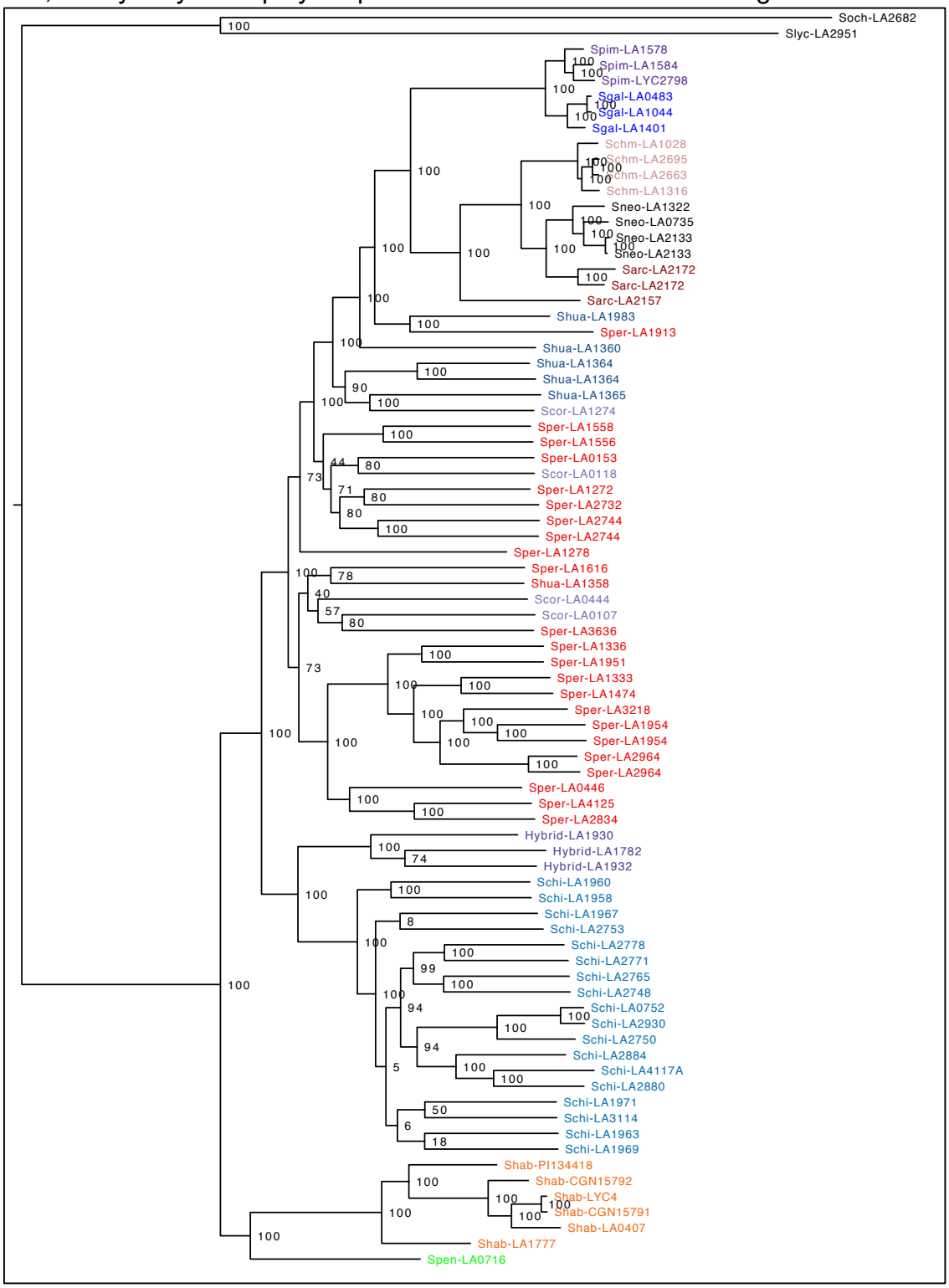
(**c**)


ch1
ch2
ch3
ch4
ch5
ch6
ch7
ch8
ch9
ch10
ch11
ch12
0
2 × 10^7
4 × 10^7
6 × 10^7
8 × 10^7
10^8
Position

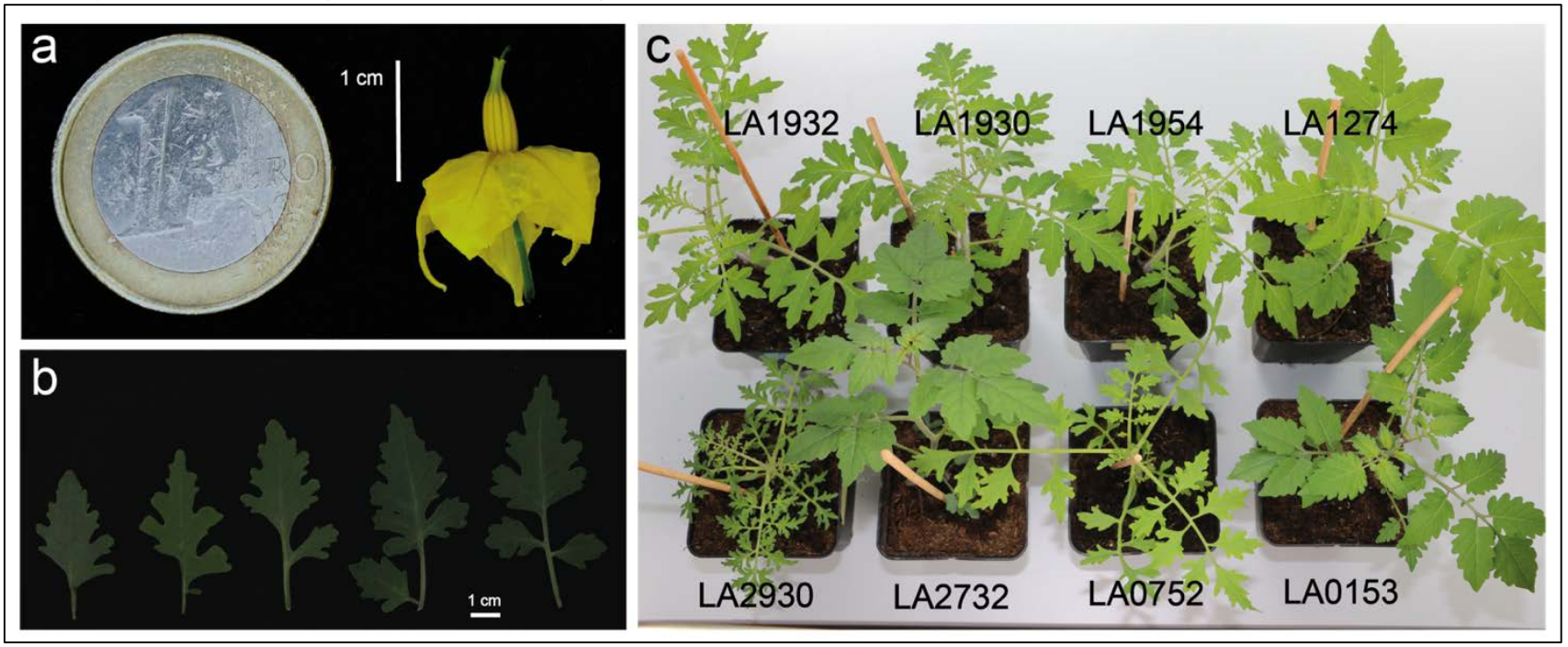
**Supplemental Figure 13** Results of the crossing study. The number of seeds and seed-like structures (SLS) for all crosses are shown as box and whisker plots. The number of fruits from which the data is derived is given as n.



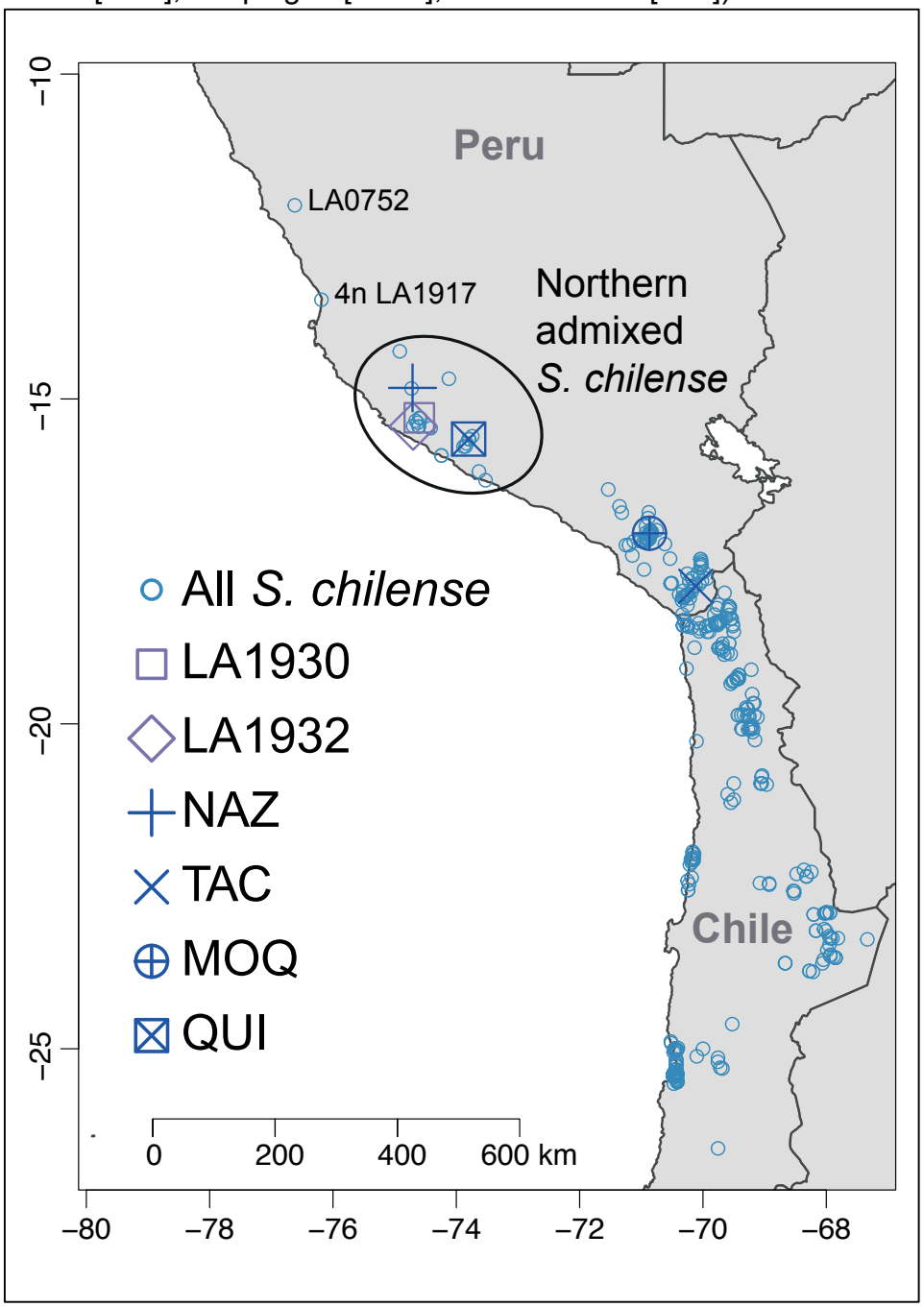
**Supplemental Figure 14** Maximum-likelihood phylogeny with all accessions using 712,432 synonymous polymorphisms with less than 50% missing data.



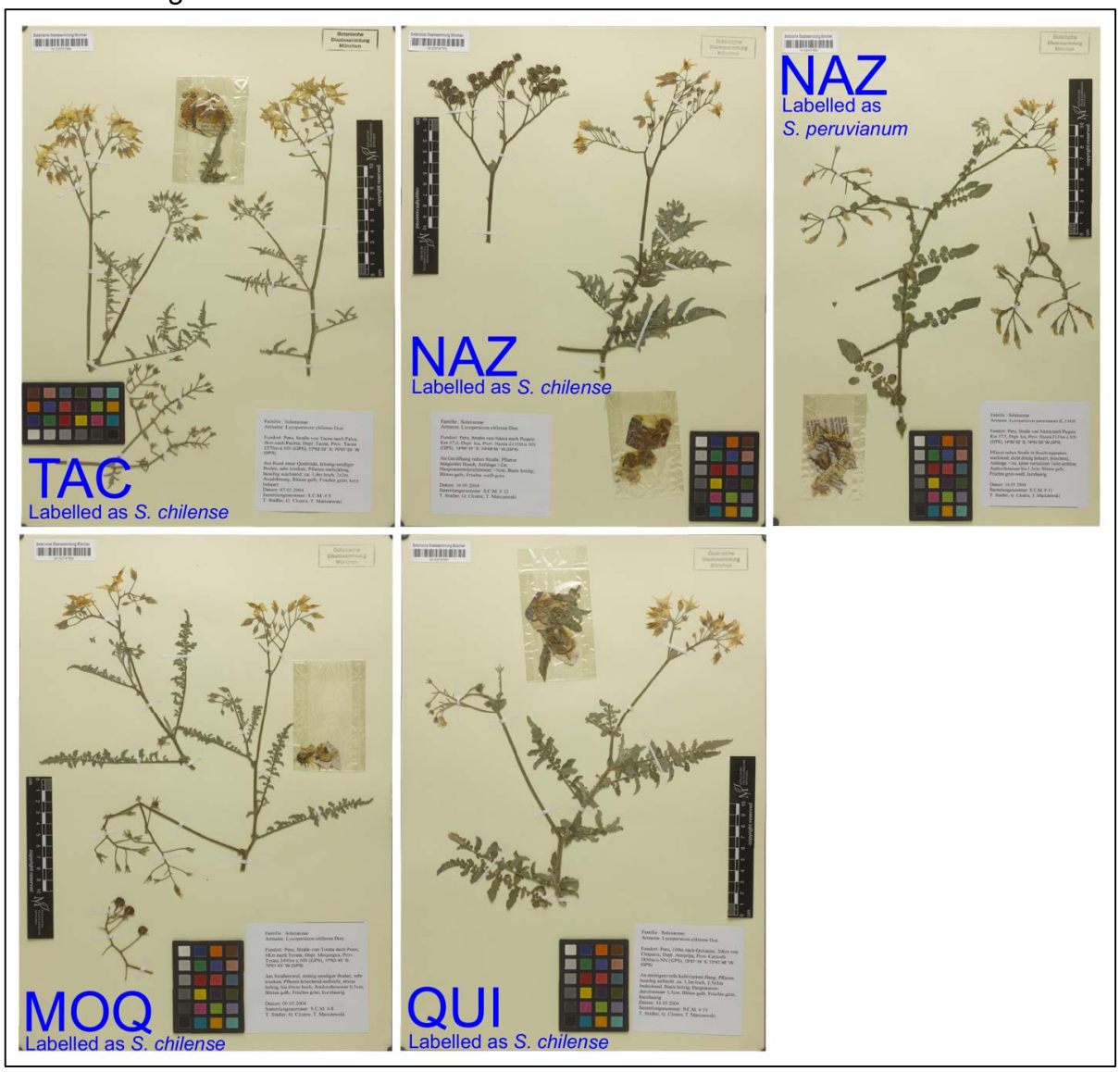
**Supplemental Figure 15** Phenotype of the cryptic *S. chilense* populations LA0752. (**a**) Typical flower of *S.chi*-LA0752. (**b**) Typical leaflet of *S.chi*-LA0752 from several individuals. (**c**) Three-week-old individuals from *S.chi*-LA0752, *S. chilense* (LA2930), *S. peruvianum* (LA0153, LA1954 and LA2732), S. corneliomulleri (LA1272), and the two *S.chi* x *S.per* hybrid accessions (LA1930 and LA1932).



**Supplemental Figure 16** Known distribution of *S. chilense*. All known collections of *S. chilense* (dataset kindly provided by Sandra Knapp, Natural History Museum, London) show the discrete northern populations identified as hybrids in this study. *S.chi*-LA0752 is shown as the most northern *S. chilense* population. The 2004 collection sites of *S. chilense* (Roselius *et al.* 2005) are also shown (Nazca [NAZ], Tacna [TAC], Moquegua [MOQ], and Quicacha [QUI])



**Supplemental Figure 17** Herbarium voucher specimens from Roselius *et al*. 2005. The location of these collections is indicated in Fig. S16. Note the similar phenotype of QUI and *S. chilense* NAZ, and the similarity of these collections to LA1930 and LA1932 in Fig. 6.



## 3.12. Supplemental Tables

**Supplemental Table 1**. Details of the populations and data analyzed in this study.

| Id | Species | Species Abbrev-iation | Study | Duplicate | # Seed-Expansion Generations | # Total Reads | # Positions with MinDepth ge 10 | # Aligned Reads | % Aligned | Province | Country | Elevation | Latitude | Longitude | Mating System |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LA1782 | Hybrid | Hybrid | c | | | 27407970 | 56256043 | 20,234,125 | 74% | Arequipa | Peru | | -15.36 | -74.62 | Allogamous-SI |
| LA1930 | Hybrid | Hybrid | | | 2 | 154021324 | 55483585 | 127,682,186 | 83% | Arequipa | Peru | 500 | -15.29 | -74.6 | Allogamous-SI |
| LA1932 | Hybrid | Hybrid | | | 3 | 70286604 | 41652129 | 58,739,548 | 84% | Arequipa | Peru | 1100 | -15.42 | -74.7 | Allogamous-SI |
| LA2157 | S. arcanum | S.arc | a | | | 336528946 | 269008348 | 175,818,877 | 52% | Cajamarca | Peru | 1600 | -6.51 | -78.81 | Facultative-SC |
| LA2172 | S. arcanum | S.arc | a | YES | | 327920754 | 286177531 | 181,859,614 | 55% | Cajamarca | Peru | | -6 | -78.91 | Allogamous-SI |
| LA2172 | S. arcanum | S.arc | c | YES | | 29277492 | 55365171 | 21,413,972 | 73% | Cajamarca | Peru | | -6 | -78.91 | Allogamous-SI |
| LA0752 | S. chilense | S.chi | | | 5 | 101970490 | 70492417 | 89,227,601 | 88% | Lima | Peru | 1200 | -12.1 | -76.6 | Allogamous-SI |
| LA1958 | S. chilense | S.chi | | | 2 | 106082222 | 63533072 | 90,323,189 | 85% | Moquegua | Peru | 1250 | -17.25 | -71.25 | Allogamous-SI |
| LA1960 | S. chilense | S.chi | | | 2 | 114906258 | 71767018 | 100,616,993 | 88% | Moquegua | Peru | 1850 | -17.08 | -70.87 | Allogamous-SI |
| LA1963 | S. chilense | S.chi | | | 1 | 90304024 | 69548517 | 78,906,333 | 87% | Tacna | Peru | 200 | -18.07 | -70.32 | Allogamous-SI |
| LA1967 | S. chilense | S.chi | | | 2 | 92496720 | 57229647 | 80,442,070 | 87% | Tacna | Peru | 1000 | -17.9 | -70.16 | Allogamous-SI |
| LA1969 | S. chilense | S.chi | | | 3 | 91197178 | 68639340 | 77,708,686 | 85% | Tacna | Peru | 3250 | -17.55 | -70.03 | Allogamous-SI |
| LA1971 | S. chilense | S.chi | | | 3 | 129813792 | 64394111 | 107,887,801 | 83% | Tacna | Peru | 3150 | -17.59 | -70.04 | Allogamous-SI |
| LA2748 | S. chilense | S.chi | | | 3 | 130209336 | 43278704 | 106,619,009 | 82% | Tarapaca | Chile | 800 | -21.21 | -69.55 | Allogamous-SI |
| LA2750 | S. chilense | S.chi | | | 2 | 207655754 | 63160144 | 38,750,231 | 19% | Antofagasta | Chile | 300 | -22.07 | -70.16 | Allogamous-SI |
| LA2753 | S. chilense | S.chi | | | 2 | 93241972 | 77690255 | 70,635,380 | 76% | Tarapaca | Chile | 1650 | -19.86 | -69.34 | Allogamous-SI |
| LA2765 | S. chilense | S.chi | | | 2 | 130974634 | 81499570 | 100,558,213 | 77% | Arica and | Chile | 2400 | -18.77 | -69.68 | Allogamous-SI |
| LA2771 | S. chilense | S.chi | | | 4 | 94589440 | 68362578 | 76,334,713 | 81% | Ãrica and | Chile | 1800 | -18.48 | -69.87 | Allogamous-SI |
| LA2778 | S. chilense | S.chi | | | 3 | 131261078 | 80666920 | 98,948,038 | 75% | Ãrica and | Chile | 2900 | -18.38 | -69.55 | Allogamous-SI |
| LA2880 | S. chilense | S.chi | | | 3 | 82840240 | 65348571 | 64,874,285 | 78% | Antofagasta | Chile | 2500 | -23.82 | -68.22 | Allogamous-SI |
| LA2884 | S. chilense | S.chi | | | 3 | 96468832 | 74182929 | 78,708,109 | 82% | Antofagasta | Chile | 2900 | -22.25 | -68.36 | Allogamous-SI |
| LA2930 | S. chilense | S.chi | | | 2 | 90023102 | 68485886 | 74,777,787 | 83% | Antofagasta | Chile | 550 | -25.5 | -70.42 | Allogamous-SI |
| LA3114 | S. chilense | S.chi | | | 2 | 105516090 | 76953459 | 75,527,578 | 72% | Tacna | Peru | 2960 | -17.68 | -70.08 | Allogamous-SI |
| LA4117A | S. chilense | S.chi | c | | | 36321890 | 62670271 | 27,729,862 | 76% | Antofagasta | Chile | 3540 | -22.9 | -67.94 | Allogamous-SI |
| LA1028 | S. chmielewskii | S.chm | c | | | 20718540 | 32504920 | 11,328,013 | 55% | Apurimac | Peru | 3000 | -13.88 | -73.01 | Facultative-SC |
| LA1316 | S. chmielewskii | S.chm | c | | | 27282862 | 55100912 | 22,331,541 | 82% | Ayacucho | Peru | 2920 | -13.39 | -73.92 | Facultative-SC |
| LA2663 | S. chmielewskii | S.chm | a | | | 341164730 | 292981762 | 187,068,096 | 55% | Cusco | Peru | 2500 | -13.7 | -71.99 | Facultative-SC |
| LA2695 | S. chmielewskii | S.chm | a | | | 354078446 | 257609756 | 193,294,895 | 55% | Cusco | Peru | 2300 | -13.96 | -71.76 | Facultative-SC |
| LA0107 | S. corneliomulleri | S.cor | c | | | 25207286 | 52928891 | 18,464,348 | 73% | Lima | Peru | 60 | -13.01 | -76.38 | Allogamous-SI |
| LA0118 | S. corneliomulleri | S.cor | a | | | 335487318 | 196520686 | 156,957,493 | 47% | | | | | | Allogamous-SI |
| LA0444 | S. corneliomulleri | S.cor | c | | | 24552910 | 52196803 | 18,856,869 | 77% | Ica | Peru | 100 | -13.43 | -76.13 | Allogamous-SI |
| LA1274 | S. corneliomulleri | S.cor | | | 2 | 94550118 | 66437945 | 75,389,974 | 80% | Lima | Peru | 1440 | -11.46 | -76.9 | Allogamous-SI |
| LA0483 | S. galapagense | S.gal | a | | | 325259616 | 653076473 | 300,903,738 | 93% | Galapagos Islands | Ecuador | | -0.37 | -91.6 | Autogamous-SC |
| LA1044 | S. galapagense | S.gal | a | | | 309118016 | 662395936 | 274,029,941 | 89% | Galapagos Islands | Ecuador | | -0.28 | -90.55 | Autogamous-SC |
| LA1401 | S. galapagense | S.gal | a | | | 332617086 | 663119779 | 319,895,925 | 96% | Galapagos Islands | Ecuador | 5 | -0.24 | -91.39 | Autogamous-SC |
| CGN15791 | S. habrochaites | S.hab | a | | | 320002236 | 155245404 | 138,286,664 | 43% | | | | | | |
| CGN15792 | S. habrochaites | S.hab | a | | | 342613468 | 155762000 | 137,105,706 | 40% | | | | | | |
| LA0407 | S. habrochaites | S.hab | a | | | 344461004 | 154612310 | 144,036,116 | 42% | Guayas | Ecuador | 70 | -2.18 | -79.91 | Facultative-SC |
| LA1777 | S. habrochaites | S.hab | a | | | 355583114 | 107272689 | 99,785,409 | 28% | Ancash | Peru | 3216 | -9.55 | -77.67 | Allogamous-SI |
| LYC4 | S. habrochaites | S.hab | a | | | 304186072 | 153083781 | 129,451,667 | 43% | | | | | | |
| PI134418 | S. habrochaites | S.hab | a | | | 333175862 | 149837952 | 141,004,224 | 42% | | | | | | |
| LA1358 | S. huaylasense | S.hua | c | | | 17653244 | 46515260 | 13,994,568 | 79% | Ancash | Peru | 750 | -9.53 | -77.96 | Allogamous-SI |
| LA1360 | S. huaylasense | S.hua | c | | | 25716052 | 54684140 | 20,772,720 | 81% | Ancash | Peru | 1490 | -9.54 | -77.93 | Allogamous-SI |
| LA1364 | S. huaylasense | S.hua | a | YES | | 334812164 | 158063838 | 155,785,273 | 47% | Ancash | Peru | 2920 | -10.13 | -77.39 | Allogamous-SI |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LA1364 | S. huaylasense | S.hua | c | YES | | 30423044 | 61774387 | 25,354,397 | 83% | Ancash | Peru | 2920 | -10.13 | -77.39 | Allogamous-SI |
| LA1365 | S. huaylasense | S.hua | a | | | 356531462 | 193579086 | 150,431,862 | 42% | Ancash | Peru | 2450 | -10.16 | -77.43 | Allogamous-SI |
| LA1983 | S. huaylasense | S.hua | a | | | 355914006 | 253148108 | 177,188,003 | 50% | Ancash | Peru | 940 | -8.69 | -77.97 | Allogamous-SI |
| LA2951 | S. lycopersicoides | S.lyc | | | 4 | 87246466 | 46040360 | 62,720,955 | 72% | Tarapaca | Chile | 2200 | -19.32 | -69.45 | Allogamous-SI |
| LA0735 | S. neorickii | S.neo | a | | | 356305676 | 287714940 | 196,633,336 | 55% | Huanuco | Peru | | -10.4 | -76.2 | Autogamous-SC |
| LA1322 | S. neorickii | S.neo | c | | | 38679884 | 69960549 | 29,641,067 | 77% | Cusco | Peru | 2380 | -13.45 | -72.43 | Autogamous-SC |
| LA2133 | S. neorickii | S.neo | a | YES | | 249439256 | 288126468 | 195,162,489 | 78% | Azuay | Ecuador | 1980 | -3.4 | -79.18 | Autogamous-SC |
| LA2133 | S. neorickii | S.neo | c | YES | | 37045366 | 71293598 | 30,268,114 | 82% | Azuay | Ecuador | 1980 | -3.4 | -79.18 | Autogamous-SC |
| LA2682 | S. ochranthum | S.och | | | 1 | 89755404 | 56376314 | 63,004,248 | 70% | Cusco | Peru | 2500 | -13.63 | -72.24 | Allogamous-SI |
| LA0716 | S. pennellii | S.pen | a | | | 329844716 | 114438071 | 99,403,464 | 30% | Arequipa | Peru | 50 | -16.2 | -73.6 | Facultative-SC |
| LA0153 | S. peruvianum | S.per | | | 6 | 90866996 | 68418645 | 77,375,883 | 85% | Ancash | Peru | 90 | -9.95 | -78.22 | Allogamous-SI |
| LA0446 | S. peruvianum | S.per | | | 4 | 105322468 | 66081629 | 76,578,198 | 73% | Arequipa | Peru | 150 | -15.78 | -74.39 | Allogamous-SI |
| LA1272 | S. peruvianum | S.per | a | | | 319497888 | 105046666 | 157,499,587 | 49% | Lima | Peru | 1000 | -11.52 | -77 | Allogamous-SI |
| LA1278 | S. peruvianum | S.per | a | | | 336836546 | 192320612 | 158,965,815 | 47% | Lima | Peru | 800 | -11.64 | -76.96 | Allogamous-SI |
| LA1333 | S. peruvianum | S.per | | | 3 | 101707500 | 67620815 | 88,724,222 | 87% | Arequipa | Peru | 700 | -16.57 | -72.63 | Allogamous-SI |
| LA1336 | S. peruvianum | S.per | | | 3 | 94154260 | 64819205 | 83,692,673 | 89% | Arequipa | Peru | 10 | -16.21 | -73.62 | Allogamous-SI |
| LA1474 | S. peruvianum | S.per | | | 2 | 77910156 | 70047648 | 68,197,814 | 88% | Arequipa | Peru | 1300 | -16.36 | -73.02 | Allogamous-SI |
| LA1556 | S. peruvianum | S.per | a | | | 277362664 | 188391821 | 150,555,118 | 54% | Lima | Peru | 250 | -12.13 | -77.03 | Allogamous-SI |
| LA1558 | S. peruvianum | S.per | a | | | 317996198 | 161196562 | 144,782,268 | 46% | Lima | Peru | | -11.44 | -76.48 | Allogamous-SI |
| LA1616 | S. peruvianum | S.per | | | 2 | 91459430 | 76729666 | 76,457,864 | 84% | Lima | Peru | 350 | -12.08 | -76.92 | Allogamous-SI |
| LA1913 | S. peruvianum | S.per | | | 2 | 82115154 | 71664926 | 71,652,758 | 87% | Ica | Peru | 900 | -14.4 | -75.2 | Allogamous-SI |
| LA1951 | S. peruvianum | S.per | | | 2 | 118039694 | 79101191 | 80,529,978 | 68% | Arequipa | Peru | 65 | -16.46 | -73.09 | Allogamous-SI |
| LA1954 | S. peruvianum | S.per | | YES | 1 | 78228862 | 60737907 | 68,249,537 | 87% | Arequipa | Peru | 50 | -17.02 | -72.08 | Allogamous-SI |
| LA1954 | S. peruvianum | S.per | a | YES | | 350310992 | 190291649 | 160,499,579 | 46% | Arequipa | Peru | 50 | -17.02 | -72.08 | Allogamous-SI |
| LA2732 | S. peruvianum | S.per | | | 1 | 74996958 | 68517416 | 63,277,237 | 84% | Tarapaca | Chile | 1750 | -19.42 | -69.58 | Allogamous-SI |
| LA2744 | S. peruvianum | S.per | | YES | 2 | 82672378 | 48268384 | 29,669,894 | 36% | Arica and | Chile | 400 | -18.55 | -70.15 | Allogamous-SI |
| LA2744 | S. peruvianum | S.per | c | YES | | 35864470 | 52467877 | 23,167,401 | 65% | Arica and | Chile | 400 | -18.55 | -70.15 | Allogamous-SI |
| LA2834 | S. peruvianum | S.per | | | 3 | 90252506 | 76002730 | 66,065,914 | 73% | Ica | Peru | 1200 | -14.77 | -74.82 | Allogamous-SI |
| LA2964 | S. peruvianum | S.per | | YES | 3 | 86482844 | 69973707 | 71,257,540 | 82% | Tacna | Peru | 100 | -18.03 | -70.84 | Allogamous-SI |
| LA2964 | S. peruvianum | S.per | c | YES | | 24875770 | 54127010 | 20,010,110 | 80% | Tacna | Peru | 100 | -18.03 | -70.84 | Allogamous-SI |
| LA3218 | S. peruvianum | S.per | | | 2 | 82629604 | 74323523 | 67,628,976 | 82% | Arequipa | Peru | 600 | -16.95 | -72.08 | Allogamous-SI |
| LA3636 | S. peruvianum | S.per | | | 1 | 94711590 | 70088210 | 74,591,151 | 79% | Lima | Peru | | -12.68 | -76.4 | Allogamous-SI |
| LA4125 | S. peruvianum | S.per | | | 1 | 83898054 | 44871634 | 66,134,403 | 79% | Tarapaca | Chile | 2510 | -19.31 | -69.42 | Facultative-SC |
| LA1578 | S. pimpinellifolium | S.pim | a | YES | | 339542086 | 588512441 | 306,955,744 | 90% | La Libertad | Peru | | -7.33 | -79.58 | Autogamous-SC |
| LA1584 | S. pimpinellifolium | S.pim | a | | | 339788834 | 657286470 | 258,871,034 | 76% | Lambayeque | Peru | | -6.37 | -79.79 | Autogamous-SC |
| LYC2798 | S. pimpinellifolium | S.pim | a | | | 346344194 | 444970160 | 293,051,229 | 85% | | | | | | |

**a** Aflitos S, Schijlen E, de Jong H, et al. (2014) Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. Plant Journal 80, 136-148.
**c** Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLoS Biology 14.

**Supplemental Table 2**. Additional data from Lin T, Zhu G, Zhang J, et al. (2014) Genomic analyses provide insights into the history of tomato breeding. Nature Genetics 46, 1220-1226 used for a ∂a∂i analysis.

| Id | Species | Species Abbreviation | Duplicate | Total Reads | Province | Country | Elevation | Longitude | Latitude | Mating System |
|---|---|---|---|---|---|---|---|---|---|---|
| LA1237 | S. pimpinellifolium | *S.pim* | | 61858842 | Esmeraldes | Ecuador | 5 | 0.87 | 79.85 | Autogamous-SC |
| LA1242 | S. pimpinellifolium | *S.pim* | | 67157190 | Morona-Santiago | Ecuador | 900 | | | Autogamous-SC |
| LA1246 | S. pimpinellifolium | *S.pim* | | 57831786 | Loja | Ecuador | 1200 | -3.99 | -79.36 | Autogamous-SC |
| LA1341 | S. pimpinellifolium | *S.pim* | | 49877478 | Lima | Peru | 500 | -11.97 | -76.79 | Autogamous-SC |
| LA1547 | S. pimpinellifolium | *S.pim* | | 58611708 | Carchi | Ecuador | 3000 | 0.58 | -77.93 | Autogamous-SC |
| LA1578 | S. pimpinellifolium | *S.pim* | YES | 55566326 | La Libertad | Peru | | -7.33 | -79.58 | Autogamous-SC |
| LA1591 | S. pimpinellifolium | *S.pim* | | 59333440 | La Libertad | Peru | | -7.72 | -79.12 | Autogamous-SC |
| LA1595 | S. pimpinellifolium | *S.pim* | | 56572400 | Ancash | Peru | 10 | -9.27 | -78.47 | Autogamous-SC |
| LA1596 | S. pimpinellifolium | *S.pim* | | 76571072 | Ancash | Peru | 15 | -8.93 | -78.57 | Autogamous-SC |
| LA1933 | S. pimpinellifolium | *S.pim* | | 59626428 | Arequipa | Peru | 140 | -15.46 | -74.45 | Autogamous-SC |
| LA2147 | S. pimpinellifolium | *S.pim* | | 55487778 | Cajamarca | Peru | 550 | -7.2 | -78.98 | Autogamous-SC |
| LA2173 | S. pimpinellifolium | *S.pim* | | 59361256 | Cajamarca | Peru | 1200 | | | Autogamous-SC |
| LA2181 | S. pimpinellifolium | *S.pim* | | 62378096 | Cajamarca | Peru | 850 | -5.78 | -78.78 | Autogamous-SC |
| LA2184 | S. pimpinellifolium | *S.pim* | | 55498628 | Amazonas | Peru | 450 | -5.59 | -78.55 | Autogamous-SC |
| LA2187 | S. pimpinellifolium | *S.pim* | | 56411540 | Amazonas | Peru | 650 | -5.93 | -78.05 | Autogamous-SC |
| LA2656 | S. pimpinellifolium | *S.pim* | | 65242236 | Tumbez | Peru | 70 | -3.8 | -80.7 | Autogamous-SC |
| LA2857 | S. pimpinellifolium | *S.pim* | | 55283766 | Galapagos Islands | Peru | 5 | -0.95 | -90.97 | Autogamous-SC |

**Supplemental Table 3**. Heterozygosity for each individual.

| Species | Id | Heterozygosity |
|---|---|---|
| Hybrid | LA1782 | 0.38% |
| Hybrid | LA1930 | 0.66% |
| Hybrid | LA1932 | 0.65% |
| *S.arc* | LA2157 | 0.06% |
| *S.arc* | LA2172 | 0.15% |
| *S.arc* | LA2172 | 0.22% |
| *S.chi* | LA0752 | 0.10% |
| *S.chi* | LA1958 | 0.54% |
| *S.chi* | LA1960 | 0.57% |
| *S.chi* | LA1963 | 0.56% |
| *S.chi* | LA1967 | 0.55% |
| *S.chi* | LA1969 | 0.39% |
| *S.chi* | LA1971 | 0.54% |
| *S.chi* | LA2748 | 0.50% |
| *S.chi* | LA2750 | 0.26% |
| *S.chi* | LA2753 | 0.53% |
| *S.chi* | LA2765 | 0.43% |
| *S.chi* | LA2771 | 0.54% |
| *S.chi* | LA2778 | 0.51% |
| *S.chi* | LA2880 | 0.31% |
| *S.chi* | LA2884 | 0.38% |
| *S.chi* | LA2930 | 0.17% |
| *S.chi* | LA3114 | 0.54% |
| *S.chi* | LA4117A | 0.35% |
| *S.chm* | LA1028 | 0.04% |
| *S.chm* | LA1316 | 0.08% |
| *S.chm* | LA2663 | 0.04% |
| *S.chm* | LA2695 | 0.06% |
| *S.cor* | LA0107 | 0.39% |
| *S.cor* | LA0118 | 0.64% |
| *S.cor* | LA0444 | 0.68% |
| *S.cor* | LA1274 | 0.61% |
| *S.gal* | LA0483 | 0.01% |
| *S.gal* | LA1044 | 0.02% |
| *S.gal* | LA1401 | 0.03% |
| *S.hab* | CGN15791 | 0.08% |
| *S.hab* | CGN15792 | 0.15% |
| *S.hab* | LA0407 | 0.14% |
| *S.hab* | LA1777 | 0.24% |
| *S.hab* | LYC4 | 0.08% |
| *S.hab* | PI134418 | 0.27% |
| *S.hua* | LA1358 | 0.57% |
| *S.hua* | LA1360 | 0.51% |
| *S.hua* | LA1364 | 0.49% |
| *S.hua* | LA1364 | 0.51% |
| *S.hua* | LA1365 | 0.73% |
| *S.hua* | LA1983 | 0.45% |

| | | |
|---|---|---|
| *S.lyc* | LA2951 | 0.31% |
| *S.neo* | LA0735 | 0.04% |
| *S.neo* | LA1322 | 0.04% |
| *S.neo* | LA2133 | 0.05% |
| *S.neo* | LA2133 | 0.04% |
| *S.och* | LA2682 | 0.11% |
| *S.pen* | LA0716 | 0.11% |
| *S.per* | LA0153 | 0.32% |
| *S.per* | LA0446 | 0.39% |
| *S.per* | LA1272 | 0.58% |
| *S.per* | LA1278 | 0.55% |
| *S.per* | LA1333 | 0.41% |
| *S.per* | LA1336 | 0.40% |
| *S.per* | LA1474 | 0.32% |
| *S.per* | LA1556 | 0.38% |
| *S.per* | LA1558 | 0.72% |
| *S.per* | LA1616 | 0.64% |
| *S.per* | LA1913 | 0.66% |
| *S.per* | LA1951 | 0.47% |
| *S.per* | LA1954 | 0.40% |
| *S.per* | LA1954 | 0.37% |
| *S.per* | LA2732 | 0.55% |
| *S.per* | LA2744 | 0.66% |
| *S.per* | LA2744 | 0.60% |
| *S.per* | LA2834 | 0.60% |
| *S.per* | LA2964 | 0.28% |
| *S.per* | LA2964 | 0.32% |
| *S.per* | LA3218 | 0.44% |
| *S.per* | LA3636 | 0.78% |
| *S.per* | LA4125 | 0.53% |
| *S.pim* | LA1578 | 0.14% |
| *S.pim* | LA1584 | 0.05% |
| *S.pim* | LYC2798 | 0.03% |

**Supplemental Table 4**. Highest log-likelihood parameters inferred from ∂a∂i with confidence intervals from 100 conventional bootstraps.

| Parameter | Value | Description |
|---|---|---|
| Maximum Log-Likelihood | -11,786 | |
| Theta | (24,409, 25,758) | Ancestral population size ($4 \cdot N_{ref} \cdot L \cdot \mu$) |
| $N_{ref}$ | (417,486, 440,556) | Individuals in the ancestral population (Theta / $L \cdot \mu \cdot 4$) |
| L | 2,866,063.30 | Effective length. The number of synonymous nucleotides sequenced to a depth of 3 reads in all individuals |
| Tau | (1.46, 1.56) | Speciation event in $2 \cdot N_{ref}$ |
| Split Time | (1,216,455, 1,374,108) | Speciation event in years ($2 \cdot N_{ref} \cdot Tau \cdot g$), if g =1 |
| Nu1F | (1.26, 1.32) | *S. chilense* population size relative to the ancestral population ($N_{ref}$) |
| Nu2F | (3.69, 3.87) | *S. peruvianum* population size |
| m12 | (0.26, 0.28) | *S. chilense* individuals introduced into S. peruvianum population in 1 generation |
| m21 | (0.13, 0.14) | *S. peruvianum* individuals introduced into S. chilense population in 1 generation |
| M12 | ($3.11 \times 10^{-7}$, $3.23 \times 10^{-7}$) | Fraction of *S. chilense* individuals derived from *S. peruvianum* in 1 generation |
| M21 | ($1.54 \times 10^{-7}$, $1.61 \times 10^{-7}$) | Fraction of *S. peruvianum* individuals derived from *S. chilense* in 1 generation |
| outgr_misid | (0.11, 0.15) | Outgroup misidentification rate |
| μ | $5.1 \cdot 10^{-9}$ | Mutation rate (Roselius et al. 2005) |

**Supplemental Table 5**. Number of seeds and seed-like structures (SLS) for individual crosses.

| Cross | Parent1 | Parent2 | SpeciesParent1 | SpeciesParent2 | Seed | SLS |
|---|---|---|---|---|---|---|
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 2 | 9 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 1 | 10 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 1 | 11 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 3 | 4 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 0 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 2 | 4 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 29 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 31 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 34 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 23 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 21 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 6 |
| LA0153xLA1932 | LA0153 | LA1932 | Sper | Hybrid | 0 | 2 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 54 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 42 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 46 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 50 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 35 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 39 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 48 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 29 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 32 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 36 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 32 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 38 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 36 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 31 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 28 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 26 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 27 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 33 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 27 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 29 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 23 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 25 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 29 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 27 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 23 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 15 |
| LA1954xLA1932 | LA1954 | LA1932 | Sper | Hybrid | 0 | 26 |

| | | | | | | |
|---|---|---|---|---|---|---|
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 1 | 6 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 6 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 21 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 16 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 19 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 14 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 16 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 18 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 1 | 32 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 34 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 38 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 36 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 26 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 34 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 31 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 28 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 27 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 22 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 31 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 38 |
| LA2732xLA1932 | LA2732 | LA1932 | Sper | Hybrid | 0 | 32 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 1 | 30 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 1 | 15 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 2 | 15 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 1 | 16 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 0 | 14 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 1 | 18 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 0 | 19 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 0 | 16 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 1 | 39 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 2 | 28 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 0 | 0 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 0 | 0 |
| LA2930xLA1932 | LA2930 | LA1932 | Schi | Hybrid | 3 | 14 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 5 | 9 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 1 | 13 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 5 | 6 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 1 | 12 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 3 | 3 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 3 | 2 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 2 | 9 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 1 | 6 |
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 2 | 7 |

| | | | | | | |
|---|---|---|---|---|---|---|
| LA2964xLA1932 | LA2964 | LA1932 | Sper | Hybrid | 1 | 6 |
| LA0153xLA2930 | LA0153 | LA2930 | Sper | Schi | 0 | 8 |
| LA0153xLA2930 | LA0153 | LA2930 | Sper | Schi | 0 | 3 |
| LA0153xLA2930 | LA0153 | LA2930 | Sper | Schi | 0 | 2 |
| LA0153xLA2930 | LA0153 | LA2930 | Sper | Schi | 0 | 3 |
| LA0153xLA2930 | LA0153 | LA2930 | Sper | Schi | 0 | 4 |
| LA0153xLA2930 | LA0153 | LA2930 | Sper | Schi | 0 | 1 |
| LA1274xLA2732 | LA1274 | LA2732 | Scor | Sper | 42 | 0 |
| LA1274xLA2732 | LA1274 | LA2732 | Scor | Sper | 39 | 0 |
| LA1274xLA2930 | LA1274 | LA2930 | Scor | Schi | 0 | 6 |
| LA1274xLA2930 | LA1274 | LA2930 | Scor | Schi | 0 | 8 |
| LA1274xLA2930 | LA1274 | LA2930 | Scor | Schi | 0 | 7 |
| LA1274xLA2930 | LA1274 | LA2930 | Scor | Schi | 0 | 4 |
| LA1274xLA2930 | LA1274 | LA2930 | Scor | Schi | 0 | 4 |
| LA1274xLA2930 | LA1274 | LA2930 | Scor | Schi | 0 | 2 |
| LA1954xLA1274 | LA1954 | LA1274 | Sper | Scor | 8 | 12 |
| LA1954xLA1274 | LA1954 | LA1274 | Sper | Scor | 8 | 17 |
| LA1954xLA1274 | LA1954 | LA1274 | Sper | Scor | 3 | 12 |
| LA1954xLA1274 | LA1954 | LA1274 | Sper | Scor | 0 | 7 |
| LA1954xLA2930 | LA1954 | LA2930 | Sper | Schi | 0 | 21 |
| LA1954xLA2930 | LA1954 | LA2930 | Sper | Schi | 0 | 27 |
| LA1954xLA2930 | LA1954 | LA2930 | Sper | Schi | 0 | 13 |
| LA1954xLA2930 | LA1954 | LA2930 | Sper | Schi | 0 | 6 |
| LA1954xLA2930 | LA1954 | LA2930 | Sper | Schi | 0 | 23 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 24 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 10 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 7 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 48 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 34 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 2 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 0 | 1 |
| LA2732xLA2930 | LA2732 | LA2930 | Sper | Schi | 43 | 0 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 23 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 19 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 0 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 0 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 24 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 22 |
| LA2930xLA1274 | LA2930 | LA1274 | Schi | Scor | 0 | 23 |
| LA2930xLA1954 | LA2930 | LA1954 | Schi | Sper | 0 | 9 |
| LA2930xLA1954 | LA2930 | LA1954 | Schi | Sper | 0 | 22 |
| LA2930xLA1954 | LA2930 | LA1954 | Schi | Sper | 0 | 16 |
| LA2930xLA1954 | LA2930 | LA1954 | Schi | Sper | 0 | 23 |

| | | | | | | |
|---|---|---|---|---|---|---|
| LA2930xLA1954 | LA2930 | LA1954 | Schi | Sper | 0 | 20 |
| LA2930xLA1954 | LA2930 | LA1954 | Schi | Sper | 0 | 15 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 26 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 21 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 30 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 21 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 18 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 23 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 17 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 25 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 1 | 16 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 13 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 15 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 17 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 1 | 8 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 6 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 5 |
| LA2930xLA2732 | LA2930 | LA2732 | Schi | Sper | 0 | 5 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 25 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 28 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 26 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 22 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 33 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 26 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 28 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 33 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 30 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 27 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 25 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 29 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 31 |
| LA2930xLA2964 | LA2930 | LA2964 | Schi | Sper | 0 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 68 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 61 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 67 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 29 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 61 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 15 | 0 |
| LA0153xLA1954 | LA0153 | LA1954 | Sper | Sper | 23 | 0 |
| LA0153xLA2732 | LA0153 | LA2732 | Sper | Sper | 51 | 0 |
| LA0153xLA2732 | LA0153 | LA2732 | Sper | Sper | 47 | 0 |
| LA0153xLA2732 | LA0153 | LA2732 | Sper | Sper | 32 | 0 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 9 | 48 |

| | | | | | | |
|---|---|---|---|---|---|---|
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 10 | 45 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 30 | 29 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 2 | 8 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 33 | 14 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 36 | 10 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 25 | 19 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 11 | 8 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 8 | 9 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 25 | 23 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 33 | 23 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 15 | 25 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 8 | 32 |
| LA1954xLA0153 | LA1954 | LA0153 | Sper | Sper | 10 | 30 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 35 | 7 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 37 | 15 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 34 | 9 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 21 | 9 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 21 | 7 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 21 | 11 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 18 | 9 |
| LA1954xLA2732 | LA1954 | LA2732 | Sper | Sper | 22 | 7 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 37 | 7 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 22 | 16 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 14 | 16 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 9 | 15 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 14 | 3 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 6 | 3 |
| LA1954xLA2964 | LA1954 | LA2964 | Sper | Sper | 3 | 4 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 79 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 54 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 56 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 53 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 54 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 52 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 57 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 61 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 56 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 53 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 67 | 0 |
| LA2732xLA0153 | LA2732 | LA0153 | Sper | Sper | 49 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 25 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 35 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 33 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 40 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 36 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 29 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 55 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 57 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 55 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 56 | 0 |
| LA2732xLA1954 | LA2732 | LA1954 | Sper | Sper | 51 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 74 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 54 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 53 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 43 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 60 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 66 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 52 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 59 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 48 | 0 |
| LA2732xLA2964 | LA2732 | LA2964 | Sper | Sper | 61 | 0 |
| LA2964xLA0153 | LA2964 | LA0153 | Sper | Sper | 21 | 5 |
| LA2964xLA0153 | LA2964 | LA0153 | Sper | Sper | 15 | 11 |
| LA2964xLA0153 | LA2964 | LA0153 | Sper | Sper | 12 | 8 |
| LA2964xLA0153 | LA2964 | LA0153 | Sper | Sper | 18 | 8 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 20 | 0 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 24 | 3 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 15 | 9 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 22 | 2 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 15 | 3 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 15 | 6 |
| LA2964xLA2732 | LA2964 | LA2732 | Sper | Sper | 17 | 4 |

# 4. Appendix 1: The Hybrids Database Website

## 4.1. Motivation for the website

In the process of estimating the hybridization frequency in the vascular plants of Michigan, we realized that biogeography is a critical factor that cannot be represented by one flora. What this means is that some species that do not hybridize in Michigan do hybridize elsewhere. Thus, the fundamental question of our study – what proportion of Michigan's plant species hybridize – can only be answered if we also consider other floras that include species found in Michigan. This is also fundamental to answering what the frequency of hybridization in the wild is. For example, two floras share 50% of their species, but 90% of their hybrids, then there is a diminishing return on hybrids per new species and we will be overestimating hybridization by only looking at one flora.

Thus, a website database was built to facilitate future studies in plant hybridization. The website was built using python and the Django framework. The website functions over a data model. The model has instances for all species and all of their unique hybrids. Not all of the data is available for all species, but this is the information that we think is important for future studies of wild hybridization. The data is searchable and downloadable. Currently, the following fields are included in the model:

1. Taxonomic Order
2. Taxonomic Family
3. *Genus Species* Author
4. Common Name
5. Physiognomy
6. Sexual System
7. Coefficient of Wetness
8. Coefficient of Conservatism
9. Hybrid Species True/False
10. Unique Hybrid True/False
11. If 'Unique Hybrid' or 'Hybrid Species' Parent 1
12. If 'Unique Hybrid' or 'Hybrid Species' Parent 2
13. Hybridizes True/False
14. Coefficient of Conservatism
15. Ploidy
16. Citation
17. If 'Unique Hybrid' Sterility
18. Image Field
19. Notes

## 4.2. Example pages from the website

### 4.2.1. Home page

plant hybrids database

Home About Browse the data Search for hybrids Download Beddows *et al.* 2017 data Contact Connect

**Welcome to the plant hybrids database.**

This website is a databae of wild plant hybridization and was created to facilitate research on plant hybrids. We want to help answer two questions:

1. What is the frequency of plant hybridization in the wild?
2. What physiological and ecological factors contribute to hybridization propensity?

Read more about hybrids and the idea behind this website here.

Currently, this website has data on species that hybridize in Michigan (incl. hybrid species) and with whom they hybridize (data from the Field Manual of Michigan Flora). You can search for hybrids by clicking here. Or you can browse all of the data by clicking here.

If you would like to contribute to this project or contact us, then please send us an email.

Poplars, cottonwoods, and aspens are all in the genus *Populus* L. How do these species maintain their integrity with widespread hybridization? Search for *Populus* hybrids

### 4.2.2. About page

plant hybrids database

Home About Browse the data Search for hybrids Download Beddows *et al.* 2017 data Contact Connect

**What is a hybrid?**

The term 'hybrid' indicates a product of two heterogeneous things, and today it usually refers to vehicles with both a petrol engine and an electric motor. In biology, the Oxford dictionary defines a hybrid as the offspring of two plants or animals of different species or varieties. G.L. Stebbins defined a hybridization as "crossing between individuals belonging to separate populations which have different adaptive norms." Hybridization is an important process in evolution. Did you know that all non-Africans have 1-2% Neanderthal DNA and that this is the result of hybridization? In plants, hybridization also has cool outcomes, including the origin of new species!

This is a database of hybrids between two plant species. These are termed interspecific hybrids.

**What this website does**

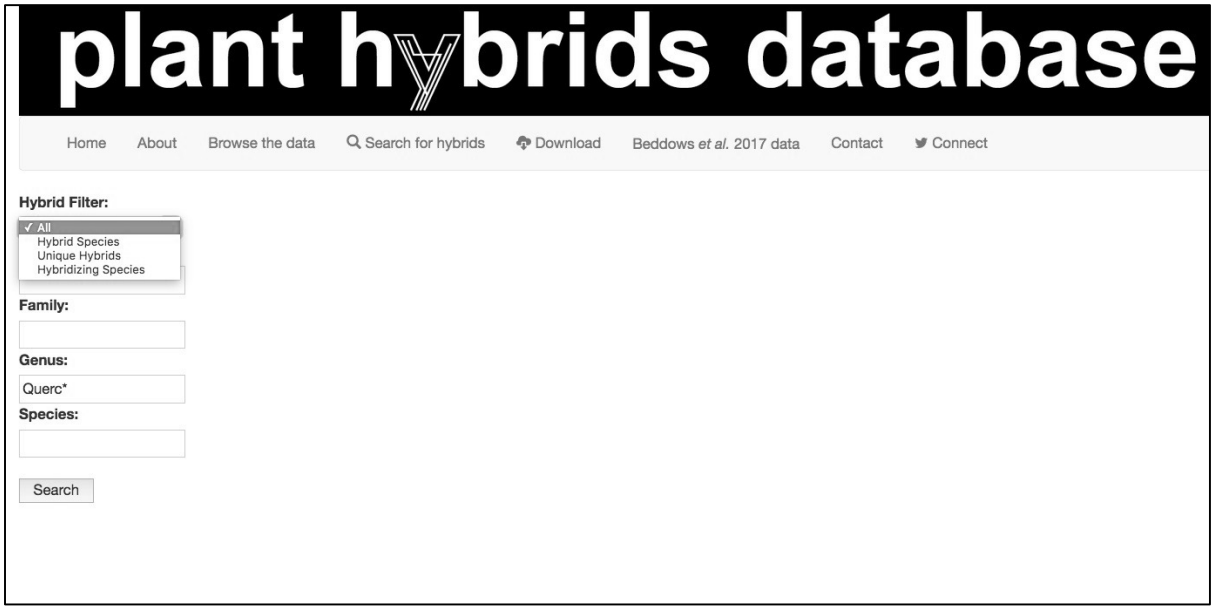
Currently this website has data from the Field manual of Michigan flora. The data is based on a model with the following fields for all species and their unique hybrids. Not all of the data is available for all species, but this is the information that we think is important for future studies of wild hybridization.

1. Taxonomic Order
2. Taxonomic Family
3. *Genus Species* Author
4. Common Name
5. Physiognomy
6. Sexual System
7. Coefficient of Wetness
8. Coefficient of Conservatism
9. Hybrid Species True/False
10. Unique Hybrid True/False
11. If 'Unique Hybrid' or 'Hybrid Species' Parent 1
12. If 'Unique Hybrid' or 'Hybrid Species' Parent 2
13. Hybridizes True/False
14. Coefficient of Conservatism
15. Ploidy
16. Citation
17. If 'Unique Hybrid' Sterility
18. Image Field
19. Note

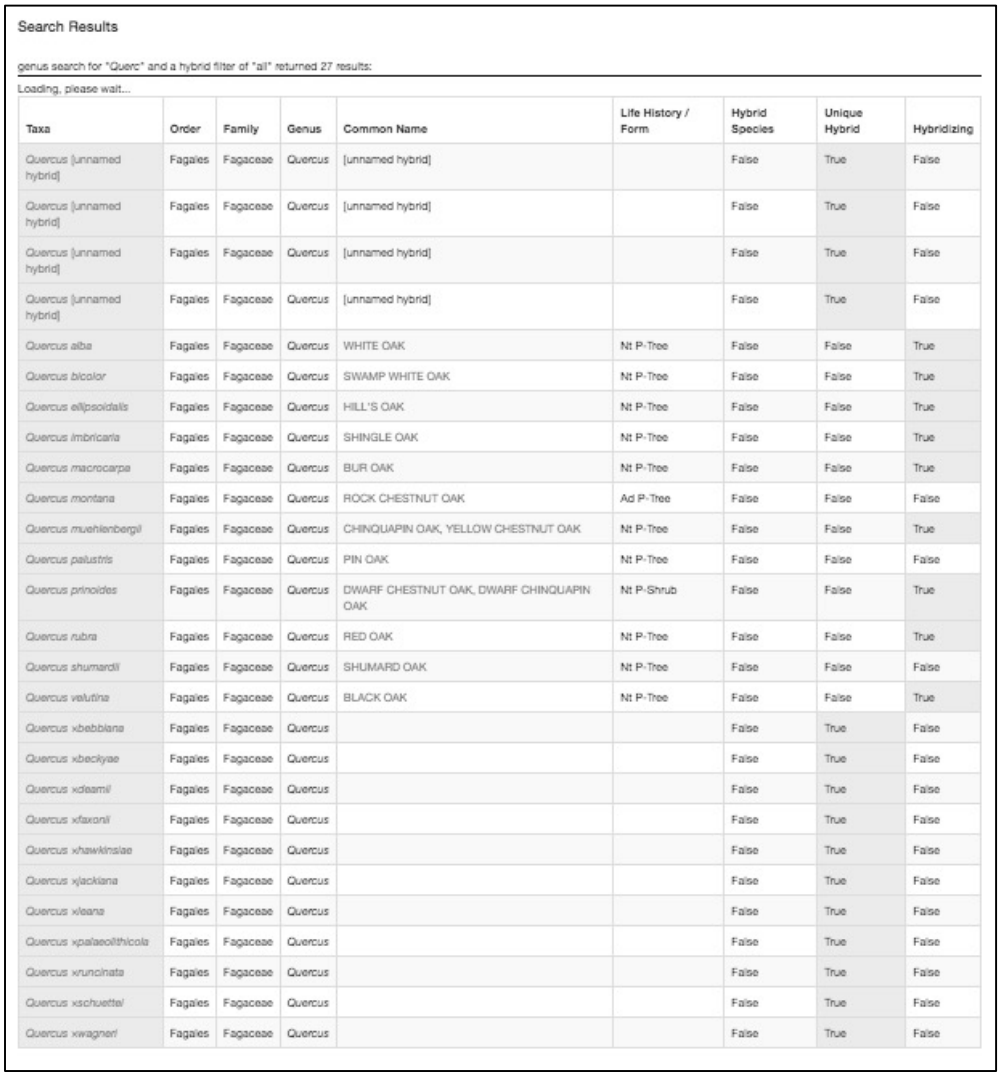
Oaks (*Quercus* L.) are well-known for hybridizing, and were used by Grant (1981) as an example of a syngameon. Search for oak hybrids

The violets (*Viola* L.) of Michigan were called a "pack of incorrigible wolves" by H. E. Ballard (1994) in regards to their difficult taxonomy - a difficulty due to widespread hybridization! Search for *Viola* hybrids

### 4.2.3. Search for data




### 4.2.4. Search results



Search Results

genus search for "Querc" and a hybrid filter of "all" returned 27 results:

Loading, please wait...

| Taxa | Order | Family | Genus | Common Name | Life History / Form | Hybrid Species | Unique Hybrid | Hybridizing |
|---|---|---|---|---|---|---|---|---|
| *Quercus* [unnamed hybrid] | Fagales | Fagaceae | *Quercus* | [unnamed hybrid] | | False | True | False |
| *Quercus* [unnamed hybrid] | Fagales | Fagaceae | *Quercus* | [unnamed hybrid] | | False | True | False |
| *Quercus* [unnamed hybrid] | Fagales | Fagaceae | *Quercus* | [unnamed hybrid] | | False | True | False |
| *Quercus* [unnamed hybrid] | Fagales | Fagaceae | *Quercus* | [unnamed hybrid] | | False | True | False |
| *Quercus alba* | Fagales | Fagaceae | *Quercus* | WHITE OAK | Nt P-Tree | False | False | True |
| *Quercus bicolor* | Fagales | Fagaceae | *Quercus* | SWAMP WHITE OAK | Nt P-Tree | False | False | True |
| *Quercus ellipsoidalis* | Fagales | Fagaceae | *Quercus* | HILL'S OAK | Nt P-Tree | False | False | True |
| *Quercus imbricaria* | Fagales | Fagaceae | *Quercus* | SHINGLE OAK | Nt P-Tree | False | False | True |
| *Quercus macrocarpa* | Fagales | Fagaceae | *Quercus* | BUR OAK | Nt P-Tree | False | False | True |
| *Quercus montana* | Fagales | Fagaceae | *Quercus* | ROCK CHESTNUT OAK | Ad P-Tree | False | False | False |
| *Quercus muehlenbergii* | Fagales | Fagaceae | *Quercus* | CHINQUAPIN OAK, YELLOW CHESTNUT OAK | Nt P-Tree | False | False | True |
| *Quercus palustris* | Fagales | Fagaceae | *Quercus* | PIN OAK | Nt P-Tree | False | False | False |
| *Quercus prinoides* | Fagales | Fagaceae | *Quercus* | DWARF CHESTNUT OAK, DWARF CHINQUAPIN OAK | Nt P-Shrub | False | False | True |
| *Quercus rubra* | Fagales | Fagaceae | *Quercus* | RED OAK | Nt P-Tree | False | False | True |
| *Quercus shumardii* | Fagales | Fagaceae | *Quercus* | SHUMARD OAK | Nt P-Tree | False | False | False |
| *Quercus velutina* | Fagales | Fagaceae | *Quercus* | BLACK OAK | Nt P-Tree | False | False | True |
| *Quercus xbebbiana* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xbeckyae* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xdeamii* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xfaxonii* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xhawkinsiae* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xjackiana* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xleana* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xpalaeolithicola* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xruncinata* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xschuettei* | Fagales | Fagaceae | *Quercus* | | | False | True | False |
| *Quercus xwagneri* | Fagales | Fagaceae | *Quercus* | | | False | True | False |

## 4.2.5. Individual data record for a hybridizing species

### *Betula ×purpusii*

*Betula ×purpusii*
is the **unique hybrid** of *Betula pumila* and *Betula alleghaniensis*

| | |
|---|---|
| **Order** | Fagales |
| **Family** | Betulaceae |
| **Common Name** | |
| **Physiognomy** | - |
| **Sexual System** | |
| **Coef of Wetness** | |
| **Coef of Conservatism** | |
| **Hybrid Species** | False |
| **Unique Hybrid** | True |
| **Hybridizes** | False |
| **Data Source** | Field Manual of Michigan Flora |
| **Ploidy** | |
| **Citation** | |
| **Note** | |

Back to top

## 4.2.6. Individual data record for a hybrid species

### *Betula murrayana* B. V. Barnes & Dancik

*Betula murrayana* is the **hybrid species** between *Betula ×purpusii* and *Betula alleghaniensis*
hybridizes with 4 species:

*Betula murrayana* × *Betula papyrifera* = *Betula* [unnamed hybrid]
*Betula murrayana* × *Betula populifolia* = *Betula* [unnamed hybrid]
*Betula murrayana* × *Betula pumila* = *Betula* [unnamed hybrid]
*Betula murrayana* × *Betula alleghaniensis* = *Betula* [unnamed hybrid]

| | |
|---|---|
| **Order** | Fagales |
| **Family** | Betulaceae |
| **Common Name** | MURRAY BIRCH |
| **Physiognomy** | Nt P-Tree |
| **Sexual System** | Monoecious |
| **Coef of Wetness** | FACW |
| **Coef of Conservatism** | 9 |
| **Hybrid Species** | True |
| **Unique Hybrid** | False |
| **Hybridizes** | False |
| **Data Source** | Field Manual of Michigan Flora |
| **Ploidy** | |
| **Citation** | |
| **Note** | |

Back to top

## 4.2.7. Individual data record for a unique hybrid

*Quercus* [unnamed hybrid]
is the **unique hybrid** of *Quercus alba* and *Quercus muehlenbergii*

| | |
|---|---|
| **Order** | Fagales |
| **Family** | Fagaceae |
| **Common Name** | |
| **Physiognomy** | - |
| **Sexual System** | |
| **Coef of Wetness** | |
| **Coef of Conservatism** | |
| **Hybrid Species** | False |
| **Unique Hybrid** | True |
| **Hybridizes** | False |
| **Data Source** | Field Manual of Michigan Flora |
| **Ploidy** | |
| **Citation** | |
| **Note** | |

Back to top

Copyright 2017 Ian Beddows

# 5. Appendix 2: Additional Tomato Data

The following data is relevant to a full understanding of the wild tomato clade and in particular the species *S. peruvianum*, *S. chilense*, and their hybrids. These analyses include additional information on the RNA-seq data which forms the basis of Chapter 3; a figure of the *S. chilense* and *S. peruvianum* speciation event incorporating parameters from the ∂a∂i model; population statistics for individual genes in *S. chilense* and *S. peruvianum*; results of the population branch statistic test for all genes in *S. chilense*, *S. peruvianum*, and the two genetic demes of *S. peruvianum*; a figure of the distribution of *S. chilense* and *S. peruvianum* haplotype lengths in the hybrid populations identified in this study; two tables of the distribution of fixed SNPs between *S.chi* and *S.per* in two hybrid individuals; a table of germination results for intraspecific, interspecific, and crosses involving hybrid populations; and the results of additional wild tomato crosses that were not included in Chapter 2.

## 5.1. Additional figures explaining the RNA-seq data

**Figure 5.1.1** Distribution of SNPs per Mb.

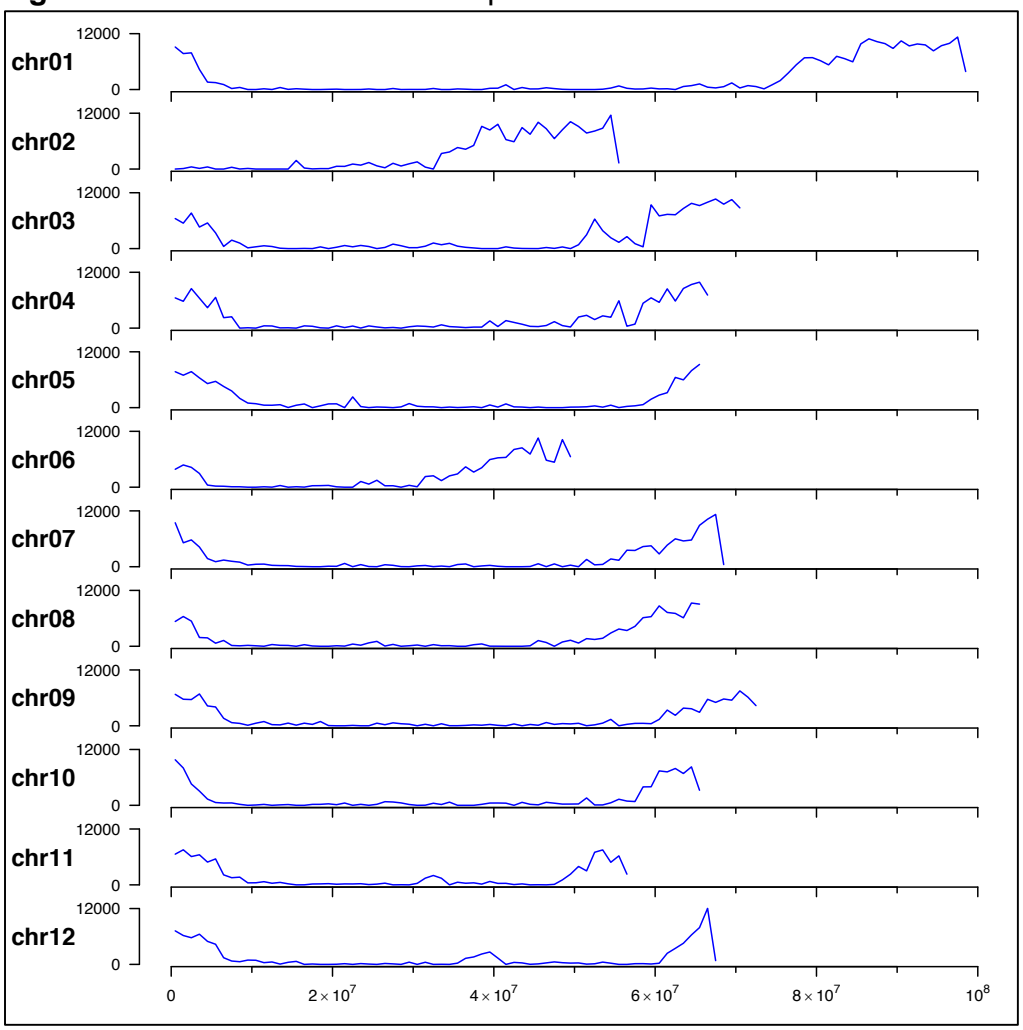


**Figure 5.1.2** Missing data per individual. Following filtering, the missing data rate per individual was not greater than 12%.

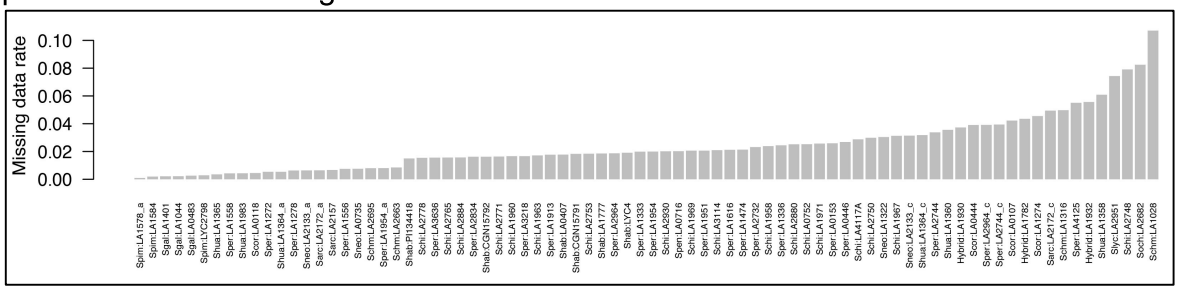
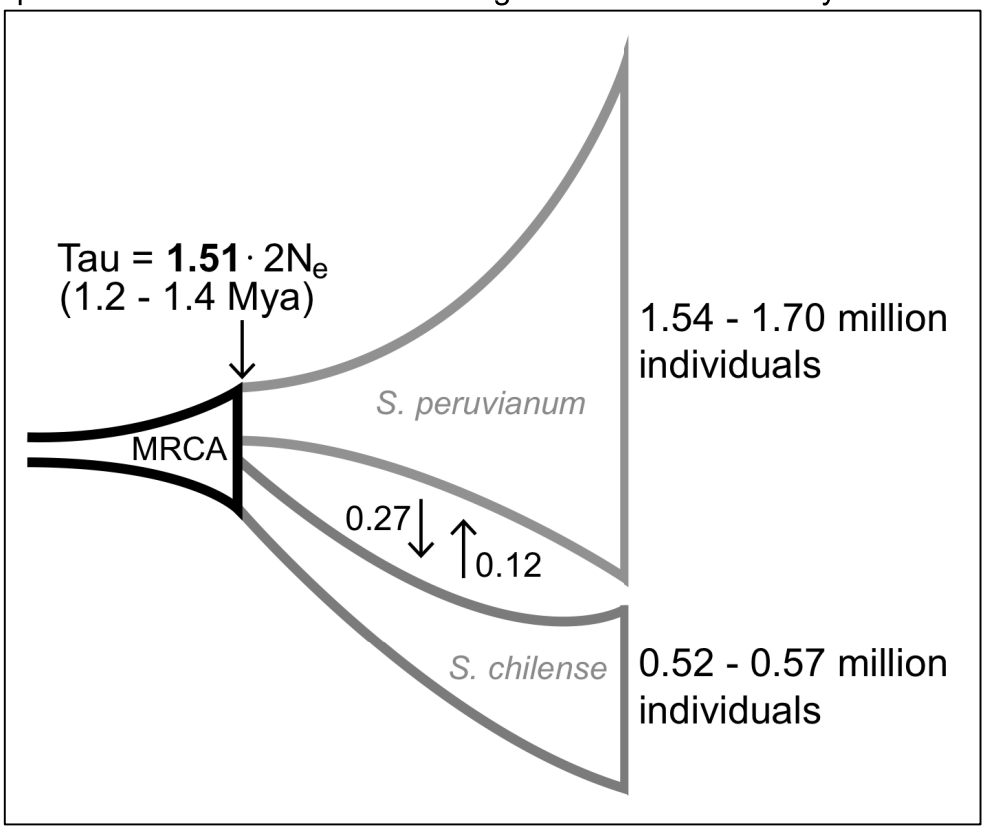



**Figure 5.1.3** Missing data rates per SNP across all accessions. Following filtering, the highest rate of missing data (# accessions with missing data at the SNP position) was 35%. This decreased rapidly and most SNPs had 0% missing data.

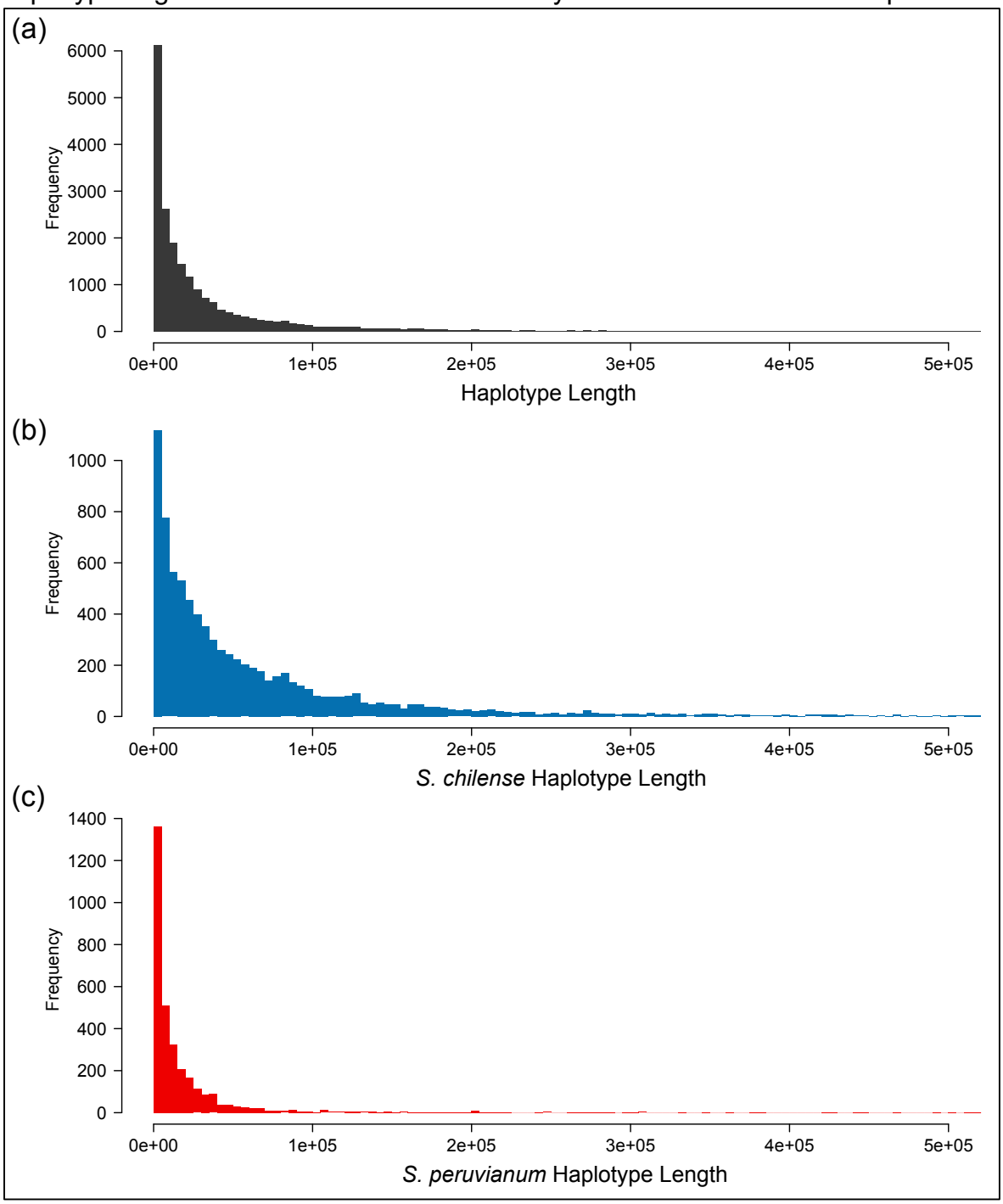

## 5.2. Model of S. chilense and S. peruvianum speciation

**Figure 5.2.1** The maximum-likelihood demographic parameters inferred using ∂a∂i for the current and past populations of *S. chilense and S. peruvianum*. The migration rates between populations are in units of individuals per generation. The speciation event is calibrated for a generation time of one year.



## 5.3. Haplotype lengths in the hybrid populations

**Fig. 5.3.1** *Solanum chilense* and *S. peruvianum* haplotypes were inferred in the hybrid individuals using HAPMIX. The length distribution for all (**a**), *S. chilense* (**b**), and *S. peruvianum* (**c**) haplotypes. On average, *S. peruvianum* haplotype segments are smaller than the *S. chilense* ones. This is evidence of a close relationship between *S. chilense* and the hybrid individuals; the relatively small haplotype segments overall indicate that the hybrids formed in the distant past.



## 5.4. Fixed differences between *S. chilense* and *S. peruvianum*

In total, 2,284 SNP positions were identified as fixed between *S. chilense* and *S. peruvianum* (allele frequency of 0 in one population and 1 in the other). For this analysis, only SNPs that had a minimum depth of 10 reads in ≥10 individuals of *S. chilense* and *S. peruvianum* were considered, and individuals with a depth of less than 10 at the position in question were masked as missing data. This resulted in 2,055,869 filtered positions of which 2,284 were fixed differences. By extension, only 0.11% of SNPs were fixed between *S. chilense* and *S. peruvianum*. The distribution of hybrid genotypes at fixed SNPs is given in Table 5.4.1. The genotype of both hybrids for individual SNPs is given in Table. 5.4.2. Together, this data indicates i) that there are very few fixed differences between *S. chilense* and *S. peruvianum* (0.11% of SNPs), ii) that the hybrids are about 50:50 *S. chilense* and *S. peruvianum* at fixed differences, and iii) the hybrids have nearly identical genotypes in regards to the fixed differences.

**Table 5.4.1.** Hybrid genotypes at fixed differences between *S. chilense* and *S. peruvianum*. CC = homozygous *S. chilense* allele, CP = heterozygous, PP = homozygous *S. peruvianum* allele.

| | *Hybrid Genotype* | | | | |
|---|---|---|---|---|---|
| | *CC* | *CP* | *PP* | *Missing data - depth <10* | *Total* |
| *LA1930* | 714 | 292 | 793 | 485 | 2284 |
| | 31% | 13% | 35% | 21% | 100% |
| *LA1932* | 692 | 167 | 670 | 755 | 2284 |
| | 30% | 7% | 29% | 33% | 100% |

**Table 5.4.2.** The two hybrid populations LA1930 and LA1932 had nearly identical genotypes in regards to SNPs fixed between *S. chilense* and *S. peruvianum.* CC = homozygous S. chilense allele, PP = homozygous S. peruvianum allele.

| | | |
|---|---|---|
| *Both hybrids CC* | 766 | 34% |
| *One or both hybrids heterozygous* | 161 | 7% |
| *Both hybrids PP* | 543 | 24% |
| *One CC and one PP* | 0 | 0% |
| *Missing data (depth <10 reads) in one or both hybrids* | 814 | 36% |
| *Total* | 2284 | 100% |

## 5.5. Seed germination rates for all crosses

**Table 5.5.1.** Seed germination for all crosses which produced seeds. The deme column indicates if crosses were within one deme of *S. peruvianum or between demes. S.cor = S. corneliomulleri*, *S.chi = S. chilense, S. per = S. peruvianum*.

| *Cross* | *Species Female Parent* | *Species Male Parent* | *Tested Number* | *Germinating Number* | *Germination Rate* | *Deme* |
|---|---|---|---|---|---|---|
| LA2964×LA1932 | *S.per* | Hybrid | 5 | 3 | 60% | n/a |
| LA2732×LA2930 | *S.per* | *S.chi* | 5 | 5 | 100% | n/a |
| LA0153×LA1932 | *S.per* | Hybrid | 5 | 2 | 40% | n/a |
| LA1954×LA0153 | *S.per* | *S.per* | 7 | 7 | 100% | Inter |
| LA1274×LA2732 | *S.cor* | *S.per* | 10 | 7 | 70% | n/a |
| LA0153×LA1954 | *S.per* | *S.per* | 10 | 10 | 100% | Inter |
| LA0153×LA2732 | *S.per* | *S.per* | 9 | 8 | 89% | Intra |
| LA2732×LA2930 | *S.per* | *S.chi* | 7 | 4 | 57% | n/a |
| LA2732×LA0153 | *S.per* | *S.per* | 10 | 8 | 80% | Intra |
| LA2732×LA2964 | *S.per* | *S.per* | 10 | 3 | 30% | Inter |
| LA2732×LA1954 | *S.per* | *S.per* | 9 | 7 | 78% | Inter |
| LA1954×LA2964 | *S.per* | *S.per* | 10 | 5 | 50% | Intra |
| LA1954×LA2732 | *S.per* | *S.per* | 10 | 10 | 100% | Inter |
| LA1954×LA1274 | *S.per* | *S.cor* | 10 | 0 | 0% | n/a |
| LA1954×LA0153 | *S.per* | *S.per* | 5 | 5 | 100% | Inter |
| LA2964×LA0153 | *S.per* | *S.per* | 10 | 9 | 90% | Inter |
| LA2964×LA2732 | *S.per* | *S.per* | 10 | 2 | 20% | Inter |
| LA2930×LA1932 | *S.chi* | Hybrid | 5 | 0 | 0% | n/a |

## 5.6. Hybrid viability

The viability of pollen from *S. peruvianum*, *S. chilense*, and the hybrid population LA1932 was tested using Alexander stain. Eight populations were tested in total, including pollen which had been frozen at -20°C. There were no significant differences between accessions (Chi-sqr test P>0.05) nor between fresh and frozen pollen (Wilcox test P>0.05).

**Table 5.6.1.** Pollen viability in the tested accessions.

| *Population* | *Viable* | *Non-viable* | *% Viable* |
|---|---|---|---|
| LA0153 | 253 | 14 | 95% |
| LA1274 | 132 | 9 | 94% |
| LA1954 | 162 | 6 | 96% |
| LA2732 | 86 | 5 | 95% |
| LA2930 | 213 | 13 | 94% |
| LA2964 | 289 | 16 | 95% |
| LA1932* | 186 | 6 | 97% |
| LA1954* | 216 | 13 | 94% |
| LA2930* | 126 | 10 | 93% |

*Pollen was frozen at -20°C

## 5.7. Additional tomato crossing data

Seeds of *S. peruvianum* mature 21 d after pollination. Seed maturity is followed by fruit development. For the crosses in 2016, fruits developed for a minimum of 50d post pollination. For crosses in 2017, fruits developed for a minimum of 28 d. The crossing method is detailed in Figure 5.7.1. The results of the 2016 crosses are in Figure 5.7.2. *Currently, these data represent crosses included in Chapter 3. However, the data is to be updated in the final version of the thesis with additional crossing data available 9/2017*.

**Figure 5.7.1.** Crossing methods and examples of seeds vs. seed-like structures. (**a**) Pollen is first collected in the cap of a 1.5 mL tube. (**b**) The stigma is dipped into the pollen. (**c**) Stigma with pollen. (**d**) Flowers are then bagged which helps identify them and protect against further cross pollination. (**e**) Examples of fully developed seeds and seed-like structures. (**f**) Visual overview of different compatibilities using accession LA1954 as an example: fully compatible (all seeds and no seed-like structures, LA2732 × LA1954), mostly compatible (LA1954 × LA2732), half-way compatible (LA1954 × LA0153), and fully incompatible (LA1954 × LA1932). Fruits of LA2732 × LA1954 (**g**) and LA2732 × LA2964 (**h**).

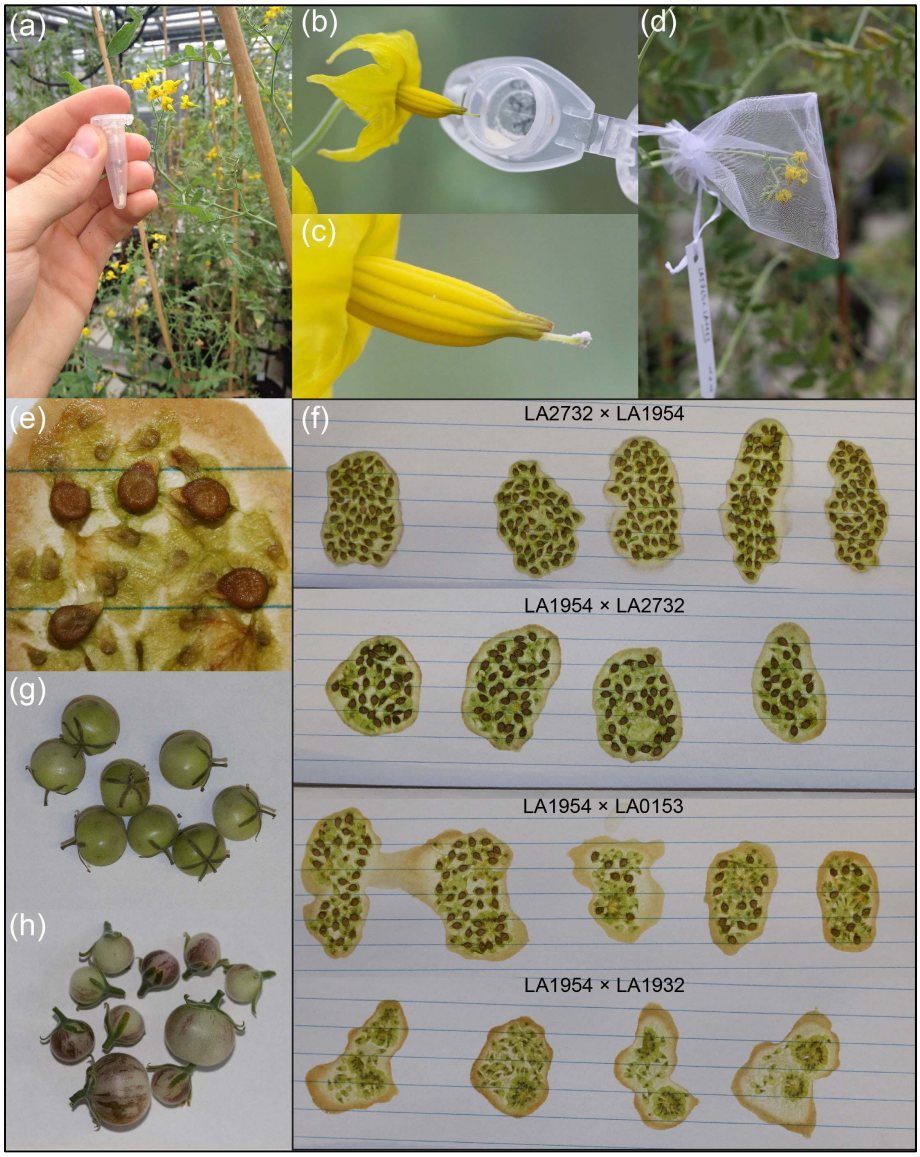


**Figure 5.7.2** The number of seeds (top) and seed-like structures (bottom) for all tested crosses are given. Crosses involving the hybrid population and the parental species are in red (left). Crosses between species are in blue (middle). Crosses within a species are in green (right). *Currently, these data represent crosses included in Chapter 3. However, the data is to be updated in the final version of the thesis with additional crossing data available 9/2017*.