

**THE GENETICS OF CONTEMPORARY EVOLUTION  
IN AN INVASIVE PERENNIAL SUNFLOWER**

by

Dan Gabriel Bock

B.Sc., Babeş-Bolyai University, Cluj-Napoca, Romania, 2008

M.Sc., University of Windsor, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2017

© Dan Gabriel Bock, 2017

## Abstract

While it is now widely accepted that contemporary evolution is common, we have limited information on the genetic underpinnings of these transitions. Here, I investigate this topic in invasive perennial sunflowers. First, I review the plant evolutionary biology literature to assess the validity of a common assumption, that plant organelle genome variation is selectively neutral. I show that organelle-encoded adaptations are likely, and supported both theoretically and empirically. I then rely on genome skimming to clarify the origin of *Helianthus tuberosus*, a hexaploid perennial sunflower and the study system I used for the rest of this thesis. Based on phylogenomic evidence, I show that *H. tuberosus* is an auto-allopolyploid that formed by hybridization between diploid and auto-tetraploid perennial sunflowers. This study provides an early example of the use of genome skimming for the identification of the progenitors of polyploid taxa, and facilitates studies on genome reorganization following polyploid speciation. Finally, I investigate the genetic architecture of invasiveness in *H. tuberosus*. I use genomic data to show that invasive genotypes originated repeatedly, and that most derive from hybridization between native and cultivated material. I then combine information from the greenhouse and from a replicated common garden, and show that increased clonal propagation is a major invasiveness trait in this system. I present evidence that high invasiveness in *H. tuberosus* can be achieved by hybridization and heterosis, or independent of hybridization, through the action of two major additive effect loci. Moreover, I find that these different genetic mechanisms can act synergistically, and that both have been exploited by widespread invasive clones. Collectively, these results show that invasiveness can be achieved *via* multiple genetic routes in the same system and during the same biological invasion event. These findings contribute to our understanding of the diverse genetic basis of contemporary evolutionary transitions.

## **Lay abstract**

With this dissertation, I aimed to contribute to our understanding of the genetic basis of rapid evolution. The first study summarizes evidence that DNA variation maintained in plant organelle genomes can be adaptive, and may be involved in these contemporary transitions. In the second study, I use a technique known as genome skimming to clarify the ancestry of the sunflower *Helianthus tuberosus*, the species I focused on for the rest of this thesis. In the third study, I investigate the genetic basis of invasiveness in *H. tuberosus*. I first rely on genetic data to clarify the origin of invasive European genotypes. I then use trait data to show that invasiveness in this system is achieved through the production of a large number of tubers. Finally, I combine genetic and phenotype data to show that invasiveness in *H. tuberosus* has evolved through two distinct genetic mechanisms.

## Preface

From Chapter 1, section 1.2 has been published in:

- Bock DG, Caseys C, Cousens R, Hahn MA, Heredia SM, Hübner S, Turner KG, Whitney K, Rieseberg LH (2015) What we still don't know about invasion genetics. *Molecular Ecology*, **24**, 2277–2297.

Loren H. Rieseberg organized the reading group and that contributed this review. All authors performed the writing. I contributed the section on genomic variation, and that on genetic architecture (which was included in this thesis). Authors are arranged alphabetically, except LH Rieseberg.

Chapter 2, the summary for this chapter (section 1.3), and the future directions related to this chapter (section 5.1) have been published as:

- Bock DG, Andrew RL, and Rieseberg LH (2014) On the adaptive value of cytoplasmic genomes in plants. *Molecular Ecology*, **23**, 4899–4911.

All authors developed the plan for this study. As the first author, I researched the relevant literature and led the writing, with revision advice from the two co-authors.

Chapter 3, as well as the summary related to this work (section 1.3), have been published as:

- Bock DG, Kane NC, Ebert DP, and Rieseberg LH (2014) Genome skimming reveals the origin of the Jerusalem artichoke tuber crop species: neither from Jerusalem nor an Artichoke. *New Phytologist*, **201**, 1021–1030.

All authors contributed to the design of this study. As the first author, I performed all data collection and analyses, and wrote the paper with input from my co-authors.



For Chapter 4, I developed the plan for the study together with Loren H. Rieseberg, who also provided guidance and feedback throughout the course of the study. I performed the sampling, greenhouse propagations, phenotyping, molecular work, and bioinformatic analyses. This work was, however, highly collaborative, and could not have been completed without the help of a large number of graduate and undergraduate volunteers and research assistants, listed in the Acknowledgements. Substantial contributions were made by Michael B. Kantar and Mirela G. Bock, who helped with plant propagation, the harvest the common garden experiment, and with phenotyping. Mikey also provided analysis advice at various stages of the project. Also, the allelopathy experiments were performed in collaboration with Céline Caseys. Lastly, I developed the strategy for analyzing the phenotype data together with Remi Matthey-Doret, who then performed the ANOVA and MCMCglmm analyses.

# Table of Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Preface</b> .....	<b>iii</b>
<b>Table of Contents</b> .....	<b>v</b>
<b>List of Tables</b> .....	<b>ix</b>
<b>List of Figures</b> .....	<b>x</b>
<b>List of Abbreviations</b> .....	<b>xiii</b>
<b>Acknowledgements</b> .....	<b>xv</b>
<b>Dedication</b> .....	<b>xvii</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1    Background.....	1
1.2    The genetics of biological invasions.....	3
1.2.1    The top-down approach.....	4
1.2.2    The bottom-up approach.....	5
1.2.3    Conclusions.....	7
1.3    Breakdown of chapters.....	8
<b>Chapter 2: On the adaptive value of cytoplasmic genomes in plants</b> .....	<b>10</b>
2.1    Introduction.....	10
2.1.1    Purifying selection in organelle genomes.....	11
2.2    Arguments for the neutrality of plant organelle DNA variation.....	13
2.2.1    Nonsynonymous DNA polymorphism should be rare in organelle genomes.....	13
2.2.2    Organelle genomes have limited coding potential.....	14

2.2.3	Mutation rates are reduced for plant organelle DNA.....	15
2.2.3.1	Variation in organelle DNA mutation rates.....	16
2.3	Evidence for an adaptive value of plant cytoplasm.....	18
2.3.1	Observational evidence: studies of organelle genome capture.....	19
2.3.2	Experimental evidence: studies of cytonuclear interactions.....	22
2.3.3	Statistical evidence: studies of positive selection at the molecular level.....	24
2.4	Conclusions.....	26

**Chapter 3: Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop**

**species: neither from Jerusalem nor an Artichoke.....31**

3.1	Introduction.....	31
3.2	Materials and Methods.....	35
3.2.1	Molecular techniques.....	35
3.2.2	Assembly of plastid and mitochondrial genomes.....	35
3.2.3	Assembly of 35S and 5S rDNA regions.....	36
3.2.4	Alignment and phylogenetic analyses.....	37
3.2.5	Survey of diagnostic polymorphism in rDNA data.....	38
3.3	Results and Discussion.....	39
3.4	Conclusions.....	44

**Chapter 4: Multiple genetic routes to the evolution of invasiveness in a perennial**

**sunflower.....51**

4.1	Introduction, Results and Discussion.....	51
4.2	Detailed materials and methods.....	59
4.2.1	Sampling.....	59

4.2.1.1	Public collections – <i>H. tuberosus</i> .....	59
4.2.1.2	Public collections – progenitor species.....	59
4.2.1.3	Sampling of invasive populations.....	60
4.2.2	Ploidy inference.....	60
4.2.3	Genotyping-by-sequencing.....	60
4.2.4	Alignment of sequence data and variant calling.....	61
4.2.5	Variant filtration and genotyping error rates.....	62
4.2.6	Clonal relationships in the <i>H. tuberosus</i> collection.....	63
4.2.7	Confirmation of <i>H. tuberosus</i> ancestry and taxonomic identifications.....	64
4.2.8	Isolation of <i>H. tuberosus</i> subgenome-specific markers.....	65
4.2.9	<i>H. tuberosus</i> phylogenetic analyses.....	67
4.2.10	<i>H. tuberosus</i> population genetic structure.....	68
4.2.11	Classification of native, cultivated, and invasive <i>H. tuberosus</i> .....	68
4.2.12	Identification of invasiveness traits in <i>H. tuberosus</i> .....	72
4.2.12.1	Phenotype data collection.....	72
4.2.12.2	Phenotype data analyses.....	76
4.2.12.3	Effect size estimates and comparisons with previous studies.....	78
4.2.13	Genetic architecture of invasiveness.....	79
4.2.13.1	Relationship between heterozygosity and trait values.....	79
4.2.13.2	Association mapping.....	80
<b>Chapter 5: Conclusions.....</b>		<b>88</b>
5.1	Chapter 2 future directions.....	88
5.2	Chapter 3 future directions.....	90

5.3	Chapter 4 future directions.....	91
<b>References.....</b>		<b>94</b>
<b>Appendices.....</b>		<b>125</b>
Appendix A - Chapter 3 Supplementary Material .....		125
Appendix B - Chapter 4 Supplementary Material .....		136

## List of Tables

Table 2.1: Studies reporting accelerated rates of nucleotide substitution in plant organelle genomes .....	28
Table 2.2: Studies using neutrality tests to infer positive selection at organelle genes.....	29
Table A.1: Details of samples used for sequencing.....	133
Table A.2: Summary of sequencing data.....	134
Table A.3: Total reads and mean coverage for high-copy regions.....	135
Table B.1: Sampling information for all <i>H. tuberosus</i> accessions.....	154
Table B.2: Sampling for <i>H. grosseserratus</i> , <i>H. divaricatus</i> , and <i>H. hirsutus</i> accessions.....	170
Table B.3: Traits measured.....	172
Table B.4: Results of fixed-ANOVAs performed for the between-species comparisons for allelopathy metrics.....	174
Table B.5: Results of fixed-ANOVAs for comparisons at the six allelopathy metrics performed within <i>H. tuberosus</i> .....	175
Table B.6: Results of mixed-ANOVAs for 13 common garden traits scored in <i>H. tuberosus</i> ....	176
Table B.7: Results of MCMCglmm.....	178
Table B.8: Hedges' effect size estimates calculated in <i>H. tuberosus</i> for tuber number and three domestication traits, as well as for 18 invasiveness traits identified from other studies.....	180
Table B.9: Results from the K + Q association mapping models.....	181
Table B.10: Clonal series identified for <i>H. tuberosus</i> samples.....	182
Table B.11: Strategy used for diploidization and tetraploidization of subgenome SNPs.....	184

## List of Figures

Figure 2.1: Example of experimental confirmation of the adaptive contribution of the plant plasmotype.....	30
Figure 3.1: Geographic distribution of the perennial <i>Helianthus</i> accessions sequenced.....	46
Figure 3.2: Phylogenetic reconstruction of plastid and mitochondrial genomes.....	47
Figure 3.3: Pairwise sequence divergences calculated within species for whole-plastome haplotypes and partial mitochondrial haplotypes.....	48
Figure 3.4: Phylogenetic reconstruction of concatenated rDNA sequences.....	49
Figure 3.5: Survey of species-diagnostic rDNA polymorphism.....	50
Figure 4.1: Population structure and genetic diversity of <i>Helianthus tuberosus</i> .....	82
Figure 4.2: Phenotypic determinants of invasiveness in <i>Helianthus tuberosus</i> .....	84
Figure 4.3: Genetic architecture of invasiveness in <i>Helianthus tuberosus</i> .....	86
Figure A.1: Number of SNPs called vs. ploidy setting for the 35S segment.....	125
Figure A.2: Number of SNPs called vs. ploidy setting for the 5S segment.....	126
Figure A.3: Singletons and shared polymorphism for each surveyed region.....	127
Figure A.4: Unrooted phylogeny of complete plastid genomes.....	128
Figure A.5: Unrooted phylogeny of partial mitochondrial genomes.....	129
Figure A.6: Unrooted phylogeny of concatenated rDNA sequences.....	130
Figure A.7: Phylogenetic reconstruction of plastid and mitochondrial genomes for diploids....	131
Figure A.8: Phylogenetic reconstruction of 35S rDNA and 5S rDNA sequences.....	132
Figure B.1: Map of North America showing <i>H. tuberosus</i> USDA sampling locations.....	136
Figure B.2: Map of Europe showing invasive <i>H. tuberosus</i> sampling locations.....	137

Figure B.3: Clonality in the <i>H. tuberosus</i> collection.....	138
Figure B.4: PCA including all <i>Helianthus</i> taxa based on 2,916 shared SNPs.....	139
Figure B.5: PCA including only samples sequenced using WGSS based on 3,916 SNPs.....	140
Figure B.6: PCA including the 427 <i>H. tuberosus</i> non-clonal genotypes based on the filtered hexaploid SNP set as well as those assigned to the two subgenomes.....	141
Figure B.7: Phylogeny including the 427 <i>H. tuberosus</i> non-clonal genotypes based on 4,700 SNPs assigned to the diploid subgenome.....	142
Figure B.8: Pedigree information for cultivated samples clustering with invasive genotypes...	143
Figure B.9: Bayesian STRUCTURE analysis of 427 <i>H. tuberosus</i> samples based on 5,000 SNPs randomly selected from the filtered hexaploid SNP set.....	144
Figure B.10: Bayesian STRUCTURE analysis of 427 <i>H. tuberosus</i> samples based on 4,700 SNPs assigned to the diploid subgenome.....	145
Figure B.11: Bayesian STRUCTURE analysis of 427 <i>H. tuberosus</i> samples based on 4,770 SNPs assigned to the tetraploid subgenome.....	146
Figure B.12: Heterozygosity calculated based on the filtered hexaploid SNP set in 5MB non-overlapping windows along the genome.....	147
Figure B.13: Genome-wide heterozygosity calculated for 427 <i>H. tuberosus</i> samples across 82,957 SNPs with genotypes in hexaploid format.....	148
Figure B.14: PCA based on 11 quantitative traits for 305 <i>H. tuberosus</i> samples included in the greenhouse and common garden experiments.....	149
Figure B.15: Means and 95% bootstrapped confidence intervals for six metrics of allelopathy in <i>H. tuberosus</i> and progenitor species.....	150



Figure B.16: Means and 95% bootstrapped confidence intervals for 13 traits measured in a common garden for *H. tuberosus*..... 151

Figure B.17: Pearson correlation coefficients between all trait values and heterozygosity estimated based on all markers, diploid subgenome markers, and tetraploid subgenome markers..... 152

Figure B.18: Correlations between tuber number and heterozygosity of whole-genome genotypes for samples assigned to the native, cultivated, or invasive categories..... 153

## List of Abbreviations

AIC	Akaike information criterion
ANOVA	Analysis of variance
BLAST	Basic local alignment search tool
CNI	Cytonuclear interactions
DNA	Deoxyribonucleic acid
ETS	External transcribed spacer
GATK	Genome Analysis Toolkit
GBS	Genotyping by sequencing
GWA	Genome-wide association
IBS	Identity-by-state
ILS	Incomplete lineage sorting
INRA	French National Institute for Agricultural Research
IPK	The Leibniz Institute of Plant Genetics and Crop Plant Research
ITS	Internal transcribed spacer
IUPAC	International Union of Pure and Applied Chemistry
LG	Linkage group
MB	Megabase
MCMC	Markov chain Monte Carlo
MQ	Mapping quality
mtDNA	Mitochondrial DNA
NSERC	National Science and Engineering Research Council

NTS	Nontranscribed spacer
PCA	Principal component analysis
PCR	Polymerase chain reaction
PGRC	Plant Gene Resources of Canada
QTL	Quantitative trait locus
QUAL	Quality
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
USDA	United States Department of Agriculture
VCF	Variant call format
WGSS	whole genome shotgun sequencing

## Acknowledgements

First and foremost, I would like to thank my advisor Loren Rieseberg. Not only that Loren was always available to give tremendously valuable research advice, but he is among the most patient and supportive people I know. Working in Loren's group for the past six years has been truly transformative, I couldn't have hoped for a more rewarding experience. I also thank the members of my advisory committee, Keith Adams and Quentin Cronk, for their timely advice and feedback.

One of the aspects that I most enjoyed about the research performed for this dissertation is that it was collaborative. I therefore have a large number of people to thank for their contribution to different components of this study, and I apologize if I forgot to mention anyone. USDA samples were obtained with the help of Laura Marek. PGRC samples were obtained with the help of Dallas Kessler and Axel Diederichsen. At IPK, sampling was facilitated by Markus Oppermann and Helmut Knuepffer. At INRA, samples were obtained with the help of Nicolas Langlade and Anne Zanetto. Invasive samples were collected with help from Mirela Bock, Cristina Bock, Iulian Bock, Alexandru Stermin and Rita Filep.

Molecular and greenhouse work was performed with assistance from Christine Phillips, Chloe Tashlin Fluegel, and Caroline Ramsay. My former officemate Saemundur Sveinsson and Chris Sears introduced me to flow cytometry. Andy Johnson provided technical flow cytometry advice. Jaroslav Doležel provided seeds for flow cytometry standards. The allelopathy bioassays were performed with the help of Céline Caseys, Pengfei Qiao, Anita Poon, Stella Li, Kate MacDonald, Kelly Borkowski, and Madeline Iseminger. I would also like to thank Melina Biron for substantial help with greenhouse work, and Anastasia Kuzmin for assisting with sequencing.

The common garden experiment was performed with help from Michael Kantar, Mirela Bock, Olga Helmy, Jeddediah Brodie, Dylan Burge, Sariel Hubner, Sylvia Heredia, Jess Henry, Winnie Cheung, Kasey Moran, Min Hahn, Kate Ostevik, Marco Todesco, Emily Drummond, Mariana Pascual, John Lee, Kurtis Baute, Greg Baute, Kasia Stepien, Azra Lalji, Wonbin Choi, Dominique Skonieczny, Ismael Hamadi, Vincent Cha, and Christina Jewell, and Seane Trehearne.

Members of the Rieseberg lab helped shape my thinking about evolutionary biology, and introduced me to genomics and bioinformatics. I greatly enjoyed getting to know all Rieseberg-lars, including but not limited to Rose Andrew, Greg Baute, Dylan Burge, Céline Caseys, Emily Drummond, Dan Ebert, Chris Grassa, Sylvia Heredia, Nolan Kane, Michael Kantar, Brook Moyers, Greg Owens, Kate Ostevik, Sébastien Renaut, Marco Todesco, and Kathryn Turner. Special thanks to Rose Andrew, Nolan Kane, Greg Baute, Chris Grassa and Sariel Hubner for bioinformatics advice, and to officemate Remi Matthey-Doret for general statistics advice.

My loving wife, Mirela Bock (née Metea), has not only directly contributed to this project, but she has, at many times, lifted my spirits when experiments did not go as planned. Thank you for all the sacrifices you make for me, you are the most generous person I know. I would also like to take this opportunity to thank my family, especially mom (Cristina), dad (Iulian), and sister (Alexandra), for their continued support throughout my academic career.

Lastly, I am tremendously thankful for the generous financial support that I received during this research, including a UBC Four Year Doctoral Fellowship, a Killam Doctoral Fellowship, and an NSERC Vanier Canada Graduate Scholarship.

## **Dedication**

To my family, both near and far, who constantly encouraged me during this adventure.

# Chapter 1: Introduction

## 1.1 Background

Ecological and evolutionary forays into the mechanisms that generate and maintain biological diversity have traditionally been considered as distinct enterprises. This is because, for much of the past century, the prevailing view has been that these processes are relevant at very different timescales (Slobodkin 1961). Compared to contemporary ecological change, evolutionary transitions were principally assumed to occur over millennia or longer. This paradigm has been lost during the past two decades, due to abundant reports of dramatic heritable shifts in phenotype that occur over only a few years or generations (e.g. Conover & Munch 2002; Grant & Grant 2002; Koskinen et al. 2002; Stuart et al. 2014).

Currently, the study of contemporary evolution is one of the most active areas of biological research. Several important topics are being investigated. For instance, there is great interest in understanding the interplay between rapid evolution and ecological speciation (reviewed in Hendry et al. 2007). The contemporary (i.e. occurring over tens to hundreds of generations) evolution of reproductive isolation *via* adaptive divergence is bolstered by mathematical models (e.g. Fry 2003) and examples from nature, in both plants (e.g. Rieseberg et al. 2003; James & Abbott 2005) and animals (e.g. Bearhop et al. 2005). Limited information is available, however, on the relative speed and types of contributing reproductive barriers, the strength of selection, or the genetic architectures involved (Hendry et al. 2007).

Recent work has also investigated rapid evolution occurring within even narrower time frames, in the context of bi-directional interactions between ecological and evolutionary processes (i.e. eco-evolutionary feedbacks; Pelletier et al. 2009; Schoener 2011). Here, important

unknowns concern the relative contribution of evolutionary (e.g. changes in adaptive traits) vs. ecological parameters (e.g. predation pressure) in determining population growth rates and higher ecosystem-level effects (Schoener 2011). Also, with increasing numbers of empirical examples that use genomics to dissect the adaptive traits involved, the hope is that eco-evolutionary dynamics will become increasingly predictable (Rodríguez-Verdugo et al. 2017).

As noted above, a recurring theme for contemporary evolution research has been understanding the genomic basis of rapidly evolving traits. We would like to know, for instance, the number and effect sizes of contributing loci (e.g. Jones et al. 2012; Levy et al. 2015), and their genetic origin (e.g. Yeaman et al. 2016). As well, important progress is to be made towards understanding the extent to which these changes are repeatable between parallel evolutionary transitions within taxa (e.g. Pascoal et al. 2014; Lescak et al. 2015) or between taxa that are exposed to similar selection pressures (Yeaman et al. 2016). Answers to these questions have important fundamental and applied bearing. From a fundamental standpoint, generalizations are expected to help guide the parameterization and predictive ability of mathematical models (Messer et al. 2016). From a practical standpoint, understanding the conditions that facilitate or impede rapid evolution can help guide the management of taxa exposed to novel environments (Stockwell et al. 2003).

The genetic study of contemporary evolution can be approached from multiple angles. One of these, commonly referred to as the 'evolve and resequence' approach (Schlötterer et al. 2014), involves interrogating the genomes of experimental populations that are grown under controlled laboratory settings. These studies capitalize on the availability of genome-enabled experimental organisms that also have rapid generational turnover, such as yeast or fruit flies (e.g. Levy et al. 2015; Graves et al. 2017). Naturally occurring populations of non-indigenous



species have frequently been proposed as an alternative study system, in the context of the rapid evolution of invasive potential (e.g. Prentis et al. 2008; Bock et al. 2015; Colautti & Lau 2015). This is not surprising, given that a large fraction of early examples of contemporary evolution involved long-distance human-mediated introductions (e.g. Williams & Moore 1989; Weber & Schmid, 1998; Lee 1999; Huey et al. 2000). Compared to the 'evolve and resequence' approach, the study of biological invasions offers the possibility of a wider snapshot, both in terms of evolutionary time and range of organisms that can be considered. Repeated introductions to geographically disparate areas are common, and so these natural experiments may frequently be replicated (e.g. Huey et al. 2000).

With this thesis, I took this latter approach. Specifically, using an invasive species as a study system, I aimed to contribute to our understanding of the genetic underpinnings of contemporary evolution. I first place this work within the context of current research on the genetics of biological invasions. I then briefly discuss the focus of each of the component studies.

## **1.2 The genetics of biological invasions**

Understanding the genetic and molecular mechanisms that underlie the formation of invasive genotypes has been a central goal of invasion genetics since the inception of the field (Baker & Stebbins 1965), yet knowledge on the topic remains limited. The few currently available examples indicate that invasiveness is often underpinned by a small number of genes. Moreover, rapid evolution in invasive taxa does not appear to be mutation limited. Below, I discuss the genetic architecture of invasiveness in the framework of two general approaches, top-down (or forward) genetics and bottom-up (or reverse) genetics.

### 1.2.1 The top-down approach

The top-down approach starts with knowledge on the phenotypic traits that vary between invasive and noninvasive genotypes, or that have been targets of selection during the evolution of invasiveness. The task then becomes to identify loci that underlie those traits. This can be achieved through candidate gene analyses and through genomewide association or quantitative trait locus (QTL) mapping.

In some cases, dissecting the genetic basis of invasiveness can be relatively straightforward, if a list of candidate genes known to affect the phenotypes under investigation is available. Some of the best-known invasiveness genes come from studies in this category. One example comes from studies of the fire ant (*Solenopsis invicta*), in which multi-queened introduced populations are more ecologically destructive and show less aggression to conspecifics than single-queened native populations (Porter & Savignano 1990). Krieger & Ross (2002) were able to identify Gp-9, a gene that encodes an odorant-binding protein, as the locus underlying polymorphism in this social behaviour in *S. invicta*. Another example is the dopamine receptor D4 gene, which is associated with novelty seeking and activity behaviour in introduced populations of yellow-crowned bishops (Mueller et al. 2014).

More often than not, no information is available on the likely genetic underpinnings of invasiveness. In this case, efforts have been directed towards finding associations between genetic markers and phenotypes of interest in pools of unrelated individuals, or in experimental populations derived from crosses between parents that show extreme trait values.

This latter approach, known as QTL-mapping, has been used with some success in weed genomics (Basu et al. 2004). In allopolyploid invasive Johnson grass (*Sorghum halepense*),

Paterson et al. (1995) used crosses between the two species progenitors to understand the genetic basis of rhizomatousness, a weediness trait in this system. A small number of QTLs, most of which show additive or dominant gene action, were identified. More recently, Whitney et al. (2015) investigated loci involved in adaptive introgression associated with range expansion in the natural hybrid sunflower *H. annuus texanus*. Three donated QTLs were found that increased components of male and female fitness in the recipient species, likely as pleiotropic effects of phenological and architectural trait QTLs that colocalized with the fitness QTLs.

### **1.2.2 The bottom-up approach**

The bottom-up approach does not require prior knowledge on traits that contribute to the propensity to invade. Instead, this strategy involves searching for changes in gene expression or allele frequency between pools of native and invasive genotypes, and making inferences about the traits involved based on knowledge of gene function.

Transcriptome analyses use microarrays or direct sequencing of RNA to identify genes that are differentially expressed in native and invasive genotypes. Lockwood & Somero (2011), for example, investigated the transcriptional response to low-salinity stress in two species of blue mussels (genus *Mytilus*). One of these, *M. galloprovincialis*, is invasive and has spread along the Pacific coast of California except areas North of Bodega Bay. This area is characterized by lower salinity and is still dominated by the native species *M. trossulus*. The authors performed a microarray analysis of *M. galloprovincialis* and *M. trossulus* individuals grown under benign conditions as well as those simulating abrupt decreases of salinity. Results revealed that most differentially expressed genes in response to salt stress are shared between the two species. Thus, either a small number of genes limit the spread of the invader, or most species-specific

differences in tolerance to osmotic stress are mediated downstream of transcription (Lockwood & Somero 2011).

Similar studies have been performed for invasive plants. Hodgins et al. (2013), for example, examined differential gene expression between native and invasive genotypes of common ragweed (*Ambrosia artemisiifolia*) across 45 062 unigenes. In this case as well, a small fraction of the genes were differentially expressed between native and invasive samples. The functional categories over-represented among the differentially expressed genes were also in agreement with results from a common garden experiment in this system (Hodgins & Rieseberg 2011) and highlighted genes involved in oxidoreductase activity, response to blue light, as well as abiotic and biotic stress response, as strong candidates for invasiveness genes in this system.

At the genome level, bottom-up approaches rely on finding the signature of positive selection, which can include regions that show high levels of genetic differentiation or shifts in the site frequency spectrum of mutations. Puzey & Vallejo-Marín (2014), for example, performed one such genome scan analysis to detect the signature of positive selection during the invasion of monkeyflowers (*Mimulus guttatus*) in the UK. While a specific target of selection was not identified, genes located in swept regions were shown to be associated with flowering time, as well as biotic and abiotic stress (Puzey & Vallejo-Marín 2014). Moreover, two of these regions were positioned near or at a chromosomal inversion polymorphism associated with a number of morphological and life history differences in monkeyflowers (Puzey & Vallejo-Marín 2014).

In another recent example, Vandepitte et al. (2014) investigated the genetic basis of adaptation following the 1824 introduction of the Pyrenean rocket (*Sisymbrium austriacum* subsp. *chrysanthum*) in Belgium using native, contemporary invasive samples and herbarium specimens collected in the introduced area. Six genes involved in flowering were identified as outliers of

genetic differentiation and experienced allele frequency changes over the course of the invasion process.

A concern with the bottom-up approach is false positives, which can arise due to nonequilibrium demographic histories (Lotterhos & Whitlock 2014), as well to genomic heterogeneity in mutation and recombination rates (Renaut et al. 2014). These issues can be especially problematic in invaders, as generally little is known about their genomes. Also, as previously discussed, populations at the invasion front undergo extreme drift, allowing neutral and deleterious alleles to surf to high frequency, mimicking the signature of selection. Further, the loci identified as ‘invasion loci’ remain hypotheses until further work confirms that they control actual invasiveness in the field.

### **1.2.3 Conclusions**

The small number of studies investigating the genetic architecture of invasiveness currently precludes the making of many generalizations. It is unclear, for example, whether and how often the genetic architecture of invasiveness traits differs from that of other traits differentiating natural populations or species. For example, are recessive QTLs more frequently established in invasive populations? Theory predicts that the probability of fixation for advantageous mutations is higher if they are dominant (Haldane’s sieve; Turner 1977). Because of frequent bottlenecks, this process might be less effective in invasive populations. Also, the extent to which evolution re-uses the same genes or genomic regions during the evolution of invasiveness remains unclear.

### 1.3 Breakdown of chapters

Chapter 2 is a review that attempts to answer the question of whether DNA variation maintained in plant organelle genomes is selectively neutral. The answer to this question has implications that range from inferences of evolutionary history (e.g. Percy et al. 2014) to contemporary adaptation (e.g. Bashalkhanov et al. 2013). While plant organelle variation has traditionally been assumed to be neutral, recent studies in animals have shown that, on the contrary, mitochondrial DNA polymorphism is frequently adaptive (e.g. Ballard & Whitlock 2004; Dowling et al. 2008; Galtier et al. 2009). In plants, however, the neutrality assumption has not been strongly challenged. I begin with a critical evaluation of arguments in favor of this long-held view. I then discuss the latest empirical evidence for the opposing prediction that sequence variation in plant cytoplasmic genomes is frequently adaptive. While outstanding research progress is being made towards understanding this fundamental topic, I highlight the need for studies that combine information ranging from field experiments to physiology to molecular evolutionary biology. Such an interdisciplinary approach provides a means for determining the frequency, drivers and evolutionary significance of adaptive organelle DNA variation.

Chapter 3 presents work that I performed to understand the origin of *H. tuberosus* (Jerusalem artichoke), the study system used in the bulk of this thesis. Despite the cultural and economic importance of this hexaploid tuber crop, its origin is debated. Competing hypotheses implicate the occurrence of polyploidization with or without hybridization, and list the annual sunflower *H. annuus* and five distantly related perennial sunflower species as potential parents. I test these scenarios by skimming the genomes of diverse populations of Jerusalem Artichoke and its putative progenitors. I identify relationships among *Helianthus* taxa using complete plastomes

(151 551 bp), partial mitochondrial genomes (196 853 bp) and 35S (8196 bp) and 5S (514 bp) ribosomal DNA. The results I present refute the possibility that Jerusalem Artichoke is of *H. annuus* ancestry. I provide the first genetic evidence that this species originated recursively from perennial sunflowers of central-eastern North America *via* hybridization between tetraploid Hairy Sunflower and diploid Sawtooth Sunflower.

Chapter 4 describes work that I performed to understand the origin of invasive *H. tuberosus*, the phenotypic basis of invasive potential, as well as the genetic architecture of invasiveness in this system. I first clarify the origin of invasive genotypes using genotyping-by-sequencing data and population genetic analyses. I then describe a common garden and a greenhouse experiment that were used to show that clonality is a major invasiveness trait in this species. I further demonstrate that invasiveness in *H. tuberosus* can result from hybrid vigor and/or the action of two major additive-effect loci. I find that these non-exclusive genetic mechanisms can act synergistically, and that both have been exploited during the recent European range expansion of this species. These results therefore collectively demonstrate that even during the same biological invasion event, multiple genetic solutions may exist for the evolution of invasiveness.

In the final chapter of the thesis I present general conclusions that can be drawn based on the research described in this thesis. I also highlight potentially fruitful directions for future research.

## Chapter 2: On the adaptive value of cytoplasmic genomes in plants

### 2.1 Introduction

For more than four decades, phylogenetic and phylogeographic studies in animals and plants have relied heavily on variation in organellar genomes (Avice et al. 1979; Scowcroft 1979), with the assumption that sequence polymorphism maintained at the level of the plasmotype is selectively neutral. This assumption is based to a large extent on the fact that organelle genes have repeatedly been shown to evolve under strong purifying selection (see section 2.1.1 below) and – in agreement with the neutral theory of molecular evolution (Kimura 1983) – under such conditions the fate of persisting variation should be dominated by genetic drift. While purifying selection does not exclude the possibility of positive selection, neutrality conditions for organelle DNA variation are often implicitly assumed to be met.

Studies in animals have increasingly contested the generality of this assumption (Ballard & Whitlock 2004; Dowling et al. 2008; Galtier et al. 2009; Balloux 2010). First, parallels have been reported between mitochondrial DNA (mtDNA) haplotype frequencies and natural or laboratory-manipulated conditions (reviewed in Toews & Brelsford 2012). Second, functional differences with potential fitness implications have been assigned to naturally-occurring mtDNA variants (Toews et al. 2013; reviewed in Ballard & Melvin 2010). Third, the signs of positive selection at mtDNA have been detected using neutrality tests (Ruiz-Pesini et al. 2004; Bazin et al. 2006; da Fonseca et al. 2008; Llopart et al. 2014).

With some exceptions (Budar & Roux 2011; Greiner & Bock 2013), the plant literature remained relatively quiescent to these developments. Assessing the validity of the neutrality assumption for plant organelle genetic variation is opportune for two reasons. First, lessons from



plant physiology teach us that if organelle genomes are under positive selection, organelle-encoded adaptations are likely to be involved in traits of major conservation, ecological, and economical importance, especially in the context of accelerating climate warming, such as tolerance to drought, light, and salt stress (Atkin & Macherel 2009; Chaves et al. 2009). Second, advances in sequencing technology have contributed to increased interest in using complete organelle genomes in studies of phylogenetics, phylogeography, and population genetics (Straub et al. 2012; Bock et al. 2014; Mariac et al. 2014). Non-neutrality of organelle DNA polymorphism might, in this case, have important bearing on our ability to infer evolutionary processes from patterns of genomic variation.

We begin this review with a critical evaluation of arguments at the center of the assumption that plant organelle DNA variation is neutral. We then discuss why theory predicts adaptive evolution at plant organelle DNA is possible, and highlight the strengths and limitations of the latest supporting empirical evidence. Some aspects of our argumentation are necessarily based on a limited number of examples currently available. Nevertheless, we hope information presented here can serve as an impetus for future studies aiming to advance our understanding of this fundamental topic further.

### **2.1.1 Purifying selection in organelle genomes**

Mitochondria and chloroplasts are firmly positioned at the hub of cellular metabolism. The critical importance of both organelles and the genes they retain has been confirmed repeatedly by observations that organelle malfunction and minute changes at organelle DNA can have severely debilitating consequences, and may even culminate in lethality (Wallace 2005; Greiner 2012). These and other observations, such as the fact that organelle genomes are highly

conserved in structure, have contributed to the view that purifying selection is the predominant force shaping organelle DNA evolution.

This assumption has been substantiated by multiple lines of evidence. For one, early comparisons of rates of synonymous and nonsynonymous substitution at organelle DNA reported an excess of nonsynonymous changes within species, as compared to those detected between species (Nachman 1998; Rand & Kann 1998). This indicated that many organelle DNA nonsynonymous mutations that contribute to intra-specific polymorphism are evolutionarily ephemeral, and removed by purifying selection before they can accumulate as inter-specific divergences.

Consistent results were provided by empirical studies. For example, in the near absence of natural selection, mutation accumulation experiments reported a considerable increase in the number of nonsynonymous mutations in mtDNA as compared to estimates obtained using phylogenetic comparisons of species pairs (Haag-Liautard et al. 2008). Also, studies using mice mutator lines, which express a proofreading-deficient mitochondrial DNA polymerase, revealed that over the course of only two generations, a large proportion of nonsynonymous changes in organelle protein-coding genes are eliminated (Stewart et al. 2008). Recent evidence from *Drosophila* suggests this rapid purging is achieved via selection at the organelle level, through the preferential propagation of unimpaired haplotypes during oogenesis (Hill et al. 2014).

In plants, examples analogous to mutation-accumulation experiments can be observed in the wild, in species that have made the evolutionary leap to parasitism. By relying on their hosts for nutrient and carbon uptake, parasitic plants have partially or completely shed the need to fix carbon autotrophically *via* photosynthesis, thereby loosening the selective clench on chloroplast genome variation (Krause 2012). Studies of organelle genome evolution in parasitic plants report

extensive genome rearrangements, gene losses, and increased rates of base substitution (Krause 2012). These results are consistent with evolution under relaxed purifying selection, although other factors, including positive selection, reduced effective population sizes, or increased mutation rates, also appear to be at least partially implicated (Bromham et al. 2013).

## **2.2 Arguments for the neutrality of plant organelle DNA variation**

Neutralist interpretations for plant organelle genetic variation can be traced to a series of three arguments. While the first two of these are shared with the animal mitochondrial genome, the third is applicable to organelle genetic variation in plants.

### **2.2.1 Nonsynonymous DNA polymorphism should be rare in organelle genomes**

According to this argument, chloroplast and mitochondrial genomes are unlikely to undergo adaptive evolution, since they should retain limited amounts of nonsynonymous DNA polymorphism within populations (Dowling et al. 2008; Galtier et al. 2009; Budar & Roux 2011). This is because organelle genes are under strong purifying selection (see section 2.1.1), and nonsynonymous mutations that occur in a haploid genome should be continuously exposed to selection.

On careful examination, it is clear that this argument does not fully consider the biology of cytoplasmic genomes, and that other attributes of plastomes and chondriomes suggest there should be scope for nonsynonymous organelle DNA variation. For example, on account of haploidy and generally uniparental inheritance, the effective population size of organelle genomes is reduced relative to that of the nuclear genome (Birky et al. 1983; Dowling et al. 2008). While low effective population size is often associated with a reduction in genetic

diversity (Cutter & Payseur 2013), it is also predicted to reduce the efficiency of selection. Moreover, because of complete linkage, selective interference (Hill & Robertson 1966) should be high in organelle DNA. From this perspective, organelle genomes should behave similarly to regions of the nuclear genome that have a long history of reduced recombination, such as the dot chromosome in *Drosophila*, or the degenerate sex chromosomes of dioecious animals and plants, which show accelerated accumulation of nonsynonymous polymorphisms (Betancourt et al. 2009; Hough et al. 2014).

Empirical evidence supports these predictions. One example is the study by Drouin et al. (2008). The authors surveyed the rates of synonymous and nonsynonymous substitutions in three mitochondrial, five chloroplast, and four nuclear genes for 27 seed plant species. While rates of nonsynonymous polymorphisms were 46 times lower than rates of synonymous polymorphisms for nuclear genes, these differences were considerably reduced for organelle genes. Specifically, nonsynonymous rates estimated at 0.042 substitutions per site for mitochondrial genes and at 0.082 substitutions per site for chloroplast genes were only 6 and 7 times lower than synonymous rates inferred for the same loci (Drouin et al. 2008).

### **2.2.2 Organelle genomes have limited coding potential**

According to this argument, because mitochondria and chloroplasts relinquished most of their genes to the nuclear genome during endosymbiotic gene transfer (Timmis et al. 2004), most organelle functions are under nuclear control. Therefore, even if local adaptation requires changes in organelle function, these are most likely to be encoded in nuclear DNA.

Of course, we already know that the apparent simplicity of organelle genomes cannot be taken as *prima facie* evidence for the adaptive neutrality of organelle DNA variation. A number

of adaptive responses have been traced back to the compact mitochondrial and chloroplast genomes of animals and plants. In animals, for which the coding capacity of the chondriome is even more reduced than it is in plants (Timmis et al. 2004), naturally-occurring mtDNA variants have been shown to differentially affect a host of traits including lifespan, fecundity, or starvation resistance (Toews et al. 2013; reviewed in Ballard & Melvin 2010). In plants, the case of weed resistance to triazine herbicides constitutes a textbook example. Over the past 40 years, persistent application of triazine herbicides has imposed a strong selective pressure for the evolution of resistance on weed populations globally (Powles & Yu 2010). Since it was initially reported in the 1970s, triazine resistance has been described in at least 68 weed species (Powles & Yu 2010). In the majority of these cases, the resistance trait has been mapped to a point mutation in the plastome *psbA* gene (Powles & Yu 2010). Another compelling example from plants is chilling tolerance in cucumber. Gordon & Staub (2011) used reciprocal backcrosses between chilling-sensitive and chilling-tolerant lines to show that tolerance to reduced temperature is inherited maternally, with the nuclear genome having a negligible contribution. The causative mutations for this trait are most likely located in the chloroplast genome, since only the plastome is inherited maternally in cucumber, while the chondriome is inherited paternally (Gordon & Staub 2011). This possibility is reinforced by the fact that strong associations between three single nucleotide polymorphisms (SNPs) in the cucumber plastome and chilling tolerance have been reported previously (Chung et al. 2007).

### **2.2.3 Mutation rates are reduced for plant organelle DNA**

Contrary to animal mtDNA, plant organelle DNA often shows markedly reduced mutation rates (see section 2.2.3.1 for a description of patterns reported as well as mechanistic

explanations). Because the rate of adaptation is limited by the supply of mutations, a third argument that can be made is that low mutational input limits adaptive evolution of plant organelle genomes.

There are three caveats to this argument. First, while mutation rates will determine the amount of standing or *de novo* variation available for adaptive evolution, we know that a third source of variation, introgression, is common for plant organelle genomes (Rieseberg & Soltis 1991). Indeed, evidence has been provided for trans-species selective sweeps at plant organelle DNA (see ‘Observational evidence’ below; Muir & Filatov 2007). Second, we do not know how much organelle DNA variation is needed for an adaptive response under changing environments. The examples of resistance to triazine herbicides in weeds or chilling tolerance in cucumber (Powles & Yu 2010; Gordon & Staub 2011) suggest the slightest alterations in organelle DNA can have important adaptive consequences. Third, comprehensive analyses of plant organelle DNA variation have been uncommon until recently and, with some exceptions, low mutation rates have been substantiated by surveys of few genes and/or a limited number of samples. With the accumulation of DNA sequence over the past decade, mutation rate speed-ups have been described at multiple levels of biological organization (Table 2.1), and the generality of this assumption has been shaken. Thus, it is not reasonable to attribute low adaptability of plant organelle DNA to its universally low rate of mutation.

### **2.2.3.1 Variation in organelle DNA mutation rates**

While in animals, mtDNA mutation rates are 5-50 times faster than for nuclear DNA (Brown et al. 1979), the situation is often inversed in plants. In a pioneering study 27 years ago,

Wolfe et al. (1987) showed that genes in the mitochondrial and chloroplast genomes of plants evolve at roughly 6-fold and 2-fold slower rates, respectively, than genes in the nuclear genome.

More recent studies have made use of an increasing amount of DNA sequence data to further revise these estimates. Drouin et al. (2008), for example, documented levels of polymorphism retained in three mitochondrial, five chloroplast, and four nuclear genes for 27 seed plant species. Results confirmed the patterns initially reported by Wolfe et al. (1987), although the magnitude of differences was shown to differ between plant groups. For instance, the ratios of mitochondrial to chloroplast to nuclear DNA synonymous substitutions were estimated to be 1:3:16 for angiosperms, compared to 1:2:4 for gymnosperms (Drouin et al. 2008). Consistent with the view that organelle DNA evolves mainly under purifying selection (section 2.1.1), the rates of nonsynonymous substitution for the same organelle genes and taxa were 6-7 times lower than rates of synonymous substitution (Drouin et al. 2008).

Mechanistic explanations for the discrepancies between animal and plant systems in organelle DNA mutation rates have generally revolved around differences that exist between animals and plants in the nuclear-encoded machinery of organelle DNA replication and repair. For instance, high mutation rates in animal mtDNA have been suggested to be at least partially caused by the absence, in animal nuclear genomes sequenced to date, of homologs for the mutS and recA genes (Lin et al. 2006; Sloan & Taylor 2012). Both of these are classic players in bacterial DNA recombination and mismatch repair that seem to have been lost in animals during endosymbiotic gene transfers (Lin et al. 2006; Sloan & Taylor 2012).

In plants on the other hand, homologs for mutS and recA are present in multiple active copies (e.g. Lin et al. 2006; Maréchal & Brisson 2010). Studies using mutants and RNA interference have illustrated that products of these genes limit the frequency of illegitimate

recombination and genome rearrangements in the plastome and chondriome (Maréchal & Brisson 2010; Sloan & Taylor 2012). Given that gene conversion, a process relying on recombination, has been shown experimentally to contribute to the elimination of *de novo* base pair substitutions in tobacco plastomes (Khakhlova & Bock 2006), it is likely that a similar mechanism also contributes to the maintenance of reduced point mutation rates at plant organelle genomes.

### **2.3 Evidence for an adaptive value of plant cytoplasm**

The arguments outlined above, while firmly grounded in organelle biology, then do not seem to exclude the possibility that positive selection may shape plant organelle DNA diversity. So is there evidence that plant organelle genetic variation is adaptive?

Characteristics of plant dispersal and gene flow suggest the maternal contribution of the genome should be a prime target for adaptive divergence. Plant dispersal is mediated by pollen and seed. Among these, seed is known to have a disproportionately lower contribution to dispersal (Petit et al. 2005). In agreement with this observation, maternally inherited markers show more subdivision among plant populations than paternally or biparentally inherited ones (Petit et al. 2005). Chloroplasts may be inherited paternally, as in conifers (Neale et al. 1986), or biparentally, as in *Passiflora* (Hansen et al. 2007), and patterns of mitochondrial variation in some cases suggest occasional leakage and recombination of paternal mitochondrial genomes (Jaramillo-Correa & Bousquet 2005; McCauley 2013). Nevertheless, inheritance of organelles is overwhelmingly uniparental and typically maternal in most groups. Given that the diversifying effects of positive selection are hindered by the homogenizing effects of gene flow, high genetic



subdivision of maternally inherited genomes may mean that even weak selection at plant organelle DNA can be sufficient to drive local adaptation.

Computer simulations support this argument. Irwin (2012) used individual-based modeling to track the genealogy of a uniparentally-inherited locus and the distribution of the phenotypic values linked to it, under spatially varying selection. Under conditions of high dispersal and extremely weak selection, the locus behaved neutrally. When dispersal was moderate, however, even fairly weak selection led to the formation of locally adapted clades (Irwin 2012). We next discuss empirical evidence that supports this theoretical prediction, pointing to adaptive plant organelle DNA variation.

### **2.3.1 Observational evidence: studies of organelle genome capture**

The replacement of one species' or population's organelle genomes with those of another has been observed in a range of taxa (Rieseberg & Soltis 1991). Commonly referred to as chloroplast or mitochondrial capture, this phenomenon is thought to originate by hybridization followed by repeated backcrossing to the pollen donor or via asexual transfer of organelles across natural grafts (Stegemann et al. 2012; Fuentes et al. 2014).

While the mechanisms by which capture takes place are clear, the evolutionary contexts it occurs in are less well understood. A common interpretation is that these events are selectively neutral, resulting from incomplete lineage sorting (Comes & Abbott 2001), from stochastic surfing of alien cytoplasm during range expansions (Neiva et al. 2010), from differential allocation to female reproductive functions (Tsitroni et al. 2003) or because reproductive barriers between species are asymmetric (McKinnon et al. 2004). Another possibility is that organelle capture events are adaptive, with plasmotypes being transferred because they confer a selective

advantage (Toews & Brelsford 2012; Greiner & Bock 2013). This scenario is used to interpret cases where captured haplotypes are associated with geography. Examples of such associations abound in animals (Toews & Brelsford 2012). In plants, however, they are less frequently reported, potentially because landscape-level surveys of organelle DNA variation have been relatively rare in plants (Schaal et al. 1998).

Some of the best known plant examples of chloroplast introgression come from studies of European white oaks. Petit et al. (2002a) performed what is to date one of the most ambitious landscape-level surveys of plant organelle DNA variation. The authors sampled over 2,600 European populations of eight white oak species and typed 12,214 individuals at chloroplast DNA. Chloroplast capture was inferred to be extensive, as haplotypes did not group by species. Instead, six chloroplast DNA clades were distributed along a longitudinal gradient across the continent (Petit et al. 2002a). This result built on previous findings in white oaks of such associations at the regional and local scales as well (Dumolin-Lapegue et al. 1997; Petit et al. 1997; Petit et al. 2002b). Other similar examples have since been provided by studies in European *Betula* (Palme et al. 2004), or South American *Nothofagus* (Acosta & Premoli 2010).

On a cursory examination, it seems reasonable to assume that positive selection was involved. Transferred haplotypes could be more fit if, for example, they are less mutationally loaded than haplotypes being replaced, or they could be better adapted to local environments. Although intuitively appealing, this adaptive designation is premature, as alternative neutral scenarios can generate similar patterns. For example, it may be that the front of organelle DNA introgression is moving, and overlaps with environment by chance. Indeed, the dominant interpretation of these data has been adaptively neutral, with introgression driven by invasion of the pollen parent (e.g. Petit et al. 2004). In light of the growing evidence for adaptive evolution

of organelle genomes, additional analyses are required to identify determinants of organelle capture.

Scenarios of neutral and adaptive organelle introgression can be tested by looking for the DNA footprints of positive selection. For example, neutral population growth and positive selection are both expected to lead to an excess of rare polymorphisms. One can differentiate between the two scenarios using coalescent simulations, by comparing levels of observed variation with those expected under neutrality (Llopart et al. 2014). Alternatively, tests of neutrality can be applied concomitantly to organelle and nuclear DNA data. Contrary to the signature for positive selection, which should be found only in organelle DNA if capture events are adaptive, neutral factors such as population expansion should leave a trace in both organelle and nuclear DNA.

This approach was used by Muir & Filatov (2007). The authors sampled populations of the hybridizing angiosperm species *Silene latifolia* and *S. dioica* across Eurasia, and typed specimens at organelle and nuclear DNA. Consistent with chloroplast capture, there was extensive haplotype sharing between species in regions of range overlap. Also, analyses rejected neutrality for chloroplast genes. By contrasting these results with those obtained for nuclear DNA, which behaved according to neutral expectations, the authors were able to exclude the possibility that organelle capture was neutral (Muir & Filatov 2007). Instead, a selective sweep was inferred to have occurred in the *Silene* plastome between 0.16 and 1.06 million years ago, which then crossed the *Silene* species boundaries (Muir & Filatov 2007).

### 2.3.2 Experimental evidence: studies of cytonuclear interactions

Strong evidence for local cytoplasmic adaptation has been provided by studies aiming to understand the determinants of cytonuclear interactions (CNI) (Burton et al. 2013). These studies use crosses or *in vitro* manipulation to make lines for which the native plasmotype has been replaced with the plasmotype of a different species or ecotype (Burton et al. 2013; Greiner & Bock 2013).

The fitness of such alloplasmic lines can be reduced irrespective of environment. In this case, intrinsic selection is thought to act against dissonant interactions between nuclear and organelle genomes that are not adapted to function in the same cell (Burton et al. 2013; Greiner & Bock 2013). Occasionally, genes involved in these interactions have been identified (Maheshwari & Barbash 2011). For example, the albino phenotype of hybrids carrying the chloroplast genome of tobacco and the nuclear genome of deadly nightshade was shown to result from defective RNA editing of the tobacco plastid *atpA* gene by nightshade nuclear-encoded enzymes (Schmitz-Linneweber et al. 2005). If, however, the fitness of alloplasmic lines is contingent on environment, extrinsic ecological selection is inferred to contribute to CNI (Burton et al. 2013; Greiner & Bock 2013). Under this scenario, CNI results because organelle genes are locally adapted, and potentially also involved in maladaptive cross-talk with nuclear genes, either directly or through linkage.

Sambatti et al. (2008) is a well-known example of studies in this category. The authors investigated the contribution of extrinsic ecological selection to CNI between *Helianthus petiolaris* and *H. annuus*, two hybridizing annual sunflowers that occupy contiguous and contrasting habitats in North America (Figure 2.1a). The authors carried out a reciprocal transplant experiment using 5,600 seedlings of the two species, their reciprocal F1s, and eight

backcross combinations of nuclear and cytoplasmic genomes (Figure 2.1b). Analysis of the survivorship of transplanted genotypes revealed a significant interaction between habitat and the fraction of *H. annuus* nuclear genome, as well as between habitat and the plasmotype of both species (Fig. 2.1c). These results are a strong indication that ecological differentiation (e.g. drought adaptation) in *H. petiolaris* and *H. annuus* is underpinned not only by nuclear genes, but also by organelle genes (Sambatti et al. 2008). Reciprocal transplant experiments have since shown that environment-dependent selection on the cytoplasm contributes to CNI between *Ipomopsis aggregata* and *I. tenuituba* (Campbell et al. 2008), and between *Penstemon newberryi* and *P. davidsonii* (Kimball et al. 2008), two pairs of species that hybridize along altitudinal clines.

Compelling examples exist at the infraspecific level as well. Leinonen et al. (2011), for instance, performed a reciprocal transplant of *Arabidopsis lyrata* subspecies that diverged in allopatry in Europe and North America, as well as their F1 and F2 reciprocal hybrids. As expected if cytoplasmic genomes - either alone or *via* their interaction with the nuclear genome - contribute to local adaptation, a strong positive effect on fitness was observed for the local cytoplasm (Leinonen et al. 2011). This pattern was pursued further in a follow-up study that used quantitative trait locus (QTL) mapping to understand the number and genomic location of nuclear genes that interact with cytoplasmic genomes during local adaptation (Leinonen et al. 2013). A fitness advantage of local nuclear alleles was associated with the local cytoplasm only at some QTLs, and only in European samples. These results showed that fitness advantages of local cytoplasm observed by Leinonen et al. (2011) are largely conferred by variation in organelle genomes, and not by CNIs (Leinonen et al. 2013).

Studies such as those outlined above, which assess the fitness of experimental crosses under natural settings, are a powerful way of determining whether plant cytoplasm contributes to local adaptation. One limitation of this approach is that, unless environmental variables are experimentally manipulated, it does not by itself provide any measure of the selection pressures that may be driving ecological divergence. Also, unless used in species for which the plastome and chondriome have opposite modes of inheritance, the experimental approach does not allow inferences to be made regarding which of the two organelle genomes is the target of selection.

### **2.3.3 Statistical evidence: studies of positive selection at the molecular level**

The adaptive contribution of plant organelle genetic variation has also been studied by looking for footprints left by positive selection in patterns of DNA variation. One of the first and most taxonomically diverse studies in this category is that by Kapralov & Filatov (2007). The authors leveraged the wealth of sequence data generated for phylogenetic purposes for the plastome *rbcL* gene, which encodes the large subunit of the photosynthetic enzyme Rubisco. Their dataset included 3,228 sequences obtained from all lineages of green plants, and some lineages of brown and red algae, diatoms, euglenids and cyanobacteria. Contrary to the traditional view that plant organelle DNA variation is neutral, the  $dN/dS$  ratio test provided evidence for positive selection at *rbcL* in as many as 75-88% of land plants (Kapralov & Filatov 2007). This result was followed by a number of other similar studies. While some provided results consistent with neutral organelle polymorphism (e.g. Wright et al. 2008), others reported patterns indicative of non-neutrality (Table 2.2). Similarly to the Kapralov & Filatov (2007) example, many of these used the  $dN/dS$  ratio test to look for sites under positive selection along a gene of interest and across a phylogeny (Table 2.2).

Loci involved in local adaptation can also be identified by searching for correlations between allele frequencies and environmental variables (Coop et al. 2010). This is a powerful way to study adaptive organelle DNA evolution, as it allows inferences to be made not only on the genetic basis of local adaptation, but also on likely agents of selection. Ideally, to understand the full genetic architecture of adaptive responses, genome-wide surveys of polymorphism should be performed. A less comprehensive but still valuable approach is to rely on knowledge of gene function, and to select a subset of loci suspected to be involved in the adaptations of interest. One example is the study of Bashalkhanov et al. (2013). The authors performed an environmental correlation analysis in red spruce, using SNPs from 36 nuclear and plastome candidate genes, chosen for their likely involvement in adaptation to climate and human-induced air pollution. Polymorphism at six nuclear genes, as well as the plastome *chlB* gene, which encodes for the light-independent protochlorophyllide reductase, was strongly associated with 19 climatic variables, suggesting these loci have been targets of spatially variable selection (Bashalkhanov et al. 2013).

Ideally, statistical inferences of positive selection should be interpreted in conjunction with experimental evidence. This is because false positive rates of neutrality tests can be high if the underlying demographic assumptions are unrealistic (Nielsen 2001; Beaumont & Balding 2004; Beaumont 2005). Moreover, interdisciplinary approaches are more likely to paint a complete picture of the genetic and ecological contexts of adaptive evolution. Galmes et al. (2014) used this strategy to investigate whether positive selection at the plastome *rbcL* gene contributed to adaptation to drought conditions during the recent diversification of the perennial angiosperm genus *Limonium* in the Balearic Islands. Two derived substitutions at functionally important Rubisco residues, I309M and S328A, were inferred to have been the result of positive

selection according to the  $dN/dS$  ratio test. In vitro enzymatic assays confirmed that these substitutions are associated with increased CO<sub>2</sub> affinity and reduced carboxylase efficiency of Rubisco. By rearing plants with both derived and ancestral *rbcL* haplotypes under irrigated and water-limited conditions, the authors were able to identify that the levels of CO<sub>2</sub> available in the chloroplast stroma during periods of drought was the likely selective agent driving these substitutions (Galmes et al. 2014).

Neutrality tests can be used to dissect episodes of adaptive evolution at the molecular level. Apart from the scarcity of experimental confirmation of putative examples of molecular adaptation, the greatest limitation of studies in this category performed so far is that they have relied on a limited number of genes. Such approaches may allow, at best, only incomplete glimpses of non-neutrality of organelle DNA variation. Analyses of complete or nearly-complete organelle genomes, which have now become increasingly accessible, should provide a more unbiased look at positive selection at the level of the plasmotype.

## **2.4 Conclusions**

In this Review, we presented experimental evidence for why it should no longer be assumed that plant organelle DNA variation is selectively neutral. In doing so, our aim was not to dampen recent excitement about the use of complete sequences of plant organelle genomes in studies of plant phylogenetics and phylogeography. Rather, we hope to caution against the use of these data without testing beforehand that neutrality assumptions are met. This is particularly relevant for studies using population samples, for which neutrality violations are expected to have a disproportionately larger effect. More generally, we aimed to highlight the neglected



possibility that local adaptation of plant populations is underpinned by both nuclear and cytoplasmic genes.

Given that sequence data are being collected at an unprecedented pace, we predict that in the near future, evidence for non-neutrality of plant organelle DNA variation, and in particular that obtained from patterns of organelle capture and tests of neutrality, will continue to accumulate. We can also expect that concomitant improvements in analytical approaches will increase the reliability of inferences of positive selection drawn from sequence data alone. Even so, by relying on isolated examples obtained from distantly connected branches in the tree of life, we are unlikely to obtain a complete picture of positive selection at plant organelle DNA. This is because answers to many of the currently outstanding questions on this topic, a few of which are highlighted in Chapter 5, are likely to depend on the species under consideration.

Future approaches should therefore aim to integrate observational, experimental, and statistical evidence in multiple systems. Moreover, physiological experiments and functional studies should be implemented to connect molecular and experimental evidence of adaptive evolution with differences in fitness. It is only by using an interdisciplinary approach that we can hope to move from documenting isolated examples of adaptive organelle DNA evolution, to understanding its frequency, drivers, and evolutionary significance.

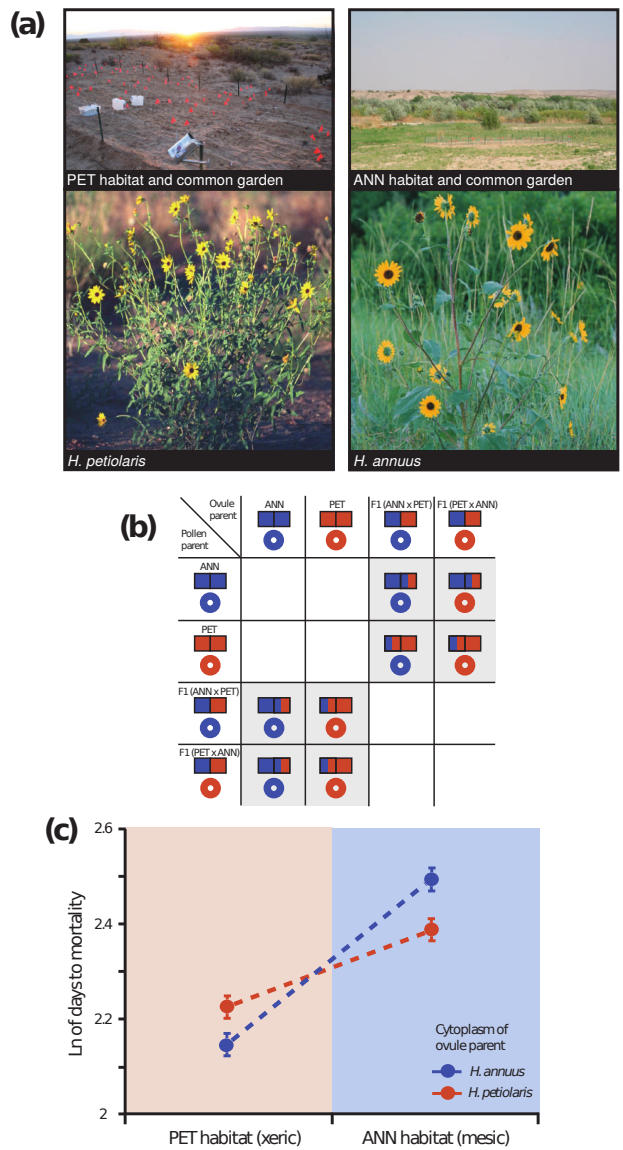
**Table 2.1:** Examples of studies reporting accelerated rates of nucleotide substitution in plant organelle genomes. “P” is used to indicate loci in the plastome, and “C” is used to indicate loci in the chondriome.

Taxa investigated	Regions with accelerated rates in one or more taxa	<i>dN</i> elevated	<i>dS</i> elevated	Reference
Flowering plants (50 taxa)	C: <i>atp1, cob, cox1, cox2, LSU rDNA, SSU rDNA</i>	•	•	Cho <i>et al.</i> (2004)
Flowering plants (58 taxa)	C: <i>atp1, cob, cox1, cox2, cox3, nad1, SSU rDNA, LSU rDNA</i>	•	•	Parkinson <i>et al.</i> (2005)
Flowering plants (127 taxa)	C: <i>nad1</i>		•	Bakker <i>et al.</i> (2006)
<i>Silene vulgaris</i> (25 samples)	C: <i>atp1</i>		•	Barr <i>et al.</i> (2007)
Land plants (306 – 578 taxa, depending on gene used)	C: <i>atp1, cob, cox1, cox2, cox3, matR, LSU rDNA, SSU rDNA</i>	•	•	Mower <i>et al.</i> (2007)
<i>Sileneae</i> (21 taxa) and <i>Oenothera</i> (4 taxa)	P: <i>clpP1</i>	•	•	Erixon & Oxelman (2008)
Flowering plants (47 taxa)	P: <i>rpl-, rps-, rpo-, psb-genes</i>	•	•	Guisinger <i>et al.</i> (2008)
<i>Silene</i> (4 species)	C: complete genomes	•	•	Sloan <i>et al.</i> (2012)
<i>Pelargonium</i> (58 species)	P: <i>rpoC1</i> C: <i>nad5</i>	•	•	Weng <i>et al.</i> (2012)
<i>Sileneae</i> (7 species)	P: <i>clpP, ycf1, ycf2</i>	•	•	Sloan <i>et al.</i> (2014b)
Geraniales (11 species)	P: complete genomes	•		Weng <i>et al.</i> (2014)
<i>Ajuga reptans</i>	C: <i>atp9, rps3, rps12</i> P: <i>atpH</i>	•	•	Zhu <i>et al.</i> (2014)

**Table 2.2:** Examples of studies using neutrality tests of DNA polymorphism to infer positive selection at organelle genes. All loci investigated were located in the plastome.

Taxa included in analysis	Genes investigated	Neutrality test used	Selected genes	Putative agent of selection	Reference
<i>Schiedea</i> (27 taxa)	<i>matK, psbA, rbcL</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	Photosynthetic performance under dry sunny conditions	Kapralov & Filatov (2006)
Green plants, brown and red algae, diatoms, euglenids and cyanobacteria (3228 taxa)	<i>rbcL</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	Photosynthetic performance under fluctuating thermal and gaseous conditions in terrestrial environments	Kapralov & Filatov (2007)
<i>Silene latifolia</i> (75 samples) and <i>S. dioica</i> (29 samples)	<i>trnL, matK, rbcL</i>	HKA and Tajima's <i>D</i>	<i>matK, rbcL, trnL + matK</i>	Not discussed	Muir & Filatov (2007)
Commelinoid monocots (338 taxa)	<i>rbcL, ndhF</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	Photosynthetic performance in CO <sub>2</sub> -rich bundle sheath cells of C <sub>4</sub> plants.	Christin <i>et al.</i> (2008)
<i>Sileneae</i> (21 taxa) and <i>Oenothera</i> (4 taxa)	<i>clpP1</i>	<i>dN/dS</i> ratio test	<i>clpP1</i>	Not discussed	Erixon & Oxelman (2008)
Flowering plants (47 taxa)	72 plastid genes	<i>dN/dS</i> ratio test	<i>rpoB, rpoC1, rpoC2</i>	Not discussed	Guisinger <i>et al.</i> (2008)
<i>Potamogeton</i> (18 taxa)	<i>rbcL, atpB, petA</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	Photosynthetic performance under environmental variation in temperature and dryness	Iida <i>et al.</i> (2009)
Green plants (31 taxa)	75 plastid genes	<i>dN/dS</i> ratio test	<i>atpE, cemA, clpP, rpoB, rps11</i>	Not discussed	Zhong <i>et al.</i> (2009)
<i>Pinus</i> (37 taxa)	nearly-complete plastomes	<i>dN/dS</i> ratio test	<i>ycf1, ycf2</i>	Not discussed	Parks <i>et al.</i> (2009)
Green plants (2279 taxa)	<i>matK</i>	<i>dN/dS</i> ratio test	<i>matK</i>	Not discussed	Hao <i>et al.</i> (2010)
<i>Flaveria</i> (15 taxa)	<i>ndhF, psbA, rbcL</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	Not discussed	Kapralov <i>et al.</i> (2011)
Ferns (27 taxa)	<i>psbA</i>	<i>dN/dS</i> ratio test	<i>psbA</i>	Photosynthetic performance under modified light conditions caused by angiosperm diversification	Sen <i>et al.</i> (2011)
<i>Amaranthaceae</i> sensu lato (179 taxa)	<i>rbcL</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	Photosynthetic performance in warm climates	Kapralov <i>et al.</i> (2012)
<i>Pelargonium</i> (58 species)	<i>rbcL, matK, ndhF, rpoC1, trnL-F</i>	<i>dN/dS</i> ratio test	<i>rpoC1</i>	Not discussed	Weng <i>et al.</i> (2012)
<i>Picea rubens</i>	<i>chlB</i>	environmental correlation	<i>chlB</i>	Photosynthetic performance under changing climatic conditions	Bashalkhanov <i>et al.</i> (2013)
<i>Sileneae</i> (7 species)	complete plastomes	<i>dN/dS</i> ratio test	<i>clpP, ycf1, ycf2</i>	Not discussed	Sloan <i>et al.</i> (2014b)
<i>Limonium</i> (42 species)	<i>rbcL</i>	<i>dN/dS</i> ratio test	<i>rbcL</i>	CO <sub>2</sub> availability under drought conditions	Galmes <i>et al.</i> (2014)

**Figure 2.1:** Example of experimental confirmation of the adaptive contribution of the plant plasmotype. (a) *Helianthus petiolaris* (PET), *H. annuus* (ANN), and common garden locations used in Sambatti et al. (2008). (b) Crosses used in Sambatti et al. (2008) to obtain different nuclear genome – organelle genome combinations. For the F1s, the maternal parent is listed first. The eight possible backcross combinations are indicated with grey shading. Squares represent the nuclear genomes, while open circles represent the plasmotype for PET (red) and ANN (blue). (c) Cytoplasm by habitat interaction for survivorship expressed as the mean  $\pm$  SE of the natural log of days to mortality for individuals sharing the same cytoplasm (reproduced with permission from Sambatti et al. (2008); photo credits: JBM Sambatti, GJ Seiler, J Rick).



## **Chapter 3: Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an Artichoke**

### **3.1 Introduction**

The perennial sunflower *Helianthus tuberosus* is a taxon with a rich human-connected history. The Cree and Huron Indians of Eastern North America, who referred to this plant, respectively, as ‘askipaw’ and ‘skibwan’ (“raw thing”), grew it for its large tubers before the first European contact (Heiser, 1976; Kosaric et al. 1984; Kays & Nottingham, 2008). As such, although tuber archaeological remains are yet to be recovered for this species, *H. tuberosus* represents one of the few domesticates that can support Eastern North America as one of the world’s cradles of domestication. After being transferred to the Old World in the early 1600s, it was readily adopted as a food plant (Heiser, 1976; Kosaric et al. 1984; Kays & Nottingham, 2008). In the process, it acquired an impressive assortment of common names that vary in botanical accuracy (Heiser, 1976; Kosaric et al. 1984; Kays & Nottingham, 2008), such as “Jerusalem Artichoke” or “Sunchoke”. Among these, “Jerusalem Artichoke”, thought to be a corruption of the Italian ‘girasole articiocco’ (“sunflower artichoke”; Smith 1807), is its most widely used appellation. By the mid 18th century, as farming of potato became widespread, the relative importance of Jerusalem Artichoke as a food plant decreased (Kays & Nottingham, 2008). Even so, it remains a globally-cultivated multifunctional crop, well adapted to diverse geoclimatic regions (Kosaric et al. 1984) including dry climates with nutrient-poor soils (Kays & Nottingham, 2008). Recent surges in its production have been prompted by the health benefits associated with the consumption of inulin (Kleessen et al. 2007; Roberfroid, 2007), the reserve carbohydrate stored in Jerusalem Artichoke tubers, and the utility of its below-ground and above-

ground parts for biofuel production and livestock feed (Bajpai & Bajpai, 1991; Cheng et al. 2009).

Despite its cultural and economic significance, important aspects of the origin of the Jerusalem Artichoke, with implications for germplasm preservation and cultivar improvement, remain unanswered. Specifically, although it is currently agreed that the Jerusalem Artichoke species originated in Central-Eastern North America, where its wild populations abound (Rogers et al. 1982; Kays & Nottingham, 2008), other details of its evolution remain a mystery. For instance, it is uncertain whether this hexaploid species ( $2n = 6x = 102$ ) is monophyletic (i.e. autopolyploid) or polyphyletic (i.e. allopolyploid or auto-allopolyploid; Kostoff, 1934; Kostoff, 1939; Darlington, 1956; Heiser & Smith, 1964). Among these, the polyphyletic auto-allopolyploid scenario appears to be the most likely, as it is supported by the cytogenetic observation that two of the three chromosome sets of Jerusalem Artichoke are homologous (Kostoff, 1939). Aside from the mechanism of formation, also unknown is the identity of the progenitor species. Two competing hypotheses have been proposed, each based on different lines of evidence. The first hypothesis, drawing on the fact that the Jerusalem Artichoke can be crossed readily with the annual sunflower *H. annuus* (Kostoff, 1939) and shows similarity to this species based on immunochemistry data (Anisimova, 1982), posits that one parent of Jerusalem Artichoke is the annual Common Sunflower *H. annuus*. The alternative hypothesis is that the Jerusalem Artichoke originated strictly from perennial sunflowers, most likely *via* hybridization between tetraploid ( $2n = 4x = 68$ ) and diploid species ( $2n = 2x = 34$ ; Heiser & Smith, 1964; Heiser, 1976). This hypothesis implicates as potential progenitors a group of five perennial sunflower taxa whose morphology and North American ranges overlap with that of Jerusalem

Artichoke (Heiser et al. 1969; Heiser, 1976; Kays & Nottingham, 2008). Of these, the Hairy Sunflower (*H. hirsutus*), a species whose rhizomes are often thickened terminally (Heiser et al. 1969) which has been proposed as an autopolyploid of *H. divaricatus* (Heiser et al. 1969), is seen as the most likely tetraploid progenitor (Heiser, 1976). The Sawtooth Sunflower (*H. grosseserratus*) and the Giant Sunflower (*H. giganteus*) are similarly considered the most likely diploid progenitors (Heiser, 1976).

Molecular phylogenetics has so far remained inconclusive in establishing the origin of the Jerusalem Artichoke (Gentzbittel et al. 1992; Schilling, 1997; Schilling et al. 1998; Timme et al. 2007). This is because the diversification of perennial *Helianthus* species is characterized by several processes known to confound phylogenetic inference. These include their recent, rapid radiation (Schilling, 1997; Timme et al. 2007), the formation of diploid hybrids (Long, 1955; Timme et al. 2007) and of polyploids *via* whole-genome duplication with or without hybridization (Timme et al. 2007), and the prevalence of post-speciation gene flow facilitated by high levels of interspecies fertility (Heiser & Smith, 1964). In addition, taxonomic ambiguity is common among perennial sunflowers, given their frequently overlapping morphologies (Heiser et al. 1969).

Phylogenomics is an effective means of addressing complex phylogenetic questions. Although traditionally used to resolve deep splits in the tree of life, this approach is now being applied to shallow phylogenetic divisions (Emerson et al. 2010; Wagner et al. 2013). For recently diverged plant species in particular, and for those with large genomes, a genome-skimming approach has been advocated (Straub et al. 2012). Also known as ultra-barcoding, or UBC, (Kane & Cronk, 2008; Kane et al. 2012), genome skimming consists of the assembly and analysis of the high-copy genomic fraction, consisting of plastid and mitochondrial genomes as

well as nuclear ribosomal DNA (rDNA). Aside from the large amount of data generated, the value of this approach stems from the complementary utility of the two marker categories. The non-recombining and uniparentally inherited organellar genomes allow the matrilineal genealogy to be recovered. In cases of reticulate speciation, organellar DNA can be used to discern between single versus multiple origin scenarios (Soltis & Soltis, 1989; Schwarzbach & Rieseberg, 2002; Guggisberg et al. 2006; Slotte et al. 2006), and to clarify whether maternal parentage was reciprocal or unidirectional (Soltis & Soltis, 1989). The biparentally inherited rDNA is ideally suited for inferring species-level phylogenies. Provided that concerted evolution has not homogenized divergent parental genotypes, rDNA can readily reveal evidence of hybridization (Malinska et al. 2010; Malinska et al. 2011). In perennial sunflowers in particular, rDNA has proven to be the most phylogenetically informative region studied so far (Timme et al. 2007).

Here, we use a genome-skimming approach to investigate the origin of the Jerusalem Artichoke. We collected the largest dataset used to date in *Helianthus* phylogeny, consisting of complete plastid genomes as well as partial sequences for the mitochondrial genome and nuclear-encoded 35S and 5S rDNA. We screened 38 accessions, representing geographically diverse populations of eight species (Fig. 3.1; Table A.1), including the Jerusalem Artichoke and all diploid and tetraploid perennial sunflowers that have been proposed as its progenitors. We supplement these data with corresponding sequences from *H. annuus* such that all proposed parents of the Jerusalem Artichoke are represented in our dataset.



## 3.2 Materials and Methods

### 3.2.1 Molecular techniques

The accessions used in this study were obtained from the US Department of Agriculture (USDA) collections held at Ames, Iowa, and were chosen to maximize geographical representation for each species within Central-Eastern North America (Fig. 3.1; Table A.1). The ploidy of each accession (Fig. 3.1; Table A.1) was determined using flow cytometry, with the internal standards *Zea mays* ( $2C = 5.43$  pg), *Secale cereale* ( $2C = 16.19$  pg) and *Vicia faba* ( $2C = 26.90$  pg; Dolezel et al. 2007). DNA was extracted from leaf tissue of single individuals using established procedures (Doyle & Doyle, 1987). Illumina paired-end (PE) libraries (100 bp read length) were prepared from fragmented genomic DNA (fragment size  $\sim 400$  bp) following standard protocols. With the exception of the four *H. maximiliani* accessions which were sequenced with samples from a related project, all libraries were run on one lane on an Illumina HiSeq 2000 machine, with pooling designed to achieve comparable total coverage for each species and ploidy level (Table A.2).

### 3.2.2 Assembly of plastid and mitochondrial genomes

Prior to *de novo* assembly, we reduced the complexity of each library by aligning quality-filtered reads to the *H. annuus* plastid (GenBank accession NC007977) and mitochondrial (GenBank accession KF815390) genomes using Bowtie2 (Langmead & Salzberg, 2012). Apart from simplifying the assembly task, this step was used to gauge the average coverage across each genome (Table A.3), and to calibrate the fragment length of each library for *de novo* assembly. Reads corresponding to organellar genomes were assembled using the *de novo* de Bruijn graph-based tool VELVET (version 1.2.06; Zerbino & Birney, 2008). We used a hash length of 21, and

a minimum contig length of 100 bp. For the plastid assembly, for which average coverage depth was 95x (Table A.3), we set the coverage cutoff to 15. For the mitochondrial genome, for which average coverage depth was 9x (Table A.3), we allowed VELVET to automate the coverage cutoff. Resulting contigs were aligned to the corresponding organellar genome of *H. annuus*, ordered, and merged (when overlapped) using CodonCode Aligner (version 2.0.4; CodonCode Corporation, Dedham, MA, USA). For the plastid genome, small gaps were filled using trimmed Illumina reads. Mononucleotide repeats that could not be bridged in all samples were collapsed, for all samples, to the smallest repeat size present in the dataset. For the mitochondrial genome, gaps that could not be bridged by Illumina reads were coded as missing data. Draft assemblies for the plastid and mitochondrial genomes of each accession were validated by mapping quality-filtered Illumina reads and visually inspecting the coverage distribution using Tablet (version 1.12.03.26; Milne et al. 2010). The full-length plastid genome of each accession was annotated using DOGMA (Wyman et al. 2004).

### **3.2.3 Assembly of 35S and 5S rDNA regions**

Quality-filtered reads for each accession were assembled using the de novo de Bruijn graph-based tool Trinity (version R2012-06-08; Grabherr et al. 2011) at default parameters. Contigs for 35S and 5S rDNA were identified based on alignments to the corresponding *H. annuus* references for 35S (GenBank accession KF767534) and 5S (GenBank accession HM638217). Preliminary inspection of rDNA contigs revealed three regions that could be aligned unambiguously across all samples: a 7,457 bp stretch of 35S rDNA (consisting of partial ETS, 18S, ITS1, 5.8S, ITS2, 26S, and partial NTS), an additional 739 bp stretch of the NTS associated with 35S, and a 514 bp stretch of 5S rDNA (consisting of 5S and its corresponding

NTS region). To incorporate intra-individual polymorphism between rDNA repeats, we aligned quality-filtered reads to each of the three regions using Bowtie2, and called SNPs using Unified Genotyper from the Genome Analysis Toolkit (GATK; version 2.1-13; DePristo et al. 2011). Because of the repetitive nature of rDNA, we treated all samples as polyploids at the SNP-calling step. To determine the ploidy setting, we surveyed 23 distinct values in Unified Genotyper (range 2x – 200x; Figs. A.1 and A.2) for each accession and rDNA region and recorded the number of SNPs called given the filtering criteria (i.e. GATK confidence score > 10; mapping quality > 15). For the final analysis, we used 100x, the ploidy setting for which the maximum number of SNPs was called, and beyond which the number of SNPs remained relatively constant (Figs. A.1 and A.2). Polymorphisms scored under these conditions and filtering criteria were incorporated in the de novo assemblies using IUPAC ambiguity codes.

### **3.2.4 Alignment and phylogenetic analyses**

For each region, we retained full sequence data, consisting of both variable and invariable sites. Alignments performed in MAFFT (version 6.814b; Katoh & Toh, 2008) with default settings were inspected and edited in CodonCode Aligner. For the alignment of draft mitochondrial genomes, we removed sites with missing data in more than five samples, and excluded singleton SNPs, due to the low coverage obtained for this region (Table A.3). We also excluded 15 segments that were classified, according to BLAST searches against the *H. annuus* plastid genome, as likely integrants of plastid DNA in the mitochondrial genome. For the 35S and 5S rDNA alignments, we excluded singleton SNPs identified (Fig. A.3), to address the possibility that false positive calls may have been incorporated in the assemblies at the SNP-calling step.

Maximum likelihood (ML) phylogenies were inferred using PhyML Best AIC Tree (version 1.02b), implemented in Phylemon (version 2.0; Sánchez et al. 2011). PhyML Best AIC Tree uses PhyML (version 3.0; Guindon & Gascuel, 2003) to select the best model of sequence evolution under AIC and build ML phylogenies. ML branch support was estimated using the Shimodaira–Hasegawa-like (SH-like) procedure implemented in PhyML (Guindon & Gascuel, 2003). The SH-like procedure assesses whether the branch being studied provides a significant likelihood gain compared to the null hypothesis that involves collapsing that branch (Guindon & Gascuel, 2003). It is a fast method for branch support estimation suitable for large datasets, which provides similar results to bootstrap (Anisimova & Gascuel, 2006). Bayesian inference analyses were conducted with MrBayes (version 3.2.1; Ronquist & Huelsenbeck, 2003), with parameters of sequence substitution set to follow as closely as possible the model inferred by PhyML. We used four runs, each with four Markov chains initiated from a random tree and run until the average standard deviation of split frequencies remained below 0.01 (range 1,000,000 to 4,000,000 generations). Trees were sampled every 500 generations. The first 25% of all trees sampled before convergence were discarded as burn-in. Mean level of sequence divergence between organellar haplotypes within each species was calculated in MEGA 5 (Tamura et al. 2011), using the Tamura-Nei model (Tamura & Nei, 1993).

### **3.2.5 Survey of diagnostic polymorphism in rDNA data**

Diagnostic sites, defined here as sites that are fixed in a given species at its lowest ploidy level, were identified by scanning the rDNA alignments in CodonCode Aligner. When such diagnostic sites showed intra-individual polymorphism (i.e. were coded in IUPAC degenerate

bases), we obtained the frequency of each underlying allele from individual quality-filtered VCF files.

### 3.3 Results and Discussion

Unrooted phylogenies of organellar genomes and rDNA revealed extensive sequence divergence between *H. annuus* and perennial sunflowers, including the Jerusalem Artichoke (Figs. A.4-A.6). Two scenarios are compatible with this observation. The first is that *H. annuus* was involved in the parentage of Jerusalem Artichoke as the pollen donor, but concerted evolutionary forces acting since the polyploidization event have overwritten the *H. annuus* - derived rDNA to the maternal type. Homogenization of rDNA arrays has been documented in other polyploids (Wendel et al. 1995), and can occur over the course of only a few generations (Malinska et al. 2010; Malinska et al. 2011). Nevertheless, in the case of Jerusalem Artichoke frequent vegetative reproduction should have resulted in the retention of both parental sequences for prolonged periods of time. The alternative scenario is that *H. annuus* did not contribute any of the three genomes in Jerusalem Artichoke. The two species are cross-fertile, and this possibility has been exploited in the past to transfer resistance to pathogens from Jerusalem Artichoke into cultivated sunflower (Atlagić et al. 1993; Atlagić & Škorić, 2006). However, the resulting hybrids often show greatly reduced fertility (Heiser & Smith, 1964; Atlagić et al. 1993; Atlagić & Škorić, 2006). Cytogenetic observations of Jerusalem Artichoke x *H. annuus* progeny have also documented a high frequency of meiotic abnormalities linked to faulty homolog recognition, including univalent and multivalent formation (Atlagić et al. 1993; Atlagić & Škorić, 2006). By contrast, hybrids between Jerusalem Artichoke and diploid perennial sunflowers such as *H. divaricatus* show more regular meiosis, with reduced univalent formation

(Chandler, 1991). These studies suggest that differences in chromosomal structure between Jerusalem Artichoke and *H. annuus* are more pronounced than those between Jerusalem Artichoke and perennial sunflowers, and as such lend further weight to the view that the formation of Jerusalem Artichoke entailed the exclusive contribution of perennial sunflowers, and not *H. annuus*.

The organellar phylogenies rooted with *H. annuus* did not recover any perennial sunflower species as reciprocally monophyletic (Fig. 3.2). Incomplete lineage sorting (ILS) and reticulation, two alternative but not mutually exclusive processes, can be invoked to explain this pattern. Caused by the retention of ancestral polymorphism, ILS should be common in perennial sunflowers, given their recent, rapid radiation (Schilling, 1997; Timme et al. 2007). Because ILS is stochastic in nature, it should result in discordant associations between accessions of different species, within and between the two organellar phylogenies. As expected, such discordant associations are a pervasive occurrence across both plastid and mitochondria phylogenies (Fig. 3.2). A similar trend was found for the diploid-only subset (Fig. A.7). Given that diploid-only phylogenies exclude the contribution of polyploid taxa of possible reticulate ancestry, this finding lends further support to the view that ILS is a major contributor to the discordances we observe.

In contrast to ILS, reticulation should result in systematic associations between species, reflecting the prevalence of post-speciation organelle capture among pairs of taxa that are inter-fertile and/or the maternal ancestry of hybrid species with extant progenitors. Such consistent associations include those between *H. giganteus* and *H. decapetalus* accessions (Fig. 3.2). Two of these groupings were corroborated by geography (Fig. 3.2), indicating they are likely instances of recent organelle capture, a phenomenon that is widespread in the genus (Rieseberg

& Soltis, 1991). The only groupings of Jerusalem Artichoke accessions recovered repeatedly across analytical methods and organellar phylogenies are those with *H. hirsutus* and *H. divaricatus* (Fig. 3.2).

The survey of cytoplasmic genomes further revealed high levels of organellar genetic variation in Jerusalem Artichoke. Each accession had a unique plastid and mitochondrial haplotype (Fig. 3.2). The mean level of sequence divergence between Jerusalem Artichoke organellar haplotypes was also within the range of those recovered between the geographically diverse accessions of other perennial sunflowers (Fig. 3.3). This is in stark contrast with the nearly complete lack of plastid variation reported in polyploids thought to have single origins (Guggisberg et al. 2006; Slotte et al. 2006). Under the assumption of Jerusalem Artichoke formation through a single genetic event, post-speciation organelle capture from other perennial sunflowers could be invoked as the source of this variation. However, given that most perennial sunflower species are diploid or tetraploid (Heiser et al. 1969) and considering that strong pre- and post-zygotic barriers are typically associated with inter-cytype gene exchange (Husband & Sabara, 2003), organelle capture from other species is expected to be limited in Jerusalem Artichoke. An alternative, more plausible explanation is that, similar to many other polyploid taxa (reviewed in Soltis & Soltis, 1999), the Jerusalem Artichoke experienced multiple independent origins, each time sequestering different organellar haplotype combinations from its maternal parent.

The rDNA phylogenies showed, in agreement with previous studies (Schilling et al. 1998; Timme et al. 2007), that no single rDNA region can resolve relationships among all perennial sunflowers (Fig. A.8). The concatenated rDNA phylogeny was nevertheless highly informative, providing unprecedented resolution for this group. Notably, most taxa that have long been

recognized as distinct based on morphology formed monophyletic groups, some with high support (Fig. 3.4). Two major clades, A and B, were recovered (Fig. 3.4). Clade A comprised all *H. giganteus* and *H. decapetalus* accessions, in line with organellar phylogenies which repeatedly group these species. Within clade B, the morphologically distinct *H. maximiliani* was recovered as highly divergent and monophyletic. All *H. divaricatus* and *H. hirsutus* accessions, along with the tetraploid accession of *H. strumosus* formed another group within clade B (Fig. 3.4). The grouping of *H. hirsutus* with *H. divaricatus* represents the first molecular phylogenetic support of the morphology-based assumption that *H. hirsutus* is an autotetraploid of *H. divaricatus* (Heiser et al. 1969). However, in consideration of the fact that rDNA may underestimate the frequency of allopolyploid speciation events (Kim et al. 2008), this possibility should be investigated further, to exclude the possibility that divergent rDNA arrays in *H. hirsutus* were homogenized to the *H. divaricatus* type. The placement of the *H. strumosus* accession with *H. divaricatus* and *H. hirsutus* is also in agreement with previous taxonomical work. Notably, tetraploid *H. strumosus* was proposed to be included with *H. hirsutus*, based on its high morphological resemblance to *H. hirsutus*, and the fact that its cross with *H. hirsutus* results in highly fertile progeny (Rogers et al. 1982; Heiser et al. 1969). In line with this observation, for the remaining analyses, we treated the *H. strumosus* accession as *H. hirsutus*.

The Jerusalem Artichoke accessions were part of a polytomy within clade B, and were closely related to the monophyletic *H. divaricatus*/*H. hirsutus* and *H. grosseserratus* clades (Fig. 3.4). This indicates that few autapomorphies separate the Jerusalem Artichoke from *H. hirsutus* and *H. grosseserratus*, the two species considered its most likely progenitors based on morphology and overlapping geographical ranges (Heiser, 1976). The unresolved phylogenetic placement of Jerusalem Artichoke accessions relative to *H. hirsutus* and *H. grosseserratus*



further indicates the possibility that Jerusalem Artichoke rDNA contains alleles that are diagnostic for each of these putative progenitors, which have not been homogenized by concerted evolutionary forces.

To test this hypothesis we defined diagnostic alleles as those that are present in all accessions of a species at the lowest ploidy level. Because these alleles must co-occur in populations sampled from disparate geographical regions (Fig. 3.1) they likely originated early in the formation of each species and would represent a small fraction of the phylogenetic variation analyzed here. Nonetheless, they should be highly informative, particularly with regards to the ancestry of taxa that arose *via* hybridization. In all, we identified 30 diagnostic sites across the rDNA regions (Fig. 3.5). In agreement with our expectation formulated on the basis of the phylogenetic reconstruction, the Jerusalem Artichoke was revealed as containing diagnostic sites from both *H. hirsutus* and *H. grosseserratus*. All alleles diagnostic of *H. hirsutus* and of *H. grosseserratus* are present in Jerusalem Artichoke (Fig. 3.5b). By contrast, with the exception of two alleles diagnostic of *H. decapetalus* which were present at low frequency in one Jerusalem Artichoke accession, no alleles diagnostic of other putative progenitors segregated in Jerusalem artichoke. Lastly, only two alleles present in all Jerusalem Artichoke accessions were not observed in any putative parental species.

The phylogenetic placement of Jerusalem Artichoke (Fig. 3.4), as well as the pattern of hybridity revealed for geographically diverse accessions analyzed here (Fig. 3.5b), indicate a monophyletic, autopolyploid origin of this species is highly unlikely. Instead, our results provide strong support that the origin of Jerusalem Artichoke was polyphyletic, involving hybridization between *H. hirsutus* and *H. grosseserratus*. Among two polyploidization scenarios that can be characterized as polyphyletic, allopolyploidization and auto-allopolyploidization, the latter is the

most plausible according to results presented here, which indicate the speciation event involved the merger of two duplicate genomes contributed by the *H. hirsutus* parent, and a third differentiated genome contributed by the *H. grosseserratus* parent. This auto-allopolyploidization scenario is also in agreement with previous cytogenetic observations pointing to a high degree of homology between two of the three chromosome complements of Jerusalem Artichoke (Kostoff, 1939).

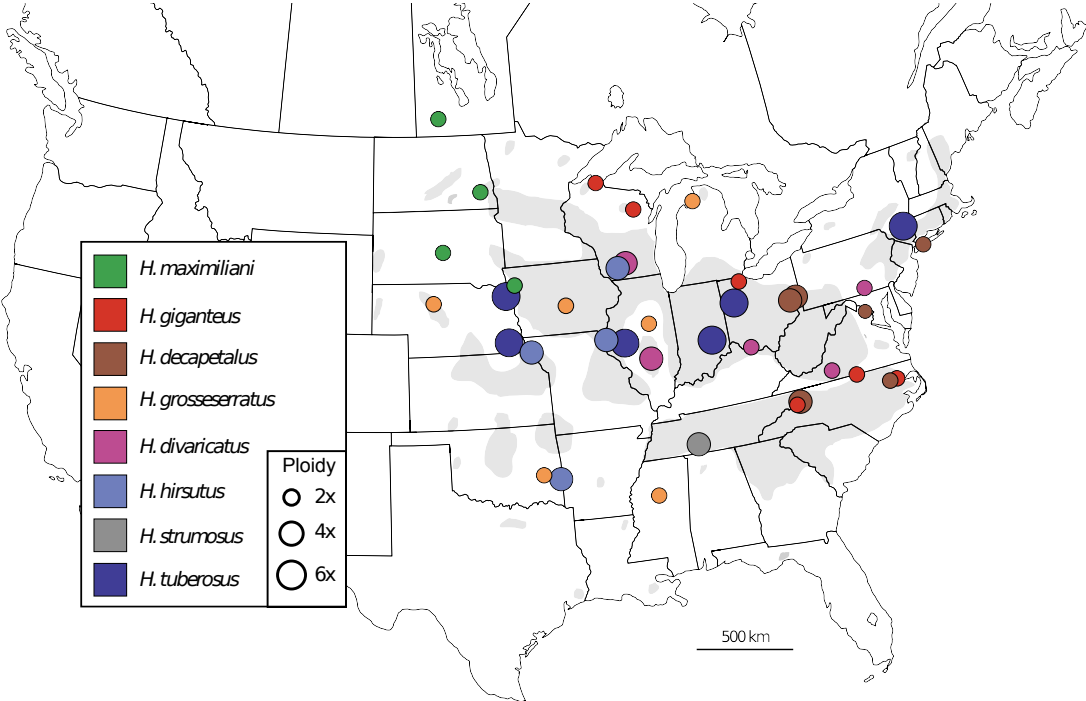
### 3.4 Conclusions

The origin of Jerusalem Artichoke, a tuber-producing species that is widely grown as a cultivated plant, has long fascinated botanists. The dataset and analyses presented here provide strong genetic evidence that the origin of Jerusalem Artichoke is polyphyletic, from perennial sunflowers in Central-Eastern North America. The likely progenitors of this species, as indicated by additive patterns of rDNA variation, were the Hairy Sunflower (*H. hirsutus*) - which was supported as a likely autotetraploid of *H. divaricatus* - and the diploid Sawtooth Sunflower (*H. grosseserratus*). Additional information was provided by organellar phylogenies. Notably, high levels of organellar genome variation indicate that Jerusalem Artichoke likely experienced recurrent formation. Furthermore, maternal origins of Jerusalem Artichoke appear to have been unidirectional, from *H. hirsutus*. This conclusion is supported by the fact that while Jerusalem Artichoke – *H. hirsutus* groupings were recovered repeatedly across analytical methods and organellar phylogenies, there was no case where organellar genomes of Jerusalem Artichoke were grouped with those of *H. grosseserratus*.

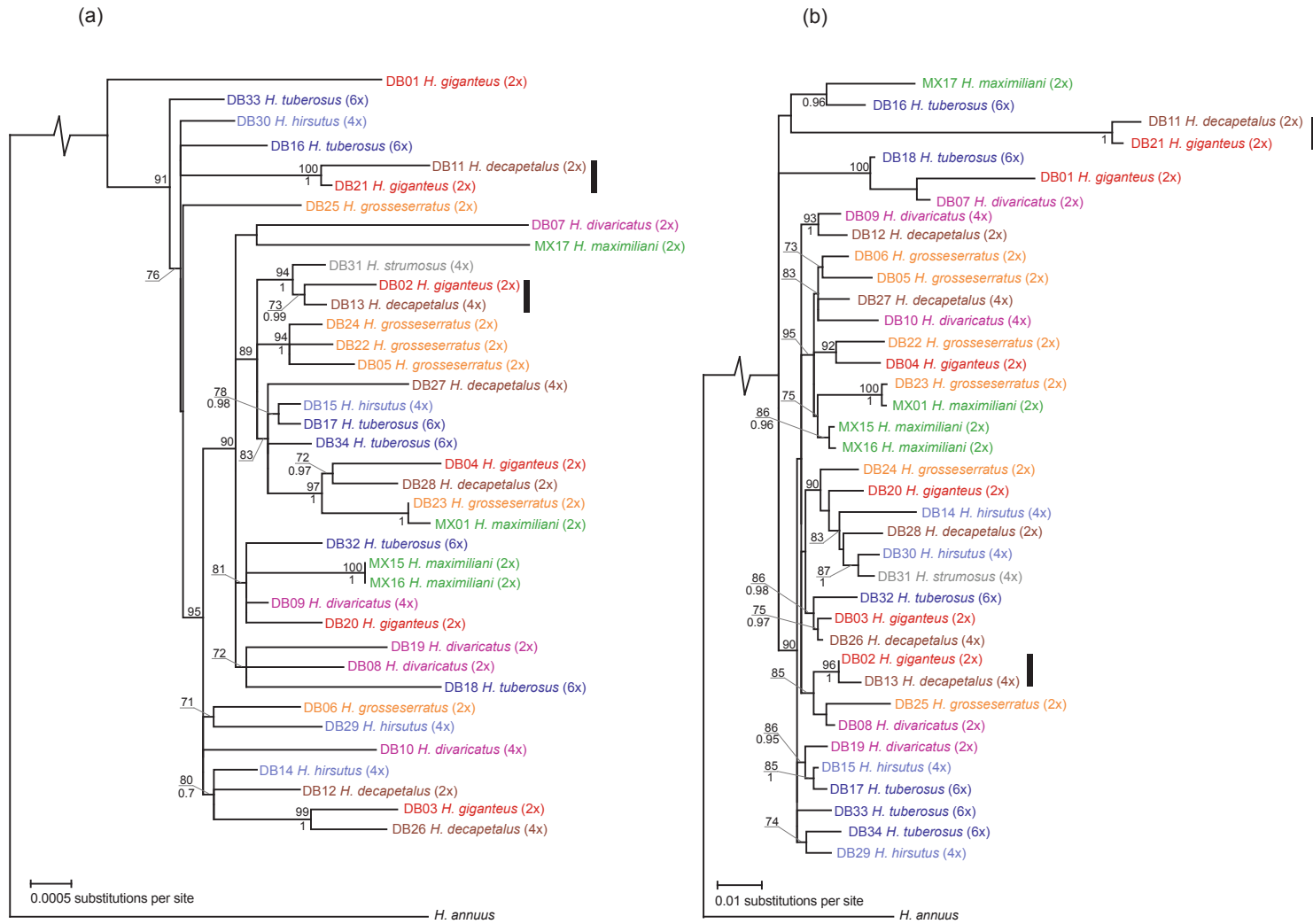
This information can be used to direct efforts of germplasm preservation for Jerusalem Artichoke and its wild species progenitors, and should form the foundation of future

improvement programs aiming to add novel valuable diversity in Jerusalem Artichoke cultivars from closely related congeners. Our findings also provide a previously lacking evolutionary framework that allows us to investigate the evolution and genetic architecture of perennial life habit and tuber production in sunflowers. Beyond these considerations, results presented here highlight the promise and applicability of next-generation sequencing technologies in general, and the genome skimming approach in particular, for resolving species boundaries, origins and relationships in previously intractable polyploid complexes.

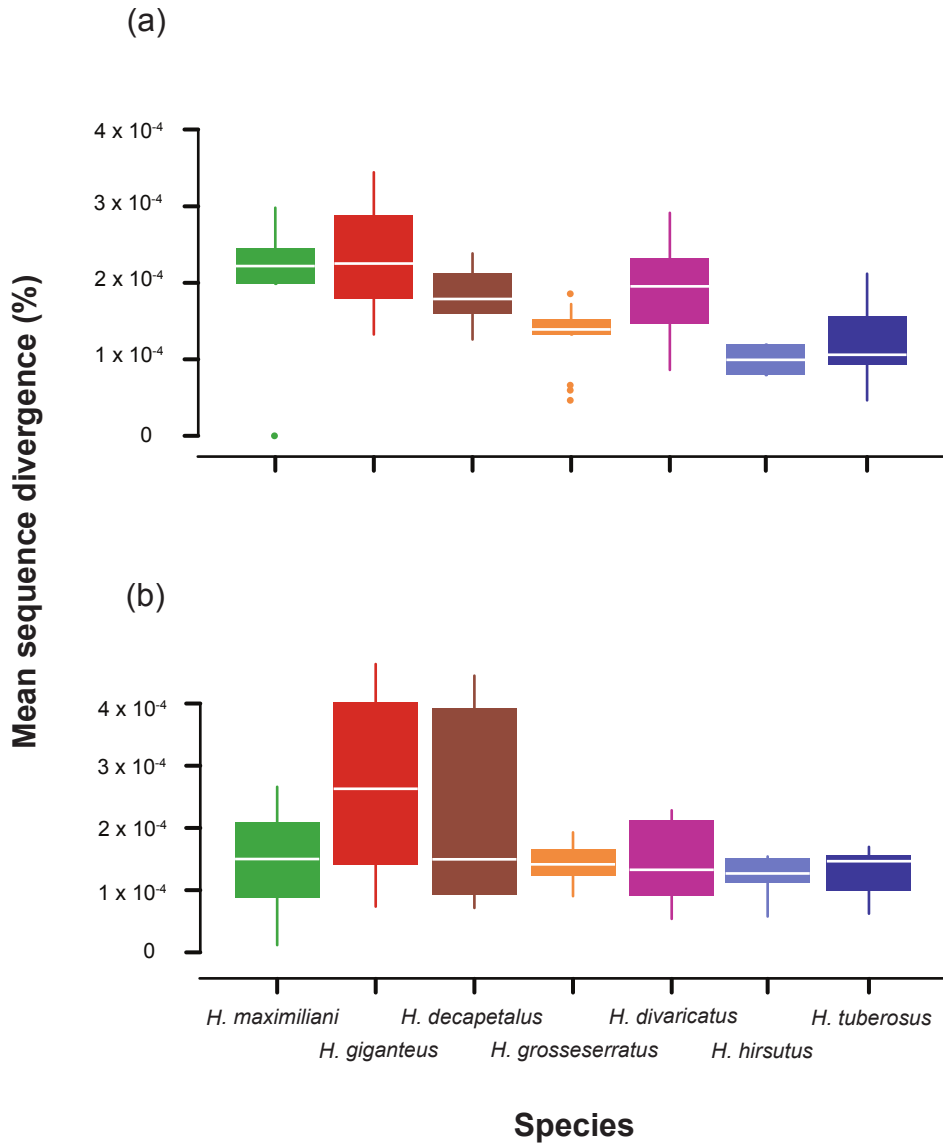
**Figure 3.1:** Geographic distribution of the perennial *Helianthus* accessions sequenced. Grey shading is used to illustrate the range of Jerusalem Artichoke in the United States [redrawn from Rogers et al. (1982)].



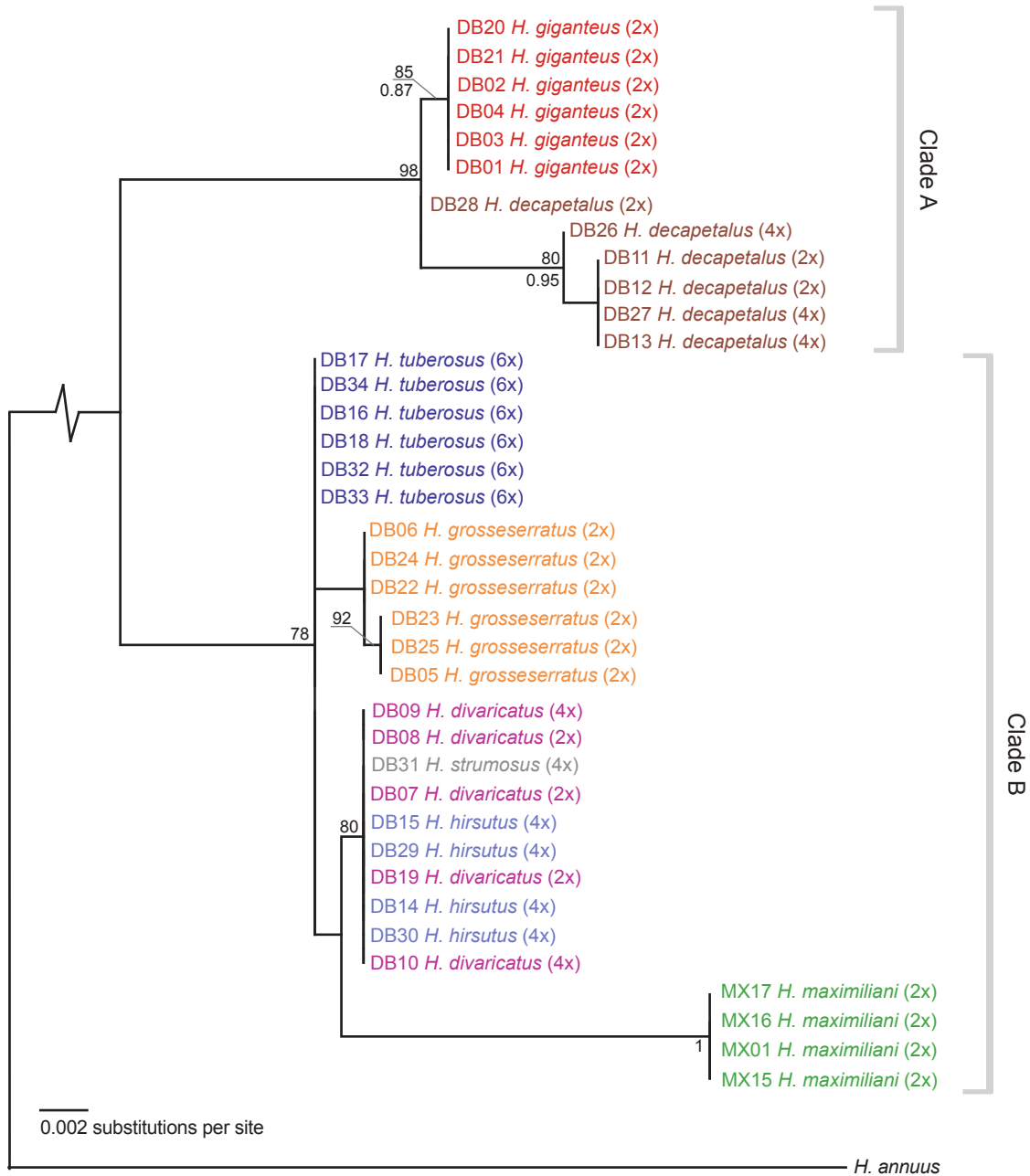
**Figure 3.2:** Maximum likelihood phylogenetic reconstruction of (a) complete plastid genomes (151,552 bp) and (b) partial mitochondrial genomes (196,853 bp) for perennial *Helianthus* accessions sequenced. Groupings supported by geography are indicated by black vertical bars.



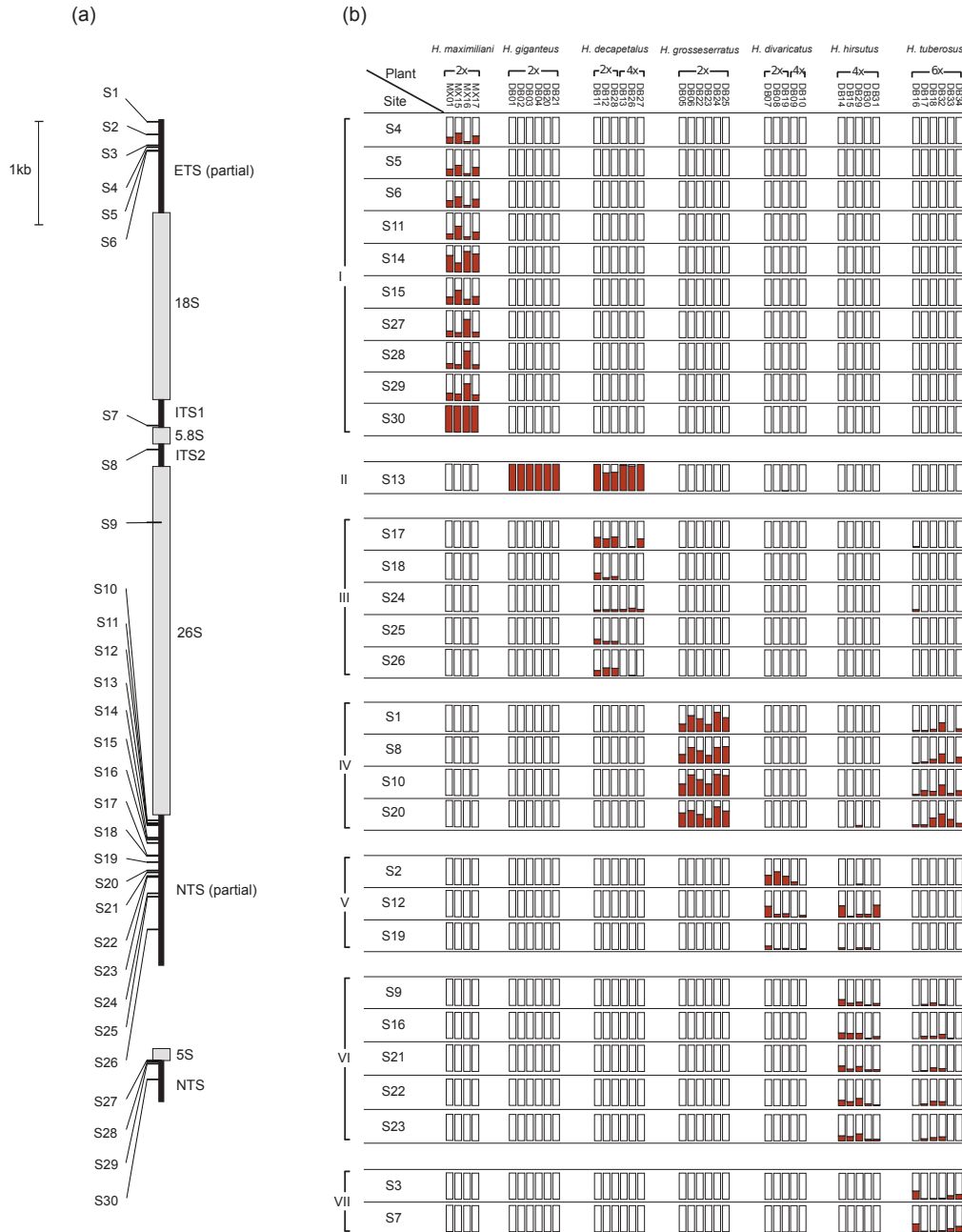
**Figure 3.3:** Mean of pairwise sequence divergences calculated between all pairs of accessions within each species for (a) whole-plastome haplotypes (151,552 bp) and (b) partial mitochondrial haplotypes (196,853 bp). *Helianthus strumosus* was excluded from this analysis since only one accession was available for this species.



**Figure 3.4:** Maximum likelihood phylogenetic reconstruction of concatenated rDNA sequences (8,710 bp) for perennial *Helianthus* accessions sequenced. Support is shown for nodes with SH-like values > 70% (above) and Bayesian posterior probabilities > 0.7 (below).



**Figure 3.5:** Survey of diagnostic rDNA polymorphism with position of diagnostic sites along the 35S and 5S rDNA regions (a) and allelic profiles for diagnostic sites (S1-S30) identified (b). All sites were bi-allelic. Bar plots show the relative proportion of the common allele (white segments) and the species-diagnostic allele (red segments) in each accession and site. Diagnostic sites are grouped by species: *H. maximiliani* (I), *H. giganteus* (II), *H. decapetalus* (III), *H. grosseserratus* (IV), *H. divaricatus* (V), *H. hirsutus* (VI), and *H. tuberosus* (VII).





## Chapter 4: Multiple genetic routes to the evolution of invasiveness in a perennial sunflower

### 4.1 Introduction, Results and Discussion

The long-held view that evolutionary change is a slow process and therefore unlikely to contribute to the success of biological invasions has now repeatedly been challenged (Prentis *et al.* 2008; Bock *et al.* 2015; Colautti & Lau 2015). Increasingly often, studies are demonstrating that contemporary evolution of invasives can have a large impact on their performance, either independent of the local environment (Krieger & Ross 2002; Perkins *et al.* 2013), or through adaptation across climatic gradients (Dlugosch & Parker 2008; Colautti & Barrett 2013). However, with some notable exceptions (Krieger & Ross 2002), few studies have succeeded in identifying the genetic mechanisms involved. As such, we do not know if evolution of invasiveness is genetically constrained, if standing genetic variation or *de novo* mutations are typically at play, or to what extent invasiveness is conferred by a few large-effect genes, or, alternatively, by many small-effect genes scattered throughout the genome (Bock *et al.* 2015). Interest in the genetic mechanisms of invasion success is strong because this information could eventually be used to curtail further spread of harmful non-indigenous species (Champer *et al.* 2016), or, alternatively, to maximize evolutionary potential in endangered or economically-relevant ones.

Here, we investigated the genetic architecture of invasion success in the perennial sunflower *Helianthus tuberosus* (Jerusalem artichoke). This species was first introduced to Europe as a tuber crop in 1607, and remained a minor cultigen until the 1900s (Kays & Nottingham 2008). At this time, a number of breeding programs and field trials were established

across the continent, with the aim of developing *H. tuberosus* as a bioenergy crop (Konvalinková 2003). Towards the end of the century, as breeding efforts were abandoned, the spread of the species in natural habitats started to be reported (Konvalinková 2003). These expansions became more aggressive during the past three decades, such that *H. tuberosus* is now considered one of the most invasive plants in Europe (Anastasiu & Negrean 2009; Filep *et al.* 2010; Fehér 2007, Konvalinková 2003).

To clarify the origin of invasive *H. tuberosus*, we used genotyping-by-sequencing (see section 4.2, Detailed Materials and Methods). We sampled the genomes of 691 individuals obtained from 49 native sites, from 21 invasive sites, and from three major international repositories of cultivated *H. tuberosus* (Appendix Table B.1; Appendix Figures B.1 & B.2). We additionally included 175 samples of the proposed progenitors of *H. tuberosus*, represented by the diploids *H. grosseserratus* and *H. divaricatus*, and by *H. hirsutus*, an autotetraploid of *H. divaricatus* (see section 4.2, Detailed Materials and Methods; Appendix Table B.2). Lastly, we incorporated SNP data from 150 samples of the annual congener *H. annuus*. Recent phylogenetic evidence indicates *H. tuberosus* is an auto-allopolyploid (Bock *et al.* 2014; Baute *et al.* 2016). We confirmed these findings using a principal component analysis (PCA) in which *H. tuberosus* clustered between its proposed progenitors (Figure 4.1A). While interspecific *H. annuus* x *H. tuberosus* hybrids have reportedly been used in *H. tuberosus* breeding (Kays & Nottingham 2008), we did not recover evidence for a *H. annuus* contribution to the ancestry of samples included in our dataset (see Appendix Figure B.4).

Within *H. tuberosus*, the PCA shows that the main axis of divergence is between wild samples obtained from North America (hereafter 'native samples') and cultivated samples. While invasive *H. tuberosus* spanned a large fraction of the PC space occupied by native and cultivated

samples, most grouped as intermediate (Figure 4.1B; Appendix Figures B.4 & B.6). This is consistent with a diverse and predominantly admixed origin for invasive genotypes, a result substantiated by multiple additional lines of evidence. A maximum-likelihood phylogeny based on diploid subgenome markers (see section 4.2, Detailed Materials and Methods) grouped invasive genotypes in two native and in two cultivated clades (Figure 4.1C; Appendix Figure B.7), indicating that invasive *H. tuberosus* originated on at least four different occasions. One of the inferred origins (clade 3, Figure 4.1C) comprises over 85% of the invasive genotypes. These samples group as basal relative to cultivated *H. tuberosus*, and are part of a larger clade containing breeding lines that were obtained from crosses (Appendix Figure B.8). Bayesian clustering also supported both the multiple origins of invasive genotypes and their admixed ancestry. While native and cultivated individuals were mainly assigned to one of two genetic clusters, invasive individuals had significant ancestry from both groups (Figure 4.1D; Appendix Figures B.9-B.11).

The levels of genetic diversity that we observe are consistent with a mainly admixed origin of invasive *H. tuberosus*. Specifically, heterozygosity was significantly elevated in invasive samples relative to that of native or cultivated samples (Figure 4.1E). This result is consistent throughout much of the genome (Appendix Figure B.12), and is robust to calculations of heterozygosity based on the complete marker set, or the subsets of SNPs assigned to the two subgenomes (Appendix Figure B.13). Also, with the exception of diploid subgenome markers, heterozygosity of cultivated samples was not significantly reduced compared to that of native *H. tuberosus*. Thus, while we did not explicitly set out to investigate the genomic consequences of artificial selection in this system, our results provide an early indication that the domestication

bottleneck in cultivated *H. tuberosus* was limited, a result that is consistent with observations in other clonal crops (e.g. Cornille et al. 2012).

Previous studies have proposed that invasion success in *H. tuberosus* could be determined by increased allelopathic potential or clonal growth (Tesio *et al.* 2010; Filep *et al.* 2016; Konvalinková 2003). To evaluate these possibilities, we compared allelopathy, clonality (estimated here as the number of tubers), and 18 other traits among 225 native, cultivated and invasive samples (section 4.2, Detailed Materials and Methods; Appendix Table B.3). For allelopathy, we additionally included 135 samples of the progenitor species of *H. tuberosus*, which are not reported as invasive. Overall, the first two PC axes of morphological variation within *H. tuberosus* recapitulated the intermediacy of invasive genotypes (Figure 4.2A; see also Appendix Figure B.14). While allelopathic potential was differentiated between species (Appendix Table B.4), we did not recover evidence for a contribution of allelopathy to invasiveness. This is because samples of *H. tuberosus* did not exceed values observed in non-invasive parental species (Figure 4.2B; see also Appendix Figure B.15), and invasive *H. tuberosus* genotypes were not extreme relative to native or cultivated conspecifics (Figure 4.2B; Appendix Table B.5). The same trend was observed for most of the other traits, with invasive genotypes displaying largely intermediate values (Appendix Figure B.16). Invasive *H. tuberosus* did show extreme phenotypes for three of the measured traits, including the number of branches, the number of inflorescences, and the number of tubers (Appendix Figure B.16). Among these, tuber number was the only one consistently identified as a significant outlier in invasive samples (Appendix Tables B.6 & B.7). The extent of divergence in tuber number between invasive and non-invasive *H. tuberosus* is substantial, and comparable with domestication-related divergence in this species for the total weight of tubers, as well as divergence at other traits associated with

invasiveness in a variety of taxa (Figure 4.2C; Appendix Table B.8; section 4.2, Detailed Materials and Methods).

Clonal relationships inferred from marker data further corroborated the important role of vegetative propagation in the invasive spread of *H. tuberosus*. Clones were common in invasive populations, with 84% (112 of 133) of samples obtained from tubers sharing a clonal connection (Appendix Table B.10). Also, clones were not restricted to single sampling locales, and in some cases were shared between sites separated by over 500 km (e.g. clonal series 3, Appendix Table B.10). Moreover, two of the clonal series contain invasive samples obtained from 2013, as well as multiple accessions that were collected in Europe since at least 1975, and archived in public repositories of *H. tuberosus* (clonal series 3 and 4; Appendix Table B.10, Appendix Figure B.3). This indicates invasive *H. tuberosus* clones can be both widespread and long-lived, a finding that mirrors results from other important weeds that rely predominantly on vegetative propagation (e.g. Hollingsworth & Bailey 2000; Kliber & Eckert 2005). Hereafter, we considered tuber number a major invasiveness trait in this system.

We next investigated the genetic architecture of invasiveness in *H. tuberosus*. One means by which we may expect invasive samples to benefit from elevated heterozygosity is through heterosis, or hybrid vigor (Lippman & Zamir, 2007). Experimental evolution evidence suggests that this mechanism can have an important contribution to the success of biological invasions (e.g. Turgeon et al. 2011; Hahn & Rieseberg 2016). When determined by dominance or epistasis, heterosis can be fixed by natural selection. The general view, however, is that it erodes quickly with sexual reproduction, and as such contributes only to the short-term success of invasives (Rius & Darling 2014). Important exceptions to this view apply to the *H. tuberosus* system. For instance, even with frequent sexual reproduction, polyploidy will delay the loss of genome-wide

heterozygosity, therefore prolonging any heterotic effects. Also, in clonally propagated species such as *H. tuberosus*, heterosis can be stabilized indefinitely, irrespective of its genetic basis.

As expected if overall heterozygosity has an important fitness contribution in this system, significant positive correlations between heterozygosity and trait values were recovered for 10 of 19 (52.6%) measured phenotypes (Appendix Figure B.17). Among these, the number of branches, the number of inflorescences, and the number of tubers consistently showed the strongest correlations across marker sets (Appendix Figure B.17). For tuber number, correlations were significant for all samples combined ( $r = 0.41$ ;  $P < 0.001$ ), as well as for the native ( $r = 0.37$ ;  $P < 0.001$ ) and the cultivated subsets ( $r = 0.31$ ;  $P < 0.05$ ; Figure 4.3A; see also Appendix Figure B.18). Invasive samples, by comparison, did not show a significant correlation, with most individuals maintaining high tuber number production irrespective of heterozygosity (Figure 4.3A; see also Appendix Figure B.18). The lack of a significant correlation for invasive *H. tuberosus* could have resulted if there is no added benefit for tuber number production beyond a given heterozygosity level, or because of the reduced samples size available for this group. Indeed, the estimated power to detect an effect similar to the one we observed in non-invasive samples was low for invasive *H. tuberosus* (0.3; section 4.2, Detailed Materials and Methods). A third possibility is that additional genetic factors have an important contribution to tuber number production in invasive samples, therefore weakening the heterosis signal.

To investigate this possibility, we performed genome-wide association (GWA) mapping. We used 43,276 biallelic SNPs, with genotypes in hexaploid format (section 4.2, Detailed Materials and Methods). After Bonferroni correction, we detected 30 associations across all traits (Appendix Table B.9). For tuber number, two highly significant QTLs were identified on two linkage groups (LGs) of the *H. annuus* HA412 genome (Badouin et al. 2017): one on LG9 ( $P =$

1.2 x 10<sup>-6</sup>), and one on LG14 ( $P = 7.94 \times 10^{-7}$ ; Figure 4.3B). We note that these marker-trait associations cannot be attributed to unaccounted population stratification, as indicated by the inflation factor ( $\lambda = 0.99$ ). The two QTLs show additive mode of gene action and have large effect sizes, jointly explaining an estimated 31% of variation in tuber number (13% for L9 QTL; 18% for LG14 QTL). Furthermore, we found clear differences in the frequency of superior alleles at these QTLs between invasive and non-invasive *H. tuberosus*. While 78.6% of invasive samples contained at least one high tuber number allele, this frequency dropped to 29.6% and 25.6% for native and cultivated samples, respectively. We also performed a set of association analyses for tuber number that incorporate genome-wide heterozygosity as an additional covariate. As expected if the signal identified at the two additive QTLs is independent of hybrid vigor, this resulted in improved association signals for both loci (Appendix Table B.9).

We further evaluated the relative contribution of the two genetic mechanisms identified here, additive QTLs and hybrid vigor, as well as their interaction, to tuber number production in *H. tuberosus*. To do this, we fit three nested linear models that explain tuber number as follows. The 'qtl-only' model included one term for genotypes at each of the LG9 and LG14 QTLs (adjusted  $R^2 = 26.1\%$ ;  $F_{7,268} = 14.89$ ;  $P < 2.2 \times 10^{-16}$ ; AIC = 655.05). The 'qtl and hybrid vigor' model additionally included a term for genome-wide heterozygosity (adjusted  $R^2 = 34.56\%$ ;  $F_{8,267} = 19.16$ ;  $P < 2.2 \times 10^{-16}$ ; AIC = 622.54). Finally, the 'interaction' model included all terms from the preceding models, as well as interaction terms between genotypes at each of the two QTLs and heterozygosity (adjusted  $R^2 = 35.59\%$ ;  $F_{13,262} = 12.69$ ;  $P < 2.2 \times 10^{-16}$ ; AIC = 622.96). AIC scores and likelihood ratio tests confirmed that the 'qtl and hybrid vigor' model outperforms the 'qtl-only' model ( $P = 2.47 \times 10^{-9}$ ), and that the 'interaction' model does not provide a significantly better fit ( $P = 0.09$ ). These results therefore support the conclusion that, for tuber number

production in *H. tuberosus*, additive QTLs and hybrid vigor act synergistically, with maximal tuber number production being achieved through the combined action of both mechanisms (Figure 4.3C). This result is corroborated by the relative abundance and range of invasive *H. tuberosus* clones. The most abundant genotypes, both in terms of overall number of clones (not shown) and number of populations covered (Figure 4.3D), were those characterized by increased heterozygosity as well as superior tuber number alleles.

While the role of evolutionary change in the success of biological invasions is well established, the genetic factors involved are often unknown. Using a top-down approach, we identified two genetic mechanisms that have contributed to the evolution of a major invasiveness trait in *H. tuberosus*, a rapidly spreading perennial plant. This finding parallels results from recent studies, which highlight that independent evolution of invasiveness between species and between populations might have a diverse genetic basis (Hodgins *et al.* 2015; Qi *et al.* 2015). Our results further show how different genetic determinants of invasiveness can interact during the same biological invasion event. Collectively, these results confirm worries formulated during early surveys of molecular variation, which indicated invasives are unlikely to be genetically constrained (e.g. Kolbe *et al.* 2004). The success of genetic control strategies for invasive species will therefore depend on a thorough understanding of the range of genetic solutions that can be exploited during the evolution of invasiveness.



## 4.2 Detailed materials and methods

### 4.2.1 Sampling

#### 4.2.1.1 Public collections – *H. tuberosus*

We obtained *H. tuberosus* samples from four major public collections: the US Department of Agriculture (USDA) in Ames, USA, the Plant Gene Resources of Canada (PGRC) in Saskatoon, Canada, the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben, Germany, and the French National Institute for Agricultural Research (INRA) in Montpellier, France (Appendix Table B.1). The USDA samples ( $N = 241$ ) were collected from 49 sites in Central-Eastern US (see Appendix Figure B.1 for map of sampling locations). The PGRC ( $N = 138$ ), IPK ( $N = 70$ ), and INRA ( $N = 130$ ) collections comprise mainly breeding lines and cultivars, as well as samples of unknown origin ( Kays & Nottingham, 2008).

#### 4.2.1.2 Public collections – progenitor species

We additionally obtained samples of the progenitor species of *H. tuberosus*, collected by USDA, also from the Central-Eastern US (Appendix Table B.2). These included samples of the diploid subgenome donor, *H. grosseserratus* ( $2n = 2x = 34$ ), and samples representative of the tetraploid subgenome donor, which include both *H. hirsutus* ( $2n = 4x = 68$ ) and *H. divaricatus* ( $2n = 2x = 34$ ; Bock *et al.* 2014). We used these samples to verify the ancestry of *H. tuberosus*, to identify markers likely to segregate within the diploid subgenome of *H. tuberosus*, and to compare allelopathy of *H. tuberosus* to that of its non-invasive progenitor species (detailed below). In all, we obtained samples from 19 populations of *H. grosseserratus* ( $N = 95$ ), eight populations of *H. hirsutus* ( $N = 40$ ), and eight populations of *H. divaricatus* ( $N = 40$ ).

#### 4.2.1.3 Sampling of invasive populations

In August-September 2013, we obtained tubers from plants growing in typical invasive *H. tuberosus* habitat at 19 sites from Central-Eastern Europe ( $N = 105$ ; Appendix Table B.1; see Appendix Figure B.2 for map of sampling locations). At each location, we used a linear transect of 60 meters and collected randomly selected plants along the transect. Collected plants at each site were separated by an average distance of 11 meters (range 7.5 – 14.5). Samples from an additional two populations (one in Romania and one in Hungary;  $N = 7$ ) were provided by Rita Filep (University of Pécs, Hungary).

#### 4.2.2 Ploidy inference

We confirmed expected ploidy for seven accessions of *H. tuberosus*, 10 accessions of *H. grosseserratus*, and all accessions of *H. divaricatus* and *H. hirsutus* (Appendix Table B.1) using flow cytometry and the internal genome size standards *Zea mays* ( $2C = 5.43$  pg), *Secale cereale* ( $2C = 16.19$  pg), and *Vicia faba* ( $2C = 26.90$  pg; Dolezel *et al.* 2007). Following these analyses, in agreement with previous phylogenetic results that indicate *H. hirsutus* is an autotetraploid of the diploid *H. divaricatus* (Bock *et al.* 2014), we reclassified four tetraploid USDA accessions initially labeled as *H. divaricatus* to *H. hirsutus* (Appendix Table B.2).

#### 4.2.3 Genotyping-by-sequencing

Genomic DNA was isolated from fresh leaf tissue for all samples of all four species ( $N = 866$ ) using a modified CTAB protocol (Doyle & Doyle, 1987). For *H. tuberosus*, of the 691 samples, we selected four genotypes for which library preparation and sequencing was

performed in duplicate (Appendix Table B.1). These were used to assess genotyping error rates and to facilitate identification of clonal genotypes (see below).

We constructed genotyping-by-sequencing (GBS) libraries using the *Pst*I and *Msp*I restriction enzymes, following the protocol described in Poland *et al.* (2012), with modifications. Firstly, we multiplexed libraries after the PCR step. Prior to pooling, the amplification products were cleaned and quantified individually. Concentration estimates were accounted for at the pooling step, to minimize between-sample variation in sequencing output. Secondly, to maximize between-library representation of GBS tags, we selected multiplexed libraries for fragments in the range of 400-700 bp. Thirdly, to reduce high-abundance (e.g. organelle or transposable element DNA) sequence representation, we used a duplex-specific nuclease step, following a modified version of the protocol in Matvienko *et al.* (2013). For all four species, we sequenced 96 samples per lane on an Illumina HiSeq 2000. Sequence reads were sorted according to adapter barcodes using pyRAD v2.17 (Eaton 2014). Adapter contamination in demultiplexed reads was removed with cutadapt v1.8 (Martin 2012). We allowed a maximum error rate of 0.15 and retained reads of a minimum length of 30 nucleotides.

#### **4.2.4 Alignment of sequence data and variant calling**

For all taxa, we aligned GBS reads to the *H. annuus* line HA412 reference genome (v1.1.bronze; <http://www.sunflowergenome.org>) using BWA-MEM v0.7.10 (Li 2013) at default parameters. Local realignment around indels was performed using the *RealignerTargetCreator* and *IndelRealigner* utilities of the Genome Analysis Toolkit v3.4 (McKenna *et al.* 2010). We sorted and merged BAM files with added read group information using Picard tools v1.79 (<http://broadinstitute.github.io/picard/>). We called sequence polymorphisms using the FreeBayes

v1.0.1 haplotype-based variant detector (Garrison & Marth 2012). To optimize memory usage and to remove noisy alleles, we required that a minimum of 10 reads support an allele in order for it to be considered. Also, only the best seven alleles were evaluated for each locus. Lastly, the minimum fraction of alternate to reference reads was adjusted depending on ploidy, according to manual recommendations (<https://github.com/ekg/freebayes>). We used the default setting of 0.2 for diploids, and values of 0.1 and 0.05 for tetraploids and hexaploids, respectively.

For *H. tuberosus*, all individuals were genotyped as hexaploids. From this set, we isolated markers likely to segregate within the diploid subgenome (detailed below), which were used for a subset of the phylogenetics and population genetics analyses. For *H. grosseserratus*, *H. hirsutus* and *H. divaricatus* samples, two FreeBayes runs were performed, with the aim of maximizing variant detection. Specifically, all *H. grosseserratus*, *H. hirsutus* and *H. divaricatus* samples were genotyped simultaneously, once as diploids and once as tetraploids. Genotypes for each species were subsequently extracted from the call set of correct ploidy.

#### **4.2.5 Variant filtration and genotyping error rates**

Genotype calls for all taxa and ploidies were filtered using the *vcfilter* utility of *vcflib*, part of the FreeBayes package (<https://github.com/ekg/vcflib>). We retained biallelic SNPs with values for QUAL and MQ > 30. We also capped minimum depth per genotype at 6, 8, and 10 reads for diploids, tetraploids, and hexaploids, respectively, and coded genotype inferences based on fewer reads as missing data. We also selected only loci with data in at least 60% of samples and minor allele frequency > 1%. For the *H. tuberosus* call set, to remove likely between-homeologue variants, we further pruned sites with observed heterozygosity > 60%. Lastly, we removed samples with over 50% missing data at filtered SNPs.

We evaluated post-filtering genotyping error rates by tallying the percentage of sites at which replicates of the four *H. tuberosus* samples were assigned different genotypes. When considering each of the five heterozygote classes in a hexaploid distinct, inferred error rates were 0.1. Note that these estimates are comparable to or lower than those previously reported for polyploid taxa (e.g. Cornille *et al.* 2016). Most genotyping errors were cases in which replicates of the same genotype were assigned to different but adjacent polyploid heterozygote classes (e.g. simplex – duplex; duplex – triplex), and as such are unlikely to have a large impact on the accuracy of our population genomics or genome-wide association (GWA) analyses. When considering heterozygote classes separated by 1 allele to be equivalent, inferred hexaploid error rates decreased to 0.04. Likewise, when considering all heterozygotes equivalent, as is the case in the diploidized *H. tuberosus* SNP set, or for the subset of diploid-like genetic models used in our GWA, error rates were 0.04.

#### **4.2.6 Clonal relationships in the *H. tuberosus* collection**

To identify clonal genotypes, we used the R package SNPRelate (Zheng *et al.* 2012). Because SNPRelate does not support polyploid VCF files, we used a diploidized version of the filtered SNP set, for which all heterozygotes were considered equivalent. We computed pairwise identity-by-state (IBS) values for all *H. tuberosus* samples, including the four replicates. The smallest IBS value obtained between DNA replicates was 97.7%, while the average IBS between all samples was 85.6%. All IBS values calculated between replicate DNA samples were part of a peak formed at the upper tail of the distribution (Appendix Figure B.3). With this information, we set a conservative IBS threshold of 97% above which genotype pairs were considered to be clones. As expected, with this criterion, no *H. tuberosus* accession originally obtained from seeds

was inferred to be part of a clonal relationship. For all subsequent analyses, unless otherwise specified, we randomly selected one sample from each clonal series (Appendix Table B.10), such that no two *H. tuberosus* genotypes had an IBS value larger than 0.97. When greenhouse-propagated plants were available for a clonal series, the clone selection was done from among these samples, such that clones included in the unique genotype set were also used for the common garden and GWAS experiments (see below). The unique genotype set consisted of 428 *H. tuberosus* samples and 84,365 SNPs filtered using the criteria above.

#### **4.2.7 Confirmation of *H. tuberosus* ancestry and taxonomic identifications**

We previously showed, using organelle DNA and rDNA data, that *H. tuberosus* ( $2n = 6x = 102$ ) is an autoallohexaploid that resulted from hybridization between the diploid *H. grosseserratus* ( $2n = 2x = 34$ ) and the tetraploid *H. hirsutus* ( $2n = 4x = 68$ ). We also showed that *H. hirsutus*, the donor of the tetraploid subgenome of *H. tuberosus*, is a chromosome-doubled offspring of *H. divaricatus* ( $2n = 2x = 34$ ; Bock *et al.* 2014). These relationships have more recently been confirmed using broader taxonomic sampling across the *Helianthus* genus (Baute *et al.* 2016).

To verify these results and to detect hybrids or potentially mis-identified samples that may have been inadvertently included in our dataset, we performed a series of principal component analyses (PCAs). In addition to the four perennial *Helianthus* taxa, we included a whole genome shotgun sequencing (WGSS) - derived call set consisting of 150 samples of the annual congener *H. annuus*, as well as 10 samples of *H. divaricatus*, and 10 samples of *H. grosseserratus*. The *H. annuus* samples, used in the first and second PCAs, were included because hybrids between *H. tuberosus* and *H. annuus* resulting from breeding efforts may be

present in *H. tuberosus* genebank material (Kays & Nottingham, 2008). The *H. divaricatus* and *H. grosseserratus* WGSS-derived variants, used in the second PCA, were included to verify that no major biases were introduced when jointly analyzing SNPs obtained through GBS and WGSS. All call sets were filtered as above, with the exception of the 1% minor allele frequency filter, which was applied across call sets in a PCA. We used the R package adegenet v2.0.1 (Jombart & Ahmed 2011), which allows for multiple ploidies to be represented in the same analysis.

#### **4.2.8 Isolation of *H. tuberosus* subgenome-specific markers**

We relied on SNP data for *H. grosseserratus*, *H. hirsutus* and *H. divaricatus* to isolate markers likely to segregate within the diploid and the tetraploid subgenomes of *H. tuberosus*. To isolate diploid subgenome markers, we first identified biallelic variants that are polymorphic ( $10\% < \text{allele frequency} < 90\%$ ) in samples of the diploid subgenome donor species (*H. grosseserratus*) but monomorphic in samples of the tetraploid subgenome donor species (*H. hirsutus*/*H. divaricatus*). We asked that these markers be scored in at least five samples of each progenitor species, and identified the subset shared with *H. tuberosus*. On average, selected SNPs were scored in 47 (50 %) samples of *H. grosseserratus* and 46 (58 %) samples of *H. hirsutus*/*H. divaricatus*. For *H. tuberosus*, variants were filtered as above, with the exception of maximum observed heterozygosity, for which we did not set an upper bound. This was because we expected observed heterozygosity to be elevated for markers fixed for one allele in the tetraploid progenitor, and polymorphic for the other allele in the diploid progenitor. A total of 35,705 SNPs were polymorphic in *H. grosseserratus* but not in *H. hirsutus*/*H. divaricatus*. Of these, 6,190 (17%) segregated among *H. tuberosus* samples with a minor allele frequency larger

than 1%. We considered these markers as candidate diploid subgenome SNPs. We used the same approach to identify markers likely to segregate within the tetraploid subgenome. A total of 37,129 SNPs were polymorphic in *H. hirsutus*/*H. divaricatus* but not in *H. grosseserratus*. Of these, 5,366 (14%) were scored in *H. tuberosus* samples at a minor allele frequency larger than 1%. These were therefore considered candidate tetraploid subgenome SNPs.

We further excluded candidate diploid and tetraploid subgenome SNPs for which the dosage of the progenitor-derived alleles departed from expectations for a diploid or a tetraploid subgenome. For candidate diploid subgenome SNPs, we requested that 90% or more of *H. tuberosus* genotypes at each marker contain at most two copies of the allele inferred to have originated from *H. grosseserratus*. Similarly, for candidate tetraploid subgenome SNPs, we requested that 90% or more of *H. tuberosus* genotypes at each marker contain at most four copies of the allele inferred to have originated from *H. hirsutus*/*H. divaricatus*. Our reasoning was that candidate subgenome SNPs with large differences between observed and expected dosage can be the result of polymorphism erroneously assigned as diagnostic of one parental taxon. The majority of SNPs met the dosage filtering criterion (76% of diploid subgenome SNPs, and 89% of tetraploid subgenome SNPs). For downstream analyses using subgenome-assigned markers, *H. tuberosus* genotypes were converted from hexaploid to diploid or tetraploid format (Appendix Table B.11).

We note that our strategy for assigning SNPs to the two subgenomes of *H. tuberosus* relies on important assumptions that might not be met across all sites. For instance, we assumed that allele frequencies observed in contemporary samples of *H. grosseserratus* and *H. hirsutus*/*H. divaricatus* are representative of the genotypes that were involved in the ancestry of *H. tuberosus*. While evidence from organelle genomes indicates that *H. tuberosus* most likely



originated repeatedly (Bock *et al.* 2014), it is currently not known which genotypes of each parental species are most similar to the ones that participated in the polyploidization events. As well, our dosage filtering criteria does not account for the possibility of partial or complete homeolog loss, which has previously been shown to occur rapidly after polyploid formation (e.g. Buggs *et al.* 2012). Under this latter scenario, our filtering criteria are conservative, and would have retained only loci without extensive evidence of loss of parental species contributions. Considering these limitations, wherever possible, we compare subgenome-specific analyses with those based on the complete filtered call set.

#### **4.2.9 *H. tuberosus* phylogenetic analyses**

We performed all phylogenetic analyses using the filtered diploid subgenome call set (4,700 SNPs). To convert the data from VCF format to aligned PHYLIP format, we used the SNPhylo pipeline v20140701 (Lee *et al.* 2014). Briefly, SNPs for each sample were concatenated, and heterozygous positions were coded using IUPAC ambiguity codes. Sequences were then aligned using MUSCLE v3.8.31 (Edgar 2004). For phylogeny reconstruction, we used the IQ-TREE v1.4.2 software (Nguyen *et al.* 2015), which accounts for variability in the rate of sequence evolution by incorporating a FreeRate model as well as gamma rate distribution models. We identified the best model of evolution for the data using the -m TESTNEWONLY +ASC option of IQ-TREE. Note that the +ASC flag is used to account for ascertainment bias. We then inferred maximum-likelihood trees with node support evaluated using 10,000 ultra-fast bootstrap replicates (Minh *et al.* 2013). The final tree was viewed and manipulated for figure generation using FigTree v1.4.2 (Rambaut 2009).

#### **4.2.10 *H. tuberosus* population genetic structure**

To further understand genetic structure among sampled *H. tuberosus* individuals, we used PCA and the Bayesian model-based clustering implemented in STRUCTURE v2.3.4 (Pritchard *et al.* 2000). Both analyses were performed using the filtered *H. tuberosus* hexaploid call set (82,946 SNPs), as well as the diploid subgenome call set (4,700 SNPs) and the tetraploid subgenome call set (4,770 SNPs). The PCAs were conducted in the R package adegenet as above. For the STRUCTURE analysis based on the filtered hexaploid call set only, to optimize run times, we used a random subset of 5,000 SNPs, selected using the *vcfrandomsample* utility of vcflib. For each STRUCTURE run, we used the admixture model with correlated allele frequencies, with  $10^5$  burn-in replicates, followed by  $10^5$  replicates. We performed 20 runs for each value of  $K$  ranging from 1 to 10. To select the most likely value of  $K$  according to Evanno *et al.* (2005) and Pritchard *et al.* (2000), we used Structure Harvester (Earl *et al.* 2012). Results were plotted using the pophelper v1.2.1 R package (Francis 2016).

#### **4.2.11 Classification of native, cultivated, and invasive *H. tuberosus***

Because of the extensive cultivation history of *H. tuberosus* across its ancestral distribution range, it can be difficult to unambiguously identify which of the naturally-occurring North American populations contain native wild material or escapes from cultivation. Classification can be challenging for cultivated lines as well. This is because cultivated *H. tuberosus* has not been subjected to prolonged and intense selection, and as a result is more morphologically similar to its wild progenitor than most modern crops are (Kays & Nottingham 2008). Also, while the bulk of samples maintained by PGRC, IPK, and INRA should represent cultivated accessions that have been selected for increased tuber and general biomass production

(Kays & Nottingham 2008), wild or invasive samples, as well as curation errors, may still be present. Here, we classified samples as native, cultivated, invasive, or unknown using a combination of collection notes and genetic data.

To identify native samples, we first labeled the 49 USDA accessions as “natural habitat” ( $N = 34$  accessions) or “artificial habitat” ( $N = 15$  accessions; Appendix Table B.1). The “natural habitat” accessions had no notes indicative of possible contribution from cultivated material, while the “artificial habitat” accessions were collected from disturbed sites like railroad tracks, edges of crop fields, and abandoned farms, or had information that was otherwise indicative of cultivated status. We then compared the grouping of samples from these two habitat categories in the PCA that included only *H. tuberosus* samples, and using the best supported STRUCTURE model ( $K = 2$ ). “Natural habitat” samples clustered towards negative values on PC1 (average PC1 score -7.7), whereas “artificial habitat” samples were more variable, and pulled towards positive PC1 values (average PC1 score -3.9), similarly to the cultivated samples described below (Appendix Table B.1). The STRUCTURE membership coefficients followed the same pattern (Appendix Figure B.1). With this information, we set a PC1 score of -5 as threshold for the identification of native material. All USDA-collected individuals with PC1 scores below this value were designated as wild ( $N = 174$ ), irrespective of whether they were originally obtained from natural or artificial North American habitats. The remaining USDA-collected samples ( $N = 67$ ) were left as unknown.

To identify which of the PGRC, IPK, and INRA accessions are likely of cultivated status, we first used accession passport data, or information detailed in Kays & Nottingham (2008). For INRA accessions, we also used the tuber images presented for this collection in Serieys *et al.* (2010). Aside from origin information and commercial clone name, we required that notes or

images confirming the production of round or ovoid tubers be available. Because *H. tuberosus* has been selected as a tuber crop, tuber shape is one of the few morphological criteria that can be used to separate cultivated from wild samples. For accessions with no tuber morphology information, we used documentation from clones of those samples, when clones were available. We note that five of the seed-propagated INRA accessions represented in our dataset were described as wild material (Serieys et al. 2010). These samples also grouped with the rest of the wild samples in the PCA (PC1 score below -10), and had tuber morphologies characteristic of wild *H. tuberosus* (Serieys et al. 2010). We therefore labeled these five accessions as wild. For *H. tuberosus* maintained at PGRC, we also labeled as cultivated those accessions that had notes indicating they were breeding and research clones and/or originated from the Morden Research Station, in Manitoba, Canada. The Morden Station was one of the main research centers developing *H. tuberosus* lines with increased tuber yields (Kays & Nottingham 2008). Lastly, we excluded from the cultivated category five accessions that had PC1 scores lower than -5, and were therefore more similar to wild samples. The remaining cultivated samples ( $N = 275$ ) were distinct according to PC1 scores from wild material (average cultivated PC1 score 12.4).

We identified as invasive all *H. tuberosus* samples obtained in 2013 from European populations ( $N = 112$ ), as well as two PGRC and 19 IPK accessions. The two PGRC accessions and nine of the 19 IPK accessions were clones of material we obtained from European populations. The remaining 10 invasive IPK accessions grouped with high support with the majority of invasive samples in our dataset (Clade 3, Figure 4.1C), and were distinct from cultivated or wild material, according to the phylogenetic reconstruction. We used three additional lines of evidence when including these samples as invasive. First, five of the 10 samples had collection notes indicating they were obtained from European sites (Poland).

Second, there were no passport or collection notes indicating wild or cultivated status for these samples. Third, neither of the 10 IPK accessions had clones present in the PGRC or INRA collections. Considering the high degree of redundancy between the three germplasm collections for other accessions, we considered it unlikely that these are unlabeled wild or cultivated lines.

Finally, we note that even if sample categorization using the criteria outlined above may have resulted in the misclassification of a small number of samples, these are unlikely to affect our results, which are based on overall trends between the wild, cultivated, and invasive categories. Several lines of evidence indicate that mis-classifications are rare, if present. First, a PCA performed for *H. tuberosus* and its progenitor species showed that, among all *H. tuberosus* samples, the material we labeled as wild is most similar to samples of the diploid and tetraploid progenitor species. This is expected if the wild group is representative of the ancestral state for the species (see Figure 4.1A). Second, analysis of tuber morphology for the subset of wild and cultivated samples represented in our common garden experiment are in agreement with patterns expected for a tuber crop. Specifically, the tubers of wild samples were narrow and elongated, indicating that these plants are likely not of cultivation value. By contrast, cultivated samples produced tubers that were significantly larger and more round (see Common garden section below). Third, there was a significant latitudinal cline for flowering time among wild samples. While recent studies have shown that this pattern can also evolve relatively quickly in introduced species as well (Colautti & Barrett 2013), it has most commonly been reported for locally adapted native plant populations, including in annual and perennial sunflowers (e.g. Blackman et al. 2011; Kawakami et al. 2011).

#### **4.2.12 Identification of invasiveness traits in *H. tuberosus***

To investigate whether certain traits have a disproportionate contribution to invasion success in *H. tuberosus*, we compared phenotypes of invasive and non-invasive samples using a greenhouse and a common garden experiment. In line with previous studies (Brown & Eckert 2005; Lavergne & Molofsky 2007; Pyšek & Richardson 2007), we expect important phenotypic determinants of invasiveness to show extreme values for invasive *H. tuberosus*.

##### **4.2.12.1 Phenotype data collection**

To minimize maternal effects, prior to the greenhouse and common garden experiments, we propagated experimental plants at the University of British Columbia (UBC) horticulture greenhouse for two generations from tubers. We scored allelopathy and investment in clonal growth (quantified as the number of tubers), as well as additional traits that have previously been considered to drive invasiveness in other species (Muth & Pigliucci 2006; Pyšek & Richardson 2007; Razanajatovo *et al.* 2016). These included stem diameter, plant height, number of branches, flowering time, number of inflorescences, inflorescence diameter, or self-compatibility (see Appendix Table B.3 for the complete list of traits). As positive controls of phenotypic differentiation, and to confirm the validity of our classification for wild and cultivated samples, we additionally scored descriptors of tuber shape, size, and yield. Given that large tubers have been the main focus of selection during *H. tuberosus* domestication (Kays & Nottingham 2008), we expected these traits to be diverged between wild and cultivated genotypes.

To quantify allelopathic potential, we followed previous work on the system (e.g. Vidotto *et al.* 2008; Tesio *et al.* 2010; Filep *et al.* 2016), and used germination bioassays. We tested the effect of aqueous leaf extracts from greenhouse-propagated plants on germination and seedling

growth of tomato. Tomato is a commonly-used bioassay test species that has previously been shown to be sensitive to *H. tuberosus* extracts prepared using a similar approach (Vidotto *et al.* 2008). We used 160 *H. tuberosus* genotypes, of which 119 were classified as either wild ( $N = 43$ ), cultivated ( $N = 50$ ), or invasive ( $N = 26$ ). Additionally, to compare allelopathy of *H. tuberosus* to that of its non-invasive progenitor species, we included samples of *H. grosseserratus* ( $N = 76$ ), *H. hirsutus* ( $N = 30$ ), and *H. divaricatus* ( $N = 29$ ).

All experimental plants were grown at the UBC horticulture greenhouse in pots (15 cm diameter, 18 cm height) with potting soil, and maintained under a 12-hour light cycle with a minimum temperature of 20° C and a maximum temperature of 25° C. We collected 4 grams of leaf tissue from 3.5 month-old plants, which we dried for 2.5 days in an incubator at 37° C, in the presence of silica gel. We ground the dried leaf tissue for 1 min using a coffee grinder, and soaked the powdered material in 10 mL of MilliQ water to obtain a dilution of 40% w/v of fresh material. Following incubation for 24h at 4° C, we thoroughly mixed the samples using a vortex, and separated the extract from the leaf material using centrifugation. Extracts were stored at 4° C until further use.

Bioassays were set up in 60 x 15 mm Petri dishes, on No. 1 Whatman filter paper (Whatman International Ltd.). In each Petri dish, we evenly placed 9 tomato seeds that had previously been sterilized by soaking in 70% ethanol for 1 min, and added 1 mL leaf extract. With each experimental batch, we included a control bioassay, for which we substituted the leaf extracts with an equal volume of MilliQ water. Petri dishes were sealed with parafilm, and stored in a closed seed germination box in the dark at room temperature for seven days. Each control and test treatment was replicated four times (304 experiments x 4 replicates = 1,216 bioassays). We recorded germination daily. Also, on day 7, we scanned 5,032 of the available tomato

seedlings and used image analysis to quantify their size. We used these data to assess allelopathy using six metrics. These included total germination, the speed of germination, the speed of accumulated germination, the coefficient of rate of germination, tomato seedling radicle length, and tomato seedling plumule length (Appendix Table B.3).

We scored 14 other traits using a common garden established from spring to fall 2015 at the UBC Totem Field research station (49°16' N, 123°14' W, 45 m elevation). Plants were sprouted at the UBC horticulture greenhouse from 2-3 cm tubers or from tuber fragments of the same size in the case of accessions for which only larger tubers were available. Between April 17<sup>th</sup> and May 21<sup>st</sup>, we transplanted 300 accessions replicated 1-3 times to the field. Prior to the transplant, to facilitate acclimatization of the plants to field conditions, we gradually lowered the greenhouse maximal and minimal temperatures for nine days, down to a maximal temperature of 18° C and a minimal temperature of 9° C. In all, 726 1-month old plants (15-30 cm tall) were included in the transplant after this hardening step.

Genotype replicates were randomly assigned a position in one of three incomplete blocks, each of which consisted of 14 rows and 18 columns. Blocks were separated by a minimal distance of 5 meters. Prior to planting, to limit competition from weeds, we covered the experimental plots with DeWitt Pro 5 weed barrier (DeWitt Co., Sikeston, MO). Within each block, to minimize plant-to-plant competition and to prevent overlap in tuber production between adjacent accessions, we spaced plants on 1.2 meter centers. We decided on this planting distance based on a preliminary field trial that we performed at the same site during the previous year. After transplanting, we monitored and hand-watered the plants as needed for one week. During this time, we replaced any plants that died with greenhouse-propagated replicates of the same genotypes, when those were available. To minimize wind damage during the growing season, for



all three blocks, we set up 4x4 wooden fence posts at both ends of each row and tied nylon rope at low, mid, and high levels along all rows. Experimental plants were individually stabilized using bamboo stakes tied to these ropes. During the growing season, we watered the plants as needed using sprinkler irrigation.

We monitored the onset of flowering three times per week, and recorded the date at which reproductively active flowers were present. To assess self-incompatibility, we bagged three flower heads per plant before anthesis, and recorded the presence or absence of seeds at the end of the growing season. By the end of the experiment, of the 726 initial plants, we excluded 117 that either died because of transplant shock or herbivory, or suffered considerable damage during a windstorm that occurred in Vancouver on August 29<sup>th</sup>. We harvested the remaining 609 plants, which represented 298 genotypes, between October 20<sup>th</sup> and November 20<sup>th</sup> 2015. At this time, most (74%) had flowered. We decided to process all plants regardless of their flowering status, to avoid the increasingly frequent frosts occurring towards the end of November, which would have impeded the harvest of below-ground biomass. A number of measurements were collected before plants were cut down. These included the diameter (in mm) of the main stem recorded at 10 cm above ground level with a digital caliper, the height (in cm) of the primary stem recorded from ground level to the apex, as well as the total number of branches and inflorescences. Note that we only counted branches that were at least 2 cm in length. For the plants that flowered during the growing season, we also collected five randomly selected inflorescences (where available), which we dried for 3 days in an incubator at 37° C. We then measured (in mm) the disk diameter of the dried material using a digital caliper, and recorded for each plant the values averaged across the five inflorescences (Appendix Table B.3).

We harvested the tubers by loosening the soil around each plant, and digging up the plants. For most samples, tubers were concentrated at the base of the main stem and/or were still attached to rhizomes. In the few cases where tubers were more dispersed, we used the direction of growth, which is outward from the plant, as well as differences in tuber shape and color to correctly identify the tuber yields of neighboring plants. The tubers from each plant were washed, counted, and weighed in the field. In all, for the 609 plants, we obtained 75,915 tubers with a total weight of 1.02 tones. From the total tuber yield of each plant, we randomly selected five tubers (609 plants x 5 tubers per plant = 3,045 tubers) which we scanned. We processed the resulting images using the software Tomato Analyzer (Brewer et al. 2006). We visually inspected the processed images and correct automatically-inferred tuber edges, whenever necessary. We then calculated four parameters of tuber shape and size. These included tuber area, perimeter, curved length, and fruit-shape index (the ratio of curved length to maximum width; Appendix Table B.3). Tuber measurements for each trait were averaged over the five tubers scanned per accession.

#### **4.2.12.2      Phenotype data analyses**

Data were analyzed using R version 3.1.2 (R Core Team 2015). For the common garden experiment, we removed the self-incompatibility trait, since no seeds were recovered for any of the samples. To improve assumptions of normality, we rank-transformed the data using the GenABEL package (Aulchenko et al. 2007). We quantified overall phenotype differentiation between samples using a PCA computed using the prcomp function in R. For this, we included 11 of the 19 traits with data for 90% or more of samples, and replaced missing values with the average observed across the population.

To identify traits with significantly extreme values in invasive *H. tuberosus*, we compared 225 samples that could be classified as wild, cultivated, or invasive using the criteria outlined above. We excluded unknown samples because this category is likely to include a mix of material from the other three categories. For allelopathy, an additional level of comparison between *H. tuberosus* and its non-invasive progenitor species was used. For each of the six allelopathy traits, we ran two types of fixed ANOVAs. The first contained data from all four taxa, and had species (*H. grosseserratus*, *H. tuberosus*, *H. divaricatus*, and *H. hirsutus*) as predictor variable. The second contained only *H. tuberosus* samples, and had sample category (native, invasive, cultivated) as the predictor. For the traits scored in common garden plants, we used mixed ANOVAs implemented with the `lme` function in the `nlme` package (Pinheiro et al 2015). We treated sample category as a fixed effect, and replicate nested within genotype as random effects. We ran all models using only *H. tuberosus* data with and without correction for population structure. The versions corrected for population structure also included as covariates the scores for the first two PC axes calculated using all filtered markers (82,957 SNPs). Significance levels were adjusted using the sequential Bonferroni method (Holm 1979).

We additionally implemented generalized linear mixed models (glmm) using Markov chain Monte Carlo in the `MCMCglmm` R package (Hadfield, 2010). For these analyses, we treated sample category as the response variable, and concomitantly considered all traits as predictor variables. Prior to the `MCMCglmm` analysis, we removed missing data as follows. First, for the 13 traits measured in the common garden experiment, we collapsed the 1-3 replicates available for each genotype to one value per genotype. To do this, we verified whether there was a significant block effect by fitting linear mixed effect models to the rank-transformed data for each trait with maximum likelihood. We used the `lme` function of the `nlme` package,

treating genotype as fixed effect and block as random effect. We compared the fit of these models to the fit of linear models without a block term. For flowering time, plant height, disk diameter, tuber number, and tuber fruit-shape index, block was not significant. For these traits, we used the average of rank-transformed trait values for the replicates available for a given genotype. For the eight other traits for which block was significant, we used the least-squares means, which was calculated based on the linear mixed effect models described above, using the package `lsmeans` (Lenth and Hervé 2015). We note that the same approach for calculating mean trait values was used for the association mapping. Lastly, we included as predictors only the 11 traits with data in 90% or more of samples. For these traits, we replaced missing values with the average recorded across the experiment. Of all phenotypes, we selected seven through a stepwise withdrawal of traits that have a Variance Inflation Factor greater than 5. Similarly to the mixed ANOVAs described above, we fit equivalent models with correction for population structure, by including genetic PC1 and PC2 scores as additional fixed effects. MCMCglmm ran for  $10^9$  iterations.

#### **4.2.12.3 Effect size estimates and comparisons with previous studies**

For traits that we infer to be associated with invasion success and domestication in *H. tuberosus*, we calculated Hedges' *g* unbiased effect size estimator using the mean, standard deviation, and sample sizes for each sample group in the `compute.es` package in R (Del Re 2013). For invasiveness traits, effect sizes were conservatively obtained from comparisons between the invasive category and the phenotypically most similar non-invasive category. For domestication traits, effect sizes were obtained from comparisons between the cultivated and

native categories. In all cases, we used the mean values obtained from replicates available for a given genotype as described above.

We additionally searched the literature for estimates of differentiation in traits associated with invasion success in other systems. We used Google Scholar, references from published papers, and the dataset reported in a recent meta-analysis (Colautti et al. 2009). Papers were trimmed using a series of selection criteria. First, we only considered studies that compared trait differentiation between invasive and non-invasive samples grown in a common greenhouse or field environment. Second, we limited our selection to studies that reported trait values for plants grown under benign conditions (similarly to those reported here), and where inferences were based on a minimum of 10 native and 10 invasive populations. Third, we included only examples in which candidate invasiveness traits were confirmed to be significantly differentiated in the expected direction between the native and invasive ranges. Because raw phenotype data is rarely given in materials made available at publication stage, we used ANOVA F-statistics reported for comparisons between native and invasive ranges. This statistic was used along with reported sample sizes to calculate Hedges' *g* effects in R, with *compute.es*.

#### **4.2.13 Genetic architecture of invasiveness**

##### **4.2.13.1 Relationship between heterozygosity and trait values**

We used Pearson's correlation analyses to test the relationship between genome-wide heterozygosity estimated using the 6x, 4x, and 2x call sets and trait values. For tuber number, to account for the possibility that population structure may be driving the observed patterns, we repeated these analyses for each of the three sample categories separately. For invasive samples, we additionally investigated the power available to detect an effect of the same magnitude to the

one observed in non-invasive (native, cultivated and unknown) samples. We used the function 'pwr.f2.test' from the *pwr* R package (Champely 2012). We set the critical  $\alpha$  level to 0.05. Effect size was estimated as the  $R^2$  from a linear model between rank-transformed heterozygosity and tuber number values. The model considered all non-invasive samples, and was built using the linear model-fitting R function 'lm'.

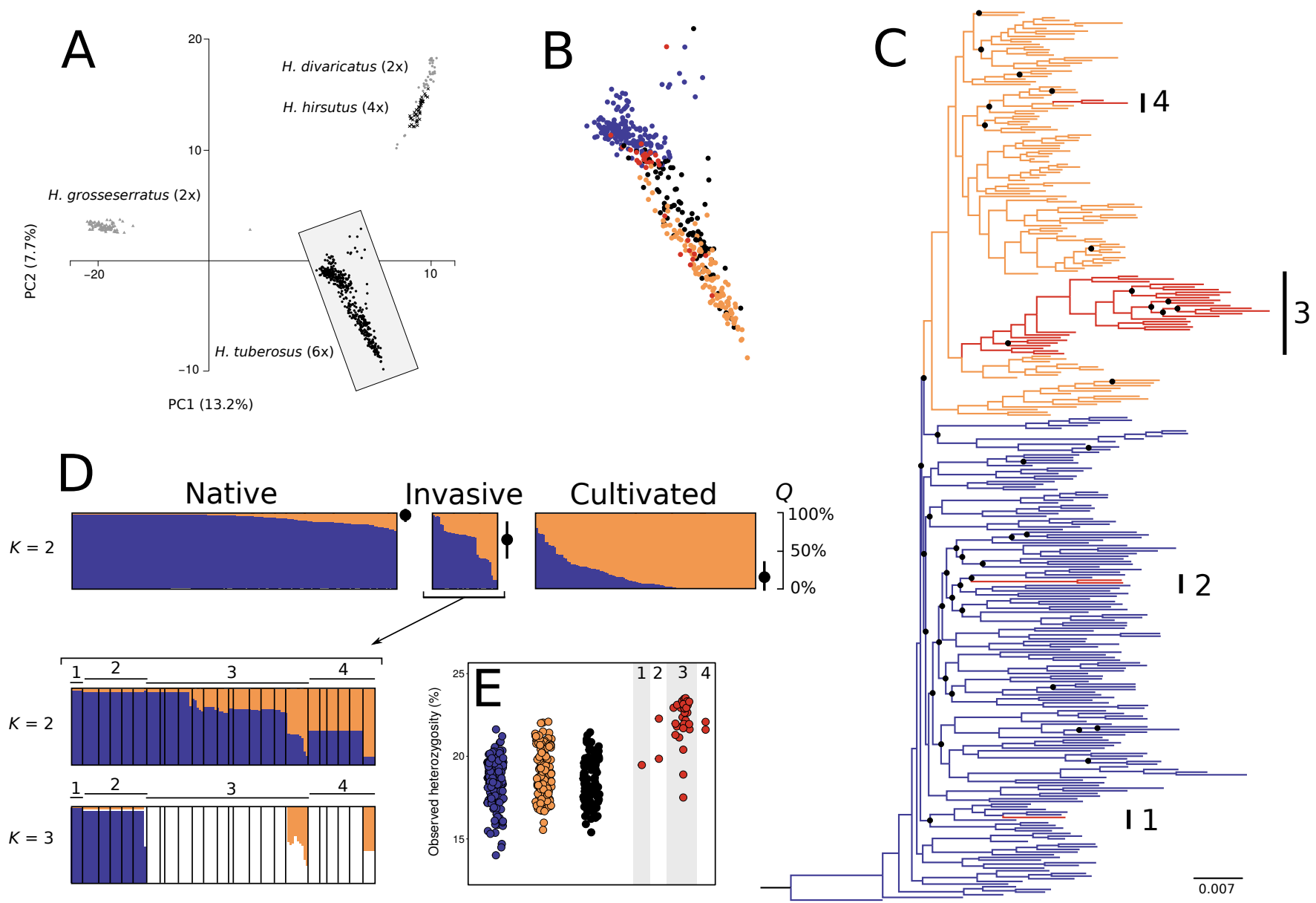
#### **4.2.13.2 Association mapping**

We used the R package GWASpoly (Rosyara *et al.* 2016), which extends the Q (or P) + K mixed linear model for association analyses to allow the use of diploid and polyploid data. We included 305 *H. tuberosus* accessions for which genotype and phenotype data was available. The genotypes were in hexaploid format, and were based on 43,276 SNPs. These markers were obtained using the same filtering criteria as above, with the exception of the call rate threshold, which we increased from 60% to 90%. We ran mixed linear models that account for kinship (K), kinship and population structure as estimated using the top two principal components (K + P), or kinship and population structure as estimated using membership coefficients for the two-populations STRUCTURE model (K + Q). The kinship matrix, which is included as a random effect in the associations, was calculated in GWASpoly as the realized relationship matrix, computed after imputing missing marker data with the population mean. We used additive, dominant, and diplo-general (diploid-like) marker-effect models. When the same associations were identified by different marker-effect models, we recorded the associations with the lowest *P* value. To establish the significance level, we used a conservative 5% Bonferroni threshold (0.05 / number of tested SNPs). All association models were evaluated using quantile-quantile plots. Additionally, to verify that population stratification is correctly accounted for, we

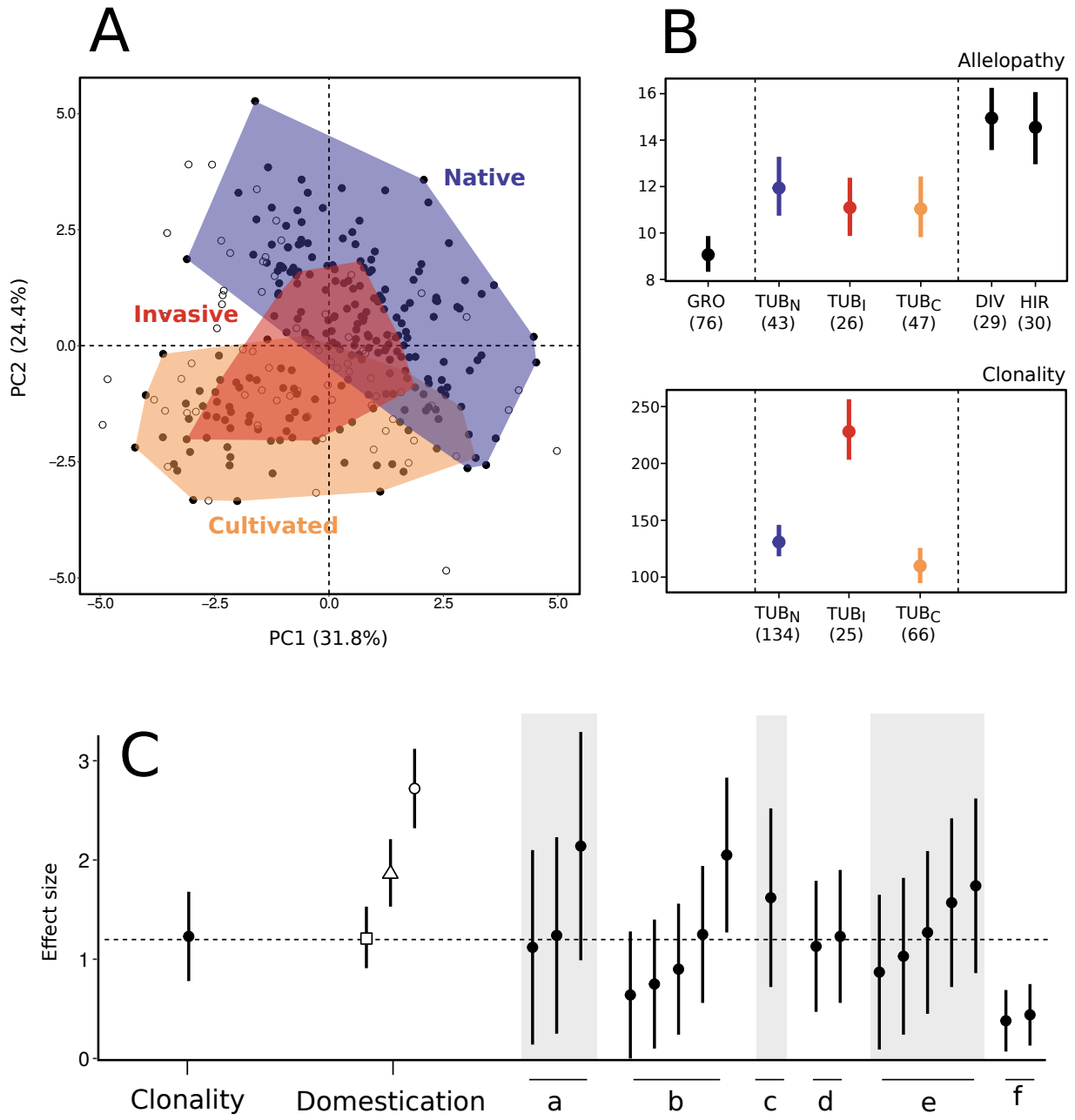
calculated inflation factors ( $\lambda$  values) in R using the GenABEL package. The proportion of variance explained ( $R^2$ ) for the top most significant SNP for each association was calculated by fitting linear models with trait values as response variable and genotype as explanatory variable. Finally, we generated manhattan and quantile-quantile plots using the R package qqman (Turner 2014).

**Figure 4.1: Population structure and genetic diversity of *Helianthus tuberosus*.** (A) PCA of 427 *H. tuberosus* non-clonal genotypes and 175 genotypes of the progenitor species based on 27,396 shared SNPs. (B) Magnification of *H. tuberosus* genotypes used in the PCA, with samples classified as native (blue), invasive (red), cultivated (orange), or unknown (black). (C) Maximum-likelihood phylogeny based on 4,700 diploid subgenome SNPs, for the subset of 333 non-clonal native, invasive, and cultivated genotypes. Numbered clades indicate inferred origins of invasive genotypes. The tree is rooted with a sample of *H. grosseserratus*, and black circles indicate bootstrap values below 70%. (D) Bayesian clustering of native, invasive and cultivated genotypes. The top barplot indicates individual membership coefficients ( $Q$ ) for 333 non-clonal genotypes excluding unknown samples. Black circles indicate mean  $Q$  ( $\pm$  SD) for one of the two inferred clusters calculated for each of the native, invasive, and cultivated groups. Kruskal-Wallis tests for differences in  $Q$  among groups:  $\chi^2_2 = 251.69$ ,  $P < 2.2 \times 10^{-16}$ ; pairwise Kruskal–Wallis comparisons using Nemenyi-test, and sequential Bonferroni adjusted  $P$  values:  $P_{\text{native-cultivated}} < 2.2 \times 10^{-16}$ ,  $P_{\text{native-invasive}} = 7.8 \times 10^{-8}$ ,  $P_{\text{cultivated-invasive}} = 2.57 \times 10^{-4}$ . The bottom barplots indicate admixture proportions at  $K=2$  and  $K=3$  for the invasive group, with each genotype multiplied by the number of observed clones. Numbers above the plots indicate the four inferred origins of invasive genotypes, and vertical black lines are used to delineate sampling locations. (E) Genome-wide heterozygosity for native, cultivated, unknown and invasive samples calculated using 82,957 filtered SNPs. Invasive samples are divided by inferred origin. Kruskal-Wallis tests for differences among groups:  $\chi^2_3 = 88.705$ ,  $P < 2.2 \times 10^{-16}$ ; pairwise Kruskal–Wallis comparisons using Nemenyi-test, and sequential Bonferroni adjusted  $P$  values:  $P_{\text{native-cultivated}} = 0.06$ ,  $P_{\text{native-invasive}} = 6.66 \times 10^{-16}$ ,  $P_{\text{cultivated-invasive}} = 6.65 \times 10^{-9}$ .

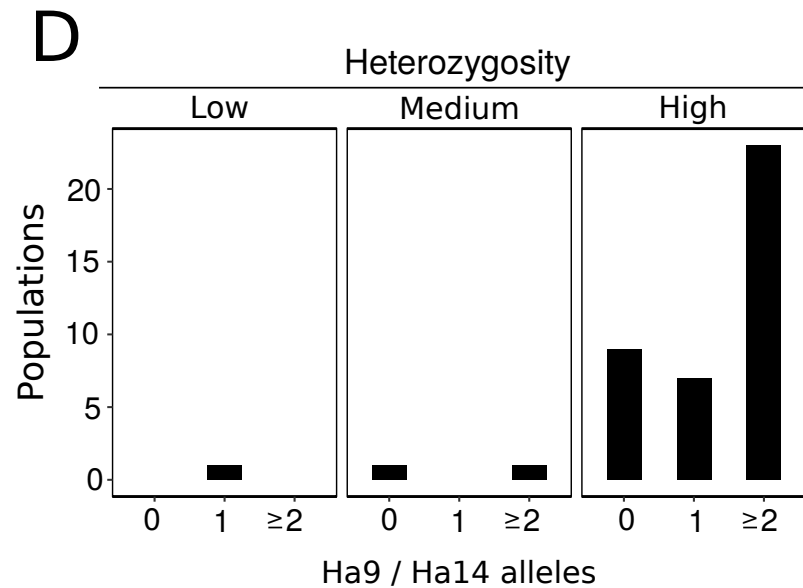
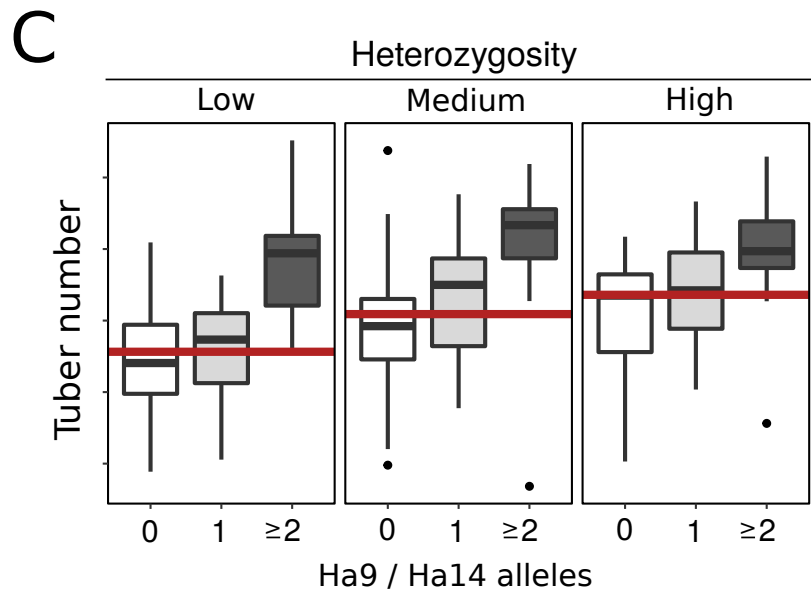
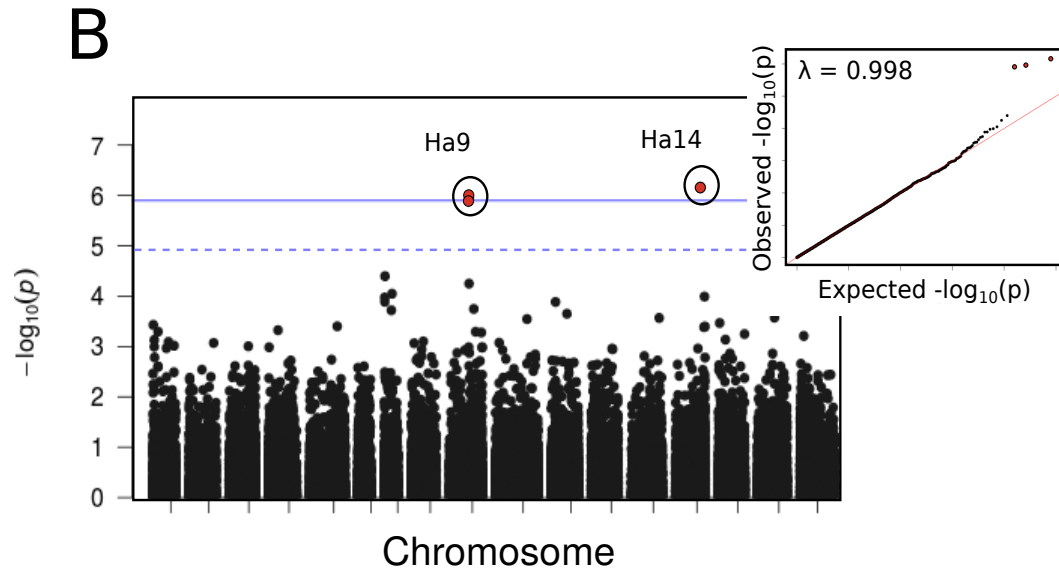
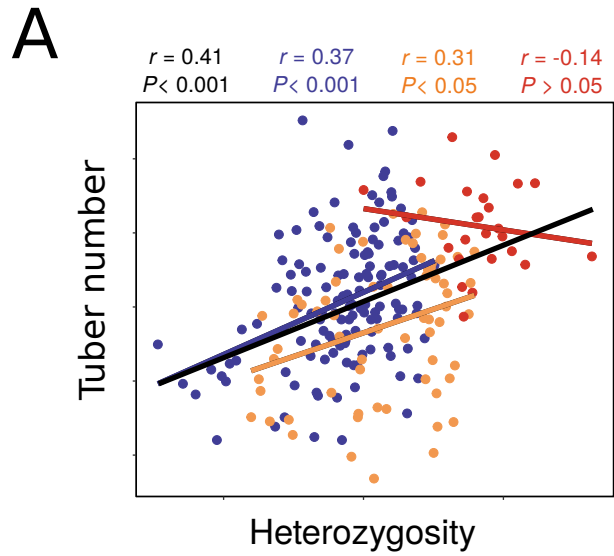




**Figure 4.2: Phenotypic determinants of invasiveness in *Helianthus tuberosus*.** (A) PCA based on 11 quantitative traits with data in 90% or more of samples for 225 native, invasive, and cultivated *H. tuberosus* samples (filled circles), and 80 unknown *H. tuberosus* samples (empty circles). Polygons enclose native (blue), invasive (red), and cultivated (orange) samples. (B) Trait means and 95% bootstrapped confidence intervals for allelopathy and clonality in *H. grosseserratus* (GRO), *H. divaricatus* (DIV), *H. hirsutus* (HIR), and for native (TUBN), invasive (TUBI) or cultivated (TUBC) *H. tuberosus*. Allelopathy is given as speed of accumulated germination estimated using bioassays, with lower values indicating higher toxicity. Clonality is quantified as the number of tubers. Sample sizes for each group are given in parentheses. Kruskal-Wallis tests for differences among groups: allelopathy between species ( $\chi^2_3 = 54.02$ ,  $P = 1.1 \times 10^{-11}$ ), allelopathy within *H. tuberosus* ( $\chi^2_2 = 1.98$ ,  $P = 0.371$ ), clonality within *H. tuberosus* ( $\chi^2_2 = 36.94$ ,  $P = 9.47 \times 10^{-9}$ ). (C) Hedges' effect size estimates (+/- 95% CI) for clonality (tuber number), three *H. tuberosus* domestication traits including total tuber weight (empty square), average tuber weight (empty triangle), and tuber shape index (empty circle), and 18 proposed drivers of invasiveness in other systems (references a-f; Appendix Table B.8). The dashed line marks the effect size estimate for clonality in *H. tuberosus*.



**Figure 4.3: Genetic architecture of invasiveness in *Helianthus tuberosus*.** (A) The relationship between rank-transformed values for heterozygosity estimated from 82,957 filtered SNPs and rank-transformed values for tuber number. Correlations are given for all samples combined (black), as well as native (blue), cultivated (orange), and invasive (red) samples. (B) Manhattan and quantile-quantile plots for tuber number association analyses. The level of statistical significance is given using Bonferroni (full line) and FDR (dotted line). The inflation factor ( $\lambda$ ) is shown in the quantile-quantile plot. (C) Boxplot of tuber number for low, medium, and high heterozygosity classes. Within each heterozygosity class, samples are divided based on the number of alleles associated with increased tuber number production. Red lines indicate average tuber number production per heterozygosity class. (D) Number of populations covered by invasive clones, plotted based on heterozygosity class and number of alleles associated with increased tuber number production.



## **Chapter 5: Conclusions**

With this dissertation, I aimed to improve our understanding of the genetics of rapid evolutionary transitions. I started this research with a review of the validity of a common assumption in the plant evolutionary biology literature, that organelle DNA variation is selectively neutral. As outlined in Chapter 2, this assumption, if incorrect, has broad implications, that include understanding contemporary evolution in response to anthropogenic stressors. In Chapter 3, I use genome skimming to identify the parentage of *H. tuberosus*, a perennial sunflower that formed *via* polyploidization, the most rapid form of speciation. This research highlights the utility of genome skimming for clarifying the evolutionary history of phylogenetically challenging taxa (see also Straub et al. 2012; Kane et al. 2012). Finally, in Chapter 4, I combine greenhouse and common garden data with large scale genotyping in *H. tuberosus*. I identify a major invasiveness trait in this system and clarify its diverse genetic architecture. I outline below a series of outstanding questions that remain to be addressed in future research, on each of the topics covered. Where possible, I also identify limitations of the approaches I used.

### **5.1 Chapter 2 future directions**

Current work on the adaptive contribution of plant organelle genomes has been disproportionately focused on the plastome. This bias is evident when considering studies that used tests of neutrality on organelle genes (Table 2.2), which have so far only considered loci in the chloroplast genome. There is therefore a need to expand these surveys to the mitochondrial genome. Beyond analyses of DNA polymorphism, substitution crosses and reciprocal transplant

experiments should be used in species for which the plastomes and chondriomes have opposite modes of inheritance, to isolate the fitness effects of mitochondrial DNA variation.

What agents of selection cause adaptive evolution in plant organelle DNA? This topic has attracted surprisingly little explicit interest so far. Most of the information currently available is indirect, and stems from knowledge of the ecology of species under investigation (e.g. Sambatti et al. 2008), or from correlations between putatively selected haplotypes and environmental variables (e.g. Bashalkhanov et al. 2013). Future studies of organelle-encoded fitness effects should aim to test these predictions experimentally, by manipulating biotic and abiotic factors suspected to act as agents of selection. The relationship between agents of selection, the strength of selection, and the likelihood of adaptive evolution in organelle DNA should then also be investigated.

Because of their tight functional integration, it has been hypothesized that coevolution between the nuclear and organelle genomes is common (Burton et al. 2013). Examples available so far appear to be cases where deleterious mutations at organelle loci have resulted in selection for compensatory mutations at nuclear loci, to maintain organelle function (Osada & Akashi 2012; Sloan et al. 2014). It is unknown, however, whether such interactions contribute to local adaptation as well. Provided that dense taxonomic sampling is available for organelle and nuclear loci of interest, future studies could investigate this possibility in a phylogenetic framework, to identify which mutations are causal and which mutations are correlated (e.g. Osada & Akashi 2012).

## 5.2 Chapter 3 future directions

The origin of *H. tuberosus* as inferred *via* genome skimming has since been confirmed based on GBS data using broader taxonomic sampling within the *Helianthus* genus (Baute et al. 2016). In chapter 4, I recovered further support for this result, using more comprehensive sampling for both *H. tuberosus* and the inferred progenitors. While the identity of the parental taxa appears to be resolved, other questions regarding the formation of *H. tuberosus* remain. For example, we do not yet know how many speciation events there were, or which parental lineages were involved. Multiple origins of polyploid taxa are now considered to be the norm (Soltis & Soltis 2009). In *H. tuberosus* as well, the levels of genetic diversity present in organellar genomes (Fig. 3.3) indicate the species is polyphyletic. Because of widespread incomplete lineage sorting (Schilling, 1997; Timme et al. 2007), pinpointing the number and identity of the contributing *H. tuberosus* parental lineages will, most likely, require the use of whole genome sequencing data.

Knowledge of the parental species of *H. tuberosus* allows additional questions to be asked, with relevance to our current understanding of polyploid speciation (reviewed in Soltis et al. 2010). For instance, is there gene flow between *H. tuberosus* and its diploid and tetraploid progenitors? If so, what is the frequency, directionality, and adaptive contribution of these events? The increase in chromosome number for polyploids is expected to provide instantaneous reproductive isolation from the sympatric parental taxa, thereby facilitating the establishment of the new polyploid species (Coyne & Orr ). Even so, we know that gene flow can occur between ploidal levels (Stebbins 1971). Recent studies have provided preliminary genomic evidence that post-polyploidization gene flow can be unidirectional from the progenitor to the polyploid (e.g.



Zohren et al. 2016), as well as bi-directional (Arnold et al. 2015), and that some of the introgressed regions may have been under selection (Arnold et al. 2015).

Other questions concern the dynamics of genome reorganization in *H. tuberosus*. A common observed phenomenon following the formation of polyploid taxa is genome downsizing, whereby the observed DNA content of the polyploid derivative is smaller than the sum of progenitor genome sizes (Leitch & Bennett 2004). This appears to be the case for *H. tuberosus* as well. Specifically, 2C DNA content estimates that I obtained for *H. tuberosus* using flow cytometry (23.4 pg) are smaller than what we would predict (24.2 pg) strictly based on values measured for *H. hirsutus* (15.8 pg) and *H. grosseserratus* (8.4 pg). In the context of polyploid speciation, we would like to know if this DNA loss is random with respect to the constituent subgenomes, or if fragments from one progenitor species are preferentially lost (Soltis et al. 2010). Evidence from *Nicotiana* and *Tragopogon* polyploids point towards non-random DNA loss occurring more frequently from the paternal parent (Renny-Byfield et al. 2012; Soltis et al. 2012). These results mirror those observed with regards to gene silencing in *Gossypium* polyploids, where maternal gene expression dominates (Gong et al. 2012). One explanation for these patterns is that genomic reorganization following polyploid formation will favor the maternal contribution to minimize cytonuclear incompatibilities and increase polyploid stability (Gong et al. 2012). Additional examples corroborating this possibility from a range of polyploid taxa are needed.

### **5.3 Chapter 4 future directions**

The study presented in Chapter 4 advances our understanding of the genetic architecture of invasive potential, and more generally, of contemporary evolution. Nevertheless, several

unknowns remain. First, we currently do not have information on the identity of the genes and mutations that drive the two association signals for tuber number production. Therefore, additional molecular work remains to be done. While this will likely be a challenging task due to the hexaploid genome of *H. tuberosus*, other characteristics of the system, including the availability of a transformation protocol (Kim et al. 2016), could facilitate experimentation. Important questions that could be addressed with this information include the relative contribution of regulatory versus protein-coding changes, and the extent of parallel use of the same gene(s), both between different biological invasion events in *H. tuberosus*, and between *H. tuberosus* and other clonal invasives.

Also, we do not currently know why the invasive *H. tuberosus* phenotype identified here, as well as the alleles associated with this trait, is being maintained at low frequencies in the native range. Previous studies have highlighted that increased performance of invasive individuals compared to native conspecifics might be due to the release from tradeoffs. These may include growth and reproductive output on one hand, and herbivore defense (Blossey & Notzold 1995) or stressful abiotic conditions (Bossdorf et al. 2005; Turner et al. 2014; Hodgins & Rieseberg 2011) on the other. Moreover, recent work has highlighted that the performance of invasive genotypes can be context-dependent (e.g. Molofsky et al. 2017). We could not investigate these possibilities because the logistic requirements for maintaining and harvesting multiple common garden experiments would have been too large. Future studies could investigate the importance of trade-offs to the evolution of invasiveness in *H. tuberosus* by comparing the performance of invasive and non-invasive genotypes under control and stress treatments.

Finally, results presented in Chapter 4 indicate that certain *H. tuberosus* clones are particularly widespread. A promising future line of investigation could therefore be to investigate whether epigenetic modifications facilitate post-establishment spread in this system. This could be accomplished by studying the association between clonal phenotypic variation scored in a common environment and epigenotypes (e.g. Zhang et al. 2016). The contribution of epigenetic modifications to invasion success and rapid adaptation in general is an area for which considerable additional research is needed (Bossdorf et al. 2008).

## References

- Acosta MC, Premoli AC (2010) Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Molecular Phylogenetics and Evolution*, **54**, 235–242.
- Anastasiu P, Negrean G (2009) Neophytes in Romania. In: Rakosy L, Momeu L (eds) Neobiota din Romania. Presa Universitară Clujeană, Cluj-Napoca, pp 66–97.
- Anisimova IN (1982) [Nature of the genomes in polyploid sunflower species]. *Byulleten' Vsesoyuznogo Ordena Lenina i Ordena Druzhby Narodov Instituta Rastenievodstva Imeni N.I. Vavilov* **118**, 27–29.
- Anisimova M, Gascuel O (2006) Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Systematic Biology*, **55**, 539–552.
- Arnold B, Kim ST, Bomblies K (2015) Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Molecular Biology and Evolution*, **32**, 1382–1395.
- Atkin OK, Macherel D (2009) The crucial role of plant mitochondria in orchestrating drought tolerance. *Annals of Botany*, **103**, 581–597.
- Atlagić J, Dozet B, Škorić D (1993) Meiosis and pollen viability in *Helianthus tuberosus* L. and its hybrids with cultivated sunflower. *Plant Breeding* **111**, 318–324.
- Atlagić J, Škorić D (2006) Cytogenetic study of hexaploid species *Helianthus tuberosus* and its F1 and BC1F1 hybrids with cultivated sunflower, *H. annuus*. *Genetika*, **38**, 203–213.
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.

- Awise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA (1979) Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences USA*, **76**, 6694–6698.
- Bajpai PK, Bajpai P (1991) Cultivation and utilization of Jerusalem artichoke for ethanol, single cell protein, and high fructose syrup production. *Enzyme and Microbial Technology*, **13**, 359–362.
- Baker HG, Stebbins GL (1965) *The Genetics of Colonizing Species*. Academic Press, New York, NY, USA.
- Bakker FT, Breman F, Merckx V (2006) DNA sequence evolution in fast evolving mitochondrial DNA *nadl* exons in Geraniaceae and Plantaginaceae. *Taxon*, **55**, 887–896.
- Ballard JWO, Melvin RG (2010) Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology*, **19**, 1523–1539.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.
- Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. *Heredity*, **104**, 419–420.
- Barr CM, Keller SR, Ingvarsson PK, Sloan DB, Taylor DR (2007) Variation in mutation rate and polymorphism among mitochondrial genes in *Silene vulgaris*. *Molecular Biology and Evolution*, **24**, 1783–1791.
- Barney JN, Whitlow TH, DiTommaso A (2009) Evolution of an invasive phenotype: shift to belowground dominance and enhanced competitive ability in the introduced range. *Plant Ecology*, **202**, 275–284.

- Bashalkhanov S, Eckert AJ, Rajora OP (2013) Genetic signatures of natural selection in response to air pollution in red spruce (*Picea rubens*, Pinaceae). *Molecular Ecology*, **22**, 5877–5889.
- Basu C, Halfhill MD, Mueller TC, Stewart CN (2004) Weed genomics: new tools to understand weed biology. *Trends in Plant Science*, **9**, 391–398.
- Baute GJ, Owens GL, Bock DG, Rieseberg LH (2016) Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *American Journal of Botany*, 103, 1-8.
- Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–571.
- Bearhop S, Fiedler W, Furness RW *et al.* (2005) Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science*, **310**, 502–504.
- Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, **20**, 435–440.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Bennett MD, Leitch IJ (2010) Plant DNA C-values database (release 6.0, Dec 2012). <http://data.kew.org/cvalues/>. Accessed on 19 February 2013.
- Betancourt AJ, Welch JJ, Charlesworth B (2009) Reduced effectiveness of selection caused by a lack of recombination. *Current Biology*, **19**, 655–660.
- Birky CW Jr, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513–527.

- Blackman BK, Michaels SD, Rieseberg LH (2011) Connecting the sun to flowering in sunflower adaptation. *Molecular Ecology*, **20**, 3503–3512.
- Blair AC, Wolfe LM (2004) The evolution of an invasive plant: An experimental study with *Silene latifolia*. *Ecology*, **85**, 3035–3042.
- Blossey B, Notzold R (1995) Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. *Journal of Ecology*, **83**, 887–889.
- Bock DG, Kane NC, Ebert DP, Rieseberg LH (2014) Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. *New Phytologist*, **201**, 1021–1030.
- Bock DG, Caseys C, Cousens RD *et al.* (2015) What we still don't know about invasion genetics. *Molecular Ecology*, **24**, 2277–2297.
- Bossdorf O, Auge H, Lafuma L *et al.* (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, **144**, 1–11.
- Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. *Ecology Letters*, **11**, 106–115.
- Brewer MT, Lang L, Fujimura K *et al.* (2006) Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiology*, **141**, 15–25.
- Bromham L, Cowman PF, Lanfear R (2013) Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evolutionary Biology*, **13**, 126.
- Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences USA*, **76**, 1967–1971.

- Brown JS, Eckert CG (2005) Evolutionary increase in sexual and clonal reproductive capacity during biological invasion in an aquatic plant *Butomus umbellatus* (Butomaceae). *American Journal of Botany*, **92**, 495–502.
- Budar F, Roux F (2011) The role of organelle genomes in plant adaptation. *Plant Signaling & Behavior*, **6**, 635–639.
- Buggs RJA, Chamala S, Wei W et al. (2012) Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Current Biology*, **22**, 248–252.
- Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 281–302.
- Campbell DR, Waser NM, Aldridge G, Wu CA (2008) Lifetime fitness in two generations of *Ipomopsis* hybrids. *Evolution*, **62**, 2616–2627.
- Champely S (2012) pwr: basic functions for power analysis. R package version 1.1.1
- Champer J, Buchman A, Akbari OS (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics*, **17**, 146–159.
- Chandler JM (1991) Chromosome evolution in sunflower. In: Tsuchiya P, Gupta PK, eds. *Chromosome engineering in plants: genetics, breeding, evolution, part B*. Amsterdam, the Netherlands: Elsevier, 229–249.
- Chaves MM, Flexas J, Pinheiro C (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany*, **103**, 551–560.
- Cheng Y, Zhou W, Gao C, Lan K, Gao Y, Wu Q (2009) Biodiesel production from Jerusalem artichoke (*Helianthus tuberosus* L.) tuber by heterotrophic microalgae *Chlorella protothecoides*. *Journal of Chemical Technology and Biotechnology*, **84**, 777–781.



- Chiapusio G, Sánchez AM, Reigosa MJ, González L, Pellisier F (1997) Do germination indices adequately reflect allelochemical effects on the germination process? *Journal of Chemical Ecology*, **23**, 2445–2453.
- Cho Y, Mower JP, Qiu YL, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences USA*, **101**, 17741–17746.
- Christin PA, Salamin N, Muasya AM *et al.* (2008) Evolutionary switch and genetic convergence on *rbcL* following the evolution of C4 photosynthesis. *Molecular Biology and Evolution*, **25**, 2361–2368.
- Chung S-M, Gordon VS, Staub JE (2007) Sequencing cucumber (*Cucumis sativus* L.) chloroplast genomes identifies differences between chilling-tolerant and -susceptible cucumber lines. *Genome*, **50**, 215–225.
- Colautti RI, Maron JL, Barrett SCH (2009) Common garden comparisons of native and introduced plant populations: latitudinal clines can obscure evolutionary inferences. *Evolutionary Applications*, **2**, 187–199.
- Colautti RI, Barrett SC (2013) Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, **342**, 364–366.
- Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, **24**, 1999–2017.
- Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution*, **55**, 1943–1962.
- Conover DO, Munch SB (2002) Sustaining fisheries yields over evolutionary time scales. *Science*, **297**, 94–96.

- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.
- Coyne J, Orr H (2004) Speciation. Sinauer associates Sunderland, Sunderland, MA.
- Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, **14**, 262–274.
- Cornille A, Gladieux P, Smulders MJM et al. (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics*, **8**, e1002703.
- Cornille A, Salcedo A, Kryvokhyzha D et al. (2016) Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Molecular Ecology*, **25**, 616–629.
- Darlington CD (1956) Chromosome botany. London, UK: Allen and Unwin.
- Del Re AC (2013) compute.es: Compute Effect Sizes. R package version 0.2-2.
- DePristo MA, Banks E, Poplin R et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.
- Dlugosch KM, Parker IM (2008) Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecology Letters*, **11**, 701–709.
- Doležel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, **2**, 2233–2244.
- Dowling DK, Friberg U, Lindell J (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology & Evolution*, **23**, 546–554.

- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.
- Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution*, **49**, 827–831.
- Dumolin-Lapegue S, Demesure B, Fineschi S, Come VL, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.
- Earl, Dent A. and vonHoldt, Bridgett M (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
- Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences, USA*, **107**, 16196–16200.
- Erixon P, Oxelman B (2008) Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS One*, **3**, e1386.
- 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.

- Fehér, A. 2007. Historical reconstruction of expansion of non-native plants in the Nitra river basin (SW Slovakia). *Kanitzia*, **15**, 47–62.
- Filep, R, Balogh L, Csergő AM (2010) Perennial *Helianthus* taxa in Târgu-Mureş city and its surroundings. *Journal of Plant Development*, **17**, 69–74.
- Filep R, Pal RW, Balázs VL et al. (2016) Can seasonal dynamics of allelochemicals play a role in plant invasions? A case study with *Helianthus tuberosus* L.. *Plant Ecology*, **217**, 1489–1501.
- Flory SL, Long F, Clay K (2011) Invasive *Microstegium* populations consistently outperform native range populations across diverse environments. *Ecology*, **92**, 2248–2257.
- da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, **9**, 119.
- Francis RM (2017) pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, **17**, 27–32.
- Fuentes I, Stegemann S, Golczyk H, Karcher D, Bock R (2014) Horizontal genome transfer as an asexual path to the formation of new species. *Nature*, **511**, 232–235.
- Fry JD (2003) Multilocus models of sympatric speciation: Bush versus Rice versus Felsenstein. *Evolution*, **57**, 1735–1746.
- Galmes J, Andralojc PJ, Kapralov MV et al. (2014) Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus *Limonium* (Plumbaginaceae). *New Phytologist*, **203**, 989–999.
- Galtier N, Nabholz B, Glemin S, Hurst GD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907.
- Gentzbittel L, Perrault A, Nicolas P (1992) Molecular phylogeny of the *Helianthus* genus, based on nuclear restriction fragment-length-polymorphism (RFLP). *Molecular Biology and Evolution*, **9**, 872–892.
- Gong L, Salmon A, Yoo M-J et al. (2012) The cytonuclear dimension of allopolyploid evolution: an example from cotton using Rubisco. *Molecular Biology and Evolution*, **29**, 3023–3036.
- Gordon VS, Staub JE (2011) Comparative analysis of chilling response in cucumber through plastidic and nuclear genetic effects component analysis. *Journal of the American Society for Horticultural Science*, **136**, 256–264.
- Grabherr MG, Haas BJ, Yassour M et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.
- Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, **296**, 707–711.
- Graves JL, Hertweck KL, Phillips MA et al. (2017) Genomics of Parallel Experimental Evolution in *Drosophila*. *Molecular Biology and Evolution*, doi: 10.1093/molbev/msw282.
- Greiner S (2012) Plastome mutants of higher plants. In: Genomics of Chloroplasts and Mitochondria (eds Bock R, Knoop V), pp. 237–266. Springer, Dordrecht, Heidelberg, New York, London, the Netherlands.
- Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organelle genomes in plants. *BioEssays*, **35**, 354–365.

- Guggisberg A, Bretagnolle F, Mansion G (2006) Allopolyploid origin of the Mediterranean endemic, *Centaureum bianoris* (Gentianaceae), inferred by molecular markers. *Systematic Botany*, **31**, 368–379.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences USA*, **105**, 18424–18429.
- Guesewell S, Jakobs G, Weber E (2006) Native and introduced populations of *Solidago gigantea* differ in shoot production but not in leaf traits or litter decomposition. *Functional Ecology*, **20**, 575–584.
- Haag-Liautard C, Coffey N, Houle D *et al.* (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biology*, **6**, e204.
- Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.
- Hahn MA, Rieseberg LH (2017) Genetic admixture and heterosis may enhance the invasiveness of common ragweed. *Evolutionary Applications*, **10**, 241–250.
- Hansen AK, Escobar LK, Gilbert LE, Jansen RK (2007) Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): implications for phylogenetic studies. *American Journal of Botany*, **94**, 42–46.
- Hao DC, Chen SL, Xiao PG (2010) Molecular evolution and positive Darwinian selection of the chloroplast maturase *matK*. *Journal of Plant Research*, **123**, 241–247.

- Heiser CB, Smith DM, Clevenger S, Martin WC (1969) The North American sunflowers. *Memoirs of the Torrey Botanical Club*, **22**, 1–218.
- Heiser CB (1976) The sunflower. Norman, OK, USA: University of Oklahoma Press.
- Heiser CB, Smith DM (1964) Species crosses in *Helianthus*: II. Polyploid species. *Rhodora*, **66**, 344–358.
- Hendry AP, Nosil P, Rieseberg LH (2007) The speed of ecological speciation. *Functional Ecology*, **21**, 455–464.
- Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genetical Research*, **8**, 269–294.
- Hill JH, Chen Z, Xu H (2014) Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. *Nature Genetics*, **46**, 389–392.
- Hodgins KA, Rieseberg LH (2011) Genetic differentiation in life-history traits of introduced and native common ragweed (*Ambrosia artemisiifolia*) populations. *Journal of Evolutionary Biology*, **24**, 2731–2749.
- Hodgins KA, Lai Z, Nurkowski K, Huang J, Rieseberg LH (2013) The molecular basis of invasiveness: differences in gene expression of native and introduced common ragweed (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology*, **22**, 2496–2510.
- Hodgins KA, Bock DG, Hahn MA et al. (2015) Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Molecular Ecology*, **24**, 2226–2240.

- Hollingsworth ML, Bailey JP (2000) Evidence for massive clonal growth in the invasive weed *Fallopia japonica* (Japanese Knotweed). *Botanical Journal of the Linnean Society*, **133**, 463–472.
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70.
- Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI (2014) Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proceedings of the National Academy of Sciences USA*, **111**, 7713–7718.
- Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science*, **287**, 308–309.
- Husband BC, Sabara HA (2003) Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist*, **161**, 703–713.
- Iida S, Miyagi A, Aoki S *et al.* (2009) Molecular adaptation of *rbcL* in the heterophyllous aquatic plant *Potamogeton*. *PLoS One*, **4**, e4633.
- Irwin DE (2012) Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *The American Naturalist*, **180**, 35–49.
- James JK, Abbott RJ (2005) Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution*, **59**, 2533–2546.
- Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. *Genetics*, **171**, 1951–1962.



- Jombart T, Ahmed I (2011) adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
- Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.
- Joshi J, Vrieling K (2005) The enemy release and EICA hypothesis revisited: incorporating the fundamental difference between specialist and generalist herbivores. *Ecology Letters*, **8**, 704–714.
- Kawakami T, Morgan TJ, Nippert JB *et al.* (2011) Natural selection drives clinal life history patterns in the perennial sun- flower species, *Helianthus maximiliani*. *Molecular Ecology*, **20**, 2318–2328.
- Kane NC, Cronk Q (2008) Botany without borders, barcoding in focus. *Molecular Ecology*, **17**, 5175–5176.
- Kane NC, Sveinsson S, Dempewolf H *et al.* (2012) Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, **99**, 320–329.
- Kapralov MV, Filatov DA (2006) Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. *PLoS One*, **1**, e8.
- Kapralov MV, Filatov DA (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evolutionary Biology*, **7**, 73.
- Kapralov MV, Kubien DS, Andersson I, Filatov DA (2011) Changes in Rubisco kinetics during the evolution of C4 photosynthesis in Flaveria (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Molecular Biology and Evolution*, **28**, 1491–1503.

- Kapralov MV, Smith JAC, Filatov DA (2012) Rubisco evolution in C4 eudicots: an analysis of *Amaranthaceae sensu lato*. *PLoS One*, **7**, e52974.
- Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.
- Kays SJ, Nottingham SF (2008) Biology and chemistry of the Jerusalem Artichoke: *Helianthus tuberosus* L. Boca Raton, FL, USA: CRC Press.
- Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal*, **46**, 85–94.
- Kliber A, Eckert CG (2005) Interaction between founder effect and selection during biological invasion in an aquatic plant. *Evolution*, **59**, 1900–1913.
- Kim, M-J, An D-J, Moon K-B et al. (2016) Highly efficient plant regeneration and *Agrobacterium*-mediated transformation of *Helianthus tuberosus* L.. *Industrial Crops and Products*, **83**, 670-679.
- Kim S-T, Sultan SE, Donoghue MJ (2008) Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences, USA* **105**, 12370–12375.
- Kimball S, Campbell DR, Lessin C (2008) Differential performance of reciprocal hybrids in multiple environments. *Journal of Ecology*, **96**, 1306–1318.
- Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Kleessen B, Schwarz S, Boehm A, Fuhrmann H, Richter A, Henle T, Krueger M (2007) Jerusalem artichoke and chicory inulin in bakery products affect faecal microbiota of healthy volunteers. *British Journal of Nutrition*, **98**, 540–549.

- Kolbe JJ, Glor RE, Schettino LRG et al. (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.
- Kosaric N, Cosentino GP, Wieczorek A, Duvnjak Z (1984) The Jerusalem Artichoke as an agricultural crop. *Biomass*, **5**, 1–36.
- Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary Fisherian life-history evolution in small salmonid populations. *Nature*, **419**, 826–830.
- Kostoff D (1934) A contribution to the meiosis of *Helianthus tuberosus* L. *Zeitschr für Pflanzenzüchtung* **19**, 423–438.
- Kostoff D (1939) Autosyndesis and structural hybridity in F1-hybrid *Helianthus tuberosus* L. × *Helianthus annuus* L. and their sequences. *Genetica*, **21**, 285–300.
- Konvalinková P (2003) Generative and vegetative reproduction of *Helianthus tuberosus*, an invasive plant in Central Europe. *Plant Invasions: Ecological Threats and Management Solutions*. Backhuys, Leiden, pp. 289-299.
- Krause K (2012) Plastid genomes of parasitic plants: a trail of reductions and losses. In: *Organelle Genetics* (ed. Bullerwell C). Springer, Berlin.
- Krieger MJB, Ross KG (2002) Identification of a major gene regulating complex social behavior. *Science*, **295**, 328–332.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.
- Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences, USA*, **104**, 3883–3888.

- Lee CE (1999) Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution*, **53**, 1423–1434.
- Lee TH, Guo H, Wang X, Kim CH, Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, **15**,162.
- Leger EA, Rice KJ (2003) Invasive California poppies (*Eschscholzia californica* Cham.) grow larger than native individuals under reduced competition. *Ecology Letters*, **6**, 257–264.
- Leinonen PH, Remington DL, Savolainen O (2011) Local adaptation, phenotypic differentiation and hybrid fitness in diverged natural populations of *Arabidopsis lyrata*. *Evolution*, **65**, 90–107.
- Leinonen P, Remington DL, Leppala J, Savolainen O (2013) Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*. *Molecular Ecology*, **22**, 709–723.
- Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society*, **82**, 651 – 663.
- Lenth RV, Hervé M (2015) Package lsmeans: R package versions 2.19. Available: <http://cran.r-project.org/package=lsmeans>.
- Lescak EA, Bassham SL, Catchen J *et al.* (2015) Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences of the USA*, **112**, 201512020.
- Levy SF, Blundell JR, Venkataram S *et al.* (2015) Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, **519**, 181–186.
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997.

- Lin Z, Kong H, Nei M, Ma H (2006) Origins and evolution of the *recA/RAD51* gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences USA*, **103**, 10328–10333.
- Lippman ZB, Zamir D (2007) Heterosis: Revisiting the magic. *Trends in Genetics*, **23**, 60–66.
- Llopart A, Herrig D, Brud E, Stecklein Z (2014) Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Molecular Ecology*, **23**, 1124–1236.
- Lockwood BL, Somero GN (2011) Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Molecular Ecology*, **20**, 517–529.
- Long RW (1955) Hybridization between the perennial sunflowers *Helianthus salicifolius* A. Dietr. and *H. grosseserratus* Martens. *American Midland Naturalist*, **54**, 61–64.
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.
- Maheshwari S, Barbash DA (2011) The genetics of hybrid incompatibilities. *Annual Review of Genetics*, **45**, 331–355.
- Malinska H, Tate JA, Matyasek R et al. (2010) Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon* (Asteraceae) allotetraploids. *BMC Evolutionary Biology*, **10**, 291.
- Malinska H, Tate JA, Mavrodiev E et al. (2011) Ribosomal RNA genes evolution in *Tragopogon*: a story of New and Old World allotetraploids and synthetic lines. *Taxon*, **60**, 348–354.
- Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, **186**, 299–317.

- Mariac C, Scarcelli N, Pouzadou J *et al.* (2014) Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Molecular Ecology Resources*, **14**, 1109–1113.
- Martin M (2012) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Bioinformatics in Action*, **17**, 10.
- Matvienko M, Kozik A, Froenicke L *et al.* (2013) Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS ONE*, **8**, e55913.
- McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist*, **200**, 966–977.
- McKenna A, Hanna M, Banks E *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
- McKinnon GE, Vaillancourt RE, Steane DA, Potts BM (2004) The rare silver gum, *Eucalyptus cordata*, is leaving its trace in the organellar gene pool of *Eucalyptus globulus*. *Molecular Ecology*, **13**, 3751–3762.
- Messer PW, Ellner SP, Hairston NG (2016) Can Population Genetics Adapt to Rapid Evolution? *Trends in Genetics*, **32**, 408–418.
- Milne I, Bayer M, Cardle L *et al.* (2010) Tablet-next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.
- Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, **30**, 1188–1195.

- Molofsky J, Collins AR, Imbert E, Bitinas Tadas, Lavergne S (2017) Are Invasive Genotypes Superior? An Experimental Approach Using Native and Invasive Genotypes of the Invasive Grass *Phalaris Arundinacea*. *Open Journal of Ecology*, **7**, 125-139.
- Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology*, **7**, 135.
- Mueller JC, Edelaar P, Carrete M *et al.* (2014) Behaviour-related DRD4 polymorphisms in invasive bird populations. *Molecular Ecology*, **23**, 2876–2885.
- Muir G, Filatov D (2007) A selective sweep in the chloroplast DNA of dioecious *Silene* (section *Elisanthe*). *Genetics*, **177**, 1239–1247.
- Muth NZ, Pigliucci M (2006) Traits of invasives reconsidered: phenotypic comparisons of introduced invasive and introduced noninvasive plant species within two closely related clades. *American Journal of Botany*, **93**, 188–196.
- Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. *Genetica*, **102–103**, 61–69.
- Neale DB, Wheeler NC, Allard RW (1986) Paternal inheritance of chloroplast DNA in Douglas-fir. *Canadian Journal of Forest Research*, **16**, 1152–1154.
- Neiva J, Pearson G, Valero M, Serrão E (2010) Surfing the wave on a borrowed board: range expansion and spread of introgressed organellar genomes in the seaweed *Fucus ceranoides* L. *Molecular Ecology*, **19**, 4812–4822.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

- Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641–647.
- Osada N, Akashi H (2012) Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Molecular Biology and Evolution*, **29**, 337–346.
- Palme A, Su Q, Palsson S, Lascoux M (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology*, **13**, 167–178.
- Parkinson CL, Mower JP, Qiu YL *et al.* (2005) Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evolutionary Biology*, **5**, 73.
- Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology*, **7**, 84.
- Pascoal S, Cezard T, Eik-Nes A *et al.* (2014) Rapid convergent evolution in wild crickets. *Current Biology*, **24**, 1369–1374.
- Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995) The weediness of wild plants—molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum Halepense* (L) Pers. *Proceedings of the National Academy of Sciences, USA*, **92**, 6127–6131.
- Pelletier F, Garant D, Hendry AP (2009) Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society of London Series B*, **364**, 1483–1489.



- Percy DM, Argus GW, Cronk QC *et al.* (2014) Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? *Molecular Ecology*, **23**, 4737–4756.
- Perkins TA, Phillips BL, Baskett ML, Hastings A (2013) Evolution of dispersal and life history interact to drive accelerating spread of an invasive species. *Ecology Letters*, **16**, 1079–1087.
- Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducouso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA*, **94**, 9996–10001.
- Petit RJ, Csaikl UM, Bordacs S *et al.* (2002a) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5–26.
- Petit RJ, Latouche-Hallé C, Pemonge M-H, Kremer A (2002b) Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. *Forest Ecology and Management*, **156**, 115–130.
- Petit RJ, Bodénès C, Ducouso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.
- Petit RJ, Duminil J, Fineschi S *et al.* (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.
- Pinheiro J, Bates D, DebRoy S, Sarkar D (2015) R Development Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–109.
- Pohlert T (2015) Package ‘PMCMR’. <http://cran.r-project.org/web/packages/PMCMR/PMCMR.pdf>

- Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.
- Porter SD, Savignano DA (1990) Invasion of polygyne fire ants decimates native ants and disrupts arthropod community. *Ecology*, **71**, 2095–2106.
- Powles SB, Yu Q (2010) Evolution in action: plants resistant to herbicides. *Annual Review of Plant Biology*, **61**, 317–347.
- Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.
- Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Puzey J, Vallejo-Marín M (2014) Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Molecular Ecology*, **23**, 4472–4485.
- Pyšek P, Richardson DM (2007). Traits associated with invasiveness in alien plants: where do we stand? In: Biological Invasions (ed. Nentwig, W.). Springer, New York, pp. 97–125.
- Qi X, Liu Y, Vigueira CC et al. (2015) More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. *Molecular Ecology*, **24**, 3329–3344.
- Rambaut A (2009) FigTree. Available from <http://tree.bio.ed.ac.uk/software/figtree/>
- Rand DM, Kann LM (1998) Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. *Genetica*, **102–103**, 393–407.
- Razanajatovo M, Maurel N, Dawson W et al. (2016). Plants capable of selfing are more likely to become naturalized. *Nature Communications*, **7**, 13313.

- Renaut S, Owens GL, Rieseberg LH (2014) Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*, **23**, 311–324.
- Renny-Byfield S, Kovarik A, Chester M et al. (2012) Independent, rapid and targeted loss of a highly repetitive DNA sequence derived from the paternal genome donor in natural and synthetic *Nicotiana tabacum*. *PLoS ONE*, **7**, e36963.
- Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5**, 65–84.
- Rieseberg LH, Raymond O, Rosenthal DM et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.
- Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, **29**, 233–242.
- Roberfroid MB (2007) Inulin-type fructans: functional food. *Journal of Nutrition*, **137**, 2493S–2502S.
- Rodríguez-Verdugo A, Buckley J, Stapley J (2017) The genomic basis of eco-evolutionary dynamics. *Molecular Ecology*. doi:10.1111/mec.14045.
- Rogers C, Thompson T, Seiler GJ (1982) Sunflower species of the United States. Bismark, ND, USA: National Sunflower Association.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. *The Plant Genome*. doi: 10.3835/plantgenome2015.08.0073.

- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223–226.
- Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2008) Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters*, **11**, 1082–1091.
- Sánchez R, Serra F, Tárraga J et al. (2011). Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research*, **39**, 1–5.
- Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.
- Schilling EE (1997) Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction site data. *Theoretical and Applied Genetics*, **94**, 925–933.
- Schilling EE, Linder CR, Noyes R, Rieseberg LH (1998) Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. *Systematic Botany*, **23**, 177–187.
- Schlötterer C, Kofler R, Versace E, Tobler R, Franssen S (2014) Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, **114**, 431–440.
- Schmitz-Linneweber C, Kushnir S, Babiychuk E et al. (2005) Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. *Plant Cell*, **17**, 1815–1828.
- Schoener TW (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science*, **331**, 426–429.

- Schwarzbach AE, Rieseberg LH (2002) Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology*, **11**, 1703–1715.
- Scowcroft WR (1979) Nucleotide polymorphism in chloroplast DNA of *Nicotiana debneyi*. *Theoretical and Applied Genetics*, **55**, 133–137.
- Sen L, Fares M, Liang B *et al.* (2011) Molecular evolution of *rbcL* in three gymnosperm families: identifying adaptive and coevolutionary patterns. *Biology Direct*, **6**, 29.
- Serieys H, Souyris I, Gil A, Poinso B, Bervillé A (2010). Diversity of Jerusalem artichoke clones (*Helianthus tuberosus* L.) from the INRA-Montpellier collection. *Genetic Resources and Crop Evolution*, **57**, 1207–1215.
- Sloan DB, Taylor DR (2012) Evolutionary rate variation in organelle genomes: the role of mutational processes. In: *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*(ed. Bullerwell CE), pp. 123–146. Springer-Verlag, Berlin.
- Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biology and Evolution*, **4**, 294–306.
- Sloan DB, Triant DA, Forrester NJ *et al.* (2014a) A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Molecular Phylogenetics and Evolution*, **72**, 82–89.
- Sloan DB, Triant DA, Wu M, Taylor DR (2014b) Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Molecular Biology and Evolution*, **31**, 673–682.
- Slobodkin LB (1961) *Growth and Regulation of Animal Populations*. Holt, Rinehart and Winston, New York, NY, USA.

- Slotte T, Ceplitis A, Neuffer B, Hurka H, Lascoux M (2006) Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *American Journal of Botany*, **93**, 1714–1724.
- Soltis DE, Soltis PS (1989). Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. *American Journal of Botany*, **76**, 1119–1124.
- Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution*, **14**, 348–352.
- Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annual Review of Plant Biology*, **60**, 561–588.
- Soltis DE, Buggs RJA, Doyle JJ, Soltis PS (2010) What we still don't know about polyploidy. *Taxon*, **59**, 1387–1403.
- Soltis DE, Buggs RJA, Barbazuk B et al. (2012) The early stages of polyploidy: Rapid and repeated evolution in *Tragopogon*. In PS Soltis and DE Soltis [eds.], *Polyploidy and genome evolution*, 271–292. Springer, New York, New York, USA.
- Smith JE (1807) *An introduction to physiological and systematic botany*. London, UK: Longman, Hurst, Reese, Orme, Paternoster Row, White.
- Stebbins GL (1971) *Chromosomal evolution in higher plants*. London: Addison-Wesley.
- Stegemann S, Keuthe M, Greiner S, Bock R (2012) Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of Sciences USA*, **109**, 2434–2438.
- Stewart JB, Freyer C, Elson JL *et al.* (2008) Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biology*, **6**, e10.

- Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**, 94–101.
- Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next generation sequencing for plant systematics. *American Journal of Botany*, **99**, 349–364.
- Stuart YE, Campbell TS, Hohenlohe PA, Revell LJ, Losos JB (2014) Rapid evolution of a native species following invasion by a congener. *Science*, **346**, 463–466.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Tamura K, Peterson D, Peterson N et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Tesio F, Weston LA, Vidotto F, Ferrero A (2010) Potential allopathic effects of Jerusalem artichoke (*Helianthus tuberosus*) leaf tissues. *Weed Technology*, **24**, 378–385.
- Timme RE, Simpson BB, Linder CR (2007) High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18S-26S ribosomal DNA external transcribed spacer. *American Journal of Botany*, **94**, 1837–1852.
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, **5**, 123–135.
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

- Toews DPL, Mandic M, Richards JG, Irwin DE (2013) Migration, mitochondria and the yellow-rumped warbler. *Evolution*, **68**, 241–255.
- Tsitroni A, Kirkpatrick M, Levin DA (2003) A model for chloroplast capture. *Evolution*, **57**, 1776–1782.
- Turgeon J, Tayeh A, Facon B et al. (2011). Experimental evidence for the phenotypic impact of admixture between wild and biocontrol Asian ladybird (*Harmonia axyridis*) involved in the European invasion. *Journal of Evolutionary Biology*, **24**, 1044–1052.
- Turner JR (1977) Butterfly mimicry: the genetical evolution of an adaptation. *Evolutionary Biology*, **10**, 163–206.
- Turner SD (2014) qqman: an R package for visualizing GWAS results using QQ and Manhattan plots. Preprint at bioRxiv <http://dx.doi.org/10.1101/005165>.
- Turner KG, Hufbauer RA, Rieseberg LH (2014) Rapid evolution of an invasive weed. *New Phytologist*, **202**, 309–321.
- Vandepitte K, De Meyer T, Helsen K et al. (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular Ecology*, **23**, 2157–2164.
- Vidotto F, Tesio F, Ferrero A (2008) Allelopathic effects of *Helianthus tuberosus* L. on germination and seedling growth of several crops and weeds. *Biological Agriculture and Horticulture*, **26**, 55–68.
- Wagner CE, Keller I, Wittwer S et al. (2013) Genome-wide RAD sequence data provides unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
- Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annual Review of Genetics*, **39**, 359–407.



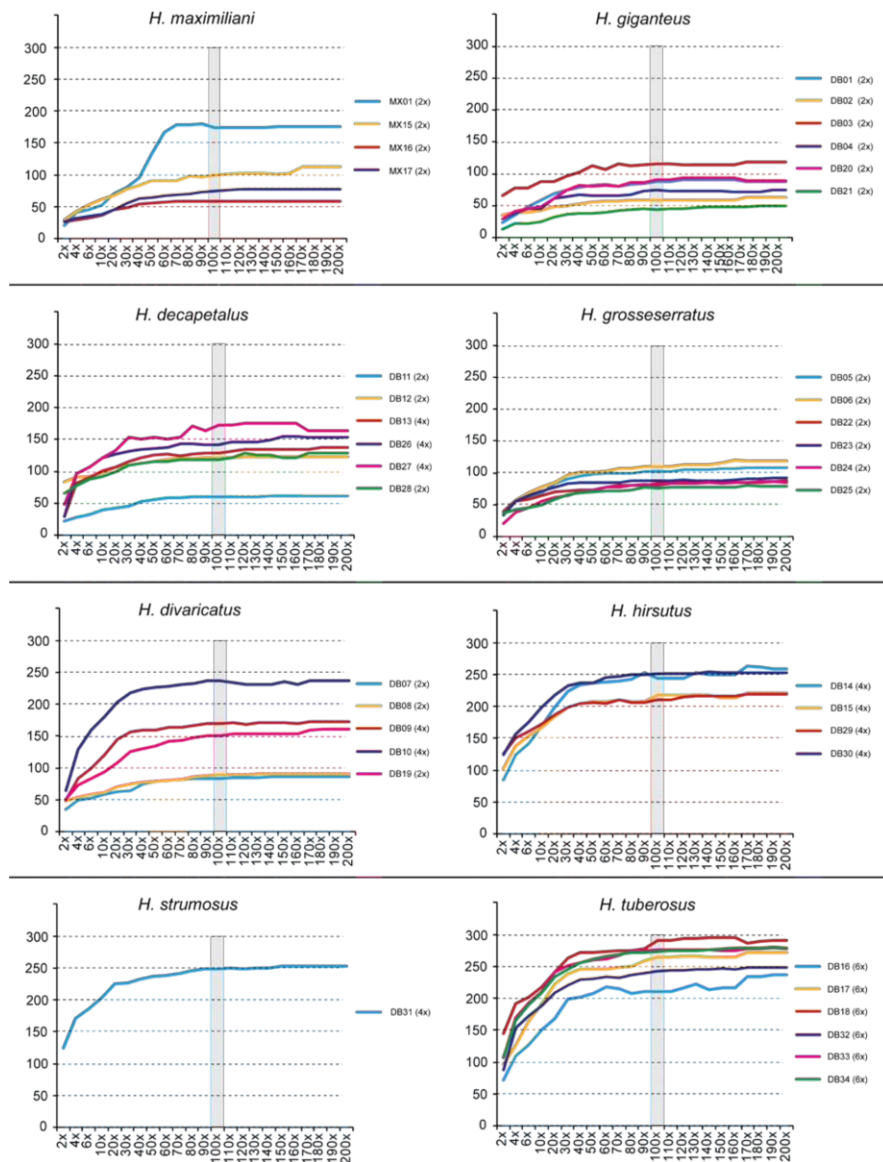
- Weber E, Schmid B (1998) Latitudinal population differentiation in two species of *Solidago* (Asteraceae) introduced into Europe. *American Journal of Botany*, **85**, 1110–1121.
- Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA*, **92**, 280–284.
- Weng ML, Ruhlman TA, Gibby M, Jansen RK (2012) Phylogeny, rate variation, and genome size evolution in *Pelargonium* (Geraniaceae). *Molecular Phylogenetics and Evolution*, **64**, 654–670.
- Weng ML, Blazier JC, Govindu M, Jansen RK (2014) Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution*, **31**, 645–659.
- Whitney KD, Broman KW, Kane NC *et al.* (2015) QTL mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Molecular Ecology* **24**, 2194–2211.
- Williams C, Moore R (1989) Phenotypic adaptation and natural selection in the wild rabbit, *Oryctolagus cuniculus*, in Australia. *Journal of Animal Ecology* **58**, 495–507.
- Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitutions vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA*, **84**, 9054–9058.
- Wright SI, Nano N, Foxe JP, Dar VN (2008) Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*. *Genetics Research*, **90**, 119–128.
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.

- Yeaman S, Hodgins KA, Lotterhos KE *et al.* (2016) Convergent local adaptation to climate in distantly related conifers. *Science* **353**, 1431–1433.
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.
- Zhang Yuan-Ye, Parepa Madalin, Fischer Markus, Bossdorf Oliver (2016) Epigenetics of colonizing species? A study on Japanese knotweed in Central Europe. S.C.H. Barrett, R.I. Colautti, K.M. Dlugosch, L.H. Rieseberg (Eds.), *Invasion Genetics: The Baker and Stebbins Legacy*, Wiley-Blackwell, Oxford (2016), pp. 328–340.
- Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.
- Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2009) Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. *PLoS One*, **4**, e5297.
- Zhu A, Guo W, Jain K, Mower JP (2014) Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Molecular Biology and Evolution*, **31**, 1228–1236.
- Zohren J, Wang N, Kardailsky I, et al. (2016) Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by RAD markers. *Molecular Ecology*, **25**, 2413–2426.

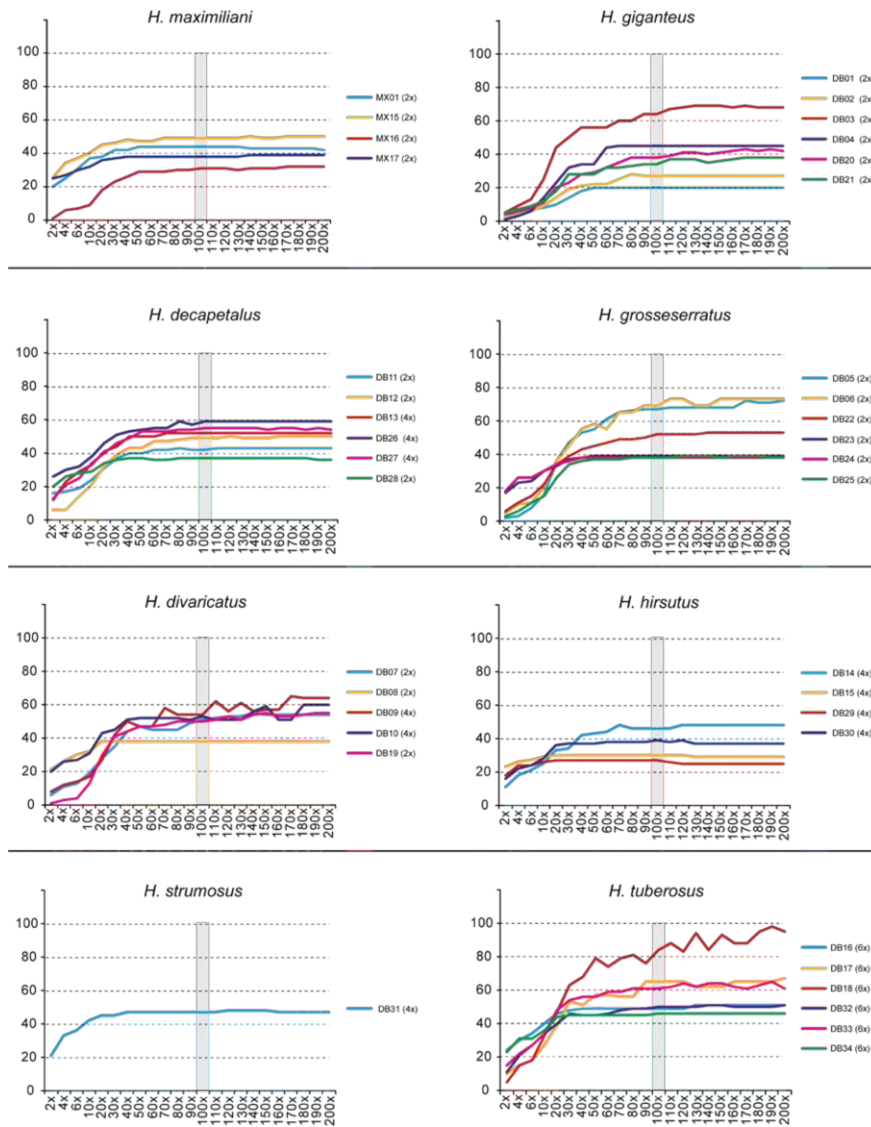
# Appendices

## Appendix A - Chapter 3 Supplementary Material

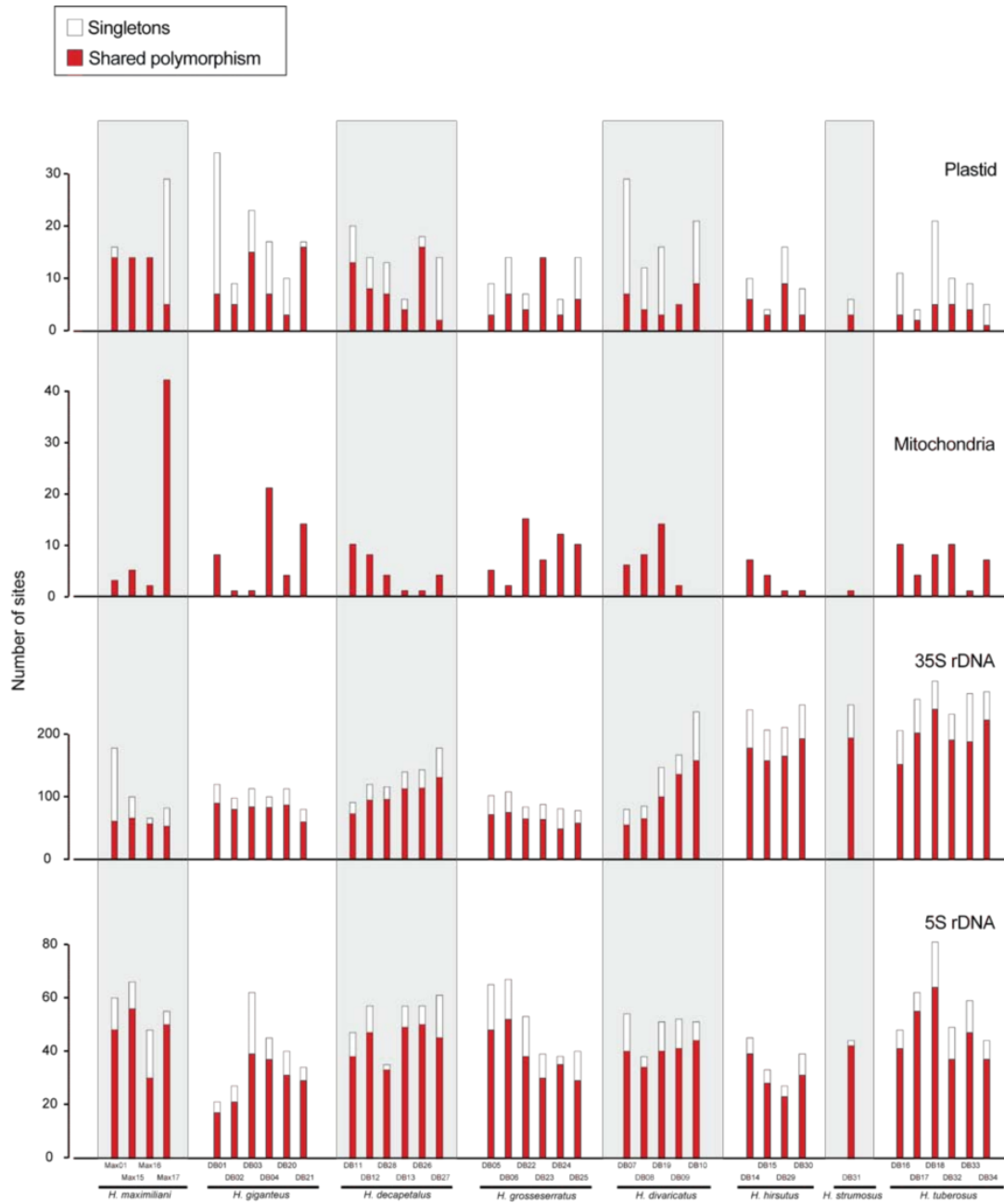
**Figure A.1:** Number of SNPs called (y axis) vs. ploidy setting used in Unified Genotyper (x axis) for the 35S segment in all species and accessions. For the final analysis we set ploidy to 100x.



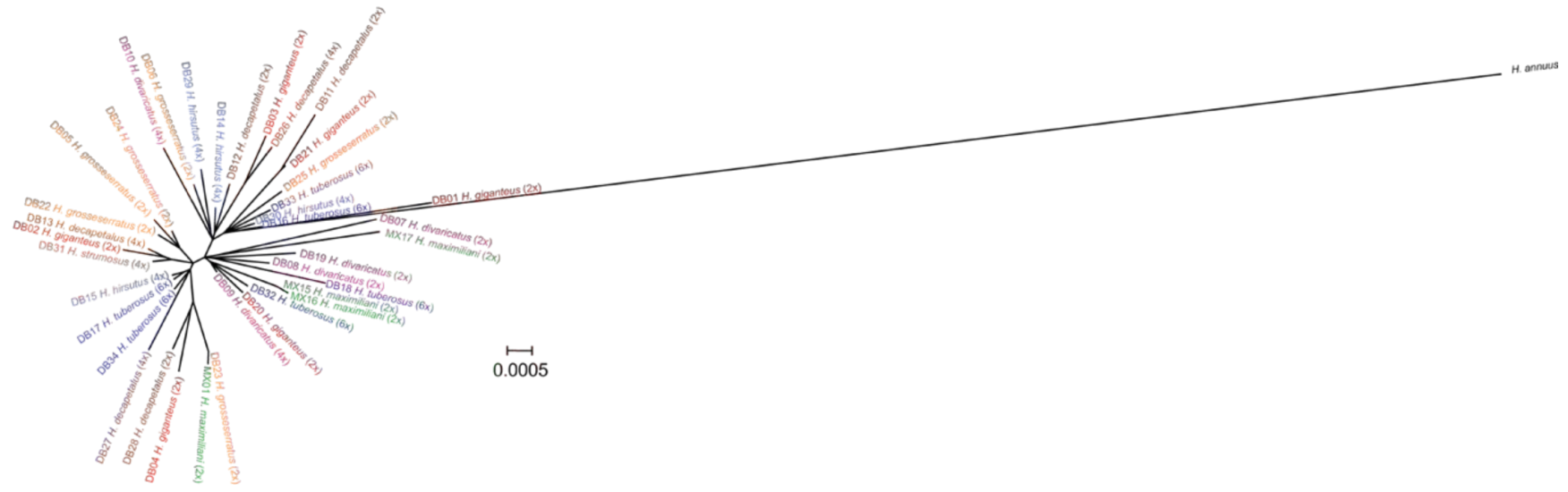
**Figure A.2:** Number of SNPs called (y axis) vs. ploidy setting used in Unified Genotyper (x axis) for the 5S segment (consisting of 5S and its associated NTS region) in all species and accessions. For the final analysis we set ploidy to 100x.



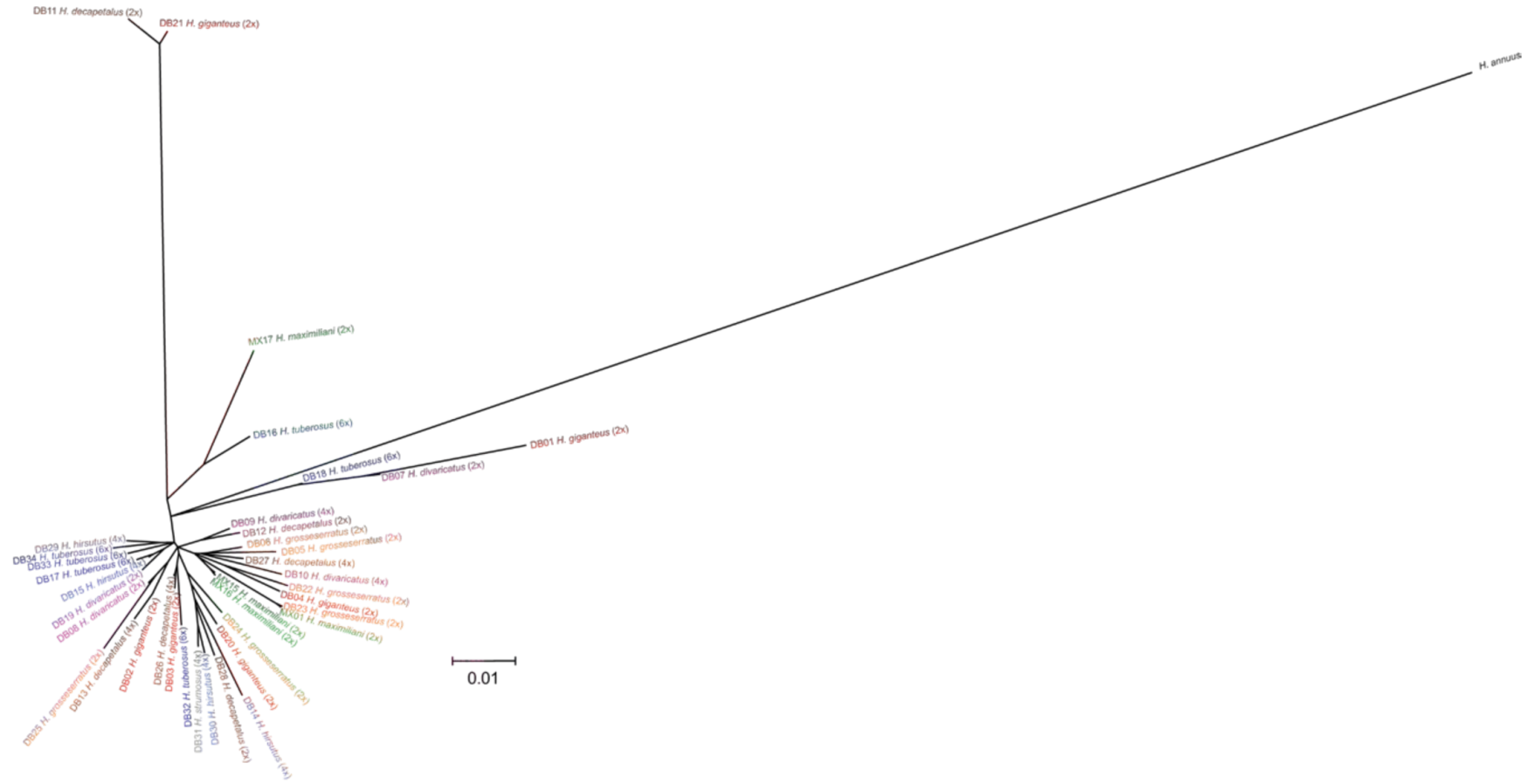
**Figure A.3:** Number of singletons (white segments) and polymorphism shared between two or more samples (red segments) for each surveyed region in all species and accessions. Counts do not include sites with missing data or insertion-deletion polymorphisms.



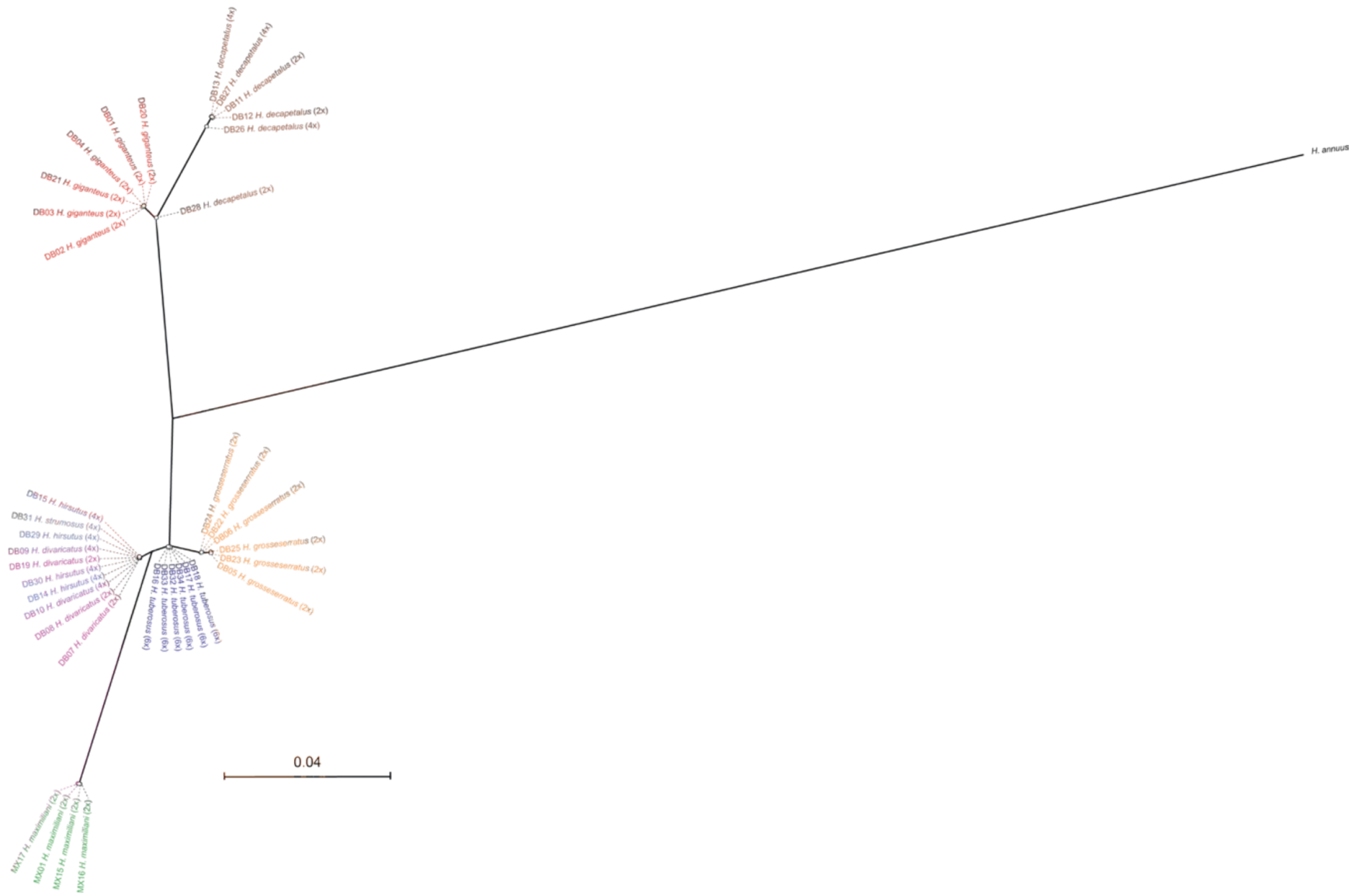
**Figure A.4:** Unrooted maximum likelihood phylogenetic reconstruction of complete plastid genomes (151,551 bp).



**Figure A.5:** Unrooted maximum likelihood phylogenetic reconstruction of partial mitochondrial genomes (196,853 bp).

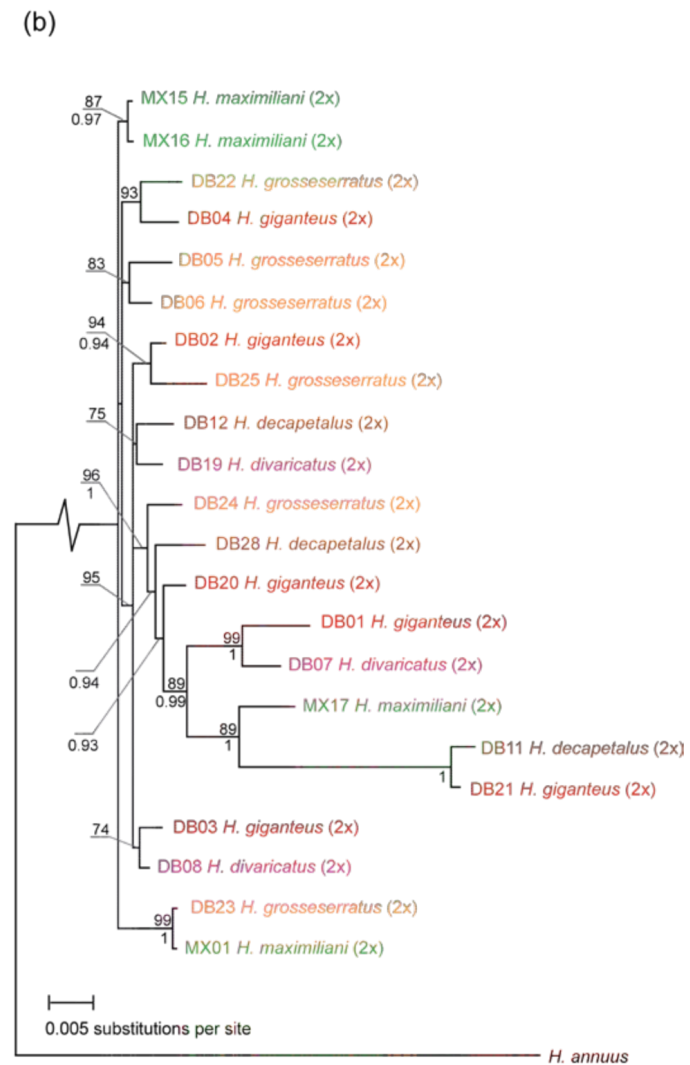
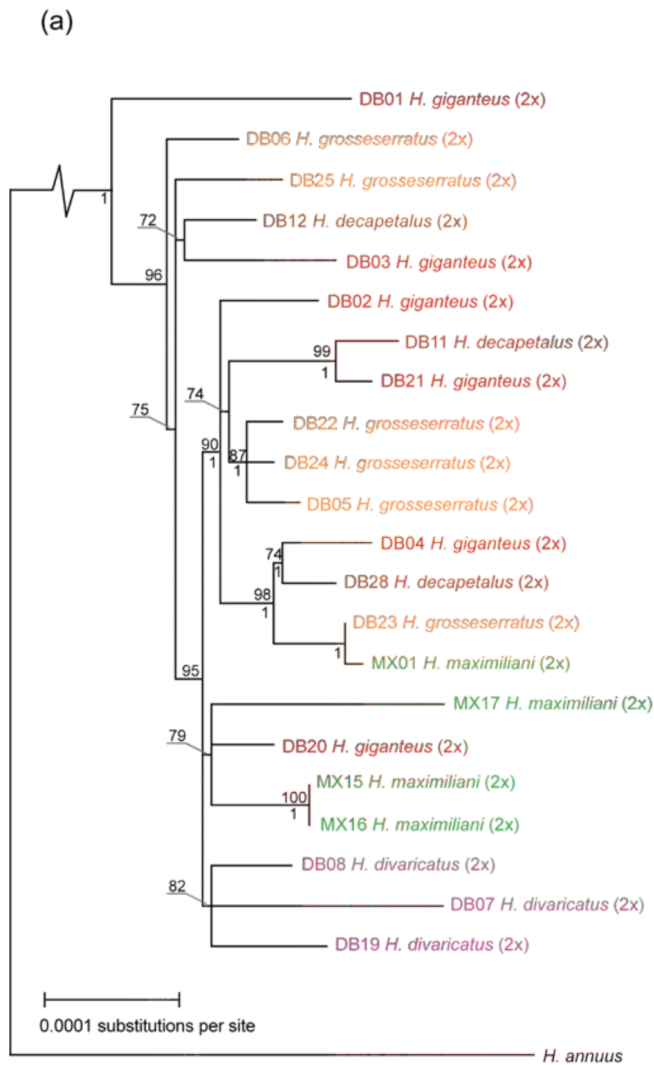


**Figure A.6:** Unrooted maximum likelihood phylogenetic reconstruction of concatenated rDNA sequences (8,710 bp).

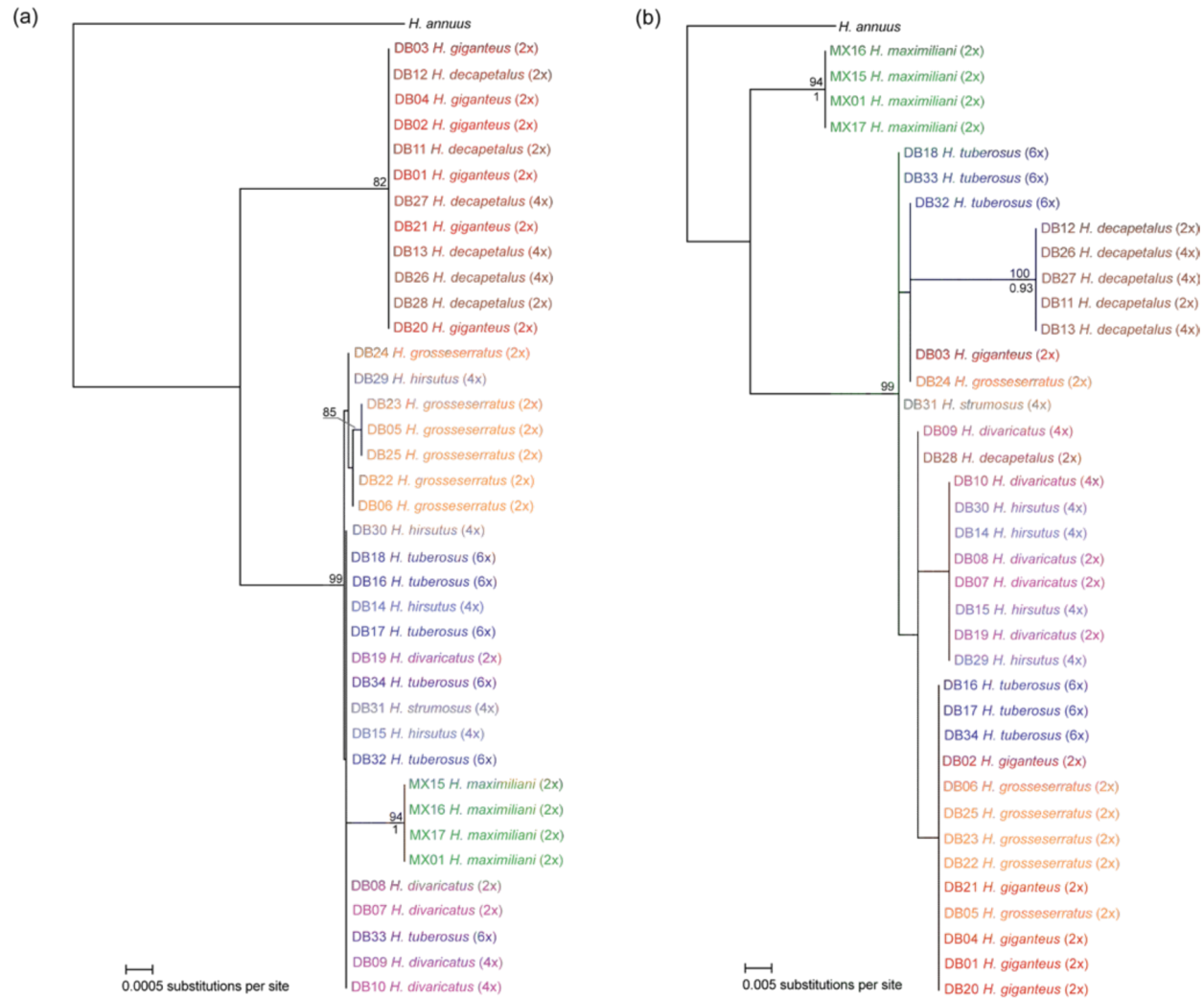




**Figure A.7:** Maximum likelihood phylogenetic reconstruction of (a) complete plastid genomes (151,551 bp), and (b) partial mitochondrial genomes (196,853 bp) for diploid-only accessions. Support is shown for nodes with SH-like values > 70% (above) and Bayesian posterior probabilities > 0.7 (below).



**Figure A.8:** Maximum likelihood phylogenetic reconstruction of (a) 35S rDNA (8,196 bp) and (b) 5S rDNA (514 bp) regions. Support is shown for nodes with SH-like values > 70% (above) and Bayesian posterior probabilities > 0.7 (below).



**Table A.1:** Details of samples used for sequencing.

<b>Taxon</b>	<b>Accession</b>	<b>Sample ID</b>	<b>Ploidy</b>	<b>Sample locality</b>	<b>Latitude</b>	<b>Longitude</b>
<i>H. maximiliani</i>	PI 601812	MX01	2x	Hughes Co., S. Dakota	44° 25' 00" N	100° 00' 00" W
"	PI 592333	MX15	"	Brandon, Manitoba	49° 42' 33" N	99° 57' 44" W
"	PI 650010	MX16	"	Cass Co., N. Dakota	46° 39' 00" N	97° 14' 00" W
"	PI 613794	MX17	"	Woodbury Co., Iowa	42° 27' 04" N	96° 11' 39" W
<i>H. giganteus</i>	PI 547177	DB01	"	Ashland Co., Wisconsin	46° 37' 00" N	90° 46' 00" W
"	PI 664647	DB02	"	Lucas Co., Ohio	41° 35' 27" N	83° 45' 43" W
"	PI 664710	DB03	"	Yancey Co., N. Carolina	35° 48' 42" N	82° 11' 50" W
"	PI 468719	DB04	"	Granville Co., N. Carolina	36° 18' 00" N	78° 35' 00" W
"	PI 547178	DB20	"	Menominee Co., Wisconsin	45° 15' 00" N	88° 36' 00" W
"	PI 503223	DB21	"	Dinwiddie Co., Virginia	36° 00' 00" N	77° 00' 00" W
<i>H. decapetalus</i>	PI 503243	DB11	"	Dinwiddie Co., Virginia	36° 00' 00" N	77° 00' 00" W
"	PI 649972	DB12	"	Fairfax Co., Virginia	38° 50' 00" N	77° 18' 00" W
"	PI 503244	DB28	"	Litchfield Co., Connecticut	41° 00' 00" N	73° 00' 00" W
"	PI 547169	DB13	4x	Columbiana Co., Ohio	40° 38' 00" N	80° 39' 00" W
"	PI 547170	DB26	"	Jefferson Co., Ohio	40° 30' 00" N	80° 55' 00" W
"	PI 468697	DB27	"	Avery Co., N. Carolina	36° 03' 00" N	81° 52' 00" W
<i>H. grosseserratus</i>	PI 468726	DB05	2x	Montgomery Co., Mississippi	33° 28' 00" N	89° 43' 00" W
"	PI 468725	DB06	"	Latimer Co., Oklahoma	34° 55' 00" N	95° 18' 00" W
"	PI 547202	DB22	"	Jasper Co., Iowa	41° 41' 00" N	93° 08' 00" W
"	PI 586890	DB23	"	Cherry Co., Nebraska	42° 10' 00" N	100° 23' 00" W
"	PI 547195	DB24	"	Charlevoix Co., Michigan	45° 12' 00" N	85° 10' 00" W
"	PI 547192	DB25	"	Livingston Co., Illinois	40° 44' 00" N	88° 46' 00" W
<i>H. divaricatus</i>	PI 503218	DB07	"	Huntindon Co., Pennsylvania	40° 00' 00" N	77° 00' 00" W
"	PI 664645	DB08	"	Adams Co., Ohio	38° 48' 39" N	83° 31' 49" W
"	PI 503209	DB19	"	Craig Co., Virginia	37° 00' 00" N	80° 00' 00" W
"	PI 547174	DB09	4x	Effingham Co., Illinois	39° 11' 00" N	88° 48' 00" W
"	PI 664604	DB10	"	Dane Co., Wisconsin	43° 04' 00" N	89° 26' 00" W
<i>H. hirsutus</i>	PI 468739	DB14	"	Le Flore Co., Oklahoma	34° 42' 00" N	94° 32' 00" W
"	PI 495610	DB15	"	Holt Co., Missouri	40° 06' 00" N	95° 13' 00" W
"	PI 547204	DB29	"	Adams Co., Illinois	40° 02' 00" N	90° 54' 00" W
"	PI 468735	DB30	"	Dane Co., Wisconsin	43° 04' 00" N	89° 26' 00" W
<i>H. strumosus</i>	PI 435888	DB31	"	Maury Co., Tennessee	35° 28' 00" N	87° 15' 00" W
<i>H. tuberosus</i>	PI 503279	DB16	6x	Schoharie Co., New York	42° 00' 00" N	74° 00' 00" W
"	PI 547243	DB17	"	Rush Co., Indiana	39° 34' 00" N	85° 27' 00" W
"	PI 650105	DB18	"	Johnson Co., Nebraska	40° 28' 00" N	96° 22' 00" W
"	PI 547248	DB32	"	Cass Co., Illinois	39° 54' 00" N	90° 07' 00" W
"	PI 547230	DB33	"	Allen Co., Ohio	40° 44' 00" N	84° 01' 00" W
"	PI 613795	DB34	"	Woodbury Co., Iowa	42° 22' 10" N	96° 22' 44" W

**Table A.2:** Summary of sequencing data, including reads that passed quality control (i.e. reads with > 36 bp of quality above 10, and any 3 bp sub-sequences of above 20 average quality). Coverage for 2x, 4x, and 6x accessions was estimated based the 2C genome size values for *H. giganteus* (2x; 9.4 Gbp), *H. divaricatus* (4x; 16.5 Gbp) and *H. tuberosus* (6x; 25.3 Gbp; Bennett & Leitch, 2010).

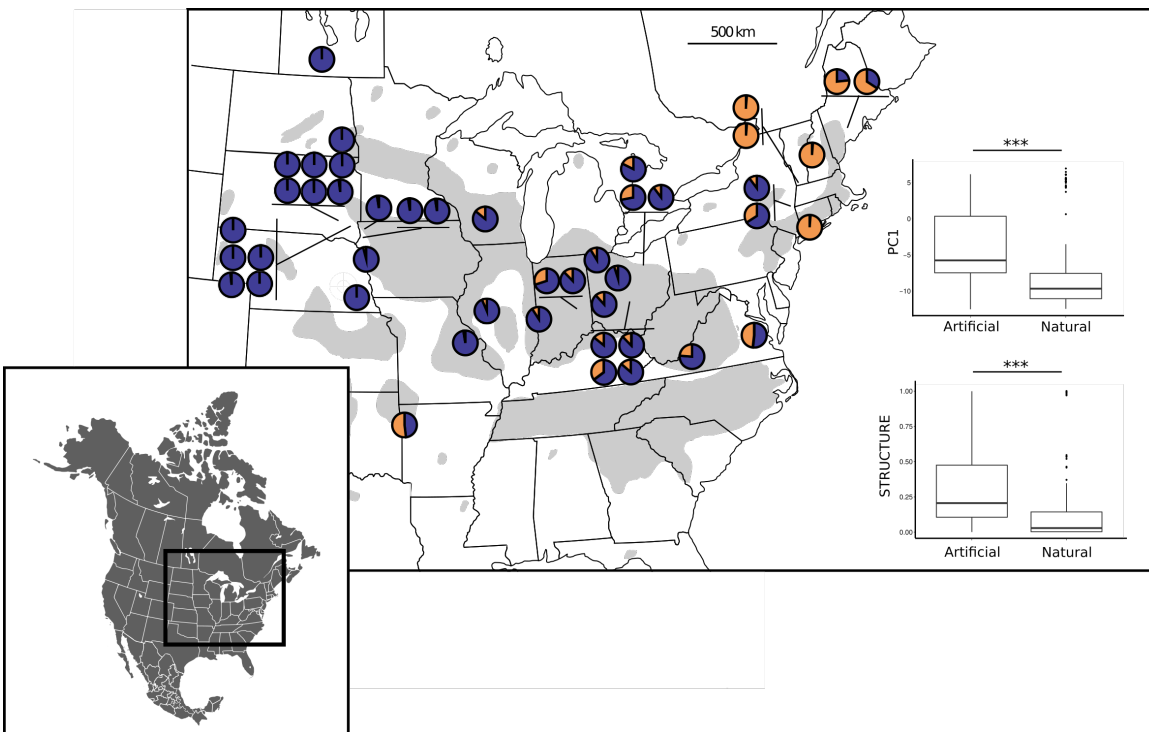
Sample ID	Ploidy	Reads	Total length (Mbp)	Average read length (bp)	Total coverage
MX01	2x	25,696,125	2,399	93	0.25 x
MX15	"	6,424,846	603	94	0.06 x
MX16	"	6,110,243	572	94	0.06 x
MX17	"	5,486,297	517	94	0.05 x
DB01	"	6,899,823	659	95	0.07 x
DB02	"	9,448,256	894	95	0.09 x
DB03	"	8,205,921	777	95	0.08 x
DB04	"	7,417,402	699	94	0.07 x
DB20	"	6,921,829	653	94	0.07 x
DB21	"	6,405,530	604	94	0.06 x
DB11	"	6,835,529	644	94	0.07 x
DB12	"	6,032,053	568	94	0.06 x
DB28	"	8,049,819	759	94	0.08 x
DB13	4x	13,706,235	1295	94	0.08 x
DB26	"	10,518,696	993	94	0.06 x
DB27	"	12,697,192	1198	94	0.07 x
DB05	2x	6,313,648	596	94	0.06 x
DB06	"	9,363,912	886	95	0.09 x
DB22	"	4,736,616	446	94	0.05 x
DB23	"	7,771,209	732	94	0.08 x
DB24	"	5,058,722	477	94	0.05 x
DB25	"	4,294,040	405	94	0.04 x
DB07	"	6,235,212	588	94	0.06 x
DB08	"	6,622,086	623	94	0.07 x
DB19	"	7,756,751	731	94	0.08 x
DB09	4x	12,794,832	1208	94	0.07 x
DB10	"	15,022,683	1417	94	0.09 x
DB14	"	16,734,141	1578	94	0.10 x
DB15	"	12,013,247	1133	94	0.07 x
DB29	"	10,291,521	969	94	0.06 x
DB30	"	11,867,147	1121	94	0.07 x
DB31	"	13,967,130	1316	94	0.08 x
DB16	6x	20,055,926	1891	94	0.08 x
DB17	"	16,704,014	1574	94	0.06 x
DB18	"	18,507,147	1740	94	0.07 x
DB32	"	13,626,680	1279	94	0.05 x
DB33	"	14,423,711	1360	94	0.06 x
DB34	"	10,172,519	948	93	0.04 x

**Table A.3:** Total reads and mean coverage for high-copy regions, estimated based on alignments to the *H. annuus* reference. Only reads that passed quality control are considered.

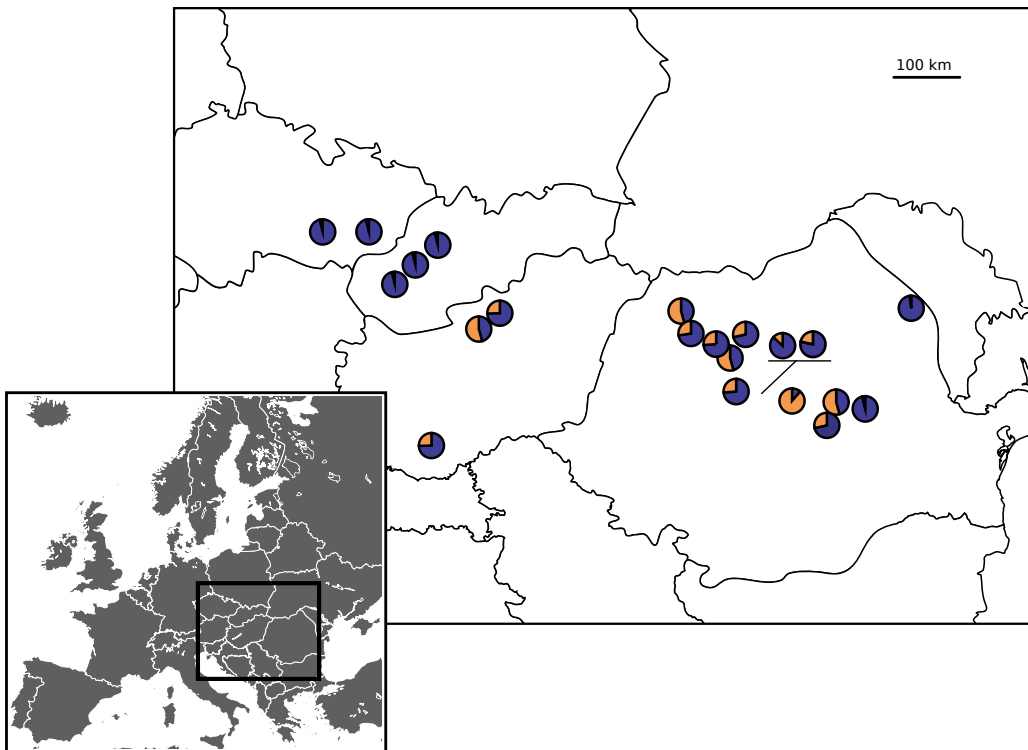
Sample ID	Plastid		Mitochondria		35S rDNA		5S rDNA	
	Total reads	Coverage	Total reads	Coverage	Total reads	Coverage	Total reads	Coverage
MX01	575,399	355 x	57,721	18 x	79,927	740 x	6,695	1,168 x
MX15	143,152	89 x	20,885	6 x	35,950	328 x	1,216	207 x
MX16	148,266	92 x	18,372	6 x	14,824	135 x	1,680	286 x
MX17	128,265	80 x	10,965	3 x	17,286	159 x	809	139 x
DB01	68,561	43 x	19,561	6 x	28,429	258 x	2,106	352 x
DB02	84,850	53 x	24,161	7 x	32,712	297 x	1,972	329 x
DB03	92,754	58 x	23,382	7 x	45,343	412 x	2,036	342 x
DB04	48,686	30 x	13,656	4 x	29,131	263 x	2,268	383 x
DB20	144,245	90 x	21,516	7 x	34,667	314 x	1,563	262 x
DB21	137,795	86 x	17,062	5 x	21,396	194 x	1,878	313 x
DB11	53,303	33 x	20,119	6 x	13,695	123 x	1,525	255 x
DB12	84,898	53 x	20,412	6 x	13,310	119 x	1,928	319 x
DB28	258,432	161 x	23,519	7 x	32,485	294 x	2,099	355 x
DB13	77,665	48 x	30,218	9 x	28,793	261 x	2,852	473 x
DB26	173,137	108 x	33,506	10 x	25,258	229 x	2,161	360 x
DB27	308,258	192 x	33,864	10 x	49,042	445 x	5,031	833 x
DB05	52,399	33 x	17,380	5 x	23,030	208 x	1,948	327 x
DB06	193,449	121 x	34,462	11 x	32,959	299 x	2,625	443 x
DB22	107,105	67 x	14,702	4 x	26,978	243 x	1,482	248 x
DB23	262,641	164 x	30,250	9 x	34,657	313 x	1,606	270 x
DB24	154,448	96 x	18,545	6 x	17,856	161 x	1,198	202 x
DB25	137,912	86 x	17,844	5 x	25,505	230 x	838	142 x
DB07	56,450	35 x	21,122	6 x	17,827	161 x	1,657	280 x
DB08	50,918	32 x	18,009	5 x	20,308	183 x	1,031	177 x
DB19	134,407	84 x	26,847	8 x	23,681	214 x	1,665	275 x
DB09	113,612	71 x	30,002	9 x	28,623	258 x	2,715	458 x
DB10	129,070	80 x	41,821	13 x	29,768	269 x	4,274	726 x
DB14	99,686	62 x	54,710	17 x	44,613	401 x	2,383	401 x
DB15	84,312	52 x	41,345	13 x	46,212	417 x	1,311	215 x
DB29	130,348	81 x	32,562	10 x	28,742	259 x	781	132 x
DB30	199,269	124 x	55,990	17 x	28,092	255 x	1,474	253 x
DB31	105,508	66 x	38,205	12 x	18,275	164 x	2,668	447 x
DB16	169,874	106 x	39,151	12 x	50,245	452 x	3,188	536 x
DB17	131,994	82 x	39,652	12 x	31,853	287 x	3,292	553 x
DB18	169,072	105 x	40,237	12 x	51,139	459 x	4,453	750 x
DB32	435,900	270 x	36,401	11 x	25,179	225 x	2,136	357 x
DB33	123,477	77 x	43,369	13 x	23,237	210 x	1,926	324 x
DB34	239,738	147 x	32,918	10 x	21,396	189 x	1,700	274 x

## Appendix B - Chapter 4 Supplementary Material

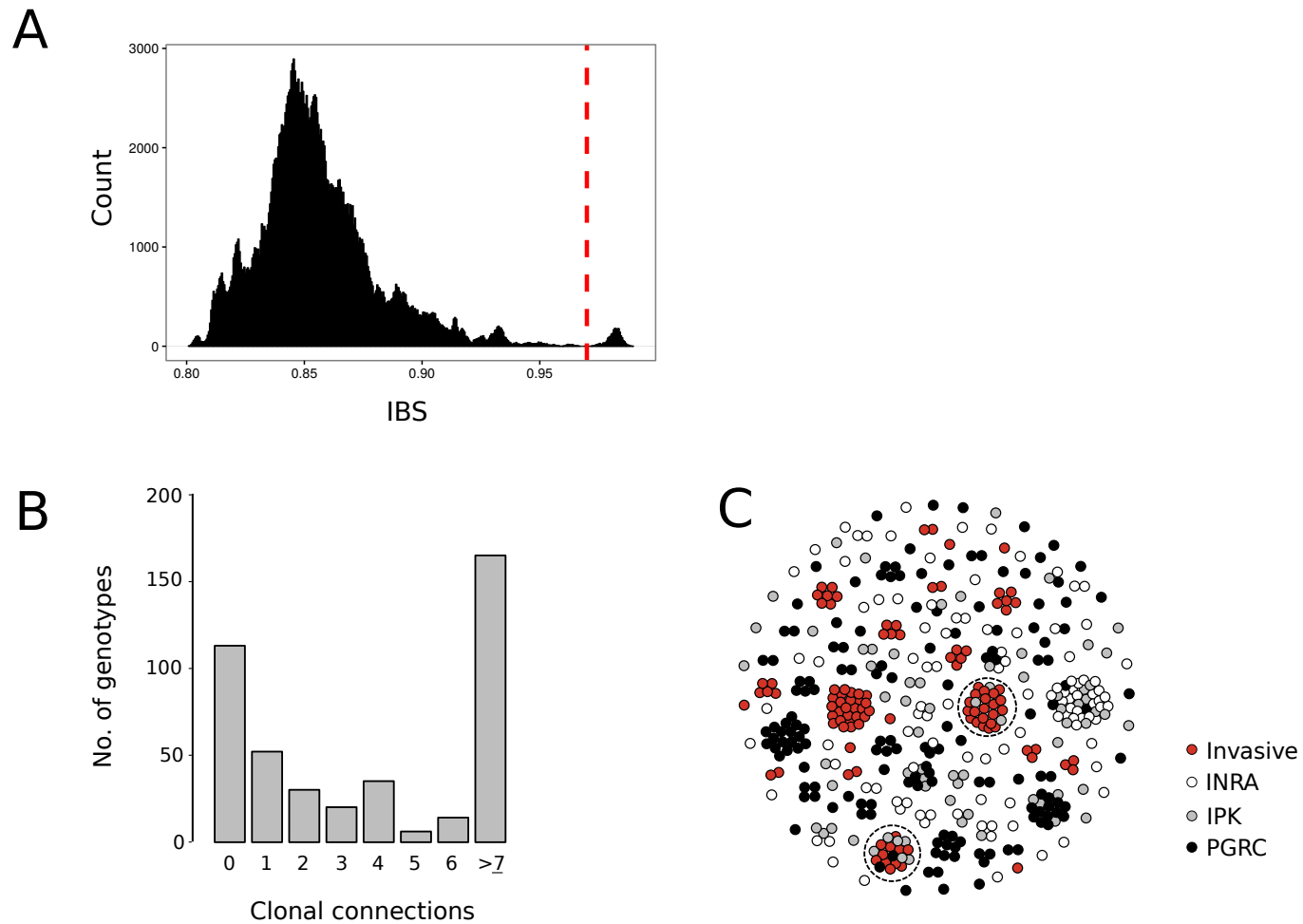
**Figure B.1:** Map of North America showing *H. tuberosus* USDA sampling locations for which GPS coordinates were available (45 of 49 accessions; Table B.1). Light gray shading is used for the US range of *H. tuberosus*, following Rogers *et al.* (1982). Pie charts represent average membership proportions for the two main genetic clusters identified in the STRUCTURE analysis. The two boxplots represent PC1 scores (top plot) and membership coefficients for the STRUCTURE cluster that is dominant in cultivated *H. tuberosus* (bottom plot) for samples collected from artificial and natural habitats. Both comparisons between site categories are statistically significant (PC1: Mann-Whitney *U* test,  $P < 2.5 \times 10^{-10}$ ; STRUCTURE: Mann-Whitney *U* test,  $P < 1.1 \times 10^{-8}$ ).



**Figure B.2:** Map of Europe showing sampling locations for which GPS coordinates were available (21 of 42 accessions; Table B.1). Pie charts represent average membership proportions for the two main genetic clusters identified in the STRUCTURE analysis.

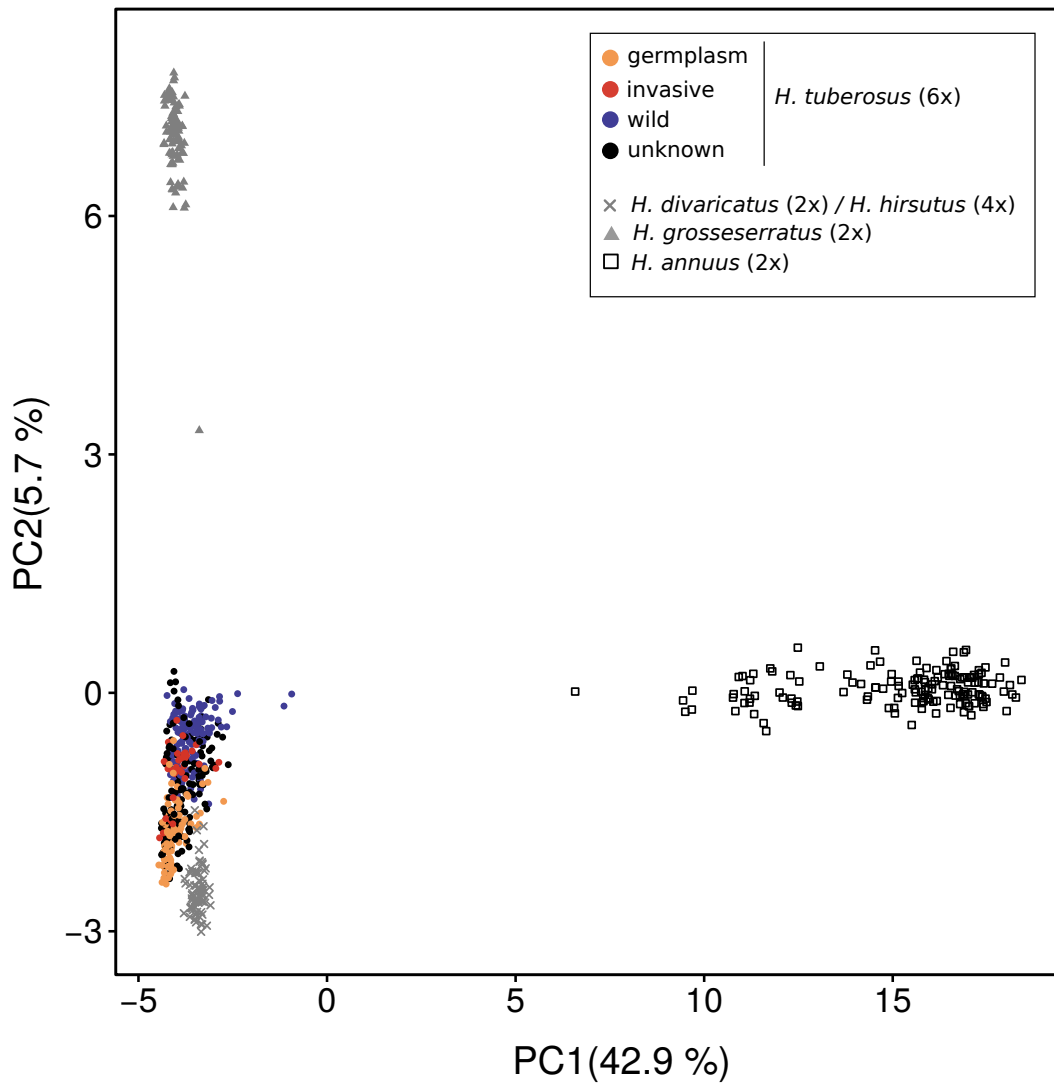


**Figure B.3:** Clonality in the *H. tuberosus* collection. (A) IBS distribution from pairwise comparisons among 695 samples, calculated using 78,839 diploidized SNPs. The red dashed line marks the 0.97 threshold set to delineate clonal and non-clonal relationships. (B) Number of clonal connections evaluated for 435 tuber-propagated accessions. (C) Clonal relationships represented as clusters for the same set of 435 samples. Single data points represent unique accessions. The dashed circles delineate the two clonal series containing invasive samples obtained in 2013, as well as PGRC or IPK accessions (see text for details).

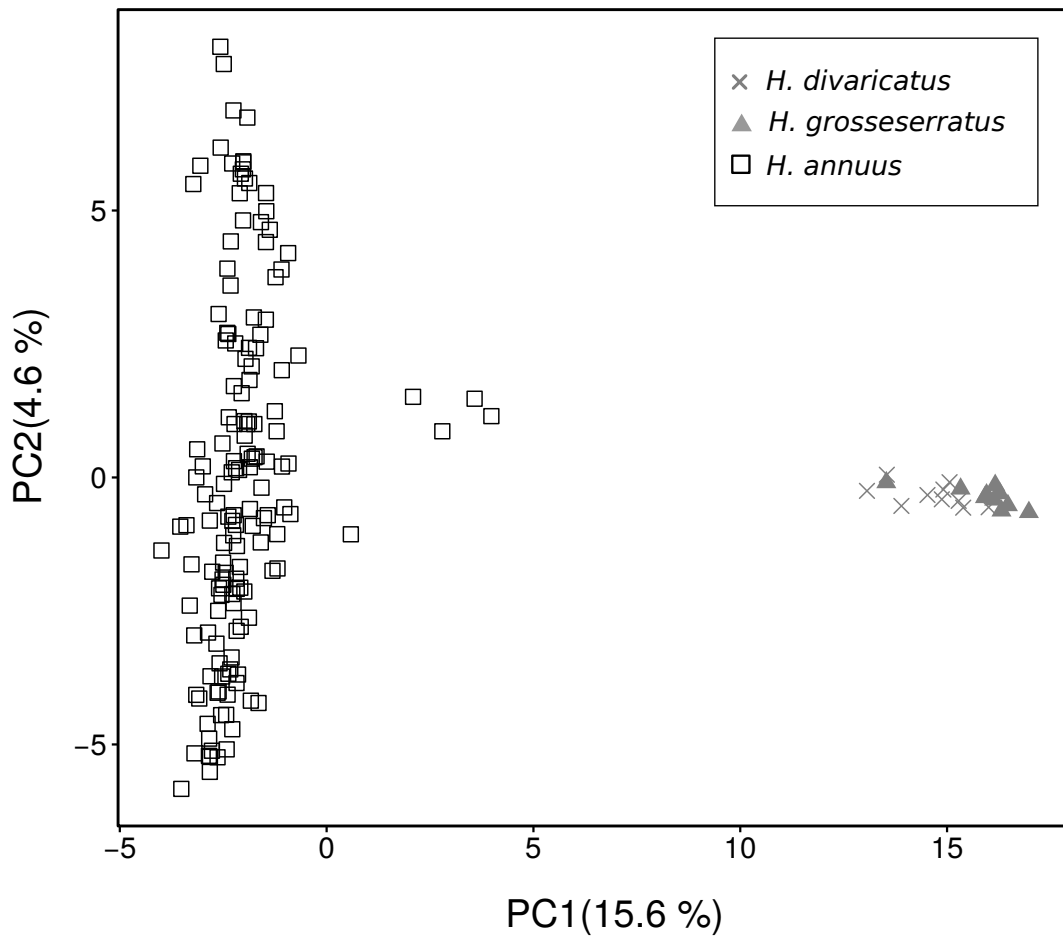




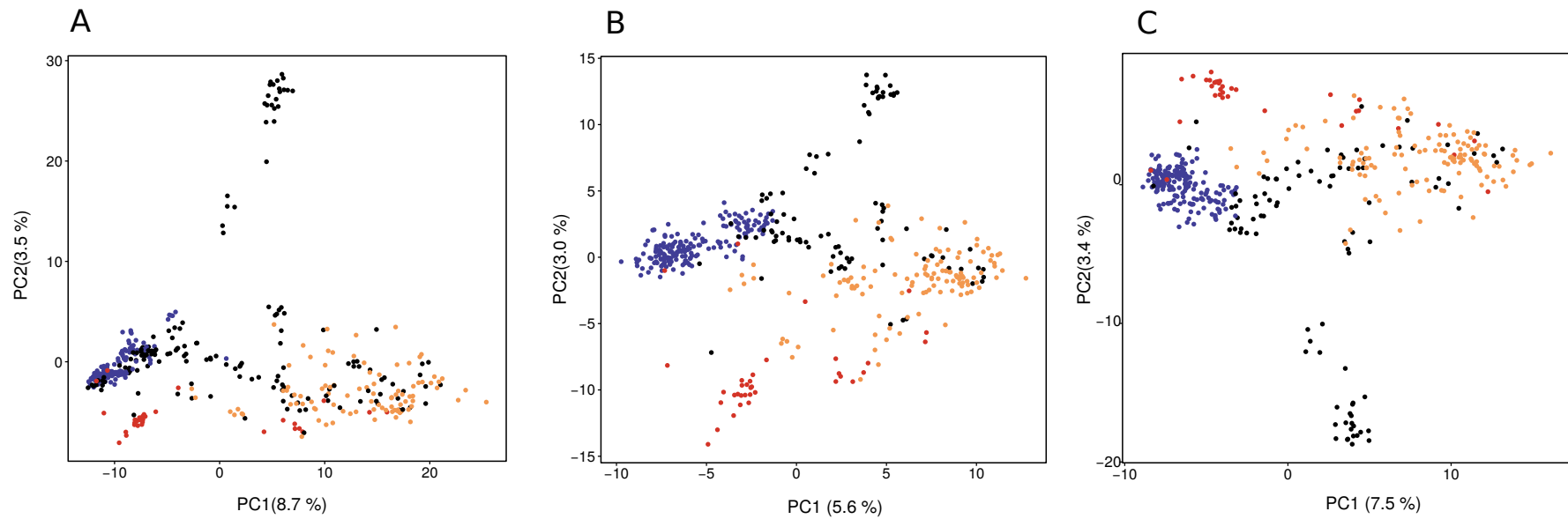
**Figure B.4:** PCA including 602 samples of *H. tuberosus*, *H. grosseserratus*, *H. divaricatus* and *H. hirsutus*, and 150 samples of *H. annuus* based on 2,916 SNPs. PC1 corresponds to divergence between annual and perennial species.



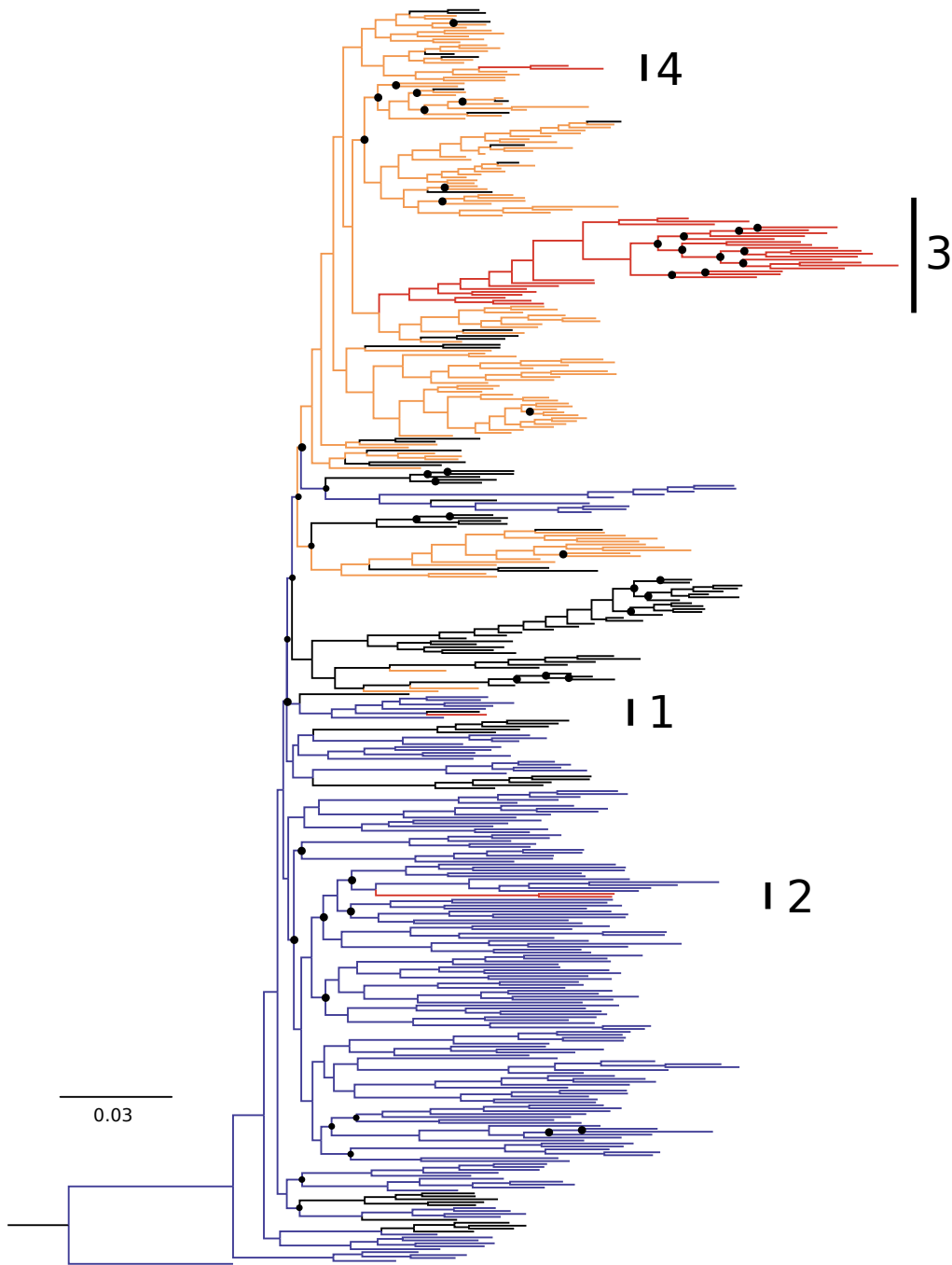
**Figure B.5:** PCA including only samples sequenced using WGSS, including 150 samples of *H. annuus*, 10 samples of *H. grosseserratus*, and 10 samples of *H. divaricatus*. The analysis is based on 3,916 SNPs. PC1 summarizes divergence between perennial (*H. grosseserratus* and *H. divaricatus*) and annual (*H. annuus*) species.



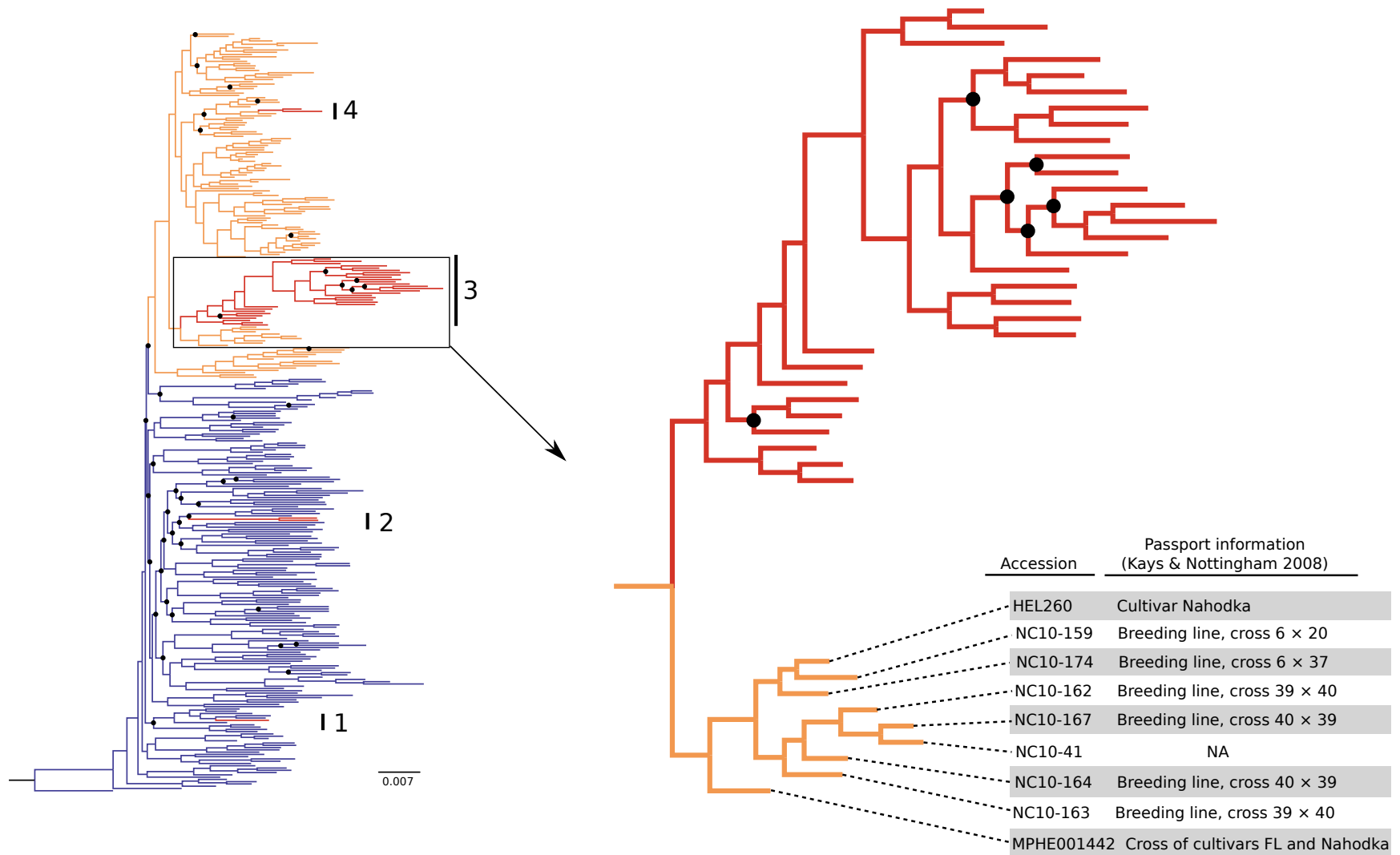
**Figure B.6:** PCA including 427 *H. tuberosus* samples assigned to the native (blue), cultivated (orange), invasive (red), or unknown (black) categories. (A) Analysis based on all 82,957 SNPs and genotypes in hexaploid format. (B) Analysis based on 4,700 SNPs assigned to the diploid subgenome, with genotypes converted to diploid format. (C) Analysis based on 4,770 SNPs assigned to the tetraploid subgenome, with genotypes converted to tetraploid format.



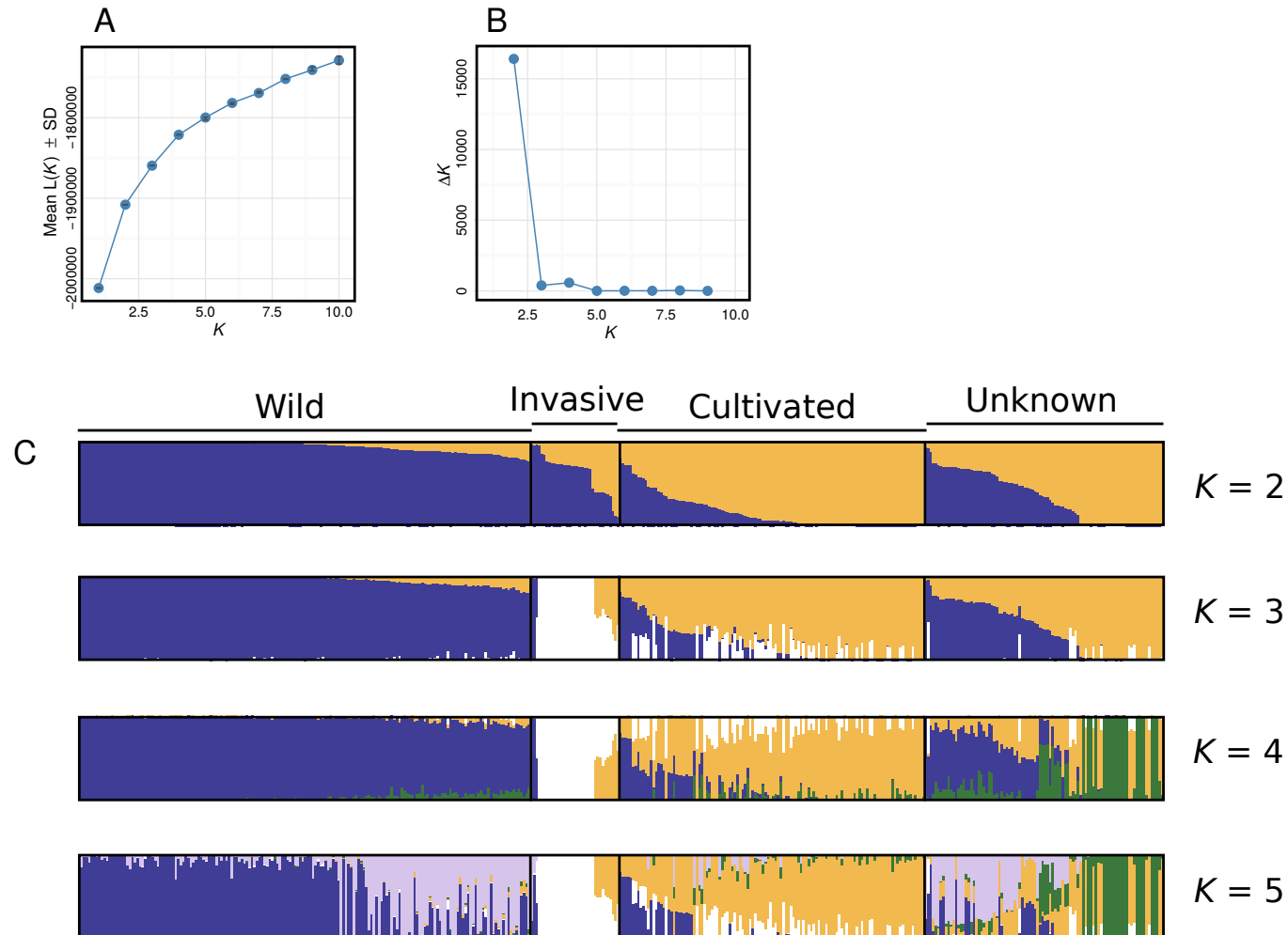
**Figure B.7:** Maximum-likelihood phylogeny for the *H. tuberosus* unique genotype set (427 samples), including native (blue) cultivated (orange) invasive (red) or unknown (black) samples. The analysis is based on 4,700 SNPs assigned to the diploid subgenome. The tree is rooted with a sample of *H. grosseserratus*. Black circles indicate bootstrap values lower than 70%. Clades 1-4 indicate the four inferred origins of invasive genotypes.



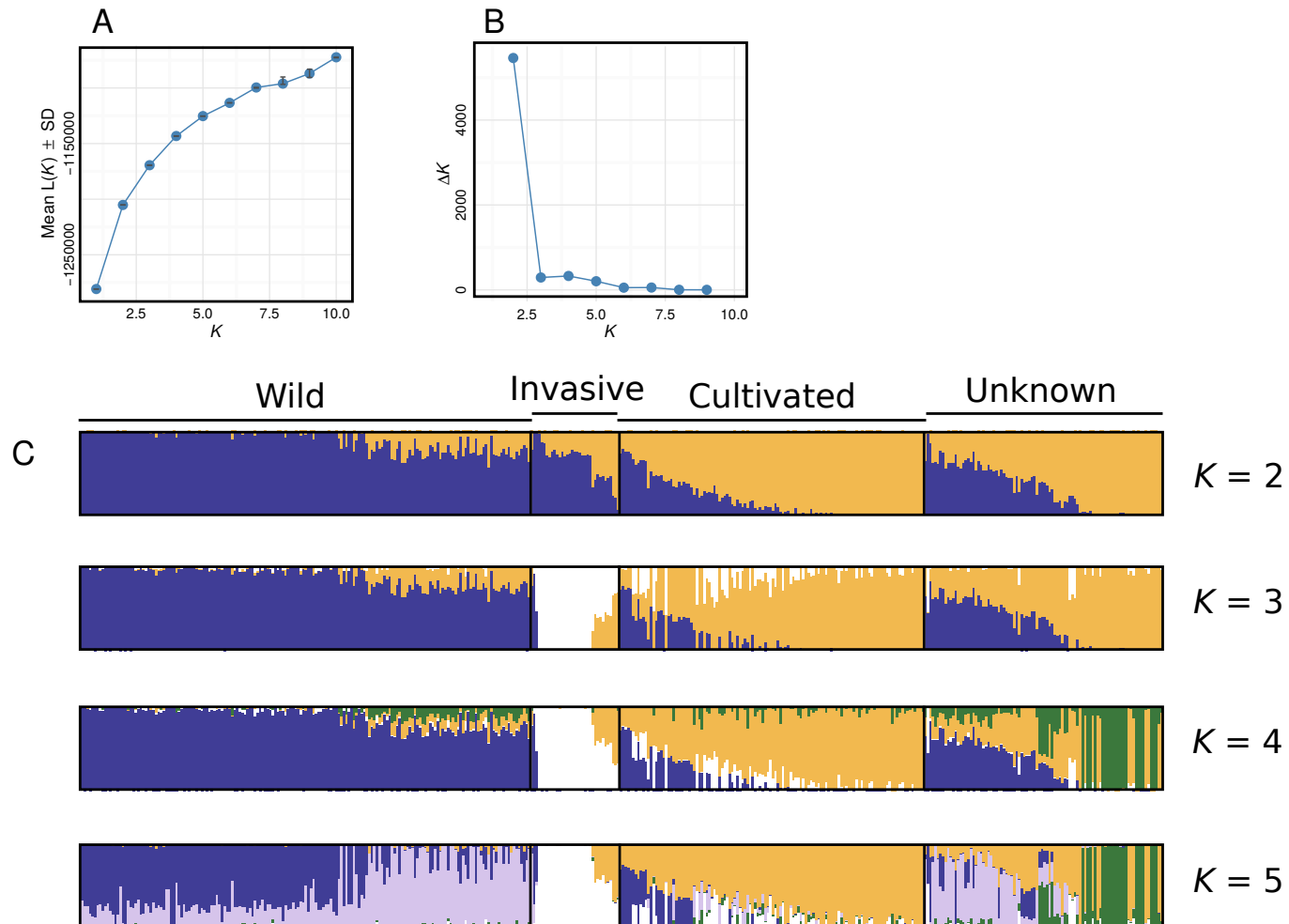
**Figure B.8:** Pedigree information (detailed in Kays & Nottingham 2008) for the cultivated samples clustering with invasive *H. tuberosus* genotypes in clade 3 (as identified in Fig. 4.1C and Appendix Figure B.7).



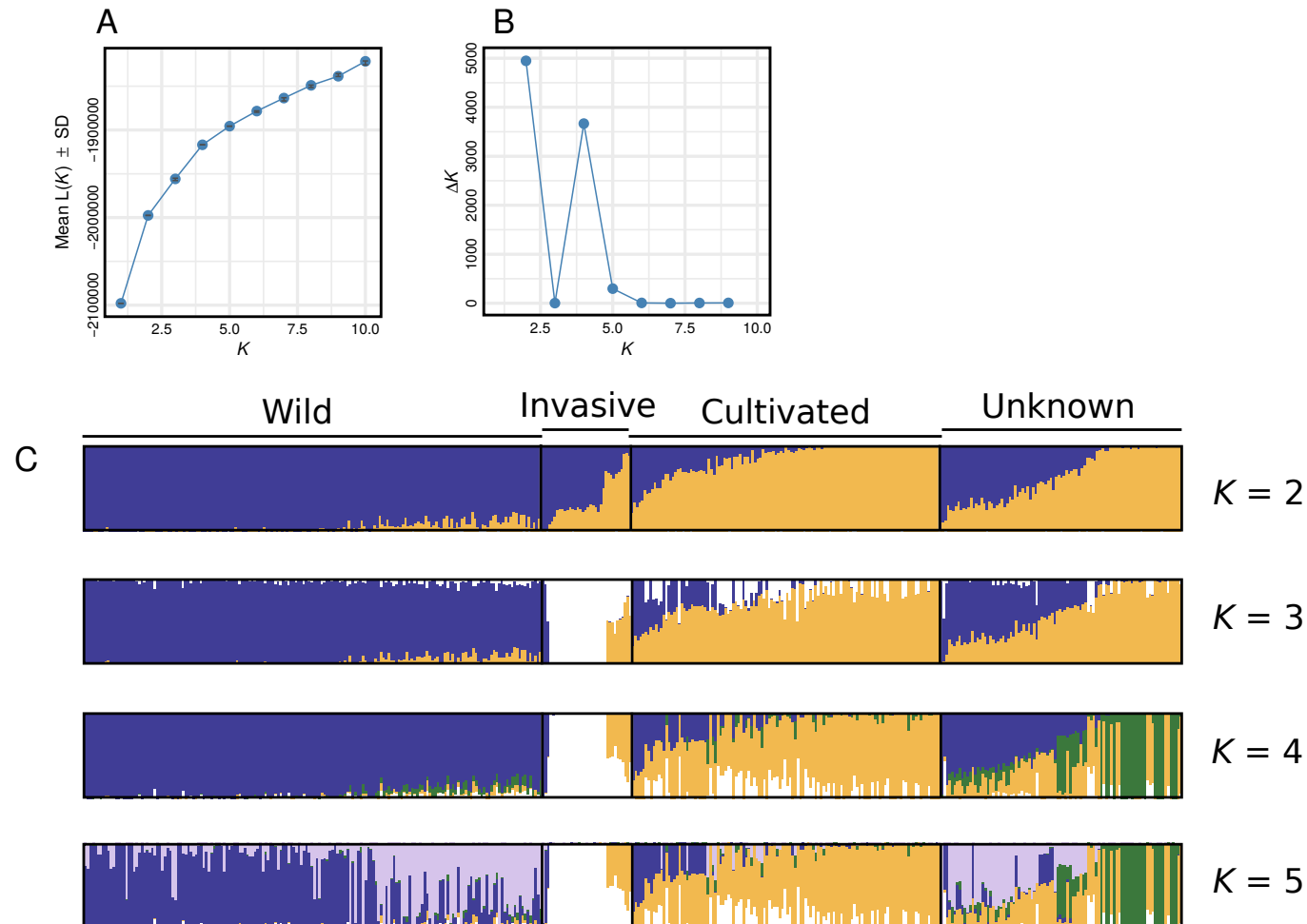
**Figure B.9:** Bayesian STRUCTURE analysis of 427 *H. tuberosus* samples based on 5,000 SNPs that were randomly selected from the filtered hexaploid SNP set. (A) Log probability of the data, and (B)  $\Delta K$  as a function of number of  $K$  inferred clusters. (C) Plots of individual membership coefficients for values of  $K$  ranging from 2 to 5. Each of the 427 samples is represented by a vertical stacked bar, with colors indicating genetic clusters. Black vertical lines separate samples assigned to the four categories.



**Figure B.10:** Bayesian STRUCTURE analysis of 427 *H. tuberosus* samples based on 4,700 SNPs assigned to the diploid subgenome, for which genotypes were converted to diploid format. (A) Log probability of the data, and (B)  $\Delta K$  as a function of number of  $K$  inferred clusters. (C) Plots of individual membership coefficients for values of  $K$  ranging from 2 to 5. Each of the 427 samples is represented by a vertical stacked bar, with colors indicating genetic clusters. Black vertical lines separate samples assigned to the four categories.

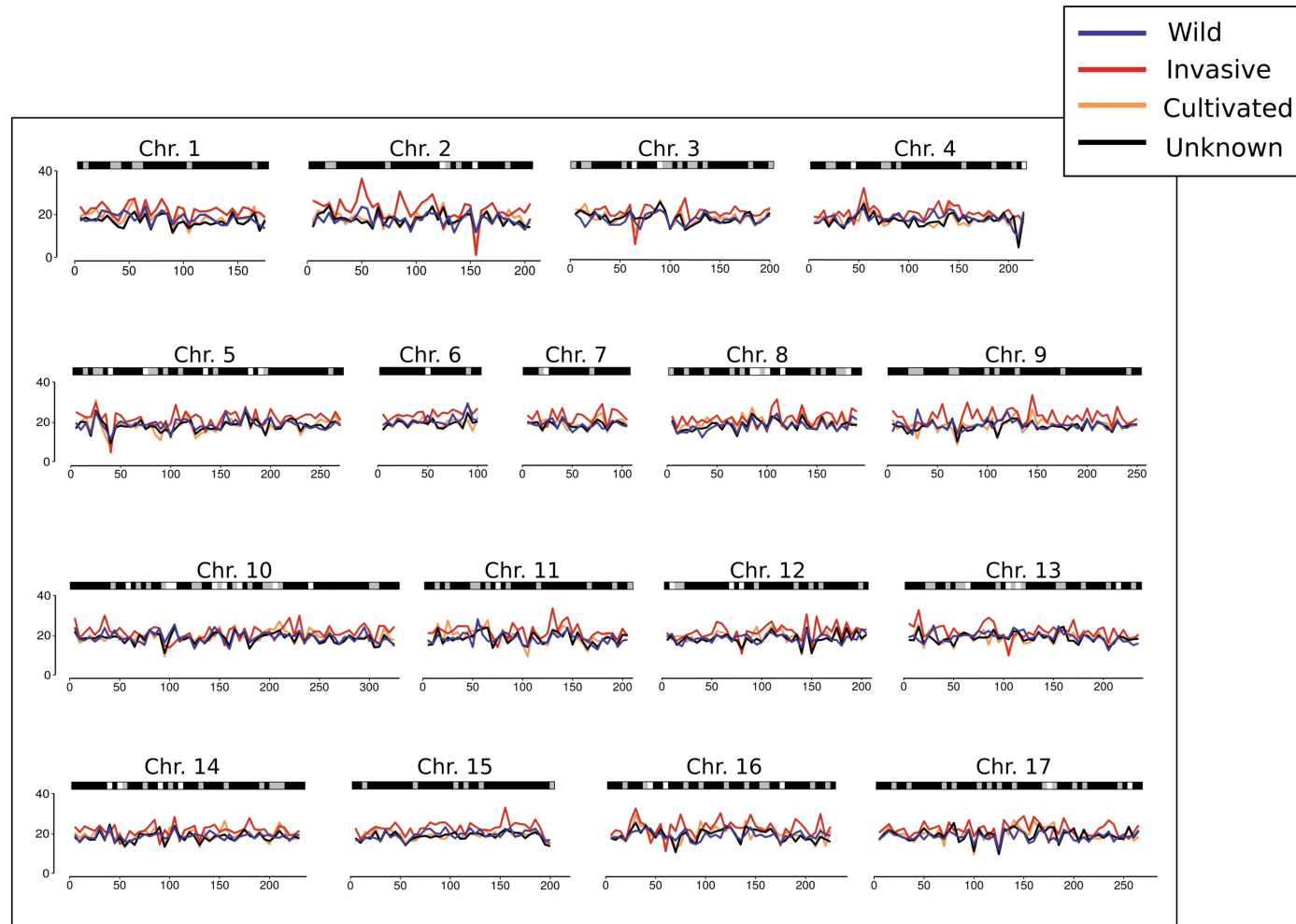


**Figure B.11:** Bayesian STRUCTURE analysis of 427 *H. tuberosus* samples based on 4,770 SNPs assigned to the tetraploid subgenome, for which genotypes were converted to tetraploid format. (A) Log probability of the data, and (B)  $\Delta K$  as a function of number of  $K$  inferred clusters. (C) Plots of individual membership coefficients for values of  $K$  ranging from 2 to 5. Each of the 427 samples is represented by a vertical stacked bar, with colors indicating genetic clusters. Black vertical lines separate samples assigned to the four categories.

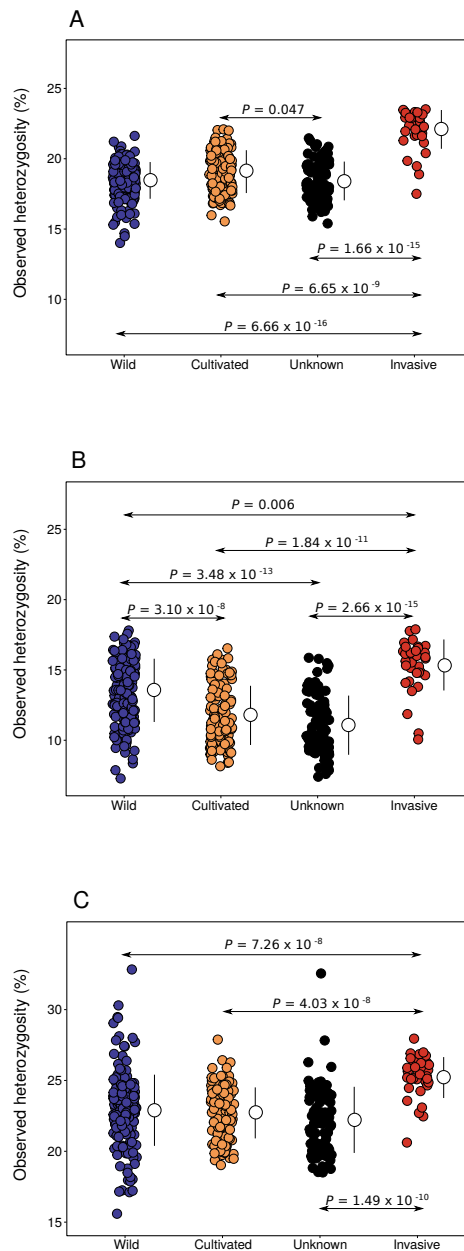




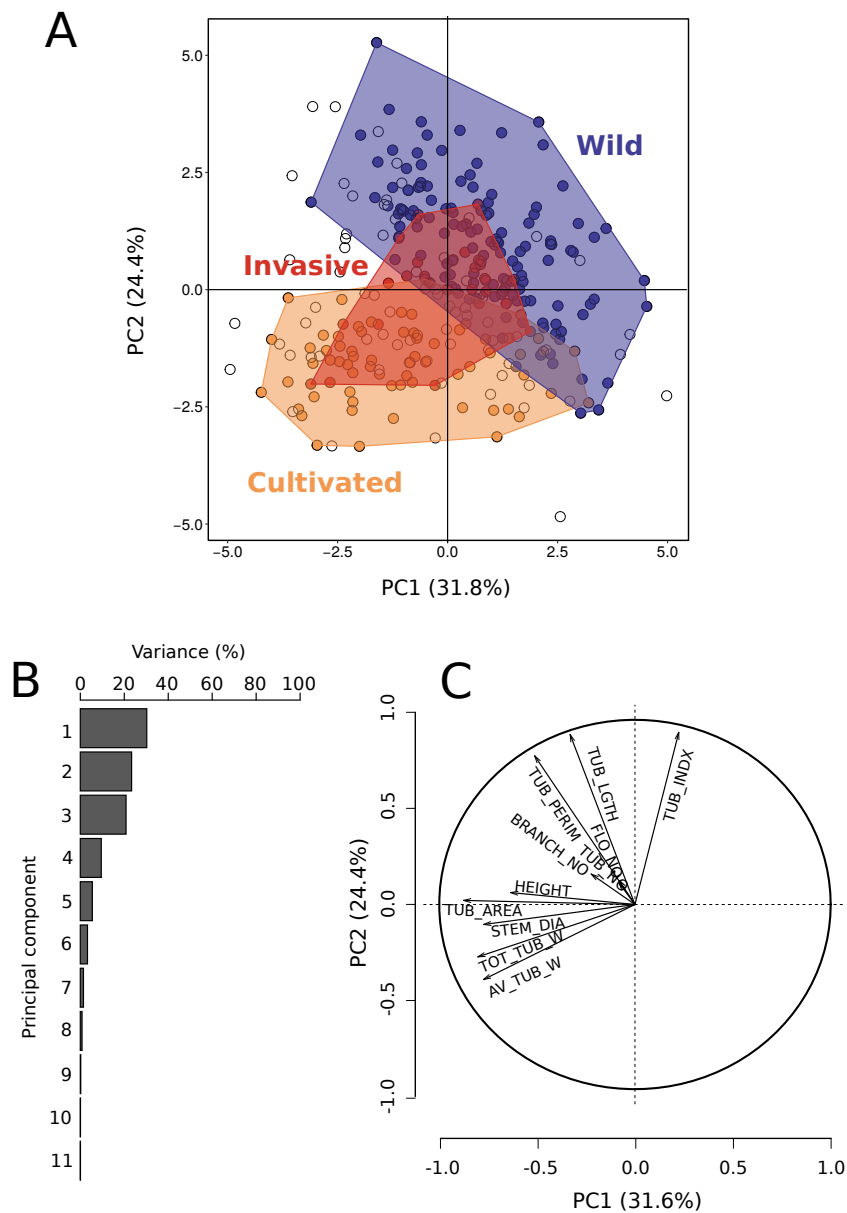
**Figure B.12:** Heterozygosity calculated in 5MB non-overlapping windows along the genome. The X axis indicates genome position (MB), and the Y axis indicates heterozygosity (%) calculated within each window for each of the native, invasive, cultivated, or unknown sample categories. The bar above each plot indicates windows for which heterozygosity of invasive samples was highest (black), lowest (white), or intermediate between other sample categories (grey).



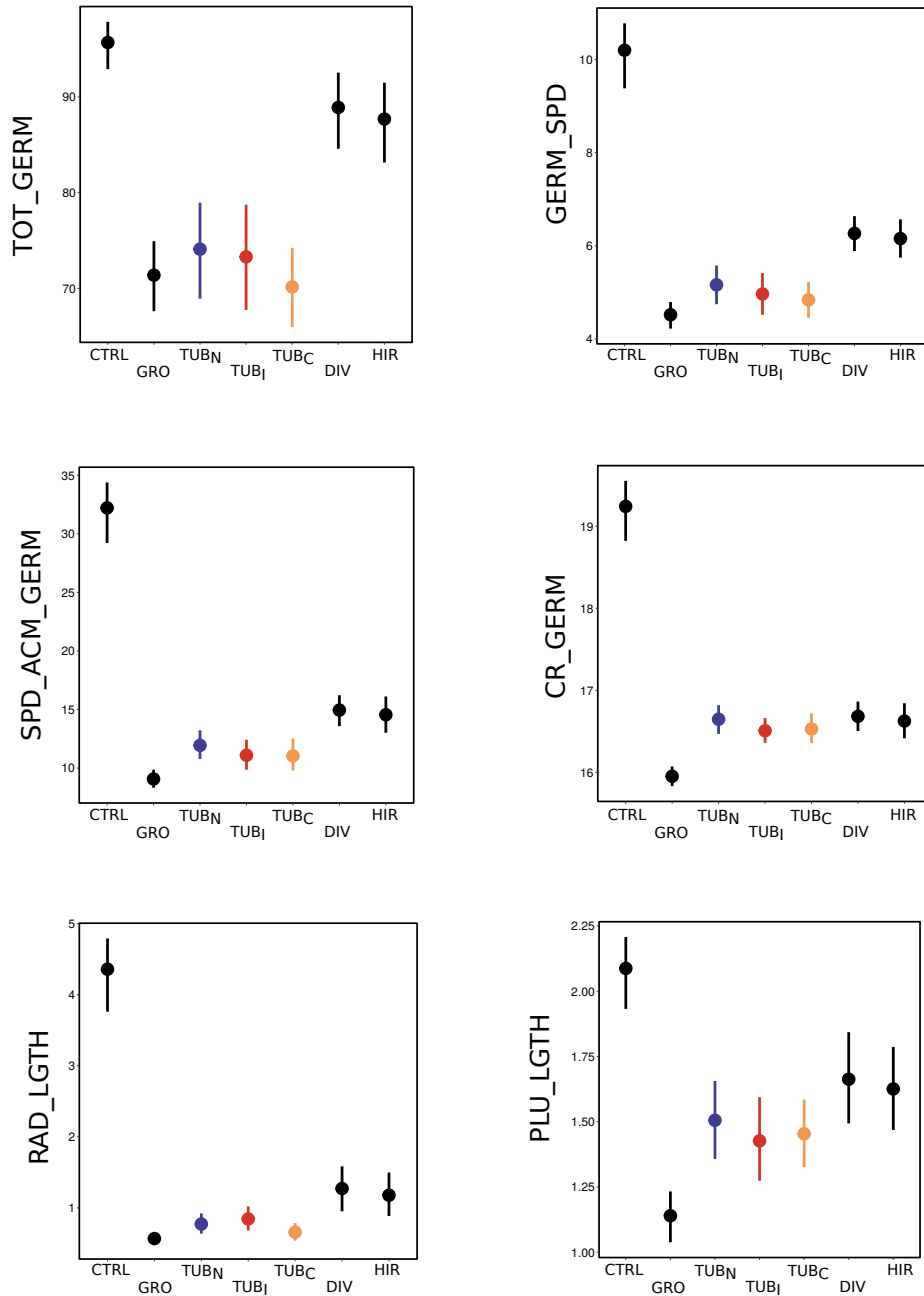
**Figure B.13:** Genome-wide heterozygosity calculated for 427 *H. tuberosus* samples across 82,957 SNPs with genotypes in hexaploid format (A), 4,700 SNPs assigned to the diploid subgenome with genotypes converted to diploid format (B), and 4,770 SNPs assigned to the tetraploid subgenome with genotypes converted to tetraploid format (C). White circles are used to indicate mean ( $\pm$  SD) for each of the native, cultivated, invasive, or unknown sample categories. Significance was tested using Kruskal–Wallis and posthoc Nemenyi-tests implemented in the PMCMR (Pohlert, 2015) package in R. *P* values are presented for significant comparisons, after adjustment using Holm (1979).



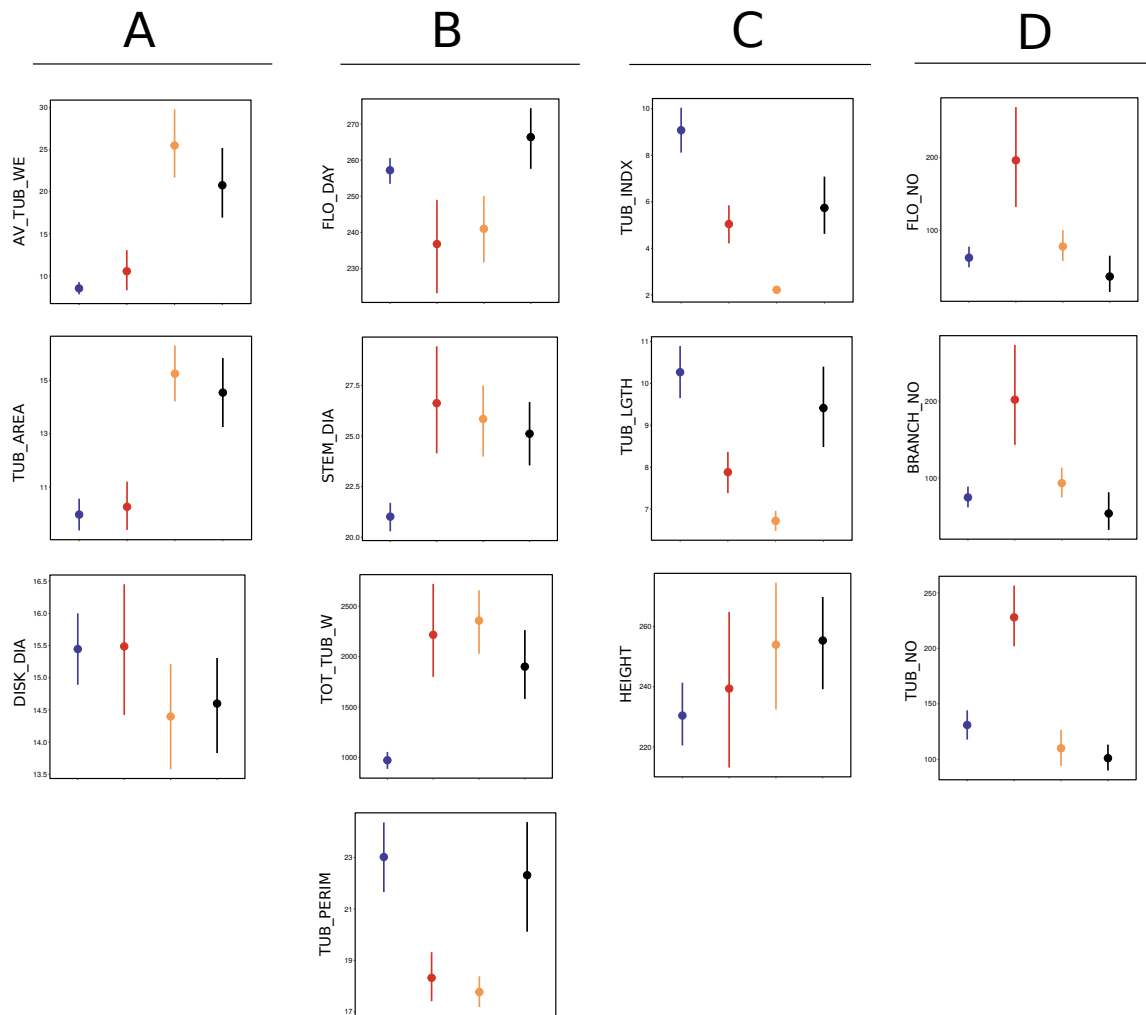
**Figure B.14:** (A) PCA for 305 *H. tuberosus* samples included in the greenhouse and common garden experiments. The analysis is based on 11 quantitative traits with data in 90% or more of samples, for which missing data was replaced with the average value observed across the collection. Samples are colored based on membership to the native (blue), cultivated (orange), invasive (red), or unknown (white) categories. Polygons enclose native, cultivated, and invasive samples. (B) Percentage of the variance accounted by the 11 principal components. (C) Loading plot showing the correlation between the factor loadings and the first two principal components. Trait abbreviations are as per Table B.3.



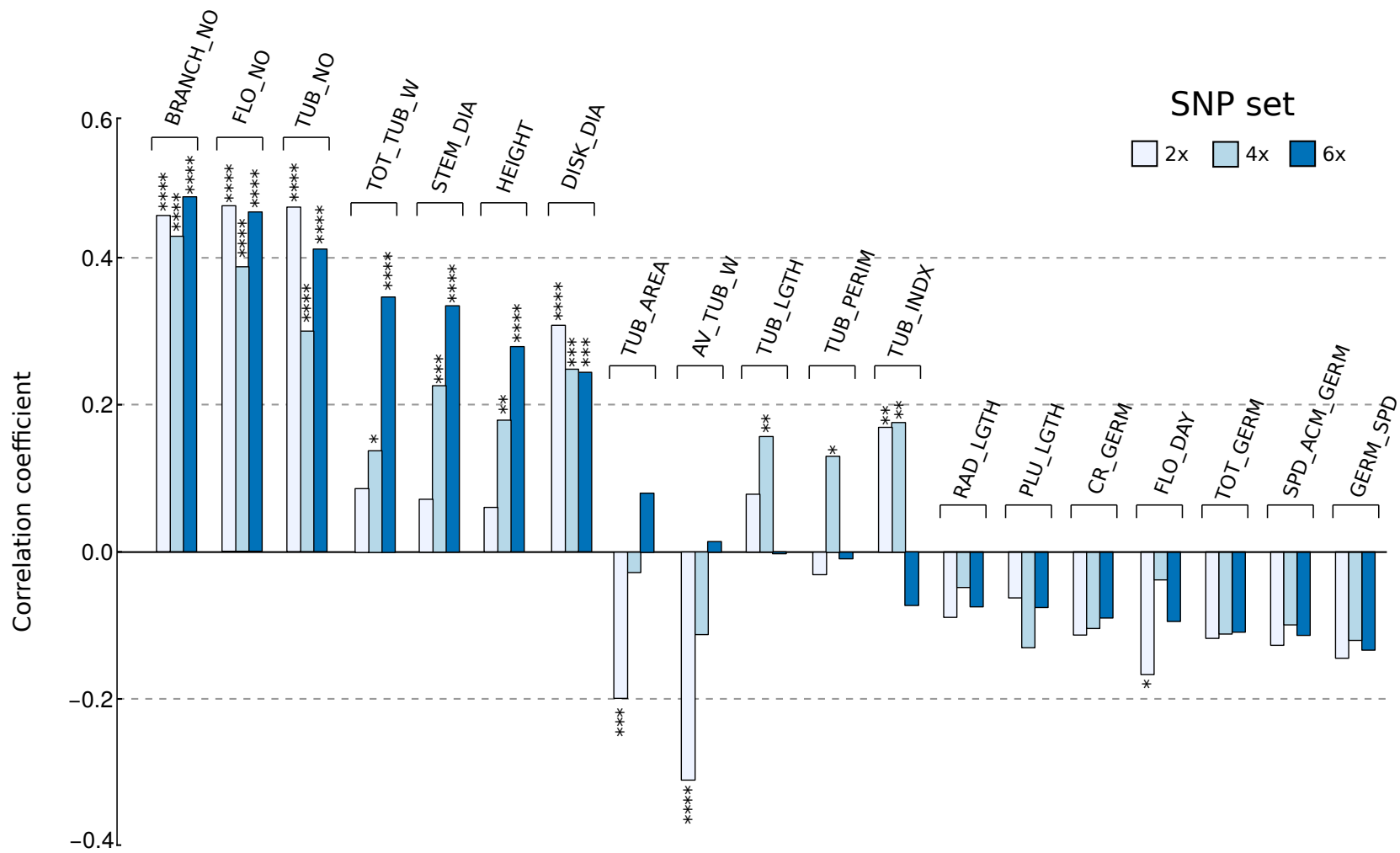
**Figure B.15:** Means and 95% bootstrapped confidence intervals for six metrics of allelopathy in *H. tuberosus* [native (TUB<sub>N</sub>), invasive (TUB<sub>I</sub>) or cultivated (TUB<sub>C</sub>) ] and progenitor species [*H. grosseserratus* (GRO), *H. divaricatus* (DIV), *H. hirsutus* (HIR)]. Trait abbreviations are as per Table B.3. CTRL represents the the control extracts, which contained only water.



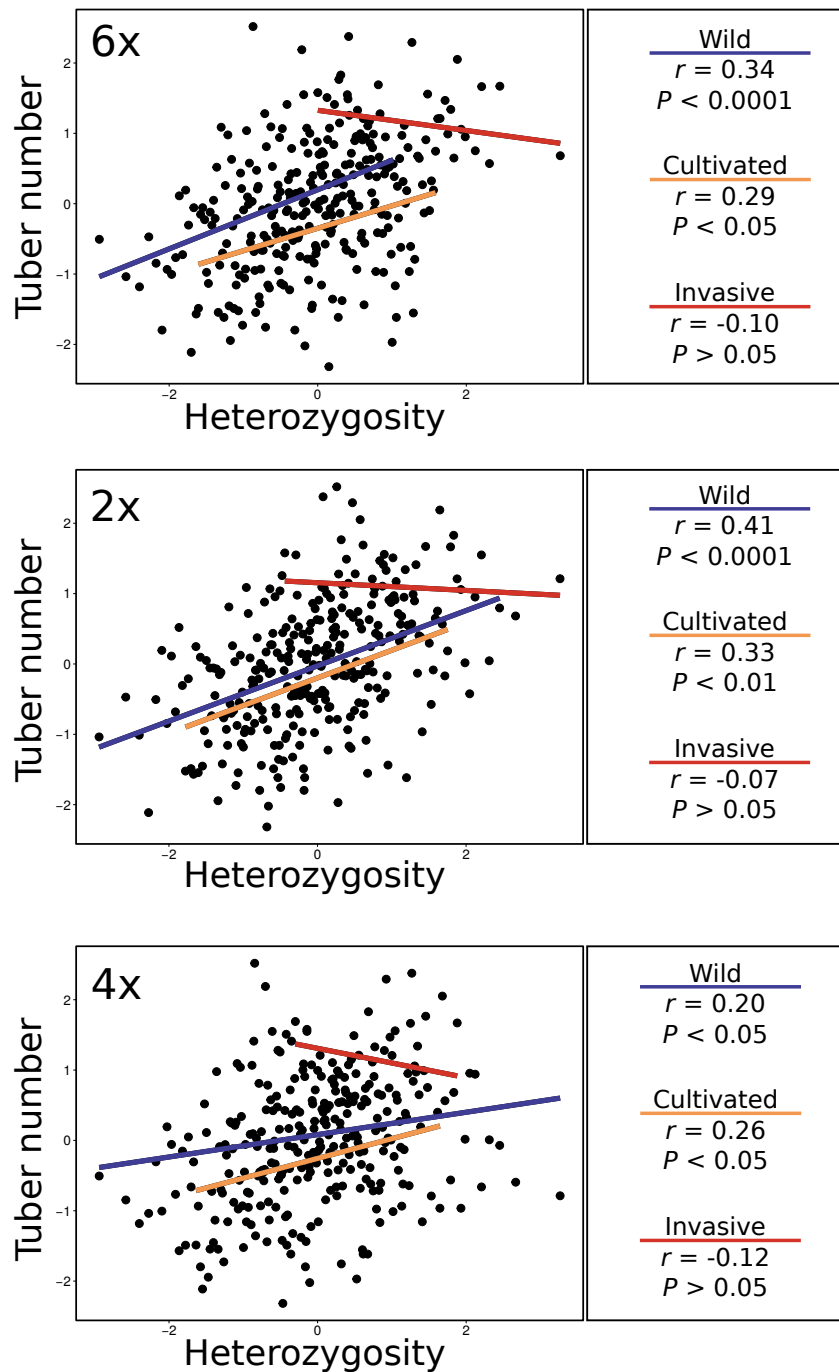
**Figure B.16:** Means and 95% bootstrapped confidence intervals for 13 traits measured in a common garden in native (blue), invasive (red) cultivated (orange), and unknown (black) *H. tuberosus*. Trait abbreviations are as per Table B.3. Column A are traits for which invasive samples are similar to the native group. Column B are traits for which invasive samples are similar to the cultivated group. Column C group traits for which the invasive group is intermediate. Column D is for traits showing extreme values in invasive samples.



**Figure B.17:** Pearson correlation coefficients between all trait values as heterozygosity estimated per sample based on all markers (6x), diploid subgenome markers (2x), and tetraploid subgenome markers (4x). Asterisks are used to indicate significance for each correlation. Trait abbreviations are as per Table B.3.



**Figure B.18:** Correlations between tuber number and heterozygosity of whole-genome genotypes for samples assigned to the native (blue), cultivated (orange), or invasive (red) categories. Heterozygosity is calculated based on all hexaploid-format markers (6x), diploid subgenome markers (2x), or tetraploid subgenome markers (4x).



**Table B.1:** Sampling information for all *H. tuberosus* accessions. For USDA accessions, the asterisk is used to denote artificial sampling habitats. *N* denotes sample size per accession. For sample IDs, letters in parentheses are used to indicate samples categorized as native (n), cultivated (c), invasive (i), or unknown (u). Sample IDs for USDA and Rieseberg Lab accessions include only intra-population/accession identifiers [e.g. 503262\_1(n) is presented as 1(n); ]

Collection	Accession	Origin	Lat.	Long.	<i>N</i>	Sample IDs	Obtained from
USDA	PI 503262	USA	37	-80	5	1(n), 4(n), 10(u), 12(n), 15(n)	seeds
USDA	PI 503265 *	USA	37	-75	5	2(u), 3(u), 5(u), 6(u), 8(u)	seeds
USDA	PI 503272	USA	41	-73	6	1(u), 4(u), 5(u), 6(u), 7(u), 11(u)	seeds
USDA	PI 503274 *	USA	44	-72	5	2(u), 3(u), 23(u), 24(u), 26(u)	seeds
USDA	PI 503276	USA	43	-73	5	1(u), 2(u), 4(u), 11(u), 14(u)	seeds
USDA	PI 503277	USA	43	-73	5	9(u), 10(u), 11(u), 12(u), 13(u)	seeds
USDA	PI 503278	USA	42	-74	6	1(n), 6(n), 7(n), 9(n), 18(n), 20(n)	seeds
USDA	PI 503279	USA	42	-74	5	2(u), 3(u), 4(u), 8(u), 12(u)	seeds
USDA	PI 547227 *	USA	43.4	-89.71	5	5(n), 7(n), 8(n), 9(n), 11(n)	seeds
USDA	PI 547228	USA	41.23	-84.35	4	22(n), 24(n), 25(n), 26(n)	seeds
USDA	PI 547230	USA	40.73	-84.01	5	4(n), 5(n), 12(n), 15(n), 20(n)	seeds
USDA	PI 547232 *	USA	40.6	-82.16	5	20(n), 21(n), 23(n), 24(n), 25(n)	seeds
USDA	PI 547233	USA	40.61	-81.85	5	4(n), 8(n), 13(n), 14(n), 15(n)	seeds
USDA	PI 547234	USA	40.86	-81.83	5	6(n), 20(u), 21(u), 22(u), 23(u)	seeds
USDA	PI 547237	USA	39.41	-83.03	5	6(n), 7(n), 8(n), 14(n), 15(n)	seeds
USDA	PI 547238 *	USA	39.3	-83.15	4	21(n), 22(n), 23(n), 24(n)	seeds
USDA	PI 547239 *	USA	39.01	-83.2	5	6(u), 7(u), 12(u), 14(u), 15(u)	seeds
USDA	PI 547241 *	USA	39.31	-83.7	5	21(n), 22(n), 23(u), 24(n), 25(n)	seeds
USDA	PI 547242 *	USA	39.6	-84.35	5	21(n), 22(n), 23(n), 24(n), 25(n)	seeds
USDA	PI 547243 *	USA	39.56	-85.45	5	3(n), 6(n), 10(n), 13(n), 14(n)	seeds



Collection	Accession	Origin	Lat.	Long.	<i>N</i>	Sample IDs	Obtained from
USDA	PI 547244 *	USA	39.16	-85.61	5	1(u), 2(u), 3(u), 8(u), 13(u)	seeds
USDA	PI 547247	USA	39.1	-87.63	5	8(n), 9(n), 11(n), 13(n), 14(n)	seeds
USDA	PI 547248	USA	39.9	-90.11	5	6(n), 12(n), 13(n), 14(n), 15(n)	seeds
USDA	PI 613795	USA	42.39	-96.37	5	3(n), 9(n), 12(n), 13(n), 20(n)	seeds
USDA	PI 613796	USA	41.35	-95.90	5	7(n), 13(n), 15(n), 20(n), 21(n)	seeds
USDA	PI 650089	USA	42.81	-96.68	5	20(n), 22(n), 23(n), 24(n), 25(n)	seeds
USDA	PI 650090	USA	42.81	-96.68	5	20(n), 21(n), 22(n), 23(n), 24(n)	seeds
USDA	PI 650091	USA	42.81	-96.68	4	21(n), 22(n), 24(n), 25(n)	seeds
USDA	PI 650092 *	USA	42.91	-96.95	5	21(n), 22(n), 23(n), 24(n), 25(n)	seeds
USDA	PI 650093	USA	42.91	-96.95	5	21(n), 22(n), 23(n), 24(n), 25(n)	seeds
USDA	PI 650094	USA	42.91	-96.95	3	21(n), 23(n), 24(n)	seeds
USDA	PI 650095	USA	42.91	-96.95	5	4(n), 8(n), 12(n), 21(n), 23(n)	seeds
USDA	PI 650096	USA	42.91	-96.95	5	20(n), 21(n), 22(n), 23(n), 25(n)	seeds
USDA	PI 650097	USA	42.81	-96.68	5	1(n), 2(n), 11(n), 12(n), 15(n)	seeds
USDA	PI 650098	USA	42.73	-96.23	5	20(n), 21(n), 22(n), 23(n), 25(n)	seeds
USDA	PI 650099	USA	43.08	-96.18	4	3(n), 21(n), 22(n), 23(n)	seeds
USDA	PI 650100	USA	42.91	-96.95	5	3(n), 4(n), 6(n), 7(n), 8(n)	seeds
USDA	PI 650101	USA	42.81	-96.68	5	20(n), 21(n), 22(n), 24(n), 25(n)	seeds
USDA	PI 650102	USA	42.91	-96.95	5	20(n), 21(n), 22(n), 23(n), 24(n)	seeds
USDA	PI 650104 *	USA	NA	NA	5	20(u), 21(u), 22(u), 23(u), 25(u)	seeds
USDA	PI 650105	USA	40.46	-96.36	5	4(n), 6(n), 11(n), 13(n), 22(n)	seeds
USDA	PI 650107	USA	NA	NA	5	2(n), 3(n), 6(n), 12(n), 13(n)	seeds
USDA	PI 650108	USA	NA	NA	5	4(n), 6(n), 8(n), 10(n), 11(n)	seeds
USDA	PI 664597 *	USA	44.63	-70	5	1(u), 8(u), 9(u), 11(u), 12(u)	seeds
USDA	PI 664611	USA	34.58	-94.23	5	2(u), 5(u), 6(u), 11(u), 15(u)	seeds
USDA	PI 664616	USA	NA	NA	5	4(n), 5(n), 9(n), 14(n), 27(n)	seeds

Collection	Accession	Origin	Lat.	Long.	N	Sample IDs	Obtained from
USDA	PI 664621 *	USA	44.7	-70.06	5	21(u), 22(u), 23(u), 24(u), 25(u)	seeds
USDA	PI 664624 *	Canada	49.39	-98.69	5	11(n), 19(n), 20(n), 23(n), 26(n)	seeds
USDA	PI 664625	USA	38.43	-91.06	5	1(n), 4(n), 7(n), 8(n), 9(n)	seeds
INRA	325	France	NA	NA	1	INRA325_1(u)	seeds
INRA	326	France	NA	NA	1	INRA326_1(u)	seeds
INRA	327	France	NA	NA	1	INRA327_2(u)	seeds
INRA	328	France	NA	NA	1	INRA328_1(u)	seeds
INRA	570	USA	NA	NA	1	INRA570_1(n)	seeds
INRA	571	USA	NA	NA	1	INRA571_1(n)	seeds
INRA	572	USA	NA	NA	1	INRA572_2(n)	seeds
INRA	732	USA	NA	NA	1	INRA732_2(n)	seeds
INRA	1013	USA	NA	NA	1	INRA1013_1(n)	seeds
INRA	1234	France	NA	NA	1	INRA1234_1(u)	seeds
INRA	MPHE001361	France	NA	NA	1	TOP1(c)	tubers
INRA	MPHE001362	France	NA	NA	1	TOP2(c)	tubers
INRA	MPHE001363	France	NA	NA	1	TOP3(c)	tubers
INRA	MPHE001364	France	NA	NA	1	TOP4(c)	tubers
INRA	MPHE001365	France	NA	NA	1	TOP5(c)	tubers
INRA	MPHE001366	France	NA	NA	1	TOP6(c)	tubers
INRA	MPHE001367	France	NA	NA	1	TOP7(c)	tubers
INRA	MPHE001368	France	NA	NA	1	TOP8(c)	tubers
INRA	MPHE001369	France	NA	NA	1	TOP9(c)	tubers
INRA	MPHE001370	France	NA	NA	1	TOP10(c)	tubers
INRA	MPHE001371	France	NA	NA	1	TOP11(c)	tubers
INRA	MPHE001372	France	NA	NA	1	TOP12(c)	tubers
INRA	MPHE001373	France	NA	NA	1	TOP13(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
INRA	MPHE001374	France	NA	NA	1	TOP14(c)	tubers
INRA	MPHE001375	France	NA	NA	1	TOP15(c)	tubers
INRA	MPHE001376	France	NA	NA	1	TOP16(c)	tubers
INRA	MPHE001377	France	NA	NA	1	TOP17(c)	tubers
INRA	MPHE001378	France	NA	NA	1	TOP18(c)	tubers
INRA	MPHE001379	France	NA	NA	1	TOP19(c)	tubers
INRA	MPHE001380	France	NA	NA	1	TOP20(c)	tubers
INRA	MPHE001381	France	NA	NA	1	TOP21(c)	tubers
INRA	MPHE001382	France	NA	NA	1	TOP22(c)	tubers
INRA	MPHE001383	France	NA	NA	1	TOP23(c)	tubers
INRA	MPHE001385	France	NA	NA	1	TOP25(c)	tubers
INRA	MPHE001386	France	NA	NA	1	TOP26(c)	tubers
INRA	MPHE001387	France	NA	NA	1	TOP27(c)	tubers
INRA	MPHE001388	France	NA	NA	1	TOP28(c)	tubers
INRA	MPHE001389	France	NA	NA	1	TOP29(c)	tubers
INRA	MPHE001390	France	NA	NA	1	TOP30(c)	tubers
INRA	MPHE001391	France	NA	NA	1	TOP31(c)	tubers
INRA	MPHE001392	France	NA	NA	1	TOP32(c)	tubers
INRA	MPHE001393	France	NA	NA	1	TOP33(c)	tubers
INRA	MPHE001394	France	NA	NA	1	TOP34(c)	tubers
INRA	MPHE001395	France	NA	NA	1	TOP35(c)	tubers
INRA	MPHE001396	France	NA	NA	1	TOP36(c)	tubers
INRA	MPHE001397	France	NA	NA	1	TOP37(c)	tubers
INRA	MPHE001398	France	NA	NA	1	TOP38(c)	tubers
INRA	MPHE001399	France	NA	NA	1	TOP39(c)	tubers
INRA	MPHE001400	France	NA	NA	1	TOP40(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
INRA	MPHE001401	France	NA	NA	1	TOP41(c)	tubers
INRA	MPHE001402	France	NA	NA	1	TOP42(c)	tubers
INRA	MPHE001403	France	NA	NA	1	TOP43(c)	tubers
INRA	MPHE001404	France	NA	NA	1	TOP44(c)	tubers
INRA	MPHE001405	France	NA	NA	1	TOP45(c)	tubers
INRA	MPHE001406	France	NA	NA	1	TOP46(c)	tubers
INRA	MPHE001407	France	NA	NA	1	TOP47(c)	tubers
INRA	MPHE001408	France	NA	NA	1	TOP48(c)	tubers
INRA	MPHE001410	France	NA	NA	1	TOP50(c)	tubers
INRA	MPHE001411	France	NA	NA	1	TOP51(c)	tubers
INRA	MPHE001412	France	NA	NA	1	TOP52(c)	tubers
INRA	MPHE001413	France	NA	NA	1	TOP53(c)	tubers
INRA	MPHE001414	France	NA	NA	1	TOP54(c)	tubers
INRA	MPHE001415	France	NA	NA	1	TOP55(c)	tubers
INRA	MPHE001416	France	NA	NA	1	TOP56(c)	tubers
INRA	MPHE001417	France	NA	NA	1	TOP57(c)	tubers
INRA	MPHE001418	France	NA	NA	1	TOP58(c)	tubers
INRA	MPHE001419	USA	NA	NA	1	TOP59(c)	tubers
INRA	MPHE001421	France	NA	NA	1	TOP61(c)	tubers
INRA	MPHE001423	France	NA	NA	1	TOP63(c)	tubers
INRA	MPHE001426	France	NA	NA	1	TOP66(c)	tubers
INRA	MPHE001427	France	NA	NA	1	TOP67(c)	tubers
INRA	MPHE001432	France	NA	NA	1	TOP72(c)	tubers
INRA	MPHE001434	France	NA	NA	1	TOP74(c)	tubers
INRA	MPHE001435	France	NA	NA	1	TOP75(c)	tubers
INRA	MPHE001436	France	NA	NA	1	TOP76(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
INRA	MPHE001437	France	NA	NA	1	TOP77(c)	tubers
INRA	MPHE001439	France	NA	NA	1	TOP79(c)	tubers
INRA	MPHE001440	France	NA	NA	1	TOP80(c)	tubers
INRA	MPHE001441	France	NA	NA	1	TOP81(c)	tubers
INRA	MPHE001442	France	NA	NA	1	TOP82(c)	tubers
INRA	MPHE001443	France	NA	NA	1	TOP83(c)	tubers
INRA	MPHE001444	France	NA	NA	1	TOP84(c)	tubers
INRA	MPHE001445	France	NA	NA	1	TOP85(c)	tubers
INRA	MPHE001446	France	NA	NA	1	TOP86(c)	tubers
INRA	MPHE001447	France	NA	NA	1	TOP87(c)	tubers
INRA	MPHE001448	France	NA	NA	1	TOP88(c)	tubers
INRA	MPHE001449	France	NA	NA	1	TOP89(c)	tubers
INRA	MPHE001451	Belgium	NA	NA	1	TOP91(c)	tubers
INRA	MPHE001453	Germany	NA	NA	1	TOP93(c)	tubers
INRA	MPHE001454	Germany	NA	NA	1	TOP94(c)	tubers
INRA	MPHE001455	Germany	NA	NA	1	TOP95(c)	tubers
INRA	MPHE001456	Germany	NA	NA	1	TOP96(c)	tubers
INRA	MPHE001457	Germany	NA	NA	1	TOP97(c)	tubers
INRA	MPHE001458	Germany	NA	NA	1	TOP98(c)	tubers
INRA	MPHE001459	Germany	NA	NA	1	TOP99(c)	tubers
INRA	MPHE001461	Hungary	NA	NA	1	TOP101(c)	tubers
INRA	MPHE001462	Hungary	NA	NA	1	TOP102(c)	tubers
INRA	MPHE001463	Former Yugoslavia	NA	NA	1	TOP103(c)	tubers
INRA	MPHE001464	Germany	NA	NA	1	TOP104(c)	tubers
INRA	MPHE001466	Europe	NA	NA	1	TOP106(c)	tubers
INRA	MPHE001467	Former USSR	NA	NA	1	TOP107(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
INRA	MPHE001468	Former USSR	NA	NA	1	TOP108(c)	tubers
INRA	MPHE001469	Former USSR	NA	NA	1	TOP109(c)	tubers
INRA	MPHE001470	Former USSR	NA	NA	1	TOP110(c)	tubers
INRA	MPHE001472	Former USSR	NA	NA	1	TOP112(c)	tubers
INRA	MPHE001473	Ukraine	NA	NA	1	TOP113(c)	tubers
INRA	MPHE001474	Former USSR	NA	NA	1	TOP114(c)	tubers
INRA	MPHE001475	Former USSR	NA	NA	1	TOP115(c)	tubers
INRA	MPHE001476	Former USSR	NA	NA	1	TOP116(c)	tubers
INRA	MPHE001477	Former USSR	NA	NA	1	TOP117(c)	tubers
INRA	MPHE001478	Former USSR	NA	NA	1	TOP118(c)	tubers
INRA	MPHE001479	Former USSR	NA	NA	1	TOP119(c)	tubers
INRA	MPHE001480	Former USSR	NA	NA	1	TOP120(c)	tubers
INRA	MPHE001481	Former USSR	NA	NA	1	TOP121(c)	tubers
INRA	MPHE001482	Former USSR	NA	NA	1	TOP122(c)	tubers
INRA	MPHE001484	Former USSR	NA	NA	1	TOP124(c)	tubers
INRA	MPHE001486	Former USSR	NA	NA	1	TOP126(c)	tubers
INRA	MPHE001487	Former USSR	NA	NA	1	TOP127(c)	tubers
INRA	MPHE001488	USA	NA	NA	1	TOP128(c)	tubers
INRA	MPHE001489	USA	NA	NA	1	TOP129(c)	tubers
INRA	MPHE001490	Canada	NA	NA	1	TOP130(c)	tubers
INRA	MPHE001491	Canada	NA	NA	1	TOP131(c)	tubers
INRA	MPHE001492	Guadeloupe	NA	NA	1	TOP132(c)	tubers
INRA	MPHE001493	Former USSR	NA	NA	1	TOP133(c)	tubers
INRA	MPHE001494	Former USSR	NA	NA	1	TOP134(c)	tubers
INRA	MPHE001495	Iran	NA	NA	1	TOP135(c)	tubers
INRA	MPHE001496	France	NA	NA	1	TOP136(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b>N</b>	<b>Sample IDs</b>	<b>Obtained from</b>
INRA	MPHE001497	Morocco	NA	NA	1	TOP137(c)	tubers
INRA	MPHE001499	France	NA	NA	1	TOP139(c)	tubers
INRA	MPHE001500	France	NA	NA	1	TOP140(c)	tubers
IPK	HEL53	Germany	NA	NA	1	Hel53_2(c)	tubers
IPK	HEL54	Germany	NA	NA	1	Hel54_2(i)	tubers
IPK	HEL55	Germany	NA	NA	1	Hel55_1(i)	tubers
IPK	HEL56	Germany	NA	NA	1	Hel56_2(u)	tubers
IPK	HEL57	Germany	NA	NA	1	Hel57(i)	tubers
IPK	HEL58	Germany	NA	NA	1	Hel58_1(i)	tubers
IPK	HEL60	Former USSR	NA	NA	1	Hel60(c)	tubers
IPK	HEL61	Former USSR	NA	NA	1	Hel61_1(c)	tubers
IPK	HEL62	Former USSR	NA	NA	1	Hel62_1(c)	tubers
IPK	HEL63	Former USSR	NA	NA	1	Hel63(u)	tubers
IPK	HEL64	Former USSR	NA	NA	1	Hel64(u)	tubers
IPK	HEL65	Former USSR	NA	NA	1	Hel65(c)	tubers
IPK	HEL66	Former USSR	NA	NA	1	Hel66_1(c)	tubers
IPK	HEL67	Former USSR	NA	NA	1	Hel67(c)	tubers
IPK	HEL68	NA	NA	NA	1	Hel68(c)	tubers
IPK	HEL231	Germany	NA	NA	1	Hel231(c)	tubers
IPK	HEL243	Germany	NA	NA	1	Hel243(c)	tubers
IPK	HEL244	NA	NA	NA	1	Hel244(c)	tubers
IPK	HEL247	Germany	NA	NA	1	Hel247(c)	tubers
IPK	HEL250	France	NA	NA	1	Hel250(c)	tubers
IPK	HEL251	Former USSR	NA	NA	1	Hel251_2(c)	tubers
IPK	HEL252	Germany	NA	NA	1	Hel252_1(u)	tubers
IPK	HEL254	NA	NA	NA	1	Hel254_1(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b>N</b>	<b>Sample IDs</b>	<b>Obtained from</b>
IPK	HEL258	NA	NA	NA	1	Hel258(c)	tubers
IPK	HEL259	France	NA	NA	1	Hel259_2(c)	tubers
IPK	HEL260	Former USSR	NA	NA	1	Hel260(c)	tubers
IPK	HEL261	France	NA	NA	1	Hel261_1(c)	tubers
IPK	HEL263	Canada	NA	NA	1	Hel263_2(u)	tubers
IPK	HEL264	Hungary	NA	NA	1	Hel264(c)	tubers
IPK	HEL266	Former Yugoslavia	NA	NA	1	Hel266_2(u)	tubers
IPK	HEL270	France	NA	NA	1	Hel270_1(c)	tubers
IPK	HEL272	France	NA	NA	1	Hel272_2(c)	tubers
IPK	HEL274	France	NA	NA	1	Hel274_1(c)	tubers
IPK	HEL278	NA	NA	NA	1	Hel278_1(i)	tubers
IPK	HEL279	Germany	NA	NA	1	Hel279_2(i)	tubers
IPK	HEL280	Germany	NA	NA	1	Hel280_2(i)	tubers
IPK	HEL281	Germany	NA	NA	1	Hel281(u)	tubers
IPK	HEL282	Germany	NA	NA	1	Hel282_1(u)	tubers
IPK	HEL286	Germany	NA	NA	1	Hel286(u)	tubers
IPK	HEL287	NA	NA	NA	1	Hel287(u)	tubers
IPK	HEL288	Poland	NA	NA	1	Hel288(i)	tubers
IPK	HEL289	Poland	NA	NA	1	Hel289_1(i)	tubers
IPK	HEL291	Poland	NA	NA	1	Hel291(i)	tubers
IPK	HEL292	Poland	NA	NA	1	Hel292_2(i)	tubers
IPK	HEL293	Poland	NA	NA	1	Hel293(i)	tubers
IPK	HEL294	Poland	NA	NA	1	Hel294_1(i)	tubers
IPK	HEL297	NA	NA	NA	1	Hel297_2(i)	tubers
IPK	HEL298	NA	NA	NA	1	Hel298_1(i)	tubers
IPK	HEL309	NA	NA	NA	1	Hel309_2(c)	tubers



<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
IPK	HEL311	NA	NA	NA	1	Hel311(c)	tubers
IPK	HEL312	NA	NA	NA	1	Hel312_1(c)	tubers
IPK	HEL313	NA	NA	NA	1	Hel313_1(u)	tubers
IPK	HEL315	NA	NA	NA	1	Hel315_1(c)	tubers
IPK	HEL317	NA	NA	NA	1	Hel317_2(c)	tubers
IPK	HEL319	NA	NA	NA	1	Hel319_2(u)	tubers
IPK	HEL320	NA	NA	NA	1	Hel320(i)	tubers
IPK	HEL321	NA	NA	NA	1	Hel321_1(i)	tubers
IPK	HEL324	NA	NA	NA	1	Hel324_1(i)	tubers
IPK	HEL325	NA	NA	NA	1	Hel325_1(i)	tubers
IPK	HEL329	NA	NA	NA	1	Hel329_1(c)	tubers
IPK	HEL331	NA	NA	NA	1	Hel331_2(c)	tubers
IPK	HEL333	NA	NA	NA	1	Hel333_1(c)	tubers
IPK	HEL338	NA	NA	NA	1	Hel338_2(u)	tubers
IPK	HEL339	NA	NA	NA	1	Hel339_1(u)	tubers
IPK	HEL340	NA	NA	NA	1	Hel340(u)	tubers
IPK	HEL341	NA	NA	NA	1	Hel341_2(u)	tubers
IPK	HEL343	NA	NA	NA	1	Hel343(u)	tubers
IPK	HEL344	NA	NA	NA	1	Hel344_1(u)	tubers
IPK	HEL345	NA	NA	NA	1	Hel345(c)	tubers
IPK	HEL519	NA	NA	NA	1	Hel519_2(u)	tubers
PGRC	NC10-3	Canada	NA	NA	1	Plot1_G(c)	tubers
PGRC	NC10-4	Canada	NA	NA	1	Plot2_3(c)	tubers
PGRC	NC10-5	Canada	NA	NA	1	Plot3_G(c)	tubers
PGRC	NC10-6	Canada	NA	NA	1	Plot4_3(c)	tubers
PGRC	NC10-7	Canada	NA	NA	1	Plot5_4(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
PGRC	NC10-8	Canada	NA	NA	1	Plot6_2(c)	tubers
PGRC	NC10-9	Canada	NA	NA	1	Plot7_4(c)	tubers
PGRC	NC10-10	Canada	NA	NA	1	Plot8_3(c)	tubers
PGRC	NC10-11	Canada	NA	NA	1	Plot9_1(c)	tubers
PGRC	NC10-12	Canada	NA	NA	1	Plot10_4(c)	tubers
PGRC	NC10-13	Canada	NA	NA	1	Plot11_G(c)	tubers
PGRC	NC10-14	Canada	NA	NA	1	Plot12_3(c)	tubers
PGRC	NC10-15	Canada	NA	NA	1	Plot13_3(c)	tubers
PGRC	NC10-16	Canada	NA	NA	1	Plot14_3(c)	tubers
PGRC	NC10-17	Canada	NA	NA	1	Plot15_4(c)	tubers
PGRC	NC10-18	Canada	NA	NA	1	Plot16_3(c)	tubers
PGRC	NC10-19	Canada	NA	NA	1	JA17(c)	tubers
PGRC	NC10-20	Canada	NA	NA	1	Plot18_2(c)	tubers
PGRC	NC10-21	Canada	NA	NA	1	Plot19_4(c)	tubers
PGRC	NC10-22	Canada	NA	NA	1	Plot20_4(c)	tubers
PGRC	NC10-23	Canada	NA	NA	1	Plot21_1(c)	tubers
PGRC	NC10-24	Canada	NA	NA	1	JA22(c)	tubers
PGRC	NC10-25	Canada	NA	NA	1	Plot23_G(c)	tubers
PGRC	NC10-26	Canada	NA	NA	1	Plot24_1(c)	tubers
PGRC	NC10-27	Canada	NA	NA	1	JA25(c)	tubers
PGRC	NC10-28	Canada	NA	NA	1	Plot26_2(c)	tubers
PGRC	NC10-29	Canada	NA	NA	1	Plot27_2(c)	tubers
PGRC	NC10-30	Canada	NA	NA	1	Plot28_1(c)	tubers
PGRC	NC10-31	Canada	NA	NA	1	Plot29_2(c)	tubers
PGRC	NC10-32	Canada	NA	NA	1	Plot30_4(c)	tubers
PGRC	NC10-33	Canada	NA	NA	1	Plot31_1(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
PGRC	NC10-35	Canada	NA	NA	1	JA33(c)	tubers
PGRC	NC10-36	Canada	NA	NA	1	Plot34_1(c)	tubers
PGRC	NC10-37	Canada	NA	NA	1	Plot35_1(c)	tubers
PGRC	NC10-38	Canada	NA	NA	1	Plot36_2(c)	tubers
PGRC	NC10-40	Canada	NA	NA	1	JA37(c)	tubers
PGRC	NC10-41	Canada	NA	NA	1	Plot38_1(c)	tubers
PGRC	NC10-42	Canada	NA	NA	1	Plot39_G(c)	tubers
PGRC	NC10-43	USA	NA	NA	1	JA40(c)	tubers
PGRC	NC10-44	USA	NA	NA	1	JA41_1_2(c)	tubers
PGRC	NC10-45	Canada	NA	NA	1	Plot42_1(c)	tubers
PGRC	NC10-46	Canada	NA	NA	1	Plot43_2(c)	tubers
PGRC	NC10-48	Canada	NA	NA	1	JA44(u)	tubers
PGRC	NC10-49	Canada	NA	NA	1	Plot45_1(u)	tubers
PGRC	NC10-52	Canada	NA	NA	1	JA46(c)	tubers
PGRC	NC10-53	Canada	NA	NA	1	Plot47_2(c)	tubers
PGRC	NC10-54	Canada	NA	NA	1	JA48(u)	tubers
PGRC	NC10-55	Canada	NA	NA	1	Plot49_1(c)	tubers
PGRC	NC10-60	Canada	NA	NA	1	JA51(c)	tubers
PGRC	NC10-61	Canada	NA	NA	1	Plot52(u)	tubers
PGRC	NC10-65	USA	NA	NA	1	Plot54(u)	tubers
PGRC	NC10-67	USA	NA	NA	1	JA55_2_4(u)	tubers
PGRC	NC10-68	USA	NA	NA	1	Plot173(c)	tubers
PGRC	NC10-70	Former USSR	NA	NA	1	Plot56_2(c)	tubers
PGRC	NC10-71	Former USSR	NA	NA	1	JA57_1(c)	tubers
PGRC	NC10-72	Former USSR	NA	NA	1	JA58_1(c)	tubers
PGRC	NC10-73	Former USSR	NA	NA	1	Plot59_1(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
PGRC	NC10-74	Former USSR	NA	NA	1	JA60(c)	tubers
PGRC	NC10-75	Former USSR	NA	NA	1	Plot174(c)	tubers
PGRC	NC10-76	Former USSR	NA	NA	1	JA61_2(c)	tubers
PGRC	NC10-77	Japan	NA	NA	1	JA62(u)	tubers
PGRC	NC10-78	Japan	NA	NA	1	Plot63_2(c)	tubers
PGRC	NC10-79	Canada	NA	NA	1	Plot64(u)	tubers
PGRC	NC10-80	Canada	NA	NA	1	JA65(u)	tubers
PGRC	NC10-81	USA	NA	NA	1	JA66_1(c)	tubers
PGRC	NC10-82	USA	NA	NA	1	JA67(u)	tubers
PGRC	NC10-83	Canada	NA	NA	1	JA68(c)	tubers
PGRC	NC10-92	Canada	NA	NA	1	Plot75_1(i)	tubers
PGRC	NC10-94	Canada	NA	NA	1	JA76(u)	tubers
PGRC	NC10-95	Canada	NA	NA	1	Plot77_1(c)	tubers
PGRC	NC10-96	France	NA	NA	1	JA78(c)	tubers
PGRC	NC10-97	Canada	NA	NA	1	JA79(c)	tubers
PGRC	NC10-103	France	NA	NA	1	JA81_1(c)	tubers
PGRC	NC10-104	Former USSR	NA	NA	1	JA82_2(c)	tubers
PGRC	NC10-105	France	NA	NA	1	JA83(c)	tubers
PGRC	NC10-106	France	NA	NA	1	Plot84_1(c)	tubers
PGRC	NC10-107	Former USSR	NA	NA	1	JA85(c)	tubers
PGRC	NC10-108	France	NA	NA	1	JA86(c)	tubers
PGRC	NC10-109	France	NA	NA	1	Plot87_1(u)	tubers
PGRC	NC10-110	Former USSR	NA	NA	1	JA88(i)	tubers
PGRC	NC10-111	France	NA	NA	1	JA89_1(c)	tubers
PGRC	NC10-113	Ukraine	NA	NA	1	JA91_2(c)	tubers
PGRC	NC10-114	Former USSR	NA	NA	1	Plot92_2(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
PGRC	NC10-119	Former USSR	NA	NA	1	Plot95(c)	tubers
PGRC	NC10-121	France	NA	NA	1	Plot97_1(c)	tubers
PGRC	NC10-123	France	NA	NA	1	Plot99(c)	tubers
PGRC	NC10-129	Germany	NA	NA	1	Plot102(c)	tubers
PGRC	NC10-130	Canada	NA	NA	1	Plot103(c)	tubers
PGRC	NC10-131	Canada	NA	NA	1	Plot104(c)	tubers
PGRC	NC10-140	Former USSR	NA	NA	1	Plot105(c)	tubers
PGRC	NC10-145	Canada	NA	NA	1	Plot108_2(c)	tubers
PGRC	NC10-146	Canada	NA	NA	1	Plot109_2(c)	tubers
PGRC	NC10-150	Canada	NA	NA	1	Plot113_1(c)	tubers
PGRC	NC10-153	Canada	NA	NA	1	Plot116_1(c)	tubers
PGRC	NC10-154	Canada	NA	NA	1	Plot117_2(c)	tubers
PGRC	NC10-155	Canada	NA	NA	1	Plot118_2(c)	tubers
PGRC	NC10-156	Canada	NA	NA	1	Plot119_1(c)	tubers
PGRC	NC10-157	Canada	NA	NA	1	Plot120_2(c)	tubers
PGRC	NC10-159	Canada	NA	NA	1	Plot122(c)	tubers
PGRC	NC10-160	Canada	NA	NA	1	Plot123_3(c)	tubers
PGRC	NC10-162	Canada	NA	NA	1	Plot125_1(c)	tubers
PGRC	NC10-163	Canada	NA	NA	1	Plot126_3(c)	tubers
PGRC	NC10-164	Canada	NA	NA	1	Plot127_3(c)	tubers
PGRC	NC10-166	Canada	NA	NA	1	Plot129_1(c)	tubers
PGRC	NC10-167	Canada	NA	NA	1	Plot130_1(c)	tubers
PGRC	NC10-168	Canada	NA	NA	1	Plot131_3(c)	tubers
PGRC	NC10-169	Canada	NA	NA	1	Plot132_2(c)	tubers
PGRC	NC10-171	Canada	NA	NA	1	Plot133_2(c)	tubers
PGRC	NC10-172	Canada	NA	NA	1	Plot134_3(c)	tubers

<b>Collection</b>	<b>Accession</b>	<b>Origin</b>	<b>Lat.</b>	<b>Long.</b>	<b><i>N</i></b>	<b>Sample IDs</b>	<b>Obtained from</b>
PGRC	NC10-173	Canada	NA	NA	1	Plot135_3(c)	tubers
PGRC	NC10-174	Canada	NA	NA	1	Plot136_1(c)	tubers
PGRC	NC10-175	Canada	NA	NA	1	Plot137_3(c)	tubers
PGRC	NC10-176	Canada	NA	NA	1	Plot138_2(c)	tubers
PGRC	NC10-177	Canada	NA	NA	1	Plot139_3(c)	tubers
PGRC	NC10-179	Canada	NA	NA	1	Plot141_2(c)	tubers
PGRC	NC10-181	Canada	NA	NA	1	Plot143(c)	tubers
PGRC	NC10-182	Canada	NA	NA	1	Plot144_2(c)	tubers
PGRC	NC10-183	Canada	NA	NA	1	Plot145_3(c)	tubers
PGRC	NC10-189	Canada	NA	NA	1	Plot151_3(c)	tubers
PGRC	NC10-190	Canada	NA	NA	1	Plot152_4(c)	tubers
PGRC	NC10-192	Canada	NA	NA	1	Plot153_3(c)	tubers
PGRC	NC10-193	Canada	NA	NA	1	Plot154_1(c)	tubers
PGRC	NC10-194	Canada	NA	NA	1	Plot155_3(c)	tubers
PGRC	NC10-196	Canada	NA	NA	1	Plot157_1(c)	tubers
PGRC	NC10-198	Canada	NA	NA	1	Plot158(c)	tubers
PGRC	NC10-199	Canada	NA	NA	1	Plot159_1(c)	tubers
PGRC	NC10-202	Canada	NA	NA	1	Plot161_1(c)	tubers
PGRC	NC10-204	Canada	NA	NA	1	Plot163_4(c)	tubers
PGRC	NC10-205	Canada	NA	NA	1	Plot164_3(c)	tubers
PGRC	NC10-208	Canada	NA	NA	1	Plot175(c)	tubers
PGRC	NC10-209	Canada	NA	NA	1	Plot167_3(c)	tubers
PGRC	NC10-210	Canada	NA	NA	1	Plot168_2(c)	tubers
PGRC	NC10-212	Canada	NA	NA	1	Plot170_1(c)	tubers
PGRC	NC10-213	Canada	NA	NA	1	Plot171(c)	tubers
PGRC	SR	NA	NA	NA	1	Plot172_1(c)	tubers

Collection	Accession	Origin	Lat.	Long.	<i>N</i>	Sample IDs	Obtained from
PGRC	Group 2	NA	NA	NA	1	JA177_1(c)	tubers
PGRC	Group 3	NA	NA	NA	1	Plot178(u)	tubers
PGRC	Group 4	USA	NA	NA	1	JA179_2(c)	tubers
Rieseberg Lab	STP	Romania	47.04	27.74	5	7_1(i), 18_1(i), 27_1(i), 39_2(i), 50_2(i)	tubers
Rieseberg Lab	TGS	Romania	45.99	26.16	6	6_1(i), 17_1(i), 20_1(i), 25_1(i), 40_1(i), 56(i)	tubers
Rieseberg Lab	ILI	Romania	45.81	25.78	5	8_1(i), 12_1(i), 27_1(i), 36(i), 60_1(i)	tubers
Rieseberg Lab	BAR	Romania	46.07	25.61	5	4(i), 16_3(i), 47_2(i), 50(i), 52_1(i)	tubers
Rieseberg Lab	ODO	Romania	46.25	25.23	5	1_2(i), 10_2(i), 21_2D(i), 39_2(i), 59(i)	tubers
Rieseberg Lab	BAL1	Romania	46.39	24.73	6	13_1(i), 14_1(i), 26(i), 40_2(i), 47_1(i), 60_1(i)	tubers
Rieseberg Lab	BAL2	Romania	46.39	24.69	5	1(i), 5(i), 9_1(i), 13_1(i), 18(i)	seeds
Rieseberg Lab	MUR	Romania	46.47	24.14	7	10_1(i), 20_1(i), 28_1(i), 37_1(i), 54(i), 56(i), 57_1(i)	tubers
Rieseberg Lab	RAS	Romania	46.90	23.78	5	9_1(i), 25_1(i), 46_1(i), 54_1(i), 59_1(i)	tubers
Rieseberg Lab	MAN	Romania	47.12	23.91	6	8_1(i), 26_1(i), 29_1(i), 41_1(i), 53_1(i), 59_1(i)	tubers
Rieseberg Lab	BEC	Romania	47.17	24.16	5	7_1(i), 23_1(i), 39_1(i), 49_1(i), 55(i)	tubers
Rieseberg Lab	LAP	Romania	47.62	23.48	5	4_1(i), 16_1(i), 23_1(i), 38_1(i), 60_1(i)	tubers
Rieseberg Lab	ILE	Romania	47.36	23.54	5	8_2(i), 21_2(i), 39_1(i), 49(i), 55(i)	tubers
Rieseberg Lab	IVA	Czech Republic	49.10	16.37	5	7(i), 16(i), 29(i), 38(i), 58(i)	tubers
Rieseberg Lab	KOS	Czech Republic	49.06	17.41	5	15_1(i), 27_1(i), 37_1(i), 47_1(i), 57_1(i)	tubers
Rieseberg Lab	VEL	Slovakia	48.78	18.68	5	3_1(i), 14_1(i), 28_1(i), 40_1(i), 59_1(i)	tubers
Rieseberg Lab	BRO	Slovakia	48.61	18.34	7	13_1(i), 25_1(i), 36_1(i), 39_1(i), 48_1(i), 54_1(i), 58_1(i)	tubers
Rieseberg Lab	JEL	Slovakia	48.38	18.07	6	2_1(i), 11(i), 19_1(i), 38_1(i), 50_1(i), 60_1(i)	tubers
Rieseberg Lab	SZI	Hungary	47.82	19.53	6	7(i), 23_1(i), 28_1(i), 40_1(i), 54_1(i), 60_1(i)	tubers
Rieseberg Lab	PAS	Hungary	47.91	19.68	6	10_1(i), 25_1(i), 38_1(i), 41_1(i), 48_2(i), 58_1(i)	tubers
Rieseberg Lab	BUK	Hungary	46.08	17.98	2	19(i), 20(i)	tubers

**Table B.2:** Sampling information for *H. grosseserratus*, *H. divaricatus*, and *H. hirsutus* accessions. The four *H. divaricatus* accessions that we reclassified to *H. hirsutus* following ploidy inference are indicated in bold.

Species	Ploidy	Collection	Accession	Lat.	Long.	N
<i>H. grosseserratus</i>	2x	USDA	PI 435697	37.16	-94.85	5
<i>H. grosseserratus</i>	2x	USDA	PI 468725	34.91	-95.3	5
<i>H. grosseserratus</i>	2x	USDA	PI 468727	35.4	-94.58	5
<i>H. grosseserratus</i>	2x	USDA	PI 468726	33.46	-89.71	5
<i>H. grosseserratus</i>	2x	USDA	PI 649998	46.65	-97.23	5
<i>H. grosseserratus</i>	2x	USDA	PI 649995	43.3	-97.11	5
<i>H. grosseserratus</i>	2x	USDA	PI 547185	43.58	-88.93	5
<i>H. grosseserratus</i>	2x	USDA	PI 547186	42.95	-89.78	5
<i>H. grosseserratus</i>	2x	USDA	PI 547188	42.51	-89.68	5
<i>H. grosseserratus</i>	2x	USDA	PI 649997	40.46	-96.36	5
<i>H. grosseserratus</i>	2x	USDA	PI 586890	42.16	-100.38	5
<i>H. grosseserratus</i>	2x	USDA	PI 547193	40.75	-86.93	5
<i>H. grosseserratus</i>	2x	USDA	PI 547195	45.2	-85.16	5
<i>H. grosseserratus</i>	2x	USDA	PI 649992	42.73	-96.23	5
<i>H. grosseserratus</i>	2x	USDA	PI 547202	41.68	-93.13	5
<i>H. grosseserratus</i>	2x	USDA	PI 547197	38.91	-87.7	5
<i>H. grosseserratus</i>	2x	USDA	PI 547200	39.98	-90.83	5
<i>H. grosseserratus</i>	2x	USDA	PI 547201	40.05	-91.15	5
<i>H. grosseserratus</i>	2x	USDA	PI 547192	40.73	-88.76	5
<i>H. hirsutus</i>	4x	USDA	PI 468739	34.7	-94.53	5
<i>H. hirsutus</i>	4x	USDA	PI 495610	40.1	-95.21	5
<i>H. hirsutus</i>	4x	USDA	PI 547204	40.03	-90.9	5
<i>H. hirsutus</i>	4x	USDA	PI 468735	43.06	-89.43	5
<b><i>H. hirsutus</i></b>	<b>4x</b>	<b>USDA</b>	<b>PI 547173</b>	<b>38.91</b>	<b>-87.7</b>	<b>5</b>
<b><i>H. hirsutus</i></b>	<b>4x</b>	<b>USDA</b>	<b>PI 547174</b>	<b>39.18</b>	<b>-88.8</b>	<b>5</b>
<b><i>H. hirsutus</i></b>	<b>4x</b>	<b>USDA</b>	<b>PI 547171</b>	<b>40.06</b>	<b>-81.38</b>	<b>5</b>
<b><i>H. hirsutus</i></b>	<b>4x</b>	<b>USDA</b>	<b>PI 664604</b>	<b>43.06</b>	<b>-89.43</b>	<b>5</b>
<i>H. divaricatus</i>	2x	USDA	PI 435675	34.96	-94.71	5
<i>H. divaricatus</i>	2x	USDA	PI 649973	38.66	-78.15	5
<i>H. divaricatus</i>	2x	USDA	PI 503214	41	-73	5
<i>H. divaricatus</i>	2x	USDA	PI 503216	41	-71	5
<i>H. divaricatus</i>	2x	USDA	PI 503215	41	-73	5



<b>Species</b>	<b>Ploidy</b>	<b>Collection</b>	<b>Accession</b>	<b>Lat.</b>	<b>Long.</b>	<b>N</b>
<i>H. divaricatus</i>	2x	USDA	PI 503218	40	-77	5
<i>H. divaricatus</i>	2x	USDA	PI 664645	38.81	-83.53	5
<i>H. divaricatus</i>	2x	USDA	PI 468709	35.81	-93.75	5

**Table B.3:** Traits measured, with abbreviations used, details of scoring, and corresponding experiment.

<b>Trait</b>	<b>Abbreviation</b>	<b>Description</b>	<b>Experiment</b>
Allelopathy – total germination	TOT_GERM	Percentage of germinated tomato seeds on day 7 of the bioassay	Greenhouse
Allelopathy – speed of germination	GERM_SPD	See Chiapusio et al. 1997 for index formula	Greenhouse
Allelopathy – speed of accumulated germination	SPD_ACM_GERM	See Chiapusio et al. 1997 for index formula	Greenhouse
Allelopathy – coefficient of the rate of germination	CR_GERM	See Chiapusio et al. 1997 for index formula	Greenhouse
Allelopathy – radicle length	RAD_LGTH	Length of tomato seedling radicle measured in cm at day 7 of the bioassay	Greenhouse
Allelopathy – plumule length	PLU_LGTH	Length of tomato seedling plumule measured in cm at day 7 of the bioassay	Greenhouse
Flowering time	FLO_DAY	Date of anthesis of the first flower, recorded as Julian date	Common garden
Self incompatibility	SI	Presence or absence of seeds at the end of the growing season for inflorescences that were bagged prior to anthesis	Common garden
Stem diameter	STEM_DIA	Stem diameter measured at ground level in cm, recorded with a digital caliper at harvest	Common garden
Plant height	HEIGHT	Plant height measured in cm along the main stem from ground level to the apex	Common garden
Number of branches	BRANCH_NO	Number of branches longer than 2 cm in length, counted at harvest	Common garden
Number of inflorescences	FLO_NO	Total number of inflorescences estimated at the time of harvest	Common garden
Disk diameter	DISK_DIA	The inflorescence diameter measured in mm and averaged for five randomly selected inflorescences per plant. Prior to measurement, the plant material was dried for 3 days in an incubator at 37° C.	Common garden
Number of tubers	TUB_NO	Total number of tubers per plant counted at harvest	Common garden
Total tuber weight	TOT_TUB_W	Total weight (in g) of the tuber yield per plant	Common garden
Average tuber weight	AV_TUB_W	Ratio of the total tuber yield weight to the number of tubers	Common garden

<b>Trait</b>	<b>Abbreviation</b>	<b>Description</b>	<b>Experiment</b>
Tuber area	TUB_AREA	The average of the area of five randomly selected tubers that were scanned. Measurements are given by Tomato Analyzer.	Common garden
Tuber perimeter	TUB_PERIM	The average of the perimeter of the five randomly selected tubers that were scanned. Measurements are given by Tomato Analyzer.	Common garden
Tuber curved length	TUB_LGTH	The average of the maximal curved length of the five randomly selected tubers that were scanned. Measurements are given by Tomato Analyzer.	Common garden
Tuber fruit-shape index	TUB_INDX	The ratio between the maximal tuber curved length and maximal tuber width, averaged for five randomly selected tubers that were scanned. Measurements are given by Tomato Analyzer.	Common garden

**Table B.4:** Results of fixed-ANOVAs performed for the between-species comparisons (including the controls) for the six allelopathy metrics. Trait abbreviations are as per Table B.3. Post-hoc comparisons between species [*H. tuberosus* (T), *H. grosseserratus* (G), *H. divaricatus* (D), *H. hirsutus* (H)] were performed in the *lsmeans* package. *P* values are corrected for multiple comparisons using the sequential Bonferroni method. Bold font indicates significance.

ANOVA						<i>P</i> (post-hoc tests)					
Trait	Source of variation	df	Mean square	<i>F</i>	<i>P</i>	T-D	T-H	T-G	D-H	D-G	H-G
TOT_GERM	Species Residuals	4 255	18 0.973	18.499	<b>2.31 x 10<sup>-13</sup></b>	<b>5 x 10<sup>-7</sup></b>	<b>8.5 x 10<sup>-6</sup></b>	1	1	<b>5 x 10<sup>-7</sup></b>	<b>8.5 x 10<sup>-6</sup></b>
GERM_SPD	Species Residuals	4 255	21.2722 0.6809	31.243	< <b>2.2 x 10<sup>-16</sup></b>	<b>1.7 x 10<sup>-5</sup></b>	<b>5.5 x 10<sup>-5</sup></b>	<b>2.4 x 10<sup>-2</sup></b>	0.755	<b>1.5 x 10<sup>-8</sup></b>	<b>6.3 x 10<sup>-8</sup></b>
SPD_ACM_GERM	Species Residuals	4 255	22.3399 0.6641	33.637	< <b>2.2 x 10<sup>-16</sup></b>	<b>0.00016</b>	<b>0.00067</b>	<b>0.00016</b>	0.67733	<b>6.6 x 10<sup>-10</sup></b>	<b>6.1 x 10<sup>-9</sup></b>
CR_GERM	Species Residuals	4 255	23.684 0.643	36.831	< <b>2.2 x 10<sup>-16</sup></b>	0.87	1	<b>1.8 x 10<sup>-12</sup></b>	1	<b>1 x 10<sup>-8</sup></b>	<b>1.8 x 10<sup>-7</sup></b>
RAD_LGTH	Species Residuals	4 212	15.1691 0.7313	20.742	<b>1.92 x 10<sup>-14</sup></b>	<b>0.0174</b>	0.05138	0.0514	0.62483	<b>8.1 x 10<sup>-5</sup></b>	<b>8.3 x 10<sup>-6</sup></b>
PLU_LGTH	Species Residuals	4 212	11.3877 0.8027	14.187	<b>2.81 x 10<sup>-10</sup></b>	0.0828	0.1222	<b>0.0001</b>	0.786	<b>2.3 x 10<sup>-6</sup></b>	<b>8.3 x 10<sup>-6</sup></b>

**Table B.5:** Results of fixed-ANOVAs (with and without corrections for population structure) for comparisons at the six allelopathy metrics performed within *H. tuberosus*, including the native, invasive and cultivated sample categories. Trait abbreviations are as per Table B.3. Significant *P* values are given in bold font.

Trait	Analysis	Source of variation	df	Mean square	<i>F</i>	<i>P</i>
TOT_GERM	No PCA	Category	2	0.61619	0.5581	0.5738
		Residuals	113	1.10402		
	With PCA	PCA1	1	2.67167	2.51	0.1156
		PCA2	1	0.00790	0.0074	0.9314
Category Residuals		2 107	0.02589 1.06144	0.0244	0.9759	
GERM_SPD	No PCA	Category	2	0.56596	0.563	0.5711
		Residuals	113	1.00527		
	With PCA	PCA1	1	2.22502	2.2607	0.1356
		PCA2	1	0.07575	0.0770	0.7820
Category Residuals		2 107	0.10314 0.98423	0.1048	0.9006	
SPD_ACM_GERM	No PCA	Category	2	0.65827	0.6559	0.5209
		Residuals	113	1.00364		
	With PCA	PCA1	1	2.39599	2.3877	0.1252
		PCA2	1	0.11236	0.1120	0.7386
Category Residuals		2 107	0.21954 1.00346	0.2188	0.8039	
CR_GERM	No PCA	Category	2	0.55448	0.5515	0.5776
		Residuals	113	1.00545		
	With PCA	PCA1	1	0.72866	0.7148	0.3997
		PCA2	1	0.20581	0.2019	0.6541
Category Residuals		2 107	0.16305 1.01937	0.1600	0.8524	
RAD_LGTH	No PCA	Category	2	1.77272	1.8092	0.1697
		Residuals	90	0.97985		
	With PCA	PCA1	1	5.1174	5.2325	<b>0.02465</b>
		PCA2	1	0.0957	0.0978	0.75524
Category Residuals		2 85	0.1283 0.9780	0.1312	0.87719	
PLU_LGTH	No PCA	Category	2	0.32384	0.32	0.727
		Residuals	90	1.01205		
	With PCA	PCA1	1	0.44091	0.4213	0.5180
		PCA2	1	0.00139	0.0013	0.9710
Category Residuals		2 85	0.63248 1.04658	0.6043	0.5488	

**Table B.6:** Results of mixed-ANOVAs with and without corrections for population structure for 13 common garden traits. Post-hoc comparisons were performed for tests with significant Category term [native (N), invasive (I) and cultivated (C)] in the lsmeans package. *P* values are corrected for multiple comparisons using the sequential Bonferroni method. Bold font indicates significance.

ANOVA						<i>P</i> (post-hoc tests)		
Trait	Analysis	Source of variation	df	<i>F</i>	<i>P</i>	N-I	C-I	N-C
FLO_DAY	No PCA	Category	2	5.177441	<b>0.0065</b>	<b>0.0304</b>	0.4658	<b>0.0304</b>
	With PCA	PCA1	1	1.44661	0.2307	-	-	-
		PCA2	1	35.81015	<b>&lt; 0.0001</b>			
		Category	2	2.12428	0.1226			
STEM_DIA	No PCA	Category	2	19.304288	<b>&lt; 0.0001</b>	<b>0.0001</b>	0.4542	<b>0.0001</b>
	With PCA	PCA1	1	33.35873	<b>&lt; 0.0001</b>	<b>0.0012</b>	<b>0.0374</b>	0.1923
		PCA2	1	0.90449	0.3427			
		Category	2	7.14377	<b>0.0010</b>			
HEIGHT	No PCA	Category	2	3.019961	0.0509	-	-	-
	With PCA	PCA1	1	7.409922	<b>0.0070</b>	-	-	-
		PCA2	1	3.089572	0.0803			
		Category	2	2.100828	0.1249			
BRANCH_NO	No PCA	Category	2	12.376015	<b>&lt; 0.0001</b>	<b>0.0001</b>	<b>0.001</b>	0.0944
	With PCA	PCA1	1	0.02847	0.8662	<b>0.0118</b>	0.8376	<b>0.0294</b>
		PCA2	1	32.54575	<b>&lt; 0.0001</b>			
		Category	2	4.62385	<b>0.0108</b>			
FLO_NO	No PCA	Category	2	12.962011	<b>&lt; 0.0001</b>	<b>0.0001</b>	<b>0.0005</b>	0.1239
	With PCA	PCA1	1	0.47518	0.4914	<b>0.019</b>	0.7472	<b>0.019</b>
		PCA2	1	45.81762	<b>&lt; 0.0001</b>			
		Category	2	4.71907	<b>0.0099</b>			
DISK_DIA	No PCA	Category	2	2.8759434	0.0592	-	-	-
	With PCA	PCA1	1	16.141779	<b>0.0001</b>	-	-	-
		PCA2	1	0.333100	0.5646			
		Category	2	2.941007	0.0557			

ANOVA						P (post-hoc tests)		
Trait	Analysis	Source of variation	df	F	P	N-I	C-I	N-C
TUB_NO	No PCA	Category	2	24.589973	< 0.0001	0.0001	0.0001	0.0125
	With PCA	PCA1	1	15.80120	0.0001	0.0119	0.0119	0.9199
		PCA2 Category	1 2	38.82128 6.22599	< 0.0001 0.0023			
TOT_TUB_W	No PCA	Category	2	44.55717	< 0.0001	0.0001	0.9435	0.0001
	With PCA	PCA1	1	65.97273	< 0.0001	0.0004	0.3445	0.0157
		PCA2 Category	1 2	10.25421 7.66466	0.0016 0.0006			
AV_TUB_W	No PCA	Category	2	84.05633	< 0.0001	0.1179	0.0001	0.0001
	With PCA	PCA1	1	211.09275	< 0.0001	0.0776	0.0776	0.0024
		PCA2 Category	1 2	5.02684 5.77744	0.0260 0.0036			
TUB_AREA	No PCA	Category	2	45.26946	< 0.0001	0.4996	0.0001	0.0001
	With PCA	PCA1	1	117.43387	< 0.0001	-	-	-
		PCA2 Category	1 2	8.22568 2.92630	0.0045 0.0557			
TUB_PERIM	No PCA	Category	2	14.340856	< 0.0001	0.0098	0.47	0.0001
	With PCA	PCA1	1	14.860331	0.0002	0.5404	0.0362	0.2336
		PCA2 Category	1 2	31.549194 3.233743	< 0.0001 0.0413			
TUB_LGTH	No PCA	Category	2	49.82342	< 0.0001	0.0015	0.0015	0.0001
	With PCA	PCA1	1	84.66548	< 0.0001	0.7285	0.011	0.0651
		PCA2 Category	1 2	24.02577 4.53715	< 0.0001 0.0117			
TUB_INDX	No PCA	Category	2	171.82907	< 0.0001	0.0001	0.0001	0.0001
	With PCA	PCA1	1	339.3821	< 0.0001	0.2286	0.0005	0.0001
		PCA2 Category	1 2	22.5825 10.3459	< 0.0001 0.0001			

**Table B.7:** Results of MCMCglmm with and without correction for population structure. Bold font indicates significance.

Trait	Analysis	Group compared to Invasive	Posterior mean	Lower 95% CI	Upper 95% CI	Effective sample size	pMCMC
STEM_DIA	No PCA	Native	-1.49419	-3.05926	0.05364	833348	0.0541
		Cultivated	-1.0181	-2.66106	0.59532	813857	0.2159
	With PCA	Native	37.2751	10.4119	63.1968	3957.3	<b>0.004147</b>
		Cultivated	-5.1337	-9.8867	-0.78	1456.2	<b>0.007259</b>
HEIGHT	No PCA	Native	-0.44046	-1.85293	0.94807	844568	0.5381
		Cultivated	0.20101	-1.1217	1.54896	861552	0.7722
	With PCA	Native	-0.4589	-20.1066	19.3297	5987.8	0.951019
		Cultivated	2.047	-0.6917	5.0689	2801.4	0.126328
FLO_NO	No PCA	Native	-1.43803	-2.66409	-0.25291	600962	<b>0.0107</b>
		Cultivated	-0.68037	-1.99633	0.61251	571615	0.3046
	With PCA	Native	-29.1688	-48.4337	-9.8254	3711.1	<b>0.001276</b>
		Cultivated	-1.7869	-5.5018	1.711	1944.4	0.318868
TUB_NO	No PCA	Native	-2.80355	-4.29123	-1.36415	384279	<b>10<sup>-5</sup></b>
		Cultivated	-2.95858	-4.705	-1.32166	406258	<b>5 x 10<sup>-5</sup></b>
	With PCA	Native	-72.0814	-95.5221	-48.6813	3114.1	<b>6 x 10<sup>-7</sup></b>
		Cultivated	-4.6501	-9.3997	-0.6514	1209.6	<b>0.003491</b>
AV_TUB_W	No PCA	Native	-2.05067	-4.21686	0.07743	724950	0.0511
		Cultivated	-0.81831	-3.40124	1.78615	729867	0.5364
	With PCA	Native	-53.312	-93.6668	-12.3658	4239.3	<b>0.010326</b>
		Cultivated	2.2288	-5.4254	9.8933	1440	0.560983
TUB_AREA	No PCA	Native	-0.49171	-2.24931	1.24025	717038	0.579
		Cultivated	2.12503	0.08567	4.21291	682716	<b>0.0376</b>
	With PCA	Native	-46.6235	-67.7166	-25.6383	7365.4	<b>6 x 10<sup>-7</sup></b>
		Cultivated	1.0617	-3.7231	5.9882	2177.8	0.669441



Trait	Analysis	Group compared to Invasive	Posterior mean	Lower 95% CI	Upper 95% CI	Effective sample size	pMCMC
TUB_INDX	No PCA	Native	4.80108	2.65694	7.02411	341064	<b>6 x 10<sup>-7</sup></b>
		Cultivated	-5.03538	-7.32956	-2.81397	397521	<b>6 x 10<sup>-7</sup></b>
	With PCA	Native	151.3398	109.1753	188.1102	1609	<b>6 x 10<sup>-7</sup></b>
		Cultivated	-6.1136	-13.3484	0.45	1079	<b>0.048</b>

**Table B.8:** Hedges' effect size estimates calculated for one invasiveness trait (tuber number) and three domestication traits (total tuber weight, average tuber weight, tuber fruit-shape index) in *H. tuberosus*, as well as for 18 invasiveness traits identified from other studies. *N1* and *N2* are the sample sizes for groups 1 and 2.

Reference	Fig. 2 ID	Species	Trait	Group 1	Group 2	<i>N1</i>	<i>N2</i>	Hedges' <i>g</i> (95% CI)
this study	TUB	<i>H. tuberosus</i>	Number of tubers	Invasive (Europe)	Native (North America)	25	134	1.23 (+/-0.45)
			Total tuber weight	Cultivated	Native (North America)	66	134	1.22 (+/-0.31)
			Average tuber weight					1.87 (+/-0.34)
			Tuber fruit-shape index					2.72 (+/-0.4)
Flory et al. (2011)	a	<i>Microstegium vimineum</i>	Survival Tiller length Biomass	Invasive (North America)	Native (China)	10	10	1.12 (+/-0.98) 1.24 (+/-0.99) 2.14 (+/-1.15)
Blair & Wolfe (2004)	b	<i>Silene latifolia</i>	Flower production Plant size at flowering First flower day Plant size at 1 month Germination	Invasive (North America)	Native (Europe)	20	20	0.64 (+/-0.64) 0.75 (+/-0.65) 0.9 (+/-0.66) 1.25 (+/-0.69) 2.05 (+/-0.78)
Barney et al. (2009)	c	<i>Artemisia vulgaris</i>	Seedling emergence	Invasive (North America)	Native (Europe)	12	15	1.62 (+/-0.9)
Gusewell et al. (2006)	d	<i>Solidago gigantea</i>	Shoot number Biomass per plant	Invasive (Europe)	Native (North America)	20	22	1.13 (+/-0.66) 1.23 (+/-0.67)
Joshi & Vrieling (2005)	e	<i>Senecio jacobaea</i>	Leaves per plant Rosettes per plant Root biomass Diameter of root crown Reproductive biomass	Invasive (North America, Australia, New Zealand)	Native (Europe)	16	13	0.87 (+/-0.78) 1.03 (+/-0.79) 1.27 (+/-0.82) 1.57 (+/-0.85) 1.74 (+/-0.88)
Leger & Rice (2003)	f	<i>Eschscholzia californica</i>	Shoot mass Number of seed capsules	Invasive (Chile)	Native (North America)	80	80	0.38 (+/-0.31) 0.44 (+/-0.31)

**Table B.9:** Results from the K + Q association mapping models. Results from K and K + P models were consistent (data not shown). For tuber number, P values are given for models with and without genome-wide heterozygosity as an additional covariate ( $P_{no\_het} / P_{het}$ ).

Trait	Marker effect	Chromosome	Position	P value
PLU_LGTH	additive	2	181968628	$8.7 \times 10^{-7}$
FLO_DAY	2-dom-alt	5	261437060	$2.57 \times 10^{-6}$
	2-dom-alt	17	113735999	$9.12 \times 10^{-7}$
	3-dom-alt	5	236363257	$4.36 \times 10^{-6}$
STEM_DIA	additive	2	104053555	$4.89 \times 10^{-8}$
	2-dom-alt	17	93955538	$2.63 \times 10^{-6}$
	3-dom-alt	7	51615318	$7.07 \times 10^{-6}$
	3-dom-alt	11	122058172	$3.23 \times 10^{-6}$
	3-dom-alt	17	31015890	$6.6 \times 10^{-8}$
	3-dom-alt	2	162138168	$4.78 \times 10^{-7}$
	3-dom-alt	9	37411999	$6.3 \times 10^{-7}$
	3-dom-alt	13	226766305	$2.95 \times 10^{-5}$
HEIGHT	1-dom-alt	5	8000822	$8.31 \times 10^{-7}$
	2-dom-alt	3	163099831	$1.9 \times 10^{-6}$
	2-dom-alt	11	33713608	$1.81 \times 10^{-6}$
	2-dom-ref	3	163099834	$1.9 \times 10^{-6}$
	3-dom-ref	3	163099889	$1.73 \times 10^{-6}$
	3-dom-ref	6	98762091	$1.62 \times 10^{-6}$
	3-dom-ref	7	66987988	$7.24 \times 10^{-6}$
BRANCH_NO	2-dom-alt	11	149087831	$5.37 \times 10^{-7}$
FLO_NO	diplo-additive	14	228189510	$1.17 \times 10^{-6}$
TUB_NO	additive	9	151578474	$1.2 \times 10^{-6} / 5.62 \times 10^{-7}$
	additive	14	192636564	$7.94 \times 10^{-7} / 3.38 \times 10^{-7}$
TOT_TUB_W	2-dom-alt	13	41483357	$2.51 \times 10^{-6}$
	2-dom-ref	3	202011788	$8.51 \times 10^{-6}$
	3-dom-ref	9	37411999	$1.34 \times 10^{-5}$
TUB_AREA	3-dom-ref	4	4985299	$1.77 \times 10^{-6}$
	3-dom-ref	9	41143496	$2.88 \times 10^{-5}$
TUB_LGTH	1-dom-ref	16	21520736	$3.02 \times 10^{-5}$

**Table B.10:** Clonal series identified for *H. tuberosus* samples. Details regarding sample IDs are given in Table B.1. For each clonal series, bold font is used to indicate the sample included in the unique genotype set.

Clonal series	Genotypes											
1	Hel231	Hel243	Hel254_1	Hel264	Hel311	Hel312_1	Hel345	Hel53_2	Hel61_1	Hel66_1	Plot102	Plot104
	<b>Plot174</b>	TOP102	TOP104	TOP106	TOP108	TOP113	TOP114	TOP116	TOP119	TOP120	TOP122	TOP129
	TOP130	TOP131	TOP132	TOP137	TOP139	TOP140	TOP19	TOP42	TOP43	TOP46	TOP53	TOP59
	TOP91	TOP94	TOP95	TOP97								
2	BRO13_1	BRO25_1	BRO36_1	BRO39_1	BRO48_1	BRO54_1	BRO58_1	IVA16	IVA29	IVA38	IVA58	IVA7
	JEL11	JEL2_1	JEL38_1	JEL50_1	JEL60_1	KOS15_1	<b>KOS27_1</b>	KOS37_1	KOS47_1	KOS57_1	VEL14_1	VEL28_1
	VEL3_1	VEL40_1	VEL59_1									
3	BAR16_3	BAR4	BAR47_2	BAR50	BAR52_1	Hel320	Hel321_1	Hel324_1	LAP16_1	LAP23_1	LAP38_1	LAP4_1
	LAP60_1	RAS25_1	RAS46_1	RAS54_1	RAS59_1	RAS9_1	SZI23_1	<b>SZI28_1</b>	SZI40_1	SZI54_1	SZI60_1	SZI7
4	BAL13_1	BAL14_1	<b>BAL26</b>	BAL40_2	BAL47_1	Hel288	Hel325_1	Hel54_2	Hel55_1	Hel57	Hel58_1	JA88
	Plot75_1	TGS17_1	TGS20_1	TGS25_1	TGS40_1	TGS56	TGS6_1					
5	Plot8_3	Plot13_3	JA17	Plot18_2	Plot23_G	Plot31_1	JA33	Plot34_1	JA46	Plot47_2	Plot56_2	JA57_1
	JA58_1	<b>Plot59_1</b>	JA68	JA82_2	JA83	Plot84_1						
6	Hel258	Hel309_2	<b>Hel315_1</b>	Hel317_2	TOP101	TOP107	TOP109	TOP115	TOP39	TOP41	TOP44	TOP45
	TOP51	TOP52	TOP54	TOP61	TOP96							
7	Hel244	Hel329_1	Hel331_2	Hel333_1	JA37	JA51	Plot105	Plot172_1	<b>Plot38_1</b>	Plot39_G	TOP99	
8	Plot63_2	<b>JA66_1</b>	Plot92_2	Plot97_1	Plot108_2	Plot120_2	Plot133_2	Plot144_2	Plot155_3			
9	<b>JA89_1</b>	JA91_2	Plot103	Hel247	TOP1	TOP32	TOP98					
10	MUR10_1	MUR20_1	MUR28_1	<b>MUR37_1</b>	MUR54	MUR56	MUR57_1					
11	MAN8_1	<b>MAN26_1</b>	MAN29_1	MAN41_1	MAN53_1	MAN59_1						
12	<b>Hel250</b>	Hel251_2	Hel68	TOP127	TOP48							
13	JA44	Plot45_1	Plot52	Plot54	<b>JA55_2_4</b>							

Clonal series	Genotypes				
14	Plot2_3	JA79	JA81_1	Plot113_1	<b>Plot135_3</b>
15	ODO1_2	ODO10_2	ODO21_2D	ODO39_2	<b>ODO59</b>
16	PAS10_1	PAS25_1	PAS38_1	PAS41_1	<b>PAS48_2</b>
17	Plot27_2	Plot28_1	JA41_1_2	<b>Plot42_1</b>	Plot43_2
18	STP7_1	<b>STP18_1</b>	STP27_1	STP39_2	STP50_2
19	<b>Hel270_1</b>	TOP6	TOP12	TOP20	
20	<b>Hel274_1</b>	TOP3	TOP5	TOP30	
21	<b>Plot24_1</b>	JA25	Plot26_2	Plot49_1	
22	<b>Plot11_G</b>	Plot14_3	Plot29_2	JA40	
23	JA60	JA61_2	<b>Plot77_1</b>	TOP124	
24	Hel56_2	Hel252_1	<b>Hel313_1</b>		
25	<b>Hel62_1</b>	Hel261_1	TOP134		
26	Plot95	<b>Hel260</b>	TOP93		
27	Plot64	JA65	<b>Hel263_2</b>		
28	<b>Plot173</b>	Hel67	TOP117		
29	ILE8_2	ILE21_2	<b>ILE49</b>		
30	IL18_1	<b>ILI12_1</b>	ILI27_1		
31	<b>Plot5_4</b>	Plot16_3	Plot30_4		
32	TOP11	<b>TOP126</b>	TOP128		
33	TOP14	TOP31	<b>TOP118</b>		
34	BEC23_1	<b>BEC49_1</b>			
35	<b>BEC39_1</b>	BEC55			
36	<b>Hel259_2</b>	TOP47			

Clonal series	Genotypes	
37	<b>Hel272_2</b>	TOP22
38	Hel281	<b>Hel340</b>
39	<b>Hel63</b>	Hel64
40	<b>ILI36</b>	ILI60_1
41	<b>Plot171</b>	JA177_1
42	Plot175	<b>JA179_2</b>
43	JA78	<b>Plot137_3</b>
44	<b>JA85</b>	TOP112
45	<b>JA86</b>	TOP10
46	Plot116_1	<b>Plot117_2</b>
47	Plot123_3	<b>Plot125_1</b>
48	<b>Plot127_3</b>	Plot129_1
49	Plot12_3	<b>Plot15_4</b>
50	Plot131_3	<b>Plot154_1</b>
51	Plot161_1	<b>Plot163_4</b>
52	<b>Plot35_1</b>	Plot36_2
53	TOP17	<b>TOP18</b>
54	<b>TOP25</b>	TOP50
55	<b>TOP33</b>	TOP34
56	TOP9	<b>TOP35</b>
57	<b>TOP40</b>	TOP57
58	<b>BUK19</b>	BUK20
59	<b>Plot99</b>	TOP7

**Table B.11:** Strategy used for diploidization and tetraploidization of subgenome SNPs. We consider cases in which the alternate allele (B) or the reference allele (A) is assigned to the diploid and the tetraploid subgenome. For each scenario, we present genotypes observed in samples of the progenitor species, original (6x) and 2x- or 4x-converted genotypes of *H. tuberosus* samples, and observed frequency of each genotype in the *H. tuberosus* SNP set used for the conversion.

subgenome	2x progenitor species genotypes	4x progenitor species genotypes	original genotypes	converted genotypes	Frequency (%)
2x	AA AB BB	AAAA	AAAAAA AAAAAB AAAABB AAABBB AABBBB ABBBBB BBBBBB	AA AB BB BB BB BB BB	72.25 11.62 4.69 1.52 0.46 0.16 0.14
2x	AA AB BB	BBBB	AAAAAA AAAAAB AAAABB AAABBB AABBBB ABBBBB BBBBBB	AA AA AA AA AA AB BB	0.03 0.02 0.04 0.12 0.37 1.01 7.58
4x	AA AA AA	AAAA AAAB AABB ABBB BBBB	AAAAAA AAAAAB AAAABB AAABBB AABBBB ABBBBB BBBBBB	AAAA AAAB AABB ABBB BBBB BBBB BBBB	71.38 11.81 6.39 3.02 1.25 0.46 0.34
4x	BB BB BB	AAAA AAAB AABB ABBB BBBB	AAAAAA AAAAAB AAAABB AAABBB AABBBB ABBBBB BBBBBB	AAAA AAAA AAAA AAAB AABB ABBB BBBB	0.04 0.05 0.16 0.31 0.53 0.85 3.42