Interploid reproductive isolation in the *Campanula rotundifolia* polyploid complex

Brittany Lynne Sutherland
Lawrenceburg, KY

M.A. in Teaching, Morehead State University, 2010
B.S. in Biology, Western Kentucky University, 2005

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Biology
University of Virginia
December, 2017

ABSTRACT

Polyploidy, or whole-genome duplication, has long been considered an important factor in plant speciation due to the tendency of newly formed polyploids to be reproductively isolated from their diploid progenitors. Recent evidence has called into question both the incidence of polyploidy-induced reproductive isolation, particularly among higher ploidy levels, and the role of whole-genome duplication in the diversification of angiosperms. To better understand the role of polyploidy in plant divergence and speciation, I used the *Campanula rotundifolia* autopolyploid complex to compare reproductive isolation between diploids (2X) and tetraploids (4X), and between tetraploids and hexaploids (6X). I evaluated postzygotic isolation, prezygotic isolation due to pollinator preference and pollen choice, and interploid reproduction occurs in nature. Because the evolution and mating systems of the complex are poorly understood, I constructed chloroplast and nuclear phylogenies and performed a survey of self-compatibility. These demonstrated that both tetraploids and hexaploids had multiple origins, and that self-incompatibility was gradually lost with repeated genome duplication and as the complex expanded from Europe into North America. Reciprocal 2X-4X and 4X-6X crosses demonstrated that while diploids and tetraploids have high postzygotic isolation, tetraploids and hexaploids are capable of interbreeding. I then confirmed these results, and the fertility of newly formed tetraploids and pentaploids, in backcrosses. Artificial mixed-ploidy arrays exposed to natural pollinators demonstrated that pollinator-mediated reproductive isolation was modest in *C. rotundifolia*. These arrays further supported two different patterns of reproduction based on parental ploidy, and demonstrated that interploid reproduction is most likely for rare cytotypes. Lastly, surveys of natural mixed-ploidy

contact zones found interploid reproduction in 4X-6X contact zones, as demonstrated by extensive presence of pentaploids and pentaploid-derived aneuploids, and by lack of genetic differentiation between tetraploid and hexaploid parents. By contrast, 2X-4X contact zones contained no triploids and no evidence of gene flow aside from one spontaneous neo-tetraploid. Taken together, these studies clearly demonstrate that reproductive isolation in polyploid complexes follows two patterns: strong (but not complete) isolation between diploids and polyploids, and weak or no isolation between polyploids. The strong reproductive isolation found between diploids and tetraploids supports the “instant speciation” hypothesis, but ongoing gene flow between tetraploids and hexaploids demonstrates that this concept is not universal. These findings suggest that interploid gene flow, particularly between different polyploid cytotypes, may constrain divergence, and may provide a causal mechanism for lower observed diversification of polyploids relative to diploids.

DEDICATION

To Bryan Cottrell, in memoriam.

You were instrumental to so much of this work. I wish you could have seen it completed.

ACKNOWLEDGEMENTS

A dissertation is not the product of just one person. There are many people to whom I am forever grateful for their support, professionally or personally or both, who made this work possible. First, I would like to thank my advisor, Laura Galloway, without whom none of this would have been possible. She has been a constant source of guidance and support throughout this process, and has been far more patient than I deserve. Many times, she has kept me from getting lost in minutiae and helped me see the big picture. She has served as a role model for how to be an effective scientist, leader, and communicator, and I am the better for having worked with her these last several years.

I would also like to thank the members of my committee, who have been instrumental in planning, executing, and completing this work. Ben Blackman and Doug Taylor both helped tremendously as first readers. Ben helped in the early stages to plan experiments and steer me away from troublesome methods, and Doug has given immensely helpful feedback on analyzing some of the later experiments and on the dissertation as a whole. Janis Antonovics and Bob Cox have both been extremely helpful in getting me to think outside my intellectual bubble, bringing in broader evolutionary ideas that I would have never gotten to on my own. Dave Carr has helped keep my experiments grounded in the realm of the feasible and has been critical to both my understanding of and proper application of statistics. I have been very lucky to have you all involved in this undertaking.

I would also like to thank the members of the Galloway Lab, who have been friends and colleagues, helping me with greenhouse and fieldwork and giving valuable

feedback on much of this work. Thank you to Holly Prendeville and Karen Barnard-Kubow for helping me be less clueless in my early days, and for enduring the occasional insect bite on dangerous terrain on my behalf. Thank you to Janet Steven, Catherine Debban, Matt Koski, and Hanqin Wu for allowing me to bounce ideas off you, even when you're busy doing other things. Thank you also to those of you who joined our lab meeting group: Carolyn Beans, Melissa Aiken, Corlett Wood, Brian Sanderson, Malcolm Augat, Ray Watson, and Alyssa Bangerter. You have all been valuable sources of ideas and feedback on my projects, professional development, and science in general over the years.

None of the research presented here would have been possible without the help of several very talented and dedicated undergraduate students. First, I would like to thank Bryan Cottrell. As my first student, he helped me pilot many of the projects shown here and took on a leadership role, mentoring those students who came after him. Bria Friestad also contributed extensively to crosses and seed counts, was a wonderful field assistant, and came out of minion retirement when I had hundreds of transplants and no help. Candy Doane, Gretchen Kuhn, and Emma Kitchen all let me strain their eyesight counting thousands of miniscule seeds. Tomas Miranda-Katz helped collect pollinator data at MLBS and helped build deer exclosures for array testing (and rebuilt them when deer tore them down). Brandie Quarles was an REU at MLBS and made the selfing project her own, crossing hundreds of plants as an inaugural project in the new MLBS greenhouse. I am truly grateful to all of you!

I would like to thank several members of the Biology department for their help with research and logistics. Wendy Crannage kept my plants in far better shape than I

could have on my own, and AnhThu Nguyen was instrumental in getting my chloroplast and RAD sequencing projects completed. Thank you to Butch Brodie, Eric Nagy, and Jaime Jones for letting me put plots all over Mountain Lake and helping me keep my plants happy while at the station.

I would also like to thank Claire Cronmiller for her support in my teaching endeavors here at UVA. Working as a Genetics TA helped confirm that teaching others is an important part of what makes science fulfilling for me, and hopefully it remains a component of my professional life moving forward.

Lastly, I would like to thank those who are important to me outside of the lab. My parents, Sharon and Dave Sutherland, have been steadfast in their support and encouragement throughout my life. They fostered my early interest in science, and have served as valuable role models for what nontraditional educational paths can look like. I would also like to thank Rivka Suparna, Andee Murray, Kirsten Thorsen, Jenn and Zach Sanger, Kristen Gallagher, Jaye Crues, and Karen Chester for being my social outlet outside of UVA. It’s been a weird trip, and you all kept me grounded.

TABLE OF CONTENTS

INTRODUCTION

Flowering plants have experienced one of the most rapid and variable diversifications of all eukaryotes. Angiosperms originated relatively recently, approximately 167-199 MYA (Bell et al., 2010), but are one of the most speciose groups of eukaryotes on Earth with over 250,000 named species (Futuyma and Agrawal, 2009). Diversification rates vary widely throughout angiosperms. For example, Asteraceae is one of the largest plant families despite originating less than 50 MYA (Tank et al., 2015). Despite extensive research, the root causes of variation in rates and patterns of diversification in angiosperms are poorly understood.

One of the commonly invoked mechanisms for angiosperm diversification is whole-genome duplication, or polyploidy (Soltis et al., 2009). Since its discovery in 1907 (Lutz, 1907), polyploidy has been identified in many plant families, and shown to cause substantial changes to morphology and physiology. These changes frequently induce complete or near-complete postzygotic reproductive isolation between diploids and related tetraploids, and polyploidy is considered one of the clearest examples of sympatric speciation (Coyne and Orr, 2004). Additionally, whole-genome duplication can result in multiple prezygotic reproductive barriers. It can result in phenological changes that make polyploids flower either earlier or later than diploids (Ramsey and Schemske, 2002; Castro et al., 2010) and can cause morphological changes to floral size and structure that alter pollinator preferences (Segraves and Thompson, 1999; Husband and Schemske, 2000). Whole-genome duplication has also been shown to increase self-compatibility (Mable, 2004; Barringer, 2007). In many systems, all or most of these mechanisms act sequentially to produce essentially complete reproductive isolation

between polyploids and diploids (Husband and Schemske, 2000; Lowry et al., 2008). Polyploidy is also thought to increase angiosperm diversification based on broad phylogenetic evidence (Soltis et al., 2009) and simulations of diversification rates relative to diploidy (Scarpino and Levin, 2014).

While reproductive isolation between diploids and tetraploids in many angiosperm groups is well-studied, the causative role polyploidy plays in plant speciation is still up for debate. Recent evidence has shown that genome duplication may not universally confer complete reproductive isolation. Occasional historical gene flow has been found between diploids and tetraploids (e.g. Slotte et al., 2008; Jørgensen et al., 2011; Zohren et al., 2016). Additionally, the role of polyploidy in reproductive isolation among higher-order cytotypes has received little attention and remains poorly understood despite increasing evidence that higher-order polyploids may experience interploid gene flow differently than diploids and tetraploids (e.g. Greiner and Oberprieler, 2012; Hülber et al., 2015). These recent findings challenge the idea that changes in ploidy automatically induce reproductive isolation. Recent phylogenetic studies likewise challenge the hypothesis that polyploidization is a major factor in angiosperm diversification (Wood et al., 2009; Mayrose et al., 2011). Comparison of diversification and extinction rates has shown that polyploids may actually diversify more slowly than diploids due to increased extinction risk (Mayrose et al., 2011).

The effect of whole-genome duplication on angiosperm diversification is an emergent property driven by its proximate effects on the speciation of individual lineages over time. Therefore, we must understand how changes in ploidy limit reproduction and gene flow between ploidy levels, including among higher-order polyploids, to understand

how whole-genome duplication has affected the diversification of angiosperms. This dissertation seeks to quantify postzygotic reproductive isolation, prezygotic reproductive isolation, and gene flow between several ploidy levels within a polyploid complex. The *Campanula rotundifolia* polyploid complex, with a large geographic range and three common ploidy levels (Kovanda, 1977; Stevens et al., 2012), is an ideal system in which to quantify interploid reproductive isolation and gene flow. However, because the system has historically proven phylogenetically difficult (Witasek, 1902; Kovanda, 1977) and is suspected of having variable mating systems (Bingham and Ranker, 2000), it is first necessary to clarify the evolutionary history of the lineage and determine the degree of self-incompatibility prior to quantifying interploid reproduction and gene flow. The dissertation therefore addresses the broad question of the effect of polyploidy on reproductive isolation and divergence in six chapters. The first two provide background for the polyploid complex, and detail its phylogeography and mating system evolution, respectively. Chapters Three and Four examine possible postzygotic reproductive isolation in F1 and backcrosses between 2X-4X and 4X-6X ploidy pairs. Chapter Five assesses prezygotic isolation due to pollinator preference and postzygotic isolation in the face of mixed-pollen loads. Lastly, Chapter Six determines interploid reproduction and gene flow in 2X-4X and 4X-6X mixed-ploidy contact zones.

Campanula rotundifolia is a short-lived perennial wildflower commonly found on calcareous soils and rock outcrops (Stevens et al., 2012). It has a circumboreal distribution in the Northern hemisphere, spanning from western Russia westward through much of Europe (Böcher, 1936; Kovanda, 1966), Greenland, and the northern latitudes of North America, though alpine populations along the Appalachian and Rocky Mountains

are found as far south as North Carolina and northern Mexico, respectively (Shetler, 1982). Due to its tolerance of a wide range of habitats and variable ploidy, *C. rotundifolia* exists as a polyploid complex comprising over 40 named taxa, including species, subspecies, and varieties (Witasek, 1902; Kovanda, 1966). Although the precise number of species within the complex is not agreed upon, for the purposes of this work we consider all taxa within section Heterophylla, roughly corresponding to the large polytomy in Mansion et al. (2012), to be *C. rotundifolia* sensu lato.

Campanula rotundifolia primarily exists as three dominant cytotypes: diploid (2n = 34), tetraploid (2n = 68), and hexaploid (2n = 102). Cytotypes are nonrandomly distributed throughout the range. Diploids are primarily found in central Europe, with smaller ranges in far northern Europe and northern Greenland (Shetler, 1982). Diploids have been documented in a single location in New England (Löve and Löve, 1966), but multiple attempts to resample this populations did not locate plants in the last known location. Tetraploids are the dominant cytotype and are common throughout both Europe and North America. Hexaploids were previously reported only from the western British Isles, British Columbia, and Alaska (Shetler, 1982). However, as part of this dissertation research, hexaploids were found in Michigan and southern Ontario. In most locations, populations are cytotypically homogeneous, although the distribution of cytotypes is such that numerous instances of sympatry or parapatry between ploidy levels may be possible. Prior to this work, however, only one contact zone had been documented (Shepherd, 2007).

Chapter One addresses the taxonomic and cytotypic complexity of *C. rotundifolia* via construction of molecular phylogenies using chloroplast and nuclear DNA. Prior

taxonomic treatments that focused on *C. rotundifolia* were limited to morphological evidence, and were confounded by phenotypic plasticity within the *C. rotundifolia* complex as well as geographically biased sampling (Witasek, 1902; Böcher, 1936; Kovanda, 1977). Molecular phylogenies of the genus *Campanula* have likewise proved insufficient to resolve the *C. rotundifolia* complex; the most recent comprehensive treatment of the genus collapses *C. rotundifolia* and most allies into a single unresolved polytomy (Mansion et al., 2012). Although some ambiguities remain, the nuclear and chloroplast phylogenies I constructed resolve multiple clades in *C. rotundifolia*, and show a southern European origin with a general trend of westward and northward expansion, as well as multiple diploid to tetraploid and tetraploid to hexaploid genome duplications. This study provides baseline information on genetic distance between populations, and demonstrates that many polyploid populations within the complex arose independently.

Campanula rotundifolia has long been considered self-incompatible (Böcher, 1936), but anecdotal reports suggest that occasional selfing occurs (Bielawska, 1973; Giblin, 2005). Because selfing decreases gene flow and creates more structured populations (Wright et al., 2013), differences in self-incompatibility could affect interploid reproductive isolation. Chapter Two tests self-compatibility throughout the *C. rotundifolia* complex. Both dispersal and ploidy change have been implicated in the transition to self-compatibility, so I assessed loss of self-incompatibility across the geographic and cytotypic range of *C. rotundifolia*. Populations ranged from fully self-incompatible to self-compatible, with central European diploids and tetraploids being most self-incompatible, and North American hexaploids being least. These patterns suggest that both genome duplication and long-distance, intercontinental dispersal

contributed to loss of self-incompatibility. In North America, the marked decrease in self-incompatibility is attributable to the additive effects of both processes.

Whole genome duplication is commonly assumed to rapidly create reproductive isolation through a variety of prezygotic and postzygotic mechanisms, but postzygotic barriers were first identified between diploids and polyploids and led to the concepts of triploid block (Marks, 1966) and instant speciation (Coyne and Orr, 2004). Chapter Three assesses the strength of these postzygotic barriers in *C. rotundifolia*. Because postzygotic barriers following polyploidization are a product by errors in meiosis and gene dosage (Stoute et al., 2012; Zielinski and Scheid, 2012), they often manifest in similar ways across disparate taxa. This is in contrast to prezygotic barriers which often result from gene by environment interactions (Husband et al., 2016), and are therefore more system-specific. In multiple controlled crosses, I found that while fruit set and seed number were comparable between 2X-4X and 4X-6X crosses, reproductive isolation was lower between tetraploids and hexaploids because germination in 4X-6X crosses was much higher. In addition, this experiment yielded the first evidence that unreduced gametes and other meiotic irregularities may play a role in interploid reproduction. While 4X-6X crosses produced 5X offspring as expected, 2X-4X crosses produced a modest number of 2X, 3X, and 4X offspring, with triploids and tetraploids in almost equal measure. The unexpected tetraploids had comparable fertility to controls, and 5X offspring from 4X-6X crosses likewise showed unexpectedly high fertility (though approximately 50% lower than controls). This chapter was the first indication that reproductive isolation between polyploids may be lower than between diploids and tetraploids. It was also the first

evidence that unreduced gametes in diploids may allow for gene flow into tetraploid populations via neotetraploids.

In order for interploid reproduction to result in gene flow, interploid hybrids must be capable to reproducing with their parental cytotypes. To determine if interploid hybrids may contribute to gene flow, or are merely a genetic sink, Chapter Four reports backcrosses a subset of existing 2X, 3X, 4X, and 5X hybrids (Chapter Three) to individuals of both parental ploidies. I found that 4X hybrids of 2X-4X crosses had normal fertility and generated tetraploid offspring, while 2X hybrids showed some loss of fitness. Triploid hybrids, by contrast, had almost no seed set with either parental ploidy, and appeared to confirm triploid block as a barrier to interploid gene flow. Among the 5X hybrids, crosses to both parents showed partial fitness, but those with 6X parents performed about twice as well. Among the backcross offspring, those between 5X and their parents produced aneuploids. This experiment bolsters the idea that neotetraploids may contribute to interploid reproduction between diploids and tetraploids, and that pentaploids do maintain substantial fertility.

Prezygotic reproductive barriers are known to also have a profound effect on interploid reproduction, and in many systems, are much stronger than postzygotic barriers (Husband and Sabara, 2004). Furthermore, plants that differ in cytotype are mostly likely to encounter each other in sympatric or parapatric contact zones, where they will be exposed to mixtures of both interploid and intraploid pollen. Chapter Five assesses both the strength of prezygotic isolation due to pollinator preference as well as varying interploid and intraploid pollen composition on interploid reproduction. I exposed mixed-ploidy 2X-4X and 4X-6X arrays of varying cytotypic frequencies to pollinators at

Mountain Lake Biological Station. I found that pollinators exhibit little cytotypic preference, but differential seed set between minority and majority cytotypes suggests that modest post-pollination prezygotic isolation is occurring. The most striking outcome was the wide disparity in F1 cytotypes. 2X-4X crosses overwhelmingly formed offspring that corresponded to maternal cytotype, with less than 3% of all offspring being triploids or 2X-derived tetraploids. However, 4X-6X crosses formed approximately 40% pentaploid offspring, and that incidence varied inversely with maternal cytotype frequency. This reiterates the findings of the greenhouse experiments, and demonstrates that genetic and developmental processes, rather than pollinator choice, drive 2X-4X reproductive isolation in this system, and suggests gene flow may be common in 4X-6X contact zones.

Chapter Six assesses interploid reproduction and gene flow in mixed-ploidy contact zones, bringing together the mechanistic components of the previous three chapters and testing their effects in natural populations. I sampled leaf tissue from two 2X-4X and two 4X-6X contact zones for flow cytometric and population genetic analysis. The 2X-4X contact zones maintained distinct subpopulations of diploids and tetraploids. Not only were these individuals in most cases spatially segregated, but there was no evidence of triploid or aneuploid offspring, and diploid and tetraploid subpopulations were genetically distinct. By contrast, 4X-6X contact zones showed evidence of interploid reproduction and gene flow in multiple ways. Intermediate pentaploids were a plurality in both 4X-6X contact zones, and individuals putatively derived from backcrosses (with genetic content approximately equal to 4.5X or 5.5X) were found in both contact zones. Interestingly, this interploid reproduction appears to

favor gene flow toward the hexaploid. Furthermore, no genetic clustering was observed in 4X-6X contact zones, and more genetic variance existed within each cytotype then between them. In total, this study found that interploid reproduction and gene flow are dependent on cytotype. Interploid reproduction is rare, and therefore gene flow is limited, between diploids and tetraploids, while interploid reproduction and gene flow are common between tetraploids and hexaploids.

In total, this research demonstrates that interploid reproductive barriers are lower between tetraploids and hexaploids than between diploids and tetraploids, and that these lowered barriers result in greater interploid gene flow at higher ploidy levels. Understanding how polyploidy affects speciation and divergence in angiosperms requires knowledge of how genome duplication shapes interploid reproduction and gene flow early in divergence—on the scale of populations and incipient species. Along with other recent work, this dissertation makes plain that different cytotypes within the same polyploid complex experience reproductive isolation differently, and that divergence should not be expected to be universal across a polyploid complex.

References

Barringer, B. C. 2007. Polyploidy and self-fertilization in flowering plants. *American Journal of Botany* 94: 1527–1533.

Bell, C. D., D. E. Soltis, and P. S. Soltis. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 97: 1296–1303.

Bielawska, H. 1973. Self-fertilization in *Campanula rotundifolia* L. s. l. group. *Acta Societatis Botanicorum Poloniae* 42: 253–264.

Bingham, R. A., and T. A. Ranker. 2000. Genetic diversity in alpine and foothill populations of *Campanula rotundifolia* (*Campanula*ceae). *International Journal of Plant Sciences* 161: 403–411.

Böcher, T. W. 1936. Cytological studies on *Campanula rotundifolia*. *Hereditas* 22: 269–277.

Castro, S., Z. Münzbergová, J. Raabová, and J. Loureiro. 2010. Breeding barriers at a diploid–hexaploid contact zone in *Aster amellus*. *Evolutionary Ecology* 25: 795–814.

Coyne J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA. 545 pp.

Futuyma, D. J., and A. A. Agrawal. 2009. Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences* 106: 18054–18061.

Giblin, D. E. 2005. Variation in floral longevity between populations of *Campanula rotundifolia* (*Campanula*ceae) in response to fitness accrual rate manipulation. *American Journal of Botany* 92: 1714–1722.

Greiner, R., and C. Oberprieler. 2012. The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra- and hexaploid relatives. *Flora - Morphology, Distribution, Functional Ecology of Plants* 207: 629–635.

Husband, B. C., and D. W. Schemske. 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. *Journal of Ecology* 88: 689–701.

Husband, B. C., and H. A. Sabara. 2003. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). New Phytologist 161: 703–713

Husband, B. C., S. J. Baldwin, and H. A. Sabara. 2016. Direct vs. indirect effects of whole-genome duplication on prezygotic isolation in *Chamerion angustifolium*: Implications for rapid speciation. *American Journal of Botany* 103: 1259–1271.

Hülber, K., M. Sonnleitner, J. Suda, J. Krejčíková, P. Schönswetter, G. M. Schneeweiss, and M. Winkler. 2015. Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio carniolicus* (syn. *Jacobaea carniolica*, Asteraceae). *Ecology and Evolution* 5: 1224–1234.

Jørgensen, M. H., D. Ehrich, R. Schmickl, M. A. Koch, and A. K. Brysting. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evolutionary Biology* 11: 346.

Kovanda, M. 1977. Polyploidy and variation in the *Campanula rotundifolia* complex. Part II. (Taxonomic). *Folia Geobotanica et Phytotaxonomica* 12: 23–89.

Kovanda, M. 1966. Some chromosome counts in the *Campanula rotundifolia* complex II. *Folia Geobotanica & Phytotaxonomica Bohemoslovaca*.

Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 3009–3021.

Löve, À., and D. Löve. 1966. Cytotaxonomy of the alpine vascular plants of Mount Washington. *University of Colorado Press*.

Lutz, A. M. 1907. A preliminary note on the chromosomes of *Oenothera lamarckiana* and one of its mutants, *O. gigas*. *Science* 26: 151–152.

Mable, B. K. 2004. Polyploidy and self-compatibility: is there an association? *New Phytologist* 162: 803–811.

Mansion, G., G. Parolly, A. A. Crowl, E. Mavrodiev, N. Cellinese, M. Oganesian, K. Fraunhofer, G. Kamari, D. Phitos, R. Haberle, G. Akaydin, N. Ikinci, T. Raus, and T. Borsch. 2012. How to handle speciose clades? Mass taxon-sampling as a strategy towards illuminating the natural history of *Campanula* (Campanuloideae). *PLoS ONE* 7: e50076–23.

Marks, G. E. 1966. The origin and significance of intraspecific polyploidy: experimental evidence from *Solanum chacoense*. *Evolution* 20: 552–557.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257–1257.

Ramsey, J., and D. W. Schemske. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.

Scarpino, S. V., and D. A. Levin. 2014. Polyploid formation shapes flowering plant diversity. *The American Naturalist*. 184: 456-465.

Segraves, K. A., and J. N. Thompson. 1999. Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossulariifolia*. *Evolution* 53: 1114-1127.

Shetler, S. G. 1982. Variation and evolution of the nearctic Harebells (*Campanula* sect. Heterophylla). J. Cramer. Vaduz.

Slotte, T., H. Huang, M. Lascoux, and A. Ceplitis. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Molecular Biology and Evolution* 25: 1472–1481.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, and D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Stevens, C. J., J. Wilson, and H. A. McAllister. 2012. Biological Flora of the British Isles: *Campanula rotundifolia*. *Journal of Ecology* 100: 821–839.

Tank, D. C., J. M. Eastman, M. W. Pennell, P. S. Soltis, D. E. Soltis, C. E. Hinchliff, J. W. Brown, E. B. Sessa, and L. J. Harmon. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist* 207: 454–467.

Witasek, J. 1902. Ein beitrag zur kenntnis der gattung *Campanula*. Alfred Hölder, K.K. Hof- und Universitats-Buchhändler, Vienna, Austria.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* 106: 13875–13879.

Wright, S. I., S. Kalisz, and T. Slotte. 2013. Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B: Biological Sciences* 280: 20130133–20130133.

Zielinski, M.-L. and O. M. Scheid. 2012. Meiosis in polyploid plants. In Polyploidy and Genome Evolution, P. S. Soltis and D. E. Soltis (eds.), Springer-Verlag, Berlin, Germany

Zohren, J., N. Wang, I. Kardailsky, J. S. Borrell, A. Joecker, R. A. Nichols, and R. J. A. Buggs. 2016. Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Molecular Ecology* 25: 2413–2426.

CHAPTER ONE:

Glaciation and whole genome duplication shape the distribution of a polyploid complex

Premise of the study – Both intrinsic and extrinsic factors contribute to a species distribution. Among plants, the extrinsic effects of glaciation and the intrinsic effects of whole genome duplication are powerful drivers of biographical patterns, but the interplay of these factors is poorly understood. Here, we assess the roles glaciation and whole-genome duplication have played in the evolution of a widespread polyploid complex by elucidating the phylogeography and cytogeography of *Campanula rotundifolia*.

Methods – We assessed the cytotype of and constructed chloroplast phylogeny for 40 populations, and a nuclear RAD-seq phylogeny for 24 populations spanning the geographic and cytotypic range of the *C. rotundifolia* complex. We used chloroplast synonymous substitution rates to estimate major geographic divergence times.

Results and Conclusions– *Campanula rotundifolia* originated in southcentral Europe and underwent range expansion throughout much of Europe and North America. At least seven genome duplication events have occurred in *C. rotundifolia*—four tetraploid and four hexaploid formations. Both nuclear and chloroplast phylogenies are largely congruent with a history of populations surviving glacial maxima in known Pleistocene refugia in Europe and North America. Disjunct glacial refugia led to divergent European populations. North America was colonized by both tetraploids and hexaploids derived from Western European lineages. A glacial refugium in Midwestern North America likely facilitated post-glacial recolonization of North America and limited genetic divergence.

Introduction

The distribution of a species results from interplay between the intrinsic physiology of the species and the extrinsic abiotic and biotic environmental factors that act upon it. Tolerance for particular environmental conditions, such as minimum or maximum temperature (Araujo et al., 2013) or moisture availability (Legates et al., 2011), defines a species' fundamental niche, or the maximum range in which it is possible for that species to survive and reproduce (Kearney and Porter, 2004). Over time, the distribution of a species fluctuates due to climatic shifts that change where extrinsic conditions meet a species' fundamental niche (Martinez-Meyer and Peterson, 2006). This may be due to evolution of intrinsic factors that shift a species' abiotic requirements ((Lowry and Lester, 2006), or to the interaction of both intrinsic and extrinsic factors. Additionally, because species are often constrained from fully occupying the fundamental niche due to interspecific or intraspecific competition, changes in intrinsic factors unrelated to environmental tolerance, such as mating system, can also shift the distribution of a species (Guggisberg et al., 2006).

Among the environmental factors that can shape species distributions, few cause such widespread shifts as the climatic changes associated with cycles of glacial advance and retreat. Advancing glaciers scour pre-existing landscapes to bedrock, extirpating all flora under the ice sheet (Clements, 1916). However, climatic shifts extend far beyond the ice sheet, with unglaciated areas typically getting colder and drier (Prentice et al., 2000). In North America during the last glacial maximum, taiga and boreal forest extended as far south as Florida (Jackson et al., 2000), effectively pushing temperate flora hundreds of kilometers south of their interglacial ranges. Europe was likewise

substantially colder and drier during the last glacial maximum, forcing many temperate species into the southern peninsulas (Prentice et al., 2000).

Repeated cycles of glacial movement during the Pleistocene have dramatically shaped angiosperm biogeography across the Northern Hemisphere (Petit, 2003), but differences in topography between North America and Europe have resulted in different biogeographic patterns on each continent (Clark et al., 2009). In Europe, glaciation of the Alps effectively cut off the three southern peninsulas (Iberian, Apennine, and Balkan) as separate geographically isolated refugia (Sharbel et al., 2000), while a larger northern contiguous refugium was delineated to the north by advance of the Weichselian ice sheet and to the south by expansion of Alpine glaciation (Bhagwat and Willis, 2008). In North America, much of the continent remained unglaciated, and gene flow could continue within a large, contiguous Southern refugium (Shafer et al., 2010). Because Europe experienced more disjunct refugia during glacial advance than North America, species that inhabited both continents during glacial maxima may show different genetic signatures following post-glacial recolonization. Specifically, due to extrinsic climatic and geographic factors, North American populations are likely to be less diverged and less geographically widespread than European populations within the same species.

Species distributions can also be affected by changes in intrinsic factors. Few genetic changes in plants have such systemic effects on physiology and ecology as whole genome duplication (WGD), or polyploidy. The physiological changes can alter the fundamental niche of polyploids relative to their diploids ancestors, causing them to occupy different habitats. For instance, some polyploids have been shown to be more drought tolerant, usually due to reduced specific leaf area and fewer stomata (Li et al.,

1996; Mraz et al., 2014). Polyploidy is also associated with higher latitudes, suggesting either increased cold-tolerance (Sugiyama, 1998, but see Thompson et al., 2015) or increased rates of polyploidization in cold climates (Ehrendorfer, 1980).

In addition to changing intrinsic physiology, WGD can also affect how polyploid plants compete with other species or with diploid relatives. In some systems, polyploids have higher seed weights and faster vegetative growth than sympatric diploids (Maceira et al., 1993). However, in other systems, competition is not apparent between diploids and polyploids (Munzbergova, 2007; Thompson et al., 2015). Polyploids may compete most strongly against diploids when exploiting new niches during range expansion. This competitive edge is two-fold. First, polyploids often have increased ecological amplitude, allowing them to exploit habitats that diploids cannot (Soltis et al. 2003; Salmon et al., 2005; Meimberg et al., 2009). Second, polyploids frequently undergo mating system shifts, increasing either self-compatibility or selfing rate (Husband et al., 2008), which can help them colonize new habitats more successfully than diploids. These differences explain why WGD serves as a significant intrinsic contributor to changes in species distribution.

The widespread polyploid complex, *Campanula rotundifolia,* is a good candidate system for studying the interplay between the extrinsic effects of glaciation and the intrinsic effects of WGD on changes in biogeography. This complex is present across both Europe and North America, and has been subject to multiple cycles of glaciation on each continent. It comprises diploids, tetraploids, and hexaploids, the distribution of which varies across each continent. In this study, we seek to determine the effects of historical glaciation and WGD on the extant distribution of *C. rotundifolia*. We ask the

following questions: 1) What is the phylogeography of *C. rotundifolia*? 2) What is the cytogeography of *C. rotundifolia*? 3) How has glaciation, WGD, and their interaction influenced phylogeography and cytogeography of *C. rotundifolia* in Europe and North America?

Materials and Methods

Study system and Population Sampling

Campanula rotundifolia L. (*Campanula*ceae) is a polyploid complex of perennial wild flowers comprising multiple named species, subspecies, and varieties, though many likely represent ecotypes rather than distinct species (Böcher, 1936; Maad et al., 2013). The complex has a circumboreal distribution spanning the northern latitudes of Europe and North America, and diploids (2n=34), tetraploids (2n=68), and hexaploids (2n=102) comprise the three dominant cytotypes (Stevens et al., 2012).

Cytotypes are not uniformly distributed throughout the range. Tetraploids predominate throughout the distribution (Bocher, 1936; Kovanda, 1977; Shetler, 1982; Stevens et al., 2012), whereas diploids and hexaploids are more geographically restricted. Diploids are found throughout Central, Eastern, and far Northern Europe (Kovanda, 1977), and as narrow endemics within the Alps and Carpathian mountain ranges (Liber et al., 2008). Hexaploids in Europe are found primarily in the British Isles (Stevens et al., 2012), although they have been reported in Spain and Greece (Blionis and Vokou, 2005; Maad et al., 2013). Tetraploids overlap with both hexaploid and diploid ranges, though known areas of sympatry or parapatry are infrequent. In North America, tetraploids are most common, with hexaploids located in British Columbia and Alaska (Shetler, 1982),

though two hexaploid populations have been reported near Lake Huron in southern Ontario (K. Burgess, pers. comm.).

Populations were sampled through direct collection and by soliciting wild-gathered seeds from donors throughout the range of *C. rotundifolia* for a total of 39 populations spanning much of Europe, the United States, and Canada (Table 1). These populations broadly sample the distribution of *C. rotundifolia*, except for extreme northern Europe. In response to earlier reports of hexaploids near the Great Lakes, we increased sampling effort along the shores of and on the major islands within Lake Huron and Lake Superior (Figure S1). For each site, 10-50 plants were sampled, but no more than 50% of all fruiting individuals. One to three mature fruits were taken from each plant for all populations and 5-10 cauline leaves for a subset of populations. Leaf tissue was stored in silica gel within 1 hour of collection. Seeds were pooled across all fruits from each individual.

Ploidy estimation via flow cytometry

For each population, 10 seeds from up to 10 individuals per population were grown to obtain leaf tissue for flow cytometric analysis. Seeds were planted in two cells of five seeds each, and reared in a growth chamber on a 12/12 light/dark cycle at 21 C/15C for eight weeks, or until seeds had germinated and grown to basal rosettes of at least 10 leaves each.

Two leaves from each germinant within a cell (up to 10 leaves: 2 x 5 germinants) were harvested into one pooled sample for flow cytometry. Tissue preparation followed a modified Otto 2-step protocol (Otto, 1990); 1 ml of Otto I buffer was added to pooled leaf samples (approximately 30-50 mg of tissue), and finely chopped with a fresh razor

blade to release whole nuclei. Cell suspensions were then strained through a 30 µm nylon filter and incubated at room temperature for 1 hour. 1 ml of Otto II buffer, supplemented with 50 ng/µl propidium iodide and 50 ng/µl RNase, were added to the cell suspension and mixed gently. Samples were incubated for 10-30 minutes at room temperature before being analyzed on a BD FACSCalibur Cell Analyzer using a 488 nm laser. Two standards were used for all populations, either co-chopped with sample leaf tissue as an internal standard or run prior to all samples in one sitting as an external standard. For putative diploid and tetraploid samples, *Raphanus sativus* (2C DNA content = 1.1 pg) was used, and *Glycine max* (2C DNA content = 2.50 pg) was used for putative hexaploid samples (Dolezel et al., 2003).

Chloroplast sequencing and phylogeny

To construct the chloroplast phylogeny, 39 populations were chosen to span the geographic range and to fully capture cytotypic diversity, these included eight from the targeted Great Lakes sampling (Table 1). Seeds were grown in the University of Virginia greenhouse, and young leaves were collected from basal rosettes for DNA extraction. In cases of poor germination, dried field-collected leaf tissue was used for DNA extraction if available. *Campanula divaricata* was chosen as an outgroup because it is closely related to the *C. rotundifolia* clade (Mansion et al., 2012). DNA for chloroplast sequencing was extracted using a modified CTAB procedure (Porebski et al., 1997).

Seven chloroplast markers were chosen from either existing primer sets or were designed specifically for *Campanula rotundifolia*. Existing primer sets (Taberlet et al., 1998) showed little variation throughout the range, although trnL-F was phylogenetically informative and therefore retained for this analysis. To improve phylogenetic resolution,

five populations were sampled from throughout the distribution (2X Czech Republic, 4X Czech Republic, 4X Belgium, 4X Virginia, and 6X Alaska) to search for highly variable regions within the complete *C. rotundifolia* chloroplast genome. Chloroplast DNA was extracted using a glucose gradient protocol (Sikoskaite, 2013). Chloroplast genomes were sequenced from Nextera library constructs on an Illumina MiSeq, and were assembled using bowtie2 (Langmead and Salzberg, 2012). Large contigs from the partial assembly were screened for regions of high polymorphism, yielding nine additional loci, of which six were used in addition to trnL-F (Table S1).

For each population, up to three individuals (2.65 ± 0.2 S.D.) were sequenced for each of the seven loci. Loci were amplified using 5 PRIME HotMasterMix (5 Prime, Gaithersburg, Maryland, USA), and PCR products were purified using Exonuclease I and Shrimp Alkaline Phosphatase (Affymetrix, Santa Clara, California, USA). Sequencing was carried out using BigDye Terminator v3.1 Cycle Sequencing Kit and cleaned using Sephadex G-50 (GE-Healthcare, Little Chalfont, United Kingdom). Sequencing products were analyzed either in-house using an ABI 3130xl sequencer, or submitted to Yale Sequencing Center for analysis. Sequences were checked for accuracy and manually aligned using CodonCode Aligner version 3.5 (CodonCode Corporation, Centerville, Maine, USA). Sequences were concatenated across loci for each individual. Maximum likelihood trees were constructed using RAxML (Stamatakis, 2014) for the concatenated sequences with 100 bootstrap replicates. Divergence times were estimated based on average synonymous substitution rates across all chloroplast loci and standardized to published estimates of chloroplast nucleotide divergence (Wolfe et al., 1987).

Nuclear sequencing and phylogeny

A RAD-seq approach (Davey and Blaxter, 2011) was employed to construct the nuclear phylogeny in order to broadly sample the nuclear genome and obtain sufficient informative characters. 28 populations were chosen to span the geographic and cytotypic distribution (Table 1). For each population, leaf tissue was collected from 6-10 individuals (7.8 ± 1.7 S.D.) either from wild plants or from greenhouse plants grown from wild-collected seed. DNA was extracted using a modified CTAB protocol, and quantified using a Qubit Assay Kit. All extractions within a population were evenly pooled by DNA quantity, then barcoded using Illumina's TruSeq barcodes. Library construction was performed at the Genomics Core Facility in the University of Virginia's Department of Biology. Barcoded populations were digested to completion using SacI, then pooled for library construction using standard Illumina HiSeq protocols. The pooled library was sequenced in parallel across two lanes of an Illumina HiSeq machine. Sequencing was performed at Beckman Coulter Genomics (Danvers, Massachusetts, USA).

Sequence cleaning and alignment were performed with Stacks (Catchen et al., 2013). Using the process-radtags program, reads were demultiplexed and low quality reads were removed. Reads were run through the Stacks pipeline to identify and genotype loci. A minimum read depth of 10 was used when constructing stacks so that each individual had a theoretical opportunity of being included in that stack. Because pooled populations were used and individuals had up to six alleles per locus, it was not possible to distinguish sequencing error from rare alleles at the read depth across most SNPs. Therefore, informative characters were drawn from only those SNPs that were fixed

within a population and variable among populations. The barcode error frequency was estimated using the percentage of reads that were dropped in 'process_radtags' because of ambiguous barcodes. Although fixed SNPs were used for phylogenetic analysis, process_radtags was also run using the standard SNP model, and relative proportion of fixed and variable sites was compared across populations.

A Perl script was used to concatenate the sequences from the Stacks 'FASTA' output for each population (the FASTA output contained the full sequence of each allele from each sample locus). The concatenated sequences were then used to construct a maximum likelihood tree in RAxML version 8.0.6 (Stamatakis, 2014) using 100 bootstrap replicates and a GTR + gamma model of evolution. The tree was visualized in Figtree (http://tree.bio.ed.ac.uk/software/figtree/) and rooted according to the chloroplast phylogeny.

Results

Cytotypic distribution of C. rotundifolia

Of the 39 populations sampled for this study, 35 populations exhibited only one cytotype across all individuals sampled. Three populations were exclusively diploid, 26 populations were tetraploid, and six populations were hexaploid. Of the four mixed-ploidy populations, two contained diploids and tetraploids, and two contained tetraploids and hexaploids. No population contained diploids and hexaploids. While no triploids were found, some pentaploids and aneuploids were observed in the mixed 4X-6X populations. These individuals were omitted from cpDNA and nuclear sequencing (though see Chapter 6). The mixed-ploidy population near northern Lake Huron (115;

Figure S1) appears to be located in a zone of contact between tetraploid and hexaploid populations.

Chloroplast sequencing and phylogeny

Seven chloroplast loci were used for phylogenetic analysis: trnL-F yielded 835 bp, ycf2 yielded 641 bp, rpl22 yielded 740 bp, rps7 yielded 668 bp, rps3 yielded 617 bp, ndhB yielded 546 bp, and ycf15 yielded 488 bp for a total of 4535 bp. Maximum likelihood phylogenetic analysis found three large clades and three distinct pairs of populations not assigned to larger clades (Figure 1). The three large clades consisted of tetraploids and hexaploids, and were geographically associated with Eastern/Central Europe, Western/Northern Europe, and North America (Figure 2).

Older populations tend to occur near southern Europe, with newer populations generally more northern and western (Figures 1 and 2). Central European diploids were most closely related to the outgroup and subtended the Eastern/Central clade. Slovakian and Slovenian tetraploids occur as a grade subtending the Western/Northern clade plus the two Italian taxa broadly considered part of the *C. rotundifolia* complex, diploid *C. pollinensis* and tetraploid *C. scheuchzeri*. Within the Western/Northern clade, populations proceed northerly and westerly, with basal French populations and British and Scandinavian populations most recently derived. A total of four duplications to tetraploidy are discernible from chloroplast data—the Czech Republic (25/44), eastern Germany (43), Italy (12), and in the lineage leading to Western Europe and North America. The two hexaploid populations resolved to different locations within the Western/Northern clade: Irish hexaploids sister to the Scandinavian populations, and British hexaploids sister to sympatric tetraploids.

North American populations form a well-supported clade that appears to originate in the Midwestern United States and splits into two lineages: one broadly eastern and one western (Figures 1 and 2). Basal haplotypes in both lineages are shared between tetraploid and hexaploid populations, and are present around the Great Lakes. An independent genome duplication event occurred in western North America, making a total of four observed hexaploidy events from the observed populations (formation of Irish, British, North American, and Alaskan hexaploids).

Based on synonymous substitution rates, *C. rotundifolia* diverged from the outgroup *C. divaricata* 2.3 million years ago (Estimate range = 2,197,117 – 2,456,956). Within *C. rotundifolia*, the split between Central and Western Europe occurred 818,000 years ago (Estimate range = 774,313 – 863,657), and between Western Europe and North America occurred 220,000 years ago (207,727 – 231,696). The western and eastern North American lineages diverged from each other approximately 200,000 years ago (192,641 – 214,868).

Nuclear sequencing and phylogeny

RAD sequencing yielded 1.3 – 27.9 million reads per pooled population. Four populations retained fewer than 10 million reads after quality control, representing fewer than 10% of all loci, and were thus removed from analysis (Table 1). Within pooled individuals, the average coverage depth per locus was 46.1 reads (standard error = 8.7). Because reduced-representation sequencing is often not uniform across all samples, almost all loci had data missing from at least one population. To construct the data matrix, loci were retained if data were present for at least 12 populations per locus, yielding a data matrix comprising 25,112 loci.

The nuclear RAD tree, while broadly similar to the chloroplast tree, resolved fewer clades. However, the resulting grade topology was better supported. The nuclear phylogeny places Central European diploids basal to all other populations (Figure 3). A well-supported Central/Western European tetraploid grade is subtended by the Central European diploids.

Sister to the Central/Western European grade is a moderately well-supported Southern/Central European grade comprising Slovenian and Slovakian tetraploids as well as the two Italian taxa (Figure 3). The Italian taxa, *C. scheuchzeri* and *C. pollinensis*, form a strongly supported sister group, subtended with good support by Slovakia and with marginal support by Slovenia. When either or both Italian populations are removed, Slovenia and Slovakia collapse into a marginally well-supported clade.

Beyond this grade, a well-supported clade contains all North American populations plus Ireland. The phylogeny strongly supports basal positioning of Ireland and all Midwestern North American hexaploids (Figure 3), though the precise relationship of hexaploid Ontario (10) and Michigan (105) is unclear. Two topologies existed at this node: 53% of all bootstrap replicates resolved Michigan basal to Ontario, and 47% allied them as sister taxa.

Beyond the basal hexaploids, North American populations are strongly separated into eastern and western clades (Figure 3). The western clade is further subdivided into two putative migration routes, one southern and one northern. Along the northern route, hexaploidy has independently arisen along the far northwestern Pacific Coast. The eastern clade comprises Appalachian and Canadian Maritime populations, as well as populations north of the Great Lakes.

Discussion

Phylogeography

Campanula rotundifolia originated in southcentral Europe approximately 2.3 million years ago, and is most closely related to Mediterranean *Campanula* diploids in section Heterophylla (Kovanda, 1977; Mansion et al., 2012). From this origin, the range expanded mostly northward and westward as plants migrated into central Europe (Figure 4), though populations remained in the Balkans and in Italy. Western European populations migrated northward following the last glacial maximum, though the precise relationship between Western European and Central/Southern European populations remains unclear. Two routes of migration from southern France are suggested by the chloroplast phylogeny. One route colonized the southern British Isles while the other colonized much of Western Europe, the northern British Isles, and Scandinavia (Figure 4).

North American populations diverged from Western Europe approximately 220,000 years ago, likely migrating from northwestern Europe to northeastern North America around that time (Figure 4). The oldest observed North American populations are located in the Midwest and around the Great Lakes. From this region, two lineages diverged from each other shortly following divergence from European populations (approximately 200,000 years ago) and have resulted in colonization of much of northern North America from the upper Midwest (Figure 4). The easterly route colonized New England, the Canadian Maritimes, and the Appalachian Mountains. The westerly route colonized the Rockies, the Sierra Nevada, and the Pacific coast northward to Alaska.

Although the chloroplast and nuclear phylogenies are in broad agreement, incongruent nodes between data sets leave some evolutionary relationships unresolved. The Irish hexaploid population resolves with North America in the nuclear tree and with Western Europe in the chloroplast tree, and a western Swedish population allies with Belgium in the nuclear tree and with eastern Sweden in the chloroplast tree. These incongruities may result from gene flow between populations of different lineages that are geographically proximate. In particular, the Czech population is near populations from the Western European clade, and the Irish population is near populations from both southern and northern colonization routes to the British Isles. Long branch attraction is another likely source of incongruity. In the chloroplast phylogeny, Slovakia and Slovenia form a discrete pair, as do the Italian populations. However, the nuclear phylogeny, in which the Italian diploid population has an especially long branch length, collapses these four populations (*C. pollinensis, C. scheuchzeri*, Slovenia and Slovakia) into a single clade. When either Italian population is removed from the analysis, Slovakia and Slovenia resolve as a marginally well-supported pair. This suggests that the nuclear phylogeny incorrectly clusters these populations and that Slovakia and Slovenia should be considered a separate grouping from the Italian populations.

Extrinsic effects of glaciation

Based on its inferred date of origin, *C. rotundifolia* arose early in the Quaternary. Although the glacial maxima in the early Quaternary are poorly circumscribed, *C. rotundifolia* was likely affected by four glacial maxima in Alps (Würmian, Rissian, Mendelian, and Günzian dating back to 800 KYA) and three in northern Europe (Weichselian, Saalian, and Elsterian dating back to 500 KYA; Ehlers and Gibbard, 2008).

By contrast, the estimated age of intercontinental dispersal suggests that *C. rotundifolia* populations have experienced at most two glacial maxima (the Wisconsinan and Illinoisan) in North America. Therefore, the effects of glaciation on divergence of North American populations can be inferred more precisely, while inferences of older divergences among European populations are less certain.

Glaciation appears to have resulted in the disjunct distribution and divergence of diploid populations in European *C. rotundifolia*. Diploid individuals occur in central Europe, Italy, and far northern Europe (not sampled; Kovanda, 1977; Shetler, 1982). The basal positioning of central European diploids and the diploid nature of related species in section Heterophylla suggest that *C. rotundifolia* originated as a diploid species and was once widespread across much of Europe. The disparate locations in the cpDNA and nuclear phylogeny as well as the relatively long branch lengths on central European and particularly on Italian diploids indicates that these taxa have been genetically isolated for much of the evolutionary history of *C. rotundifolia*. While the Alps likely presented an ongoing barrier to gene flow between Italian populations and those further north, the close association of Italian populations with Slovenian and Slovakian tetraploids shows that more recent gene flow existed east of the Alps. It is likely that early *C. rotundifolia* diploid populations became disjunct north and south of Alps during an early glacial maximum in the Quaternary Period (unsampled northern diploid populations likely derive from ancestors that persisted north of the Alps during an earlier glacial cycle). During that time and subsequent glaciations, gene flow may have been maintained with other southerly diploids through calcareous refugia in southeastern Alps near the Dolomite mountains (Schönswetter, 2005), but not with more northerly diploids. Following retreat

of the Alpine glaciers (Deffontaine et al., 2005), the Italian and central European diploids were sufficiently disjunct that the Alps served as a sufficient continuous barrier to gene flow, allowing the high levels of genetic distance observed in extant populations to accumulate.

While disjunct diploids indicate that glaciation may have initiated early isolation, the rapid expansion of Western European *C. rotundifolia* reflects the effect of glacial retreat, likely predating the Günzian glaciation (approx. 1.01 MYA). Although population sampling limits the information obtainable from the nuclear phylogeny, the cpDNA phylogeny places the origin of the Western European clade in southern France, and related to the Italian populations. This suggests that during the last glacial maximum, *C. rotundifolia* tetraploids may have survived in glacial refugia in the southwestern Alps, between modern-day Nice and the Aoste valley. This refugium has been implicated in the survival of other species during the last glacial maximum (Schönswetter, 2005). From this origin, *C. rotundifolia* tetraploids expanded northward rapidly following glacial retreat, and have colonized as far north as the British Isles and Scandinavia (4, 18, 21, 58, 120; Figure 1, 2). Glaciation likely facilitated migration north by exposing land bridges connecting Scandinavia, the British Isles, and mainland Europe (Deffontaine, 2005).

North American populations are genetically and geographically consistent with long-distance dispersal of both tetraploids and hexaploids just prior to the Illinoisan glacial maximum approximately 200,000 years ago. The most basal populations of the North American clade are geographically restricted Midwestern tetraploids (cpDNA phylogeny only) and hexaploids (both phylogenies), which are seemingly incongruous with a hypothesized long-distance dispersal from northwestern Europe. However, the

glacial history of North America provides a possible cause for both the basal position of Midwestern populations and for the absence of older eastern North American populations. During the last two glacial maxima, the Wisconsinan approx. 18,000 years ago and the Illinoisian approx. 200,000 years ago, Quaternary glaciation extended as far south as Pennsylvania, the Ohio River, and Kansas (Hallberg, 1986). However, much of Minnesota, Wisconsin, and Iowa remained unglaciated (Hobbs, 1999), and the Paleozoic Plateau in southern Minnesota and northern Iowa has never been glaciated (Hobbs, 1999). This unglaciated region has served as a refugium for multiple species during glacial maxima (Ross, 1999; Hobbs, 1999; Li et al., 2013). *Campanula rotundifolia* appear to have persisted in this unglaciated region, and expanded both eastward and westward following glacial retreat. This eastward and westward expansion has been seen in genus *Smilax*, which is also thought to have survived the last glacial maximum in the Paleozoic plateau (Li et al., 2013).

Cytogeography

Campanula rotundifolia has undergone at least seven separate genome duplication events: three from diploidy to tetraploidy and four from tetraploidy to hexaploidy. In some cases, these duplication events are associated with range expansions. The tetraploidy event that occurred in Central Europe preceded expansion across much of Central and Western Europe. Similarly, tetraploids likely underwent long-distance dispersal from northwestern Europe to North America, and subsequently spread throughout the continent. Hexaploids likewise show patterns of broad range expansion, from northern Europe into Ireland, then from Europe to North America. Range expansion of hexaploids within North America is less clear; Midwestern hexaploids have a limited

distribution, and sampling density in northwestern North America makes the extent of western hexaploids unclear.

Recurrent WGD events and overlapping dispersal history create multiple regions in which two ploidy levels are either in close association, or form zones of sympatric or parapatric contact. Four contact zones are observed in this study. Parapatric diploids and tetraploids in Central Europe likely result from a recent local genome duplication. Sympatric tetraploids and hexaploids in England likewise appear to result from a local genome duplication. Two other contact zones appear to originate from true secondary contact of different cytotypes: one in eastern Germany between diploids and tetraploids, and one in southern Ontario between tetraploids and hexaploids (Figure S1). Finally, the Italian diploids and tetraploids, although not in sympatry or parapatry, have a shared history of endemism due to the Alps serving as a barrier to migration.

Effects of genome duplication

Although establishment of new polyploids is difficult due to genetic instability (Hollister, 2012) and minority cytotype exclusion (Husband, 2000), polyploids that do establish are often at a competitive and dispersal advantage relative to diploid progenitors due to wider ecological amplitude (Petit et al., 1999), as well as potential for higher seed set and faster growth (Maceira et al., 1993). Phylogenetic and geographic evidence suggest that, as in other species (Petit et al., 1999), tetraploids may have competitive advantages over diploids in most habitats in Europe, and have consequently extirpated diploids from much of the range. Among European populations, tetraploid *C. rotundifolia* generally have higher seed set than diploids (Chapter 2), and germinate and flower earlier (Stevens et al., 2012). These differences could be due to adaptive changes subsequent to

genome duplication, but may also reflect intrinsic physiological change due to WGD, as has been observed in other systems (e.g. Ramsey et al., 2011). If these biological differences have remained consistent over time, they could provide an explanation for the greater spread of tetraploids over diploids.

Tetraploids may also be widespread due to expansion into ranges outside the fundamental niche of diploids. Polyploids are noted for having a greater capacity for self-fertilization (Stupar et al., 2007; Tao and Iezzoni, 2010; Chapter 2) as well as increased ecological amplitude (Petit et al., 1999). These advantages can facilitate both establishment of polyploids at range edges and long-distance dispersal. Polyploids are often self-compatible due to loss of incompatibility systems with genome duplication (Stone, 2002, Tao and Iezzoni, 2010), and to strong selective pressures associated with overcoming minority cytotype exclusion (Mable, 2004). Because self-compatible polyploids more likely to establish and survive as small populations than diploids, WGD is thought to facilitate long-distance dispersal (Linder and Barker, 2014). Consistent with this hypothesis, North American *C. rotundifolia*, both tetraploids and hexaploids, are significantly more self-compatible than European lineages (Chapter 2). Increased self-compatibility may explain the initial intercontinental dispersal of the North American lineages as well as their widespread colonization of North America.

Conclusion

Campanula rotundifolia is a widespread polyploid complex with a complicated cytotypic distribution. While glaciation and WGD independently explain much of the biogeography of *C. rotundifolia*, some relationships are best explained by the interplay of both forces. For instance, increased ecological tolerance and loss of self-incompatibility

(Chapter 2) in polyploid *C. rotundifolia* help explain their competitive edge over diploid relatives, but these advantages may have been most beneficial in the open niches afforded by glacial retreat. Indeed, colonization of both Western Europe and North America by *C. rotundifolia* polyploids was rapid following glacial retreat, and led to the widespread and relatively genetically diverse Western European, Eastern North American, and Western North American clades. Although its phylogeography is congruent with classic signatures of glaciation, the cytogeography shows that WGD was instrumental in the dispersal and establishment success of *C. rotundifolia*. In order to fully understand the dispersal history and biogeography of widespread polyploid complexes, it is necessary to consider both extrinsic climatic factors as well as the intrinsic factors associated with whole-genome duplication.

Acknowledgements

We would like to thank A. Nguyen for help constructing the chloroplast and RAD-seq libraries as well as for technical guidance; K. Barnard-Kubow for technical guidance; and C. Debban and M. Koski for comments on an early version of this manuscript.

References

Araujo, M. B., F. Ferri-Yanez, F. Bozinovic, P. A. Marquet, F. Valladares, and S. L. Chown. 2013. Heat freezes niche evolution. *Ecology Letters* 16: 1206–1219.

Bhagwat, S. A., and K. J. Willis. 2008. Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* 35: 464–482.

Blionis, G. J., and D. Vokou. 2005. Reproductive attributes of *Campanula* populations from Mt. Olympos, Greece. *Plant Ecology* 178: 77–88.

Bocher, T. W. 1936. Cytological studies on *Campanula rotundifolia*. *Heriditas* 22: 269–277.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Clark, P. U., A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, and J. X. Mitrovica, et al. 2009. The last glacial maximum. *Science* 325: 710–714.

Clements, F. E. 1916. Plant succession: An analysis of the development of vegetation. Carnegie Institution of Washington, Washington.

Davey, J. W., and M. L. Blaxter. 2011. Radseq: next-generation population genetics. *Briefings in Functional Genomics* 9: 416–423.

Deffontaine, V., R. Libois, P. Kotlik, R. Sommer, C. Nieberding, E. Paradis, J. B. Searle, and J. R. Michaux. 2005. Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology* 14: 1727–1739.

Dolezel, J., J. Greilhuber, S. Lucretti, A. Meister, M.A. Lysak, L. Nardi, and R. Obermayer. 1998. Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Annals of Botany* 82 (Suppl. A): 17–26.

Dufresne, F., M. Stift, R. Vergilino, and B. K. Mable. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* 23: 40-69.

Ehrendorfer, F. 1980. Polyploidy and distribution. in Polyploidy: Biological Relevance, Ed. W. H. Lewis. Springer-Verlag, USA.

Guggisberg, A., G. Mansion, S. Kelso, and E. Conti. 2006. Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex of *Primula*. *New Phytologist* 171: 617-632.

Hallberg, G. R. 1986. Pre-Wisconsin glacial stratigraphy of the central plains region of Iowa, Nebraska, Kansas, and Missouri. Quaternary Science Reviews 5: 11-15.

Hobbs, H. 1999. Origin of the driftless area by subglacial drainage - a new hypothesis. In Geological Processes Past and Present. Eds. D. M. Mickelson and J. W. Attig. Geological Society of America, Boulder, CO,. USA.

Hollister, J. D., B. J. Arnold, E. Svedin, K. S. Xue, B. P. Dilkes, and K. Bomblies. 2012. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa PLOS Genetics* 8: e1003093.

Husband, B. C. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings of the Royal Society B: Biological Sciences* 267: 217–223.

Husband, B. C., B. Ozimec, S. L. Martin, and L. Pollock. 2008. Mating consequences of polyploid evolution in flowering plants: Current trends and insights from synthetic polyploids. *International Journal of Plant Sciences* 169: 195–206.

Jackson, S. T., R. S. Webb, K. H. Anderson, J. T. Overpeck, T. Webb III, J. W. Williams, and B. C. S. Hansen. 2000. Vegetation and environment in eastern North America during the last glacial maximum. *Quaternary Science Reviews* 19: 489-508.

Kearney, M., and W. P. Porter. 2004. Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology* 85: 3119–3131.

Kovanda, M. 1977. Polyploidy and variation in the *Campanula rotundifolia* complex. Part II. (Taxonomic) 2. Revision of the groups Vulgares and Scheuchzerianae in Czechoslovakia and adjacent regions. *Folia Geobotanica & Phytotaxonomica* 12: 23–89.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

Legates, D. R., R. Mahmood, D. F. Levia, T. L. deLiberty, S. M. Quiring, C. Houser, and F. E. Nelson. 2011. Soil moisture: A central and unifying theme in physical geography. *Progress in Physical Geography* 35: 65–86.

Li, W.-L., G. P. Berlyn, P. M. S. Ashton. 1996. Polyploids and their structural and physiological characteristics relative to water deficit in *Betula papyrifera* (Betulaceae). *American Journal of Botany* 83: 15-20.

Li, P., M. Li, Y. Shi, Y. Zhao, Y. Wan, C. Fu, and K. M. Cameron. 2013. Phylogeography of North American herbaceous *Smilax* (Smilacaceae): combined

AFLP and cpDNA data support a northern refugium in the driftless area. *American Journal of Botany* 100: 800-814.

Liber, Z., S. Kovačić, T. Nikolić, S. Likić, and G. Rusak. 2008. Relations between western Balkan endemic *Campanula* (*Campanula*ceae) lineages: Evidence from chloroplast DNA. *Plant Biosystems* 142: 40–50.

Linder, H. P. and N. P. Barker. 2014. Does polyploidy facilitate long-distance dispersal? *Annals of Botany* 113: 1175–1183.

Lowry, E., and S. E. Lester. 2006. The biogeography of plant reproduction: potential determinants of species' range sizes. *Journal of Biogeography* 33: 1975–1982.

Maad, J., W. S. Armbruster, and C. B. Fenster. 2013. Floral size variation in *Campanula rotundifolia* (*Campanula*ceae) along altitudinal gradients: patterns and possible selective mechanisms. *Nordic Journal of Botany* 31: 361–371.

Mable, B. K. 2004. Polyploidy and self-compatibility: is there an association? *New Phytologist* 162: 803–811.

Mable, B. K., A. V. Robertson, S. Dart, C. Di Berardo, and L. Witham. 2005. Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* 59: 1437–1448.

Maceira, N. O., P. Jacquard, and R. Lumaret. 1993. Competition between diploid and derivative autotetraploid *Dactylis glomerata* L. from Galicia. Implications for the establishment of novel polyploid populations. *New Phytologist* 124: 321-328.

Mansion, G., G. Parolly, A. A. Crowl, E. Mavrodiev, N. Cellinese, M. Oganesian, K. Fraunhofer, G. Kamari, D. Phitos, R. Haberle, G. Akaydin, N. Ikinci, T. Raus, and T. Borsch. 2012. How to handle speciose clades? Mass taxon-sampling as a strategy towards illuminating the natural history of *Campanula* (Campanuloideae). *PLoS ONE* 7: e50076–23.

Martinez-Meyer, E. and A. T. Paterson. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography* 33: 1779-1789.

Meimberg, H., K. J. Rice, N. F. Milan, C. C. Njoku, and J. K. Mckay. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* 96: 1262-1273.

Mràz, P., E. Tarbush, and H. Müller-Schärer. 2014. Drought tolerance and plasticity in the invasive knapweed *Centaurea stoebe* s.l. (Asteraceae): Effect of populations stronger than those of cytotype and range. Annals of Botany 114: 289-299.

Münzbergova, Z. 2007. No effect of ploidy level in plant response to competition in a common garden experiment. *Biological Journal of the Linnean Society* 92: 211-219.

Otto, F. J. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewickz Z, Crissman HA, eds. Methods in cell biology, vol. 33. San Diego, CA, USA: Academic Press, 105–110.

Petit, C., F. Bretagnolle, and F. Felber. 1999. Evolutionary consequences of diploid–polyploid hybrid zones in wild species. *Trends in Ecology & Evolution* 14: 306–311.

Petit, R. J. 2003. Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* 300: 1563–1565.

Porebski, S., L. G. Bailey, and B. R. Baum. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15: 8–15.

Prentice, I. C., Jolly, D., and BIOME 6000 participants. 2003. Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography* 27: 507– 519

Ramsey, J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences* 108: 7096–7101.

Ross, T. K. 1999. Phylogeography and conservation genetics of the Iowa Pleistocene snail. *Molecular Ecology* 8: 1363–1373.

Salmon, A., M. L. Ainouche, J. F. Wendel. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* 14: 1163-1175.

Schönswetter, P., I. Stehlik, R. Holderegger, and A. Tribsch. 2005. Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology* 14: 3547–3555.

Shafer, A. B. A., C. I. Cullingham, S. D. Côté, and D. W. Coltman. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* 19: 4589–4621.

Sharbel, T. F., B. Haubold, and T. Mitchell-Olds. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology* 9: 2109–2118.

Shetler, S. G. 1982. Variation and evolution of the nearctic Harebells (*Campanula* sect. Heterophylla). J. Cramer. Vaduz.

Sikorskaite, S., M.-L. Rajamaki, D. Baniulis, V. Stanys, and J. P. Valkonen. 2013. Protocol: Optimised methodology for isolation of nuclei from leaves of species in the Solanaceae and Rosaceae families. *Plant Methods* 9:31-40.

Soltis, D. E., P. S. Soltis, and J. A. Tate. 2003. Advances in the study of polyploidy since *Plant speciation. New Phytologist* 161: 173–191.

Soltis, P. S., and D. E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. *PNAS* 97: 7051–7057.

Stamatakis, A. 2014. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stevens, C. J., J. Wilson, and H. A. McAllister. 2012. Biological flora of the British Isles: *Campanula rotundifolia*. *Journal of Ecology* 100: 821–839.

Stone, J. L. 2002. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *The Quarterly Review of Biology* 77: 17-32.

Stupar, R. M., P. B. Bhaskar, B. S. Yandell, W. A. Rensink, A. L. Hart, S. Ouyang, R. E. Veilleux, et al. 2007. Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* 176: 2055–2067.

Sugiyama, S. 1998. Differentiation in competitive ability and cold tolerance between diploid and tetraploid cultivars in *Lolium perenne*. *Euphytica* 105: 55-59.

Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.

Tao, R., and A. F. Iezzoni. 2010. The S-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic and molecular features. *Scientia Horticulturae* 124: 423–433.

Thompson, K. A., B. C. Husband, and H. Maherali. 2015. No influence of water limitation on the outcome of competition between diploid and tetraploid *Chamerion angustifolium* (Onagraceae). *Journal of Ecology* 103: 733-741.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *PNAS* 9054-9058

Table 1) *Campanula rotundifolia* accessions sampled for this study. Populations are arranged by longitude and are separated by continent (black line). Locality is given to the nearest city or park in Europe. Canadian and United States collections include the province and state. Populations displaying two ploidy levels represent mixed populations. Asterisk denotes populations omitted from RAD due to low coverage. Plus denotes mixed-ploidy population in which only one ploidy level (4X) was included in RAD analysis.

Taxon	Population ID	Ploidy	RAD	Locality	Country	Latitude	Longitude
C. rotundifolia	18	4	X	Uppsala	Sweden	59.793500	17.687966
C. rotundifolia	6	4	X	Kosice	Slovakia	48.662026	17.110677
C. scheuchzeri	12	4	X	Calabria	Italy	39.292853	16.535462
C. rotundifolia	44	2/4		Prague	Czech Republic	50.122280	14.418402
C. rotundifolia	2	4	X	Lasko	Slovenia	45.746155	14.210064
C. rotundifolia	43	2/4		Mittelndorf	Germany	50.936227	14.203765
C. gentilis	24	4	X	Vinaricka Hora	Czech Republic	50.184791	14.081019
C. gentilis	25	2	X	Zamky	Czech Republic	50.429255	14.075625
C. rotundifolia	23	2	X	Dresden	Germany	50.973440	13.821118
C. pollinensis	11	2	X	Teramo	Italy	42.472284	13.576092
C. rotundifolia	21	4	X	Halmstad	Sweden	56.635697	12.912012
C. rotundifolia	33	4	X*	Bielefeld	Germany	52.035005	8.492211
C. rotundifolia	1	4		Bonn	Germany	50.703711	6.975901
C. rotundifolia	32	4	X*	Col du Lautaret	France	45.018390	6.387670
C. rotundifolia	16	4	X	Liege	Belgium	50.720598	5.881635
C. rotundifolia	4	4	X	Derbyshire	England	53.248518	-1.459668
C. rotundifolia	120	4/6		Cheddar Gorge	England	51.286392	-2.762467
C. rotundifolia	58	6	X	Leitrim	Ireland	54.448743	-8.243710
C. rotundifolia	51	4	X	Cote D'Or, NS	Canada	45.290731	-64.773962
C. rotundifolia	53	4		Gaspesie NP, QB	Canada	48.947424	-66.124616
C. rotundifolia	55	4		Quebec City, QB	Canada	46.856518	-71.481777
C. rotundifolia	30	4	X	Turner's Falls, MA	United States	42.595089	-72.581203
C. rotundifolia	26	4	X	Narrows, VA	United States	37.264392	-80.984409
C. rotundifolia	13	6	X	Port Elgin, ON	Canada	44.812101	-81.298583
C. rotundifolia	10	6	X	Red Bay, ON	Canada	44.416252	-81.458360
C. rotundifolia	116	6		South Baymouth, ON	Canada	45.558822	-82.018024
C. rotundifolia	114	4	X	Chutes PP, ON	Canada	46.226020	-82.076846
C. rotundifolia	115	4/6	X+	Misery Bay PP, ON	Canada	45.800465	-82.749153
C. rotundifolia	105	6	X	Wilderness SP, MI	United States	45.466180	-83.880487
C. rotundifolia	109	4		Gros Cap, ON	Canada	46.536313	-84.592852
C. rotundifolia	113	4		Lake Superior PP, ON	Canada	47.134278	-84.724688
C. rotundifolia	34	4	X*	Caledonia, MN	United States	43.594553	-91.433767
C. rotundifolia	56	4	X	Los Alamos, NM	United States	36.005245	-106.380679
C. rotundifolia	5	4		Gothic, CO	United States	38.958520	-106.989946
C. rotundifolia	3	4	X*	Flathead Lake, MT	United States	47.905419	-114.021168
C. rotundifolia	36	4	X	Calgary, AB	Canada	50.981862	-114.177349
C. rotundifolia	38	4	X	San Juan Islands, WA	United States	48.525043	-123.001733
C. rotundifolia	39	4	X	Trinity NF, CA	United States	40.901941	-123.156648
C. rotundifolia	31	6	X	Anchorage, AK	United States	61.524261	-149.262686

Figure 1) Chloroplast maximum likelihood phylogeny for *Campanula rotundifolia* accessions. Numbers at nodes denote bootstrap support. Colors differentiate clades and correspond to locations in Figure 2. Black bar denotes outgroup. Ploidy levels of diploid, hexaploid, and mixed-ploidy haplotypes are labeled; unlabeled haplotypes are tetraploid. Populations for each haplotype are given within bars; those listed in white are sympatric contact zones between ploidy levels.

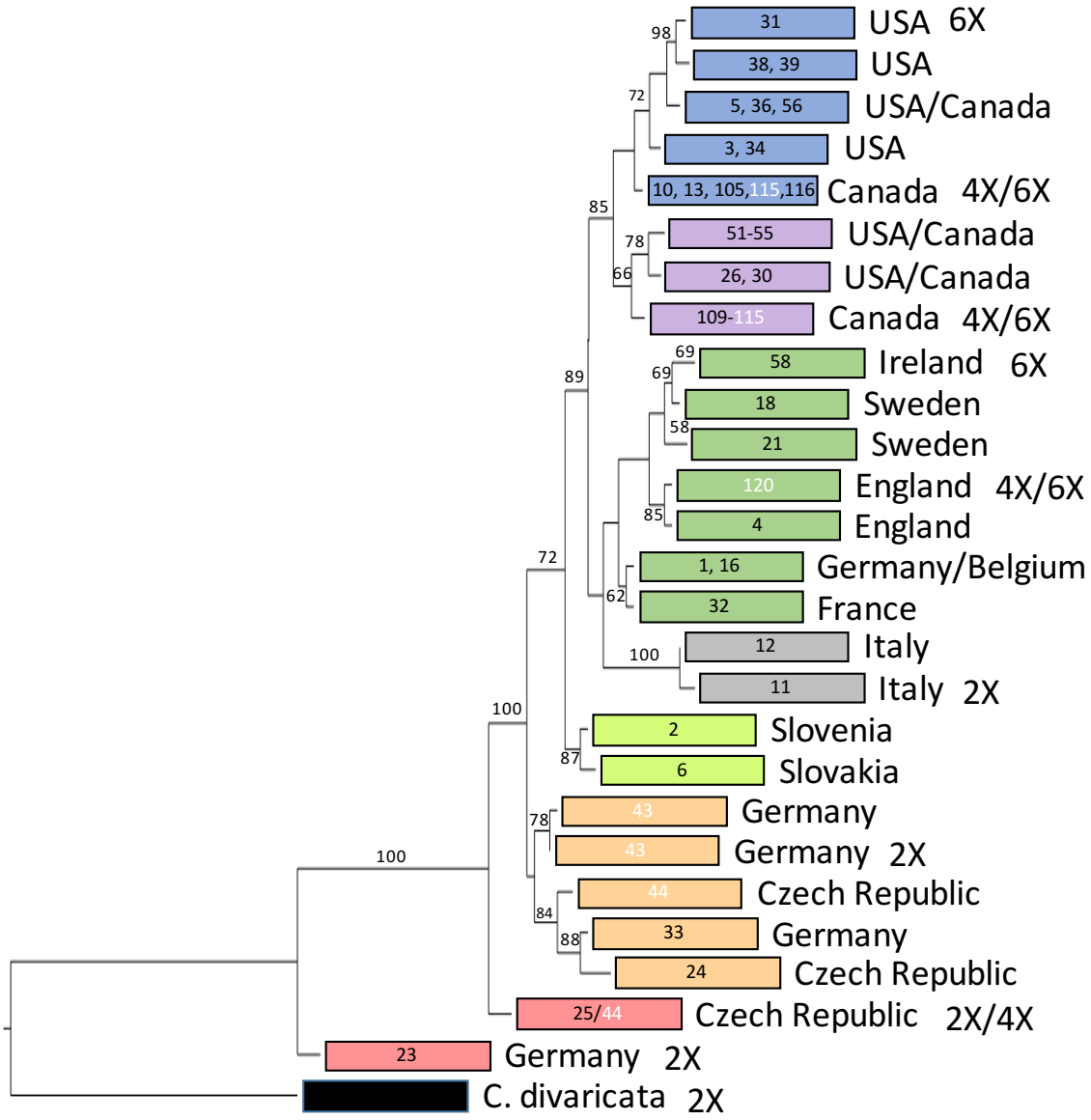


Figure 2) Map of sampled *Campanula rotundifolia* populations. Colors correspond to clades in the chloroplast phylogeny (Figure 1), while numbers indicate populations in Table 1. a) Map of North American populations, with inset map (b) showing populations around the Great Lakes. c) Map of European populations.

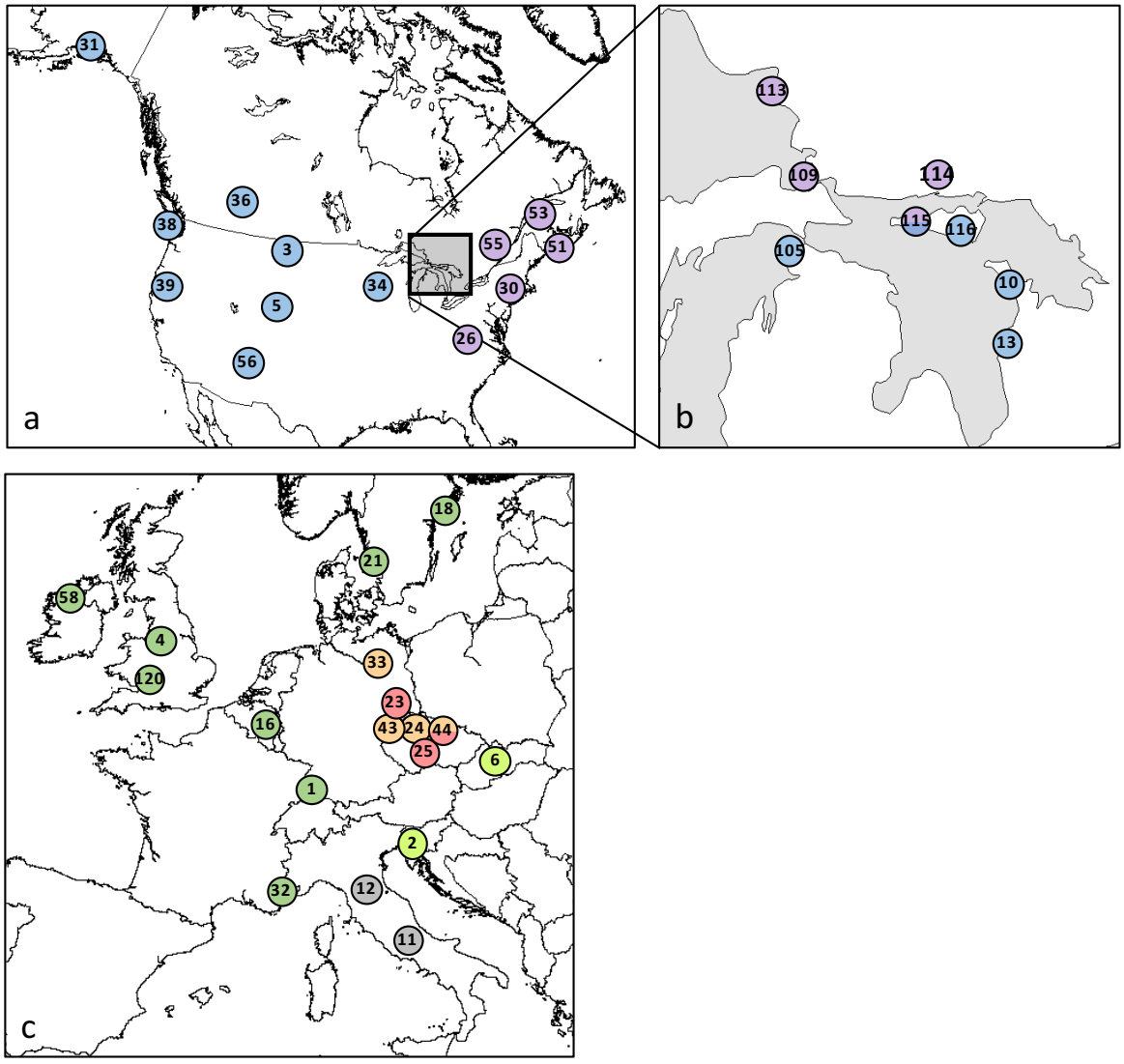


Figure 3) Nuclear maximum likelihood phylogeny for *Campanula rotundifolia* based on restriction-site associated (RAD) sequences. Colors correspond to clades in the chloroplast phylogeny (Figure 1) and numbers refer populations in Table 1. Numbers at nodes indicate bootstrap support. All populations tetraploid except as labelled.

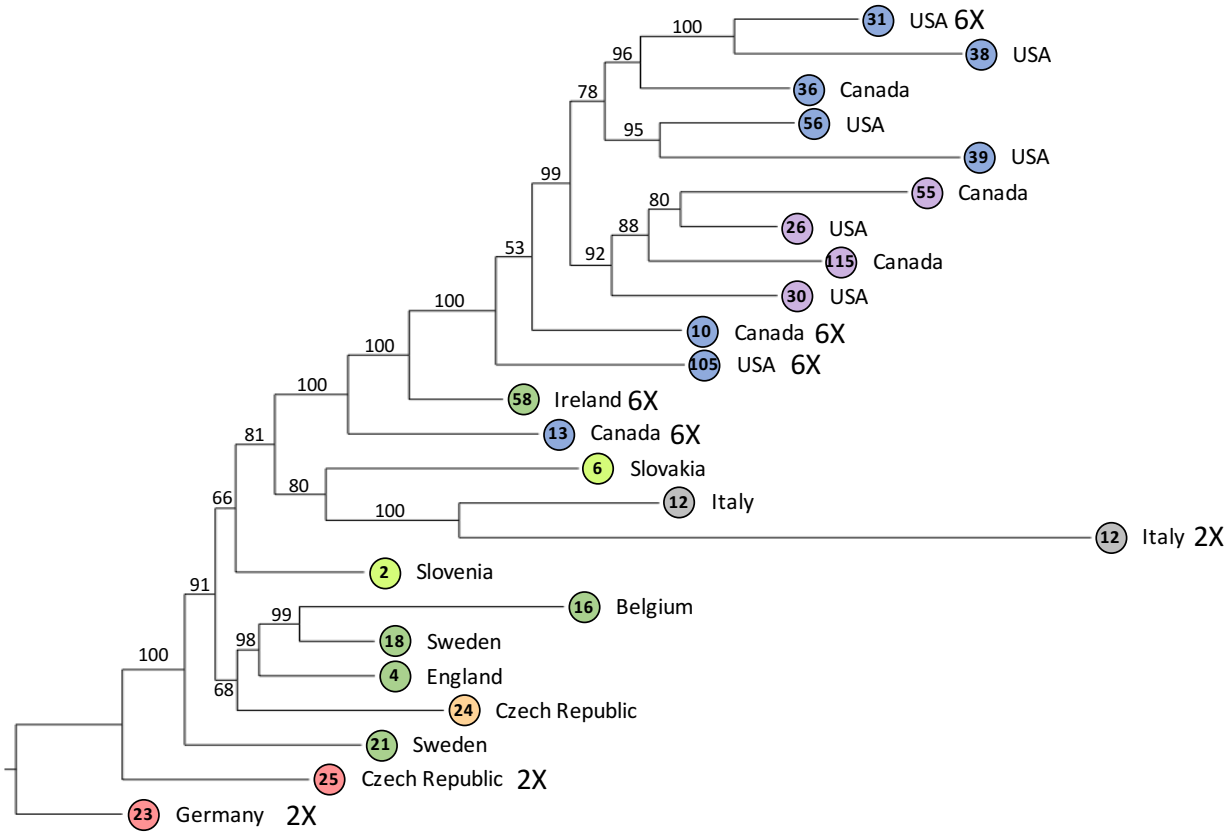


Figure 4) Map of hypothesized migration routes of North American (a) and European (b) populations of *Campanula rotundifolia*. Arrow colors denote clades derived from chloroplast phylogeny (Figures 1,2). Grey circles represent hypothesized glacial refugia: north and south of the Alps in Europe, and the Paleozoic plateau in North America.

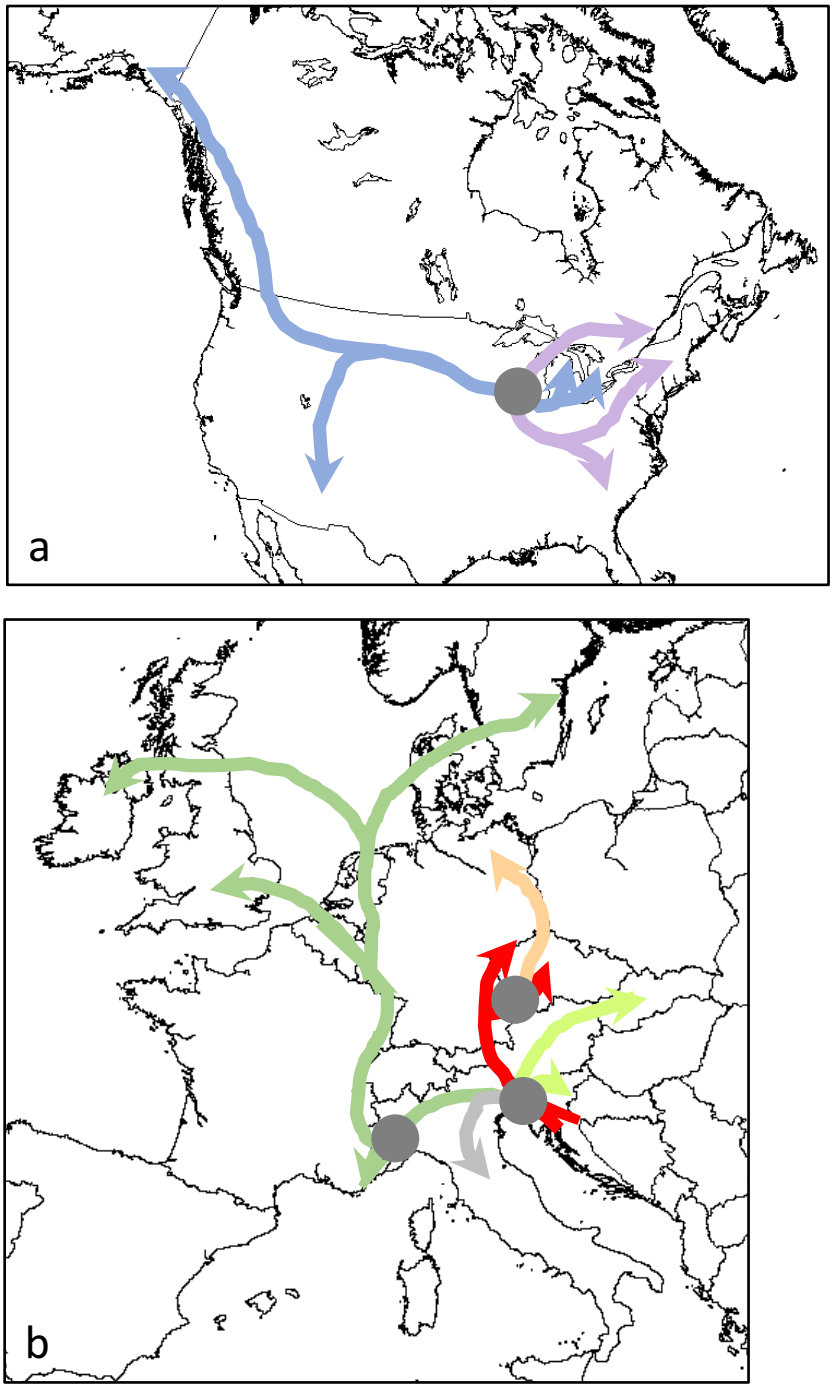
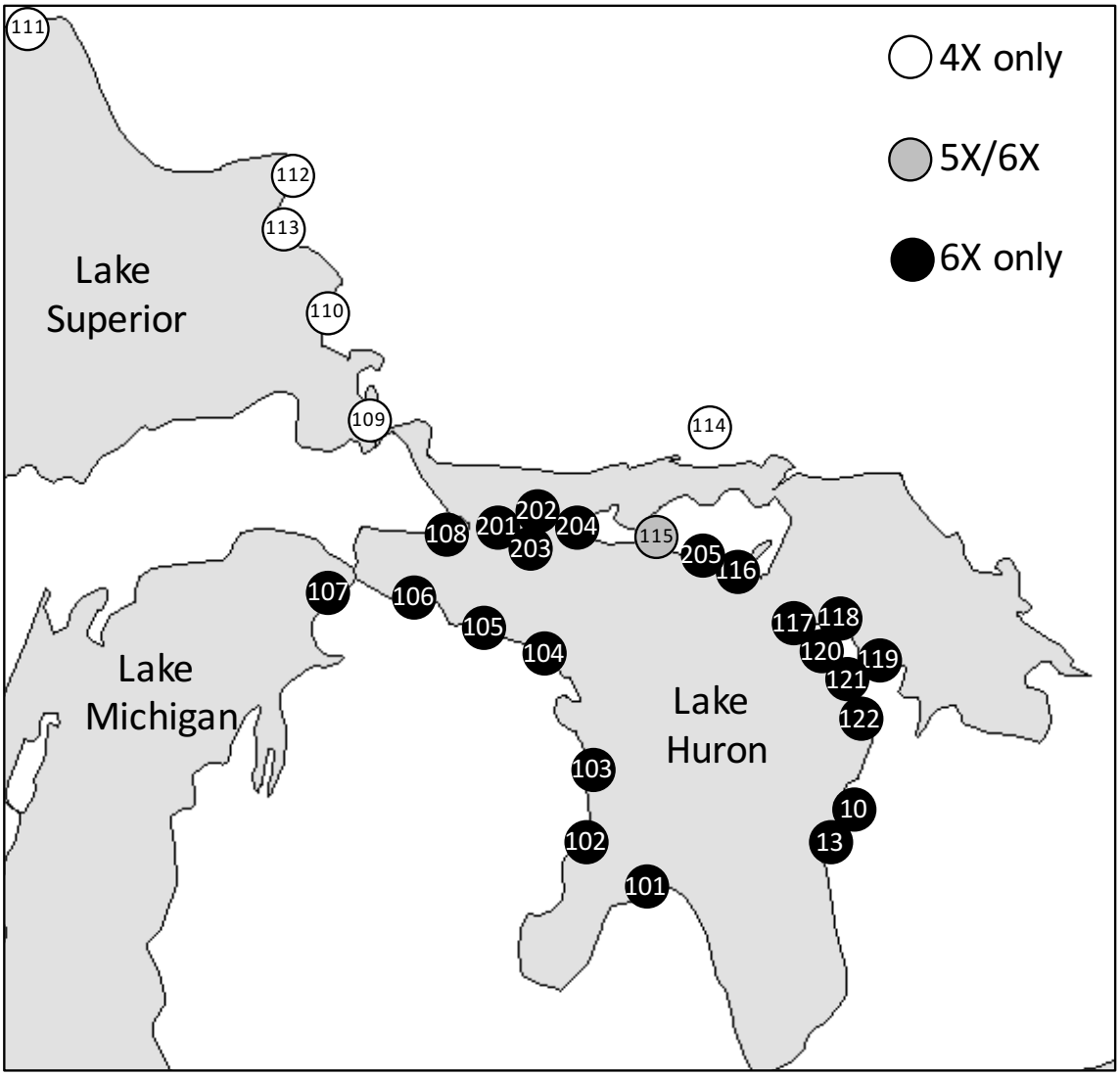


Table S1) Primer pairs derived from *Campanula rotundifolia* whole-chloroplast sequencing. Pairs above the solid line were used to generate the chloroplast DNA dataset.

cpDNA region	Primer Name	Primer Sequence
ndhB	Cro_ndhBF	CCGATTATTTCAATTGCTCAGG
ndhB	Cro_ndhBR	TTGAAGAGGATCCCTGTTAAGC
rpl22-rps23	Cro_rpl22F	TTCGTGTCCTCCAATACGTCC
rpl22-rps23	Cro_rpl22R	ATTCGGGGCCGTTCTTACC
rps3-rpl16	Cro_rps3F	ATCTTCTCAGTCTTTATTGGCTCG
rps3-rpl16	Cro_rps3R	GGTAGGCTTCCCCTACAAACC
rps7-ndhB	Cro_rps7F	ACGGGACGCAATATCTAAGG
rps7-ndhB	Cro_rps7R	TTTTTGATCAGAGGTTGAATCG
ycf-2	Cro_ycf-2aF	GATGGCTGATCAAACTGACG
ycf-2	Cro_ycf-2aR	TCCCCTGTGAAAGACTAATCG
ycf15-rpl23	Cro_ycf15F	AGTCTTAGTTAGTGATCCCGGC
ycf15-rpl23	Cro_ycf15R	CGTCCGGTTCTATCGGTGC
petA-rps4	Cro_petAF	GTTAGTCCCGGTAACGCCC
petA-rps4	Cro_petAR	CTTGGCATCTGTTCTTTTGGC
trnL-ccsA	Cro_ccsAF	GTGAGCAAGCCGCTATGG
trnL-ccsA	Cro_ccsAR	TCATGCTTACGTGCATTATTAACC
trnV-I	Cro_trnV-I F	GTATTGCTTTCATACGGCGGG
trnV-I	Cro_trnV-I R	CGGACAACACATATAAAGAGACCC

Figure S1: Collection map for targeted sampling of *Campanula rotundifolia* around the North American Great Lakes, Lake Huron and Lake Superior. See Table 1 for sample locations; samples 201-204 are all on Drummond Island and due to space constraints depicted locations are approximate.



CHAPTER TWO:

Dispersal history and whole-genome duplication contribute to loss of self-incompatibility in a polyploid complex

This chapter is formatted as a co-authored manuscript for the American Journal of Botany (Sutherland BL, Quarles BM, Galloway LF) and is under revision.

Premise of the study – Angiosperm species often shift from self-incompatibility to self-compatibility following population bottlenecks. Across the range of a species, population bottlenecks may result from multiple factors, each of which may affect the geographic distribution and magnitude of mating system shifts. In this study, we describe how intercontinental dispersal and genome duplication facilitate loss of self-incompatibility.

Methods – Self and outcross pollinations were performed on plants from 24 populations of the *Campanula rotundifolia* polyploid complex. Populations spanned the geographic distribution and three dominant cytotypes of the species (diploid, tetraploid, hexaploid).

Key results – Loss of self-incompatibility was associated with both intercontinental dispersal and genome duplication. European plants were largely self-incompatible while North American plants were intermediately to fully self-compatible. Within both European and North American populations, loss of SI increased as ploidy increased. Ploidy change and intercontinental dispersal both contributed to loss of SI in North America, but range expansion did not affect SI within Europe or North America.

Conclusions – When species are subject to population bottlenecks arising through multiple factors, each factor can contribute to SI loss. In a widespread polyploid complex, the loss if SI can be predicted by the cumulative effects of whole-genome duplication and intercontinental dispersal.

Introduction

Most angiosperms, though hermaphroditic, are predominantly or exclusively outcrossing (Barrett et al., 1996) to avoid the often high cost of inbreeding depression (Lande and Schemske, 1985; Charlesworth and Charlesworth, 1987). In many plant families, outcrossing is enforced through self-incompatibility, in which self-incompatible (SI) plants recognize and prevent fertilization by related pollen (East, 1926; Bateman, 1952). Although self-incompatibility is common (de Nettancourt, 1977), numerous losses have been documented when selection favors the capacity to self-fertilize (Barrett, 1988; Mable et al., 2005; Busch and Schoen, 2008; Igic et al., 2008). Loss of self-incompatibility frequently occurs following population bottlenecks, when the cost of inbreeding depression declines and reproductive assurance becomes critical to persistence of small, isolated populations (Baker, 1955; Busch and Schoen, 2008; Igic and Busch, 2013; Pannell et al., 2015). Population bottlenecks often arise during range expansion or following long-distance, particularly intercontinental, dispersal. In plants, they can also occur after whole-genome duplication. As drivers of population bottlenecks, each of these processes—range expansion, intercontinental dispersal, and whole-genome duplication—differ in duration, severity, and underlying genetic effects on mating system. In polyploid complexes with expansive or disjunct ranges, one or more of these processes can contribute to overall loss of self-incompatibility. Therefore, to understand mating system evolution in widespread polyploid complexes, it is necessary to assess the individual and cumulative effects of range expansion, intercontinental dispersal, and whole-genome duplication across the cytotypic and geographic range of the complex.

Both range expansion and intercontinental dispersal are known to cause population bottlenecks at range edges (Lande and Schemske, 1985; Porcher and Lande, 2005; Busch et al., 2011; Griffin and Willi, 2014), but expectations for their effects on effective population size differ. Intercontinental dispersal is a relatively rare event that often involves only one or a few founder individuals who experience little to no gene flow with the source population. Population bottlenecks during intercontinental dispersal are often profound, and create a demographic sieve in which individuals who can self are more likely to persist, thereby enriching the frequency of self-compatibility among colonists. This phenomenon has been described as Baker's Law (Baker, 1955; Pannell et al., 2015). However, during range expansion, population bottlenecks are typically less severe and edge populations may experience gene flow with more central populations (Wilson et al., 2009). Here, as leading-edge populations become progressively smaller and more genetically homogeneous, the cost of inbreeding is reduced and the need for reproductive assurance increases. This change in selective pressures can favor self-compatible individuals over obligate outcrossers (Hargreaves and Eckert, 2014). As this process is gradual, it may more time for loss of SI following intracontinental range expansion than following intercontinental dispersal (Hauck, 2006).

Severe population bottlenecks can also occur following whole-genome duplication (Husband et al., 2008). When tetraploids arise in an otherwise diploid population via genome duplication, any reproduction between cytotypes is likely to result in hybrid triploid embryos, which are subject to severe fitness deficits due to low viability and fertility (Miller and Venable, 2000; Mable et al., 2004; Husband et al., 2008). This creates substantial reproductive isolation between diploids and polyploids. Under such

conditions, polyploids experience strong selection against SI (Sabara et al., 2013). Indeed, most emergent polyploids within diploid populations are transient individuals that fall victim to minority cytotype exclusion because most mating attempts result in unfit hybrids, and only those individuals that are self-compatible tend to persist (Husband, 2000; Pannell et al., 2004). This selective pressure for loss of SI is prevalent enough that shifts in mating system have been correlated with changes in ploidy (Entani et al., 1999). In addition to strong selection for self-compatibility, polyploids may also be preadapted for SI loss (Hauck, 2006; Tao and Iezzoni, 2010). Genome duplication has been shown to reduce the efficiency of SI systems, particularly gametophytic self-incompatibility (GSI) systems in which self-recognition is mediated by the haploid pollen genotype (Hauck et al., 2006; Miller et al., 2008).

In polyploid systems that have undergone range expansion or long-distance dispersal, successive bottlenecks may result in step-wise loss of SI. Earlier bottlenecks may actually facilitate later ones. For example, following genome-duplication, a population of polyploids may be more self-compatible than diploids, and thus more likely to establish after intracontinental range expansion or intercontinental dispersal than diploids (Linder and Barker, 2014). Following such dispersal, these potential colonists may experience a severe population bottleneck and be subject to further loss of SI. The facilitation of subsequent dispersal by more self-compatible polyploids is consistent with the overrepresentation of polyploids following intercontinental (Crawford et al., 2009) or island colonizations (te Beest et al., 2012).

The *Campanula rotundifolia* polyploid complex is a tractable system in which to compare the relative effects of range expansion, intercontinental dispersal, and whole-

genome duplication on loss of SI. The complex has a broad circumboreal range encompassing the northern latitudes of Europe and North America, and comprises three dominant cytotypes (Shetler, 1982; Stevens et al., 2012). It is thought to have originated in southeastern Europe approximately 2.3 MYA. It has a history of relatively gradual range expansion throughout Europe, transatlantic dispersal to North America approximately 220 KYA, and subsequent range expansion throughout northern North America (Sutherland, 2017). Tetraploids originated from diploids and hexaploids from tetraploids at least four times each via whole-genome duplication (Sutherland, 2017). The complex is considered self-incompatible (Bielawska, 1973; Kovanda, 1977). However, anecdotal accounts of selfing (Shetler, 1982; Giblin, 2005) suggest that the mating system may be variable.

Here, we seek to understand the patterns of SI loss throughout the geographic distribution of the *C. rotundifolia* polyploid complex. We assess self-incompatibility in multiple European and North American populations that represent all dominant cytotypes to address four questions. Is the *C. rotundifolia* polyploid complex self-incompatible throughout its geographic and cytotypic range? Is range expansion or intercontinental dispersal associated with any loss of self-incompatibility? Is whole-genome duplication associated with any loss of self-incompatibility? Is loss of SI predictable based on the cumulative effects of range expansion, intercontinental dispersal, and genome duplication?

Materials and methods

Study species

Campanula rotundifolia L. is a short-lived perennial wildflower with a circumboreal distribution covering much of the northern latitudes of Europe and North America (Shetler, 1982; Stevens et al., 2012). It exists as an autopolyploid complex comprising diploids in Central and Northern Europe, tetraploids throughout the distribution, and hexaploids in the British Isles and North America (Shetler, 1982; Stevens et al., 2012, Sutherland, 2017). *Campanula rotundifolia* is protandrous; prior to anthesis, anthers deposit pollen onto stylar hairs which results in secondary pollen presentation. The stigmatic lobes remain closed for the first 24-48 hours post-anthesis. Pollinator visitation stimulates stylar hairs to retract and stigmatic lobes to open and begin curling under (Shetler, 1982), placing receptive stigmas closer to their own pollen.

The mechanism of self-incompatibility in *C. rotundifolia* is not known. However, the congener *C. rapunculoides* exhibits a GSI system in which self-recognition is mediated by coat protein alleles in the haploid pollen grain (Good-Avila et al., 2008) and it is likely *C. rotundifolia* shares this mechanism.

We chose 24 populations of *C. rotundifolia* that spanned the geographic range and ploidy variation present within the complex (Figure 1), and comprised 11 European and 13 North American populations (Table S1). To assess the effect of ploidy variation on SI, we included all dominant cytotypes. In Europe, these samples comprised three diploid, seven tetraploid, and one hexaploid population, and in North America accessions comprised eight tetraploid and five hexaploid populations. Although diploids have been reported in North America (Löve and Löve, 1966), multiple attempts to resample a

diploid population in the northeastern U.S. were unsuccessful, and the population may be extirpated. Populations were sampled across the three main clades identified in a chloroplast phylogeny (Sutherland, 2017) that roughly correspond to three locations: Central and Southern Europe, Western and Northern Europe, and North America (Figure 1). This sampling design permitted evaluating the effects of both gradual range expansion (Central/Southern to Western/Northern Europe and within North America) and intercontinental dispersal (Western/Northern Europe to North America). Additional North American populations were chosen as well as to assess SI in a narrowly defined contact zone between tetraploids and hexaploids along the Great Lakes (Figure 1b). Ploidy was assessed for greenhouse-grown plants of all populations using a modified Otto 2-step protocol (Otto, 1990, Sutherland and Galloway, 2017).

Crossing Method

Crosses were performed on greenhouse plants grown from wild-collected seed. We performed one outcross and one self-cross per plant for an average of 10.7 (± 4.2 S.D.) plants for each population (minimum 6; 248 total pairs of self-crosses and outcrosses). Pollinated flowers were surgically emasculated in the bud prior to anther dehiscence to prevent potential within-flower selfing. Preliminary crosses showed no difference in seed set between intact maternal flowers and those for which anthers had been removed. Once stigmas opened, pollen-bearing styles of paternal flowers were rubbed on stigmatic lobes of maternal flowers, depositing an excess of pollen. For outcrosses, paternal flowers were chosen from different plants within the same population, and for self crosses, paternal flowers were chosen from a different flowering stem on the same plant. Mature fruits were collected prior to dehiscence, and fully

developed brown, inflated seeds were counted. Flowers that failed to set fruit were recorded as seed set of 0. If outcrosses set fewer than 10 seed, indicating poor pollination success (6.6% of crosses), they were replaced along with their matched self-cross by an additional pair of crosses. No outcrosses set fewer than 10 seed twice consecutively.

Statistical analysis

Self-incompatibility was scored by calculating the index of self-incompatibility (ISI; Lloyd, 1969), or one minus the ratio of seeds produced by self-fertilization relative to outcrossed seed produced. Although fruit set is more commonly used to calculate ISI (Raduski et al., 2011), we used seed set because polyploids that display GSI often have multiple pollen coat genotypes in one paternal flower, some of which may successfully self-fertilize ovules (Tao and Iezzoni, 2010). This variability leads to high fruit formation, but reduced seed formation. ISI was calculated for each plant, and can range from 0 to 1. If selfed seed exceeded outcrossed seed for an individual, the individual was deemed fully SC and ISI was set to one. Individuals have historically been classified as SI, intermediately SI, or SC based on threshold values (Lloyd, 1969); SI individuals have ISI > 0.8, and SC individuals have ISI < 0.2.

To evaluate the effect of ploidy, range expansion and intercontinental dispersal on self-incompatibility, we used a generalized mixed model PROC GLIMMIX (SAS 9.3 SAS Institute, INC. 2011). A logit transformation of seed set ratio (selfed/outcrossed seed per plant) was performed to normalize the data prior to calculating ISI. The model included ploidy and location as fixed effects, and population nested within ploidy and location as a random effect. Location was treated as a categorical variable with three levels (Central/Southern Europe, Western/Northern Europe, and North America). Ploidy

is non-randomly distributed throughout the range, with most hexaploids in North America and all diploids in Central Europe (Figure 1). Because of this distribution, the model was unavoidably unbalanced, and as such, an interaction term for these factors was not included in the model.

To address this imbalance and to test the effect of range expansion apart from intercontinental dispersal, analyses were also performed on subsets of the data. First, the effect of ploidy was tested within each continent. All three cytotypes were compared in Europe, and North America was limited to tetraploids and hexaploids. Populations were compared using the same mixed model described above, but omitting location as a main effect. Second, tetraploid populations occur throughout the range and are fairly evenly distributed, permitting a test of the relationship between dispersal distance and SI. Specifically, linear regressions were performed to test the effect of distance (as a continuous variable) from the hypothesized center of origin on ISI, along hypothesized routes of migration (Sutherland, 2017). To minimize the effect of potential leverage on these regressions from having clusters of populations in North America and Europe, the distance between nearest European and North American populations was set equal to the longest distance between any other two adjacent populations (approx. 1600 km). Three regressions were performed on population mean ISI: all tetraploid populations, only European tetraploids, and only North American tetraploids.

Results

Mean population ISI ranged from self-incompatible to self-compatible, with values from 0.915 to 0.131 (Figure 2a), and location was strongly associated with ISI ($F_{2,19}$=10.17, P<0.001). Central European populations were self-incompatible, with an

average ISI of 0.847 (Figure 2b), while Western European populations had a slightly lower average ISI of 0.747, but this difference was not significant (Tukey-Kramer, $t = -0.97$, $p = 0.60$). North American populations had a significantly lower average ISI of 0.418 (Tukey-Kramer, $t = -4.50$, $p < 0.001$), indicating partial self-compatibility. One North American population, Michigan 105, was fully self-compatible with an ISI of 0.140 (Figure 2a). When analysis was limited to tetraploid populations, distance from the hypothesized center of origin was associated with decreasing self-incompatibility ($t=-4.97$, $P<0.001$, $\beta=-0.799$, $R^2=0.638$; Figure 3a). However, when samples were limited to either Europe or North America, distance was not significantly associated with loss of SI (Europe: $t=-1.08$, $P=0.323$, $\beta=-0.402$, $R^2=0.162$; North America: $t=0.28$, $P=0.788$, $\beta=0.114$, $R^2=0.013$; Figure 3b,c).

Increasing ploidy was associated with decreasing self-incompatibility (Figure 2). Diploids were self-incompatible with ISI values between 0.883 and 0.866 (mean ISI = 0.880). Tetraploid ISI values ranged from 0.915 to 0.301, displaying high variance in self-incompatibility (mean ISI = 0.709). Hexaploids were partially to fully self-compatible with ISI values between 0.529 and 0.131 (mean ISI = 0.361). This negative association between self-incompatibility and ploidy was near significant for the whole dataset ($F_{2,19}=3.38$, $P=0.056$), with tetraploids not significantly different from either diploids or hexaploids, but hexaploids having significantly lower SI than diploids (Tukey-Kramer, $t = -2.52$, $p = 0.033$). When ploidy was tested within each continent, increasing ploidy was significantly associated with decreasing self-incompatibility in both Europe and North America (Europe: $F_{2,11}= 3.82$, $P=0.032$; North America: $F_{1,12}=5.56$, $P=0.020$; Figure 2).

Discussion

The *Campanula rotundifolia* polyploid complex is not uniformly self-incompatible, and loss of self-incompatibility is consistent with a history of both whole-genome duplication and intercontinental dispersal, but not intracontinental range expansion. Some populations are strongly SI, however most have experienced some loss of self-incompatibility, which contrasts with historical accounts of *C. rotundifolia*'s mating system (Bielawska, 1973; Ægisdóttir and Thórhallsdóttir, 2006). SI loss was most pronounced in North American populations and in hexaploids, with almost complete SI loss in some North American hexaploids, which reflects the cumulative effects of bottlenecks caused by intercontinental dispersal and WGD.

SI loss in *C. rotundifolia* has been influenced by a history of intercontinental dispersal, but not gradual range expansion. Based on nuclear (Mansion et al., 2012, Sutherland, 2017) and chloroplast (Sutherland, 2017) phylogenies, *C. rotundifolia* is hypothesized to have originated in southcentral Europe. From that center of origin, the complex has undergone gradual range expansion in Europe, intercontinental dispersal to North America, then gradual range expansion in North America. However, SI loss was only significantly associated with intercontinental dispersal between Europe and North America. While Western European populations had slightly lower ISI than Central Europe, this loss was modest and not significant. In contrast, ISI in North America was 46% lower than in Europe. Although the hypothesized single dispersal event from Europe to North America (Sutherland, 2017) makes it difficult to definitively ascribe loss of SI to intercontinental dispersal, loss of SI following long-distance dispersal events, including intracontinental dispersal (Šingliarová et al., 2008), intercontinental dispersal (Costa et

al., 2017), and island colonization (Grossenbacher et al., 2017), has been observed across disparate taxa, in both diploid and polyploid systems (e.g. Barrett, 2015). These similar findings in multiple taxa support the conclusion that intercontinental dispersal has led to a reduction in self-incompatibility in *C. rotundifolia*.

The lack of significant loss of SI within European tetraploids or North American tetraploids is consistent with bottlenecks that are less severe following range expansion than following intercontinental dispersal. European taxa expanded after glacial maximum from refugia both north and south of the Alps (Sutherland, 2017), and likely experienced milder population contractions at the leading edge while still maintaining some gene flow with the core distribution (e.g. Wilson et al., 2009). In contrast, dispersal from Europe to North America required transoceanic migration over at least 5800 km, effectively cutting off gene flow from source populations (e.g. Wilson et al., 2009). This abrupt lack of gene flow coupled with small population sizes among colonists would have more strongly selected for reproductive assurance than in the leading edge populations of range expansion, and is a possible explanation for the considerable loss of SI observed in North American taxa.

Whole-genome duplication was also associated with SI loss. Diploid populations were strongly self-incompatible while tetraploid populations had ISI values 22% lower. Hexaploids were weakly SI to SC, with ISI values 42% lower than tetraploids and 60% lower than diploids. Given that newly established polyploids appear to undergo extreme population bottlenecks (Husband et al., 2008; Sabara et al., 2013), strong selection for the capacity to self is expected in order to provide reproductive assurance (Barringer, 2007). Furthermore, each successive genome duplication (for example, from tetraploidy to

hexaploidy) presents a new potential bottleneck event that can select for loss of self-incompatibility. The SI loss seen in tetraploid and hexaploid *C. rotundifolia* is consistent with selection for reproductive assurance during population bottlenecks (Guo et al., 2009) caused by successive rounds of genome duplication (Mable, 2004).

Alternatively, loss of SI following genome duplication may be due to breakdown of gametophytic self-incompatibility. Gametophytic self-incompatibility relies upon linked pollen coat proteins and stylar RNAses to prevent growth of pollen tubes from related pollen (Tao and Iezzoni, 2010). In diploid plants, haploid pollen express a single GSI coat protein, and if that protein is recognized by stylar RNAses, pollen tube arrest is complete or near-complete (Luu et al., 2000). However, recognition efficiency decreases when multiple coat proteins are present, which is possible in heterozygous pollen from tetraploids and hexaploids (Hauck et al., 2006). Because heterozygous autotetraploids will produce a mixture of heteroallelic and homoallelic pollen, self-fertilization will be reduced, but not eliminated. In hexaploids, a higher proportion of heteroallelic pollen is possible, resulting in a further decrease in pollen tube recognition and arrest. As a result, SI may be lost cumulatively through subsequent genome duplication events. This mechanism is consistent with loss of SI observed in *C. rotundifolia*, with progressively lower ISI values as ploidy increases.

GSI breakdown may also explain the unexpectedly high frequency of intermediate ISI values observed in *C. rotundifolia*. Although mixed mating systems are well-documented, in most angiosperms self-incompatibility fits a bimodal distribution tending toward either SI or SC (Raduski et al., 2011). Intermediate SI has been considered a transient state (Schemske and Lande, 1985). In contrast, only *C. rotundifolia* diploids are

universally SI; most tetraploid and hexaploid populations exhibit intermediate ISI values. The high proportion of intermediate ISI values throughout the distribution suggests that the capacity for outcrossing has been maintained in this system long-term, and that polyploids may not fit the broader pattern of bimodality observed for SI among diploids. Intermediate values of ISI may reflect GSI breakdown associated with ploidy change. As such, changes in SI are not associated with selection on mating system that may favor outcrossing or selfing, they may be more likely to be stable.

Patterns of SI loss in *C. rotundifolia* are consistent with a stepwise process in which loss of SI accumulates over successive population bottleneck events. Early in its history, *C. rotundifolia* experienced a slight loss of SI upon duplication from diploidy to tetraploidy in Central Europe. This was followed by a larger loss of SI upon ploidy change from tetraploid to hexaploid. These losses in SI in European tetraploids and hexaploids were compounded upon intercontinental dispersal, with North American tetraploids showing considerable loss of SI relative to European tetraploids, and North American hexaploids becoming mostly or completely self-compatible. This stepwise loss of self-incompatibility demonstrates that intercontinental dispersal and genome duplication both act to induce loss of SI.

Not only has WGD and intercontinental dispersal resulted in step-wise loss of SI, some bottleneck events may have been facilitated by those that preceded them. As Baker's law suggests, individuals who are more self-compatible have a higher likelihood of establishment following long-distance dispersal; therefore, colonist populations are more likely to have higher self-incompatibility overall following a demographic sieve. Loss of SI in European tetraploids and hexaploids may have increased their likelihood of

successful establishment following intercontinental dispersal, thereby setting the stage for individuals from these reduced-SI populations to undergo another population bottleneck (and subsequent loss of SI) following dispersal. Polyploid complexes often have broad ranges, and many complexes are known to span continents (Bleeker et al., 2002; Al-Shehbaz et al., 2006; Brochmann and Brysting, 2008; Montesinos et al., 2012). As such, successive bottlenecks that facilitate predictable step-wise loss of SI may be common as these lineages undergo genome duplication and long-distance dispersal events.

The high variation in self-incompatibility in *C. rotundifolia* and pattern of SI loss with ploidy increase and intercontinental dispersal suggest two things. First, loss of SI following genome duplication may be partial, and populations may persist with intermediate self-incompatibility. Second, while gradual range expansion appears to play little to no role in SI loss in *C. rotundifolia*, both intercontinental dispersal and whole-genome duplication appear to have contributed, independently and successively, to SI loss. The repeated effects of whole-genome duplication inducing SI loss in a putatively GSI system, followed by enrichment for self-compatible individuals in colonist populations appears to have mediated wide variation in mating system across the range of *C. rotundifolia*. The stepwise changes in self-incompatibility observed in this system are not consistent with either whole-genome duplication or intercontinental dispersal alone, but can be predicted as the result of both processes acting consecutively. To fully understand mating system evolution in widespread polyploid complexes, it is critical to determine how changes in both distribution and cytotype influence each other in the maintenance or loss of self-incompatibility.

Acknowledgements

We sincerely thank B. Cottrell for help with planning and executing a pilot version of this study; T. Miranda-Katz for help in the greenhouse; E. Nagy and the Mountain Lake Biological Station for technical and logistical support; M. Koski, C. Debban, K. Barnard-Kubow, and A. Bangerter for comments on an earlier version of the manuscript and NSF for funding (DBI-1461169 and DEB-1457686).

References

Ægisdóttir, H. H., and T. E. Thórhallsdóttir. 2006. Breeding system evolution in the Arctic: a comparative study of *Campanula uniflora* in Greenland and Iceland. *Arctic* 38: 306-312.

Al-Shehbaz, I. A., M. A. Beilstein, and E. A. Kellogg. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution* 259: 89-120.

Baker, H. G. 1955. Self-compatibility and establishment after "long-distance" dispersal. *Evolution* 9: 347-349.

Barrett, S. C. H., L. D. Harder, and A. C. Worley. 1996. The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions: Biological Sciences* 351: 1271-1280.

Barrett, S. C. H. 1988. The evolution, maintenance, and loss of self-incompatibility systems. In: *Plant Reproductive Ecology* (ed. Lovett Doust J), pp. 98–124. Oxford University Press, Oxford.

Barrett, S. C. H., 2015. Foundations of invasion genetics: the Baker and Stebbins legacy. *Molecular Ecology* 24: 1927-1941.

Barringer, B. C. 2007. Polyploidy and self-fertilization in flowering plants. *American Journal of Botany* 94: 1527-1533.

Bateman, A. J. 1952. Self-incompatibility systems in angiosperms: I. Theory. *Heredity* 6: 285-310.

te Beest, M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda, M. Kubešová, and P. Pyšek. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109: 19-45.

Bielawska, H. 1973. Self-fertilization in *Campanula rotundifolia* L. s.l. group. *Acta Societatis Botanicorum Poloniae* 42: 253–264.

Bleeker, W., and H. Hurka. 2001. Introgressive hybridization in *Rorippa* (Brassicaceae): gene flow and its consequences in natural and anthropogenic habitats. *Molecular Ecology* 10: 2013-2022.

Brochmann, C., and A. K. Brysting. 2008. The arctic - an evolutionary freezer? *Plant Ecology and Diversity* 1: 181-195.

Busch, J. W., and D. J. Schoen. 2008. The evolution of self-incompatibility when mates are limiting. *Trends in Plant Science* 13: 128–136.

Busch, J. W., S. Joly, and D. J. Schoen. 2011. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. *Molecular Biology and Evolution* 28: 1717–1729.

Charlesworth, D. and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* 18:237-268.

Costa, J., V. Ferrero, M. Castro, J. Louriero, L. Navarro, and S. Castro. 2017. Variation in the incompatibility reactions in tristylous *Oxalis pes-caprae*: large-scale screening in South African native and Mediterranean basin invasive populations. *Perspectives in Plant Ecology, Evolution, and Systematics* 24: 5-36.

Crawford D. J., T. K. Lowrey, G. J. Anderson, G. Bernardello, A. Santos-Guerra, and T. F. Stuessy. 2009. Genetic diversity in Asteraceae endemic to oceanic islands: Baker's law and polyploidy. In: Funk V. A., Susanna A., Stuessy T. F., and R. J. Bayer, eds. Systematics, evolution, and biogeography of Compositae. Vienna, Austria: International Association for Plant Taxonomy, 139–151.

East, E. M. 1926. The physiology of self-sterility in plants. *Journal of General Physiology* 8: 403-416.

Entani, T., S. Takayama, M. Iwano, H. Shiba, F. S. Che, and A. Isogai. 1999. Relationship between polyploidy and pollen self-incompatibility phenotype in *Petunia hybrida* Vilm. *Bioscience, Biotechnology, and Biochemistry* 63: 1882–1888.

Giblin, D. E. 2005. Variation in floral longevity between populations of *Campanula rotundifolia* (*Campanula*ceae) in response to fitness accrual rate manipulation. *American Journal of Botany* 92: 1714–1722.

Good-Avila, S. V., D. Majumder, H. Amos, and A. G. Stephenson. 2008. Characterization of self-incompatibility in *Campanula rapunculoides* (*Campanula*ceae) through genetic analyses and microscopy. *Botany* 86: 1–13.

Griffin, P. C., and Y. Willi. 2014. Evolutionary shifts to self-fertilisation restricted to geographic range margins in North American *Arabidopsis lyrata*. *Ecology Letters* 17: 484–490

Grossenbacher, D. L., Y. Brandvain, J. R. Auld, M. Burd, P.-O. Cheptou, J. K. Conner, A. G. Grant, S. M. Hovick, J. R. Pannell, A. Pauw, T. Petanidou, A. M. Randle, R. Rubio de Casas, J. Vamosi, A. Winn, B. Igic, J. W. Busch, S. Kalisz, and E. E. Goldberg. 2017. Self-compatibility is over-represented on Islands. *New Phytologist.* 215: 469–478.

Guo, Y.-L., J. S. Bechsgaard, T. Slotte, B. Neuffer, M. Lascoux, D. Weigel, and M. H. Schierup. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*,

associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences* 106: 5246–5251.

Hargreaves, A. L., and C. G. Eckert. 2014. Evolution of dispersal and mating systems along geographic gradients: implications for shifting ranges. *Functional Ecology* 28: 5-21.

Hauck, N. R. 2006. Accumulation of nonfunctional s-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid *Prunus*. *Genetics* 172: 1191–1198.

Husband, B. C. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings of the Royal Society B: Biological Sciences* 26: 217–223.

Husband, B. C., B. Ozimec, S. L. Martin, and L. Pollock. 2008. Mating consequences of polyploid evolution in flowering plants: current trends and insights from synthetic polyploids. *International Journal of Plant Sciences* 169: 195–206.

Igic, B., and J. W. Busch. 2013. Is self-fertilization an evolutionary dead end? *New Phytologist* 198: 386–397.

Igic, B., R. Lande, and J. R. Kohn. 2008. Loss of self-incompatibility and its evolutionary consequences. *International Journal of Plant Sciences*. 169: 93–104.

Kovanda, M. 1966. Some chromosome counts in the *Campanula rotundifolia* complex II. *Folia Geobotanica & Phytotaxonomica* 13: 268–273.

Kovanda, M. 1977. Polyploidy and variation in the *Campanula rotundifolia* complex. Part II. Taxonomic 2. Revision of the groups Vulgares and Scheuchzerianae in Czechoslovakia and adjacent regions. *Folia Geobotanica & Phytotaxonomica* 22: 23–89.

Lande, R., and D. W. Schemske. 1985. The evolution of self-fertilization and inbreeding depression in plants. 1. Genetic models. *Evolution* 39: 24–40.

Lloyd, D. G. 1965. Evolution of self-compatibility and racial differentiation in *Leavenworthia* (Cruciferae). *Contributions from the Gray Herbarium of Harvard University* 195: 3-134.

Löve A, and D. Löve. 1966. Cytotaxonomy of the alpine vascular plants of Mount Washington. Series in Biology. Paper 38. Boulder, CO, USA: University of Colorado Press.

Luu, D.-T., X. Qin, D. Morse, and M. Cappadocia. 2000. S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature* 407: 649-651.

Mable, B. K. 2004. Polyploidy and self-compatibility: is there an association? *New Phytologist* 162: 803–811.

Mable, B. K., J. Beland, and C. Di Berardo. 2004. Inheritance and dominance of self-incompatibility alleles in polyploid *Arabidopsis lyrata*. *Heredity* 93: 476–486.

Mable, B. K., A. V. Robertson, S. Dart, C. Di Berardo, and L. Witham. 2005. Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* 59: 1437–1448.

Mansion G., G. Parolly, A. A. Crowl, E. Mavrodiev, N. Cellinese, M. Oganesian, K. Fraunhofer, G. Kamari, D. Phitos, R. Haberle, G. Akaydin, N. Ikinci, T. Raus, and T. Borsch. 2012. How to handle speciose clades? mass taxon-sampling as a strategy towards illuminating the natural history of *Campanula* (Campanuloideae). *PLOS One* http://dx.doi.org/10.1371/journal.pone.0050076

Miller, J. S., and D. L. Venable. 2000. Polyploidy and the evolution of gender dimorphism in plants. *Science* 289: 2335-2338.

Miller, J. S., R. A. Levin, and N. M. Feliciano. 2008. A tale of two continents: Baker's rule and the maintenance of self-incompatibility in *Lycium (*Solanaceae*)*. *Evolution* 62: 1052–1065.

Montesinos, D., G. Santiago, and R. M. Callaway. 2012. Neo-allopatry and rapid reproductive isolation. *The American Naturalist* 180: 529-533.

de Nettancourt, D. 1977. The genetic basis of self-incompatibility. Monographs on theoretical and applied genetics. In: *Incompatibility in angiosperms.* Berlin: Springer, 28–57.

Otto, F. J. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewickz Z, Crissman HA, eds. Methods in cell biology, vol. 33. San Diego, CA, USA: Academic Press, 105–110.

Pannell, J. R., J. R. Auld, Y. Brandvain, M. Burd, J. W. Busch, P.-O. Cheptou, J. K. Conner, E. E. Goldberg, A.-G. Grant, D. L. Grossenbacher, S. M. Hovick, B. Igic, S. Kalisz, T. Petanidou, A. M. Randle, R. R. de Casas, A. Pauw, J. C. Vamosi, and A. A. Winn. 2015. The scope of Baker's law. *New Phytologist*. 208: 656–667.

Pannell, J. R., D. J. Obbard, and R. Buggs. 2004. Polyploidy and the sexual system: what can we learn from *Mercurialis annua*? *Biological Journal of the Linnean Society* 82: 547–560.

Porcher, E., and R. Lande. 2005. Loss of gametophytic self-incompatibility with evolution of inbreeding depression. *Evolution* 59: 46–60.

Raduski, A. R., E. B. Haney, and B. Igic. 2012. The expression of self-incompatibility in angiosperms is bimodal. *Evolution* 66: 1275–1283.

Sabara, H. A., P. Kron, and B. C. Husband. 2013. Cytotype coexistence leads to triploid hybrid production in a diploid-tetraploid contact zone of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 100: 962–970.

Schemske, D. W., and R. Lande. 1985. The evolution of self-fertilization and inbreeding depression in plants. II. Empirical observations. *Evolution* 39: 41-52.

Shetler, S. G. 1982. Variation and evolution of the nearctic Harebells (*Campanula* sect. Heterophylla). J. Cramer. Vaduz.

Šingliarová, B., J. Chrtek Jr., and P. Mráz. 2008. Loss of genetic diversity in isolated populations of an alpine endemic *Pilosella alpicola* subsp. *ullepitschii*: effect of long-term vicariance or long-distance dispersal? *Plant Systematics and Ecology* 275: 181-191.

Stevens, C. J., J. Wilson, and H. A. McAllister. 2012. Biological flora of the British Isles: *Campanula rotundifolia*. *Journal of Ecology* 100: 821–839.

Sutherland, B. S. 2017. Interploid reproductive isolation in the *Campanula rotundifolia* polyploid complex. Ph.D. Dissertation, University of Virginia, Charlottesville, Virginia, USA.

Sutherland, B. S., and L. F. Galloway. 2017. Postzygotic isolation varies by ploidy level within a polyploid complex. *New Phytologist* 13: 404-412.

Tao, R., and A. F. Iezzoni. 2010. The S-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic and molecular features. *Scientia Horticulturae* 124: 423–433.

Wilson, J. R. U., E. E. Dormontt, P. J. Prentis, A. J. Lowe, and D. M. Richardson. 2009. Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology and Evolution* 24: 136-144.

Figure 1) Maps of sampled diploid, tetraploid, and hexaploid *Campanula rotundifolia* populations. A) North American distribution of sampled populations, grey square enlarged in B) shows distribution of populations centered around the Great Lakes. C) European distribution of sampled populations; black line denotes partition between populations assigned to Central and Western European locations.

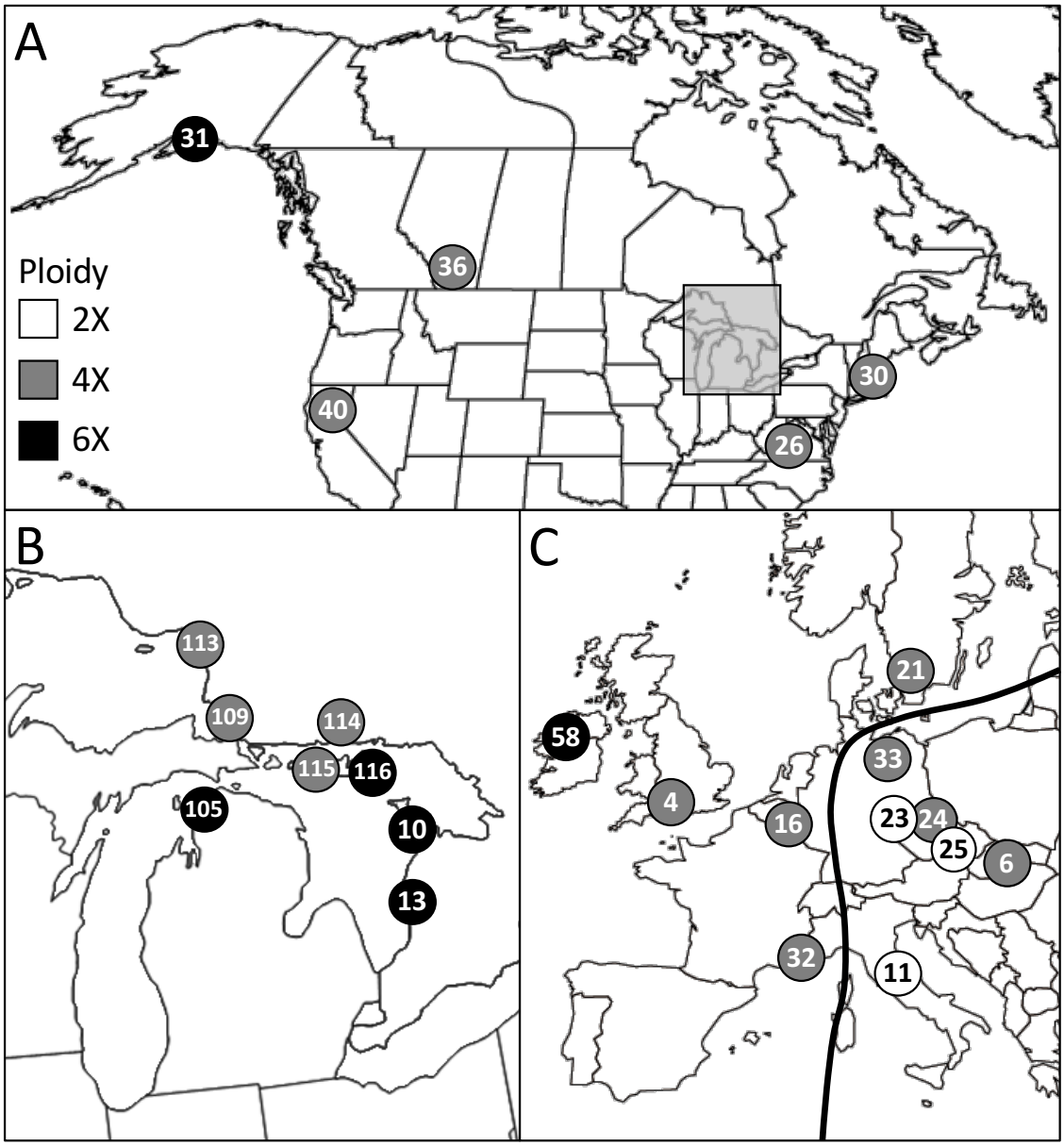
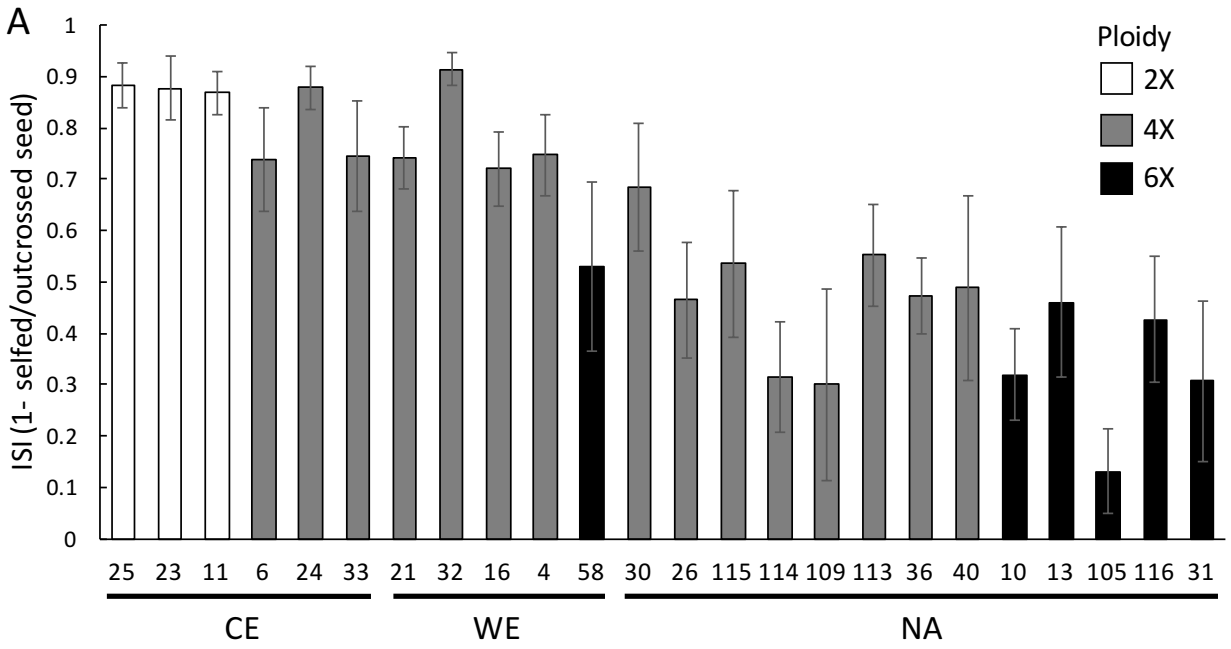


Figure 2) Index of Self-Incompatibility (ISI) in *Campanula rotundifolia* A) by population, and B) by location and ploidy. ISI averaged across all maternal plants in a population; error bars denote standard error. Location categories (CE=Central Europe, WE=Western Europe, NA=North America) are arranged east to west, and populations are arranged by east to west within each location and ploidy. See Figure 1 for population locations.



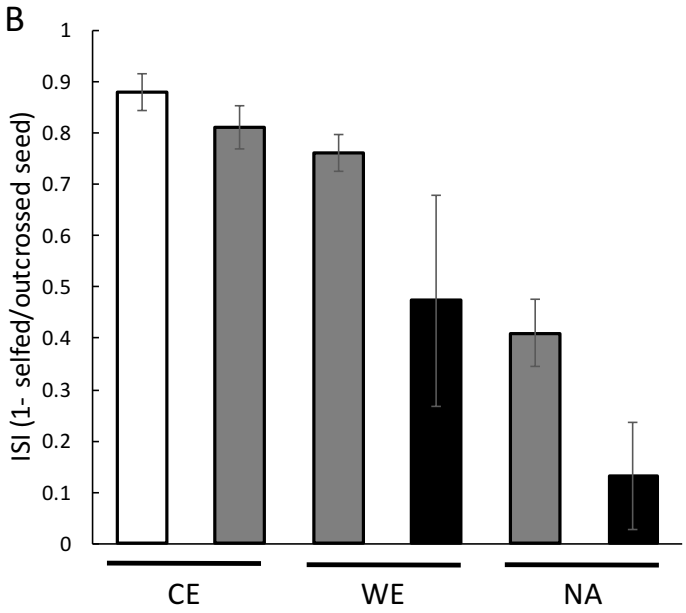


Figure 3) Effect of distance from hypothesized center of origin on Index of Self-Incompatibility (ISI) for tetraploid populations. Linear regression of population means by distance for A) all tetraploid populations, B) populations in Europe only, and C) populations in North America only.

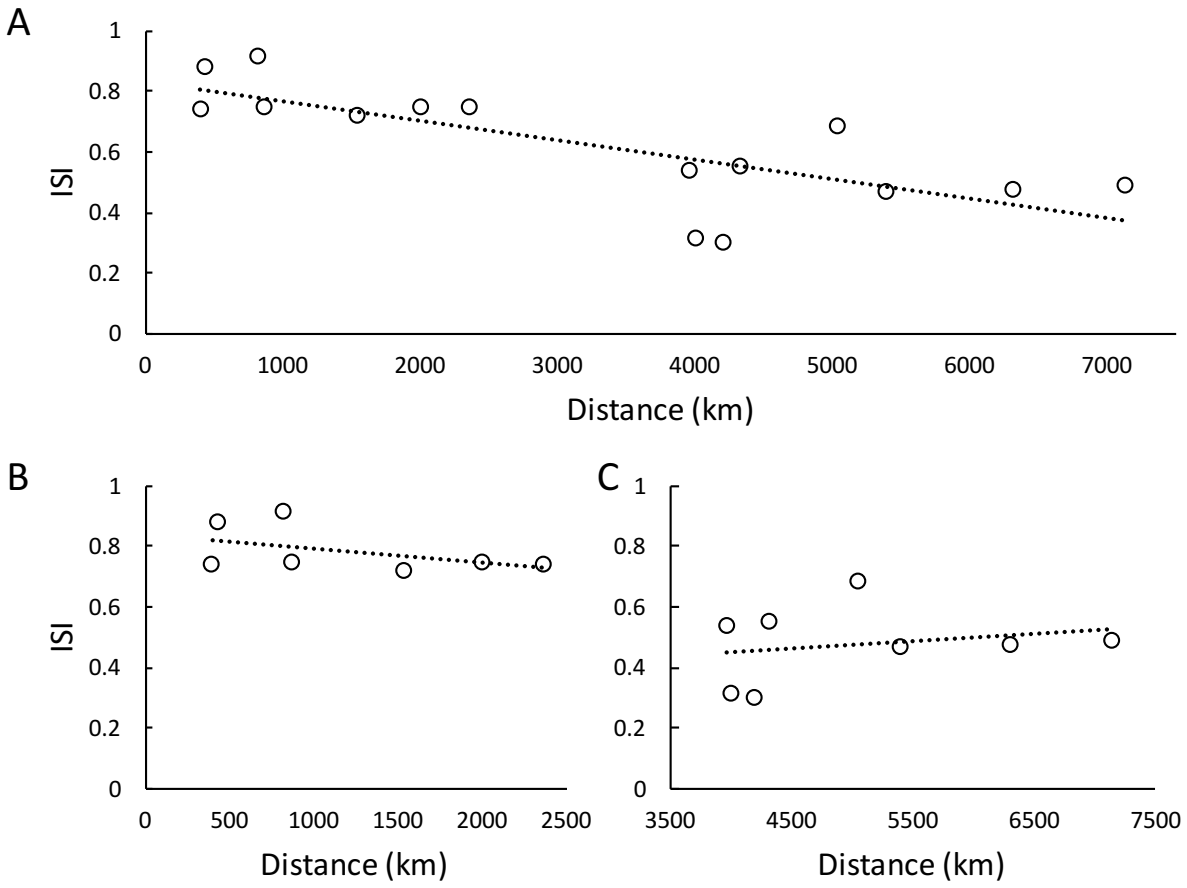


Table S1) Accessions of *Campanula rotundifolia* and allied species used for this study. Locations: CE=Central Europe, WE=Western Europe, NA=North America. Ploidy estimates obtained via flow cytometry on all plants.

Taxon	Location	Ploidy	Accession - Locality	Latitude	Longitude
C. gentilis	CE	2	25 - Czech Republic	50.4293	14.0756
C. rotundifolia	CE	2	23 - Germany	50.9734	13.8211
C. pollinensis	CE	2	11 - Italy	42.4691	13.5636
C. rotundifolia	CE	4	6 - Slovakia	48.6620	17.1107
C. gentilis	CE	4	24 - Czech Republic	50.1848	14.0810
C. rotundifolia	CE	4	33 - Germany	52.0350	8.4922
C. rotundifolia	WE	4	21 - Sweden	56.6357	12.9120
C. rotundifolia	WE	4	2 - Germany	50.7069	7.1292
C. rotundifolia	WE	4	16 - Belgium	50.7206	5.8816
C. rotundifolia	WE	4	4 - England	53.2485	-1.4597
C. rotundifolia	WE	6	58 - Ireland	54.1876	-8.6168
C. rotundifolia	NA	4	30 - Massachusetts, USA	42.5951	-72.5812
C. rotundifolia	NA	4	26 - Virginia, USA	37.2644	-80.9844
C. rotundifolia	NA	4	114 - Ontario, Canada	46.2247	-82.0751
C. rotundifolia	NA	4	115 - Ontario, Canada	45.7928	-82.7565
C. rotundifolia	NA	4	109 - Ontario, Canada	46.5360	-84.5898
C. rotundifolia	NA	4	113 - Ontario, Canada	47.4796	-84.8031
C. rotundifolia	NA	4	36 - Alberta, Canada	50.9819	-114.1774
C. rotundifolia	NA	4	40 - California, USA	40.6692	-124.1559
C. rotundifolia	NA	6	13 - Ontario, Canada	44.8121	-81.2986
C. rotundifolia	NA	6	10 - Ontario, Canada	44.4163	-81.4584
C. rotundifolia	NA	6	116 - Ontario, Canada	45.5583	-82.0185
C. rotundifolia	NA	6	105 - Michigan, USA	45.7506	-84.8650
C. rotundifolia	NA	6	31 - Alaska, USA	61.3805	-149.5380

CHAPTER THREE:
Postzygotic isolation varies by ploidy level within a polyploid complex

As published: Sutherland BL and LF Galloway. 2017. Postzygotic isolation varies by ploidy level within a polyploid complex. *New Phytologist* 213: 404-412.

- Whole genome duplication is considered to be a significant contributor to angiosperm speciation due to accumulation of rapid, strong interploid reproductive isolation. However, recent work suggests that interploid reproductive isolation may not be complete, especially among higher order cytotypes. This study evaluates postzygotic reproductive isolation among three cytotypes in a polyploid complex.
- We conducted reciprocal crosses using two diploid and two hexaploid populations each crossed to tetraploid populations spanning the geographic and phylogenetic range of the *Campanula rotundifolia* polyploid complex. Interploid and intrapopulation crosses were scored for fruit set, seed number, germination proportion and pollen viability. Postzygotic isolation was calculated for each cross as the product of these fitness components. A subset of offspring was cytotyped via flow cytometry.
- Postzygotic isolation was significantly lower in tetraploid–hexaploid crosses than diploid–tetraploid crosses, mostly due to substantially higher germination among tetraploid–hexaploid crosses. Tetraploid–hexaploid crosses produced pentaploids exclusively, whereas diploid–tetraploid crosses produced both triploids and tetraploids in high frequencies.
- Postzygotic isolation was weaker among higher order polyploids than between diploids and tetraploids, and unreduced gametes may facilitate diploid–tetraploid reproduction. This incomplete postzygotic isolation could allow ongoing interploid gene flow, especially among higher order polyploids, which may slow divergence and speciation in polyploid complexes.

Introduction

Polyploidization is considered one of the primary mechanisms of sympatric plant speciation (Bolnick & Fitzpatrick, 2007; Wood et al., 2009) due to its induction of rapid, strong reproductive isolation between diploids and related polyploids (Husband & Sabara, 2004; Rieseberg & Willis, 2007; Ramsey, 2011). However, diploids and polyploids frequently exist in complexes in which interploid gene flow has been observed (Stebbins, 1942; Petit et al., 1999), suggesting that the magnitude of interploid reproductive isolation may be overestimated.

In particular, recent work suggests that reproductive barriers between higher order polyploids may be weaker than those between diploids and polyploids (Hersch-Green, 2012; Sonnleitner et al., 2013; Hülber et al., 2015). If interploid reproductive isolation is incomplete, gene flow between cytotypes could slow diversification and speciation among polyploid lineages (Costa et al., 2014). Polyploids are an important component of angiosperm diversity; c. 35% of genera contain polyploids and polyploid complexes (Otto & Whitton, 2000; Wood et al., 2009). Yet despite this prevalence, broad patterns of polyploid evolution are not well-understood and relative diversification rates of polyploid lineages are still subject to debate (Mayrose et al., 2011; Soltis et al., 2014; Kellogg, 2016). Quantification of postzygotic interploid isolation and comparison of diploid–tetraploid reproductive barriers to those between higher order cytotypes will improve our understanding of interploid gene flow and its influence on polyploid diversification.

The mechanisms by which interploid mating can occur may differ depending on the cytotype of each parent. In diploid–tetraploid matings, fusion of a 1n haploid gamete from the diploid parent with a 2n diploid gamete from the tetraploid parent results in a

triploid embryo. These embryos face triploid block, substantial developmental defects caused by parental genomic imbalance and meiotic irregularities (Marks, 1966; Köhler et al., 2010). However, unreduced gametes offer an alternative mechanism of diploid–tetraploid reproduction (Bretagnolle & Thompson, 1995). If an unreduced 2n gamete from the diploid parent fuses with a reduced 2n gamete from the tetraploid parent, the resultant tetraploid embryo would not be subject to the decreased fitness inherent to triploid block. This mechanism of interploid compatibility has been studied as a route to neopolyploid formation (Ramsey & Schemske, 1998; Schatlowski & Köhler, 2012) but rarely as a source of ongoing interploid gene flow.

Postzygotic reproductive barriers between higher order polyploids may not be as strong as in diploid–tetraploid systems. When diploid angiosperms reproduce, both a diploid embryo and a triploid endosperm are formed. The endosperm contains a 2:1 ratio of maternal to paternal genomes (Birchler, 2014), and this ratio is the same in most homoploid systems regardless of ploidy. A 2:1 genomic ratio is critical for normal endosperm development; substantial deviation leads to aberrant endosperm development and little to no germination (Haig & Westoby, 1991; Scott et al., 1998; von Wangenheim & Peterson, 2004). The magnitude of deviation from a 2:1 parental genomic ratio in the endosperm, known as parental genomic imbalance, is associated with the severity of aberrant development. Because the magnitude of genomic imbalance in tetraploid–hexaploid hybrids is approximately one third less than in diploid–tetraploid hybrids (Sonnleitner et al., 2013) these higher-ploidy hybrids may experience fewer endosperm defects and thus may have higher fitness than diploid–tetraploid hybrids. In addition, the magnitude of imbalance is always less when maternal ploidy is greater, so we expect

crosses with maternal ploidy excess to germinate more frequently. Finally, high rates of inviable or aneuploid pollen in odd-ploidy hybrids appear to be less frequent in pentaploids than triploids (Costa et al., 2014), again supporting greater success of interploid reproduction in higher order polyploids.

Interploid reproduction is likely to be subject to the same genetic incompatibilities that affect homoploid reproduction. As populations become geographically isolated and accrue genetic divergence, Dobzhansky–Muller incompatibilities can arise and prevent hybridization following secondary contact. This effect has been well-studied in homoploid systems (Moyle et al., 2004; Nosrati et al., 2011). Although empirical data are sparse, genetic incompatibilities may arise more quickly in polyploids, either through reciprocal resolution of redundant gene copies (Lynch & Force, 2000; though see Muir & Hahn, 2015) or faster fixation of adaptive mutations in diverging polyploid lineages (Otto & Whitton, 2000). If polyploids accumulate genetic incompatibilities more rapidly than diploids, Dobzhansky–Muller incompatibilities may occur faster in polyploid lineages. Therefore, we may expect higher order polyploids to demonstrate more postzygotic reproductive isolation as genetic divergence increases than diploids and tetraploids. However, this effect may be difficult to detect given the strong postzygotic isolation expected due to genomic imbalance in interploid mating.

In order to elucidate patterns of interploid postzygotic isolation, we quantify postzygotic barriers among three cytotypes–diploid, tetraploid and hexaploid–of an outcrossing autopolyploid complex. Using an autopolyploid system allows us to reduce the conflating effects of parental genomic incompatibilities or heterosis that can occur in allopolyploids. We use a series of controlled crosses to investigate the following

questions: in a complex comprising diploids, tetraploids and hexaploids, do tetraploids exhibit different levels of postzygotic isolation when crossed with diploids than with hexaploids? Do interploid crosses show a difference in postzygotic isolation depending on parental cross-direction? Does increasing genetic divergence between populations increase interploid postzygotic isolation and are any patterns similar across cytotypes? Is there evidence of interploid reproduction via unreduced gametes?

Materials and Methods

Campanula rotundifolia exists as an autopolyploid complex comprising three dominant cytotypes: diploid (2n = 34 chromosomes), tetraploid (2n = 68 chromosomes) and hexaploid (2n = 102 chromosomes) (Kovanda, 1966; Stevens et al., 2012). Interploid reproduction between these three cytotypes, particularly diploid–tetraploid and tetraploid–hexaploid reproduction, could experience different postzygotic isolation with respect to both parental ploidy and cross-direction. Endosperm genomic imbalance is expected to be greater for diploid–tetraploid crosses than tetraploid–hexaploid crosses. When a tetraploid dam is crossed to a diploid sire, the genomic ratio is 4:1 due to fertilization of a 2n = 4x maternal polar nucleus by a 1n = 1x pollen grain. The reciprocal cross ratio is 1:1 (2n = 2x polar nucleus and 1n = 2x pollen grain). These cross ratios, 4:1 and 1:1, represent 19% and 25% deviation, respectively, from expected parental genomic dosages of 2:1. By contrast, the endosperm genomic ratios in 6x–4x and 4x–6x crosses are 3:1 and 4:3, representing 12% and 15% deviation, respectively, from the homoploid parental genomic dosages. Parental genomic imbalance and the severity of developmental defects are therefore expected to be reduced in tetraploid–hexaploid crosses relative to diploid–tetraploid crosses, and when the maternal ploidy is greater.

Campanula rotundifolia is a short-lived, self-incompatible, perennial wildflower that is thought to have originated in Central or Eastern Europe. It has a circumboreal distribution, and is common throughout much of Europe as far north as Svalbard and as far south as northern Spain, westward to Ireland and eastward to western Russia (Stevens et al., 2012). In North America, it is common in calcareous rocky outcrops and sandy lakeshores across the northern latitudes, and can extend in isolated alpine populations along the Appalachian and Rocky Mountains as far south as North Carolina and Mexico, respectively (Giblin, 2005). Due to a hypothesized recent range expansion as well as cytotypic and morphological complexity, *C. rotundifolia* is taxonomically complex, consisting of numerous named species, subspecies and varieties. For the purposes of this study, a broad definition including *C. rotundifolia* and its allies is used, roughly corresponding to the *C. rotundifolia* polytomy present in Mansion et al. (2012).

Cytotypes are nonuniformly distributed throughout the range. Diploids occur in Central and Eastern Europe as well as extreme Northern Europe. Limited diploid populations have been reported in North America (Löve & Löve, 1966), but repeated efforts to resample these populations have proven to be unsuccessful (B. Sutherland, pers. obs.). Hexaploids are limited to the British Isles in Europe (Stevens et al., 2012) and in central and western North America (Chapter 1). Tetraploids are the dominant cytotype and are common throughout the distribution, with multiple known contact zones with both diploids and hexaploids. Given the skewed geographic and phylogenetic distribution of cytotypes, it is not possible to compare all three cytotypes within the same geographic region. Therefore, to assess the effects of both ploidy and genetic divergence on postzygotic isolation, we chose 11 tetraploid populations spanning the geographic and

genetic range to cross against two diploid and two hexaploid 'test' populations (Table S1). Therefore, crosses should be viewed as demonstrating potential for reproduction, bearing in mind that prezygotic barriers untested in this study would exist in natural populations. Fruits were solicited from local collectors familiar with *C. rotundifolia* and were harvested in 2006 and 2012 (Table S1). Seeds from 30 source populations were obtained from throughout the distribution – 16 from Europe and 14 from North America. Crossing populations chosen from these samples included two North American hexaploids, two Central European diploids, seven European tetraploids and four North American tetraploids.

Before performing controlled crosses, we established the parental generation for all 15 study populations. Two hundred seeds were planted from each population in 40 cells of five seeds. Each cell contained seeds from the same family, and families were sampled evenly across all available for a given population, typically 10–15 families, but as high as 23. Seeds were surface sown on a moist 3:1 mix of Sunshine growth medium and Turface soil conditioner, then covered with a very thin layer of dry soil. Seeds were cold-stratified at 4° C for 14 d to improve germination (Drake & Ewing, 1997), then moved to a growth chamber with a 12-h light : dark cycle at 24° C : 15° C. Germination, defined here as full emergence of both cotyledons and shedding of the seed coat, was scored every 2 d for 6 wk. Once all five potential germinants had emerged in a cell, or after 6 wk had elapsed, germinants were thinned randomly to one per cell. After 6 wk of germination, plants were returned to 4° C for 6 wk for vernalization. Following vernalization, 20 germinants of each tetraploid population and 40 from each test

population were transplanted into conetainers and grown in the glasshouse where additional light was used to extend day length to 16 h.

In order to investigate interploid postzygotic isolation, the 11 tetraploid populations were each reciprocally crossed to two diploid (23 and 25) and two hexaploid (10 and 13, Table S1) 'test' populations. A total of 88 cross-types were created (11 4x populations x 2 test ploidies x 2 test populations/ploidy x 2 cross-directions). Within each tetraploid population and test diploid or hexaploid population, maternal and paternal plants were chosen randomly. Six to eight pollinations were performed per cross-type for a total of 560 experimental pollinations. In addition, 150 intrapopulation pollinations were performed, 10 per population. For each pollination, a bud was chosen before anthesis and emasculated by physically removing young anthers. Buds were then monitored daily for opening of the stigmatic lobes. Once stigmatic surfaces were exposed, a surplus of donor pollen was brushed onto the maternal flower.

Four fitness traits were measured to assess postzygotic isolation: fruit set, seed number, germination proportion and pollen viability. Fruit set was defined as the presence of a visibly inflated fruit at maturation, c. 3 wk after pollination, and was scored as a binary character. Seed number per fruit was scored by opening ripe fruits before dehiscence and counting all mature seeds, defined here as medium to dark brown and commensurate in size with those of intraploid crosses. Once counted, 25 replicates of five seeds each were sown for all cross-types, evenly distributed across all families. For cross-types with fewer seeds, all available seeds were sown. In many cases, fruits yielded only a couple of seeds, too few for planting. A total of 1374 replicates were planted across all cross-types, averaging 16 replicates per cross-type with a range of 3–25. As with parental

populations, F1 germination was scored every 2 d for 6 wk. Up to 20 seedlings were selected from all germinants within a given cross-type for growth to maturity. Due to poor germination of some cross-types, only 896 seedlings were transplanted, averaging 10 F1s per cross-type. When sufficient germinants were available, one seedling was selected randomly for transplant per replicate. Survival rates were high throughout the experiment; all cross-types experienced at least 95% survival, and no discernible patterns were found, so survivorship was not included as a fitness trait.

In order to get a measure of F1 fertility, we quantified the number of viable pollen grains on a subset of F1 plants. Pollen was sampled more thoroughly in diploid–tetraploid F1s because they showed considerable variation in cytotype (see the Results section), and we wanted to collect robust samples of each cytotype. Pollen was scored in c. 60% of diploid–tetraploid F1s, but only 30% of tetraploid–hexaploid F1s. Just before anthesis, anthers were removed from one flower per chosen plant. Anthers were dried for 72 h at room temperature, then stained with 60 µl 1% lactophenol blue solution. Stained samples were stored at least 1 wk before counting, but could be stored in the staining solution indefinitely. To count pollen grains, samples were vortexed thoroughly, then 10 µl of pollen suspension was pipetted onto a microscope slide and covered. Four randomly chosen fields of view were selected from the slide, and all stained and unstained pollen grains were counted from each slide. To calculate the number of viable pollen produced by each individual, the average proportion of stained (viable) pollen was calculated and multiplied by the total number of pollen counted from the four views.

Cytotypes of a subset of F1 plants were estimated via flow cytometry to confirm expected ploidy (Otto 2-step protocol; Otto, 1990). Because F1s from diploid–tetraploid

crosses showed considerable variation in cytotype, sampling was skewed toward diploid–tetraploid F1s in order to capture accurate estimates of cytotypic diversity. A total of 432 plants were cytotyped, 36% of the diploid–tetraploid F1s and 12% of the tetraploid–hexaploid F1s. Approximately 20 mg each of radish (*Raphanus sativus* 'Saxa': DNA content 1.11 pg/2C) and soybean (*Glycine max* 'Polanka': DNA content 2.50 pg/2C) were used as either internal standards (co-chopped with sample tissue) or external standards (prepared separately and analyzed to calibrate machine parameters for use throughout a single session). Approximately 30 mg of fresh *C. rotundifolia* leaf tissue was collected from basal rosettes and chopped finely into 1 ml of Otto I buffer. The sample then passed through a 30-l m filter and was incubated at room temperature for c. 1 h. Otto II buffer, containing 50 µg/ml of RNAse A and 50 µg/ml of propidium iodide, was then added and incubated for 10–15 min before visualization. Samples were analyzed using a FACSCalibur flow cytometer. We compared the relative fluorescence of unknown samples to our internal and external standards to estimate relative DNA quantity. Ploidy levels were assigned to be the nearest whole-number multiple of the ratio of fluorescence between a diploid *C. rotundifolia* and each standard. Because the haploid genome size for *C. rotundifolia* is c. 1.1 Gb, there was no overlap in fluorescence peaks between each cytotype (2x– 6x).

Because initial cytotype results for diploid–tetraploid F1s did not match our predictions of uniformly 3x offspring, a second round of pollinations was performed for approximately one-third of populations (underlined taxa in Table S1) to demonstrate repeatability of our results. Eight pollinations were performed for each cross-type. Up to 50 seeds per cross-type were planted in replicates of five seeds each and germinated as

above. Once sufficient leaf tissue was available, plants were cytotyped as above for a total of 240 additional individuals.

In order to calculate genetic distances, a maximum-likelihood phylogeny was generated using RADseq data for 28 populations of *C. rotundifolia* that included the 15 populations used in this study. A presence threshold of 12 populations out of 28 was set, resulting in a DNA dataset consisting of 25,762 SNPs and surrounding invariant positions which were concatenated into one continuous sequence per population (Chapter 1). Using a GTR+ gamma model of nucleotide substitution, a maximum-likelihood tree was generated using RAxML (Stamatakis, 2014). Pairwise differences per nucleotide (π) were then calculated using MEGA v.5.2.2 (Tamura et al., 2011).

Statistical analysis

Postzygotic isolation was defined as the difference in performance between within-population intraploid crosses and interploid crosses for four fitness components: fruit set (scored as a binary trait), seed number, germination proportion and viable pollen number. To standardize measures, mean values for each of the four components were calculated for all intraploid crosses, then interploid values for individual offspring were divided by intraploid means. Seed number was assumed to be influenced mainly by ovule production in the maternal plant, so seed number from each maternal population was used as the intraploid value. For fruit set, germination and pollen viability, an average of both parental populations was calculated and used as the intraploid value. The product of all four components was calculated to produce a relative cumulative fitness for each cross. Postzygotic isolation was calculated as one minus cumulative fitness – lower fitness values therefore lead to higher isolation rates, and vice versa.

We used a generalized mixed model (PROC GLIMMIX, SAS 9.3; SAS Institute Inc., Cary, NC, USA) to evaluate postzygotic isolation in interploid crosses. Analyses were performed on each fitness component as well as on cumulative fitness. Cross-direction (whether the maternal plant was high or low ploidy) was included as a main effect. We used test population as a fixed effect with four levels, for example, two hexaploid and two tetraploid populations, and tested the a priori hypothesis of ploidy with an independent contrast that compared the diploid–tetraploid results to the tetraploid–hexaploid results. Because our goal was to sample the variation present in tetraploids, tetraploid population was included as a random effect. For pollen viability, an additional analysis compared F1s by cytotype combining all cross-types; that is, pollen viability of triploid vs tetraploid offspring for all diploid–tetraploid crosses. Finally, we used ANCOVA to assess whether the genetic distance between crossed populations influenced F1 performance for each trait. We first calculated an average performance of each cross-type for each trait and regressed it on genetic distance (π) including the fixed effects of test population, cross-direction and their interaction. We also included the interactions between the fixed effects and genetic distance. These interactions indicate whether any effects of genetic distance on fitness depend on test population or cross-direction.

Results

F1 performance varied among fitness components (Table 1; Figs 1, S1). Fruit set and seed number were both greater when the paternal parent had the higher ploidy (Figure 1a,b), whereas germination and pollen viability did not depend on the direction of the cross. There was a striking difference in germination between offspring of diploid–tetraploid crosses and tetraploid–hexaploid crosses (Figure 1c). On average, seeds from

crosses with hexaploids had over six times greater germination than those from crosses with diploids. Fruit set and pollen viability was slightly higher among F1s from tetraploid–hexaploid crosses than diploid–tetraploid crosses (Figure 1a,d).

Postzygotic isolation was lower when tetraploid plants were crossed with hexaploids than with diploids, regardless of cross direction (Table 1; Figure 2). Diploid–tetraploid crosses experienced well over 90% reduction in fertile F1s relative to intraploid crosses and over 95% reduction in three of four diploid cross-types. By contrast, although tetraploid–hexaploid crosses experienced substantial postzygotic isolation, they were capable of producing more fertile F1s than diploid crosses, with reductions ranging from only 71.6% to 84.3%. Cross-direction did not affect cumulative postzygotic RI.

F1s from tetraploid–hexaploid crosses were almost uniformly pentaploid (Figure 3). However, cytotype frequencies for diploid–tetraploid crosses differed strongly from the expectation of uniform triploids. Only c. 47% of all diploid–tetraploid F1s were triploid; another 43% of these offspring were tetraploid, 8% were diploid and 1.5% pentaploid. Follow-up diploid–tetraploid crosses confirmed these results. Among follow-ups, 52% were triploid, 45% were tetraploid and 2.3% were diploid. No directional effect was observed in tetraploid crosses to diploids or hexaploids. Pollen fertility was assessed for triploid and tetraploid F1s from diploid–tetraploid crosses, and pentaploid F1s from tetraploid–hexaploid crosses. Tukey post-hoc comparisons found that pollen fertility differed with respect to cytotype, with tetraploid F1s significantly more pollen fertile than either triploids or pentaploids, and with pentaploids significantly more fertile than triploids (Figure 4; $F_{2,231} = 5.03$, $P = 0.0073$).

Genetic distance between populations only influenced fruit set of interploid crosses (Table S2). Greater genetic distance in diploid–tetraploid crosses increased fruit set, explaining 19.2% of the variance, but had little effect on tetraploid–hexaploid fruit set (Table S2; Figure S2).

Discussion

We hypothesized that within a polyploid complex, higher order cytotypes would experience weaker interploid postzygotic isolation than diploids and tetraploids. We found that tetraploids experienced 22% weaker postzygotic isolation when crossed with hexaploids than when crossed with diploids, and that this difference was due primarily to large differences in germination. A surprisingly high percentage of offspring from diploid–tetraploid crosses were fertile tetraploids; however, low germination of diploid–tetraploid F1s limited the effect of these tetraploids on postzygotic isolation. The weaker postzygotic isolation between higher order cytotypes is consistent with ongoing field data (Chapters 5,6) which show that interploid offspring are relatively common between tetraploids and hexaploids, but virtually absent between diploids and tetraploids in both experimental and natural mixed-ploidy populations. The finding that interploid postzygotic isolation is weaker among higher order cytotypes supports the possibility of greater gene flow and reduced divergence within polyploid complexes.

Diploid–tetraploid crosses experienced over 96% postzygotic isolation in both cross-directions, whereas postzygotic isolation between tetraploids and hexaploids ranged from 76% to 83%. Germination had the strongest effect of all fitness components on postzygotic isolation; significantly more seeds germinated from tetraploid–hexaploid crosses than from diploid–tetraploid crosses (61.9% and 9.2%, respectively). Differences

in interploid germination were likely caused by parental genomic imbalances in developing seeds and embryos. In interploid *Arabidopsis* crosses, reduced germination was found in diploid–tetraploid F1s, and attributed to aberrant development of endosperm caused by overexpression of genes from the parent with genomic excess (Slotte et al., 2008; Stoute et al., 2012). Crosses between tetraploids and hexaploids are expected to have reduced parental genomic imbalance (Sonnleitner et al., 2013). This reduced imbalance between higher ploidy levels suggests that endosperm development in F1s derived from tetraploid–hexaploid crosses have fewer developmental defects than those from diploid–tetraploid crosses, allowing for greater seed viability, which is congruent with our markedly higher germination proportions among tetraploid–hexaploid F1s.

Although no parental cross-direction effect on cumulative fitness was observed, three fitness components differed with cross direction. Germination was marginally higher in interploid crosses in which the maternal ploidy was greater. By contrast, fruit set and seed number were greater in crosses when the higher-ploidy parent was paternal. These opposing effects of parental direction among traits likely explain why we observed no directional effect on postzygotic isolation. Parental genomic imbalance may explain the observed directional effects of individual traits. In Brassicaceae, interploid crosses with maternal genome excess often create more viable seed than crosses with paternal genome excess (Dilkes & Comai, 2004; Stoute et al., 2012), which is consistent with our germination results. However, the higher fruit set and seed number observed in this study for crosses with a paternal genomic excess is congruent with observations from other systems. Paternal genomic excess has been linked with higher fruit set and seed number in other systems (Haig & Westoby, 1991; but see Greiner & Oberprieler, 2012). This is

thought to result from later seed abortion when the paternal genome is in excess, leading to phenotypically normal but inviable seeds, whereas maternal excess aborts seed development early enough that phenotypically normal seeds are not formed.

Genetic distance played a limited role in postzygotic isolation within the *C. rotundifolia* complex. Although tetraploid populations chosen for this study sampled broadly from the phylogeny and spanned the geographic range of the species, there was a limited effect of genetic distance between populations on fitness of crosses. We expected that postzygotic isolation would increase with greater genetic distance, as has been observed in homoploid systems (Moyle et al., 2004; Nosrati et al., 2011). We also expected that tetraploid–hexaploid crosses would accumulate genetic incompatibility more quickly with increasing genetic distance than diploid–tetraploid crosses due to resolution of redundant gene copies and faster mutation rates among higher cytotypes (Lynch & Force, 2000; Otto & Whitton, 2000). However, we found no support for these expectations. There are several possible explanations for the lack of any genetic distance effect on postzygotic isolation. First, it is possible that the magnitude of postzygotic isolation imparted by ploidy change masks any underlying effect of genetic distance. Alternatively, decreases in fitness may only become apparent in F2 recombinant hybrid individuals (Ramsey et al., 2003). Lastly, populations chosen for this study, particularly those from North America, have relatively few fixed nucleotide differences in chloroplast DNA, suggesting that populations have not diverged sufficiently to accumulate Dobzhansky–Muller incompatibilities.

Cytotypic frequencies for interploid hybrids differed substantially from expected values. Interploid crosses are expected to yield intermediate cytotypes: triploids in

diploid–tetraploid crosses, and pentaploids in tetraploid–hexaploid crosses. Although tetraploid–hexaploid crosses consistently had pentaploid offspring, only 50% of diploid–tetraploid offspring were triploid and over 44% were tetraploid. Unreduced gametes have been associated with neopolyploid formation in other systems (Ramsey, 2007) and are likely to have mediated neotetraploid formation in this study. Previous work suggests that unreduced ovules are more likely than unreduced pollen to generate neopolyploids (Ramsey, 2007). However, the high frequencies of tetraploids observed in both diploid–tetraploid crossing directions suggest that both types of unreduced gametes can facilitate neotetraploid formation in *C. rotundifolia*. Although diploid–tetraploid crosses yielded high relative rates of tetraploid germinants, the absolute number of germinants for diploid–tetraploid crosses was much lower than for tetraploid–hexaploid crosses. Poor viability of triploid offspring is likely to explain both low germination of diploid–tetraploid crosses as well as the high relative rate of tetraploid formation; because tetraploid seeds are not subject to parental genomic imbalance, a greater proportion of those seeds germinate relative to their triploid siblings. Therefore, although unreduced gamete formation occurs at a low rate, the diploid–tetraploid germinants in this study were effectively enriched for neotetraploids.

Although unreduced gametes can occur in any ploidy level, fusion of an unreduced (2x) gamete from a diploid with a reduced gamete from a tetraploid is the most likely to result in viable offspring. Not only does this combination yield even-ploidy offspring, duplication of the diploid genome effectively ameliorates parental genomic imbalance. Fusion of an unreduced (4x) gamete from a tetraploid parent with a reduced gamete from a diploid parent would result in considerable parental genomic imbalance

likely to cause reduced germination and survival. Likewise, for tetraploid–hexaploid crosses, unreduced hexaploid gametes (6x) fused with reduced tetraploid gametes (2x), would likely be inviable due to highly unbalanced parental genomes. However, unreduced tetraploid gametes (4x) with reduced hexaploid gametes (3x) would yield less parental genomic imbalance. Although these embryos are predicted to survive, none were found in this study, perhaps due to the high success of the more frequent 5x offspring.

The differential patterns of interploid postzygotic isolation found in the present study, as well as preliminary data from natural populations, suggest that rates of gene flow vary among cytotypes. If these patterns are widespread, interploid gene flow could constrain divergence within polyploid complexes, particularly among higher order cytotypes. Increased interploid gene flow among higher cytotypes may also help explain why pentaploids are found more commonly in mixed tetraploid–hexaploid populations than triploids are found in mixed diploid–tetraploid populations (e.g. Stevens et al., 2012; Hülber et al., 2015; Chapter 6). Our results also caution against assumptions of little to no gene flow in diploid–tetraploid populations based on absence of triploids. Triploid block is a powerful barrier to interploid gene flow via triploid intermediates. The diploid–tetraploid crosses in our study would, in almost all cases, experience essentially complete postzygotic isolation without formation of fertile tetraploids. Although tetraploid offspring were few in absolute terms, the fact that they comprised a substantial proportion of diploid–tetraploid offspring suggests that they may represent an alternative path to unidirectional gene flow from diploids to tetraploids. Focusing solely on the presence of triploids as a marker of interploid reproduction overlooks the potential effect of neotetraploid formation.

Interploid reproductive isolation is often considered to be strong enough to act as a sympatric speciation mechanism between related cytotypes, to the point of being dubbed ‘instant speciation’ (Coyne & Orr, 2004). Although this may be true in some taxa, our findings indicate that postzygotic barriers between higher order cytotypes may be lower than previously assumed, and suggest that unreduced gametes may mediate reproduction between diploids and tetraploids. By focusing on individual postzygotic barriers, we establish that germination greatly affects the role that ploidy level plays on interploid postzygotic isolation. If naturally occurring mixed-ploidy populations show similar differences in postzygotic isolation, the potential for differential rates of gene flow is substantial. So far, our work in natural populations aligns with the results presented herein. Greater gene flow among higher cytotypes, if found, could help explain the persistence of polyploid complexes and provide an alternative hypothesis for the finding of lower than expected diversification in some polyploid lineages.

Acknowledgements

We thank numerous collectors worldwide for help obtaining seed samples for this study, and are grateful to B. Cottrell and B. Friestad for help with plant care and crossing. We also thank K. Barnard-Kubow, C. Debban, B. Sanderson, M. Augat and R. Watson for feedback on drafts of this manuscript; as well as B. Blackman, R. Cox, D. Carr and J. Antonovics for guidance throughout this project. We would like to thank the National Science Foundation (NSF DEB-1020717), the Society for the Study of Evolution, the Torrey Botanical Society and the Botanical Society of America for funding for this project.

References

Birchler JA. 2014. Interploidy hybridization barrier of endosperm as a dosage interaction. Frontiers in Plant Science 5: 281–291.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. Annual Review of Ecology, Evolution, and Systematics 381: 459–487.

Bretagnolle F, Thompson JD. 1995. Tansley review no. 78. Gametes with the stomatic chromosome number: mechanisms of their formation and role in the evolution of autopolypoid plants. New Phytologist 129: 1–22.

Costa J, Ferrero V, Louriero J, Castro M, Navarro L, Castro S. 2014. Sexual reproduction of the pentaploid, short-styled *Oxalis pes-caprae* allows the production of viable offspring. Plant Biology 16: 208–214.

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

Dilkes BP, Comai L. 2004. A differential dosage hypothesis for parental effects in seed development. Plant Cell 16: 3174–3180.

Drake D, Ewing K. 1997. Germination Requirements of 32 Native Washington Prairie Species. In: Dunn PV, Ewing K, eds. Ecology and conservation of the South Puget sound prairie landscape. Seattle, WA, USA: The Nature Conservancy of Washington, 181–187.

Giblin DE. 2005. Variation in floral longevity between populations of *Campanula rotundifolia* (*Campanula*ceae) in response to fitness accrual rate manipulation. American Journal of Botany 92: 1714–1722.

Greiner R, Oberprieler C. 2012. The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra- and hexaploid relatives. Flora 207: 629–635.

Haig D, Westoby M. 1991. Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. Philosophical Transactions of the Royal Society B: Biological Sciences 333: 1–13.

Hersch-Green E. 2012. Polyploidy in indian paintbrush (*Castilleja*; Orobanchaceae) species shapes but does not prevent gene flow across species boundaries. American Journal of Botany 99: 1680–1690.

Hülber K, Sonnleitner M, Suda J, Krejčíková J, Schönswetter P, Schneeweiss GM, Winkler M. 2015. Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio*

carniolicus syn. *Jacobaea carniolica*, Asteraceae. Ecology and Evolution 56: 1224–1234.

Husband BC, Sabara HA. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). New Phytologist 161: 703–713.

Kellogg EA. 2016. Has the connection between polyploidy and diversification actually been tested? Current Opinion in Plant Biology 30: 25–32.

Köhler C, Scheid OM, Erilova A. 2010. The impact of the triploid block on the origin and evolution of polyploid plants. Trends in Genetics 26: 142–148.

Kovanda M. 1966. Some chromosome counts in the *Campanula rotundifolia* complex II. Folia Geobotanica Phytotaxonomica Bohemoslovaca 3: 268–273.

Löve A, Löve D. 1966. Cytotaxonomy of the alpine vascular plants of Mount Washington. Series in Biology. Paper 38. Boulder, CO, USA: University of Colorado Press.

Lynch M, Force AG. 2000. The origin of interspecific genomic incompatibility via gene duplication. American Naturalist 156: 590–605.

Mansion G, Parolly G, Crowl AA, Mavrodiev E, Cellinese N, Oganesian M, Fraunhofer K, Kamari G, Phitos D, Haberle R et al. 2012. How to handle speciose clades? mass taxon-sampling as a strategy towards illuminating the natural history of *Campanula* (Campanuloideae). PLoS ONE 7: e50076.

Marks GE. 1966. The origin and significance of intraspecific polyploidy: experimental evidence from *Solanum chacoense*. Evolution 20: 552–557.

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K. 2011. Recently formed polyploid plants diversify at lower rates. Science 333: 1257–1260.

Moyle LC, Olson MS, Tiffin P. 2004. Patterns of reproductive isolation in three angiosperm genera. Evolution 58: 1195–1208.

Muir CD, Hahn MW. 2015. The limited contribution of reciprocal gene loss to increased speciation rates following whole-genome duplication. American Naturalist 185: 70–86.

Nosrati H, Price AH, Wilcock CC. 2011. Relationship between genetic distances and postzygotic reproductive isolation in diploid *Fragaria* (Rosaceae). Biological Journal of the Linnean Society 104: 510–526.

Otto FJ. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewickz Z, Crissman HA, eds. Methods in cell biology, vol. 33. San Diego, CA, USA: Academic Press, 105–110.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. Annual Review of Genetics 34: 401–437.

Petit C, Bretagnolle F, Felber F. 1999. Evolutionary consequences of diploid–polyploid hybrid zones in wild species. Trends in Ecology and Evolution 148: 306–311.

Ramsey J. 2007. Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). Heredity 98: 143–150.

Ramsey J. 2011. Polyploidy and ecological adaptation in wild yarrow. Proceedings of the National Academy of Sciences, USA 108: 7096–7101.

Ramsey J, Bradshaw HG Jr, Schemske DW. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (phrymaceae). Evolution 57: 1520–1534.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 291: 467–501.

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

Schatlowski N, Köhler C. 2012. Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. Journal of Experimental Botany 63: 6059–6067.

Scott RJ, Spielman M, Bailey J, Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development 125: 3329–3341.

Slotte T, Huang H, Lascoux M, Ceplitis A. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). Molecular Biology and Evolution 25: 1472–1481.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA et al. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). New Phytologist 202: 1105–1117.

Sonnleitner M, Weis B, Flatscher R, Escobar Garcia P, Suda J, Krejčíková J, Schneeweiss GM, Winkler M, Schönswetter P, Hülber K. 2013. Parental ploidy strongly affects offspring fitness in heteroploid crosses among three cytotypes of autopolyploid *Jacobaea carniolica* (Asteraceae). PLoS ONE 8: e78959.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies. Bioinformatics 30: 1312–1313.

Stebbins GL Jr. 1942. Polyploid complexes in relation to ecology and the history of floras. American Naturalist 76: 36–45.

Stevens CJ, Wilson J, McAllister HA. 2012. Biological flora of the British Isles: *Campanula rotundifolia*. Journal of Ecology 100: 821–839.

Stoute AI, Varenko V, King GJ, Scott RJ, Kurup S. 2012. Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. Plant Journal 71: 503–516.

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

von Wangenheim KH, Peterson HP. 2004. Aberrant endosperm development in interploidy crosses reveals a timer of differentiation. Developmental Biology 270: 277–289.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences, USA 106: 13875–13879.

Table 1) Analysis of variance for fitness components and cumulative fitness of *Campanula rotundifolia* interploid crosses. Test population refers to the diploid and hexaploid populations crossed to each tetraploid population, and cross-direction refers to the relative ploidy of each parent in a cross, that is, if the maternal plant was a greater ploidy (4x-2x and 6x-4x crosses) or a lesser ploidy (2x-4x and 4x-6x crosses). Ploidy contrast compares the two diploid and two hexaploid populations using an a priori test. Tetraploid population was included in the model as a random effect. Table lists F-values, P-values: +, < 0.1; *, < 0.05; **, < 0.01; ***, < 0.001.

Source	df	Fruit Set	Seed Set	Germination	Pollen Viability	Cumulative Fitness
Test Population	3	2.71*	2.09	56.87***	1.34	8.05***
Cross Direction	1	26.12***	30.85***	2.79+	0.82	0.32
Test pop x Direction	3	1.2	0.48	0.83	1.37	0.94
Ploidy Contrast	1	6.6*	0.02	162.29***	4.00*	21.10***
Error		549	549	473	183	401

Figure 1) Relative fitness of interploid crosses in *Campanula rotundifolia*. Two diploid (23, 25) and two hexaploid (10, 13) test populations were each crossed to 11 tetraploid (4x) populations; crosses list maternal plant first. (a) Fruit set, (b) seed number per fruit (crosses that did not set fruit excluded), (c) proportion germination, and (d) viable pollen all reported relative to mean parental values. Bars denote SE.

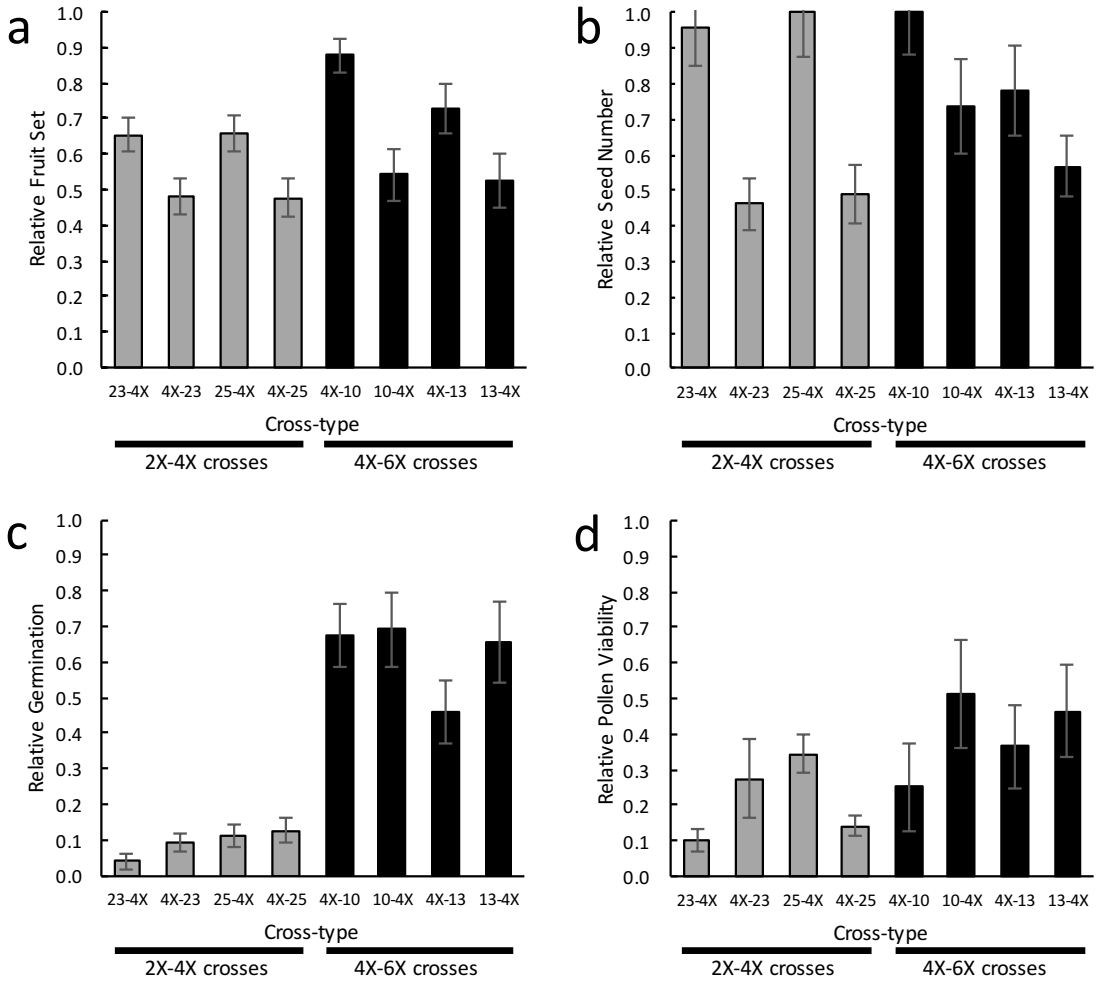


Figure 2) Index of postzygotic reproductive isolation for interploid crosses of *Campanula rotundifolia*. Two diploid (23, 25) and two hexaploid (10, 13) test populations were each crossed to 11 tetraploid (4x) populations. Reproductive isolation was calculated as the product of relative fruit set, relative seed number, relative germination proportion, and relative pollen fertility. Bars denote SE.

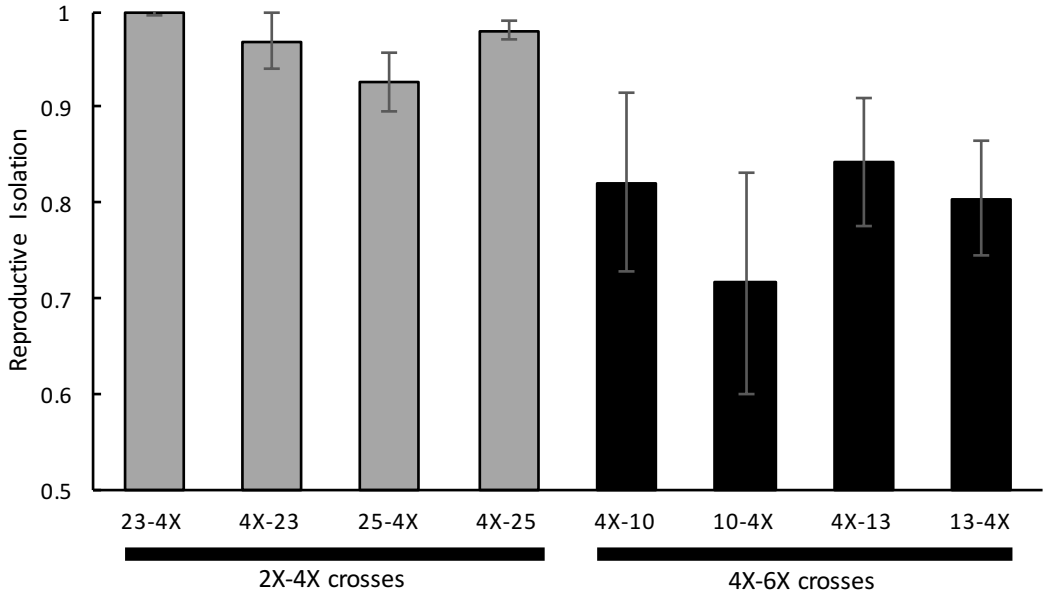


Figure 3) Combined cytotype composition of F1 offspring from two rounds of interploid crosses. Diploid and hexaploid *Campanula rotundifolia* test populations were reciprocally crossed to 11 tetraploid populations. Maternal ploidy is listed first.

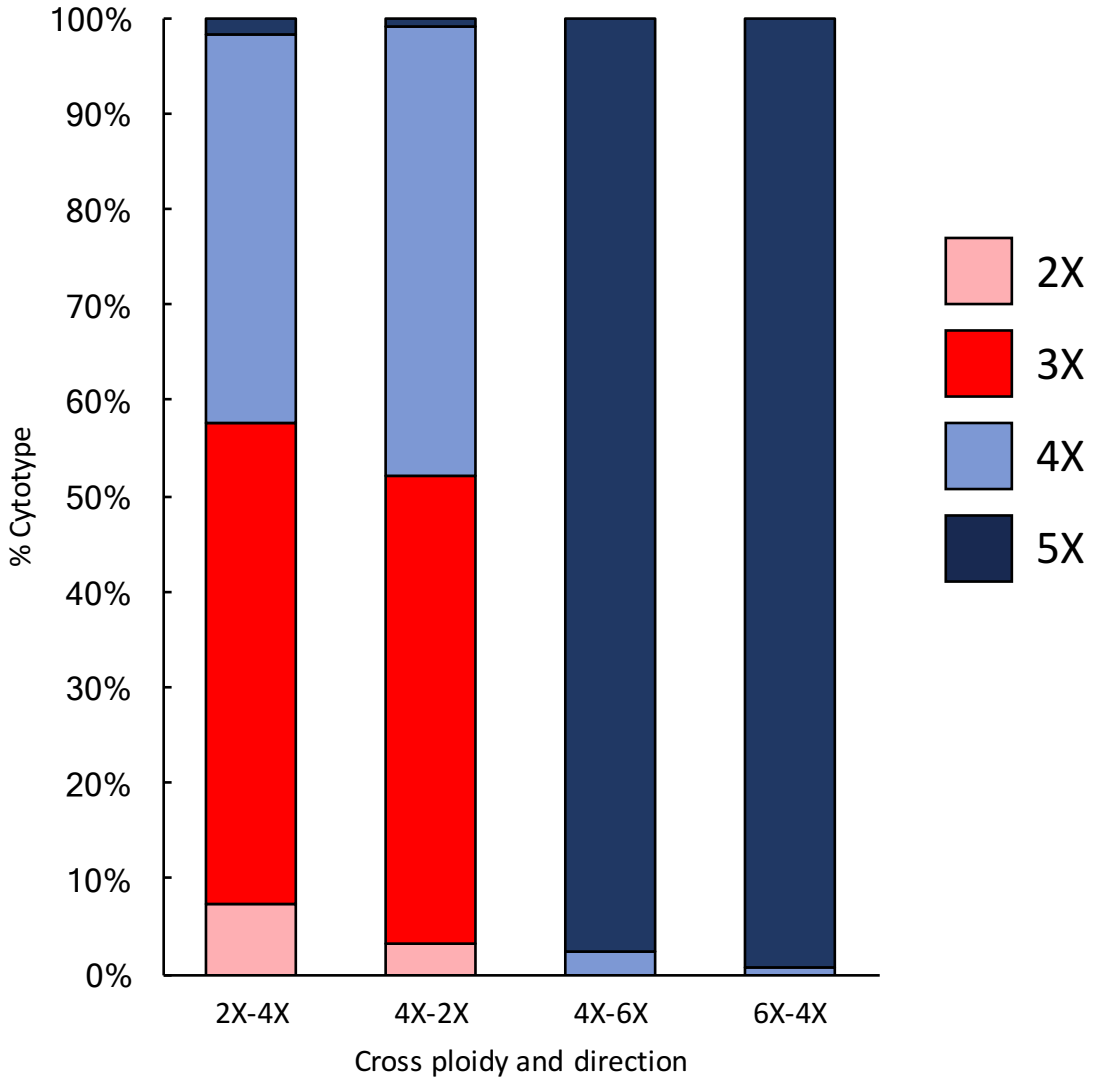


Figure 4) Pollen viability of offspring of diploid–tetraploid and tetraploid–hexaploid crosses in *Campanula rotundifolia*. Offspring of interploid crosses was separated by cytotype and pollen counts were pooled across crosses. Bars denote SE.

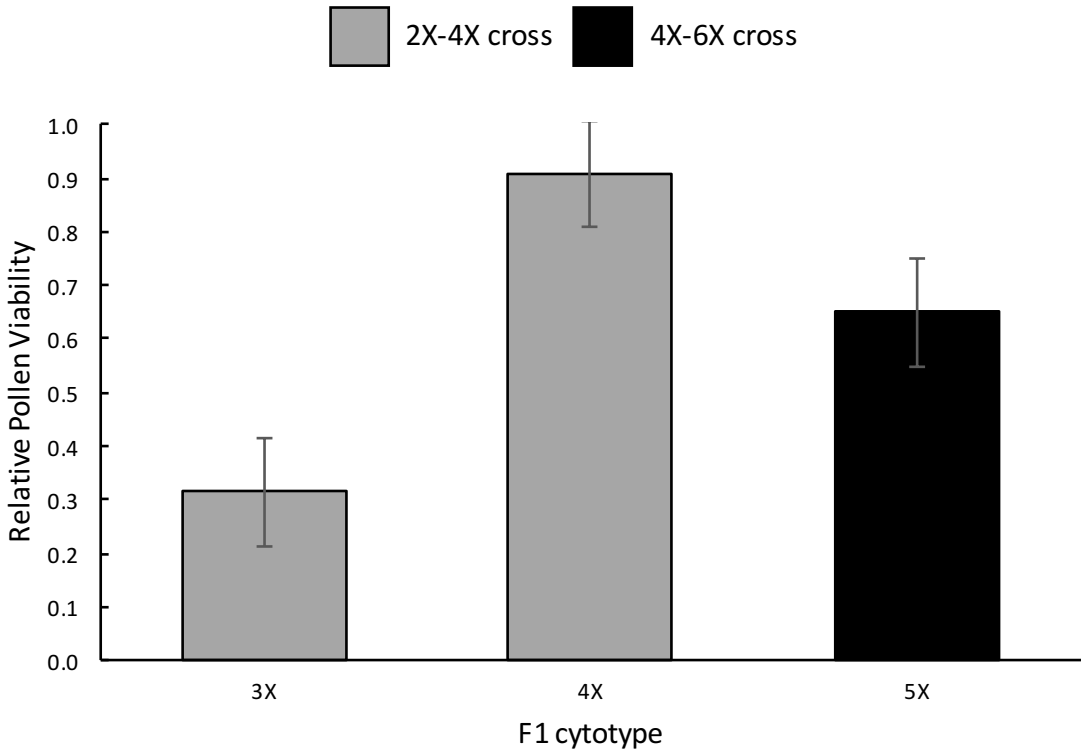


Table S1) Accessions of *Campanula rotundifolia* and allied species used for this study. Ploidy estimates obtained via flow cytometry on all plants used as parents in the crossing study.

Taxon	Accession Number	Ploidy	Locality	Latitude	Longitude
Populations under study					
C. rotundifolia	4	4	Derbyshire, England	53.24852	-1.45967
C. rotundifolia	6	4	Slovakia	48.66203	17.11068
C. rotundifolia	16	4	Plombieres, Belgium	50.72060	5.88164
C. rotundifolia	21	4	Halmstad, Sweden	56.63570	12.91201
C. gentilis	24	4	Vinaricka Hora, Czech Rep.	50.18479	14.08102
C. rotundifolia	26	4	Giles County, VA, USA	37.26439	-80.98441
C. rotundifolia	30	4	Turner's Falls, MA USA	42.59509	-72.58120
C. rotundifolia	32	4	Col du Lautaret, France	45.01839	6.38767
C. rotundifolia	33	4	Bielefeld, Germany	52.03501	8.49221
C. rotundifolia	36	4	Calgary, Alberta, Canada	50.98186	-114.17735
C. rotundifolia	38	4	San Juan Island, WA, USA	48.52504	-123.00173
Test Populations					
C. rotundifolia	23	2	Dresden, Germany	50.97344	14.07563
C. gentilis	25	2	Zamky, Czech Republic	50.42926	13.82112
C. rotundifolia	10	6	Port Elgin, Ontario, Canada	44.41625	-81.45836
C. rotundifolia	13	6	Red Bay, Ontario, Canada	44.81210	-81.29858

Table S2) Analysis of covariance on cross-type means for fitness subcomponents and cumulative fitness with respect to genetic distance. Table lists F-values, p-values: + < 0.1, * < 0.05, ** <0.01, ***<0.001.

	df	Fruit Set	Seed Number	Germination	Pollen Viability	Fitness
Test Population	3	3.61*	2.74+	5.74**	0.39	0.18
Cross Direction	1	0.08	0.36	0.29	0.09	1.28
Test Pop x Direction	3	1.30	0.95	0.14	2.39+	0.78
Genetic Distance	1	0.21	0.01	2.07	2.36	2.49
Distance x Test Pop	3	3.27*	2.62+	2.36+	0.73	0.16
Distance x Direction	1	1.13	0.38	0.16	0.08	1.89
Error df		72	72	72	72	72

Figure S1) Fitness of interploid crosses in *Campanula rotundifolia* compared to intrapopulation crosses. a) Fruit set, b) seed number per fruit (crosses that did not set fruit excluded), c) proportion germination, and d) viable pollen all reported relative to mean parental values. Bars denote standard error; standard errors for diploid and hexaploid populations (23 and 25, 10 and 13, respectively) generally larger than interploid crosses due to smaller sample sizes (10 crosses per population).

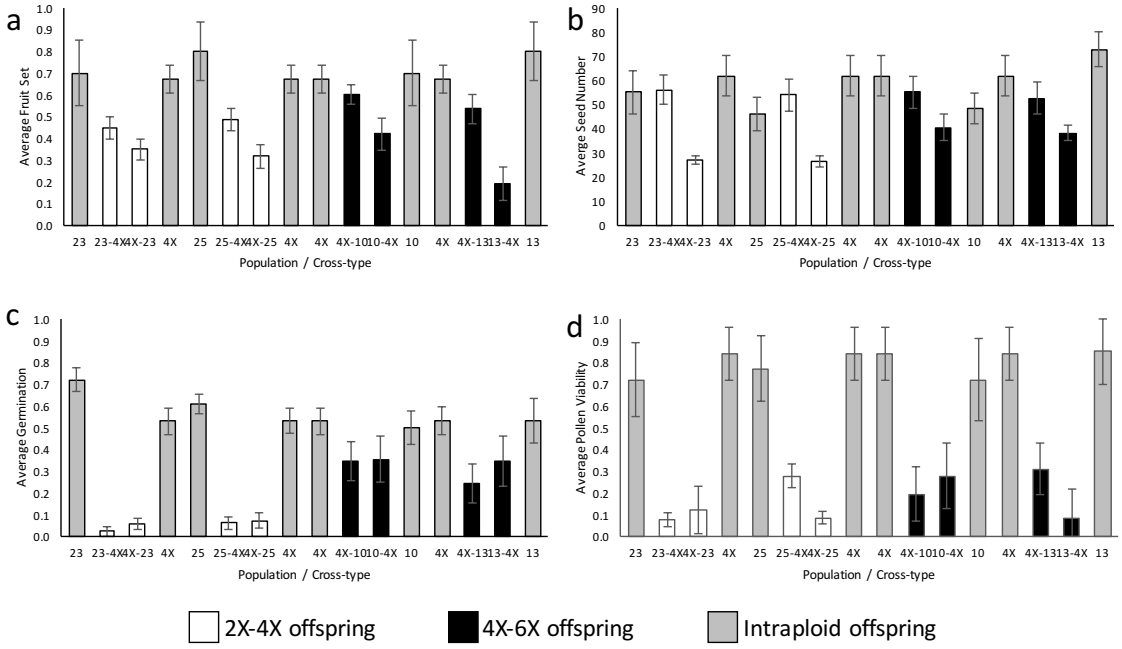
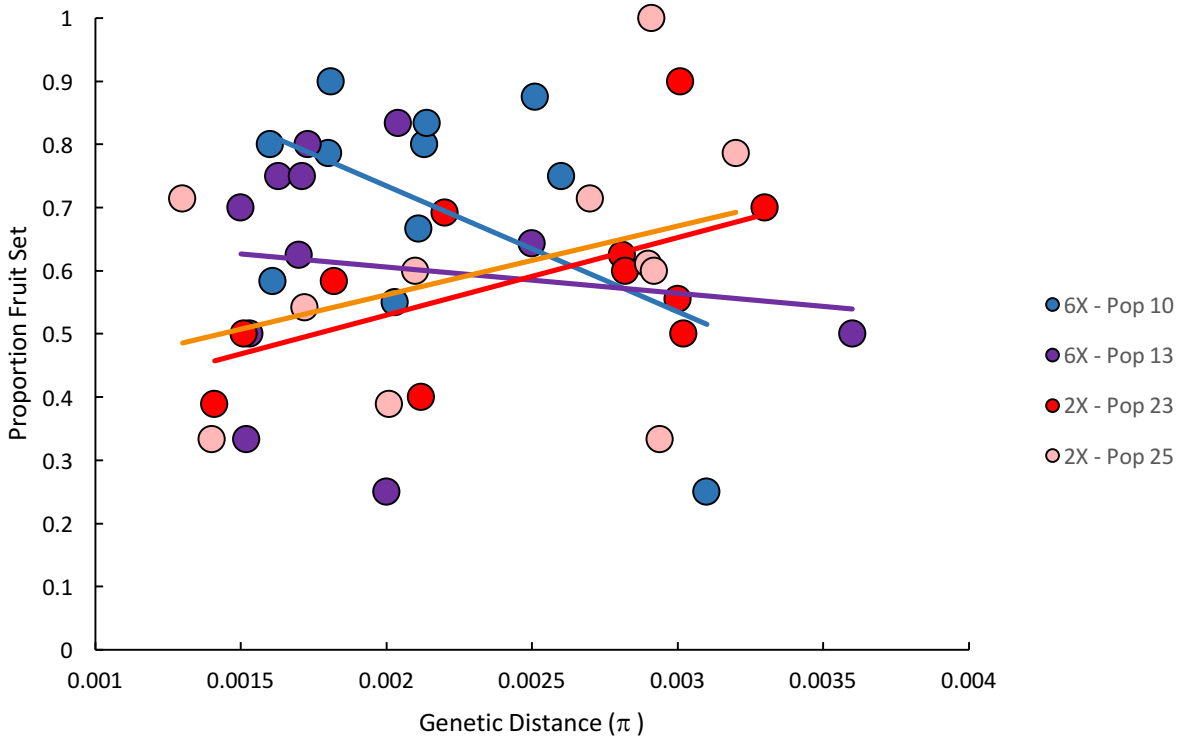


Figure S2) Analysis of covariance regressing fruit set against genetic distance with respect to test-population. Reciprocal crosses were averaged because no difference was found with respect to cross direction. Fruit set increased with genetic distance in diploid crosses (R^2=0.192) and decreased slightly in hexaploid crosses (R^2=0.047).



CHAPTER FOUR:

Hybridization and backcrossing across different ploidy levels may contribute to ongoing gene flow in a polyploid complex

Premise of the study – The evolutionary fate of lineages following secondary contact is well-studied in homoploid systems, but has been largely ignored in polyploid complexes because individuals differing in cytotype are assumed to be reproductively isolated. Yet interploid reproduction is increasingly found in polyploid systems, and may have substantial effects on gene flow and divergence in polyploid complexes.

Methods – To evaluate the possibility of gene flow between cytotypes, we used F1 interploid hybrids derived from 2X-4X and 4X-6X crosses within the *Campanula rotundifolia* autopolyploid complex. We conducted reciprocal backcrosses of the F1 hybrids to both parental cytotypes and measured seed set and germination, and estimated ploidy of backcross progeny.

Key results – 3X hybrids showed little ability to backcross with either the 2X or 4X parental cytotype. 5X hybrids, by contrast, had relatively high seed set and germination, particularly when crossed with 6X individuals. 2X and 4X interploid hybrids from 2X-4X crosses had high fitness in intraploid backcrosses (i.e. 2X with 2X, 4X with 4X), though 4X hybrids had higher fitness. F1 cytotype was consistent with parental cytotype for intraploid backcrosses of 2X and 4X F1s, while interploid backcrosses produced both 3X and 4X offspring. Among 5X F1s, backcross progeny were almost uniformly aneuploid.

Conclusions – These results demonstrate that outcomes of secondary contact between ploidy levels in an autopolyploid complex depend on parental cytotype. Low backcross ability of 3X hybrids, coupled with relative scarcity of fertile 2X and 4X hybrids, indicate that diploids and tetraploids are largely reproductively isolated and likely to develop secondary reinforcement upon contact. By contrast, high fertility in 5X hybrids and viability of backcrosses implies that interploid gene flow may be high in 4X-6X contact zones, and that high frequency of aneuploids may lead to formation of hybrid swarms. These results suggest that polyploids may diverge more quickly from diploids than from other polyploids, and that different polyploid cytotypes may not undergo the "instant speciation" characteristic of diploids and tetraploids.

Introduction

When related lineages come into secondary contact following a period of isolation, their long-term evolutionary fate is determined by the magnitude of gene flow between them (Schluter, 2001; Nosil, 2008). Gene flow between lineages in secondary contact can be limited by postzygotic barriers related to both the fitness of hybrid offspring (Schluter, 2001; Husband and Sabara, 2004) and the ability of these hybrids to backcross with parental types (Ellstrand et al., 1999; Moraes et al., 2013). When hybrid fitness is low, parental types often develop secondary reinforcement and strengthen prezygotic barriers, eventually undergoing speciation (Noor, 1999). When hybrid fitness is high, and hybrids can reproduce with parental types, the lineages may fuse back into one homogeneous population or become a hybrid swarm (Nolte and Tautz, 2010). When hybrids are fit, but cannot backcross, a novel hybrid species may develop (Rieseberg, 1997). Lastly, when hybrids maintain intermediate fitness and interfertility with parental types, the parental populations may continue to diverge into separate species, but may do so more slowly than under reinforcement (Slatkin, 1987; Nosil, 2008).

While these four outcomes of secondary contact are the subjects of frequent investigation in diploid systems, the outcomes of secondary contact between lineages that differ in ploidy level are not well known, in part because diploids and polyploids are often assumed to be reproductively isolated. Changes in ploidy are thought to rapidly confer strong reproductive isolation between cytotypes (Coyne and Orr, 2004), and there is evidence of reinforcement between cytotypes in polyploid complexes driven by selection against unfit hybrids (Husband and Sabara, 2004; Moyle et al., 2004). Within polyploids, multiple studies have demonstrated the establishment of reproductively

isolated hybrid species following hybridization and concomitant genome duplication between two diploid lineages (e.g. Abbott and Lowe, 2004; Hegarty et al., 2012). However, recent work has demonstrated that different cytotypes within polyploid complexes—groups of related individuals that differ in ploidy level—can sometimes interbreed (Hülber et al., 2015), and that the success of such interbreeding is largely based on parental cytotype, with higher-order polyploids being more capable of interbreeding than diploids and tetraploids (Sonnleitner et al., 2015; Chapter 3). This suggests that zones of secondary contact in some polyploid complexes may be similar to those observed in homoploid lineages. If so, we can use the relatively well-understood processes found in homoploid contact zones to predict rates of gene flow and divergence among polyploid complexes.

Reproduction between cytotypes in polyploid complexes depends not only on parental cytotypes, but also on the direction of between-ploidy crosses and on the resultant cytotypes of hybrid and backcross offspring. Previous work suggests that reproduction between cytotypes may depend on whether the higher-ploidy parent is maternal or paternal (Chapter 3). Crosses between diploid and tetraploid *Arabidopsis* show higher germination when the maternal plant has a higher cytotype due to amelioration of otherwise aberrant endosperm development (Scott et al., 1998; Stoute et al., 2012). If this finding is broadly applicable across polyploid complexes, it predicts that not only do higher-order polyploids have a higher likelihood of successful reproduction between ploidy levels, but that reproductive success, and therefore gene flow, may be asymmetric. If interploid hybrids (i.e. those formed between different ploidy levels) can reproduce with parental cytotypes, offspring could be expected to contain aneuploid

chromosome counts due to aberrant meiosis in the hybrid parent. For example, gametes from a 3X interploid hybrid may contain one or two copies of each chromosome. When crossed to a 2X individual that produces uniformly haploid gametes, offspring may contain two copies of some chromosomes and three copies of others. Such aneuploidy is expected to have severe fitness effects on the backcross progeny, yet examples of phenotypically normal aneuploids and apparent backcross progeny are known in polyploid complexes (McAllister, 1972; Hülber et al., 2015).

Clarifying the role of interploid hybridization is key to understanding the evolutionary history of polyploids within angiosperms. The diversification of polyploid lineages has been a recent topic of debate (Soltis et al., 2014; Mayrose et al., 2015). Although phylogenetic studies have posited that polyploidy accounts for the species richness of many large angiosperm families (Soltis et al., 2009), more recent genomic analyses suggest that polyploids diversify more slowly than diploids (Mayrose et al., 2011; Arrigo and Barker, 2012). Higher extinction rates in polyploid lineages are hypothesized to be the driving factor behind this slower diversification rate (Mayrose et al., 2011), but this hypothesis is driven exclusively by long-term evolutionary patterns, and does not explore the possibility that reproductive compatibility, and by extension gene flow, among cytotypes within polyploid complexes may slow polyploid divergence relative to diploids.

The *Campanula rotundifolia* polyploid complex offers an opportunity to explore the dynamics that govern interploid reproduction in mixed-ploidy contact zones. Previous work shows that interploid reproduction is highly restricted between diploids and tetraploids, but possible between tetraploids and hexaploids (Chapter 3). These dynamics

may create patterns of gene flow that differ with cytotypic composition of the contact zones. However, the ability of interploid hybrids to successfully backcross with parental cytotypes—a necessary condition for the establishment of interploid gene flow—is little known. Here we characterize the ability of F1 interploid hybrids to backcross with the parental cytotypes from which they derive. By quantifying the rates of backcrossing and the cytotypes of backcrossed progeny, we can assess potential for interploid gene flow in a polyploid complex. A greater understanding of this gene flow can inform investigation into the roles isolation and gene exchange play in long-term patterns of polyploid evolution and divergence. Specifically, we address the following questions: 1) Does F1 cytotype, parental cytotype, or cross direction affect the likelihood for interploid gene flow through backcrosses? 2) What are the relative effects of backcross seed set and germination rate on potential for interploid gene flow? 3) What is the cytotypic makeup of backcross progeny, and how common are aneuploids among backcrosses?

Materials and methods

Campanula rotundifolia is a short-lived perennial wildflower that exists as an autopolyploid complex with three dominant cytotypes: diploid (2n= 34 chromosomes), tetraploid (2n = 68 chromosomes), and hexaploid (2n = 102 chromosomes; Kovanda, 1966; Stevens et al., 2012). It has a circumboreal distribution, and is common throughout much of the northern latitudes of Europe (Stevens et al., 2012) and North America (Giblin, 2005). The three cytotypes are not uniformly distributed. Tetraploids are common throughout the range, but diploids are restricted almost exclusively to northern and central Europe, and hexaploids are found only in the western British Isles and in central and western North America (Shetler, 1982; Stevens et al., 2012; Chapter 1).

We assessed the probability of gene flow via interploid hybrids using backcrosses between interploid F1s generated from 2X-4X and 4X-6X crosses from a previous study (Chapter 3) and the parental populations used to create those hybrids. Each ploidy pair (2X-4X or 4X-6X) was replicated with two sets of parental populations. To approximate conditions found in natural mixed-ploidy contact zones, we chose parental population pairs that were in close geographic proximity or, if no such pairs were available, populations that were genetically similar (Table S1). From each parental population pair, two types of interploid F1 hybrids were originally created, one in each reciprocal crossing direction. These reciprocal interploid F1s differed in germination but in no later fitness traits (Chapter 3), and backcrosses derived from these F1s showed no significant differences in fitness traits based on parental crossing direction (results not shown), so the two cross directions of interploid F1s were combined within population pair for the current study.

3X and 5X cytotypes were the predicted outcomes, respectively, of 2X-4X and 4X-6X interploid crosses. However, both population pairs within the 2X-4X ploidy pair generated three cytotypes: 2X, 3X, and 4X. 4X individuals can occur through unreduced gamete formation, while the relatively rare 2X individuals may represent anomalous meiotic events. At least ten fertile individuals were available from each population pair for all 2X, 4X, and 5X interploid F1s. However, few 3X interploid hybrids were available due to extremely low germination. Therefore, six 3X plants were chosen from each population pair. Rearing of these F1 individuals is described in Chapter 3, and plants were maintained in a greenhouse on a 16/8-hour light/dark cycle for the duration of the study.

To investigate interploid backcrossing, F1 interploid hybrids were reciprocally backcrossed to parental population individuals. Parental population individuals were derived from intrapopulation crosses, but were the same generation as F1s. Each of the F1 hybrids was reciprocally crossed to two individuals from each of their parental populations, resulting in each F1 being used four times as maternal and paternal parents (Figure S1). This resulted in up to 160 planned crosses per cytotype for 2X, 4X, and 5X F1s (2 parental pairs x 10 F1 plants/pair x 2 directions x 2 parental plants x 2 replicate pollinations), and 96 planned crosses in 3X F1s (2 parental pairs x 6 F1 plants x 2 directions x 2 parental plants x 2 replicate pollinations). Insufficient flowers lowered total number of crosses slightly to 148-157 for 2X, 4X, and 5X F1s, and 88 for 3X F1s (Table S2). Because some populations of *C. rotundifolia* are self-compatible (Chapter 2), all flowers chosen for pollinations were emasculated by physically removing the anthers in the unopened bud prior to anthesis. Crosses were performed by using the pollen-covered style of a paternal flower to brush a surplus of pollen on the stigmatic lobes of the maternal flower. Mature fruits were collected just prior to dehiscence, approximately 21 days after pollination.

Backcross success was measured using two fitness traits: seed number and germination proportion. All medium to dark brown seeds at least 0.3 mm long were counted as mature seeds, with all light or uninflated structures scored as aborted. Once counted, up to ten seeds from each mature fruit were planted in two replicates of five. For crosses with fewer seeds, all available seeds were sown. Only 22 mature seeds were obtained from all 3X-2X crosses. These seeds were sown individually. Seeds were germinated on a 12/12-hour light/dark cycle at 22 C/15 C, and scored every 48 hours.

Seeds were considered to have germinated once both cotyledons were opened and the seed coat had been shed. Germination scoring continued for six weeks, and germinants were randomly thinned to one per cell. Seedlings were then grown for an additional eight weeks to obtain sufficient leaf tissue for cytometric analysis. Earlier work found no difference in survival of interploid offspring relative to controls (Chapter 3), so survival to adulthood was recorded but not used as a fitness trait.

Backcross cytotypes were estimated via flow cytometry (Otto 2-step protocol; Otto, 1990). No backcrosses derived from 3X F1s survived long enough to be cytotyped. A total of 360 plants were cytotyped, 15 from each cross type, representing 38% of all diploid-tetraploid F1s and 19% of all tetraploid-hexaploid F1s. As internal controls, approximately 20 mg of radish (*Raphanus sativus* 'Saxa': DNA content 1.11 pg/2C) was co-chopped into samples derived from crosses with a 2X parent, and 20 mg of soybean (*Glycine max* 'Polanka': DNA content 2.50 pg/2C) included with all other samples. Approximately 30 mg of fresh *C. rotundifolia* leaf tissue was collected from backcross rosettes, and chopped finely into 1 mL of Otto I buffer. The sample was then strained through a 30 μm nylon filter and incubated at room temperature for approximately 1 hr. Otto II buffer, containing 50 ug/mL of RNAse A and 50 ug/mL of propidium iodide, was then added and incubated for 10-15 minutes prior to analysis. Samples were visualized using a FACSCalibur flow cytometer with a 488-nm laser. We compared relative fluorescence of unknown samples to our internal and external standards to estimate relative DNA quantity. Plants within two standard deviations of the nearest whole-number multiple of the ratio of diploid *C. rotundifolia* over the internal standard were classified as euploid. Plants that fell outside these intervals were classified as aneuploid.

Statistical Analysis

Backcross success was assessed as the product of relative seed set and relative germination. Because maximum seed set is often determined by ovule number, which may vary among population and cytotype, backcross seed set was standardized relative to the average seed set of the maternal population in intraploid crosses. Relative germination was standardized to the average germination of both parental populations in intraploid crosses. An overall fitness measure was calculated as the product of relative seed set and relative germination.

To assess variation in seed set, germination and overall fitness, we used a generalized mixed model (PROC GLIMMIX, SAS 9.3 SAS Institute, INC. 2011). F1 cytotype (2X, 3X, 4X, or 5X), parental cytotype (2X or 4X, 4X or 6X), and backcross cross direction (F1 hybrid as maternal or paternal parent) were main effects. Parental population pair nested within F1 cytotype and F1 cross-direction were random effects. Independent contrasts were used to compare intraploid to interploid backcrosses of 2X and 4X F1s, 2X and 4X intraploid backcrosses to each other, and 5X backcrosses to each parental cytotype. For sake of clarity, crosses will be reported hereafter by listing the F1 hybrid first, then the parental cytotype (e.g. 4X F1-4X describes the intraploid cross between 4X F1 hybrids and their 4X parental population).

Results

The combination of F1 cytotype and parental cytotype influenced crossing success and offspring fitness (Table 1). Backcrosses were most successful when F1 cytotype matched parental cytotype, and when pentaploid F1 hybrids were backcrossed

against hexaploids (Figure 1). F1 cross direction (whether the F1 served as maternal or paternal parent) was not significant for any crosses.

Seed set was highest in the intraploid backcrosses, 2X F1-2X and 4X F1-4X (Figure 1a), and was significantly higher than interploid backcrosses of the same F1 cytotype (intraploid vs. interploid contrast: $F_{1,144} = 11.86$, $p < 0.001$). Although seed set was high for both intraploid crosses, it was significantly higher for 4X F1-4X crosses than 2X F1-2X crosses (Figure 1a; 2X F1-2X vs. 4X F1-4X contrast: $F_{1,144}=5.44$, $p = 0.02$). While both intraploid crosses set similar maximum quantities of seed (45 seeds in 2X F1-2X and 48 seeds in 4X F1-4X crosses), 2X F1-2X crosses had substantially more fruits that failed to set any seed. This resulted in 4X F1-4X crosses setting approximately 92% as many seed as control crosses while 2X F1-2X crosses on average set approximately 63% as many seed.

All interploid crosses between F1 hybrids and 2X or 4X parents suffered low seed set (Figure 1a). 2X F1-4X and 4X F1-2X crosses both produced less than 13% as many seed as control crosses. All crosses with 3X parents produced little to no seed; 3X F1-2X crosses combined only produced 22 seeds (1% relative seed set), and 3X F1-4X crosses combined produced 103 seeds (5.5% relative seed set). Seed set in 5X F1-6X crosses was 77.6% that of the control crosses and roughly double that of the 5X F1-4X crosses (Figure 1a; 5X F1-4X vs. 5X F1-6X contrast: $F_{1,144} = 6.32$, $p = 0.01$).

Germination largely followed the same pattern as seed set, with intraploid crosses and 5X F1-6X crosses having high germination rates (Figure 1b). Although 2X and 4X intraploid backcross had similar germination rates, 5X F1-4X crosses had approximately 25% lower germination than 5X F1-6X crosses (5X F1-4X vs. 5X F1-6X contrast: $F_{1,144}$

= 5.83, p = 0.02), however it was still higher than other interploid crosses. Both 2X F1-4X and 4X F1-2X crosses had relative germination approximately one quarter that of intraploid crosses (Figure 1b). 3X F1-2X crosses produced no germinants, while 3X F1-4X crosses had a total of 6 germinants (5.8% germination rate), none of which survived long enough to obtain tissue for flow cytometry. All other germinant classes had similar high survivorship (> 93%).

The similar patterns observed in both seed set and germination proportion were in overall backcross fitness; 2X and 4X intraploid and 5X F1-6X crosses had the greatest reproductive success. Intraploid crosses 2X F1-2X and 4X F1-4X crosses had 52.4% and 85.5% respectively fitness relative to control crosses (Figure 1c). Interploid crosses between 2X and 4X hybrids and 2X and 4X parents, by contrast, had overall fitness that was effectively zero. Among 5X F1 hybrids, backcross success was asymmetric with respect to parental cytotype. 5X F1-4X crosses were only approximately 26% as fit as controls, while the more successful 5X F1-6X crosses were approximately 67% as fit (5X F1-4X vs. 5X F1-6X contrast; $F_{1,144} = 11.62$, $p < 0.001$).

Potential interploid gene flow depends not only on backcross fitness, but also on backcross ploidy level. 2X F1-2X and 4X F1-4X crosses produced almost exclusively 2X and 4X offspring, respectively, although three backcross offspring had PI fluorescence values slightly above those expected for either 2X or 4X individuals and were scored as aneuploid (Figure 2). Interploid crosses involving 2X and 4X F1s (2X F1-4X and 4X F1-2X) produced mixtures of both 3X and 4X individuals, with no progeny displaying fluorescence values associated with aneuploidy (Figure 2). In contrast, progeny from 5X F1-4X crosses showed considerable variation; 5% were consistent with 4X, 70% with

aneuploidy between 4X and 5X, and 25% with 5X (Figure 2). Progeny from 5X F1-6X crosses were almost all aneuploid between 5X and 6X, with two individuals showing fluorescence consistent with 5X.

Discussion

The evolutionary outcome of secondary contact between cytotypes in a polyploid complex depends on whether gene flow can occur, which in turn depends on whether different cytotypes can form hybrids and if those hybrids can backcross. We found that 3X hybrids generated from 2X-4X crosses were incapable of backcrossing. By contrast, 2X and 4X hybrids of those same crosses were capable of backcrossing with an individual of their same ploidy, could not reproduce with an individual that differed in ploidy. In contrast, the 5X hybrids generated from 4X-6X crosses could backcross successfully to both parental cytotypes. These results suggest that the capacity for gene flow in mixed-ploidy contact zones depends upon the cytotypes comprising each contact zone. Contact zones composed of two polyploids that differ in cytotype may experience greater gene flow, and therefore decreased divergence, than contact zones comprising diploids and tetraploids.

Backcrosses were most successful between plants of the same ploidy, and least successful if one of the parents was triploid. The expected F1 interploid cytotypes, 3X from 2X-4X crosses and 5X from 4X-6X crosses, had markedly different reproductive success. Backcrosses with triploids produced less than 3% of the seed of either intraploid control cross, while backcrosses with pentaploids produced on average 59% relative seed set. This is consistent with previous work that found higher pollen fertility in pentaploids than triploids (Hülber et al., 2015; Chapter 3) and with the broad literature that finds poor

fitness in triploid interploid hybrids, to the point that such postzygotic reproductive isolation is termed “triploid block” (Husband and Sabara, 2004; Köhler et al., 2010). Backcrosses of hybrid 2X and 4X cytotypes had germination commensurate with controls in intraploid backcrosses, but highly reduced germination in interploid backcrosses. The poor performance of interploid 2X-4X backcrosses is consistent with the parental crosses that generated the F1 generation (Chapter 3), and likely reflects aberrant endosperm development due to parental genome imbalance (Scott et al., 1998; Stoute et al., 2012; Birchler, 2014). However, while backcross 4X cytotypes had little loss of seed set, 2X seed set was reduced, suggesting that the diploid F1s are not fully analogous to parental diploids.

The similar patterns of backcrossing on seed set and germination led to strong differences in overall reproductive success. Triploids had little ability to produce seeds capable of germinating, and all germinants failed to survive to adulthood providing no avenue for interploid gene exchange. Interploid backcrosses between 2X and 4X F1s and their parental cytotypes were only marginally more successful, and would likely be poor conduits for gene flow in a mixed-ploidy contact zone comprising diploids and tetraploids. By contrast, intraploid 2X and 4X backcross offspring had substantial reproductive success, particularly 4X offspring that were only slightly less fit than control crosses. Although 2X and 4X F1s likely resulted from unreduced gamete production and other meiotic irregularities, they may serve as an occasional mechanism for interploid gene flow in mixed-ploidy contact zones, particularly unidirectional gene flow from diploids to tetraploids. Such unidirectional interploid gene flow has occasionally been documented in other systems (Brochmann et al., 2004). 5X backcross offspring likewise

showed asymmetry in overall reproductive success, with 5X F1-6X crosses over twice as successful as 5X F1-4X crosses. This suggests that, overall, contact zones containing 4X and 6X cytotypes may experience greater interploid reproduction than those containing diploids and tetraploids, but gene flow may still be expected to proceed mostly from the 4X to the 6X cytotype.

We found that interploid hybrids derived from unreduced gametes in 2X-4X contact zones, as well as pentaploids formed in 4X-6X contact zones, may exhibit sufficient backcross success to serve as conduits for gene flow. However, the net effect of these interploid F1 hybrids on gene flow is informed by their frequency in contact zones. In *C. rotundifolia*, germination of 2X-4X F1 seeds was only one-sixth that of 4X-6X seeds, and of those few germinants, only 40% were highly fit tetraploids, and another 5% diploids (Chapter 3). Unlike that study, in which all plants were given pollen that did not match their ploidy, plants in natural mixed-ploidy populations would receive a mixture of pollen, and frequency of interploid hybrids would be further reduced. Furthermore, low germination has been observed in diploid-tetraploid interploid crosses in other systems (Sabara et al., 2013; Roccaforte et al., 2015). Therefore, although diploid and tetraploid hybrid F1s retain high capacity for backcrossing and could hypothetically mediate ongoing gene flow in mixed-ploidy contact zones, their low frequency limits their overall contribution. In contrast, pentaploids are commonly formed in 4X-6X crosses (Chapter 3), experimental mixed-ploidy arrays where plants received both inter- and intraploid pollen (Chapter 5), and natural mixed-ploidy contact zones (Chapter 6). As such, their capacity for backcrossing contributes substantially to potential gene flow in mixed 4X-6X populations.

Aneuploid and odd-numbered euploid plants often show reduced viability and fertility relative to even-numbered euploid plants. As such, the cytotypes that result from backcrosses may determine capacity for gene flow in mixed-ploidy contact zones. Euploid backcross offspring were found in four backcross classes: 2X F1-2X and 4X F1-4X, which produced nearly uniform diploid and tetraploid backcross offspring, respectively, and 2X F1-4X and 4X F1-2X, which produced both triploid and tetraploid offspring. Backcrosses between 5X F1s and either parental cytotype, by contrast, produced substantial frequencies of aneuploid offspring. The ability of 2X and 4X F1s to produce euploid offspring raises the likelihood of interploid gene flow in mixed 2X-4X contact zones, but the low frequency of such F1s means that interploid gene flow would remain relatively rare. Conversely, the aneuploids produced by backcrosses of 5X F1s suggests that gene flow may be lower than expected solely based on backcross reproductive success. However, evidence of interploid reproduction in other systems (Hülber et al., 2015; Sonnleitner et al., 2016) and apparent pollen fertility of near-5X aneuploids in *C. rotundifolia* (McAllister, 1972; Sutherland, unpub.) suggests that aneuploids derived from 5X F1s may successfully reproduce, albeit likely at lower rates than euploid individuals.

In total, the capacity for interploid reproduction in the *C. rotundifolia* polyploid complex may result in three of the four possible outcomes when different cytotypes come into secondary contact. Contact zones comprising diploid and tetraploid cytotypes will demonstrate strong interploid reproductive isolation and have little to no interploid gene flow (cf. Coyne and Orr, 2004; Husband and Sabara, 2004). These lineages should show relatively high rates of interploid divergence with only occasional signatures of

introgression and will likely develop secondary reinforcement, similar to homoploid systems in which reproductive isolation is high (Han et al., 2015). We expect introgression in 2X-4X contact zones, when it happens, to be mostly unidirectional, through rare, fertile tetraploids formed via unreduced gamete production (Zohren et al., 2016; Kreiner et al., 2017).

Alternatively, contact zones comprising tetraploids and hexaploids will demonstrate little loss of fitness in the hybrid pentaploids and low-to-moderate loss of fertility in backcrosses, particularly in crosses between pentaploids and hexaploids. This backcrossing ability will likely result in interploid reproduction and gene flow, depending on cytotype frequency and evenness in 4X-6X contact zones. In contact zones in which one parental cytotype is infrequent, formation of viable but less fertile hybrids may cause insufficient selection to drive secondary reinforcement and has been predicted to result in extirpation of the minority parental cytotype and gradual elimination or absorption of hybrids within the majority cytotype (Husband, 2000). If 4X and 6X individuals are both frequent and well-dispersed within a contact zone, the asymmetric reproductive capacity between pentaploids and their progenitor cytotypes may cause a hybrid swarm between pentaploids and hexaploids that leaves tetraploids largely intact. Regardless of the outcome of interploid reproduction in any specific contact zone, the formation of viable pentaploid hybrids in tetraploid-hexaploid contact zones that are capable of reproduction with parental types is a striking difference from diploid-tetraploid contact zones, and homogenizing interploid gene flow is likely to be much higher among polyploids than between diploids and polyploids.

The longstanding view is that strong reproductive isolation is characteristic of polyploids and contributes to high rates of diversification among lineages enriched for polyploid taxa (Meyers and Levin, 2006). From a taxonomic perspective, this view has merit since some of the most speciose angiosperm lineages are also the most polyploid (Otto and Whitton, 2000). Recent genomic work, however, suggests that polyploid lineages actually diversify less than diploids, but posits that this decreased diversification is due to higher extinction rates among polyploid taxa (Mayrose et al., 2011; Arrigo and Barker, 2012). While this conclusion has been questioned by others on methodological grounds (Soltis et al., 2014), it is possible that another explanation exists for apparent decreased polyploid divergence. If higher-order polyploid lineages are capable of higher rates of interploid gene flow than is possible between diploids and tetraploids, these lineages would be expected to diverge more slowly from each other. This slower divergence may account for some of the lack of diversification observed in polyploid lineages.

Acknowledgements

We thank B Rulik, Z. Munzbergova, and T. Schmoll for the original seed collections of European populations. We would like to thank the National Science Foundation (NSF DEB-1020717), the Society for the Study of Evolution, the Torrey Botanical Society, and the Botanical Society of America for funding for this project.

References

Abbott, R. J., and A. J. Lowe. 2004. Origins, establishment and evolution of new polyploid species: *Senecio cambrensis* and *S. eboracensis* in the British Isles. *Biological Journal of the Linnean Society* 82: 467–474.

Arrigo, N., and M. S. Barker. 2012. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* 15: 140–146.

Birchler, J. A. 2014. Interploidy hybridization barrier of endosperm as a dosage interaction. *Frontiers in Plant Science* 5: 1–4.

Brochmann, C., A. K. Brysting, D. E. Soltis, P. S. Soltis, I. J. Leitch, and C. J. Pires. 2004. Polyploidy in arctic plants. *Biological Journal of Linnaean Society* 82: 521-536.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer. New York, NY.

Ellstrand, N. C., N. C. Prentice, and J. F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30: 539–563

Giblin, D. E. 2005. Variation in floral longevity between populations of *Campanula rotundifolia* (*Campanula*ceae) in response to fitness accrual rate manipulation. *American Journal of Botany* 92: 1714–1722.

Haig, D., and M. Westoby. 1988. On limits to seed production. *The American Naturalist*. 131: 757-759

Han, T.-S., Q. Wu, X.-H. Hou, Z.-W. Li, Y.-P. Zou, S. Ge, and Y.-L. Guo. 2015. Frequent introgressions from diploid species contribute to the adaptation of the tetraploid shepherd's purse (*Capsella bursa-pastoris*). *Molecular Plant* 8: 427–438.

Hegarty, M. J., R. J. Abbott, and S. J. Hiscock. 2012. Allopolyploid speciation in action: The origins and evolution of *Senecio cambrensis*. *In* Polyploidy and Genome Evolution, 245–270. Springer Berlin Heidelberg, Berlin, Heidelberg.

Husband, B. C. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings of the Royal Society B: Biological Sciences* 267: 217–223.

Husband, B. C., and H. A. Sabara. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* 161: 703–713.

Hülber, K., M. Sonnleitner, J. Suda, J. Krejčíková, P. Schönswetter, G. M. Schneeweiss, and M. Winkler. 2015. Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio*

carniolicus (syn. *Jacobaea carniolica*, Asteraceae). *Ecology and Evolution* 5: 1224–1234.

Kovanda, M. 1966. Some chromosome counts in the *Campanula rotundifolia* complex II. *Folia Geobotanica & Phytotaxonomica Bohemoslovaca* 12: 23-89.

Köhler, C., O. Mittelsten Scheid, and A. Erilova. 2010. The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* 26: 142–148.

Kreiner, J. M., P. Kron, and B. C. Husband. 2017. Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. *New Phytologist* 1–11. doi: 10.1111/nph.14423

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S.P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257–1257.

Mayrose, I., S. H. Zhan, C. J. Rothfels, N. Arrigo, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2015. Methods for studying polyploid diversification and the dead-end hypothesis: a reply to Soltis et al. (2014). *New Phytologist* 206: 27–35.

McAllister, H. A. 1972. The Experimental Taxonomy of *Campanula rotundifolia* L. Ph.D. Thesis, University of Glasgow, Glasgow, UK.

Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. *Evolution* 60: 1198–1206.

Moraes, A. P., M. Chinaglia, and C. P. Silva. 2013. Interploidy hybridization in sympatric zones: the formation of *Epidendrum fulgens* × *E. puniceoluteum* hybrids (Epidendroideae, Orchidaceae). *Ecology and Evolution* 3: 3824 - 3837.

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

Nolte, A. W., and D. Tautz. 2010. Understanding the onset of hybrid speciation. *Trends in Genetics* 26: 54–58.

Noor, M. 1999. Reinforcement and other consequences of sympatry. *Heredity* 83: 503–508.

Nosil, P. 2008. Speciation with gene flow could be common. *Molecular Ecology* 17: 2103–2106.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual review of genetics* 34: 401–437.

Otto, F. J. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewickz Z, Crissman HA, eds. Methods in cell biology, vol. 33. San Diego, CA, USA: Academic Press, 105–110.

Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics*. 28: 359-389.

Roccaforte, K., S. E. Russo, and D. Pilson. 2015. Hybridization and reproductive isolation between diploid *Erythronium mesochoreum* and its tetraploid congener *E. albidum* (Liliaceae). *Evolution* 69: 1375–1389.

Sabara, H. A., P. Kron, and B. C. Husband. 2013. Cytotype coexistence leads to triploid hybrid production in a diploid-tetraploid contact zone of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 100: 962–970.

Schluter, D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16: 372–380.

Scott, R. J., M. Spielman, J. Bailey, H. G. Dickinson. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125: 3329-3341.

Shetler, S. G. 1982. Variation and evolution of the nearctic Harebells (*Campanula* sect. Heterophylla). J. Cramer. Vaduz.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236: 787–792.

Soltis, D. E., M. C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, M. B. Cortez, P. S. Soltis, and M. A. Gitzendanner. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytologist* 202: 1105–1117.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, C. W. dePamphilis, P. K. Wall, and P. S. Soltis. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Sonnleitner, M., K. Hülber, R. Flatscher, P. Escobar García, M. Winkler, J. Suda, P. Schönswetter, and G. M. Schneeweiss. 2015. Ecological differentiation of diploid and polyploid cytotypes of *Senecio carniolicus* sensu lato (Asteraceae) is stronger in areas of sympatry. *Annals of Botany* 117: 269-276.

Stevens, C. J., J. Wilson, and H. A. McAllister. 2012. Biological flora of the British Isles: *Campanula rotundifolia*. *Journal of Ecology* 100: 821–839.

Stoute, A. I., V. Varenko, G. J. King, R. J. Scott, and S. Kurup. 2012. Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *The Plant Journal* 71: 503-516.

Zohren, J., N. Wang, I. Kardailsky, J. S. Borrell, A. Joecker, R. A. Nichols, and R. J. A. Buggs. 2016. Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Molecular Ecology* 25: 2413–2426.

Table 1) Analysis of variance testing the effects of F1 cytotype (2X, 3X, 4X, or 5X), parental cytotype (2X or 4X for 2X, 3X, 4X F1s; 4X or 6X for 5X F1s), and cross direction (F1 plant as the mother or father) on seed set, germination and fitness. Population pair (nested within F1 cytotype) and original cross direction (of F1s) were included as random effects. F-values listed. *** p-value <0.0001, all others p > 0.05.

Source	df	Seed Set	Germination	Fitness
F1 Cytotype	3	16.76***	10.83***	15.81***
Parental Cytotype	2	16.49***	13.18***	19.03***
Cross Direction	1	0.47	1.37	1.22
F1 x Parent	2	19.79***	9.80***	14.73***
F1 x Direction	3	0.49	1.82	0.83
Parent x Direction	2	0.45	3.85	0.97
Three-Way Interaction	2	0.06	4.26	1.02
Error	145			

Figure 1) Backcross relative (a) seed set, (b) germination proportion, and (c) overall fitness by F1 cytotype and parental cytotype for *Campanula rotundifolia*. Seed set standardized based on maternal population. Germination was standardized based on both parental populations. Cytotypes crossed shown on x-axis. Error bars denote standard error.

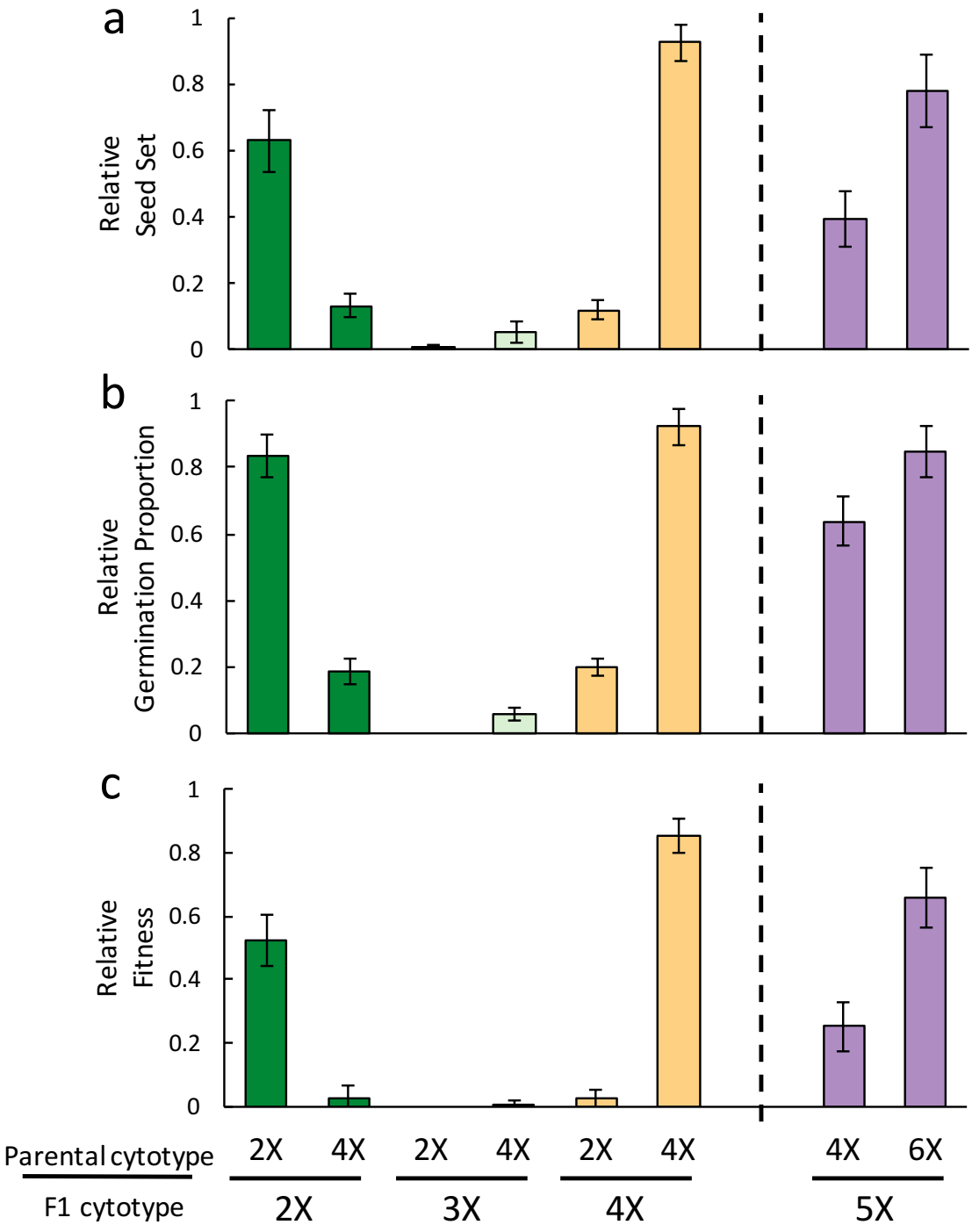


Figure 2) Cytotypic composition of *Campanula rotundifolia* backcross progeny. "+" denotes aneuploid cytotypes. Cytotypes crossed shown on x-axis.

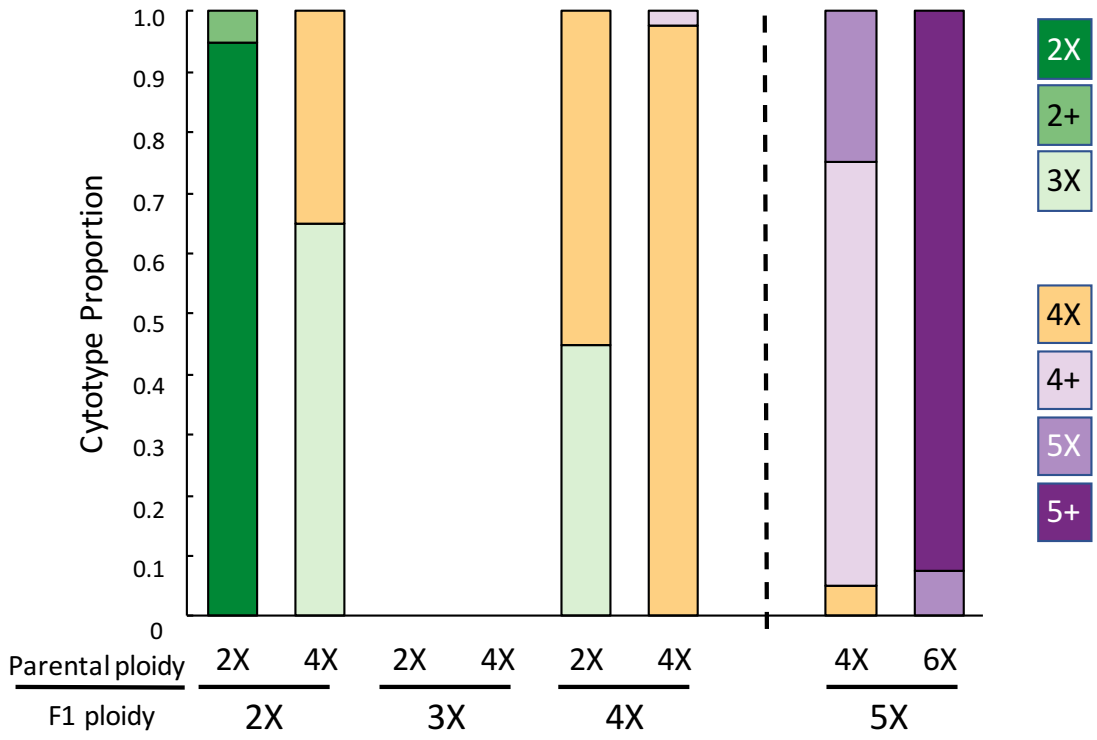


Table S1) Parental *Campanula rotundifolia* population pairs from which F1 hybrids were created. Each pair is delineated by a horizontal line.

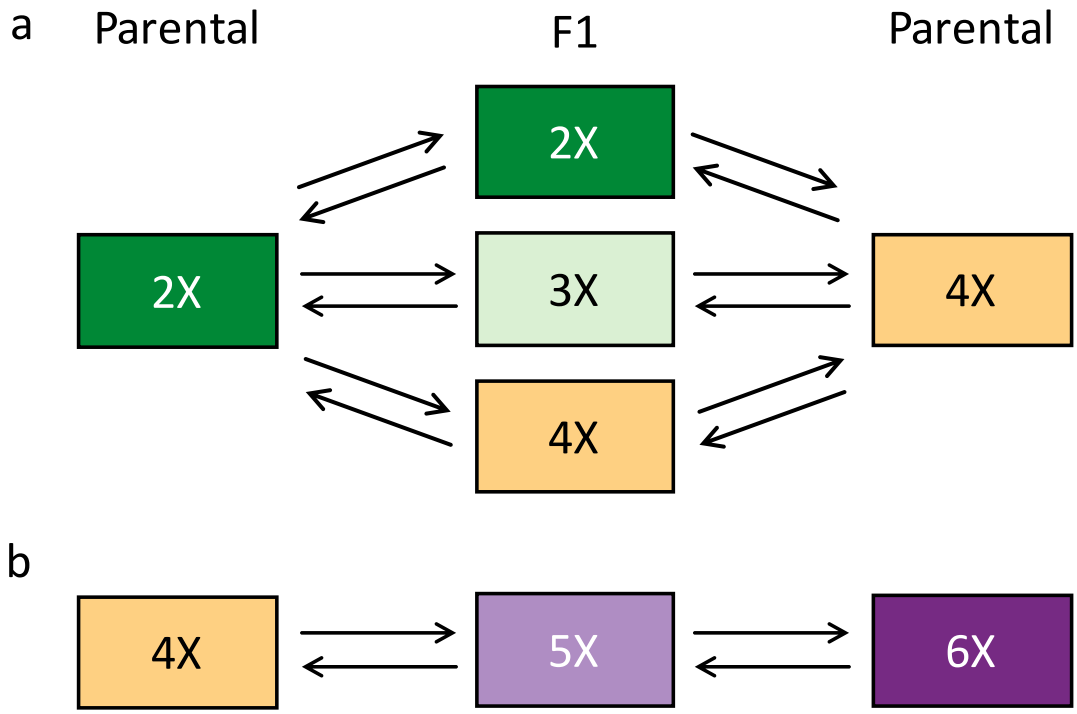
Population ID	Ploidy	Locality	Country	Latitude	Longitude
25	2	Zamky	Czech Republic	50.429255	14.075625
24	4	Vinaricka Hora	Czech Republic	50.184791	14.081019
23	2	Dresden	Germany	50.973440	13.821118
33	4	Mittelndorf	Germany	50.937567	14.203496
30	4	Turner's Falls, MA	United States	42.595089	-72.581203
10	6	Red Bay, ON	Canada	44.416252	-81.458360
115	4	Misery Bay PP, ON	Canada	45.800465	-82.749153
116	6	South Baymouth, ON	Canada	45.558822	-82.018024

Table S2) Number of backcrosses performed. Ploidy pair refers to parental cytotypes used to create F1s. Population pair refers to specific populations crossed, see Table S1.

Ploidy Pair	Population Pair	F1 Ploidy	F1 Number	2X Backcross	4X Backcross
2X - 4X	24 - 25	2X	10	18	20
		3X	6	12	12
		4X	10	17	19
2X - 4X	23 - 33	2X	10	20	20
		3X	6	12	12
		4X	10	17	20

Ploidy Pair	Population Pair	F1 Ploidy	F1 Number	4X Backcross	6X Backcross
4X - 6X	26 - 30	5X	10	20	16
4X - 6X	115 - 116	5X	10	19	20

Figure S1) Schematic diagram of crossing design. F1 plants (middle column) were reciprocally crossed to both parental cytotypes. Three different F1 cytotypes were crossed against parental 2X and 4X cytotypes (a), while only one F1 cytotype was crossed against 4X and 6X parental cytotypes (b). Crosses were repeated for both population pairs per cytotype, and for two replicates per arrow.



CHAPTER FIVE:

Reproduction between cytotypes in polyploid complexes is mediated by parental ploidy level and population composition

Premise of the study – Some researchers have called into question the assumption of strong reproductive isolation between plants differing in ploidy level, especially among higher order polyploids, a view supported by the current research (Chapters 3, 4). Here we assess patterns of interploid reproductive isolation across cytotypes in the *Campanula rotundifolia* polyploid complex, focusing on prezygotic barriers related to pollinator preference and postpollination, as well as postzygotic barriers which determine interploid hybrid production. We also investigate the contribution of cytotype frequency within contact zones to hybridization.

Methods – Experimental mixed-ploidy arrays of *C. rotundifolia* were exposed to 5-day periods of natural pollination during which insect visitation was observed. Arrays varied in which parental cytotypes were paired (2X-4X or 4X-6X) and in cytotype frequency. Seed set was scored and interploid hybridization evaluated by cytotyping offspring via flow cytometry.

Key results – Pollinator movement was random with respect to both cytotype and frequency. Germination did not differ among cytotypes or frequencies, however seed set was less when cytotypes were uncommon within arrays. Few triploid hybrids (2.75%) were formed in 2X-4X arrays, but substantial numbers of pentaploid hybrids (36%-57%) were produced 4X-6X arrays. In both cytotype combinations, interploid hybrids were most common when the maternal cytotype was least frequent.

Conclusions – Although pollinators did not differ in preference, marked differences in offspring cytotype frequencies were observed relative to that expected by random pollinator movement. This contrast indicates that in *C. rotundifolia* prezygotic barriers

are weak, while postpollination barriers are strong between diploids and tetraploids but permissive between tetraploids and hexaploids. F1 cytotypes also showed that interploid reproduction is more likely when a cytotype is rare in a population and therefore receiving proportionally more interploid pollen. These results suggest that interploid reproduction, and by extension gene flow, is greater among polyploid cytotypes than between diploids and polyploids. Such increased gene flow could provide a causal explanation for the recent findings of lower than expected diversification in many angiosperm polyploid lineages.

Introduction

Whole genome duplication, or polyploidy, is an important speciation and diversification mechanism in angiosperms (Soltis et al., 2007; Wood et al., 2009). It is often considered an "instant" speciation process (Coyne and Orr, 2004; Parisod et al., 2010), due to the rapid development of strong reproductive barriers following polyploidization. Reproductive isolation between diploids and related tetraploids is well-supported (e.g. Husband and Schemske, 2000; Thompson and Merg, 2008; Roccaforte et al., 2015), but is poorly understood between higher-order polyploids. This oversight is important because polyploid complexes, groups of interrelated taxa comprising two or more ploidy levels, are common throughout angiosperms (Stebbins, 1942; Guggisberg et al., 2009; Lo et al., 2010; Green et al., 2011). Recent evidence suggests reproductive isolation may be lower between polyploids than between diploids and polyploids (Greiner and Oberprieler, 2012; Hülber et al., 2015; Chapter 3). If reproduction between cytotypes occurs, gene flow may be ongoing within polyploid complexes. While gene flow does not always constrain divergence and speciation (Nosil, 2008), it often shapes broad patterns of diversification (Petit and Excoffier, 2009; Ravinet et al., 2017). Increased gene flow may provide a causal explanation for lower divergence rates among polyploids than among diploids across the angiosperms (Mayrose et al., 2011).

Interploid reproductive isolation results from both prezygotic and postzygotic reproductive barriers (Husband and Schemske, 2000) and can vary with cytotype and population distribution (Kolar et al., 2009). Prezygotic barriers include differences in phenology, differential pollinator preferences, and post-pollination stylar and pollen interactions (Husband and Schemske, 2000). Pollinator preference is perhaps the best

studied prezygotic barrier in polyploid complexes (Husband and Sabara, 2004; Sonnleitner et al., 2015). Pollinators have shown distinct cytotypic preferences in several diploid-tetraploid systems, including *Heuchera grossulariifolia* (Segraves and Thompson, 1999; Thompson et al., 2004), and *Chamerion angustifolium* (Husband and Schemske, 2000), resulting in low interploid pollen exchange rates. However, pollinators in other systems have little to no ploidy-specific preference, including diploid and hexaploid *Aster amellus* (Castro et al., 2010) and diploid and tetraploid *Libidibia ferrea* (Borges et al., 2012). While floral morphology changes with ploidy in some systems, providing cues by which pollinators may discriminate (Kennedy et al., 2006; Ramsey, 2011), a number of polyploid complexes, particularly autopolyploids, contain cytotypes with few apparent differences in floral display, olfactory cues, or nectar production (Soltis et al., 2007; Jersáková et al., 2010; Wefferling et al., 2017; Eriksson et al., 2017).

Post-pollination prezygotic barriers and postzygotic barriers can also contribute to interploid reproductive isolation, particularly when sympatric cytotypes are subject to mixed pollen loads. Mixed loads of conspecific and heterospecific pollen most commonly display poor growth of heterospecific pollen tubes (Wolf et al., 2001) or preferential fertilization by conspecific pollen (Chapman et al., 2005). Some polyploid systems exhibit similar preference for intraploid pollen (analogous to conspecific pollen in homoploid systems) with almost all offspring following mixed-pollen fertilization resulting from intraploid pollen (Baack, 2005). Postzygotic barriers also contribute to overall reproductive isolation and have been documented in multiple polyploid complexes (e.g. Husband and Sabara, 2004; Ståhlberg, 2007; Chapter 3). However, interploid postzygotic barriers tend to be lower between different polyploid cytotypes

than between diploids and tetraploid (Greiner and Oberprieler, 2012; Sonnleitner et al., 2015; Hülber et al., 2015, Chapter 3). Unlike prezygotic isolation that is highly individual to system, postzygotic isolation derives from molecular and developmental irregularities, and are likely to be consistent across a wide range of polyploid taxa (Stoute et al., 2012; Lu et al., 2012; Herben et al., 2016).

The likelihood of interploid gene flow rests on two factors: the ability of interploid mating to produce viable, fertile hybrids, and the frequency with which interploid mating occurs in an area of sympatry. If prezygotic and postzygotic barriers still permit some interploid reproduction, the likelihood of such reproduction is mediated by relative frequency of each cytotype in contact. Cytotypes can come into contact under two circumstances: when newly formed polyploids are surrounded and outnumbered by related diploids or lower polyploids (Husband, 2000), or during secondary contact, when relative frequency of each cytotype may vary from parity to one ploidy outnumbering the other. While newly formed polyploids are often extirpated due to minority cytotype exclusion, under secondary contact the degree of interploid reproduction and the fitness of any hybrids determines if cytotypes will coexist without interbreeding, if one cytotype will extirpate the other, or if cytotypes will establish interploid gene flow. Given that interploid hybrids between higher cytotypes tend to be more fit (Hülber et al., 2015; Chapter 3, 4), patterns of frequency-mediated interploid reproduction may vary across a polyploid complex. In studies of interploid reproductive isolation, little attention has been focused on secondary contact, and in particular the effect of relative cytotypic frequency on interploid reproduction.

Widespread polyploid complexes often contain multiple zones of contact between related cytotypes, which may differ in relative frequency and distribution of each cytotype. To assess likelihood of interploid gene flow, it is necessary to both quantify the reproductive barriers that may limit interploid reproduction and characterize the effect of cytotype frequency at contact zones. Here we assess interploid reproduction within a polyploid complex under natural pollination conditions. Specifically, we use the *Campanula rotundifolia* polyploid complex to address the following questions: Do pollinators demonstrate cytotypic preference in mixed-ploidy populations? When both intraploid and interploid pollen are available, do plants create interploid hybrids? And is the rate of interploid hybrid production affected by parental cytotype or frequency?

Materials and Methods

Study System

Campanula rotundifolia is a short-lived perennial herb native to much of the northern latitudes of Europe and North America (Shetler, 1982; Stevens et al., 2012). It exists as a polyploid complex comprising three main cytotypes: diploid (2n=34); tetraploid (2n=68), and hexaploid (2n=102). Cytotypes are non-randomly distributed throughout the range; tetraploids are widespread, while diploids are found primarily in Central and Northern Europe, and hexaploids have relatively limited distributions in the British Isles and in Central and Western North America (Shetler, 1982; Stevens et al., 2012). Multiple zones of interploid contact between diploids and tetraploids, or tetraploids and hexaploids, have been observed (McAllister, 1972; Chapter 1).

Reproductive biology and phenology are similar across *C. rotundifolia* cytotypes. All three cytotypes flower in early to mid-summer. Tetraploids have a broad flowering

time that overlaps with both diploids and hexaploids, including in sympatry (Stevens et al., 2012; B. Sutherland, pers. obs.). Flower size and number differ across continents. North American plants have fewer, larger flowers (B. Sutherland, unpub.). Flower size is consistent across ploidy levels in European populations. Flower number can differ between diploids and tetraploids, with diploids having slightly more flowers per individual (Stevens et al., 2012). Most pollinators that visit *C. rotundifolia* are generalists, including a variety of Hymenopteran and some Dipteran species, particularly solitary bees (Megachilidae, Halictidae, Andrenidae) and bumble bees (*Bombus*; Bingham and Orthner, 1998; Stevens et al., 2012).

Although *C. rotundifolia* has historically been considered largely self-incompatible (Shetler, 1982; Bingham and Ranker, 2000), recent work has shown that self-incompatibility varies (Chapter 2). The species is self-incompatible and therefore is obligately outcrossing in Central Europe, while in Western Europe individuals are moderately self-compatible. North American populations are largely self-compatible but the extent of self-fertilization in natural populations is not known.

Array Construction

We constructed mixed-ploidy experimental arrays using four pairs of *C. rotundifolia* populations. We used these population pairs to assess the likelihood of interploid mating as a function of pollinator preference, parental cytotype, and cytotypic abundance. Populations were chosen, where possible, from naturally occurring contact zones between cytotypes (Table S1). Two diploid-tetraploid pairs were selected from central Europe—one parapatric pair from the Czech Republic, and two populations in eastern Germany separated by approximately 50 km. Two tetraploid-hexaploid pairs were

chosen from North America—one was a parapatric pair from Manitoulin Island in southern Ontario. The other included populations from Massachusetts and Ontario (approx. 740 km apart), chosen for their small genetic distance (Chapter 1), to minimize divergence despite a larger geographic distance.

Seeds that had been wild-collected in each of the eight populations were sown in the University of Virginia greenhouse, transplanted as basal rosettes into 4-inch pots, and grown to flowering. Once buds appeared, plants were acclimatized for three weeks at the Mountain Lake Biological Station (MLBS) greenhouse before use in experimental arrays. MLBS is located in Giles County, Virginia, approximately 40 km from the nearest *C. rotundifolia* population.

Each pair of populations was used to construct experimental arrays of 24 plants. Arrays were constructed and observed in June and July 2015 at MLBS. For each population pair, four cytotype frequencies were tested by varying the number of each population in the 24-plant arrays: 8% (2 plants), 25% (6 plants), 75% (18 plants), and 92% (22 plants). Plants in each array were arranged 0.5 m apart in a 4-by-6 rectangle (Figure S1). Arrays were placed in habitats comparable to those of natural populations but at least 100 meters apart to minimize cross-pollination, and surrounded by netting at a distance of 1 m to prevent deer herbivory. Three replicate arrays of each diploid-tetraploid pair of populations for each cytotype frequency, and two replicate arrays of each tetraploid-hexaploid pair for each cytotype frequency were constructed, for a total of 80 arrays (2 population pairs/ploidy pair x 2 cytotype/population pair x 4 frequencies/cytotype x 5 replicates [=3 replicates for diploid-tetraploid arrays + 2 replicates tetraploid-hexaploid arrays]).

Plants with open male- and female-phase flowers were chosen for placement in arrays for 5-day exposure periods. Both cytotypes within an array were trimmed to the same number of flowers to present similar floral displays to pollinators. European plants were trimmed to five open flowers each, and North American plants to three open flowers each. Where possible, similar numbers of male and female phase flowers were retained across both populations in each array. To account for variation in floral display that arose throughout each trial due to the opening of new flowers, the number of open male-phase and female-phase flowers on each plant was recorded on the 3rd day of array exposure.

Pollinator Observation

To determine if pollinators have a preference for either ploidy, arrays were observed four times for 30 minutes each during their 5-day exposure, once between 10:00-12:00 and once between 14:00-16:00 on the 2nd and 4th days. Prior to each observation period, a general description of weather was recorded, noting presence or absence of precipitation, cloud cover, and wind. During each observation period, a pollinator entering an array was identified to order, followed visually until they left the array, and the position of each plant and number of flowers visited was recorded. Pollinator movement was calculated as plant-level transitions. Pollinator visitation was infrequent enough that multiple pollinators were rarely in an array at one time.

Following the 5-day exposure period, plants were returned to the greenhouse and each female-phase flower was tagged. Any tagged flowers that were not wilted after 48 hours, indicating fruit set, were removed, and plants were reshuffled into new arrays and placed outside for another replicate. Up to five tagged fruits per plant per replicate were

harvested when mature. Seeds were then counted and seed set averaged per plant for each replicate.

F1 production and cytotypes

In each round of replication, 34 plants were randomly chosen from all arrays within each population pair (2 plants/cytotype at 8% frequency + 5 plants/cytotype at all other frequencies). For each of these plants, up to five seeds from each of five randomly chosen fruits were planted in different cells (1 cell/fruit). This should have resulted in 1700 cells (34 plants/array x 5 cells/plant x 2 population pairs/ploidy pair x 5 replicates). However, low seed set in some fruits resulted in approximately 1500 cells planted. Germination was scored every two days for six weeks. Seeds were considered germinated once cotyledons were open and the seed coat had been shed. Germination was averaged across all fruits for a plant from a replicate. Seedlings were thinned to one per cell, then transplanted individually into conetainers to allow plants to grow to sufficient size for cytotyping.

Once seedlings had developed at least 10 rosette leaves, 20 mg of tissue was collected for cytotyping via flow cytometry. Plant nuclei were extracted and stained with propidium iodide using a modified Otto 2-step protocol, and analyzed on a BD FACSCalibur flow cytometer. For 2X-4X populations, *Raphanus* was used as either an internal or external standard, and *Glycine* was used for 4X-6X populations (Dolezel, 2005). A total of 1020 F1 offspring were cytotyped; 624 from diploid-tetraploid arrays and 396 from tetraploid-hexaploid arrays. F1 offspring were binned to the nearest euploid DNA content of known *C. rotundifolia* cytotypes (i.e. a multiple of the haploid DNA content).

Statistical analysis

Pollinator transitions were scored as either intraploid (moving between plants of the same cytotype) or interploid (moving between plants of different cytotype), and were used as a measure of pollinator preference. Therefore, foraging bouts that visited only one plant were removed from the dataset. Differences in pollinator transition were assessed with respect to ploidy pair (2X-4X vs 4X-6X), focal cytotype (whether transitions are being tracked relative to the higher or lower cytotype within a ploidy pair; assigned after data collection since observers were naïve), and cytotype frequency (8%, 25%, 75%, 92%) using a Generalized Linear Mixed Model (PROC GLIMMIX, SAS 9.3 SAS Institute, INC. 2011) with a binomial link. Ploidy pair, focal cytotype, and cytotype frequency were main effects, with population (nested within ploidy pair), replicate, pollinator type (five families or genera within Hymenoptera or Diptera), average male and female flower number, plot location, and weather (predominantly sunny, cloudy, rainy, or windy) as random effects. The last three (weather, plot location, flower number) did not have significant effects and therefore were removed from subsequent models. Pollinator preference was also compared to null expectations under random pollinator movement using G-tests. G-tests were performed separately on each ploidy pair and cytotype frequency combination.

Differences in relative seed set, relative germination, and F1 cytotype between ploidy pairs, relative cytotype (the lower or higher ploidy cytotype within an array), and cytotype frequency were tested using a generalized linear mixed model. Relative seed set was calculated by dividing average seed set per plant for each replicate array by the average seed set of intraploid control crosses from the same maternal population. Unlike

relative seed set, which is heavily influenced by maternal ovule number (Haig and Westoby, 1991) and therefore only standardized by the maternal population, relative germination was calculated by dividing average germination per plant for each replicate array by the average germination of seeds derived from intraploid crosses of both populations in that array. Relative seed set, relative germination, and F1 cytotype were analyzed with a Gaussian link and identity distribution. F1 cytotype was only analyzed in 4X-6X arrays because insufficient variation existed in 2X-4X array cytotypes. Pairwise comparisons of relative seed set across ploidy pair and cytotype frequency combinations were made using a Tukey-Kramer test.

Results

Pollinators

A total of 195 foraging bouts were recorded, with 8.1 ± 1.2 SE bouts per observation period. Twenty-one foraging bouts consisted of a visit to only one plant (10.76% of all bouts). Visitors included five families or genera in the Orders Hymenoptera and Diptera: Andrenidae (mining bees), Halictidae (sweat bees), *Bombus* (bumble bees), Megachilidae (mason bees), and Syrphidae (Syrphid flies). *Bombus* accounted for approximately 20% of all bouts, but made more transitions per bout than other pollinators, comprising 39.2% of all transitions, with Halictids and Syrphids comprising 22.1% and 21.8%, respectively.

The frequency of intraploid transitions by pollinators did not depend on ploidy pair or focal cytotype, and no difference was observed among pollinator classes ($F_{1,349}$ = 0.71, p = 0.399). However, perhaps not unexpectedly, an increase in intraploid pollinator transitions was strongly associated with an increase cytotype frequency (Table 1, Figure

1), suggesting that pollinators move randomly within arrays relative to cytotype. G-tests largely confirm that pollinator transitions do not differ from a null expectation of random movement (Table 2). However, there were slightly more transitions to the other tetraploid when the 4X cytotype was 8% of the array than would be expected by chance in 2X-4X arrays (Table 2, Figure 1).

F1 production and cytotypes

Seed set was significantly higher, almost double, in 4X-6X arrays than in 2X-4X arrays, and increased with increasing cytotype frequency though the pattern of increase depended on the ploidy composition of the arrays (Table 1; Figure 2). In 2X-4X arrays, seed set was 20% or less that of control crosses when either cytotype comprised less than 50% of an array, and approximately 40% when a cytotype was in the majority (Figure 2a). Both diploids and tetraploids followed this pattern, though diploids had slightly higher seed set overall. In 4X-6X arrays, seed set in both minority frequencies were greater than the same frequencies in 2X-4X arrays but again substantially less than control crosses (Figure 2b). Seed set increased once a cytotype comprised the majority an array and was significantly higher when a cytotype was 75% of an array relative to 25%, and higher still when a cytotype comprised 92% of an array (Figure 2b). While on average germination was lower in seeds from mixed-ploidy arrays than from intraploid controls, germination rates were consistent across ploidy pairs, maternal cytotype, and cytotype frequencies (Table 1; Figure S2).

Three ploidy levels were observed in the offspring of each ploidy pair: 2X, 3X, and 4X offspring in 2X-4X arrays, and 4X, 5X, and 6X offspring in 4X-6X arrays. The proportion of intraploid hybrids (inferred by offspring cytotype) differed strongly

between ploidy pairs; triploids were rare in 2X-4X arrays, but pentaploids were common in 4X-6X arrays (Figure 3). In 2X-4X arrays, interploid hybrids were only seen when the maternal cytotype was either 8% or 25% of the array; 3% of offspring from 2X mothers and 8.5% from 4X mothers were hybrids (Figure 3a). Of all offspring that did not match their maternal cytotype, most were triploid, but three offspring of diploid mothers were tetraploid. Conversely, pentaploid interploid hybrids were common in 4X-6X arrays, ranging from 17% to 72% of all offspring (Figure 3b). Proportion of pentaploids varied inversely with cytotype frequency (Table 3); when maternal cytotype was uncommon, most offspring were interploid hybrids. On average, hexaploids produced 25% more interploid hybrids than tetraploids (Table 3, Figure 3b). There were no observations of hexaploids from tetraploid mothers, or vice versa.

Discussion

We used experimental arrays in which both prezygotic isolation (via pollinator preference) and postzygotic isolation (via pollen preference and poor hybrid seed quality) had the potential to occur. We found that cytotypes within the *C. rotundifolia* polyploid complex differ dramatically in their capacity to produce between-ploidy hybrids, and that the probability of interploid reproduction is mediated by cytotype frequency. Pollinator preference did not contribute to interploid reproductive isolation in this system. Rather, post-pollination barriers differed in strength between cytotypes, and were more permissive of reproduction between tetraploids and hexaploids than between diploids and tetraploids. These results suggest that interploid reproduction within polyploid complexes may be best conceptualized as two separate patterns—one between diploids and

tetraploids, and one among polyploid cytotypes—and support the hypothesis that ongoing gene flow among polyploids may limit divergence of polyploid taxa in angiosperms.

Prezygotic isolation due to cytotypic preference by pollinators was not observed in either diploid-tetraploid or tetraploid-hexaploid pairs. The frequency of intraploid pollinator transitions, in which a pollinator moved from one plant to another of the same cytotype, was not significantly different from expectations of random movement in almost all cases, and the few observed differences were small in magnitude. Lack of pollinator preference is not uniform among autopolyploid taxa. Pollinators have been shown to have distinct preferences for a given ploidy when multiple cytotypes grow in sympatry, and the strength of this pollinator-driven prezygotic barrier is a major contributor to overall interploid reproductive isolation (Segraves and Thompson, 1999; Husband and Sabara, 2004). However, other autopolyploid complexes show a lack of pollinator preference (Jersáková et al., 2010; Wefferling et al., 2017), similar to what we observe in *C. rotundifolia*. These results suggest that the relative importance of pollinator preference in interploid reproductive isolation is controlled by system-specific mechanisms, such as changes in floral morphology, nectar output, or nectar composition. It is worth noting that pollinator trials were performed using insects that were naïve to the species. The nearest occurrence of *Campanula rotundifolia* to the experimental site is approximately 40 km away, making it unlikely that local pollinators have experience with the species. However, pollinators learn differences in floral types, e.g. nectar rewards, quickly, often within one exposure to novel flowers (Ings et al., 2009). Therefore, it is unlikely that the inexperience of the pollinators drove the lack of preference that we found.

Minority cytotypes set proportionately fewer seed, regardless of ploidy. Differences in seed set with respect to cytotype frequency may be due to pollen competition or self-fertilization. Although few studies have examined effects of interploid competition on seed set, interspecific pollen competition has been shown to lower seed set in multiple systems (Galen and Gregory, 1989; Tiffin et al., 2001), so the presence of abundant interploid pollen may have suppressed seed set when a cytotype was infrequent in an array. Across all frequencies, seed set was over 80% higher in 4X-6X arrays. Unlike previous studies that investigated interploid reproduction (Greiner and Oberprieler, 2012; Chapter 3), plants in this study had access to both interploid and intraploid pollen. Previous interploid crossing studies in *C. rotundifolia* (Chapter 3) found similar seed set in both 2X-4X and 4X-6X crosses. The higher seed set observed in 4X-6X arrays, particularly at low frequencies, may be due to a greater capacity for selfing in higher ploidy levels (Chapter 2).

Unlike seed set, germination did not differ with ploidy or cytotype frequency. This contrasts with previous interploid crosses performed in *C. rotundifolia*, in which seed set showed little difference between ploidy pairs but germination was markedly higher in 4X-6X crosses (Chapter 3). The change in the apparent stage of reproductive barrier (seed set vs. germination) may be due to differences in available pollen. In the greenhouse experiment (Chapter 3), maternal plants were only exposed to pollen from plants that differed in cytotype, while in the current mixed-ploidy arrays, pollen that matched maternal cytotype was also available. Interploid crosses in *Arabidopsis* have shown aberrant endosperm development, resulting in early seed abortion (Köhler et al., 2010) or poor germination of phenotypically normal or near-normal seeds (Scott et al.,

1998; Stoute et al., 2012). Because plants in the current study received pollen that both matched and differed from their cytotype, development of intraploid progeny may have preempted provisioning and development of interploid progeny in the same ovary. Although this effect has not been studied in interploid system, sibling competition and differential provisioning has been observed in other taxa (Bañuelos and Obeso, 2003). If variable resource provisioning occurred following mixed-ploidy pollination in *C. rotundifolia*, fewer seeds may have developed in crosses with strong parental genomic imbalance (i.e. 2X-4X crosses), but those that did were likely the result of intraploid pollination and therefore less likely to suffer germination deficits.

Although interploid pollen competition and seed development impart a net reproductive barrier between both pairs of cytotypes studied (2X-4X and 4X-6X), the higher seed set in 4X-6X arrays suggests that post-pollination reproductive barriers are higher between diploids and tetraploids. However, differences in seed set alone do not account for the strikingly different patterns of interploid hybridization observed among offspring in 2X-4X and 4X-6X arrays. In 2X-4X arrays, less than 5% of all cytotyped offspring were triploid interploid hybrids, while between one-third and half of all offspring in 4X-6X arrays were pentaploid interploid hybrids. These results indicate that 4X-6X mixed-ploidy contact zones have a capacity for interploid reproduction whereas 2X-4X contact zones do not, and are consistent with previous work in *C. rotundifolia* (Chapter 3) and in *Senecio carniolicus*, in which pentaploids are common but triploids rare (Hülber et al., 2015).

The differential interploid reproduction found here has implications for the evolution of *C. rotundifolia*, particularly in contact zones, and for polyploid lineages

more broadly. Backcrosses between pentaploids and parental 4X and 6X cytotypes confirm that *C. rotundifolia* pentaploid hybrids retain approximately 50% fertility relative to parental cytotypes (Chapter 4), and other systems have demonstrated low to moderate fertility in pentaploids (Hülber et al., 2015). This capacity for reproduction, coupled with high rates of pentaploid formation in 4X-6X arrays, suggests that pentaploid hybrids may serve as conduits for gene flow in mixed-ploidy contact zones. By contrast, the paucity of triploid offspring in 2X-4X arrays coupled with the low seed set observed in minority cytotypes shows that diploids and tetraploids are effectively reproductively isolated. Over time, 2X-4X contact zones may be predicted to establish secondary reinforcement to prevent wasted mating opportunities, although lack of pollinator preferences does not demonstrate presence of such reinforcement yet. By contrast, 4X-6X contact zones may establish hybrid swarms, and serve to homogenize genetic structure of nearby tetraploids and hexaploids, constraining divergence of these cytotypes.

Cytotypic frequency, not just cytotype composition, is critical to potential interploid gene flow in mixed-ploidy contact zones. While tetraploids and hexaploids were markedly more capable of hybridizing than diploids and tetraploids, frequency of hybridization was higher when cytotypes were less common regardless of ploidy combination. This suggests that if contact zones are largely spatially segregated, such that each cytotype is locally dominant, gene flow may be limited. Due to the lower seed set and low frequency of hybridization, diploids and tetraploids individuals may establish secondary reinforcement to prevent unfit interploid mating, while tetraploids and hexaploids may permit local hybridization, limited to areas of immediate contact. Conversely, if a cytotype is outnumbered within its local environment, due to either

spatial heterogeneity of the contact zone or to introduction of a small number of founders, the ability to hybridize at low frequencies may create three outcomes. If diploids are introduced into an otherwise tetraploid population, any hybrid progeny would be triploid or tetraploid, and diploids may quickly disappear from the population. However, if tetraploids are introduced into an otherwise diploid population, interploid hybrids would tend to supplement the tetraploid population, allowing establishment alongside diploids. Presence of even a small number of tetraploids or hexaploids within an otherwise different-ploidy population will allow gene flow via pentaploid hybrids. As such, evidence of introgression, even from only transient introductions of another cytotype, would be most likely between tetraploids and hexaploids, and least likely upon introduction of diploids.

Because plants that differ in ploidy level have historically been considered reproductively isolated, the dynamics of mixed-ploidy contact zones within polyploid complexes are usually couched only in terms of competition. Cytotypes could experience coexistence or extirpation, but introgression and gene flow have rarely been considered. However, it is necessary to reevaluate our assumptions regarding polyploid evolution and the importance of contact zones between cytotypes and their spatial structure. For example, new cytotypes within otherwise homogeneous populations may actually facilitate hybridization, so minority cytotypes may not be dependent upon selfing in order to persist.

Higher-order polyploids retain greater capacity for gene flow than diploids and tetraploids, and this pattern is broadly consistent across *C. rotundifolia* and other polyploid systems (Greiner and Oberprieler, 2012; Hülber et al., 2015; Chapter 3). These

dual patterns of reproductive isolation—diploids maintaining stronger reproductive barriers than polyploids—suggests that we should be looking for two different signatures of diversification in polyploid lineages. Diploids may lose gene flow with tetraploids and therefore have higher diversification rates, while polyploid taxa retain homogenizing gene flow, and therefore diversify at lower rates throughout angiosperms.

Acknowledgements

We would like to thank E. Nagy and the Mountain Lake Biological Station for field space, logistical and technical support; B. Quarles for help with plant propagation and field assistance; D. Carr for analytical consultation; and the Galloway lab group for comments on an earlier draft of this manuscript. We would also like to thank the National Science Foundation for funding (NSF DEB-1457686)

References

Abbott, R. J., and A. J. Lowe. 2004. Origins, establishment and evolution of new polyploid species: *Senecio cambrensis* and *S. eboracensis* in the British Isles. *Biological Journal of the Linnean Society* 82: 467–474.

Baack, E. J. and M. L. Stanton. 2005. Ecological factors influencing tetraploid speciation in snow buttercups (*Ranunculus adoneus*): niche differentiation and tetraploid establishment. *Evolution* 59: 1936-1944.

Bañuelos, M. J. and J. R. Obeso 2003. Maternal provisioning, sibling rivalry and seed mass variability in the dioecious shrub *Rhamnus alpinus*. *Evolutionary Biology* 17: 19-31.

Bingham, R. A. and A. R. Orthner. 1998. Efficient Pollination in Alpine Plants. *Nature* 391: 238-239.

Bingham, R. A., and T. A. Ranker. 2000. Genetic diversity in alpine and foothill populations of *Campanula rotundifolia* (*Campanula*ceae). *International Journal of Plant Sciences* 161: 403–411.

Borges, L. A., L. G. Rodrigues Souza, M. Guerra, I. C. Machado, G. P. Lewis, A. V. Lopes. 2012. Reproductive isolation between diploid and tetraploid cytotypes of *Libidibia ferrea* (= *Caesalpinia ferrea*) (Leguminosae): ecological and taxonomic implications. *Plant Systematics and Evolution* 298: 1371-1381.

Castro, S., Z. Münzbergová, J. Raabová, and J. Loureiro. 2010. Breeding barriers at a diploid–hexaploid contact zone in *Aster amellus*. *Evolutionary Ecology* 25: 795–814.

Chapman, M. A., D. G. Forbes, R. J. Abbott. 2005. Pollen competition among two species of *Senecio* (Asteraceae) that form a hybrid zone on Mt. Etna, Sicily. *American Journal of Botany* 92: 730-735.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer. New York, NY.

Dolezel, J. 2005. Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. *Annals of Botany* 95: 99–110.

Eriksson, J. S., J. L. Blanco-Pastor, F. Sousa, Y. J. K. Bertrand, and B. E. Pfeil. 2017. A cryptic species produced by autopolyploidy and subsequent introgression involving *Medicago prostrata* (Fabaceae). *Molecular Phylogenetics and Evolution* 107: 367–381.

Fowler, N. L., and D. A. Levin. 1984. Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *The American Naturalist* 124: 703–711.

Galen, C., and T. Gregory. 1989. Interspecific pollen transfer as a mechanism of

competition - consequences of foreign pollen contamination for seed set in the alpine wildflower, *Polemonium-Viscosum*. *Oecologia* 81: 120–123.

Green, A. F., T. S. Ramsey, and J. Ramsey. 2011. Phylogeny and biogeography of ivies (*Hedera* spp., Araliaceae), a polyploid complex of woody vines. *Systematic Botany* 36: 1114–1127.

Greiner, R., and C. Oberprieler. 2012. The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra- and hexaploid relatives. *Flora - Morphology, Distribution, Functional Ecology of Plants* 207: 629–635.

Guggisberg, A., G. Mansion, and E. Conti. 2009. Disentangling reticulate evolution in an arctic-alpine polyploid complex. *Systematic Biology* 58: 55–73.

Haig, D., and M. Westoby. 1991. Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philosophical Transactions: Biological Sciences*. 333: 1-13.

Herben, T., P. Trávníček, and J. Chrtek. 2016. Reduced and unreduced gametes combine almost freely in a multiploidy system. *Journal of PPEES Sources* 18: 15–22.

Hülber, K., M. Sonnleitner, J. Suda, J. Krejčíková, P. Schönswetter, G. M. Schneeweiss, and M. Winkler. 2015. Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio carniolicus* (syn. *Jacobaea carniolica*, Asteraceae). *Ecology and Evolution* 5: 1224–1234.

Husband, B. C. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings of the Royal Society B: Biological Sciences* 267: 217–223.

Husband, B. C., and D. W. Schemske. 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. *Journal of Ecology* 88: 689–701.

Husband, B. C., and H. A. Sabara. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* 161: 703–713.

Ings, T. C., N. E. Raine, and L. Chittka. 2009. A population comparison of the strength and persistence of innate colour preference and learning speed in the bumblebee *Bombus terrestris*. *Behavioral Ecology and Sociobiology* 63: 1207–1218.

Jersáková, J., S. Castro, N. Sonk, K. Milchreit, I. Schödelbauerová, T. Tolasch, and S. Dötterl. 2010. Absence of pollinator-mediated premating barriers in mixed-ploidy populations of *Gymnadenia conopsea* s.l. (Orchidaceae). *Evolutionary Ecology* 24:

1199–1218.

Kennedy, B. F., H. A. Sabara, D. Haydon, and B. C. Husband. 2006. Pollinator-mediated assortative mating in mixed ploidy populations of *Chamerion angustifolium* (Onagraceae). *Oecologia* 150: 398–408.

Kolář, F., M. Štech, P. Trávníček, J. Rauchová, T. Urfus, P. Vít, M. Kubešová, and J. Suda. 2009. Towards resolving the *Knautia arvensis* agg. (Dipsacaceae) puzzle: primary and secondary contact zones and ploidy segregation at landscape and microgeographic scales. *Annals of Botany* 103: 963-974.

Köhler, C., O. Mittelsten Scheid, and A. Erilova. 2010. The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* 26: 142–148.

Lo, E. Y. Y., S. Stefanović, and T. A. Dickinson. 2010. Reconstructing reticulation history in a phylogenetic framework and the potential of allopatric speciation driven by polyploidy in an agamic complex in *Crataegus* (Rosaceae). *Evolution* 64: 3593–3608.

Lu, J., C. Zhang, and D. C. Baulcombe. 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proceedings of the National Academy of Sciences* 109: 5529–5534.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S.P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257–1257.

McAllister, H. A. 1972. The experimental taxonomy of *Campanula rotundifolia* L. Ph.D. Thesis, University of Glasgow, Glasgow, UK.

Nosil, P. 2008. Speciation with gene flow could be common. *Molecular Ecology* 17: 2103–2106.

Oswald, B. P., and S. L. Nuismer. 2011. A unified model of autopolyploid establishment and evolution. *The American Naturalist* 178: 687–700.

Parisod, C., R. Holderegger, and C. Brochmann. 2010. Evolutionary consequences of autopolyploidy. *New Phytologist* 186: 5–17.

Petit, R. J., and L. Excoffier. 2009. Gene flow and species delimitation. *Trends in Ecology and Evolution* 24: 386-393.

Ramsey, J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences* 108: 7096–7101.

Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. A. F. Noor, B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary*

Biology 30: 1450-1477.

Roccaforte, K., S. E. Russo, and D. Pilson. 2015. Hybridization and reproductive isolation between diploid *Erythronium mesochoreum* and its tetraploid congener *E. albidum* (Liliaceae). *Evolution* 69: 1375–1389.

Sabara, H. A., P. Kron, and B. C. Husband. 2013. Cytotype coexistence leads to triploid hybrid production in a diploid-tetraploid contact zone of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 100: 962–970.

Scott R. J., Spielman M., Bailey J., and Dickinson H. G. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development 125: 3329–3341.

Segraves, K. A., and J. N. Thompson. 1999. Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossulariifolia*. *Evolution* 53: 1114-1127.

Shetler, S. G. 1982. Variation and evolution of the nearctic harebells (*Campanula* sect. Heterophylla). J. Cramer. Vaduz.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.

Sonnleitner, M., K. Hülber, R. Flatscher, P. Escobar García, M. Winkler, J. Suda, P. Schönswetter, and G.M. Schneeweiss. 2015. Ecological differentiation of diploid and polyploid cytotypes of *Senecio carniolicus* sensu lato (Asteraceae) is stronger in areas of sympatry. *Annals of Botany* 117: 269-276.

Ståhlberg, D. 2007. Habitat differentiation, hybridization and gene flow patterns in mixed populations of diploid and autotetraploid *Dactylorhiza maculata* s.l. (Orchidaceae). *Evolutionary Ecology* 23: 295–328.

Stebbins, G. L., Jr. 1942. Polyploid complexes in relation to ecology and the history of floras. *The American Naturalist*.

Stevens, C. J., J. Wilson, and H. A. McAllister. 2012. Biological Flora of the British Isles: *Campanula rotundifolia*. *Journal of Ecology* 100: 821–839.

Stoute, A. I., V. Varenko, G. J. King, R. J. Scott, and S. Kurup. 2012. Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *The Plant Journal* 71: 503-516.

Thompson, J. N., and K. F. Merg. 2008. Evolution of polyploidy and the diversification of plant-pollinator interactions. *Ecology* 89: 2197–2206.

Thompson, J. N., S. L. Nuismer, and K. Merg. 2004. Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biological Journal of the Linnean*

Society 82: 511–519.

Tiffin, P., S. Olson, and L. C. Moyle. 2001. Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B: Biological Sciences* 268: 861–867.

Wefferling, K. M., S. Castro, J. Loureiro, M. Castro, D. Tavares, and S. B. Hoot. 2017. Cytogeography of the subalpine marsh marigold polyploid complex (*Caltha leptosepala* s.l., Ranunculaceae). *American Journal of Botany* 104: 271–285.

Wolf, P. G., D. R. Campbell, N. M. Waser, S. D. Sipes, T. R. Toler, and J. K. Archibald. 2001. Tests of pre- and postpollination barriers to hybridization between sympatric species of *Ipomopsis* (Polemoniaceae). *American Journal of Botany* 88: 213-219.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* 106: 13875–13879.

Table 1) Analysis of variance for pollinator transitions, seed set, and germination in mixed-ploidy *Campanula rotundifolia* arrays. Ploidy pair refers to the two cytotypes within a given array (2X-4X or 4X-6X), cytotype denotes either the higher or lower cytotype (2X vs. 4X, 4X vs. 6X) within an array (treated as focal cytotype for pollinator visitation and maternal cytotype for seed set and germination), frequency is the frequency of each cytotype in an array (8%, 25%, 75%, of 92%). Population, replicate, and pollinator type were included as random effects. Table lists F-values; p-values: + < 0.1, * < 0.05, **<0.01, ***<0.001.

Source	Num DF	Pollination	Seed Set	Germination
Ploidy Pair	1	0.14	12.45***	3.23+
Frequency	3	4.58**	2.76+	1.87
Cytotype	1	0.3	0.05	2.76+
Pair x Freq.	1	0.71	5.51*	1.63
Pair x Cyto.	3	0.71	0.05	0.49
Freq. x Cyto.	3	0.34	1.29	1.34
Three-way	1	0.53	2.04	1.03
Error DF		349	332	340

Table 2) G-tests evaluating the proportion of pollinator transitions relative to expectations under random movement. Because tests included only two possible classes, DF for all G-tests is 1.

Cytotype Freq	Minority Ploidy	Majority Ploidy	G-value	P-value
8% - 92%	Diploid	Tetraploid	0.385	0.535
8% - 92%	Tetraploid	Diploid	3.973	0.046
25% - 75%	Diploid	Tetraploid	0.399	0.528
25% - 75%	Tetraploid	Diploid	2.696	0.101
8% - 92%	Tetraploid	Hexaploid	0.820	0.365
8% - 92%	Hexaploid	Tetraploid	0.213	0.645
25% - 75%	Tetraploid	Hexaploid	1.085	0.298
25% - 75%	Hexaploid	Tetraploid	0.426	0.514

Table 3) Analysis of variance of offspring cytotypes in tetraploid-hexaploid arrays. Cytotype frequency is the frequency of each cytotype in an array (8%, 25%, 75%, of 92%), relative cytotype indicates the cytotype (4X vs. 6X). Table lists F-values, p-values: ** <0.01, ***<0.001.

Source	Num DF	F1 cytotype
Cytotype Frequency	1	11.31**
Relative Cytotype	1	13.89***
Cyto. Freq. x Rel. Cyto.	3	0.65
Error DF	390	

Figure 1) Proportion intraploid transitions by pollinators within (a) 2X-4X and (b) 4X-6X arrays of *C. rotundifolia* where cytotypes were present in four frequencies. Transitions were considered intraploid when a pollinator moved from one plant to another of the same cytotype. Error bars denote standard error. Dashed lines denote expected proportions based on random pollinator visitation.

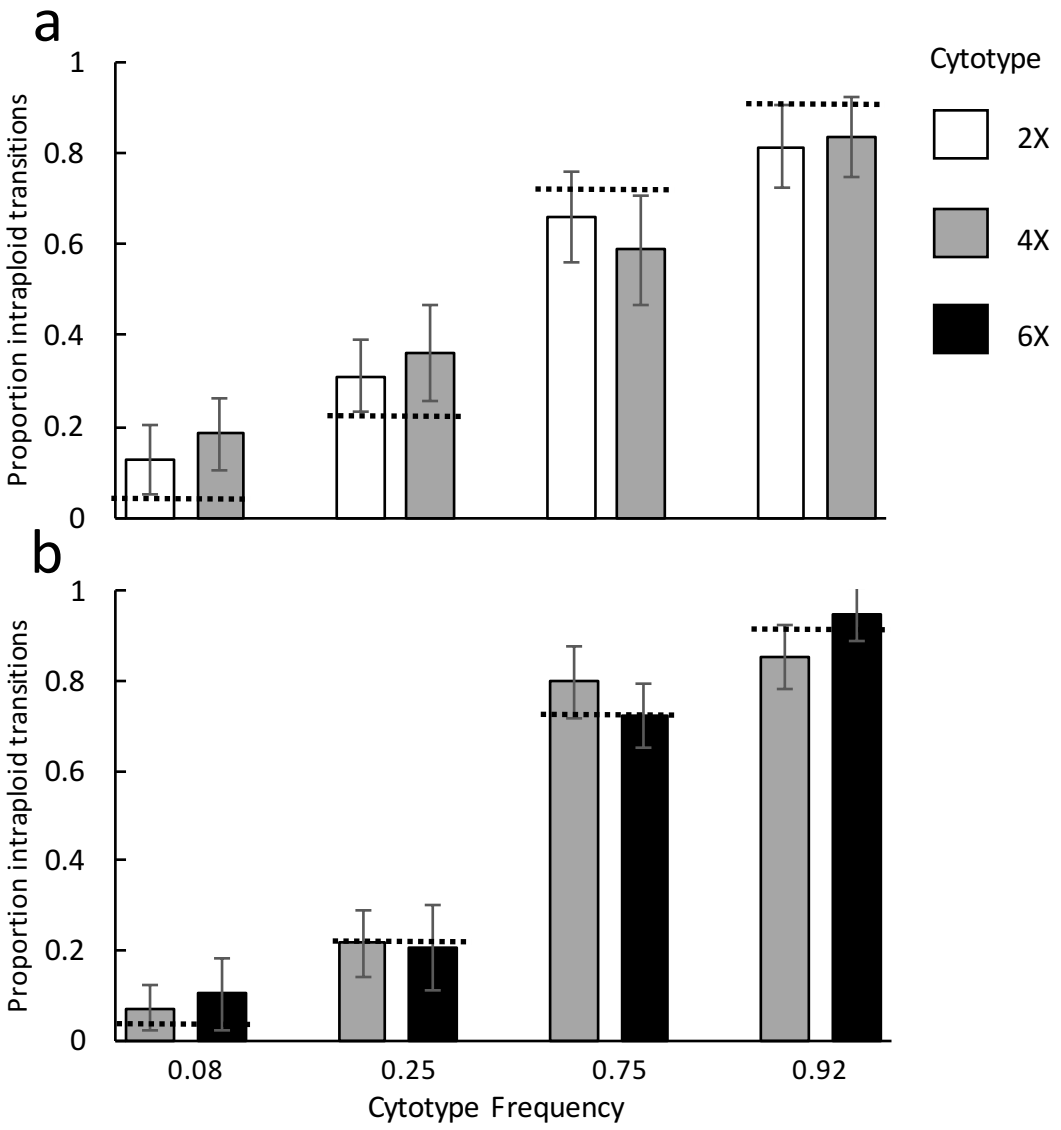


Figure 2) Relative seed set as a proportion of control seed set in *C. rotundifolia* plants exposed to different frequencies of interploid and intraploid pollen within mixed-ploidy arrays. Seeds from plants in (a) 2X-4X arrays and (b) 4X-6X arrays. Error bars denote standard error. Bars with different letters are significantly different in a Tukey multiple comparison of ploidy pair and cytotype frequency combination.

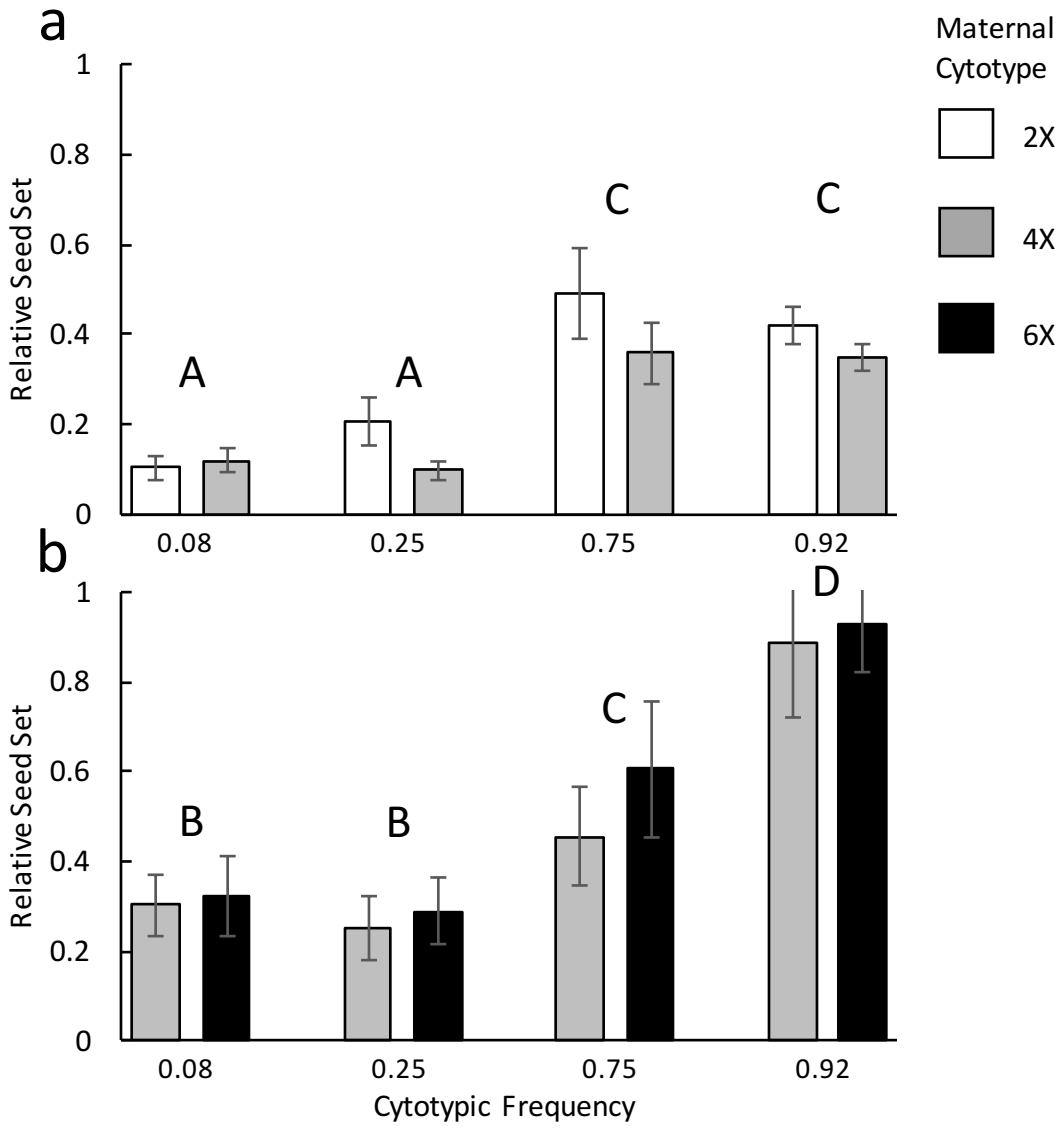


Figure 3) Proportion hybrid *C. rotundifolia* offspring by maternal cytotype and frequency, as estimated by formation of 3X offspring in 2X-4X arrays (a) or 5X offspring in 4X-6X arrays (b). Maternal cytotype denotes the cytotype of the parent from which seeds were produced. Error bars denote standard error for values > 0.

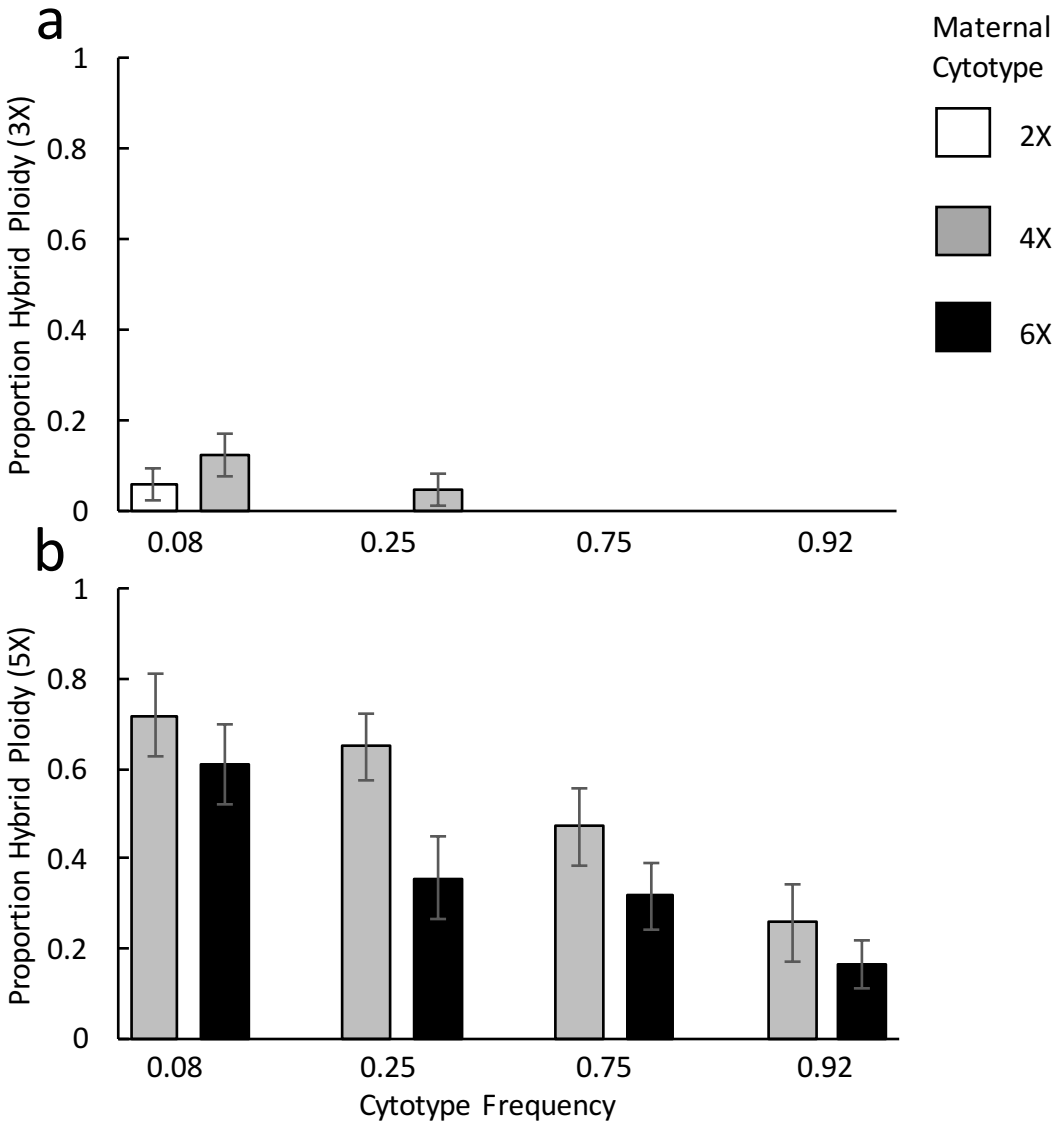
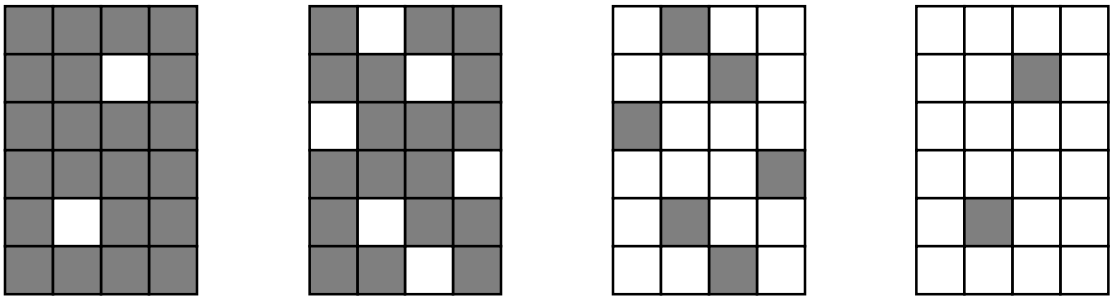


Table S1) *C. rotundifolia* population pairs used in array construction. Each pair delineated by a horizontal line.

Population ID	Ploidy	Locality	Country	Latitude	Longitude
25	2	Zamky	Czech Republic	50.429255	14.075625
24	4	Vinaricka Hora	Czech Republic	50.184791	14.081019
23	2	Dresden	Germany	50.973440	13.821118
33	4	Mittelndorf	Germany	50.937567	14.203496
115	4	Misery Bay PP, ON	Canada	45.800465	-82.749153
116	6	South Baymouth, ON	Canada	45.558822	-82.018024
30	4	Turner's Falls, MA	United States	42.595089	-72.581203
10	6	Red Bay, ON	Canada	44.416252	-81.458360

Figure S1) Schematic diagram of array construction. Diploid-tetraploid arrays are shown with grey squares indicating diploid plants and white squares indicating tetraploid plants. Tetraploid-hexaploid arrays followed same construction scheme.



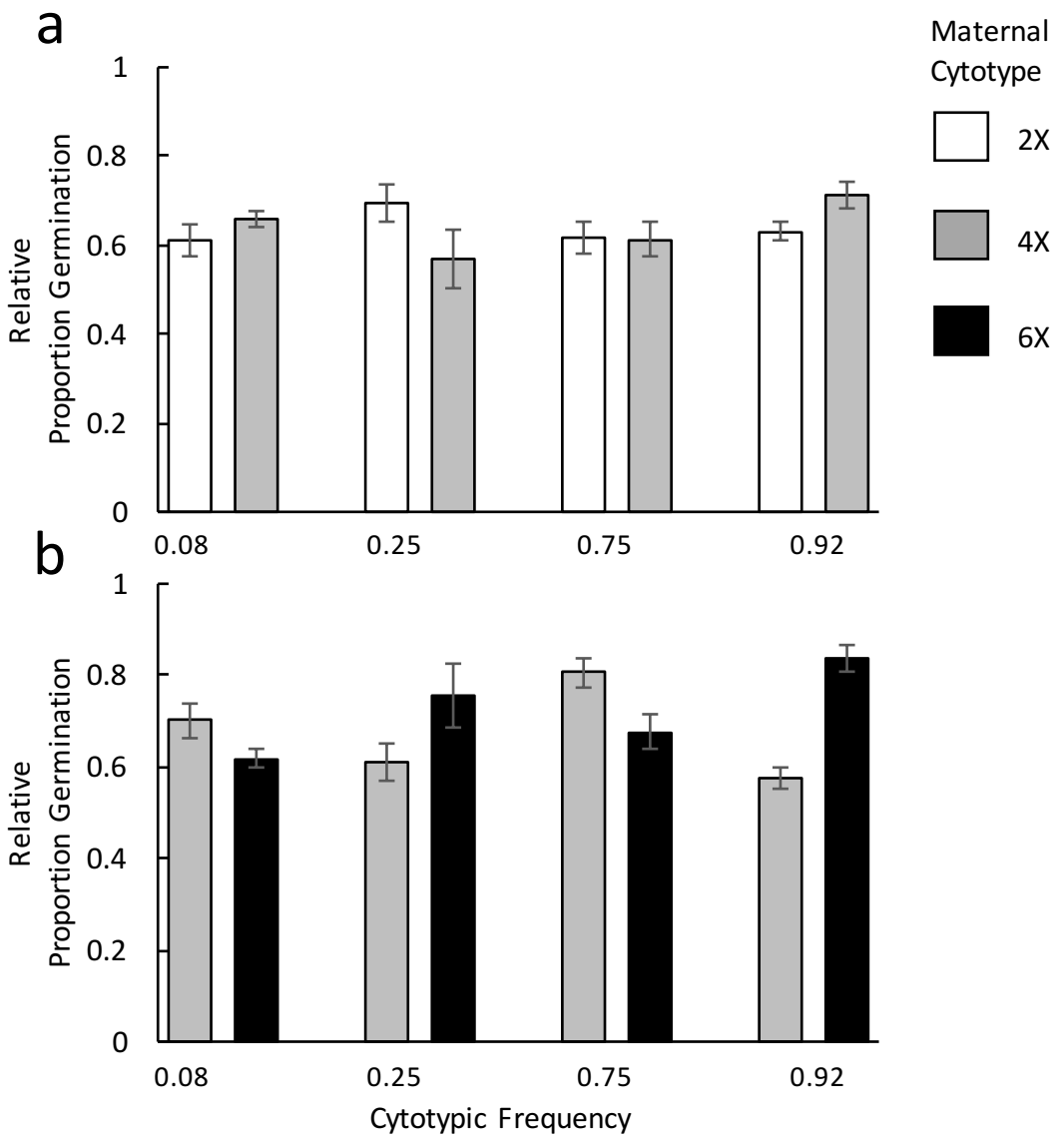
92% 2X
8% 4X

75% 2X
25% 4X

25% 2X
75% 4X

8% 2X
92% 4X

Figure S2) Germination proportion in *C. rotundifolia* seeds from (a) 2X-4X arrays and (b) 4X-6X arrays by maternal cytotype and frequency. Germination proportions relative to intraploid controls. Error bars denote standard error.



CHAPTER SIX:

Does ploidy change confer reproductive isolation? Mixed evidence from contact zones within a polyploid complex

Abstract

Premise of the study – Polyploidy is often assumed to confer immediate reproductive isolation and therefore play a significant role in plant speciation. However, recent evidence including the current research questions the importance of polyploidy in plant diversification, particularly among higher-order polyploids. To assess the role of whole genome duplication in plant speciation, it is important to evaluate the extent to which ploidy differences inhibit gene flow.

Methods – We sampled four mixed-ploidy contact zones (two 2X-4X and two 4X-6X) from the polyploid complex *Campanula rotundifolia*. Using flow cytometry to estimate ploidy and microsatellites to assess potential gene exchange, we compared rates of interploid reproduction and gene flow in contact zones.

Key results – 2X-4X and 4X-6X contact zones differ markedly in both interploid reproduction and gene flow. While no intermediate cytotypes were discovered in 2X-4X contact zones, pentaploids were common in 4X-6X zones, as were aneuploids suggestive of pentaploid backcrossing. Likewise, 2X and 4X individuals were genetically distinct, but 4X, 5X, and 6X individuals were largely genetically homogeneous.

Conclusions – These results demonstrate that diploids and tetraploids are reproductively isolated in natural contact zones but tetraploids and hexaploids are not, supporting experimental evidence in multiple systems. These findings suggest that diploids and polyploids can quickly diverge, however, differentiation of higher cytotypes in polyploid complexes may remain constrained by gene flow. Diversification in polyploid complexes may therefore be lower than would otherwise be assumed based on current accepted patterns of rapid interploid reproductive isolation.

Introduction

Reproductive isolation is one of the most common drivers of divergence and speciation. Without reproductive isolation, the homogenizing effects of even low levels of gene flow may prevent divergence (Slatkin, 1985). Whole genome duplication (WGD, or polyploidy), because of its ability to prevent reproduction between diploids and related polyploids (Coyne and Orr, 2004; Husband and Sabara, 2004), has long been considered an important speciation mechanism in angiosperms (Grant, 1981; Rieseberg and Willis, 2007). The role of WGD in speciation is further bolstered by phylogenetic evidence showing that many of the most speciose plant families are also the most heavily polyploid (Otto and Whitton, 2000; Soltis et al., 2003; 2012).

Recently, this view of WGD as a primary driver of angiosperm speciation has been questioned. Genomic analyses have suggested slower diversification rates among polyploids than diploids (Slotte et al., 2008; Mayrose et al., 2011), contrary to previous assumptions (Otto and Whitton, 2000; Soltis et al., 2003; 2012), and evidence of historical gene flow has been found in some diploid-tetraploid systems (Slotte et al., 2008; Jørgensen et al., 2011; Zohren et al., 2016). However, little is known about two components that are fundamental to understanding the role of WGD in angiosperm speciation. First, the frequency and pattern of contemporary interploid gene flow is not known in most systems. Second, interploid gene flow among different polyploid cytotypes, and how it differs from gene flow between diploids and tetraploids, remains largely unknown. Assessment of contemporary interploid gene flow in polyploid complexes, including between polyploids that differ in cytotype, is critical to

understanding the extent to which WGD limits gene flow and thereby promotes speciation in angiosperms.

Interploid reproductive isolation in polyploid complexes can manifest through numerous prezygotic and postzygotic barriers. Polyploids generally germinate later than diploids (Maceira et al., 1993), may flower earlier or later (Pires et al., 2004; Ramsey, 2011), may experience different pollinator regimes (Segraves and Thompson, 1999), and may be more likely to self, thereby reducing mating opportunities with diploids (Barringer, 2007). If diploids and polyploids do successfully produce hybrid seed, aberrant seed development and meiosis lead to the viability and fertility deficits that characterize the "triploid block" (Marks, 1966) inherent to many interploid hybrids. In most systems, a combination of prezygotic and postzygotic barriers act in concert (Husband and Sabara, 2004), effectively eliminating the possibility of reproduction and gene flow between diploids and tetraploids.

Despite this strong evidence for reproductive isolation between diploids and tetraploids, recent work demonstrates that gene flow has occurred in some polyploid complexes (Slotte et al., 2008; Jørgensen et al., 2011; Zohren et al., 2016). However, most studies that detect interploid gene flow have found evidence of historical introgression, but have not assessed contemporary gene flow. To truly understand the role WGD plays in plant speciation, in depth analysis of ongoing interploid gene flow in polyploid complexes is needed.

Patterns of interploid gene flow among higher-order polyploids is a largely overlooked problem. The few studies that have explicitly investigated higher cytotypes, suggest that interploid reproduction and gene flow may be markedly higher between

polyploids than between diploids and tetraploids (Greiner and Oberprieler, 2012; Hülber et al., 2015). The effect of these higher cytotypes on polyploid speciation is an especially salient problem given that polyploid complexes comprising three or more cytotypes are relatively common across angiosperms (Meyers and Levin, 2006). To fully understand the effect of WGD on speciation, it is necessary to characterize not only interploid reproduction and gene flow between diploids and tetraploids, but also between the higher cytotypes often found in these polyploid complexes.

Autopolyploid complexes are strong experimental systems to understand potential interploid reproduction and gene flow as they minimize the confounding effects of different parental genomes. The *Campanula rotundifolia* autopolyploid complex comprises three dominant cytotypes with multiple, though infrequent, contact zones between cytotypes, presenting a unique opportunity to compare interploid gene flow within one polyploid system. We assess the presence of gene flow in mixed ploidy populations, and compare the magnitude of gene flow between two pairs of ploidy levels: diploids and tetraploids, and tetraploids and hexaploids. We ask the following questions: 1) Does interploid reproduction occur in *C. rotundifolia* mixed-ploidy contact zones, as determined by presence of individuals of intermediate ploidy? 2) Are populations genetically structured by cytotype as expected if polyploidy is an isolating barrier? and 3) Is interploid reproduction and gene flow more common between tetraploids and hexaploids than between diploids and tetraploids?

Materials and methods

Study System

Campanula rotundifolia is a short-lived, generalist-pollinated, perennial wildflower. It has a broadly circumboreal distribution, located in the northern latitudes of North American and throughout much of Europe (Shetler, 1982; Stevens et al., 2012), and exists as a morphologically variable polyploid complex (Gadella, 1964; Kovanda, 1966; Mansion et al., 2012).

There are three main cytotypes, diploid (2n = 34), tetraploid (2n = 68), and hexaploid (2n = 102), and the cytotypic and phenotypic complexity of the group has led to a number of named species, ecotypes, and varieties (Gadella, 1964; Kovanda, 1977). We treat these herein as *C. rotundifolia* sensu lato. The three cytotypes are non-randomly dispersed throughout the distribution. Tetraploids are the dominant cytotype throughout most of the range, while diploids are restricted mainly to central and extreme northern Europe (Löve and Löve, 1943; Stevens et al., 2012), and hexaploids are restricted to the British Isles, and central and northwestern North America (Stevens et al., 2012; Chapter 1). Most populations consist of only one cytotype. Although cytotypes do not usually exist in parapatry, some contact zones between ploidy levels are known to exist (Shepherd, 2007; Chapter 5; K. Šemberová, pers. comm.). Known contact zones consist of either diploids and tetraploids, or tetraploids and hexaploids.

Sampling locations

Four mixed-ploidy contact zones were sampled: two diploid-tetraploid, and two tetraploid-hexaploid. For each sampled plant in each contact zone, approximately five to ten cauline leaves were collected and placed in silica gel within 12 hours of collection, and the GPS location noted.

The two diploid-tetraploid populations were located in central Europe, one near the village of Mittelndorf near Sächsische-Schweiz National Park in eastern Germany, and one near central and southern Prague in the Czech Republic (Table 1). In these relatively small and spatially circumscribed populations, the location where plants were found was comprehensively surveyed, and leaf tissue was collected from all plants that and were at least 1 m apart. The Mittelndorf population consisted of two disjunct groups approximately 0.5 km apart (Figure 1). The western portion of the population was primarily located in disturbed grassy meadow along a roadway, and the eastern portion a recently mown hayfield. It is likely the additional plants exist between these disjunct groups, but recent mowing made any intervening plants impossible to find. The Prague contact zone comprised two locations (Figure 2): a small (approximately 30 m x 50 m) hillside meadow near Velká skála (Figure 2a), and approx. 8 km to the south a recently mown meadow in Řeporyje, approximately 100 m x 50 m, bounded on three sides by a housing development, and to the north by a wooded area (Figure 2b).

The tetraploid-hexaploid populations were considerably larger and more spatially dispersed. Cheddar Gorge in England is approximately 2 km long, with a maximum depth of 137 m from the bottom to the southern rim (Table 1, Figure 3). *Campanula rotundifolia* is common throughout this site on exposed limestone cliff faces as well as the meadows along the rim, but the vertical cliff faces on the south side prevented comprehensive collection. Collection efforts consisted of four east-west transects: one each along the northern and southern rims, and one each along the northern and southern cliff faces at the base of the gorge to a height of 2 meters. On each transect, leaf tissue was collected from all accessible plants that presented more than one flowering stem that

were at least 1 m apart. The second 4X-6X population, Misery Bay Provincial Nature Reserve Canada, encompasses approximately 10 km^2 on the southern coast of Manitoulin Island in Lake Huron (Table 1, Figure 4). Patches of alvar—exposed limestone bedrock that forms a nutrient poor, cracked, pavement-like substrate (Lundholm and Stark, 2007)—are interspersed within the otherwise heavily forested reserve. These alvar glades provide suitable habitat for *C. rotundifolia*, with plants occupying the cracks within the bedrock. Plants were collected along one transect from the edge of reserve property to the lakeshore; most samples were found in two large glades spanning a total distance of approximate 1 km.

Flow cytometry & Cytometric analysis

Flow cytometry was used to estimate ploidy of all collected plants in each contact zone. Field-collected silica-dried tissue was used for analysis. Samples were first weighed and approximately 5 mg of tissue was reserved for DNA extraction. All remaining tissue was processed for flow cytometric analysis using a modified Otto 2-step protocol (Otto, 1990; Chapter 1). In order to prevent loss of tissue during cell disruption, dried tissue was immersed in 1 mL of Otto I buffer (0.1 M citric acid, 0.5% w/v Tween 20) for two minutes and then finely chopped for approximately three minutes to maximize release of intact nuclei. Samples were then incubated at room temperature for 1 hr before addition of Otto II buffer (0.1 M sodium phosphate) to which 50 ng/uL propidium iodide and 50 ng/uL RNase I had been added. Samples were incubated for 10-15 minutes before analysis on a BD FACSCalibur Cell Analyzer. Samples were analyzed using a 488-nm laser and run at slow speed to minimize doublets. Once samples had been gated to exclude cell debris, at least 2000 events were recorded or the sample was run to

exhaustion if tissue quantities were limited. Relative fluorescence at maximum peak height was compared to external standards with *Raphanus sativus* used for 2X-4X contact zones, and *Glycine max* for 4X-6X contact zones.

DNA content was estimated by comparing the relative fluorescence of unknown samples to that of external standards. A fluorescence ratio was calculated for each sample by dividing the sample's fluorescence by the fluorescence of the external standard. The external standards were included with every analytical session, and were re-run in the event of any changes in calibration mid-run. To assign an estimated cytotype to each individual, estimated DNA content for each plant was compared to a known diploid *C. rotundifolia* from a population in Dresden, Germany that was not part of any sampled contact zone. DNA content ratios within 25% of a whole-number multiple were considered euploid. All values outside that range were assigned as aneuploid individuals.

DNA extraction, microsatellite amplification & analysis

Gene flow was assessed using microsatellites. DNA was extracted from all samples using a modified CTAB protocol that had been optimized for plate processing (Costa and Roberts, 2014). Dried tissue was first pulverized using stainless steel shot in a bead beater, then CTAB buffer was added to each sample, and from there CTAB extraction proceeded normally. Eight microsatellite markers were chosen for amplification and analysis that were designed specifically for *C. rotundifolia* (Plue et al., 2015; Table S1). Microsatellite loci were amplified as duplexes using 5'-fluorescently labeled forward primers. Amplified loci were visualized at the Yale Genome Sequencing Center, and were scored using GeneMarker 3.4 software. Due to peak stuttering, and

because all loci were created as trimeric or tetrameric repeats, all bins were two nucleotides wide.

To investigate overall genetic similarity of cytotypes within each contact zone, private alleles were first counted and standardized for population size and cytotype using ADZE (Szpiech et al., 2008). This software (and all comparable packages) could only analyze populations that consisted of a single-cytotype. Therefore, to analyze mixed-ploidy populations, the maximum cytotype of each contact zone (either 4X or 6X) was applied to all samples and extra allele calls treated as missing data. Initial investigation of genetic distance was performed using POLYSAT 1.7 (Clark and Jasieniuk, 2011). Pairwise genetic distances were first calculated for all individuals within a contact zone using the Bruvo distance function (which accounts for allelic mutations), then a Principal Coordinates Analysis (PCA) was performed on these pair-wise distances. AMOVAs were then performed on the microsatellite data using Arlequin 3.5 (Excoffier and Lischer, 2010) to quantify degree of population differentiation. To account for the polyploid microsatellite data, in which some individuals had up to six distinct alleles per locus, each allele was treated as a dominant marker, and scored for presence or absence (Garcia-Verdugo et al., 2013). In 2X-4X contact zones, variance was partitioned between ploidy levels. In 4X-6X contact zones, variance was partitioned between ploidy levels, but all aneuploids were grouped with pentaploids. Due to spatial discontinuity in both 4X-6X contact zones, variance was also partitioned into relevant spatial groupings: Misery Bay was partitioned into northern and southern glades, and Cheddar Gorge was partitioned to separate the southern rim from the other three transects – northern rim, and both inner transects.

Results

Cytotypic results

The two contact zone types showed different patterns of fluorescence intensity (Figure 1). Diploid-tetraploid contact zones comprised individuals that clustered discretely around a fluorescence intensity indicative of either diploidy or tetraploidy (Figure 1a, b). No intermediate values suggesting triploidy or aneuploidy were observed. By contrast, two or three intermediate local maxima of fluorescence intensity were observed in Cheddar Gorge (Figure 1d) and Misery Bay (Figure 1c), approximately corresponding to expected pentaploids, as well as putative aneuploids between 4X and 5X, and between 5X and 6X. In both tetraploid-hexaploid contact zones, observed fluorescence intensities between putative 5X and 6X individuals formed a near-continuous distribution.

The diploid-tetraploid contact zones comprised 38 plants in Mittelndorf and 30 plants in Prague (Table 1). These contact zones contained only diploid and tetraploid individuals. Mittelndorf contained 23 diploid individuals, primarily located along a path in a mown field to the northeast of the population, 15 tetraploid individuals locate peripherally to the south and west (Figure 2). The Prague contact zone comprised two distinct subpopulations located approximately 8 km apart, each comprising only one observed cytotype. The southern subpopulation contained 14 tetraploids, while 16 diploids were restricted to the north (Figure 3).

The two tetraploid-hexaploid contact zones, by contrast, contained numerous individuals of intermediate cytotype. 41.0% of all individuals in Cheddar Gorge were either pentaploid or aneuploid (50 of 122 total; Figure 4), primarily aneuploid between

5X and 6X. Likewise, 48% of all individuals in Misery Bay were also pentaploid or aneuploid (24 of 50 total; Figure 5). Cytotypes were not randomly distributed in either contact zone. In Cheddar Gorge, ploidy levels at the base of the Gorge were largely uniformly distributed, with no clustering of tetraploids or hexaploids. However, individuals on the southern rim were almost exclusively tetraploid. In Misery Bay, tetraploids were primarily found in the northeastern glade while hexaploids were found to the southwest. Pentaploids and aneuploids were common in both glades.

Microsatellite results

3-18 alleles were recovered for each of the eight microsatellite loci amplified for each population (Table 2). More alleles were recovered in 4X-6X contact zones than 2X-4X contact zones, but this may have been a factor of larger sample sizes in Cheddar Gorge and Misery Bay. However, while more alleles were recovered in the 4X-6X contact zones, the 2X-4X contact zones had considerably more alleles that were private to a cytotype. In the Mittelndorf contact zone, six private alleles were present in the diploids and three in the tetraploids. In the Czech contact zone, four diploid and nine tetraploid alleles were private. By contrast, in the Cheddar Gorge contact zone, only one private allele was found in each of the three cytotypic classes (4X, 5X + aneuploids, and 6X). In Misery Bay, one private allele was found in the tetraploids, but none in the 5X, aneuploid, or 6X cytotypic classes.

PCA analyses showed differences in cytotypic clustering patterns between 2X-4X and 4X-6X contact zones (Figure 6). For both 2X-4X contact zones (Figure 6a,b), all or almost all diploids and tetraploids formed separate clusters, with considerably more variance explained by the first principal component than in 4X-6X contact zones (Table

3). By contrast, little clustering by cytotype was observed in either 4X-6X contact zone, indicating that cytotypes are not genetically distinct. Little variance in clustering was explained by the first two principal components in either 4X-6X contact zone (Table 3; Figure 6c,d).

AMOVA partitioned genetic variance differently in diploid-tetraploid and tetraploid-hexaploid contact zones. In 2X-4X contact zones, genetic variance was higher between groups (i.e. cytotypes) than within them (Table 4). In 4X-6X contact zones, however, no significant genetic variation was observed at any scale (Table 4).

Discussion

The pattern of cytotypic and genotypic variation observed in *Campanula rotundifolia*'s mixed-ploidy contact zones suggests that interploid reproduction is common when tetraploids and hexaploids are sympatric, but rarely occurs between diploids and tetraploids. Cytotypic evidence shows that tetraploids and hexaploids can not only interbreed, but that pentaploid hybrids also undergo interploid reproduction, particularly with hexaploids. Genotypic evidence finds that this reproduction through pentaploid intermediates results in homogenizing gene flow in tetraploid-hexaploid contact zones. These results demonstrate that patterns observed in controlled crossing studies (Greiner and Oberprieler, 2012; Chapter 3, 4) are borne out in natural populations, and makes the case that variable patterns of reproductive isolation may lead to differing levels of divergence among cytotypes in polyploid complexes.

Cytotypic distributions were strikingly different between the two types of contact zones. In the diploid-tetraploid contact zones, Mittelndorf and Prague, no triploids—the expected intermediate cytotype—or aneuploids were observed. The distribution of

diploid and tetraploid DNA content suggests interploid reproduction rarely occurs and doesn't lead to ongoing interploid gene flow in 2X-4X contact zones. By contrast, both histograms of tetraploid-hexaploid contact zones showed that pentaploids were a substantial component, suggesting ongoing hybridization between tetraploids and hexaploids. Furthermore, both contact zones had evidence of pentaploid-hexaploid interbreeding, as shown by presence of numerous aneuploids between 5X and 6X. These aneuploids formed a near-continuous distribution, suggesting pentaploids are fertile and that reproduction with hexaploids is ongoing. The lack of triploids and presence of pentaploids in these natural contact zones confirms the patterns of interploid reproduction found in previous studies of *C. rotundifolia* (Chapter 3, 4, 5). Under both hand-crossed and naturally pollinated conditions, interploid crosses routinely formed more pentaploids than triploids (Chapters 3 and 5), and interploid backcrosses necessary for gene flow were only observed among pentaploids (Chapter 4). Similar patterns have also been observed another polyploid complex. Despite comprehensive sampling of a diploid-tetraploid contact zone, no triploids were obtained in *Senecio carniolicus*, yet pentaploids were common in a similar tetraploid-hexaploid contact zone. Interploid reproduction involving pentaploids was likewise biased toward hexaploids (Hülber et al., 2015; Sonnleitner et al., 2015).

The triploid block hypothesis, with its loss of triploid viability due to developmental defects (Marks, 1966), sets up clear expectations for why interploid reproduction should rarely occur. The mechanisms by which *C. rotundifolia* pentaploids avoid these odd-ploidy consequences are not known, but other systems may provide explanations. Endosperm development mediated by differential gene expression occurs

multiple angiosperm families (Radchuk, 2006; Xing and Zhang, 2010; Stoute et al., 2012; Lu et al., 2012). In 2X-4X crosses, imbalanced gene expression leads to aberrant endosperm development and low germination (Stoute et al., 2012; Lu et al., 2012). Parental genomic imbalance is expected to be lower in 4X-6X crosses. Other systems show seed morphologies (Costa et al., 2013; Sekine et al., 2013) and patterns of triploid and pentaploid germination that are consistent with less aberrant seed development among pentaploids (Hegarty et al., 2012; Laport et al., 2016).

Likewise, while meiosis in pentaploids is not well-understood, meiosis tends to be less disrupted than in related triploids. In *Senecio carniolicus*, backcrosses between pentaploids and hexaploids confirm pentaploid fertility (Hülber et al., 2015), while *Vaccinium* pentaploids are known to be fertile and to cross with both tetraploids and hexaploids (Laverty and Vorsa, 1991). Increased fertility among pentaploids could be due to unequal segregation during meiosis as in the *Rosa canina* group (Lim et al., 2005), or could be due to reduced dosage sensitivity in higher polyploids (Edger and Pires, 2009). In summary, pollen fertility in *C. rotundifolia* and other systems is consistent with less aberrant meiosis in higher odd-numbered polyploids than is found in analogous triploids.

Genetic evidence is congruent with cytotypic data, and supports differences in interploid reproduction between diploid-tetraploid and tetraploid-hexaploid contact zones. Three lines of analysis—cytotypic allelic richness, PCA clustering, and AMOVA—all support the same pattern of interploid gene flow occurring in 4X-6X contact zones, but not 2X-4X contact zones. Diploid-tetraploid contact zones have more alleles that are private to cytotype than tetraploid-hexaploid contact zones, even after standardizing for differences in population size (e.g., Szpiech et al., 2008). This higher

incidence suggests that gene flow is insufficient between diploids and tetraploids to homogenize allele frequencies across all individuals. PCA and AMOVA results likewise support homogenization between cytotypes within tetraploid-hexaploid contact zones, while diploid-tetraploid contact zones remain genetically distinct. The only evidence of population differentiation in tetraploid-hexaploid contact zones appears to be caused by spatial barriers. Tetraploids along the southern rim of Cheddar Gorge show some evidence of distinct genetic clustering, which may indicate that pollinators rarely forage both within and along the rim of the gorge.

The homogenization of sympatric tetraploids and hexaploids, and the lack thereof in sympatric diploids and tetraploids, is all the more striking given the evolutionary history of each contact zone. Chloroplast DNA evidence suggests that three of the four contact zones, Mittelndorf, Prague, and Cheddar Gorge, most likely arose through in situ genome duplication and as such have not experienced substantial diverge of chloroplast haplotypes (Figure S1). By contrast, tetraploids and hexaploids along the Great Lakes show significant divergence in both chloroplast (Figure S1) and nuclear genes (Chapter 1). However, this history of divergence has not prevented individuals in the Misery Bay contact zone from homogenizing upon secondary contact. Likewise, a more recent history of divergence and continued sympatry has not fostered homogenization between diploids and tetraploids in Mittelndorf and Prague.

The Mittelndorf diploid-tetraploid contact zone did demonstrate a possible avenue of interploid reproduction and gene flow that is not reliant upon triploid intermediates. One apparent neotetraploid was observed that was geographically disjunct and genetically distinct from other nearby tetraploids. This tetraploid individual was

genetically almost identical to its most proximate diploid neighbor, suggesting that it arose via unreduced gamete formation, possibly through fusion of two unreduced gametes, or through fertilization of an unreduced ovule with normal reduced pollen from an existing tetraploid. Unreduced gametes are relatively common in multiple angiosperm systems (Bringhurst and Gill, 1970; Ramsey, 2006; Kreiner et al., 2017), and have been implicated as the most likely mechanism of polyploid formation (Oswald and Nuismer, 2011) and less commonly as a possible mechanism of interploid gene flow (Herben et al., 2016). The stark genetic division between diploids and tetraploids in the Mittelndorf contact zone suggests that such neotetraploids do not significantly contribute to interploid gene flow, but their presence at even low frequencies is theoretically capable of somewhat constraining divergence.

Because diploid-tetraploid and tetraploid-hexaploid contact zones experience different rates of interploid reproduction and gene flow, different spatial distributions of cytotypes are likely to develop. In diploid-tetraploid contact zones, lack of gene flow and triploid-block reduce the persistence of locally rare cytotypes, creating areas of local homogeneity (Hardy and Vekemans, 2001; Münzbergová et al., 2013). This cytotypically structured pattern is evident in both diploid-tetraploid contact zones studied. Conversely, because interploid reproduction is common and hybrid fitness is high in tetraploid-hexaploid contact zones, establish and persistence of cytotypically heterogeneous groupings is expected, and has been recently observed in other systems (Sonnleitner et al., 2015; Laport et al., 2016). Cytotypic spatial structure in these contact zones reflects habitat-specific gene flow barriers and interploid gene flow biased toward hexaploids. In Misery Bay, the northern glade comprises primarily tetraploids, while the southern glade

comprises pentaploids, aneuploids, and hexaploids. The southern clade likely reflects occasional pentaploid formation due to tetraploid pollination, followed on continued interploid backcrossing with hexaploids. In Cheddar Gorge, tetraploids are again more spatially clustered than other cytotypes, due to both unidirectional gene flow and pollination barriers imposed by surrounding hardwood forests and limestone cliffs. In total, patterns of gene flow inform spatial segregation of cytotypes, with different expectations for diploid-tetraploid and tetraploid-hexaploid contact zones.

The contrasting patterns of gene flow observed in this study suggest that different cytotypes within polyploid complexes face different evolutionary outcomes. The lack of gene flow between diploids and tetraploids supports the presence of strong reproductive barriers, tantamount to "instant speciation," (Coyne and Orr, 2004). Due to the rapid onset of reproductive isolation, these diploids and tetraploids may show higher speciation rates than would occur in similar homoploid systems. Conversely, tetraploids and hexaploids show little to no differentiation due to interploid reproductive isolation. The relative rarity of pentaploids and aneuploids throughout the Great Lakes region suggests that these hybrids are ultimately less fit than their tetraploid and hexaploid relatives, a hypothesis supported by backcross experiments (Chapter 5), but it is nevertheless clear that these individuals serve as a conduit for interploid gene exchange. This gene exchange may constrain divergence of higher cytotypes, leading lower speciation rates relative to homoploid systems.

Acknowledgements

We would like to thank the Ontario Ministry of Natural Resources, the United Kingdom National Trust, and Longleat Estate for sampling permission, M. Koski, C. Debban, R. Watson, and K. Kubow for helpful suggestions. We would also like to thank for the UVA Center for Global Inquiry and Innovation, the Society for the Study of Evolution, Torrey Botanical Society, and the NSF DEB-1457686 for funding.

References

Barringer, B. C. 2007. Polyploidy and self-fertilization in flowering plants. *American Journal of Botany* 94: 1527–1533.

Bringhurst, R. S., and T. Gill. 1970. Origin of *Fragaria* polyploids. II. Unreduced and doubled-unreduced gametes. *American Journal of Botany* 57: 969-976.

Clark, L. V. and M. Jasieniuk. 2011. POLYSAT: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources* 113: 562-566.

Costa, C. M., and R. P. Roberts. 2014. Techniques for improving the quality and quantity of DNA extracted from herbarium specimens. *Phytoneuron* 48: 1-8.

Costa, J., V. Ferrero, J. Loureiro, M. Castro, L. Navarro, and S. Castro. 2013. Sexual reproduction of the pentaploid, short-styled *Oxalis pes-caprae* allows the production of viable offspring D. Byers [ed.],. *Plant Biology* 16: 208–214.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer. New York, NY.

Edger, P. P., and J. C. Pires. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699–717.

Excoffier, L. and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Gadella, T. W. J. 1964. Cytotaxonomic studies in the genus *Campanula*. *Wentia* 11: 1–104.

Garcia-Verdugo, C., J. A. Calleja, P. Vargas, L. Silva, O. Moreira, and F. Pulido. 2013. Polyploidy and microsatellite variation in the relict tree *Prunus lusitanica* L.: how effective are refugia in preserving genotypic diversity of clonal taxa? *Molecular Ecology* 22: 1546–1557.

Grant, V. 1981. Plant speciation. Columbia University Press. New York, NY.

Greiner, R., and C. Oberprieler. 2012. The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra- and hexaploid relatives. *Flora - Morphology, Distribution, Functional Ecology of Plants* 207: 629–635.

Hardy, O. J., and X. Vekemans. 2001. Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. *Evolution* 55: 943.

Hegarty, M. J., R. J. Abbott, and S. J. Hiscock. 2012. Allopolyploid Speciation in Action: The Origins and Evolution of *Senecio cambrensis*. *In* Polyploidy and Genome Evolution, 245–270. Springer Berlin Heidelberg, Berlin, Heidelberg.

Herben, T., P. Trávníček, and J. Chrtek. 2016. Reduced and unreduced gametes combine almost freely in a multiploidy system. *Journal of PPEES Sources* 18: 15–22.

Husband, B. C., and H. A. Sabara. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* 161: 703–713.

Hülber, K., M. Sonnleitner, J. Suda, J. Krejčíková, P. Schönswetter, G. M. Schneeweiss, and M. Winkler. 2015. Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio carniolicus* (syn. *Jacobaea carniolica*, Asteraceae). *Ecology and Evolution* 5: 1224–1234.

Jørgensen, M. H., D. Ehrich, R. Schmickl, M. A. Koch, and A. K. Brysting. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evolutionary Biology* 11: 346.

Kovanda, M. 1977. Polyploidy and variation in the *Campanula rotundifolia* complex. Part II. (Taxonomic). *Folia Geobotanica et Phytotaxonomica* 12: 23–89.

Kovanda, M. 1966. Some chromosome counts in the *Campanula rotundifolia* complex II. *Folia Geobotanica & Phytotaxonomica Bohemoslovaca*.

Kreiner, J. M., P. Kron, and B. C. Husband. 2017. Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. *New Phytologist* 1–11.

Laport, R. G., R. L. Minckley, and J. Ramsey. 2016. Ecological distributions, phenological isolation, and genetic structure in sympatric and parapatric populations of the *Larrea tridentata* polyploid complex. *American Journal of Botany* 103: 1358–1374.

Laverty, T., and N. Vorsa. 1991. Fertility of aneuploids between the 5x and 6x levels in blueberry - the potential for gene-transfer from 4x to 6x levels. *Journal of the American Society for Horticultural Science* 116: 330–335.

Lim, K. Y., G. Werlemark, R. Matyasek, J. B. Bringloe, V. Sieber, H. El Mokadem, J. Meynet, J. Hemming, A. R. Leitch, and A. V. Roberts. 2005. Evolutionary implications of permanent odd polyploidy in the stable sexual, pentaploid of *Rosa canina* L. *Heredity* 94: 501–506.

Löve, A., and D. Löve. 1943. The significance of differences in the distribution of diploids and polyploids. *Hereditas* 29: 145–163.

Lu, J., C. Zhang, and D. C. Baulcombe. 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proceedings of the National Academy of Sciences* 109: 5529–5534

Lundholm, J. T., and K. E. Stark. 2007. Alvar seed bank germination responses to variable soil moisture. *Canadian Journal of Botany* 85: 986–993.

Maceira, N. O., P. Jacquard, and R. Lumaret. 1993. Competition between diploid and derivative autotetraploid *Dactylis glomerata* L. from Galicia. Implications for the establishment of novel polyploid populations. *New Phytologist* 124: 321–328.

Mansion, G., G. Parolly, A. A. Crowl, E. Mavrodiev, N. Cellinese, M. Oganesian, K. Fraunhofer, G. Kamari, D. Phitos, R. Haberle, G. Akaydin, N. Ikinci, T. Raus, and T. Borsch. 2012. How to handle speciose clades? Mass taxon-sampling as a strategy towards illuminating the natural history of *Campanula* (Campanuloideae). *PLoS ONE* 7: e50076–23.

Marks, G. E. 1966. The origin and significance of intraspecific polyploidy: Experimental evidence from *Solanum chacoense*. *Evolution* 20: 552–557.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257–1257.

Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. *Evolution* 60: 1198–1206.

Münzbergová, Z., M. Surinová, and S. Castro. 2013. Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity* 110: 123–130.

Oswald, B. P., and S. L. Nuismer. 2011. A unified model of autopolyploid establishment and evolution. *The American Naturalist* 178: 687–700.

Otto, F. J. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewickz Z, Crissman HA, eds. Methods in cell biology, vol. 33. San Diego, CA, USA: Academic Press, 105–110.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual review of genetics* 34: 401–437.

Plue, J., K. Vandepitte, O. Honnay, and S. A. O. Cousins. 2015. Isolation by 454-sequencing and characterization of polymorphic microsatellite markers in the tetraploid perennial herb *Campanula rotundifolia*. *Conservation Genetics Resources* 7: 721-722.

Pires, J. C., J. W. Zhao, M. E. Schranz, E. J. Leon, P. A. Quijada, L. N. Lukens, and T. C. Osborn. 2004. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biological Journal of the Linnean Society* 82: 675–688.

Radchuk, V. 2006. Jekyll encodes a novel protein involved in the sexual reproduction of

barley. *The Plant Cell Online* 18: 1652–1666.

Ramsey, J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences* 108: 7096–7101.

Ramsey, J. 2006. Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). *Heredity* 98: 143–150.

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

Segraves, K. A., and J. N. Thompson. 1999. Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossulariifolia*. *Evolution* 53: 1114.

Sekine, D., T. Ohnishi, H. Furuumi, A. Ono, T. Yamada, N. Kurata, and T. Kinoshita. 2013. Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *The Plant Journal* 76: 792–799.

Shetler, S. G. 1982. Variation and evolution of the nearctic harebells (*Campanula* sect. Heterophylla). J. Cramer. Vaduz.

Shepherd, J. R. 2007. Polyploidy and the phylogeography of *Campanula rotundifolia* L. in the British Isles and Ireland. University of Edinburgh, Masters Thesis, 132pp.

Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics*. 16:393-430.

Slotte, T., H. Huang, M. Lascoux, and A. Ceplitis. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Molecular Biology and Evolution* 25: 1472–1481.

Soltis, D. E., P. S. Soltis, and J. A. Tate. 2003. Advances in the study of polyploidy since Plant Speciation. *New Phytologist* 161: 173–191.

Soltis, D. E., R. J. A. Buggs, W. B. Barbazuk, S. Chamala, M. Chester, J. P. Gallagher, P. S. Schnable, and P. S. Soltis. 2012. The early stages of polyploidy: rapid and repeated evolution in *Tragopogon*. *In* Polyploidy and Genome Evolution, 271–292. Springer Berlin Heidelberg, Berlin, Heidelberg.

Sonnleitner, M., K. Hülber, R. Flatscher, P. Escobar García, M. Winkler, J. Suda, P. Schönswetter, and G. M. Schneeweiss. 2015. Ecological differentiation of diploid and polyploid cytotypes of *Senecio carniolicus* sensu lato (Asteraceae) is stronger in areas of sympatry. *Annals of Botany* 117: 269-276.

Stevens, C. J., J. Wilson, and H. A. McAllister. 2012. Biological flora of the British Isles: *Campanula rotundifolia*. *Journal of Ecology* 100: 821–839.

Stoute, A. I., V. Varenko, G. J. King, R. J. Scott, and S. Kurup. 2012. Parental genome

imbalance in *Brassica oleracea* causes asymmetric triploid block. *The Plant Journal* 71: 503-516.

Szpiech, Z. A., M. Jakobsson, and N. A. Rosenberg. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24: 2498-2504.

Xing, Y., and Q. Zhang. 2010. Genetic and molecular bases of rice yield. *Annual Review of Plant Biology* 61: 421–442.

Zohren, J., N. Wang, I. Kardailsky, J. S. Borrell, A. Joecker, R. A. Nichols, and R. J. A. Buggs. 2016. Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Molecular Ecology* 25: 2413–2426.

Table 1) Summary of population sampling and cytotype distribution in each of *Campanula rotundifolia*'s four contact zones. Aneuploids (all individuals with estimated DNA content between 4X and 6X not otherwise assigned to a ploidy) were included within the 5X ploidy class.

Contact Zone	Ploidy	Latitude	Longitude	Individuals	2X	4X	5X	6X
Mittelndorf	2X-4X	50.934994	14.192485	38	23	15		
Prague	2X-4X	50.067941	14.343481	30	16	14		
Cheddar Gorge	4X-6X	51.285494	2.756566	122		46	50	26
Misery Bay	4X-6X	45.797273	-82.732944	50		12	24	14

Table 2) Summary statistics describing the allelic diversity of microsatellite loci amplified for *Campanula rotundifolia*'s four contact zones. Ploidy refers to the dominant cytotypes present in the contact zone. PA refers to the count of alleles across all loci that are private to one cytotype in the contact zone. PA_R is the rarefied count of private alleles to standardize by population size.

Contact Zone	Ploidy	Total Alleles	Alleles/Locus	PA	PA_R
Mittelndorf	2X-4X	38	3.1	9	4.52
Prague	2X-4X	41	3.8	13	4.77
Cheddar Gorge	4X-6X	69	8.0	3	3.42
Misery Bay	4X-6X	60	7.1	1	1.67

Table 3) Summary of variances for first two axes of Principal Coordinates Analysis.

	1st PC	2nd PC
Mittelndorf	34.16%	8.47%
Prague	38.37%	10.13%
Misery Bay	19.60%	13.07%
Cheddar Gorge	11.77%	8.48%

Table 4) Analysis of Molecular Variance between and within cytotypes in all studied contact zones. For all contact zones, groups were defined by cytotype, and populations within groups were constructed by partitioning individuals within cytotype.

Source of variation	DF	Fixation Index	P-value
Mittelndorf			
Among groups	1	1.30492	0.048
Among populations within groups	2	-0.81727	0.256
Within populations	35	-1.08083	1
Prague			
Among groups	1	1.52061	0.031
Among populations within groups	2	-0.60321	0.326
Within populations	28	-0.01483	1
Misery Bay			
Among groups	1	-1.48126	0.338
Among populations within groups	2	0.19997	1
Within populations	47	-0.98508	1
Cheddar			
Among groups	1	-3.01402	0.657
Among populations within groups	2	0.09986	1
Within populations	118	-2.61316	1

Figure 1) Cytotypic distribution of *Campanula rotundifolia* individuals in each contact zone. X-axis denotes the estimated genome size as compared to internal standards *R. sativus* and *G. max*. Colors denote assigned ploidy level based on genome size. Contact zones are as follows: a) Mittelndorf, b) Prague, c) Misery Bay, and d) Cheddar Gorge.

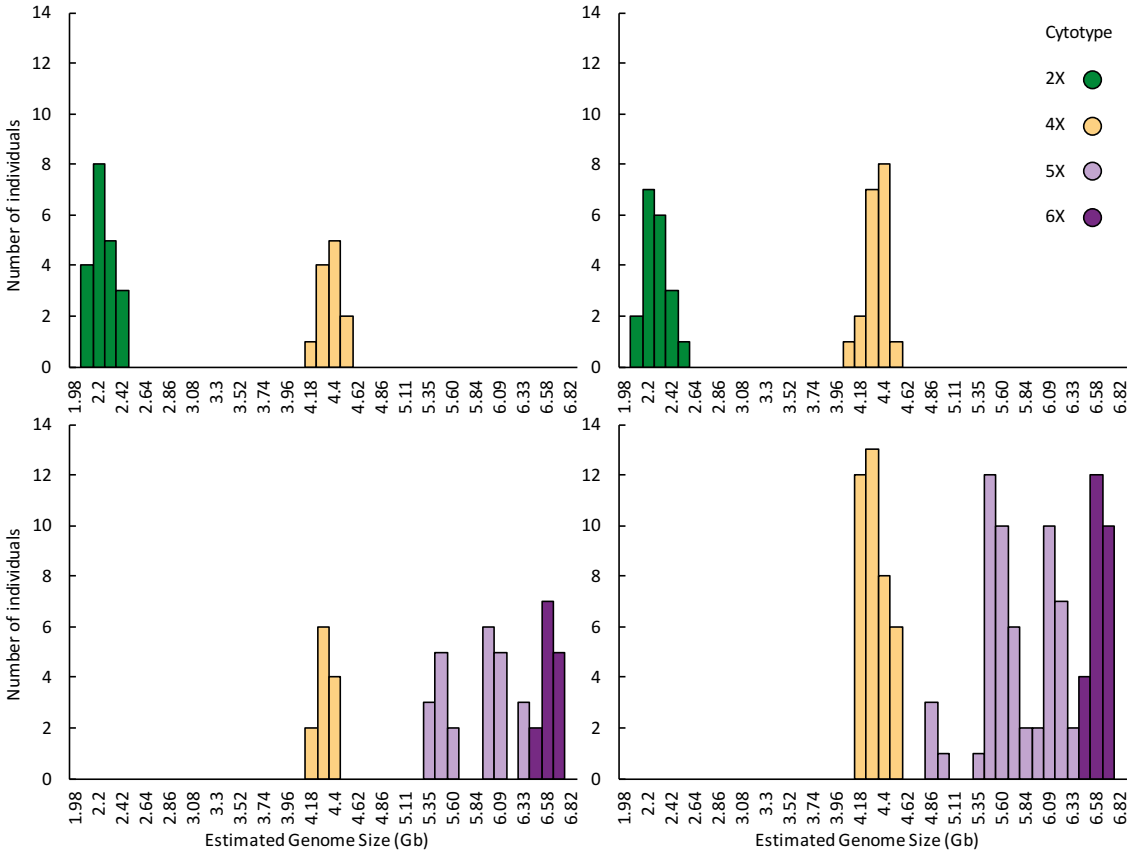


Figure 2) Coordinate plots and cytotypic assignment of all individuals in *Campanula rotundifolia*'s diploid-tetraploid contact zone near Mittelndorf, Saxony, Germany. Ploidy estimates based on flow cytometry.

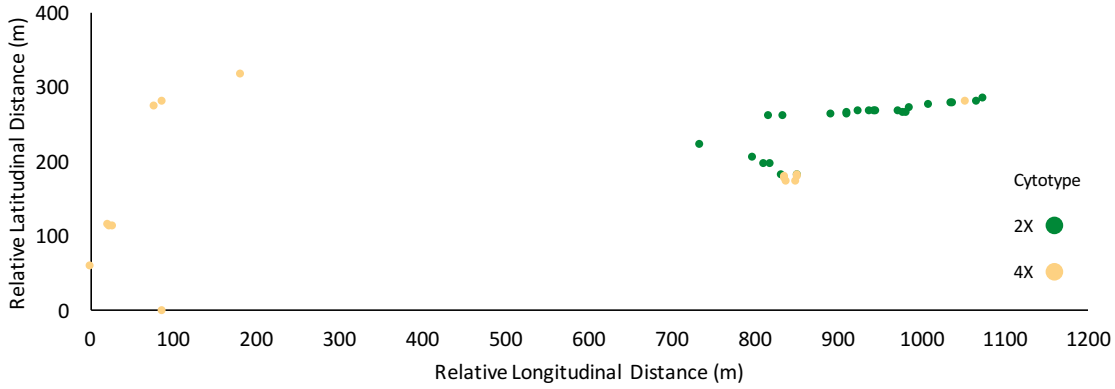


Figure 3) Coordinate plots and cytotypic assignment of all individuals in *Campanula rotundifolia*'s diploid-tetraploid contact zone in Prague in the Czech Republic. Populations shown separately due to distance between them (approx. 8 km). Diploids found in southern Prague near Řeporyje, tetraploids found in northern Prague near Velká skála. Ploidy estimates based on flow cytometry.

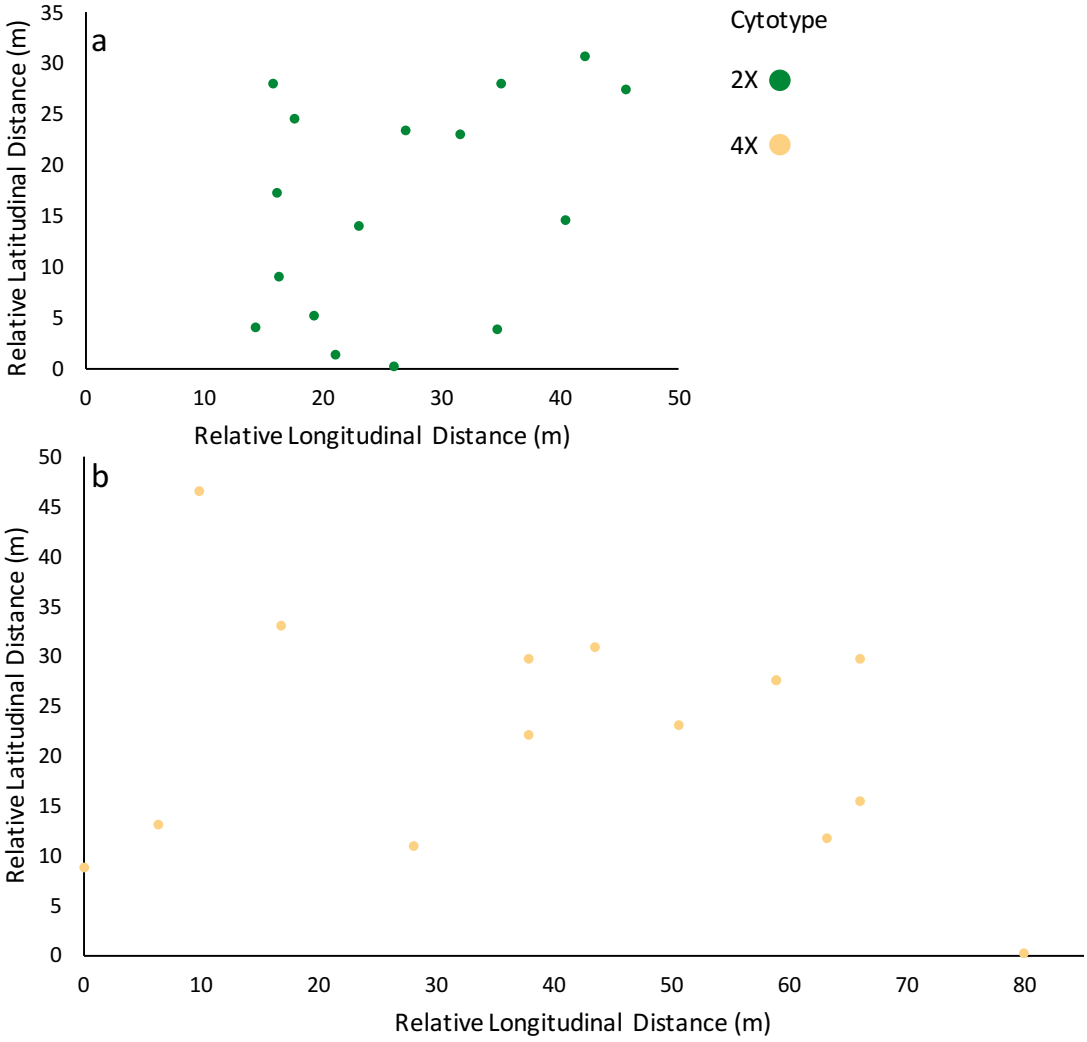


Figure 4) Coordinate plots and cytotypic assignment of all individuals in *Campanula rotundifolia*'s tetraploid-hexaploid contact zone in Cheddar Gorge, England, United Kingdom. Ploidy estimates based on flow cytometry.

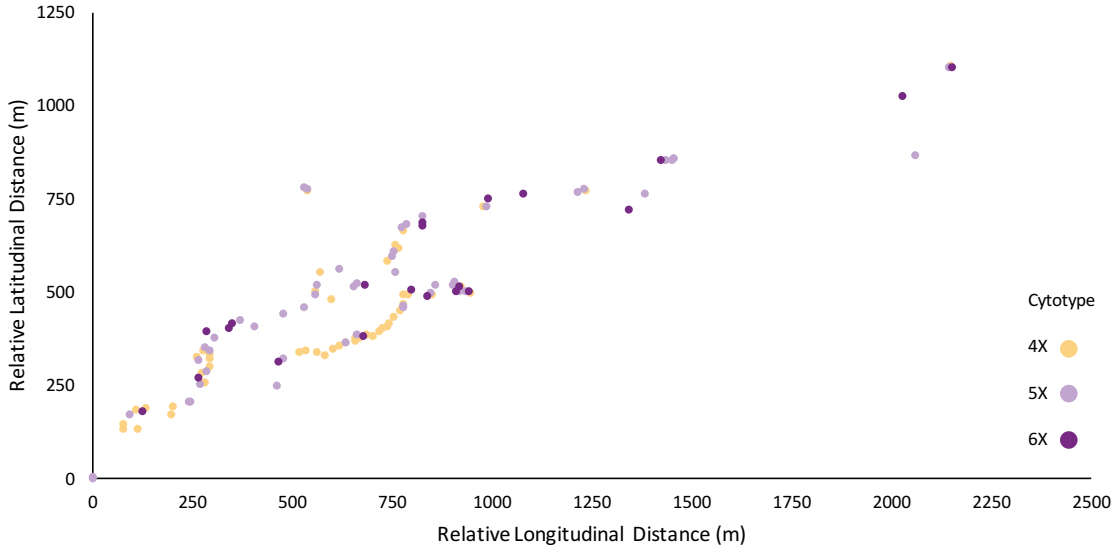


Figure 5) Coordinate plots and cytotypic assignment of all individuals in *Campanula rotundifolia*'s tetraploid-hexaploid contact zone in Misery Bay, Ontario, Canada. Ploidy estimates based on flow cytometry.

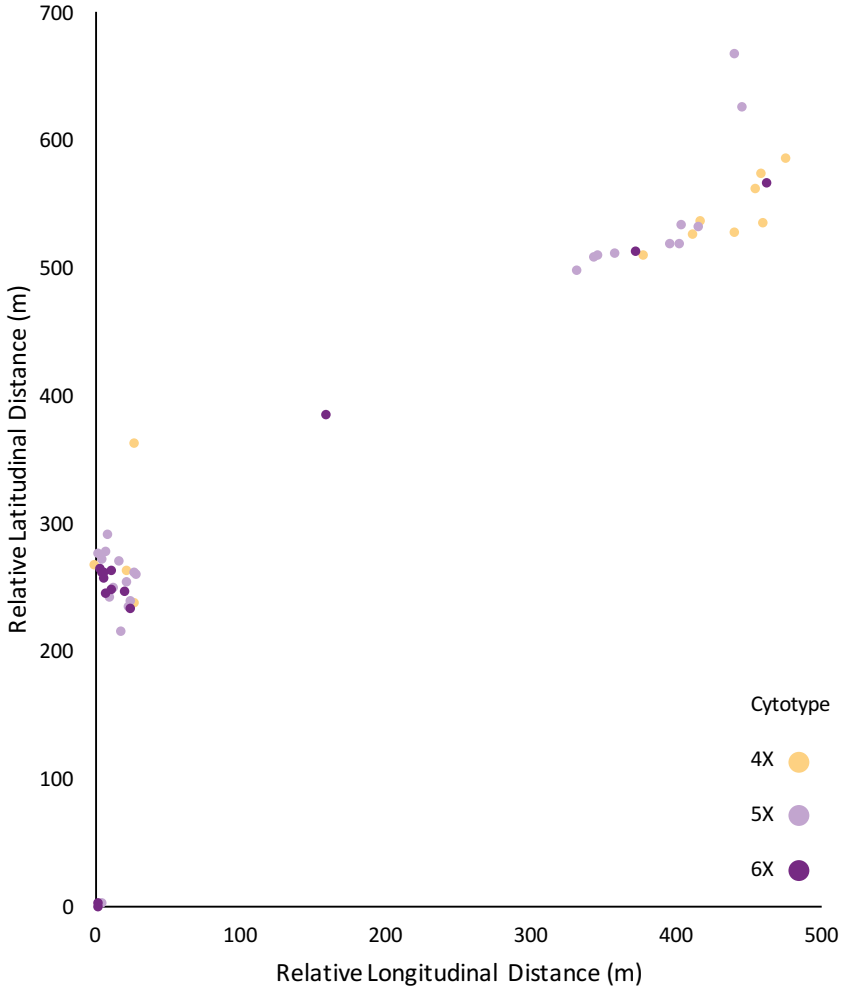
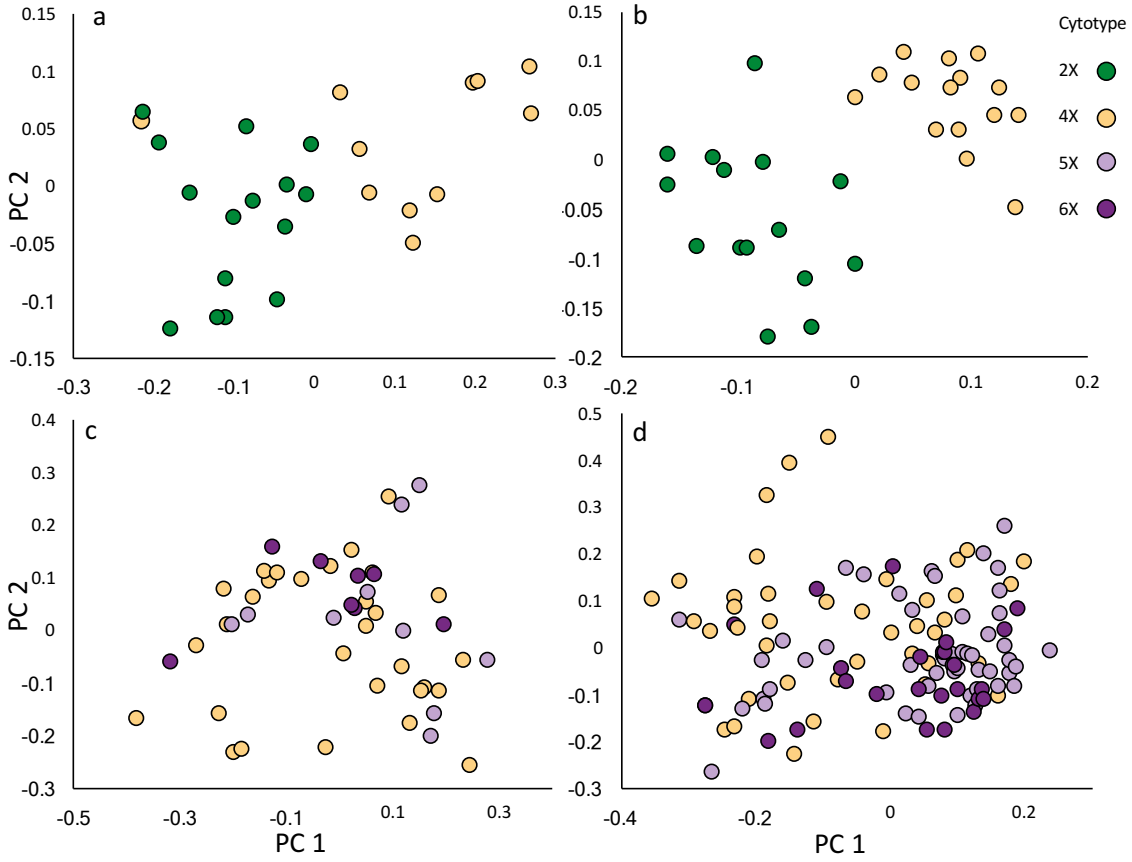


Figure 6) Principal Component Analysis (PCA) of each of the four contact zones. a) Mittelndorf and b) Prague are 2X-4X contact zones, and c) Misery Bay and d) Cheddar Gorge are 4X-6X contact zones. Triangles denote the southern rim tetraploids of the Cheddar Gorge contact zone.



Supplemental Table 1) Microsatellite loci used for gene flow assessment. All loci taken from Plue et al., 2105.

Microsatellite Loci
Camrot_002896
Camrot_003772
Camrot_010189
Camrot_010246
Camrot_011624
Camrot_013423
Camrot_015251
Camrot_019708

Supplemental Figure 1) Location of contact zone populations on chloroplast phylogeny. Labelled taxa in red comprise mixed-ploidy contact zones. See Chapter 1 for additional details.