7-16-2020

Recurrent Formation, Low Levels of Ecological Differentiation, and Secondary Dispersal Facilitate the Establishment and Persistence of Autopolyploids in *Eriophyllum lanatum*

Nicolas Alexander Diaz
Portland State University

Recurrent Formation, Low Levels of Ecological Differentiation, and Secondary Dispersal Facilitate the Establishment and Persistence of Autopolyploids in *Eriophyllum lanatum*

by
Nicolas Alexander Diaz

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biology

Thesis Committee:
Mitch Cruzan, Chair
Daniel Ballhorn
Sarah Eppley

Portland State University
2020

Abstract

The high rates of polyploidization events in angiosperms is a well-documented driver of diversification and speciation. The consequences of polyploidy—from gene expression up to ecology—and the processes facilitating the persistence of polyploidy in its early establishment in populations are poorly understood. In this thesis, I examined the role of recurrent formation, ecological differentiation, and secondary dispersal via biotic vectors in the maintenance and persistence of an intervarietal polyploid contact zone of *Eriophyllum lanatum* in Southern Oregon. Sampling 35 total populations, I used a whole chloroplast capture and flow cytometry to determine the diversity and distribution of chloroplast haplotypes and estimate the number of origins of polyploidy. Comparative ecological niche modeling was used to evaluate the relationship of the tetraploid ecological niche to the diploid niche and to measure niche overlap and niche breadth. Finally, I used a landscape genetics approach to examine patterns of seed dispersal in the contact zone. I identified 7 independent polyploidization events, indicating that recurrent formation has played an important role in maintaining polyploid populations. There was a high degree of niche overlap in diploids and tetraploids, although tetraploids occupied a slight broader niche than diploids. I found better support for an isolation by resistance model over isolation by distance model for patterns of seed dispersal. The contributions of canopy and elevation to the best supported model are consistent with secondary seed dispersal by biotic vectors, most likely hoof-epizoochory by ungulates.

Acknowledgements

I would like to thank my committee members, Dr. Sarah Eppley and Dr. Daniel Ballhorn, and especially my advisor, Dr. Mitch Cruzan, for their guidance and expertise. My lab mates, Jaime Schwoch, Elizabeth Scott, and numerous undergraduates for your assistance with field work, troubleshooting bioinformatics, and camaraderie. Dr. Sabry Elias at the Oregon State University Seed Lab for conducting flow cytometry. Dr. John Mooring for his correspondence in the initial stages of the project and for sharing his field notes. My family, friends, and cohort for their unwavering support and encouragement. I would like to acknowledge the American Society of Plant Taxonomists, the Forbes Lea Foundation Fund, and the National Science Foundation for providing funding for this work.

Table of Contents

List of Tables

List of Figures

Chapter 1: Recurrent formation, low levels of ecological differentiation, and secondary dispersal facilitate the establishment and persistence of autopolyploids in *Eriophyllum lanatum*

Introduction

Polyploidy, or the possession of more than two sets of chromosomes, has been described as the ‘‘most important amendment to Darwin and Wallace’s account of evolution” (Haldane, 1959). Advances in genomics have uncovered both recent and ancient polyploidization events across the tree of life and point to polyploidy playing a key role in driving speciation, adaptation, and complexity in biological systems (Van de Peer et al., 2017). A wealth of recent studies examining the ecological, physiological, and genomic consequences of polyploidization events has illuminated the pervasive and critical role that polyploidy has played in the evolutionary history of all plant lineages (Masterson, 1994; Soltis et al., 2009; Mayrose et al., 2011; Alix et al., 2017). A growing body of evidence suggests that polyploidy played an advantageous role for plant lineages that persisted through and beyond the Cretaceous—Paleogene extinction event, the most recent mass extinction event in which an estimated 60% of plant species went extinct, in addition to large swaths of animals and dinosaurs (Fawcett et al., 2009; Vanneste et al., 2014; Lohaus & Van de Peer, 2016; Soltis & Van de Peer, 2016). In fact, a whole genome duplication event preceding the evolution of angiosperms that was followed by additional rounds of duplications within lineages has been suggested as an explanation for the rapid diversification of angiosperms (Masterson, 1994; Jiao et al., 2011; Tank et al., 2015). Furthermore, radiations and increased rates of diversification have been demonstrated to follow polyploidization events in angiosperms (Tank et al., 2015). While

the importance of polyploidy as an evolutionary mechanism is no longer debated, both the immediate and long-term evolutionary consequences of polyploidization remain active areas of research (Spoelhof et al., 2017).

Polyploidization and subsequent genome restructuring can generate pools of novel genetic diversity for natural selection to act upon and drive speciation and adaptation (Adams & Wendel, 2005; Alix et al., 2017). This novel diversity found in polyploids is often manifested in physiological and ecological characters that are distinct from their lower-ploidy progenitors (Thompson et al., 2015; Rey et al., 2017). In addition to the divergence of physiological and ecological characters, polyploidization typically creates strong reproductive barriers between individuals of different ploidy levels—or cytotypes—and is a significant mechanism for generating reproductive isolation and sympatric speciation (Ramsey & Schemke, 1998). Herein, the use of 'cytotype' will be restricted to refer to the ploidy of an individual. However, not all polyploids are created equal: the amount of diversity and the mechanisms for maintaining diversity depend on the source of the duplicated genomes (Glover et al., 2016). Polyploids are typically characterized as either allopolyploid (merging of two divergent genomes) or autopolyploid (doubling of a single genome), however, these characterizations don't always capture the complexity of polyploidization events (Doyle & Sherman-Broyles, 2017).

Allopolyploidy is often defined as resulting from interspecific hybridization, where the allopolyploid receives two homeologous genomes, effectively becoming a fixed hybrid maintaining the parental subgenomes (Glover et al., 2016; Doyle &

Sherman-Broyles, 2017). In contrast, autopolyploidy is the result of intraspecific doubling of homologous genomes (Glover et al., 2016; Doyle & Sherman-Broyles, 2017). Taxonomy and species concepts complicate these definitions; if two morphologically distinct varieties of a single species produce polyploid progeny with intermediate morphology, should they be described as allo- or autopolyploids? An inheritance-based definition was thought to help clarify the issues with taxonomy; allopolyploidy exhibiting disomic inheritance and autopolyploidy exhibiting polysomic inheritance. Unfortunately, the allo- and autopolyploidy dichotomy sometimes fails to capture the complexity of polyploidization events, and 'mixosomic' inheritance has been documented in lineages that have not completely diverged (Soltis et al., 2016). The distinction between allo- and autopolyploidy is critical to understanding the evolutionary history of lineages: allopolyploids can maintain allelic variation from both progenitors with disomic segregation indefinitely, whereas autpolyploids will lose allelic variation through polysomic segregation over time (Doyle & Sherman-Broyles, 2017).

Most polyploid research has focused on allopolyploidy rather than autopolyploidy because of a number of historical biases (Spoelhof et al., 2017). Scientists viewed autopolyploidy as less common than allopolyploidy, likely in part because allopolyploids have received more taxonomic recognition due to their distinct morphology (i.e. resembling a combination of the two progenitors). Autopolyploids were assumed to have more disadvantages than allopolylpoids when they form and arise, for example: multivalent pairing between chromosomes resulting in aneuploid (and inviable) gametes, the loss of heterozygosity due to polysomic inheritance, and competition with similar

diploid progenitors (Ramsey & Schemske, 2002; Spoelhof et al., 2017). Recent research has suggested that auto- and allopolyploids form at similar rates (Ramsey and Schemske, 2002; Barker et al., 2016; but see Doyle and Sherman-Broyles, 2017) and that both suffer from multivalent pairing leading to reduced fertility (Zhang et al. 2013; Lloyd and Bomblies, 2016). While there have been many calls to expand research on autopolyploids (Ramsey and Shemske, 1998; Ramsey and Schemske, 2002; Soltis et al., 2007; Soltis et al., 2010), allopolyploid research remains much further ahead.

The rate of formation and the ecological and genomic processes facilitating the establishment of polyploids continues to be a fruitful avenue of research (Husband & Sabara, 2003; Kliber & Eckert, 2005; Mooring, 2008; Trávníček et al., 2011; Certner et al., 2017). When a polyploid first arises in a population, it must quickly overcome the minority cytotype exclusion principle (MCE), which holds that mixed-ploidy populations are not stable due to the low reproductive success between cytotypes (Ramsey & Schemke, 1998) and the lack of appropriate mates for the minority cytotype (Levin, 1975). Between MCE, reduced fertility due to meiotic abnormalities, and competition with diploids, the deck would seem to be stacked against neopolyploids. Allopolyploids can benefit from heterosis and are less similar to their progenitors than autopolyploids, which are highly similar and frequently indistinguishable from their progenitors (Soltis et al., 2007). Thus, autopolyploids provide an opportunity for identifying mechanisms promoting polyploid persistence because they are not confounded by hybridization, as is the case of allopolyploids.

Despite the odds, polyploids form, establish, and persist through a wide variety of mechanisms. Some of which include: a transition to self-compatibility, vegetative reproduction, a shift in ecological niche, assortative mating, immigration of similar cytotypes, and superior dispersal ability (Husband & Sabara, 2003; Munoz-Pajares et al., 2017; Certner et al., 2017; Herben et al., 2017). These mechanisms vary across studies and are often specific to a study system, as no generalizable "rules" have emerged for how important these mechanisms are or how they operate (Soltis et al., 2016). In autopolyploid systems, research has primarily focused on the role of unreduced gamete production or recurrent formation of polyploids (Oswald & Nuismer, 2011; Spoelhof et al., 2017). Factors that aid autopolyploids in escaping MCE or reduce competition with diploids, such as niche shifts, have been studied, but yield inconsistent results (Baack, 2005; Glennon et al., 2014; Visger et al., 2016; Gaynor et al., 2018). Consequently, there have been calls to expand the systems used in autopolyploid research to address basic questions regarding polyploid establishment (Soltis et al., 2016; Spoelhof et al., 2017).

Considerably less attention has been given to the ways in which polyploidy can alter biotic interactions and impact communities (Segraves, 2017). Interestingly, some hypotheses about how polyploids might escape MCE, e.g. assortative mating, allude to community effects of polyploidy; yet few studies have tried to capture the ways in which biotic interactions, in addition to other factors, might reinforce and promote the establishment of polyploids (but see Thompson et al., 2004; Kennedy et al., 2006; and Těšitelová et al., 2013). Research examining polyploid formation and persistence should consider the role of both abiotic and biotic factors. The field of landscape genetics offers

a framework that allows for the consideration of both abiotic and biotic factors in how polyploids disperse and persist (Cruzan & Hendrickson, 2020). Landscape genetics utilizes population genetics and spatial analyses to estimate the effect of the landscape features on dispersal (Manel et al., 2003). Observing pollen and seed dispersal in plants is often both difficult and impractical given the size and volume of pollen and seeds that can be produced by an individual plant. Fortunately, due to maternal inheritance of the chloroplast, chloroplast markers can be used in landscape genetics to measure effective or realized seed dispersal (Cruzan and Hendrickson, 2020). Using this framework allows for the testing of two hypotheses: isolation by distance (IBD; Wright, 1943) and isolation by resistance (IBR; McRae & Beier, 2007). IBD assumes that gene flow is more likely to occur between geographically close populations, and thus genetic distance increases with geographic distance between populations. Whereas IBR is based on a model that uses circuit theory to assign resistance values to different 'paths' that connect populations across the landscape, thus genetic distance between populations is influenced more by landscape features than geographic proximity. There are few examples of landscape genetic analyses on plants (e.g. Arredondo et al., 2018; Grasty et al. 2020), but to date, there is no published research on the landscape genetics of poylploid plants that take advantage of optimized circuit theory methods (i.e. Resistance GA, Peterman, 2018).

A landscape genetic analysis that utilizes chloroplast markers and ecological niche modelling can provide information about polyploid establishment and persistence beyond identifying how landscape features influence dispersal. While chloroplast genetic markers are not typically used in intraspecific studies due to their slow evolution, and

thus low variability, new methods for whole chloroplast sequencing and genotyping have made their use in landscape genetic analyses possible (Kohrn et al., 2017; Grasty et al., 2020). Chloroplast haplotype networks based on individuals whose ploidy has been determined through flow cytometry allow for the inference of the number of origins of polyploidy. The abiotic niche can be estimated through the construction of an ecological niche model, while biotic interactions (i.e. dispersal vectors) can be incorporated through careful selection of landscape variables that influence the movement of those vectors (Cruzan and Hendrickson, 2020). Additionally, ecological niche modelling can also be used to test the niche shift hypothesis separate from the landscape genetic model. By integrating chloroplast genetic data and ecological niche modelling, landscape genetics has the potential to offer unique insight to the establishment of poylploids.

A polyploid contact zone at the edge of the ranges of *Eriophyllum lanatum* var. leucophyllum and achillaeoides around Medford, Oregon, provides an opportunity to address questions regarding polyploid formation and establishment using both landscape genetics and ecological niche modelling. The contact zone was previously identified and described by J.S. Mooring (2008) using chromosome squashes and morphological comparisons. Polyploids can be found throughout the ranges of both varieties (Mooring, 2008), suggesting that polyploids have arisen from the union of unreduced gametes within a population and could be considered autopolyploids. However, polyploid populations increase in frequency with proximity to the contact zone where the varieties meet, with individuals appearing intermediate in their morphology (Mooring, 2001; Mooring 2008). Given the weak reproductive barriers between varieties and their highly

variable morphology both within and outside the contact zone, these polyploids more than likely fall closer to the autopolyploid end of the auto- allopolyploid continuum. Greenhouse crosses revealed that individuals —regardless of ploidy level—were self-incompatible, suggesting that polyploid individuals must rely on vegetative reproduction and/or the establishment of other nearby sexually compatible polyploids to persist (Mooring, 2001; Mooring, 2008). The distribution of polyploids was found to not be correlated with soil, climate, topography, geological history, or species interactions (Mooring, 1975; Mooring 2008). For these reasons, Mooring described the hybrid zone as "perplexing" (J.S. Mooring personal communication), and proposed that polyploidy has had a stabilizing effect on intervarietal hybrids (Mooring, 2008). Mooring's work was limited both by the "difficult preparations" of chromosome squashes (Mooring, 2008), and the lack of high throughput analytical tools that are available today, (e.g. flow cytometry to estimate ploidy).

With the foundation of Mooring's work, and the advent of new analytical tools, this contact zone can be used to ask questions about polyploid formation and persistence. Flow cytometry and whole chloroplast sequencing will provide insight as to whether recurrent formation has contributed to poylploid persistence. Examining the relationships of confirmed tetraploids to diploids in a cpDNA haplotype network will allow for an estimate of the number of times polyploids have formed independently. The construction of ecological niche models and environmental Principal Components Analysis (PCA) will address the niche shift hypothesis by quantifying niche overlap and niche breadth of both tetraploids and diploids. Finally, a landscape genetic analysis that incorporates

cpDNA, and separates the effects of habitat suitability (estimated from ecological niche modelling) from landscape features that influence dispersal, will address whether local dispersal or secondary dispersal through biotic vectors has been important for polyploid establishment and persistence. Utilizing whole chloroplast sequencing, ecological niche modeling, and landscape genetics, I will address the following questions in *Eriophyllum lanatum* var. leucophyllum and achillaeoides: (1) How many times has polyploidy arisen? (2) Is there a shift in ecological niche between diploids and tetraploids? and (3) How do biotic interactions or landscape features influence dispersal?

Methods

Sampling

To be able to identify the haplotypes associated with each variety in the contact zone, sampling was conducted from Northern California to Washington, covering the partial range of *E. lanatum* var. achillaeoides, the contact zone around Medford, Oregon, and the majority of the range of var. leucophyllum. A total of 35 populations were sampled: 5 populations within the range of var. achillaeoides, 6 populations within the range of var. leucophyllum, and 23 populations spanning the contact zone (Figure 2a, 2b). For each population, leaf tissue was collected from 20 individuals, with a minimum of 1 meter between sampled individuals. *E. lanatum* frequently clonally propagates from roots and forms clumps (Mooring, 2008); when possible, space between sampled individuals was maximized to both avoid sampling clones and to sample evenly throughout populations. Three leaves were sampled from each individual and were dried with silica beads. Linking ploidy to individual chloroplast haplotypes was necessary to understand the distribution of cytotypes across the contact zone. Accordingly, we sampled 10 individuals from 20 populations within the contact zone.

Chloroplast Capture and Haplotype Calling

DNA was extracted using the Qiagen DNeasy 96 Plant Kit (Qiagen, Redwood City, California) and subsequently quantified using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham Massachusetts). Using the Kapa HyperPlus kit for Illumina (Kapa Biosystems Inc., Wilmington, Massachusetts) and NEBNext Dual Index primers (New England BioLabs Inc., Ipswich, Massachusetts), libraries were prepared from equimolar pools for each population (20 individuals), a randomly selected individual

from each population (single sample library - SSL), and each of the cytotyped individuals (SSLs) (Kohrn et al., 2017). Following library construction, individual and pooled samples were multiplexed for chloroplast target enrichment. A whole chloroplast genome capture was performed using a custom MYBaits target enrichment kit (Arbor Sciences, Ann Arbor, Michigan). The enriched libraries were sequenced on an Illumina HiSeq 2500 (Illumina Inc., San Diego, California) at the Oregon Health and Sciences University Massively Parallel Sequencing Shared Resource facility (MPSSR, OHSU, Portland, Oregon).

Pooled and individual cpDNA was sequenced to identify single nucleotide polymorphisms (SNPs), and in turn, use these SNPs to understand gene flow and haplotype diversity. Sequence data was processed by: (1) removing adapter sequences with CutAdapt 1.13 (Martin, 2011), (2) removing low-quality base pairs with Sickle (Joshi & Fass, 2011), (3) aligning trimmed sequences to a de novo *E. lanatum* chloroplast genome with BWA-MEM 0.7.15 (Li, 2013), (4) realigning reads around indels with Picard Tools 2.9.0 (http://broadinstitute.github.io/picard/), and (5) calling SNPs with FreeBayes 1.0.2 (Garrison & Marth, 2012). SNPs were filtered at a depth of 400 base pairs using a custom python script. SSLs were used to construct an initial haplotype network phylogeny, which was then used by CallHap (Kohrn et al. 2017) to discover new cpDNA haplotypes and estimate haplotype frequencies from pooled populations. Haplotype frequencies from each pool were used to calculate pairwise N_{ST} for each population using SPAGeDi (Hardy & Vekemans, 2002).

Flow Cytometry

Flow cytometry was used to determine the distribution of cytotypes across the contact zone and to assess the cytoptype composition of populations. Leaf tissue and flower buds were collected from 14 populations with 10 individuals sampled per population. Ploidy estimation was conducted at the Oregon State University Seed Lab using a Partec PA flow cytometer. The cytometer was calibrated using leaf tissue from populations with previously determined chromosome counts by John Mooring Ph.D. (Mooring, 2008). Ploidy estimations were based upon the fluorescence peaks, an approximation of nuclear DNA content, for each sample. Chromosome squashes were attempted to further validate flow cytometry results. Immature flower buds were collected from the same individuals used in flow cytometry and fixed in Farmer's fixative for 24 hours and then transferred to 70% ethanol for long-term storage. Chromosome squashes were conducted following the protocol outlined in Windham et al. 2020.

Comparative Ecological Niche Modeling

Ecological niche modeling was conducted on diploids and tetraploids to test the niche shift hypothesis and characterize the relationship of the tetraploid niche to the diploid niche. An initial set of predictor variables (Table 1) were sourced from WorldClim and other online GIS repositories (Fick & Hijmans et al., 2017; O'Donnell & Ignizio, 2012). Due to spatial autocorrelation, a Pearson's pairwise correlation test was performed on all variables. Variable retention was determined by correlation values (< 0.8 and variables that would be biologically relevant to *E. lanatum*.

Maxent is a machine learning program that utilizes a maximum entropy algorithm to generate species distribution models or ecological niche models (Phillips et al., 2017).

Maxent has a set of default parameters that can be appropriate to use for a variety of species given that certain criteria (e.g. sample size) are met for the input files. If some or not all of the criteria are met for the input files, it is recommended to generate models using different parameters and settings and to use model selection approach to determine the "optimal" settings (Warren & Seifert, 2011; Merow et al., 2013). ENMEval is an R program that generates models for every combination of specified settings and parameters and then uses a model selection approach to determine the most appropriate settings (Muscarella et al., 2014). Given the small sample sizes for both Diploids and Tetraploids, the following model settings were tested in ENMEval: Linear and Quadratic features, n-1 jackknife, and regularization multipliers of 0.5, 1, 1.5, and 2. The optimal model settings were determined by using the model with the lowest AIC value.

Final niche models were constructed in Maxent using the settings determined from ENMEval. Each model was built using: 100 replicates, a random seed for partitioning training and testing points, bootstrapping for replicates, and jackknife to measure variable importance. The final models consisted of the average values across each replicate run.

ENMTools 1.4.4 is a program used to calculate various metrics and test hypotheses with ecological niche models (Warren et al., 2010). The ASCII files for the final models were imported into ENMTools to calculate Levin's niche breadth and Schoener's niche overlap (D) (Levin, 1968; Schoener, 1968). The difference in niche breadth was calculated as follows: $Tetraploid_{NB} - Diploid_{NB}$. Parametric statistics are not appropriate to evaluate the output of overlap and niche breadth statistics. To evaluate the

observed niche overlap value, a niche identity test with 100 replicates was run to generate a distribution of expected values for niche overlap. The same approach was used to generate a null distribution for the difference in niche breadth.

Environmental Principal Components Analysis and ANOVA

The relationship between the tetraploid niche and diploid niche was further explored using Principal Components Analysis (PCA) and ANOVAs. Extract Multi Values to Points tool in ArcMap was used to extract values from environmental variables at the geographic coordinates for each population and was exported as a table. The table was imported into R and a PCA was conducted using the 'prcomp' function. Values were centered and scaled prior to the analysis. The PCA was visualized using the 'ggbiplot' package. Following visualization, ANOVAs were run on each of the environmental variables to test for differences in variation at the sites for each ploidy. ANOVAs were run in R using the 'aov' function.

Landscape Genetics

Landscape layers (e.g. roads, rivers, and development) and cpDNA genetic distance were used to investigate how landscape features affect seed dispersal in the contact zone. Layers were selected based on the following criteria: (1) the layer must represent a feature of the landscape likely to affect the dispersal ability of the plant, and (2) the layer must be distinct from environmental data that would be considered a component of the environmental (abiotic) niche of *E. lanatum*. All layer processing was conducted in ArcMap 10.5.1, all tools mentioned herein can be found in the standard toolboxes and the Spatial Analyst toolbox. Agriculture, Canopy, Development, and

Rivers were extracted from the National Land Cover Dataset for Oregon using the extract by Attributes tool. Two Digital Elevation Model (DEM) rasters were used to construct an elevation layer using the mosaic to new raster tool. A roads layer consisting of polygons was converted to a roads raster using the polygon to raster tool. A habitat suitability layer was constructed as described in the niche modeling methods using a subset of the environmental layers: Bio1, Bio3, Bio5, Bio6, Bio14, percent sand, and soil pH. In addition to the high degree of niche overlap, a greenhouse crossing experiment revealed that intercytotype gene flow is possible, though less successful than ploidy-matched crosses. Consequently, all tetraploid and diploid occurrences were used to construct the habitat suitability layer. To account for ploidy differences at the population level, a ploidy layer was constructed using Thiessen Polygons.

Coordinates of populations with known ploidy were saved as a csv file and imported into ArcMap. The coordinates were converted from WGS84 to the NAD83 UTM Zone 10N projection. The envelope shape in the minimum bounding geometry tool was used to fit a polygon around the locations. A 5-kilometer buffer was added to the polygon, which resulted in rounded corners. The minimum bounding geometry tool was used again to eliminate the rounded corners and create the final study area. A model was constructed to batch process the landscape layers doing the following: (1) project into the NAD83 UTM Zone 10N projection, (2) resample to set all cell sizes to 30x30, (3) extract by mask to set the extent to exactly match the final study area, and (4) convert the rasters to ASCII format. The habitat suitability layer was rescaled using the Raster Calculator to

eliminate long decimals. ASCIIs were inspected in Notepad++ to ensure cell size and extent were uniform across all layers.

ResistanceGA

ResistanceGA is an R program used to optimize continuous and categorical landscape resistance values using a genetic algorithm that incorporates the pairwise genetic distance of populations (Peterman, 2018). CommuteDistance, found in the gdistance R package (van Etten, 2017), was selected to generate the resistance matrix over the popular Circuitscape (McRae et al., 2008), due to significantly faster processing time (Marotte & Bowman, 2017; Arredondo et al., 2018). CommuteDistance employs the same algorithm as Circuitscape, in which pairwise resistance distances between populations are calculated, which informs the construction of the overall resistance matrix, and finally tests the ability of the resistance matrix to predict the genetic distance matrix (N_{ST} – calculated based on cpDNA haplotype frequencies). Each individual layer was optimized using the single surface optimization function to determine the best transformation to apply to the layer for multi-surface optimization. Once each layer was optimized, the multi-surface optimization function was used to measure how much of the genetic variation could be explained by a composite layer constructed using all of the of the resistance layers. The Resist.boot function, a subsampling without replacement bootstrap analysis, was run to determine the relative support for each resistance layer. Bootstrapping was run for 10,000 iterations using a randomly selected subset samples representing 75% of the total samples. Resist.boot employs a maximum likelihood population effects parameterization model (MLPE) to fit the N_{ST} matrix to each of the

resistance layers. MLPE models are less error prone and can accommodate non-independent samples, unlike multiple regressions on genetic distance matrices (MRDM) and Mantel tests (Row et al., 2017; Grasty et al., 2020). The final resistance model was determined using AICc scores and the Top Model output from the bootstrap analysis. Linear regressions examining the relationship between genetic distance (N_{ST}) and both geographic distance (IBD) and resistance distance (IBR) were conducted in R using the 'lm' function found in the stats package.

Results

Haplotype Distribution and Diversity

After processing and filtering 106 SNPs were recovered from 700 individuals across 35 populations, resulting in 51 haplotypes. CallHap discovered 15 new haplotypes, with the remaining 36 haplotypes coming from single sample libraries. New haplotypes ranged in frequency from 5%-55% (1 to 11 out of 20 individuals) at an individual site. Haplotype diversity was highest in the Medford region, a previously identified intervarietal polyploid contact zone (Figure 1a; Mooring, 2008). Haplotypes were not associated with varietal taxonomic designations. The most abundant haplotype (salmon pink, Figure 1a, 1b), which has a central position in the overall network, was found at 17 different sites and was assigned to 181 of the total individuals sampled at the population level. The second most abundant haplotype (pale green, Figure 1a, 1b), was one mutation (SNP) away from the most abundant haplotype, and was found in 54 individuals across 7 different sites. Most haplotypes are separated by 1-2 SNPs, 8 haplotypes are separated by 3 or more SNPs from their neighboring haplotype, with the most distant haplotype having 20 SNPs between it and its neighbor.

Origins of polyploidy

Flow cytometry revealed that 16 out of the 20 populations chosen were made up of a single cytotype (Table 2). In addition to diploids and tetraploids, mixed-ploidy populations contained individuals that could not be assigned to either ploidy (based on the values and shapes of the peaks; Sabry Ellis, personal communication). Numerous difficult and time-consuming attempts at chromosome squashes failed, which is

consistent with John Mooring's description of his attempts (personal communication; images from attempts can be found in the Appendix). Sequencing of individuals of determined ploidy (via flow cytometry) allowed for a conservative inference of 7 polyploid formation events, suggesting polyploids have been recurrently formed in this region (Figure 2). In two central haplotypes (II and III; Figure 2), tetraploids outnumber diploids, and another central haplotype (V; Figure 2) was only associated with tetraploids. Based on their central positions and outnumbering of diploids, these polyploid lineages likely arose from older polyploid formation events. The formation events at the tips of the network are more recent (IV, VI, VII; Figure 2).

Comparative ecological niche modeling

Using AICc, the optimal model settings for tetraploids determined by ENMEval included: Linear features and a regularization multiplier of 1.5. With 100 bootstrapped replicates, the tetraploid model (Figure 3a) had an average AUC score of 0.844. The jackknife test of variable importance revealed Max Temperature of the Warmest Month (bio5) and Annual Mean Temperature (bio1) contributed to 90.6% of the model, with the remaining variables making up the other 10 percent (Table 3). Variable response curves indicate that tetraploid habitat suitability is highest in areas where the maximum temperature reaches at least 28.9^{o}C.

Optimal model settings for diploids determined by ENMEval included: Linear and Quadratic features and a regularization multiplier of 1.5. With 100 bootstrapped replicates, the diploid model (Figure 3b) had an average AUC score of 0.878. The jackknife test of variable importance revealed Mean Temperature of Coldest Quarter

(bio11) and Annual Mean Temperature (bio1) contributed to 90% of model, with the remaining variables making up the other 10 percent (Table 3). Variable response curves indicate that Diploid habitat suitability is strongly influenced by an average temperature of 5.23^{o}C during the three coldest months of the year.

The observed niche overlap, Schoener's D, for the diploid and tetraploid models was 0.47. The niche identity test revealed that the difference between the models was not significant ($p > 0.18$; Figure 4a). The observed difference in niche breadth ($Tetraploid_{NB}$ – $Diploid_{NB}$) was 0.11, indicating that tetraploids occupy a broader niche than diploids. However, this difference was not significant ($p > 0.17$; Figure 4b).

Environmental Principal Components Analysis and ANOVA

Two principal component axes explain most of the environmental variation (74.7%) in the contact zone. Groupings on the PCA plot were consistent with the results of the niche overlap and breadth tests. The tetraploid niche is broadly stretched across PC1 and PC2, while the diploid niche is narrower and largely overlapping with the tetraploid niche (Figure 5). The niche of the mixed populations is almost completely enveloped by the tetraploid niche and has a small overlapping region with diploids (Figure 5).

Soil pH and percent sand were the only environmental predictors that had significant differences in their variance amongst the ploidies (Table 4). Mean values for mixed populations did not uniformly fall in between tetraploid and diploid values and were more frequently higher or lower than the two ploidies. Mixed populations were

intermediate between diploid and tetraploids for: Min Temperature Coldest Month and Mean Temperature of Coldest Quarter (Table 4).

Landscape genetics

The combined multi-surface model was identified as the best supported model describing genetic distance amongst the populations of *E. lanatum* (average marginal R^2 = 0.046, average AICc = -219.21; Figure 6a). It was selected as the top model in all of the 10,000 bootstrap iterations, outperforming both geographic distance (average marginal R^2 = 0.028, average AICc = -90.57; Figure 6b) and each of the individual surface layers (Table 5). The single optimized layers of agriculture, habitat suitability, elevation, and canopy all had average marginal R^2 values greater than the combined surface, with agriculture explaining the most variation in genetic distance (average marginal R^2 = 0.12, average AICc = -88.18; Table 5). In the combined surface, canopy explained most of the variation, followed by elevation and agriculture (Table 5).

A linear regression confirmed geographic distance poorly explains genetic distance (N_{ST}) (R^2= -0.004542 F-statistic = 0.1454, DF = 1 and 188, p-value = 0.7034; Figure 6b), thus rejecting the hypothesis of isolation by distance. Resistance distance, based on the top model, could better explain genetic distance, though this relationship was not statistically significant (R^2 = 0.01248, F-statistic: 3.388, DF = 1 and 188, p-value: 0.06724; Figure 6a). Bootstrapping and linear regressions revealed IBR to be the better supported model for seed dispersal.

The optimized resistance values reflect which features of the layers are conduits or barriers to dispersal. In both the single surface and the combined multi-surface models,

rivers, development, and agriculture were consistently treated as a barriers. As a single layer, an inverse ricker transformation was applied to habitat suitability, indicating that unsuitable habitat and highly suitable habitat function as barriers whereas low to moderately suitable habitat is a conduit. However, in the multi-surface model all habitat values were set to 1, suggesting its contribution to the model was equal to that of geographic distance (although habitat contributed to less than 1% of the multi-surface model). Roads were treated as a conduit in both single-surface and multi-surface models. In the ploidy layer, regions assigned as diploid were treated as conduits in both single surface and multi-surface models, tetraploid regions were barriers in the single surface but conduits in the multi-surface, and mixed-ploidy regions were barriers in both models. As a single surface, elevation was assigned an inverse monomolecular trans formation, wherein areas of low elevation functioned as barriers and high elevation was a conduit for dispersal. Making up 18.17% of the multi-surface model, elevation was assigned an inverse ricker transformation: very low and high elevation were barriers, and moderately low elevation was a conduit. Canopy as a single surface was optimized with an inverse ricker transformation, with moderately low canopy as a conduit, and high and very low canopy as a barrier. As the largest contribution to the multi-surface model (75.33%, Table 6), a ricker transformation was applied to canopy: low canopy was a strong barrier, and high canopy a conduit.

Discussion

The polyploid contact zone at the interface of the ranges of *E. lanatum* var. achillaeoides and leucophyllum is composed of diploid, tetraploid, and mixed ploidy populations. Chloroplast haplotypes are shared amongst the two varieties, and do not correspond with taxonomic identities. The ecological niche of tetraploids is slightly broader than that of diploids, although there is a high degree of overlap. The recurrently formed tetraploids have persisted in this region longer than some diploids, and there is some evidence suggesting intercytyope gene flow or the presence of tetraploids undergoing diploidization. Finally, genetic connectivity in the contact zone is best explained by an isolation by resistance model (IBR), wherein the landscape features, primarily canopy and elevation, influence seed dispersal.

Polyploidy appears to have arisen several times in this region, which is likely one of the main drivers for the persistence of these autopolyploids. Considering *E. lanatum* is self-incompatible, recurrent formation eases the pressure of the minority cytotype exclusion (MCE) principle, creating more opportunities for successful mating between polyploids (Ramsey & Schemske 2002). The number of estimated formation events is comparable to other polyploid systems such as *Heuchera grossulariifolia* with an estimated 2-7 origins (Segraves et al., 1999) and *Astropelis integerrima* with 10 origins (Beck et al., 2012), but does not approach that of *Galax urceolata* with an estimated 47 independent origins (Servick et al., 2015). Recurrent formation may have been necessary but insufficient for the establishment of polyploids; thus, it is necessary to consider other contributing factors. Individuals that could not be confidently assigned to diploidy or

tetraploidy may represent triploid individuals, which have been found in other poylploid contact zones and are potentially a result of and bridge for intercytotype gene flow (Ramsey & Schemske, 2002; Baack, 2004; Sabara et al., 2013; Servick et al. 2015; Barringer & Galloway, 2017). *E. lanatum*'s tendency to clonally propagate from roots may have also contributed to the rate at which polyploids formed and provided additional opportunities for gene flow, although this does not do much to mitigate the self-incompatibility problem. The case of *E. lanatum* is somewhat similar to that of the classic autopolyploid systems, *Galax urceolata* and *Pilosella rhodopea,* both of which exhibit recurrent formation, clonal propagation, and the maintenance of self-incompatibility mechanisms across ploidal levels (Servick et al., 2015; Barringer & Galloway, 2017; Gaynor et al., 2018; Šingliarová et al., 2019).

With high niche overlap between the ploidies, and tetraploids occupying a slightly broader ecological niche than diploids, it is unlikely that a niche shift in tetraploids by itself facilitated the establishment and persistence of polyploidy in this region. However, some have suggested that in autopolyploids, small deviations from niche identity may be important for dampening the effects of MCE (Visger et al., 2016; Spoelhof et al., 2017). With IBR as the best supported model for dispersal, secondary seed dispersal via biotic vectors such as ungulates or avians, in addition to local passive dispersal, may have played an important role in the movement of cytotypes to establish new populations and introduce compatible mates (Heinken & Raudnitschka, 2002; Segraves, 2017). It is likely that a combination of recurrent formation, intercytotype gene flow, *E. lanatum*'s

tendency to clonally propagate, a slight shift in ecological niche, and secondary dispersal via biotic interactions enabled these polyploids to escape MCE.

With 51 unique haplotypes across 35 populations, the level of variation observed in the chloroplast genome of *E. lanatum* is interesting in and of itself. By comparison, a study that sampled 32 populations of *Ranunculus occidentalis* across a larger geographic range recovered 18 unique haplotypes (Cruzan & Hendrickson, 2020). Interestingly, the geographic location of most of the variation is consistent with the intervarietal polyploid contact zone that was previously identified (Mooring, 2008). For varieties achillaeoides and leucophyllum, these taxonomic designations are not consistent with chloroplast haplotypes. There are many chloroplast haplotypes that only occur in the contact zone, and considering the diversity of haplotypes both within and among populations there, this contact zone in Southern Oregon may be the confluence of several lineages that were previously isolated in glacial refugia (Cruzan & Templeton, 2000). With 75 binomials and trinomials having been applied to this group*, E. lanatum* has been taxonomically troubling for almost 100 years due to its highly variable morphology, low reproductive barriers between varieties, the role of polyploidy, and a large range (Constance, 1937; Cronquist, 1955; Mooring, 2008). It is possible that polyploidy has played a direct or indirect role in generating the diversity within this lineage. Polyploids were first identified in the study region in 2001 by J.S. Mooring, who hypothesized that polyploids represented the results of hybridization between varieties (i.e. allopolyploids), and nothing more (Mooring, 2008). Based on the diversity and distribution of chloroplast haplotypes, the polyploids in this region are more akin to autopolyploids than

allopolyploids. The central positions of tetraploid haplotypes in the network indicate that some polyploids have persisted longer than the diploids they arose from. The individuals of unknown ploidy, previously discussed as potential 'triploids', were assigned to some of the central haplotypes. An alternative explanation for these ploidy-undetermined individuals is that, given their age and that they are concentrated in a few populations within close proximity of each other, they may represent individuals that have undergone some diploidization – the process of genome loss and fractionation that returns polyploids to a diploid state (Soltis et al. 2016). This would need to be confirmed by chromosome squashes and/or sequencing of the nuclear genome.

The niche conservatism observed in tetraploids is not unexpected for an autotetraploid (Baack & Stanton, 2005; Visger et al., 2016), which does not acquire new alleles from another divergent genome as is the case in allopolyploidy (Stebbins, 1950). The similarity in morphology and genetics of diploids and tetraploids might increase the strength of competition between the ploidies when the co-occur, making processes that facilitate assortative mating (e.g. niche shift) even more important (Visger et al., 2016). Both niche modelling and the PCA revealed the degree of niche overlap in tetraploids and diploids and indicated that tetraploids occupy broader ecological space, which has been observed in other polyploid complexes (Glennon et al., 2014; Visger et al., 2016). With significant differences in the soil features from the ANOVA, it was surprising that in the niche modeling these features contributed to only a small percentage of the final models (Table 3; Table 4). The observed niche shift may be the direct result of whole genome duplication events (most likely through physiological changes), but could also be

explained by selection subsequent to the duplication events. A common garden with both naturally and artificially formed neotetraploids could shed light on the processes underlying niche evolution in this system.

The landscape genetic analysis revealed complex dispersal patterns that are better explained by a combination of landscape features (Isolation By Resistance – IBR; McRae and Beier, 2007) than geographic distance (Isolation By Distance – IBD; Wright, 1943). Among the variables considered in the best model, canopy and elevation were particularly important in explaining genetic variation. While these results suggest wind as a dispersal vector, the pappus of *E. lanatum* seeds is reduced to a crown of short scales, and is likely not dispersed via wind. An alternative explanation is that most seeds fall nearby their progenitor, and on occasion, secondary dispersal occurs through movement of seeds by animal vectors. While dispersal events were detected throughout the contact zone, as evidenced by haplotype sharing amongst the populations (e.g. dark brown, salmon, and pale green haplotypes; Figure 2b), landscape genetic models could only explain a small amount of the genetic variation present in these populations. Dispersal facilitated by biotic vectors such as ungulates, which are common in this region, could have been an important mechanism for connecting the recurrently formed tetraploids and promoting both their establishment and persistence (Albert et al., 2015; Segraves 2017; Baltzinger et al., 2019). Seeds that fall locally or are detached from the mother plant by ungulates can be picked up and carried in mud that clings to hooves (hoof-epizoochory). Indeed, *E. lanatum* meets 3 of 7 criteria that increase the likelihood of hoof-epizoochory – open habitat, release height, and lack of an appendage – as identified in a trait-based

meta-analysis of ungulate seed dispersal (Albert et al., 2015). High canopy was assigned as a conduit for seed dispersal in the multi-surface model, however, *E. lanatum* grows in open habitat. This suggests that seed dispersers tend to move along paths with greater canopy cover, which is consistent with ungulate preferences for cover in fragmented landscapes (Hewison et al., 2001).

The barriers of development and agriculture, which were assigned the strongest resistance values, may have further shaped the genetic structure of the contact zone by influencing the movement of seed dispersers (Hewison et al., 2001). Human modifications to the landscape can reduce habitat availability for both *E. lanatum* and its biotic counterparts, including pollinators and secondary dispersers, and thus reducing connectivity. Roads, however, are conduits to dispersal, which may be explained by the proximity of a few populations to roads and the way in which roads cut through and connect the fragmented landscape (Ansong & Pickering, 2013). Further, this result is consistent with field observations of *E. lanatum* growing along the rocky soils exposed on the sides of roads. Rivers, another barrier, may also constrain the movement of both pollen and seed dispersers that are unable or less likely to cross a river. Ploidy is likely playing a role in structuring the genetic diversity in this contact zone, although the construction of the ploidy layer with Thiessen polygons may not have been the ideal approach to approximate this relationship (as demonstrated by the ploidy layer being ranked the worst model overall and contributing only 0.33% to the multi-surface model; Table 5 and 6). Nonetheless, this novel use landscape genetics to investigate seed dispersal in a polyploid system was revealing. These results highlight the importance of

both biotic interactions as well as secondary seed dispersal in the maintenance and persistence of polyploid lineages.

Conclusions

The aim of this research was to address basic questions about the mechanisms promoting polyploid formation, establishment, and persistence. The intervarietal polyploid contact zone where *E. lanatum* var. achillaeoides and leucophyllum meet in southern Oregon provided a unique opportunity to investigate the mechanisms of recurrent formation, abiotic niche shift, and biotic interactions. Chloroplast haplotype networks and flow cytometry indicate polyploids have arisen at least four times, suggesting recurrent formation was important in the maintenance of the contact zone. Given the relative age of some tetraploid haplotypes, it does not seem that recurrent formation was enough on its own to allow polyploids to persist. The slightly broader ecological niche of tetraploids relative to diploids may have allowed for the establishment of new populations outside of the diploid niche space, thus promoting successful mating amongst tetraploids. Despite the diminutive pappus of *E. lanatum*, patterns of seed dispersal in the contact zone were best explained by the effect of landscape features rather than geographic distance. This suggests seeds are being moved by biotic vectors, such as ungulates, and that biotic interactions have been involved in the persistence of this contact zone. Polyploidy and weak reproductive isolation between varieties have undoubtedly played an important role in shaping the diversity and distribution of chloroplast haplotypes in *E. lanatum*. Whether intercytotype gene flow or diploidization of tetraploids is occurring in this region is unclear, though the use of nuclear genomic data would likely prove to be illuminating. Landscape genetics has been demonstrated to

be a useful tool for investigating the role of biotic interactions in the persistence of polyploids, future studies of polyploid systems will benefit from its utility.

Tables

Table 1. Bioclimatic and edaphic variables considered for use in ecological niche modelling. Bold font indicates layers that were retained for model construction

Layer	Description
bio 1	**Annual Mean Temperature**
bio 2	Annual Mean Diurnal Range
bio 3	**Isothermality**
bio 4	Temperature Seasonality
bio 5	**Max Temperature of Warmest Month**
bio 6	**Min Temperature of Coldest Month**
bio 7	Annual Temperature Range
bio 8	Mean Temperature of Wettest Quarter
bio 9	Mean Temperature of Driest Quarter
bio 10	Mean Temperature of Warmest Quarter
bio 11	**Mean Temperature of Coldest Quarter**
bio 12	Annual Precipitation
bio 13	Precipitation of Wettest Month
bio 14	**Precipitation of Driest Month**
bio 15	Precipitation Seasonality
bio 16	Precipitation of Wettest Quarter
bio 17	Precipitation of Driest Quarter
bio 18	Precipitation of Warmest Quarter
bio 19	Precipitation of Coldest Quarter
Clay (%)	Percentage of clay in soil
Sand (%)	**Percentage of sand in soil**
Silt (%)	Percentage of silt in soil
Soil pH	**pH of the soil**

Table 2. Population locations (WGS 84), names, and ploidy determined by flow cytometry.

Latitude	Longitude	Population	Ploidy
42.41628	-122.77571	Agate Lake	Tetraploid
42.38638	-122.89174	Central Cemetery	Tetraploid
42.46143	-122.88158	Denman	Tetraploid
42.47236	-122.79391	Eagle Hill	Tetraploid
42.50557	-122.89545	Glass	Tetraploid
42.43580	-123.0516	Gold Hill	Tetraploid
42.43775	-122.98508	Gold Ray Dam	Diploid
42.43583	-122.98819	Gold Ray Dam Railroad	Diploid
42.35139	-122.97154	John's Peak	Diploid
42.46851	-122.94812	Lower Table Rock	Mixed
42.46241	-122.94863	Lower Table Rock 2	Tetraploid
42.44535	-122.95143	Lower Table Rock 3	Diploid
42.45814	-122.95325	Lower Table Rock 4	Diploid
42.49863	-122.94333	Perry	Diploid
42.34732	-122.78734	Roxy Anne	Tetraploid
42.46958	-122.91353	Upper Table Rock 2	Mixed
42.46530	-122.89701	Upper Table Rock 3	Mixed
42.46765	-122.91064	Upper Table Rock 4	Diploid
42.46616	-122.91705	Upper Table Rock West	Mixed
42.13963	-122.59339	Songer Wayside	Tetraploid

Table 3. Analysis of variable contributions produced by Maxent models for tetraploids and diploids. Values are averages over 100 replicate bootstrapped runs.

Tetraploid		Diploid
Percent Contribution	Variable	Percent Contribution
62.4	bio5	1.7
28.2	bio1	34.1
2.8	soil pH	0.5
2.4	bio14	1.4
2.4	percent sand	4.2
0.8	bio3	0.1
0.7	bio11	57
0.2	bio6	1

Table 4. Mean values and standard deviation of environmental variables for each ploidy and reported F values and probabilities from ANOVA's comparing the variance of an environmental variable amongst the ploidies. Significant results are designated by an asterisk.

Ploidy	Soil pH	Sand (%)	bio6 Min Temperature Coldest Month (°C)	bio5 Max Temperature Warmest Month (°C)
Diploid	6.05 ± 0.4	45.71 ± 9.87	-4.54 ± 0.1	26.95 ± 0.51
Tetraploid	6.42 ± 0.52	31.72 ± 8.82	-4.84 ± 0.39	27.4 ± 1.12
Mixed	6.87 ± 0.4	25.73 ± 14.78	-4.73 ± 0.08	27.67 ± 0.25
F value	3.68	6.013	2.511	1.013
Pr (> F)	0.047*	0.0106*	0.111	0.384

Ploidy	bio3 Isothermality (%)	bio14 Precipitation of Driest Month (mm)	bio11 Mean Temperature of Coldest Quarter (°C)	bio1 Annual Mean Temperature (°C)
Diploid	46.31 ± 0.62	9.88 ± 0.72	4.99 ± 0.07	11.98 ± 0.29
Tetraploid	46.06 ± 0.78	9.84 ± 0.93	4.85 ± 0.23	12.08 ± 0.6
Mixed	45.66 ± 0.25	9.77 ± 0.62	4.97 ± 0.05	12.28 ± 0.15
F value	1.046	0.022	1.765	0.469
Pr (> F)	0.373	0.978	0.201	0.634

Table 5. Model selection using 10,000 bootstrap iterations for single and multi-surface models including the number of parameters defined in a model (k), the Akaike information criterion for small sample sizes (AICc), the average Akaike weight (Weight), the average rank (Rank), marginal R-squared (R^{2m}), and the percentage of iterations in which a surface model was identified as a top model during bootstrapping (Top Model %).

Surface	k	AICc	Weight	Rank	R^{2m}	Top Model (%)
Combined	22	-219.217	1	1	0.0461	100
Distance	2	-90.573	1.35E-28	2.515	0.0281	0
Rivers	3	-90.053	9.69E-29	3.715	0.0334	0
Development	3	-89.549	8.33E-29	4.657	0.0396	0
Roads	3	-89.39	7.45E-29	5.223	0.0281	0
Agriculture	4	-88.187	2.69E-28	6.733	0.119	0
Habitat	4	-88.468	4.65E-29	7.123	0.0535	0
Elevation	4	-88.418	4.36E-29	7.308	0.054	0
Canopy	4	-88.018	4.38E-29	7.39	0.0866	0
Ploidy	4	-86.276	2.69E-29	9.331	0.0218	0

Table 6. The relative contributions (%) of each landscape layer (Feature) in the combined multi-surface model.

Feature	Contribution (%)
Agriculture	4.92
Canopy	75.33
Development	0.22
Elevation	18.17
Rivers	0.6
Roads	0.22
Habitat	0.21
Ploidy	0.33

Figures

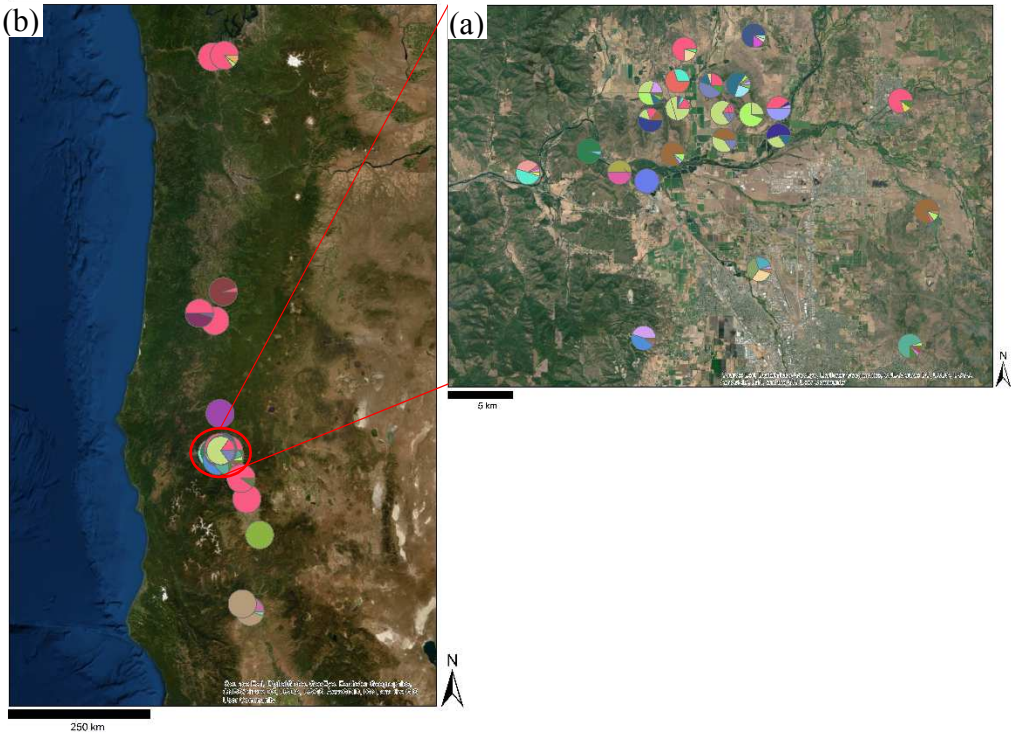

Figure 1. Haplotype diversity and distribution in the polyploid contact zone nearby Medford, Oregon (a). Pies represent haplotype frequencies at each population. Each color represents a haplotype, color scheme is maintained in both panels. Chloroplast haplotype distribution across the ranges of *E. lanatum* var. achillaeoides and leucophyllum (b).

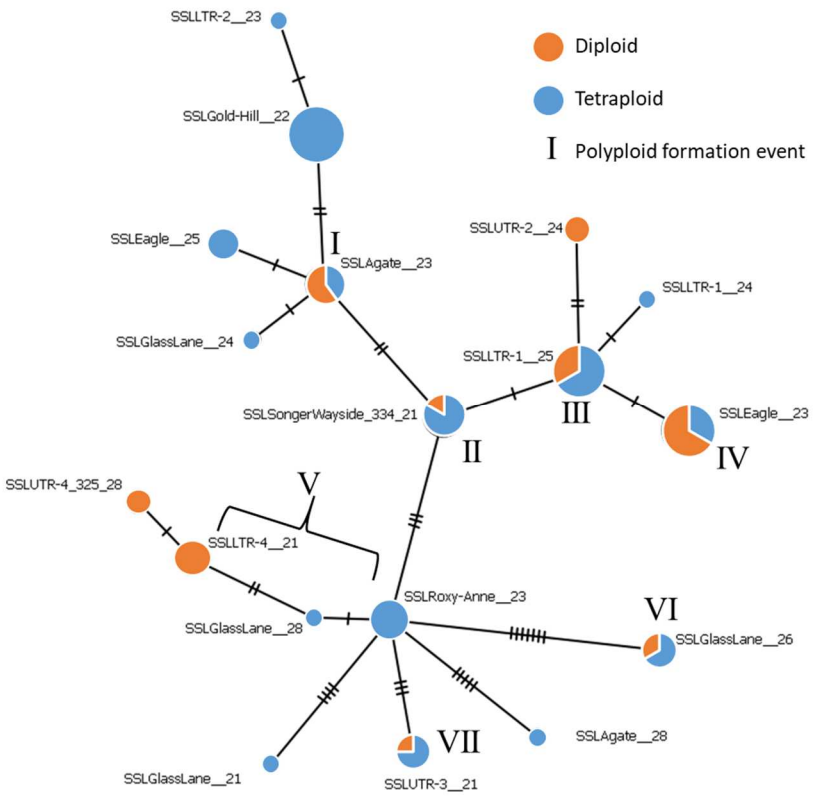

Figure 2. TCS haplotype network generated in PopArt based on 68 cytotyped individuals. Each circle represents an individual haplotype for a total of 17. The size of a circle reflects the number of individuals assigned to that haplotype and each dashed line represents 1 SNP.

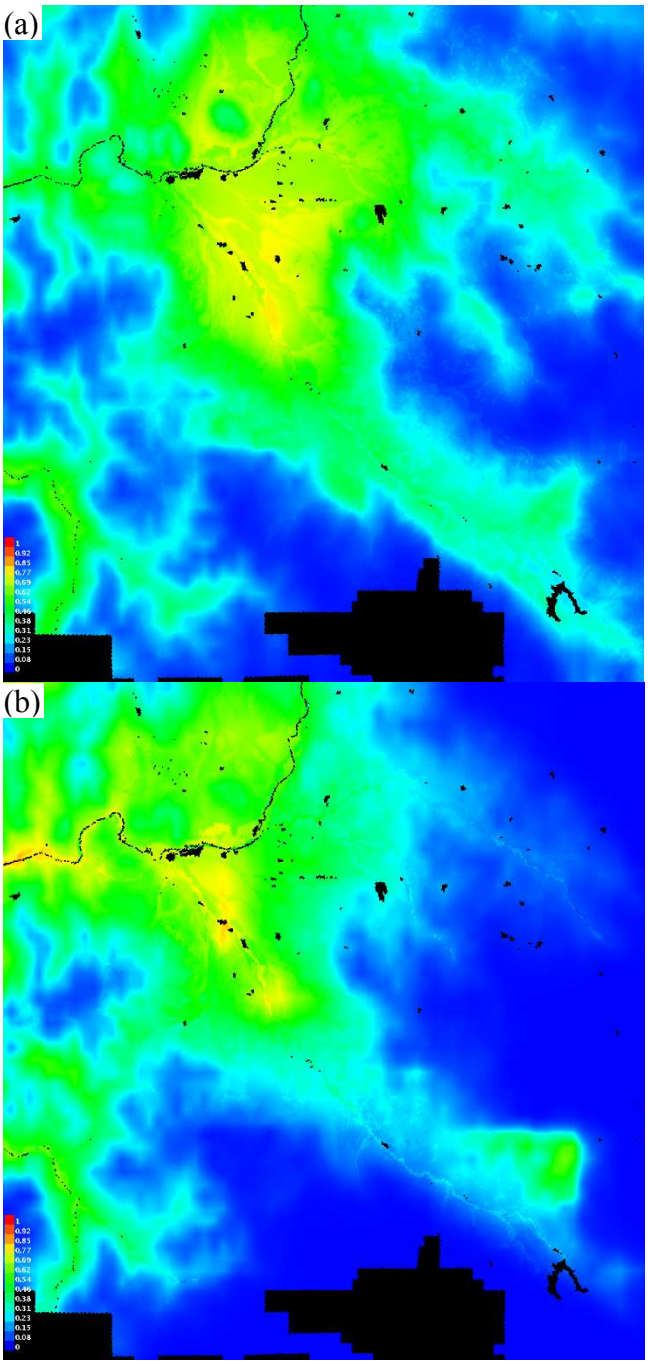

Figure 3. Tetraploid ecological niche model produced using Maxent. Diploid ecological niche model constructed using maxent (b) Warmer colors indicate suitable habitat and cooler colors indicate less suitable habitat. Black areas represent missing data.

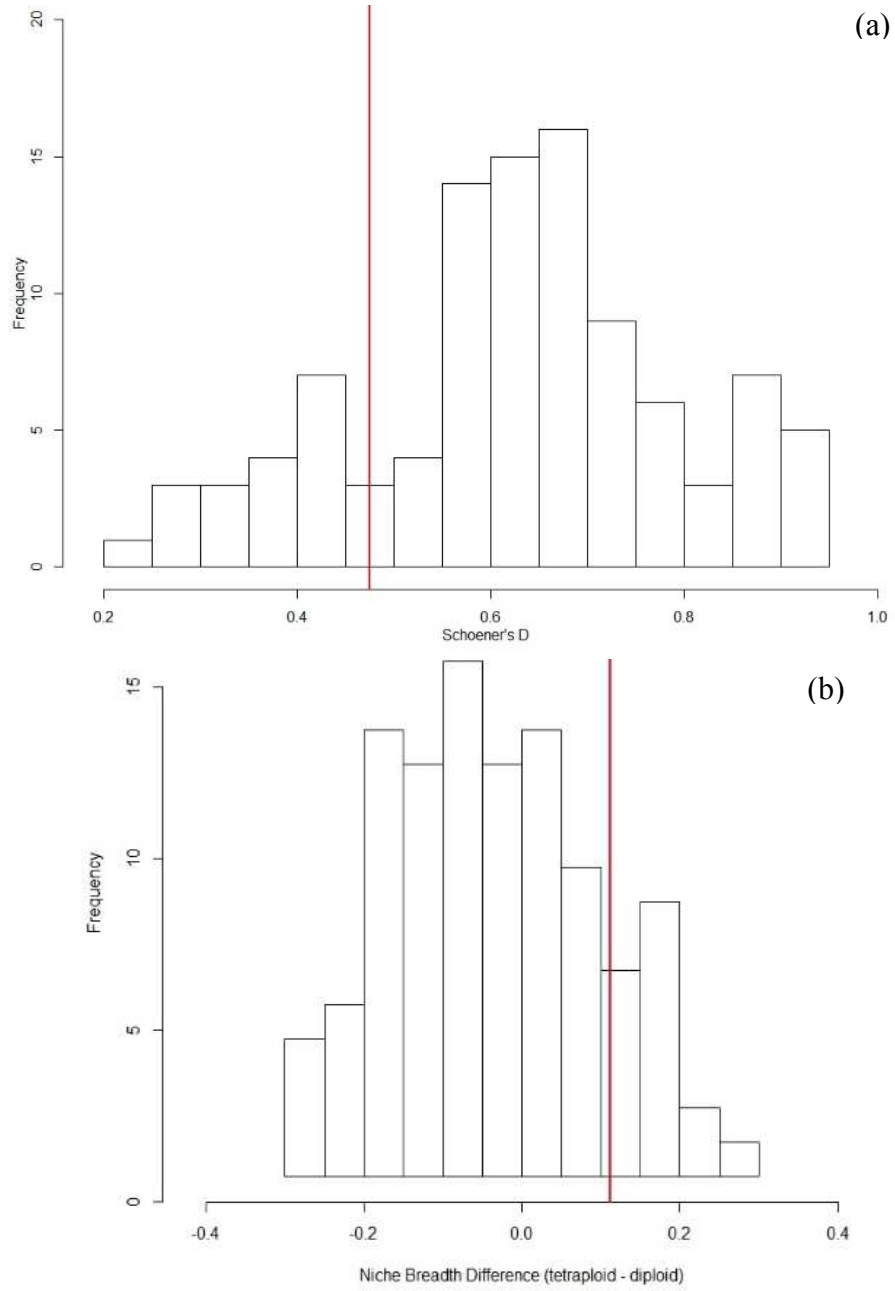

Figure 4. The result of a niche identity test (Warren 2008): a comparison of the observed niche overlap (red line, Schoener's D) to a null distribution of overlap scores generated by 100 psuedoreplicates. Observed niche overlap = 0.47, $p > 0.18$ (a). A comparison of the observed difference in niche breadth (Levins 1968) to a null distribution generated by 100 psuedoreplicates. Observed niche breadth difference = 0.11, $p > 0.17$ (b).

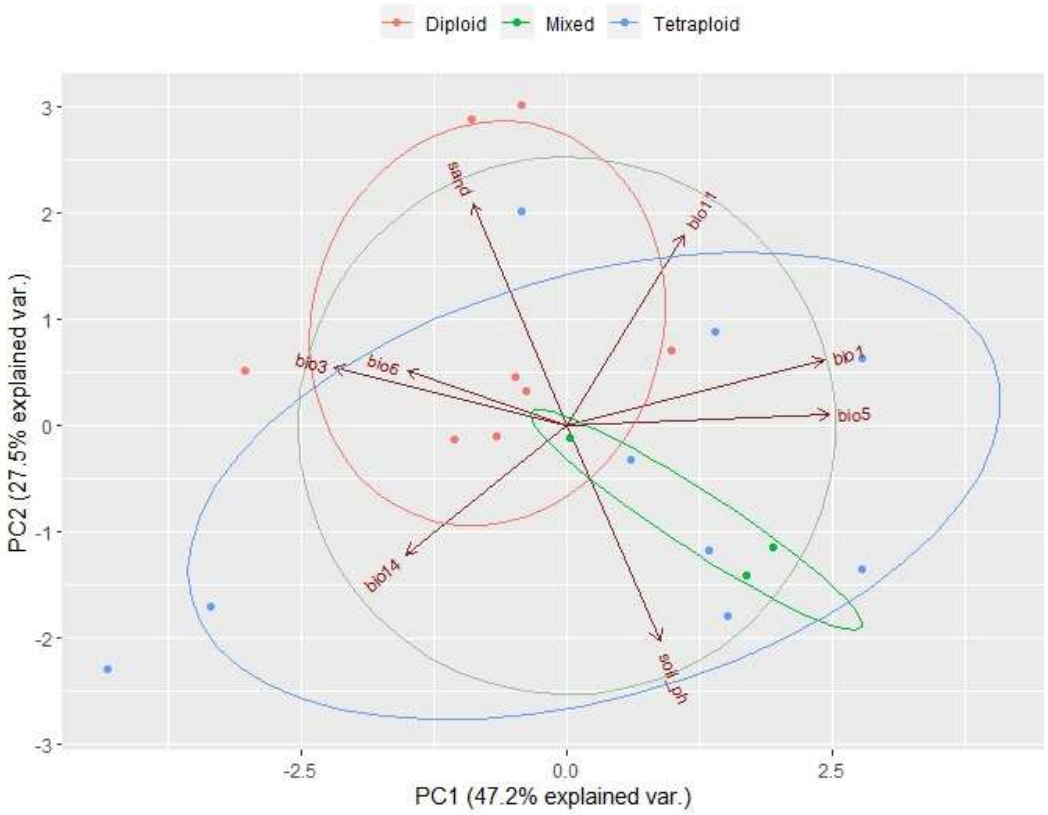

Figure 5. Diploid, Mixed, and Tetraploid populations plotted against PC1 (bio5 - Max temperature of Warmest Month, bio3 - Isothermality, bio1 - Mean Annual Temperature) and PC2 (Soil pH, Sand, bio11 -Mean Temperature of Coldest Quarter). Remaining variables include bio6 – Min Temperature of Coldest Month – and bio14 – Precipitation of Driest Month.

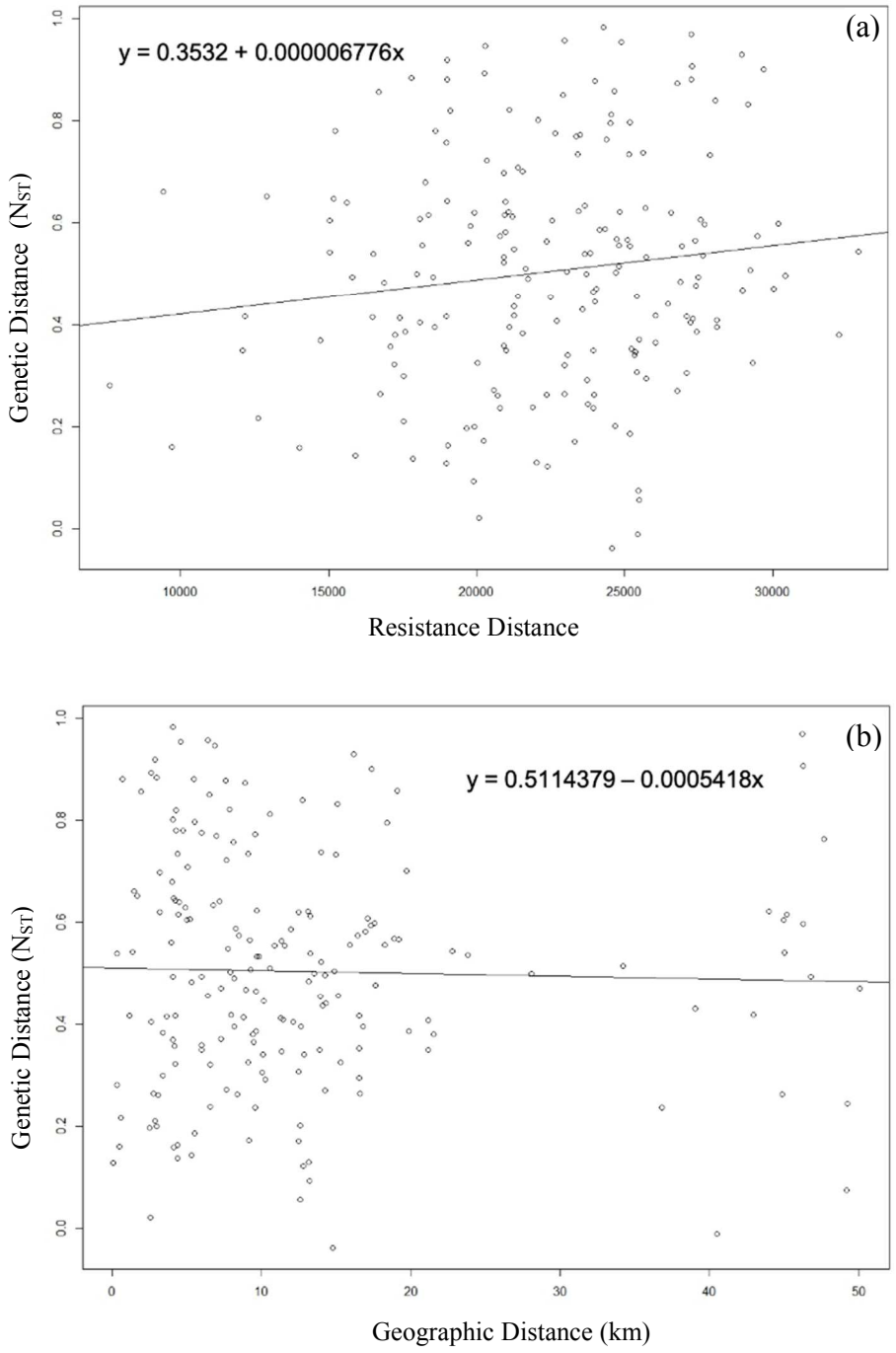

Figure 6. The relationship between resistance distance and genetic distance (N_{ST}). Fitted line and equation represent a linear regression (a). The relationship between geographic distance and genetic distance (N_{ST}). Fitted line and equation represent a linear regression (b).

References

Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, *8*(2), 135–141. https://doi.org/10.1016/j.pbi.2005.01.001

Albert, A., Auffret, A. G., Cosyns, E., Cousins, S. A., D'hondt, B., Eichberg, C., ... & Malo, J. E. (2015). Seed dispersal by ungulates as an ecological filter: a trait-based meta-analysis. *Oikos*, *124*(9), 1109-1120.

Alix, K., Gérard, P. R., Schwarzacher, T., & Heslop-Harrison, J. S. P. (2017). Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Annals of Botany*, *120*(2), 183–194. https://doi.org/10.1093/aob/mcx079

Ansong, M., & Pickering, C. (2013). Are weeds hitchhiking a ride on your car? A systematic review of seed dispersal on cars. *PLoS One*, *8*(11).

Arredondo, T. M., G. L. Marchini, and M. B. Cruzan. 2018. Evidence for human-mediated range expansion and gene flow in an invasive grass. Proceedings of the Royal Society B: Biological Sciences 285:20181125.

Baack, E. J. 2004. Cytotype segregation at regional and microgeographic scales. Am. J. Bot 91:1783-1788

Baack, E. J. 2005. Ecological factors influencing tetraploid establishment in snow buttercups (Ranunculus adoneus, Ranunculaceae): Minority cytotype exclusion and barriers to triploid formation. American Journal of Botany 92: 1827–1835

Baack, E. J. 2005. To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. Heredity 94, 538–546

Baack, E. J., & Stanton, M. L. (2005). Ecological factors influencing tetraploid speciation in snow buttercups (Ranunculus adoneus): niche differentiation and tetraploid establishment. *Evolution*, *59*(9), 1936-1944.

Baltzinger, C., Karimi, S., & Shukla, U. (2019). Plants on the move: hitch-hiking on ungulates distributes diaspores across landscapes. *Frontiers in Ecology and Evolution*, *7*, 38.

Barringer, B. C., & Galloway, L. F. (2017). The reproductive ecology of diploid and tetraploid Galax urceolata. *The American Midland Naturalist*, *177*(2), 299-308.

Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., & Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*, *210*(2), 391-398.

Beck, J. B., Allison, J. R., Pryer, K. M., & Windham, M. D. (2012). Identifying multiple origins of polyploid taxa: A multilocus study of the hybrid cloak fern (Astrolepis integerrima; Pteridaceae). *American Journal of Botany*, *99*(11), 1857-1865.

Ćertner, M., Fenclova, E., Kúr, P., Kolář, F., Koutecký, P., Krahulcová, A., & Suda, J. (2017). Evolutionary dynamics of mixed-ploidy populations in an annual herb: Dispersal, local persistence and recurrent origins of polyploids. *Annals of Botany*, *120*(2), 303–315. https://doi.org/10.1093/aob/mcx032

Constance, L. 1937. A systematic study of the genus Eriophyllum Lag. Univ. Calif. Publ. Bot. 18: 69- 136.

Cronquist, A. (1955). Phylogeny and taxonomy of the Compositae. *American Midland Naturalist*, 478-511.

Cruzan, M. B., & Templeton, A. R. (2000). Paleoecology and coalescence:

phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology & Evolution*, *15*(12), 491-496.
Cruzan, M. B., & Hendrickson, E. C. (2020) Landscape Genetics of Plants -- Challenges and Opportunities. Unpublished, in review.
Doyle, J. J., & Sherman-Broyles, S. (2017). Double trouble: taxonomy and definitions of polyploidy. *New Phytologist*, *213*(2), 487–493. https://doi.org/10.1111/nph.14276
Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379
Fawcett, J. A., Maere, S., & Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences*, *106*(14), 5737–5742. https://doi.org/10.1073/pnas.0900906106
Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086
Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv: 1207:3907 [Preprint].
Gaynor, M. L., D. B. Marchant, D. E. Soltis, and P. S. Soltis. 2018. Climatic niche comparison among ploidal levels in the classic autopolyploid system, Galax urceolata. American Journal of Botany 105(10): 1631–1642
Glennon, K. L., Ritchie, M. E., & Segraves, K. A. (2014). Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters*, *17*(5), 574–582. https://doi.org/10.1111/ele.12259
Glover, N. M., Redestig, H., & Dessimoz, C. (2016). Homoeologs: What Are They and How Do We Infer Them? *Trends in Plant Science*, *21*(7), 609–621. https://doi.org/10.1016/j.tplants.2016.02.005
Grasty, M. R., P. G. Thompson, A. E. Pheil, E. C. Hendrickson, and M. B. Cruzan. 2020. Fine-scale habitat heterogeneity and vole runways influence seed dispersal in Plagiobothrus nothofulvus. Am. J. Bot. 107:413-422.
Haldane, J. B. S. (1959). The Theory of Natural Selection To-Day. *Nature*, *183*(4663), 710–713.
Hardy, O. J., and X. Vekemans. 2002. SPAGEDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618–620.
Heinken, T., & Raudnitschka, D. (2002). Do Wild Ungulates Contribute to the Dispersal of Vascular Plants in Central European Forests by Epizoochory ? A Case Study in NE Germany. Forstwissenschaftliches Centralblatt, 121, 179–194.
Herben, T., Suda, J., & Klimešová, J. (2017). Polyploid species rely on vegetative reproduction more than diploids: A re-examination of the old hypothesis. *Annals of Botany*, *120*(2), 341–349. https://doi.org/10.1093/aob/mcx009
Hewison, A. J., Vincent, J. P., Joachim, J., Angibault, J. M., Cargnelutti, B., & Cibien, C. (2001). The effects of woodland fragmentation and human activity on roe deer

distribution in agricultural landscapes. *Canadian journal of zoology*, *79*(4), 679-689.
Husband, B., & Sabara, H. (2003). review between Reproductive isolation and their diploid autotetraploids Chamerion progenitors in fireweed ,. *New Phytologist*, *161*(3), 703–713. https://doi.org/10.1046/j.1469-8137.2003.00998.x
Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., … dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–100. https://doi.org/10.1038/nature09916
Joshi, N., and J. Fass. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files, version 1.33. Website: github.com/najoshi/sickle
Kennedy, B. F., Sabara, H. A., Haydon, D., & Husband, B. C. (2006). Pollinator-mediated assortative mating in mixed ploidy populations of Chamerion angustifolium (Onagraceae). *Oecologia*, *150*(3), 398-408.
Kliber, A., & Eckert, C. G. (2005). Interaction between founder effect and selection during biological invasion in an aquatic plant. *Evolution*, *59*(9), 1900–1913. https://doi.org/10.1111/j.0014-3820.2005.tb01060.x
Kohrn, B. F., J. M. Persinger, and M. B. Cruzan. 2017. An efficient pipeline to generate data for studies in plastid population genomics and phylogeography. Applications in Plant Sciences 5:1700053.
Levin, D.A. (1975). Minority cytotype exclusion in local plant populations. *Taxon*, 24(1), 35-43
Levins, R. (1968). Evolution in Changing Environments (Monographs in Population Biology). Vol. 2. Princeton University Press, Princeton, New Jersey, USA.
Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997 [Preprint].
Lloyd A, Bomblies K. 2016. Meiosis in autopolyploid and allopolyploid Arabidopsis. Current Opinion in Plant Biology 30: 116–122
Lohaus, R., & Van de Peer, Y. (2016). Of dups and dinos: evolution at the K/Pg boundary. *Current opinion in plant biology*, *30*, 62-69.
Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189-197.
Marrotte, R. R., and J. Bowman. 2017. The relationship between least-cost and resistance distance. *PLOS One* 12: e0174212.
Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12.
Masterson, J. (1994). Stomatal Size in Fossil Plants: Evidence for Polyploidy in Majority of Angiosperms. *Science*, *264*(5157), 421–424. https://doi.org/10.1126/science.264.5157.421
Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., & Otto, S. P. (2011). Recently Formed Polyploid Plants Diversify at Lower Rates. *Science*, *333*(6047), 1257–1257. https://doi.org/10.1126/science.1207205
McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences, 104(50), 19885–19890. http://doi.org/10.1073/pnas.0706568104

Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058-1069.

Mooring, J. S. (1975). A Cytogeographic Study of Eriophyllum lanatum (Compositae , Helenieae). *American Journal of Botany*, *62*(10), 1027–1037.

Mooring, J. S. (2001). Barriers to interbreeding in the Eriophyllum lanatum (Asteraceae, Helenieae) species complex. *American Journal of Botany*, *88*(2), 285–312.

Mooring, J. S. (2008). An Eriophyllum Lanatum (Asteraceae) Hybrid Zone in Oregon. *MADRONO*, *55*(4), 269–279.

Muñoz-Pajares, A. J., Perfectti, F., Loureiro, J., Abdelaziz, M., Biella, P., Castro, M., … Gómez, J. M. (2017). Niche differences may explain the geographic distribution of cytotypes in Erysimum mediohispanicum. *Plant Biology*, (Stebbins 1971), 1–9. https://doi.org/10.1111/plb.12605

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198-1205.

O'Donnell, M.S., and Ignizio, D.A., 2012, Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p.

Oswald, B. P., & Nuismer, S. L. (2011). A unified model of autopolyploid establishment and evolution. The American Naturalist, 178(6), 687-700.

Peterman, W. E. 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. Methods in Ecology and Evolution 9:1638-1647.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. *Ecography*, *40*(7), 887–893. https://doi.org/10.1111/ecog.03049

Ramsey, J., & Schemske, D. W. (1998). Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. *Annual Review of Ecology and Systematics*, *29*(1), 467–501. https://doi.org/10.1146/annurev.ecolsys.29.1.467

Ramsey, J., & Schemske, D. W. (2002). Neopolyploidy in flowering plants. *Annual review of ecology and systematics*, *33*(1), 589-639.

Rey, P. J., Manzaneda, A. J., & Alcántara, J. M. (2017). The interplay between aridity and competition determines colonization ability, exclusion and ecological segregation in the heteroploid Brachypodium distachyon species complex. *New Phytologist*, *215*(1), 85–96. https://doi.org/10.1111/nph.14574

Row, J. R., S. T. Knick, S. J. Oyler-McCance, S. C. Lougheed, and B. C. Fedy. 2017. Developing approaches for linear mixed modeling in landscape genetics through landscape-directed dispersal simulations. Ecology and Evolution 7: 3751–3761.

Sabara, H. A., Kron, P., & Husband, B. C. (2013). Cytotype coexistence leads to triploid hybrid production in a diploid–tetraploid contact zone of Chamerion angustifolium (Onagraceae). *American Journal of Botany*, *100*(5), 962-970.

Schoener, T. W. (1968). The Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology*, *49*(4), 704-726.

Segraves, K. A. (2017). The effects of genome duplications in a community context. *New Phytologist*, *215*(1), 57-69.
Segraves, K. A., Thompson, J. N., Soltis, P. S., & Soltis, D. E. (1999). Multiple origins of polyploidy and the geographic structure of Heuchera grossulariifolia. *Molecular Ecology*, *8*(2), 253-262.
Servick, S., Visger, C. J., Gitzendanner, M. A., Soltis, P. S., & Soltis, D. E. (2015). Population genetic variation, geographic structure, and multiple origins of autopolyploidy in Galax urceolata. *American Journal of Botany*, *102*(6), 973-982.
Šingliarová, B., Zozomová-Lihová, J., & Mráz, P. (2019). Polytopic origin and scale-dependent spatial segregation of cytotypes in primary diploid–autopolyploid contact zones of Pilosella rhodopea (Asteraceae). *Biological Journal of the Linnean Society*, *126*(2), 360-379.
Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., … Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, *96*(1), 336–348. https://doi.org/10.3732/ajb.0800079
Soltis, D. E., Buggs, R. J., Doyle, J. J., & Soltis, P. S. (2010). What we still don't know about polyploidy. *Taxon*, *59*(5), 1387-1403.
Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS. 2007. Autopolyploidy in angiosperms: have we grossly underestimatedthe number of species? Taxon 56: 13–30.
Soltis, D. E., & Soltis, P. S. (1999). Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution*, *14*(9), 348–352. https://doi.org/10.1002/bies.201300096
Soltis, D. E., Visger, C. J., Blaine Marchant, D., & Soltis, P. S. (2016). Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany*, *103*(7), 1146–1166. https://doi.org/10.3732/ajb.1500501
Spoelhof, J. P., Soltis, P. S., & Soltis, D. E. (2017). Pure polyploidy: closing the gaps in autopolyploid research. *Journal of Systematics and Evolution*, *55*(4), 340-352.
Stebbins, G. L. 1950. Variation and evolution in plants. Oxford University Press, London, UK.
Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., … Harmon, L. J. (2015). Nested radiations and the pulse of angiosperm diversification: Increased diversification rates often follow whole genome duplications. *New Phytologist*, *207*(2), 454–467. https://doi.org/10.1111/nph.13491
Těšitelová, T., Jersáková, J., Roy, M., Kubátová, B., Těšitel, J., Urfus, T., ... & Suda, J. (2013). Ploidy-specific symbiotic interactions: divergence of mycorrhizal fungi between cytotypes of the G ymnadenia conopsea group (Orchidaceae). New Phytologist, 199(4), 1022-1033.
Thompson, K. A., Husband, B. C., & Maherali, H. (2015). No influence of water limitation on the outcome of competition between diploid and tetraploid Chamerion angustifolium (Onagraceae). *Journal of Ecology*, *103*(3), 733–741. https://doi.org/10.1111/1365-2745.12384
Thompson, J. N., Nuismer, S. L., & Merg, K. (2004). Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biological Journal of the Linnean*

Society, *82*(4), 511-519.
Trávníček, P., Dočkalová, Z., Rosenbaumová, R., Kubátová, B., Szeląg, Z., & Chrtek, J. (2011). Bridging global and microregional scales: Ploidy distribution in Pilosella echioides (Asteraceae) in central Europe. *Annals of Botany*, *107*(3), 443–454. https://doi.org/10.1093/aob/mcq260
Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, *18*(7), 411–424. https://doi.org/10.1038/nrg.2017.26
van Etten, J. 2017. R package gdistance: Distances and routes on geographical grids. Journal of Statistical Software 76: 1–21.
Vanneste, K., Maere, S., & Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1648), 20130353. https://doi.org/10.1098/rstb.2013.0353
Visger, C. J., Germain-Aubrey, C. C., Patel, M., Sessa, E. B., Soltis, P. S., & Soltis, D. E. (2016). Niche divergence between diploid and autotetraploid Tolmiea. *American Journal of Botany*, *103*(8), 1396-1406.
Windham, M. D., Pryer, K. M., Poindexter, D. B., Li, F. W., Rothfels, C. J., & Beck, J. B. (2020). A step-by-step protocol for meiotic chromosome counts in flowering plants: A powerful and economical technique revisited. Applications in Plant Sciences, 8(4).
Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography*, *33*(3), 607–611. https://doi.org/10.1111/j.1600-0587.2009.06142.x
Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological applications*, *21*(2), 335-342
Wright, S. (1943). Isolation by Distance. Genetics, 28(2), 114–138.
Zhang H, Bian Y, Gou X, Zhu B, Xu C, Qi B, Li N, Rustgi S, Zhou H, Han F, Jiang J. 2013. Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proceedings of the National Academy of Sciences USA 110: 3447–3452

Appendix: Images of aceto-carmine chromosome squashes

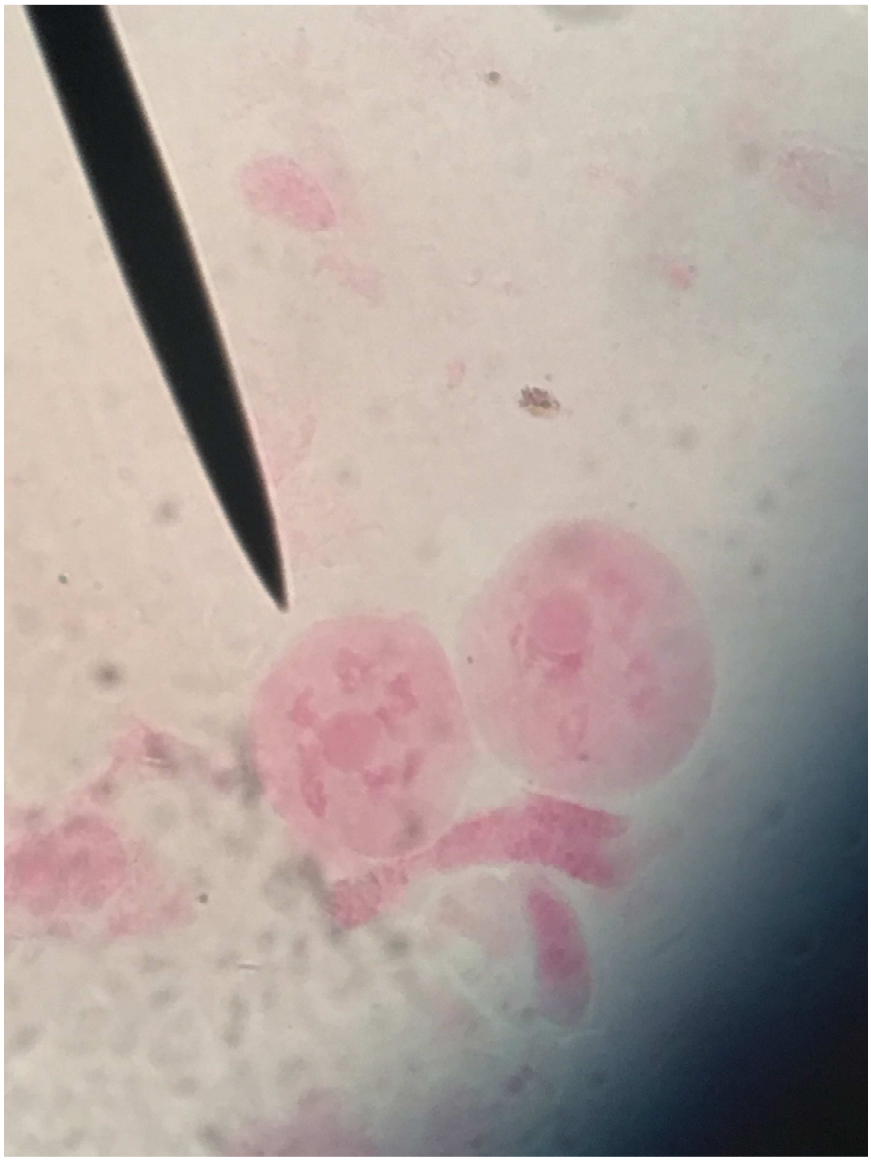

Image 1. Chromosome squash using immature anthers from individuals JP-23 and JP-24 (both identified as diploids by flow cytometry). Cells are stained with acetocarmine. Image taken with an iPhone 6S camera through the ocular lens of a compound microscope. Eight chromosomes were counted.

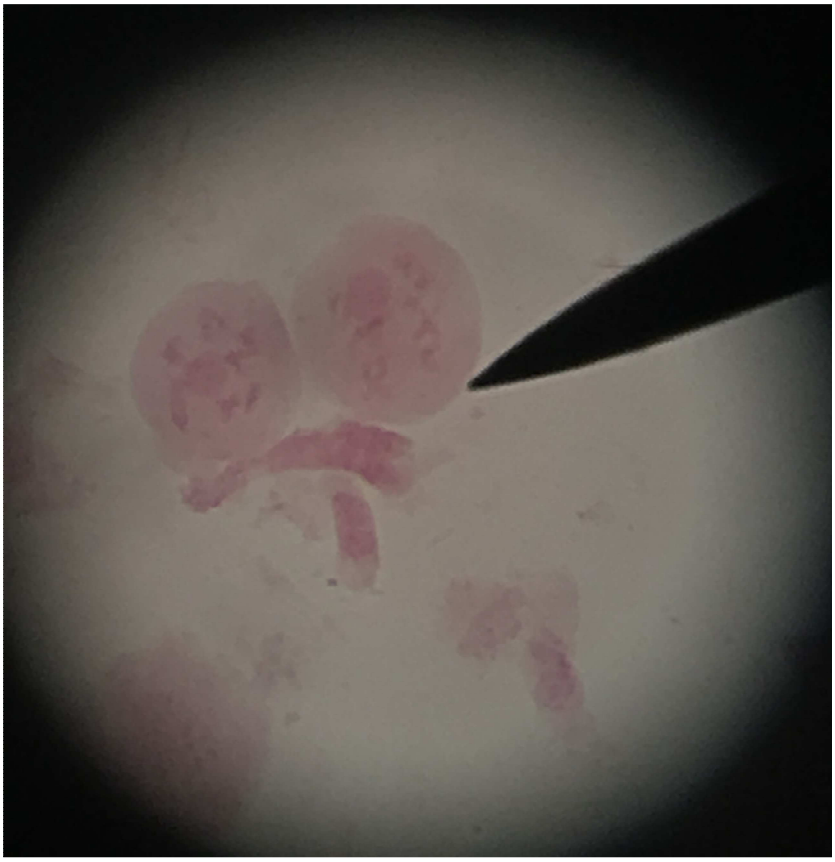

Image 2. Different view of the cells in Image 1.

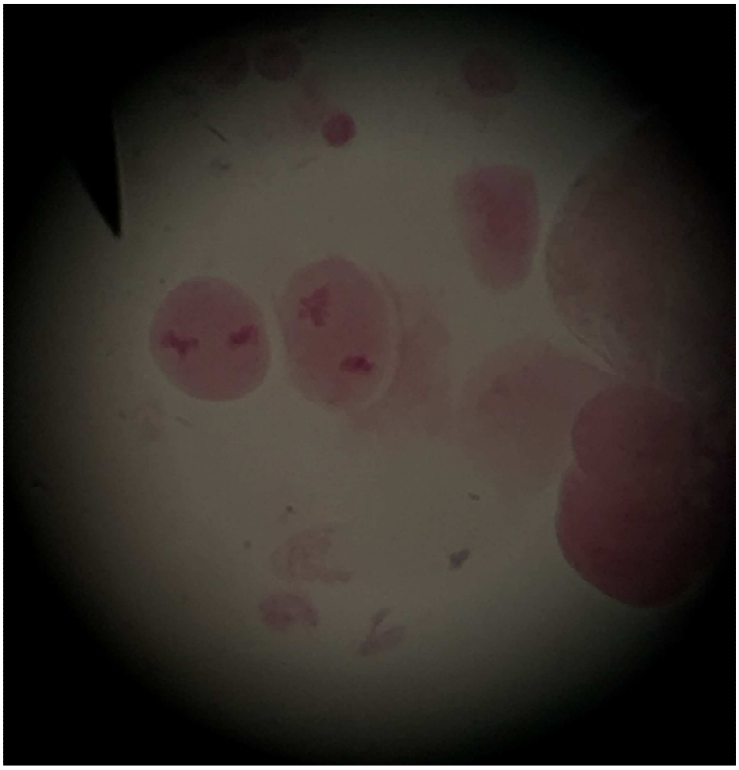

Image 3. Cells at a different cell division stage from the same chromosome squash using immature anthers from individuals JP-23 & JP-24 (Image 1; Image 2).

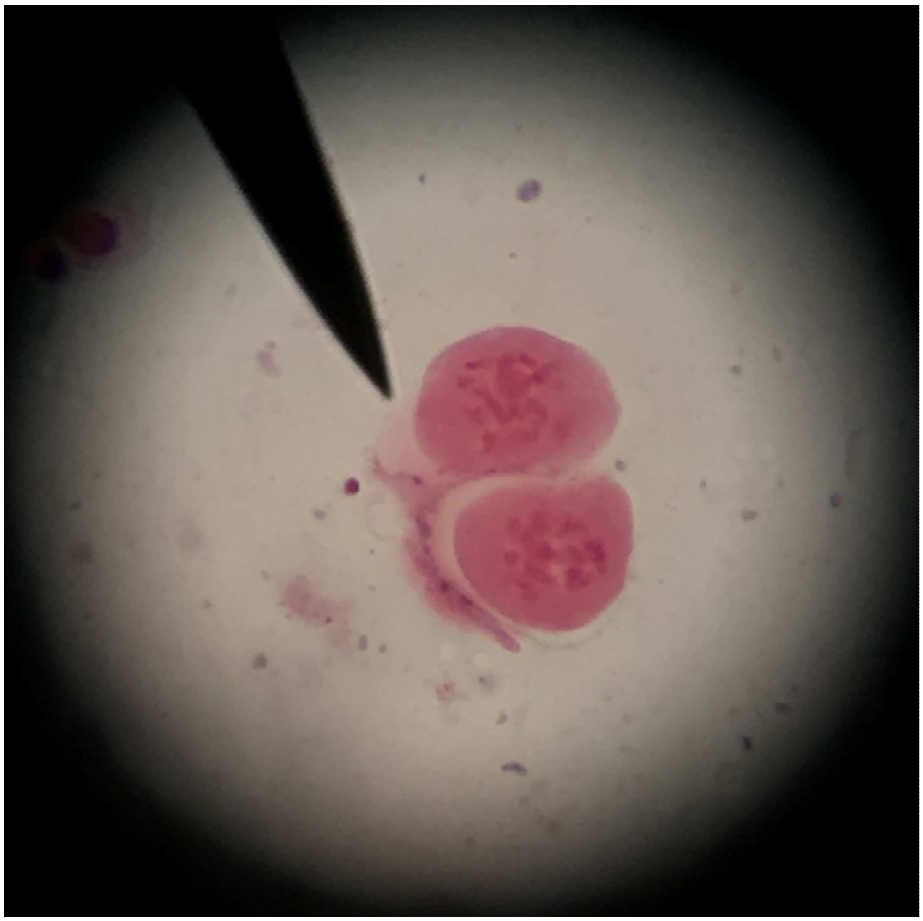

Image 4. Attempted chromosome squash using immature anthers from individuals RA-23, RA-28, RA-29. Cells are stained with acetocarmine. Image taken with an iPhone 6S camera through the ocular lens of a compound microscope.