5-2024

Environmental Factors Associated With Triploid Aspen Occurrence in Intermountain West Landscapes

James A. Walton
Utah State University, jim.walton@usu.edu

ENVIRONMENTAL FACTORS ASSOCIATED WITH TRIPLOID ASPEN OCCURRENCE IN INTERMOUNTAIN WEST LANDSCAPES

by

James A. Walton

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

of

MASTER OF SCIENCE

in

Ecology

Approved:

_____	_____
Karen E. Mock, Ph.D. Major Professor	R. Douglas Ramsey, Ph.D. Committee Member
_____	_____
R. Justin DeRose, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2024

ABSTRACT

Environmental Factors Associated with Triploid Aspen Occurrence

in Intermountain West Landscapes

by

James Walton, Master of Science

Utah State University, 2024

Major Professor: Dr. Karen Mock
Department: Wildland Resources

Polyploidy is common among plants and can contribute to physiological and morphological differences, altering how plants respond to environmental changes, promoting genetic diversification and even species radiation. Quaking aspen (*Populus tremuloides*), a keystone species associated with high plant and animal diversity is frequently found in mixed diploid/triploid populations in the Intermountain West. High mortality rates and widespread population declines in aspen are of increasing concern in the Intermountain West, and often ascribed to changing climates and drought stress events. The goal of this study was to better understand environmental factors influencing the distribution of triploid aspen population in the Intermountain West. Using restriction-site associated DNA sequencing (RAD-seq), we examined the occurrence of diploid and triploid aspen populations at various spatial scales in relation to environmental variables associated with soil moisture content. Our results suggest that triploidy in aspen on the landscape is associated with environmental variables related to soil moisture and may be influenced by specific local variations in topography, climate, and precipitation patterns.

Overall, we found that conditions associated with low soil moisture were associated with lower frequencies of triploid aspen. Our results suggest that triploid aspen clones may be more susceptible to mortality than diploid clones in the warmer, drier climates expected with climate change. We acknowledge, however, that there are many knowledge gaps regarding the generation and persistence of triploid aspen clones. Understanding the causes and consequences of triploidy in aspen will be important in predicting and managing aspen persistence in landscapes of the western U. S. under changing climate conditions.

(93 pages)

PUBLIC ABSTRACT

Environmental Factors Associated with Triploid Aspen Occurrence in Intermountain West Landscapes

James A. Walton

Polyploidy is common among plants and can contribute to physiological and morphological differences, altering how plants respond to environmental changes, promoting genetic diversification and even species radiation. Quaking aspen (*Populus tremuloides*), a keystone species associated with high plant and animal diversity is frequently found in mixed diploid/triploid populations in the Intermountain West. Triploid aspen carries an extra chromosomal copy, whereas the diploid type contains two chromosomal copies. High mortality rates and widespread population declines in aspen are of increasing concern in the Intermountain West, and often ascribed to changing climates and drought stress events. The goal of this study was to better understand environmental factors influencing the distribution of triploid aspen population in the Intermountain West. Using restriction-site associated DNA sequencing (RAD-seq), a method used to identify thousands of genetic markers from a group of individuals, we examined the occurrence of diploid and triploid aspen populations at various spatial scales in relation to environmental variables associated with soil moisture content. Our results suggest that triploidy in aspen on the landscape is associated with environmental variables related to soil moisture and may be influenced by specific local variations in topography, climate, and precipitation patterns. Overall, we found that conditions associated with low soil moisture were associated with lower frequencies of triploid

aspen. Our results suggest that triploid aspen clones may be more susceptible to mortality than diploid clones in the warmer, drier climates expected with climate change. We acknowledge, however, that there are many knowledge gaps regarding the generation and persistence of triploid aspen clones. Understanding the causes and consequences of triploidy in aspen will be important in predicting and managing aspen persistence in landscapes of the western U. S. under changing climate conditions.

ACKNOWLEDGMENTS

Without the incredible influence of Karen Mock, who introduced me to genetics more than a decade ago, this thesis wouldn't exist. Your mentorship, support, and friendship have gone far beyond the work of this thesis. You have always believed in my potential and your academic and personal influence has been deeply appreciated. Thank you for plucking me out of rabbit holes and for your patience while I explored them.

I would also like to sincerely thank my committee members, Doug Ramsey and Justin DeRose, for their insights, support, and guidance throughout this research. This project may not have been possible without the help of my colleagues in the Molecular Ecology Lab; thank you for your support.

I express my deepest gratitude to my incredible wife, Shay, and our two beautiful children, Hattie and Maple, for their unwavering support throughout this journey. Your love, patience, and understanding have been my pillars of strength during the challenges of balancing work and academics. Shay, your encouragement has been my driving force, and I am immensely grateful for your sacrifices and belief in me. Hattie and Maple, your cheerful presence and understanding when I had to prioritize my studies have been truly remarkable. I couldn't have asked for a more supportive and loving foundation to pursue my academic goals.

Jim Walton

CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

The location of species or individuals on the landscape can provide valuable insight into biogeographic history, habitat limitations, and persistence patterns. With changing climates, it is increasingly important to understand the drivers of spatial distributions, which have implications for predicting future occupancy and establishing effective management practices. Discrimination of topographic, edaphic, and climatic drivers associated with current species distributions can provide important clues about mechanisms underlying persistence and mortality. Further, intraspecific variation, including genetic and phenotypic diversity within a species is under-evaluated and often overlooked (Des Roches et al., 2021), despite its influence on an organism's distribution, movements, and survival on a landscape.

Genotypic variation can be an important factor in species distribution and intraspecific diversity affecting species presence on a landscape. One form of genotypic variation is polyploidy, or the presence of more than two sets of chromosomes. Polyploidy is common in plants and has long been recognized as an important factor contributing to physiological characteristics, promotion of species radiations, rapid plant evolution, and provide the basis for genetic diversification (Jiao et al., 2011; Levin, 1983; Otto & Whitton, 2000; Parisod et al., 2010; Ramsey & Schemske, 1998; Soltis et al., 2009; Tank et al., 2015; Van de Peer et al., 2017).

In polyploid plants, the increased genome size can contribute to larger cell sizes as well as larger and less dense stomata (Beaulieu et al., 2008; Hodgson et al., 2010). In some species whole genome duplication in polyploids may allow for occupation of new

habitats, including colonization of post-glacial regions under fluctuating climates (Dynesius & Jansson, 2000) or the ability to outcompete diploids (Edgeloe et al., 2022). Species with varying levels of intraspecific cytotypes (ploidy levels) often occupy distinct geographical and ecological spaces (McIntyre, 2012; Otto & Whitton, 2000). Triploids, which can result from unreduced gametes produced by a diploid parent, are expected to have reduced fertility due to problems with meiosis and gamete production (Comai, 2005; Van de Peer et al., 2017), compared to their diploid counterparts. There is some evidence that the production of unreduced gametes, leading to more triploid progeny, can be stress induced (Heilborn, 1934; Mason et al., 2011; Wang et al., 2016).

The distribution of foundational species is especially important as they can provide habitat for a range of other species (Ellison et al., 2005; Mills et al., 2000) and contribute disproportionately to biodiversity (Altieri et al., 2007; Angelini et al., 2011). Quaking aspen (*Populus tremuloides*) is one such foundational species that is also the most broadly distributed native tree species in North America (Little, 1971). Once established, aspen stands provide ecosystems that exist in a wide range of soil, climate, and disturbance conditions (Ally et al., 2010; Landhäusser et al., 2019; Stevens-Rumann et al., 2018). Attributes such as high seed mobility, the ability to spread clonally, broad distribution, and stand isolation (D. Chong et al., 1994; Mock et al., 2008) contribute to high genetic diversity at multiple scales in this species (Mitton & Grant, 1996). Additionally, aspen provides cultural value as part of myths and folklore, industrial value for wood products, and recreational value (G. W. Chong et al., 2001; David et al., 2001; Fechner & Barrows, 1976; Fife, 1994; Mueggler, 1985).

In the past century, fire suppression has enabled conifer establishment to displace aspen in many Intermountain West landscapes (Bartos & Campbell, 1998; Jones et al., 2005). In more recent decades, high mortality rates have been observed in western aspen due to xylem cavitation following drought (Anderegg et al., 2013; Dudley et al., 2015; Love et al., 2019). With current climate change scenarios, this trend is expected to continue (Rehfeldt et al., 2009; Worrall et al., 2013). Rapidly changing climates create an urgent need to mitigate the widespread rapid dieback of aspen in the western U.S., given its importance in biodiversity and wildlife habitat as well as its societal value.

Disturbance is vital to the survival and regeneration of aspen. Aspen is well-known as a fire-adapted species, producing a vigorous suckering response and seedling recruitment in post-fire and post-disturbance environments (Krasnow & Stephens, 2015; Kreider & Yocom, 2021). Seed-based reproduction is much less common than suckering and is believed to be restricted in some landscapes due to stringent germination requirements (Latva-Karjanmaa et al., 2003). Seed-based reproduction, however, is advantageous because it increases and/or maintains genetic diversity and allows for adaptation and dispersal that could help track climate changes. Due to its robust response to disturbance, aspen is an exceptionally good candidate for ecological restoration, including reclamation, reforestation, and afforestation following ecological disturbances such as wildfire, insects, disease, or anthropogenic causes such as timber harvest, bark beetle outbreaks, and mining (Krasnow & Stephens, 2015; Landhäusser et al., 2019).

Aspen is known to have diploid and triploid intraspecific cytotypes (2 or 3 sets of chromosomes, respectively) that coexist in many western landscapes (Mock et al., 2012), and cytotype may influence growth rate, physiology, phenology, and morphology (B. W.

Blonder et al., 2023; DeRose et al., 2015; Greer et al., 2018). Triploids are predominantly found at the lower latitudes of the species range (Mock et al., 2012). Reproduction of aspen triploids is primarily from root suckering, as triploids are expected to have reduced fertility (Levin 1983, Otto and Whitton 2000), though viable seedlings from triploid aspen individuals have been observed (Goessen et al., 2022). Existing evidence indicates in dioecious plants environmental stress can induce the production of unreduced gametes, leading to the generation of more triploid progeny (Heilborn, 1934; Mason et al., 2011; Wang et al., 2016). Physiologically, triploid aspen have been shown to have a greater percent nitrogen content, leaf size and mass, greater chlorophyll content, and a faster growth rate than their diploid counterparts (Blonder et al., 2022; Greer et al., 2018). Additionally, despite greater intrinsic water use efficiency, triploid aspen appear to have lower stomatal sensitivity to rising vapor pressure deficit, suggesting that triploid aspen may perform more poorly than diploids in drought conditions (B. Blonder et al., 2020; Dixon & DeWald, 2015; Greer et al., 2018). However, in microcosm studies (Eisenring et al., 2023), found that genet differences explain more phenotypic variation than ploidy level differences in aspen shoots. Further, Triploids in drought-stressed areas have been found to have higher mortality rates (Dixon & DeWald, 2015) than their diploid counterparts. We hypothesize that if triploid aspen is generally more vulnerable to drought cavitation and drought-induced mortality than diploids, then their persistence and distribution on the landscape would be correlated with landscape and climate variables that are associated with higher levels of soil moisture content.

Sexual dimorphism may also play a significant role in shaping the distribution patterns of aspen across the Intermountain West region of North America. Research

indicates that males may be more common in dioecious plant populations due to the reproductive burden borne by females (Barrett & Hough, 2013). Female plants may invest significant energy in seed production, potentially limiting their persistence in environmentally stressful conditions, such as drought (Barrett & Hough, 2013; Dawson & Ehleringer, 1993; Espirito-Santo, 2003; Freeman et al., 1980; Gross & Soule, 1981; C. Li et al., 2007). The landscape-level sex ratio of aspen genets tends to be male-biased, particularly in areas of higher elevation (2:1 to 3:1 as elevation increases)(Bidner, 2021; Mitton & Grant, 1996). This sex-based disparity underscores the intricate interplay between biological factors and environmental conditions shaping aspen forest dynamics, underscoring the necessity to elucidate mechanisms driving these sex-specific distribution patterns.

In this study, we aim to understand how polyploidy and sex affect the distribution of aspen at both the local and regional scales in the Intermountain West. Because soil moisture, at both local and regional scales, is expected to generally have a major influence on aspen distribution, we hypothesize that in landscapes associated with lower soil moisture content, the proportion of triploids will be lower and the proportion of males will be higher, both at local and regional scales. To address these hypotheses, we compare the incidence of diploid and triploid individuals, as well as male and female individuals, with respect to environmental variables that are linked to soil moisture content.

CHAPTER II

METHODS

Sample Collection, and Preservation

In 2016, a set of 32 30 km x 30 km sites (Figure 1, Table 1) were chosen by identifying areas in the Intermountain West which 1) had extensive presence of aspen forest cover type, 2) maximized distribution across latitudinal, longitudinal, and elevational gradients, 3) occurred on U.S. Forest Service land (to facilitate access), and 4) were within ~1 km in width to secondary road corridors in order to optimize sampling efficiency and increase sampling size. Sites ranged from 39.2° to 44.5° latitude, -120.5° to -104.9° longitude, and from 1,100 m to 3,310 m in elevation (Figure 1, Table 1).

Within each site, 21-52 trees were selected for sampling (mean = 47.4/site, n=1,518) (Table 1). Trees ranged in size from 10.2 to 21 cm in diameter at breast height (average = 13.8 cm). An attempt was made to distribute sampled trees along elevational gradients and among “wet” and “dry” areas based subjectively on surrounding plant composition and immediate proximity to surface water. Care was taken to minimize the probability of collection from identical clones by assuring trees were separated by large geographic distance (>~100 m) or were clearly partitioned by non-aspen vegetation. One or more leaves were collected from each tree, placed in a paper envelope, and stored in silica gel desiccant until DNA extraction. Trees were geo-located and identified with a uniquely numbered aluminum tag.

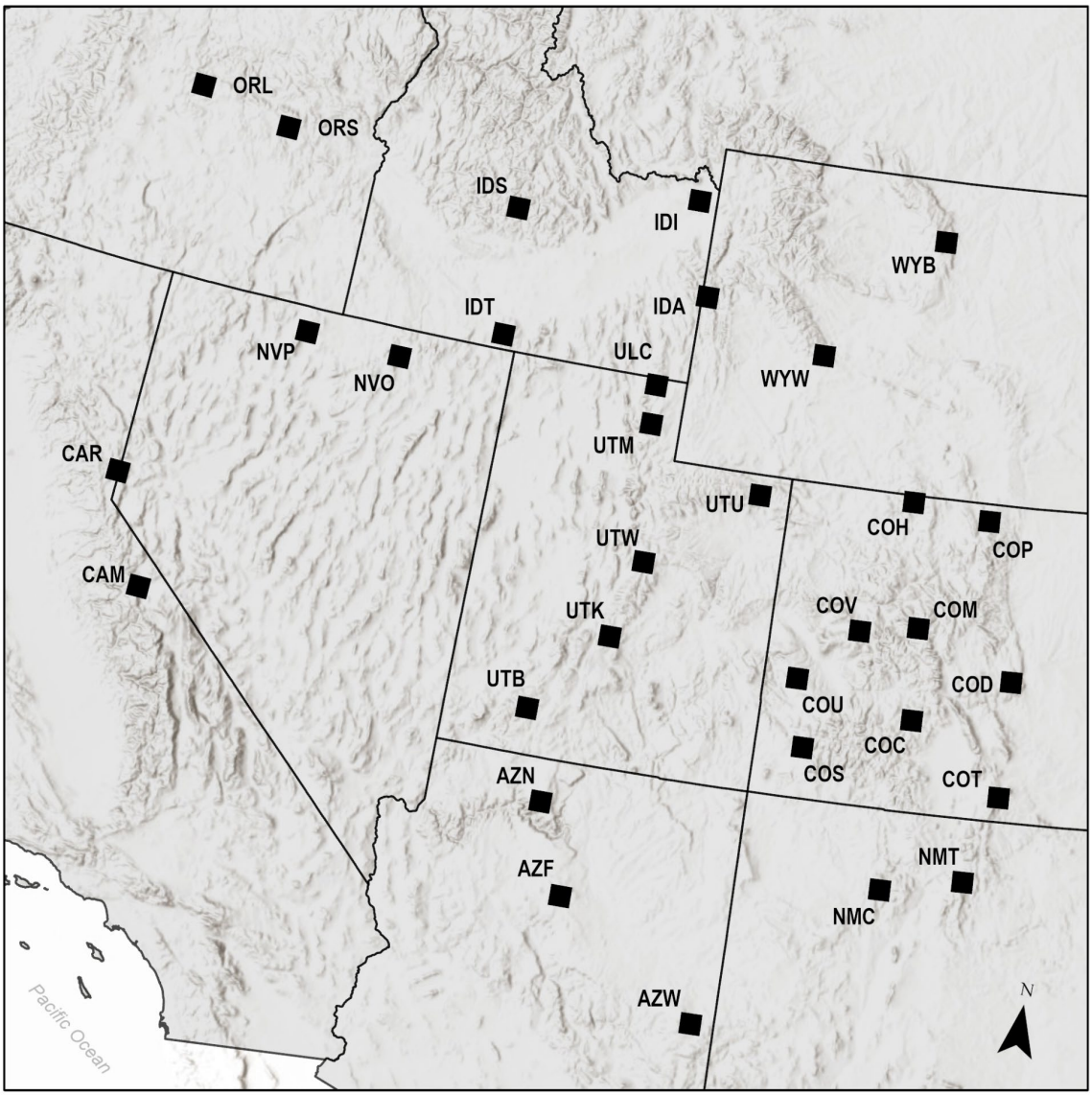

Figure 1. Location of sampling sites from across the Intermountain West (n=32), collected in 2016. Leaves were collected from 21-52 trees ranging from 4-21 cm DBH. Within each site (30 km x 30 km) samples were collected to maximize distribution across latitudinal, longitudinal, and elevational gradients.

Table 1. Table of sampling site characteristics. Site name, number of samples in site (N), elevational range (Elevation Range (m)), percent of diploids and triploids present at site, percent male and female samples at site, and approximate site center (latitude, longitude) for 32 collection sites analyzed as part of this study.

Site	N	Elevation Range (m)	Diploid/Triploid (%)	Male/Female (%)	Site Center (Approx.)
AZF	41	2250-2906	66 / 34	60 / 40	35.312, -111.687
AZN	49	2189-2790	82 / 18	60 / 40	36.449, -112.245
AZW	46	2605-2855	96 / 4	56 / 44	33.975, -109.403
CAM	43	2169-2960	91 / 9	55 / 45	38.017, -119.215
CAR	30	1715-2698	87 / 13	52 / 48	39.370, -120.016
COC	41	2756-3249	68 / 32	66 / 34	38.159, -106.599
COD	21	2083-3077	71 / 29	71 / 29	38.777, -105.067
COH	44	2141-2782	70 / 30	70 / 30	40.911, -106.983
COM	45	2348-2986	67 / 33	55 / 45	39.330, -106.667
COP	45	2261-2985	73 / 27	64 / 36	40.779, -105.694
COS	43	2286-3147	58 / 42	65 / 35	37.646, -108.281
COT	42	2472-3317	86 / 14	71 / 29	37.302, -105.086
COU	37	2492-2929	62 / 38	70 / 30	38.506, -108.513
COV	40	2350-2732	48 / 52	78 / 22	39.206, -107.610
IDA	23	1740-1904	91 / 9	50 / 50	43.115, -110.950
IDI	34	1788-2204	88 / 12	47 / 53	44.298, -111.363
IDS	40	1817-2462	83 / 17	57 / 43	43.787, -114.463
IDT	40	1608-2229	70 / 30	72 / 28	42.183, -114.280
NMC	42	2242-2872	48 / 52	81 / 19	35.976, -106.789
NMT	39	2397-3148	56 / 44	56 / 44	36.190, -105.515
NVO	39	1801-2270	69 / 31	64 / 36	41.632, -115.925
NVP	22	1653-2347	81 / 19	62 / 38	41.675, -117.537
ORL	37	1172-1959	70 / 30	68 / 32	44.368, -120.326
ORS	42	1316-1789	88 / 12	57 / 42	44.127, -118.694
ULC	16	1892-2473	81 / 19	na	41.907, -111.560
UTB	46	2418-3145	59 / 41	61 / 39	37.587, -112.720
UTK	44	2453-3143	34 / 66	82 / 18	38.665, -111.615
UTM	36	1688-2706	67 / 33	69 / 31	41.411, -111.537
UTU	46	2252-3002	41 / 59	64 / 36	40.738, -109.547
UTW	48	2628-3083	46 / 54	69 / 31	39.668, -111.278
WYB	39	1956-2765	92 / 8	46 / 54	44.229, -106.969
WYW	44	2164-2893	70 / 30	62 / 38	42.606, -108.815

In 2018, the collected leaf samples were processed for DNA extraction. LGC BioResearch Technologies (Beverley, MA) conducted the initial DNA extractions using sbeadex™ nucleic acid purification. Samples collected from California populations (CAM and CAR) were extracted in the Molecular Ecology Lab at Utah State University using Qiagen's DNeasy® 96 Plant Kits according to the manufacturer's protocol with a final elution in buffer AE.

Genomic Data Preparation and Analysis

Genomic Library Preparation

Double digest restriction-site associated sequencing (ddRAD-seq) libraries were prepared following the methods outlined in Parchman et al. (2012). Restriction enzymes EcoRI (NEB, R0101L) and MseI (NEB, R0525L) were first used to digest genomic DNA for individual samples resulting in a sticky-end restriction cut site, providing for ligation of customized adaptor sequences. DNA restrictions were carried out in 8.6 µL reactions and contained 10x Cutsmart Buffer (NEB), 0.06 M NaCl, 116.3 U/µL Msel and 581.5 U/µL EcoRl enzymes, and 6 µL DNA template with a target concentration of 15-20 ng/µL. Digestion was then completed at 37 C for 2 hours, followed by enzyme inactivation at 65 C for 20 minutes.

Following restriction, a common double-stranded Msel adaptor (Msel1 & Msel2, Table 2) was ligated to all individuals on the previously cut Msel sticky-end and a unique 8-10 base-pair barcode that contained a 4 base-pair difference between barcodes, was incorporated into on the EcoRl sticky-end using T4 Ligase (NEB, M0202L). Reactions for ligation of adaptors and barcodes contained 1.54 µL total volume and consisted of 0.898 µM each Msel forward and reverse adaptors, 0.004 M NaCl, 6014.363 U/mL T4

DNA Ligase, and 0.289 U T4 Buffer. This mix as well as 1 µL of the EcoRl barcode were added to each digest and ligated at 16 C for 2 hours, followed by enzyme inactivation at 65 C for 20 minutes.

Following restriction and ligation of the individually barcoded samples, Illumina PCR (polymerase chain reaction) primers (Illpcr1F & Illpcr2R, Table 2) were used to amplify fragments containing the ligated adaptors and barcodes. During this step, 2 separate individually barcoded restriction-ligation products were pooled into 2 20 uL PCR reactions. This step was taken to help ameliorate stochastic differences in PCR production and reduce time in the lab and the use of costly Taq polymerase (Parchman et al., 2012). PCR reactions contained .02 U/µL BioRad Iproof High Fidelity DNA Polymerase (Cat. 172-5301), 1x Iproof Buffer (HF), 0.2 mM each dNTP, 1 mM MgCl2, 0.33 µM pre-mixed Illumina primers (Illpcr1F & Illpcr2R; 5 µM stock), and 3 µL restriction-ligation product. Reactions were amplified at 98 C for 3 seconds followed by 30 cycles of 98 C for 20 seconds, 60 C for 30 seconds, and 72 C for 30 seconds, followed by a final extension of 72 C for 10 minutes.

Table 2. Primer sequences used for ligation and PCR during ddRAD-seq library preparation. Msel1 and Msel2 adaptors and Illumina PCR primers (Illpcr1F & Illpcr2R) used for amplivication. '*' indicates phosphorothioate bonds at the 3' end of each Illumina primer to inhibit exonuclease activity of proof-reading polymerase as per Parchmen at al. (2012) library preparation protocol.

Primer	Sequence (5'-3')
Msel1	GCAGAAGACGGCATACGAGCTCTTCCGATCTG
Msel2	TACAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
Illpcr1F	A*A*TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Illpcr2R	C*A*AGCAGAAGACGGCATACGAGCTCTTCCGATCTGTAAG

A second extra PCR step is used to convert the single-stranded template remaining from the first PCR into a double-stranded form. This step results in libraries with a better distribution of fragments in the correct size range (Parchman et al., 2012). The extra PCR step contained 0.1x Iproof Buffer (HF), 0.18 mM each dNTP, and 0.29 µM pre-mixed Illumina primers in a 2.125 µL total volume and was added directly to the first 20 µL PCR reaction. The thermocycling profile for this reaction was 98 C for 3 minutes, followed by 60 C for 2 minutes, and a final extension of 72 C for 10 minutes.

Amplified fragments were then pooled into a single completed library plate and combined with another 3 additional plates with unique barcodes. All 4 plates were then sequenced in two single Illumina™ HiSeq 2500 platform lanes, with single-end 100 base-pair reads. This was repeated over 18 unique sample plates. In total, 1,518 samples were extracted and sequenced for this project. Of those samples, 168 were not able to identify clone or ploidy level due to low or missing data, 11 samples had unreliable geographic coordinates, and 95 samples were identified as belonging to the same clone, resulting in 1,244 available samples for modeling and analysis. Barcode indices, access to raw sequence data, and a list of sequencing runs can be found in the Supplementary Material (Appendix A).

Data Processing, Variant Calling and Filtering

To analyze the raw sequence data generated from the ddRAD-seq libraries we first demultiplexed sequences into individual files by sample and removed cutsite adapters using Cutadapt V4.1 (Martin, 2011). Custom scripts were then used to join sequencing data from replicated samples contained in multiple libraries. Cut-site and read-through adapters were removed using Cutadapt V4.1 and FastQC (Andrews, 2010)

was implemented to evaluate read quality and depth. Sequencing-level libraries averaged 225.5 million sequences, while samples averaged 1.12 million reads per sample (min = 238, max = 6.2 million). Samples containing <50000 or >5.5 million reads were dropped from further analysis.

After removing low-quality or low-read samples using custom scripts, the remaining sample sequences were mapped to the genome ((FAIRsharing Team, 2018; Lin et al., 2018), Potrs01b-genome.fa) using the Burrows-Wheeler Aligner (BWA) (H. Li & Durbin, 2009). The resulting mapped sequence files were then sorted and indexed using SAMtools (Danecek et al., 2021). Finally, reads were stacked and variants were called using BCFtools (Danecek et al., 2021). Variant call outputs (in VCF file format) were annotated to include allelic depth, genotype depth, and strand bias to aid in further filtering of single nucleotide polymorphisms (SNP) calls. We evaluated parameters including variant depth, variant missing data, minor allele frequency (MAF), individual depth, and individual missing data within the initial VCF file using VCFtools (Danecek et al., 2011). The initial variant call dataset contained 3,090,468 SNPs, each with an average depth of 2 sequences per SNP.

Further filtering of the initial variant calls allows for the identification of high-quality SNPs and can be filtered on the count of variants, allele frequency, site quality, mean depth of coverage, and the proportion of missing data per sample and per site. Filtering of the initial variant calls was completed by limiting variant sequence depth per sample to 3 minimum and 40 maximum. SNPs with a quality score less than 20 (as rated by BCFTools), and a minor allele frequency (MAF) of 0.025 were removed prior to further analysis. The filtered variant call dataset contained 24,613 SNPs with an average

depth of 12.5 SNPs. This step ensures that only high-quality variants are used in downstream analysis and reduces the potential for false positive results.

Clone Identification

Once variants were filtered, clonal membership was determined using a bimodal distribution of pairwise Jaccard distances as demonstrated in Mock et al. (2008) which enabled the removal of genetically identical ramets. This method is based on the premise that somatic mutations within a clone will result in much lower pairwise distances than differences between clones due to distinct seed origin. To accomplish this, we implemented the vcf2Jaccard script (Rowe, 2019) to convert variant calls to a pairwise data frame of mean Jaccard similarity coefficients across all SNPs. Jaccard coefficients were then evaluated to identify coefficient values indicative of clonal relationships using custom scripts (Appendix A, S2.2). A Jaccard coefficient value of >0.897 was used to assign pairwise samples to identical clones. Of 1,518 samples evaluated, 1,362 unique clones were identified. As expected, the distribution of distances due to somatic mutations within clones was separated from the distribution of inter-clonal distances by a span of distances with very low frequencies (Appendix C, Figure C4), and clones identified using this procedure were spatially proximal ranging from 9 to 740 m apart (mean = 188 meters).

Ploidy Level Classification

We used statistical methods to infer diploid or triploid cytotypes from rates of allelic heterozygosity ratios. Our initial ploidy level calls were performed using gbs2ploidy (Gompert & Mock, 2017). However, because gbs2ploidy was developed

using a training set of samples limited to Utah and Colorado, we were concerned that there would be an acquisition bias or heterozygosity differences across the greater extent of this study. To alleviate this concern, we compared and analyzed three separate ploidy level classifiers: nQuire (Weiß et al., 2018), FastPloidy (Goessen et al., 2022), and gbs2ploidy (Gompert & Mock, 2017). To further explore this concern, gbs2ploidy was also trained on varying geographic groups of trees including grouped samples from the Intermountain West, state, and site levels. A comparison of the ploidy level callers can be found in Appendix B. Overall, gbs2ploidy was used due to its accuracy in classifying ploidy level in data sets with low SNP depths by removing samples with n<1000 heterozygous alleles. Heterozygous allele depth and formatting for gbs2ploidy input were completed using custom scripts (Appendix A). Ploidy level calls resulted in 833 diploid and 419 triploid assignments, with 105 samples having ambiguous calls (samples not showing adequate probabilities of being diploid or triploid) and 59 samples with n < 1000 heterozygous alleles and were subsequently removed from downstream ploidy level analyses.

Genetic Sex Identification

Sex determination of aspen tree samples was achieved using the male-specific *TOZ19* locus as documented in Pakull et al. (2015), which includes a control locus present in both males and females to distinguish between PCR failure and negative fragment amplification. PCR amplification and conditions, as well as final genetic sex identification, were done as part of the work completed and described in Bidner (2021). While the *TOZ19* marker is known to be a reliable sex-identifier in *Populus* (Pakull et al., 2011) in diploid aspen, it has not been thoroughly verified in triploid aspen and should be

cautiously interpreted. We conducted a limited assessment of the *TOZ19* marker on triploid aspen at two sites (UTM and UTU) in the spring of 2023 to visually verify flowers of triploid aspen to compare with genetic sex identifications. Five male and two female triploid trees were visually verified by flower.

Climate, Topographic, and Edaphic Data

Plant distributions are influenced by many environmental variables acting at different spatial scales including microclimates, seasonal variation, and geographic heterogeneity. A goal of this study was to assess the distribution of aspen cytotypes and sexes, at varying geographic scales, considering topographic and modeled climate variables that are expected to impact soil moisture. In addition to seasonal precipitation, topographical features such as elevation, aspect, and slope steepness (slope) can indirectly control variables such as solar radiation, moisture availability, and soil characteristics (Barbour & Billings, 2000).

Our analysis employed 18 variables, 14 of which were derived from monthly 30-year normals (PRISM Climate Group, 2022), 3 from digital elevation models (DEM)(ELEV, SLOPE, SL)(*U.S. Geological Survey*, 2021), and a measure of modeled high resolution soil moisture (SMOGS)(Vergopolan et al. 2020) (Table 3). SMOGS is a modeled 30m data set of the first 5cm of soil surface calculated using land surface and radiative transfer modeling, microwave data, SMAP satellites, and in situ data (Vergopolan et al., 2020). In addition to annual climate measures such as temperature (Trng, Tmin, Tmean, Tmax, VPDMax, DEW), and precipitation (MAP), we further divide variables into temporal growing season subsets (TmaxGS, PPTGS, RHGS) and consider variables representing the periodicity of precipitation (PRATIO), temperature

minimums represented by growing degree days as a yearly warmth index (DD5), as well as variables expressing interaction between precipitation and growing degree days (GSPDD5 & ADI) (Table 3). A more detailed look at how each variable was calculated can be found in Appendix D. While GIS data is a representation of on-ground spatial information, its utility is limited by pixel size, so details within that resolution are generalized across entire spatial areas, possibly leading to resolution errors in extracted data. In this study, climate data is generalized over 800 m^2 pixels, while topographic and soil moisture measures are generalized over approximately 10 m^2 (1/3rd arc second) and 30 m^2, respectively. Additionally, ensuring accuracy and consistency between tiling schemes was considered when performing spatial joins or proximity analysis.

Table 3. Environmental variable acronyms, description, and initial importance ranking to its relevance to aspen cytotype profile. More detailed variable descriptions can be found in Appendix D.

Variable	Description	Variable Importance Rank
ELEV (m)	Elevation	**1***
SLOPE (d)	Slope	**3**
SL (index)	Southerlyness	13
Trng (°C)	Temperature Range	6
TmaxGS (°C)	Max Temperature - GS	18
PPTGS (mm)	Precipitation - GS	8
VPDmax (hPa)	Vapor Pressure Deficit Maximum	15
RHGS (%)	Relative Humidity - GS	12
DD5 (index)	Degree-days > 5 °C	14
GSPDD5 (index)	(PPTGS x DD5)/1000	17
ADI (index)	Annual Dryness Index	9
PRATIO (∝)	Ratio of GS to Annual Precipitation	**4**
SMOGS (%)	Soil Moisture	**2**
Tmax (°C)	Max Temperature	16
Tmin (°C)	Minimum Temperature	7
Tmean (°C)	Mean Temperature	11
DEW(d)	Mean Dewpoint	10
MAP	Mean Annual Precipitation	**5**

Notes:
'GS' = Growing Season is calculated as April - September
'*' = Ranking includes both measures of Elevation (ELEV) and Adjusted Elevation (AELEV).

Environment Modeling and Analysis

Modeling Overview

A suite of environmental models was implemented to explore aspen ploidy level and sex as response variables, employing a range of mixed-effect (ME) modeling structures to gain insight into the factors influencing the landscape distribution of cytotype and sex (Table 4). We first examined cytotype response with the site as a mixed effect, evaluating 18 variables independently (Model 1). Models 2 through 4 continued the investigation into ploidy level response at varying geographic scales, evaluating a 4-variable model with site as a mixed effect (Model 2) and two additional 6-variable models with latitude group (LATGRP; Model 3) and ecoregion group (ECOGRP; Model 4) as mixed effects. In Models 5 and 6, we expanded our analysis to include covariates, exploring ploidy level response with latitude group as a covariate and site as a mixed effect (Model 5), as well as with ecoregion group as a covariate and site as a mixed effect (Model 6), each utilizing three additional covariates. Lastly, we concentrated on sex as a response, employing SITE as a mixed effect (Model 7). This modeling approach allowed us to analyze the environmental factors that influence the distribution of cytotype and sex in aspen at multiple spatial extents.

Table 4. Environmental model names and structure evaluating variable importance, and cytotype and sex response at varying spatial scales.

Model #	Model Name	Model
1	Variable Importance	$CYTO_i \sim Binomial(p_i)$ $Logit(p_i) = \beta_0 + \beta_1 x_{VAR} + \gamma_0 S_{SITE}$
2	4-variable SITE ME	$CYTO_i \sim Binomial(p_i)$ $Logit(p_i) = \beta_0 + \beta_1 x_{AELEV} + \beta_2 x_{SLOPE} + \beta_3 x_{PRATIO} + \beta_4 x_{SMOGS} + \gamma_0 S_{SITE}$
3 & 4	6-variable LATGRP ME & 6-variable ECOGRP ME	$CYTO_i \sim Binomial(p_i)$ $Logit(p_i) = \beta_0 + \beta_1 x_{AELEV} + \beta_2 x_{SLOPE} + \beta_3 x_{TRNG} + \beta_4 x_{PRATIO\ 2} + \beta_5 x_{SMOGS} + \beta_6 x_{TmaxGS} + \gamma_0 S_{LATGRP\ or\ ECOGRP}$
5	3-variable LATGRP covariate	$CYTO_i \sim Binomial(p_i)$ $Logit(p_i) = \beta_0 + \beta_1 x_{SLOPE} + \beta_2 x_{TRNG} + \beta_3 x_{SMOGS} + \beta_4 x_{LATGRP} + \gamma_0 S_{SITE}$
6	3-variable ECOGRP covariate	$CYTO_i \sim Binomial(p_i)$ $Logit(p_i) = \beta_0 + \beta_1 x_{SLOPE} + \beta_2 x_{PRATIO} + \beta_3 x_{SMOGS} + \beta_4 x_{ECOGRP} + \gamma_0 S_{SITE}$
7	Sex ID ME	$SEXID_i \sim Binomial(p_i)$ $Logit(p_i) = \beta_0 + \beta_1 x_{AELEV} + \gamma_0 S_{SITE}$

Variable Selection

Model predictors were selected based on descriptive statistics and graphical representation for each variable. All computations, including data cleanup and descriptive statistics, were done using R (R Core Team, 2021), including open access packages that extend the base resources in R. The main package implemented for analysis included 'lme4' (Bates et al., 2015). Spearman's correlation tests were performed to assess the relationships between variables (Appendix A; Supplementary File 3). Further, each variables (VAR) importance was analyzed individually using a Generalized Linear Mixed Model (GLMM) with ploidy level (CYTO) as a binomial response with a logit-link and site (SITE) as a random intercept (Model 1). Pseudo R^2 (R^2_{mar} & R^2_{con}) and AIC (Akaike

Information Criterion) as well as BIC (Bayesian Information Criterion) values for each model were used to aid in evaluation of variable importance and subsequently inform which variables were included in the final model (Appendix C; Table C1 and Figure C1) and ranked based on importance in Table 3.

After exploring the correlation between variables, we conducted a linear regression analysis to explore the relationship between elevation (ELEV) and geographic location, including latitude (LAT) and longitude (LON) (Model 2, Figure C2 & C3). This was done to remove elevation bias based on latitudinal and longitudinal zonation effects. Aspen data for this model was downloaded from the USDA Forest Service, Forest Inventory and Analysis program (Burrill et al., 2021). The Forest Inventory and Analysis design uses a multiphase sampling design that is geographically unbiased and encompasses all land area to select forest sampling points (Bechtold and Patterson, 2005). Forested areas visited during sampling can be considered unbiased estimates of the larger population for a given forest type. We used the elevation measured on sampling points from any inventory plot that had the aspen present. Aspen elevation was then predicted from the model producing a new adjusted elevation variable (AELEV) for use in subsequent analysis. Latitude and longitude terms were squared to capture possible nonlinear relationships in elevation with respect to geographic location.

$$Y_{ELEV} = \beta_0 + \beta_1 x_{LAT} + \beta_2 x_{LAT}^2 + \beta_3 x_{LON} + \beta_4 x_{LON}^2 + \epsilon_{ELEV} \quad (1)$$

Variables were considered for removal from the model when Spearman's correlation value was >0.6. Removal of a variable was then based on interpretation of ecological importance to aspen and priority was given to non-synthetic variables.

Ploidy Level Modeling

Following removal of correlated variables, we employed a GLMM to simplify model complexity and further analyze the relationship between ploidy level and environmental variables. All variables were standardized to a mean of 0 and 1 standard deviation. We chose a GLMM as it allows for the incorporation of both fixed environmental variables and random effects introduced by sample collection sites, accommodating the hierarchical data structure across varying spatial scales. To increase the model skill, variables were individually removed and replaced one-by-one, with replacement of the previously tested variable, in a stepwise manner within the model. Resulting models (Table 5) were evaluated based on their AIC, a measure used to balance model fit and complexity, and our current ecological understanding of aspen systems.

Table 5. Variants of stepwise models used for model optimization dependent on ploidy level (CYTO) and site random effects. The marginal R^2 (R^2_{mar}) considers only the variance of the fixed effects (without the random effects), while the conditional R^2 (R^2_{con}) takes both fixed and random effects into account. A decrease in AIC from the 'All Variables' model variation (AIC 1448.93), when the model is run without a variable, indicates a decrease in model performance resulting in removal of variable for final model fit.

Model Variations by Site Mixed Effects

CYTO ~	**R^2_{con}**	**R^2_{mar}**	**AIC**	**BIC**
All Variables	0.3945	0.1354	1448.93	1505.32
w/out AELEV	0.4141	0.0512	1455.01	1506.27
w/out SLOPE	0.3862	0.1240	1451.25	1502.51
w/out SL	0.3947	0.1354	1446.97	1498.23
w/out Trng	0.4115	0.1233	1449.04	1500.30
w/out PRATIO	0.3881	0.1362	1448.25	1499.51
w/out SMOGS	0.3674	0.1180	1453.63	1504.90
w/out TmaxGS	0.3981	0.1321	1447.25	1498.51
w/out RH	0.3950	0.1353	1446.94	1498.20
w/out GSPDD5	0.3950	0.1329	1447.08	1498.34
Best Fit Model	**AELEV, SLOPE, PRATIO, SMOGS Only**			
	0.4087	0.1230	1441.19	1471.95

The optimal 4-variable SITE ME model (Table 4) included the binary response variable ploidy level (CYTO), modeled with the predictor variables adjusted elevation (AELEV), slope (SLOPE), growing-season to annual precipitation proportion (PRATIO), and growing season soil moisture (SMOGS), and implemented a binomial distribution with a logit-link function. To account for clustering of samples within sites (SITE), it was included as a mixed effect in the model.

The same modeling process as the 4-variable SITE ME model was repeated applying mixed effects based on site groupings of latitude (LATGRP; Figure 2) and ecoregion (ECOGRP; Figure 3) in place of site mixed effect, resulting in a 6-variable LATGRP ME model and a 6-variable ECOGRP ME model (Table 4). Variables were removed in a stepwise manner, for each model, as was done in the 4-variable SITE ME model to reduce complexity and increase skill of each model (Table 6). For grouping by latitude, sites were grouped in increments of 2° latitude 35° to 37°, 37° to 39°, 39° to 41°, 41° to 43° and 43° to 45° (Figure 2). Site 'AZW', 34° latitude, was included in the 35° to 37° grouping. Ecoregion groups were first based on site location with the Environmental Protection Agency (EPA), Level III Ecoregions of the Continental United States. Sites were further grouped if ecoregions spanned >2° latitude. If a single site was identified within its own ecoregion, it was grouped with additional sites containing the most similar climate zones (Beck et al., 2018) (Figure 3).

Table 6. Variants of stepwise models used for model optimization dependent on ploidy level (CYTO) and latitude group (LATGRP) or ecoregion group (ECOGRP) random effects. The marginal R^2(R^2_{mar}) considers only the variance of the fixed effects (without the random effects), while the conditional R^2(R^2_{con}) takes both fixed and random effects into account. A decrease in AIC from the 'All Variables' model variation (AIC 1488.06 for LATGRP or 1464.32 for ECOGRP), when the model is run without a variable, indicates a decrease in model performance resulting in removal of variable for final model fit.

Model Variations by Ecoregion group and Latitude group Mixed Effects

	LATGRP Modeling				ECOGRP Modeling			
CYTO ~	**R^2_{con}**	**R^2_{mar}**	**AIC**	**BIC**	**R^2_{con}**	**R^2_{mar}**	**AIC**	**BIC**
All Variables	0.2099	0.1606	1488.06	1544.45	0.3289	0.1676	1464.32	1520.70
w/out AELEV	0.2931	0.0996	1492.95	1544.21	0.3320	0.0628	1470.88	1522.14
w/out SLOPE	0.1902	0.1460	1490.84	1542.10	0.3119	0.1534	1467.42	1518.68
w/out SL	0.2098	0.1601	1486.28	1537.55	0.3293	0.1674	1462.41	1513.67
w/out Trng	0.1850	0.1227	1501.93	1553.20	0.3414	0.1414	1468.03	1519.29
w/out PRATIO	0.2075	0.1583	1486.50	1537.76	0.3196	0.1690	1463.90	1515.16
w/out SMOGS	0.2185	0.1618	1488.61	1539.87	0.3224	0.1508	1466.34	1517.60
w/out TmaxGS	0.1970	0.1584	1491.24	1542.50	0.3267	0.1545	1465.06	1516.32
w/out RH	0.2055	0.1555	1487.13	1538.39	0.3293	0.1663	1462.42	1513.68
w/out GSPDD5	0.2110	0.1555	1486.40	1537.66	0.3295	0.1652	1462.52	1513.78
Best Fit Model	**AELEV, SLOPE, Trng, PRATIO, SMOGS, TmaxGS Only**							
	0.2066	0.1529	1483.37	1524.38	0.3299	0.1644	1458.63	1499.64

In order to further investigate patterns within latitude and ecoregion groupings we repeated the variables selection process starting with stepwise removal of variables based on AIC in two separate models, a 3-variable LATGRP covariate and 3-variable ECOGRP covariate model, which included latitude or ecoregion groups as model covariates and included site as a mixed effect (Table 7).

Table 7. Variants of stepwise models used for model optimization of latitude and ecoregion groups (LATGRP & ECOGRP) as covariates, site random effects, and dependent ploidy level (CYTO). The marginal R^2(R^2_{mar}) considers only the variance of the fixed effects (without the random effects), while the conditional R^2(R^2_{con}) takes both fixed and random effects into account. A decrease in AIC from the 'All Variables' model variation (AIC 1452.43 for LATGRP or 1447.53 for ECOGRP), when the model is run without a variable, indicates a decrease in model performance.

Model Variations by Eco group and Lat group Covariates

	LATGRP Modeling				ECOGRP Modeling			
CYTO ~	**R^2_{con}**	**R^2_{mar}**	**AIC**	**BIC**	**R^2_{con}**	**R^2_{mar}**	**AIC**	**BIC**
All Variables	0.3914	0.1829	1452.43	1529.32	0.3897	0.3206	1447.53	1560.31
w/out AELEV	0.3916	0.1787	1450.75	1522.51	0.3949	0.3138	1447.33	1554.98
w/out SLOPE	0.3843	0.1703	1455.00	1526.77	0.3815	0.3075	1450.50	1558.14
w/out SL	0.3917	0.1828	1450.47	1522.23	0.3897	0.3205	1445.56	1553.21
w/out Trng	0.4093	0.1694	1452.86	1524.63	0.3951	0.3133	1447.32	1554.97
w/out PRATIO	0.3853	0.1836	1451.44	1523.21	0.3866	0.3164	1447.75	1555.40
w/out SMOGS	0.3646	0.1725	1456.30	1528.06	0.3689	0.3122	1450.49	1558.14
w/out TmaxGS	0.3999	0.1770	1445.12	1501.51	0.3936	0.3171	1446.04	1553.69
w/out RH	0.3920	0.1825	1450.44	1522.20	0.3899	0.3202	1445.56	1553.21
w/out GSPDD5	0.3914	0.1825	1450.44	1522.21	0.3895	0.3200	1445.56	1553.21
Best Fit Model	**SLOPE, Trng, SMOGS Only**				**SLOPE, PRATIO, SMOGS Only**			
	0.3960	0.1690	1443.93	1490.07	0.3983	0.3096	1438.54	1520.56

Sex Modeling

To explore ploidy level patterns we also considered sex as a possible driver of aspen cytotype distribution due to sex-specific morphological and physiological differences associated with reproductive burden. In order to determine if ploidy level patterns were being confounded by sex we modeled sex (SEXID) as a binary response. We compared variables in a stepwise manner to optimize model performance (Table 8), repeating the processes performed for modeling of ploidy level. The best fit model included only the adjusted elevation variable. The proportion of variance in the dependent variable associated with all the predictor variables was low, with an R2con of 0.0347 and an R^2_{mar} of 0.01409.

Table 8. Variants of stepwise models used for model optimization response to sex using site random effects. The marginal R^2 (R^2_{mar}) considers only the variance of the fixed effects (without the random effects), while the conditional R^2 (R^2_{con}) takes both fixed and random effects into account. A decrease in AIC from the 'All Variables' model variation (AIC 1594.63), when the model is run without a variable, indicates a decrease in model performance resulting in removal of the variable for final model fit.

SEXID Model Variations by Site Mixed Effects

SEXID ~	R^2_{con}	R^2_{mar}	AIC	BIC
All Variables	0.0341	0.0278	1594.63	1650.66
w/out AELEV	0.0372	0.0121	1596.50	1647.43
w/out SLOPE	0.0339	0.0278	1592.64	1643.57
w/out SL	0.0343	0.0267	1592.92	1643.85
w/out Trng	0.0428	0.0250	1593.30	1644.23
w/out PRATIO	0.0336	0.0277	1592.66	1643.59
w/out SMOGS	0.0311	0.0247	1593.54	1644.47
w/out TmaxGS	0.0349	0.0275	1592.68	1643.61
w/out RH	0.0344	0.0225	1594.08	1645.01
w/out GSPDD5	0.0341	0.0267	1592.94	1643.88
Best Fit Model		**AELEV Only**		
	0.0347	0.0141	1582.34	1597.62

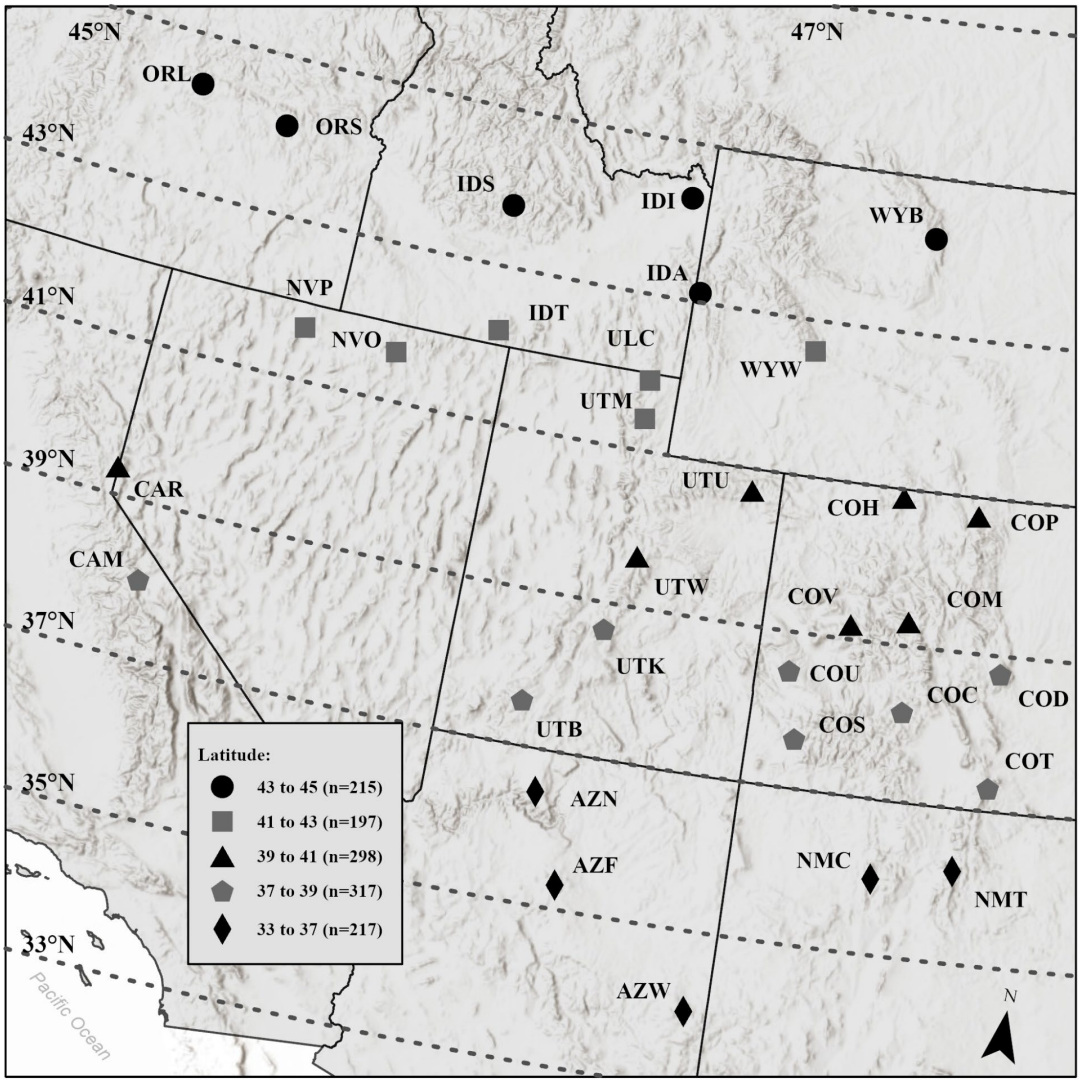

Figure 2. Sampling sites grouped by latitude (LATGRP). Sites were grouped in increments of 2° latitude 35° to 37°, 37° to 39°, 39° to 41°, 41° to 43° and 43° to 45° (Figure 2). Site 'AZW', at ~34° latitude, was included in the 35° to 37° grouping. n = the number of samples in each group.

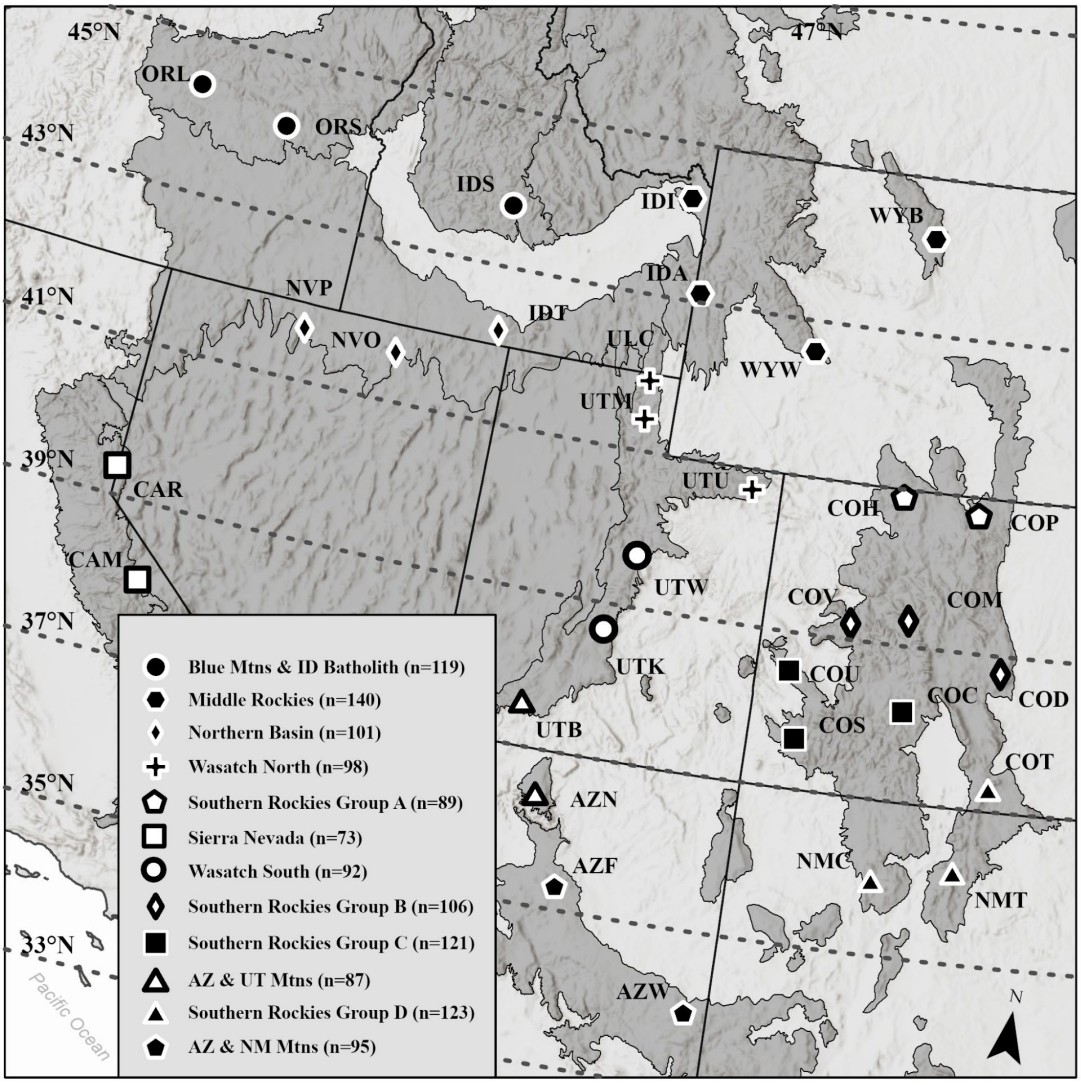

Figure 3. Sampling sites grouped by ecoregion (ECOGRP). Ecoregion groups were first based on site location with the Environmental Protection Agency (EPA), Level III Ecoregion of the Continental United States as seen in dark shaded regions. Sites were further grouped if ecoregions spanned >2° latitude. If a single site was identified in its own ecoregion it was grouped with sites with similar climate zones (sites IDS and AZN). n= the number of samples in each group.

CHAPTER III

RESULTS

Cytotype Distribution

Ploidy Level Classification

Classification of ploidy levels resulted in 865 diploids and 379 triploids represented in the dataset. Overall, the rate of detected triploidy was highest in Colorado and Utah. The occurrence of triploids decreased in sites north, south, and west of those in Colorado and Utah, and the largest ratios of diploids to triploids were found in California, Sierra Nevada range and sampled sites in Oregon. Site ‘AZW’, the study's most southern site, had the highest diploid to triploid ratio (22:1) and ‘UTK’, a site in Utah had the lowest ratio at 0.5:1, where triploids outnumbered sampled diploids (Figure 4). The distribution of triploid aspen were consistent with the findings of Mock et al. (2012), with the highest triploid proportions occurring at sites within Utah and Colorado; comparisons north of the last glacial maximum and south of the 33rd parallel (latitude) could not be made. Further, triploid cytotypes were found to be more common at more central latitudes and longitudes of the study area extent.

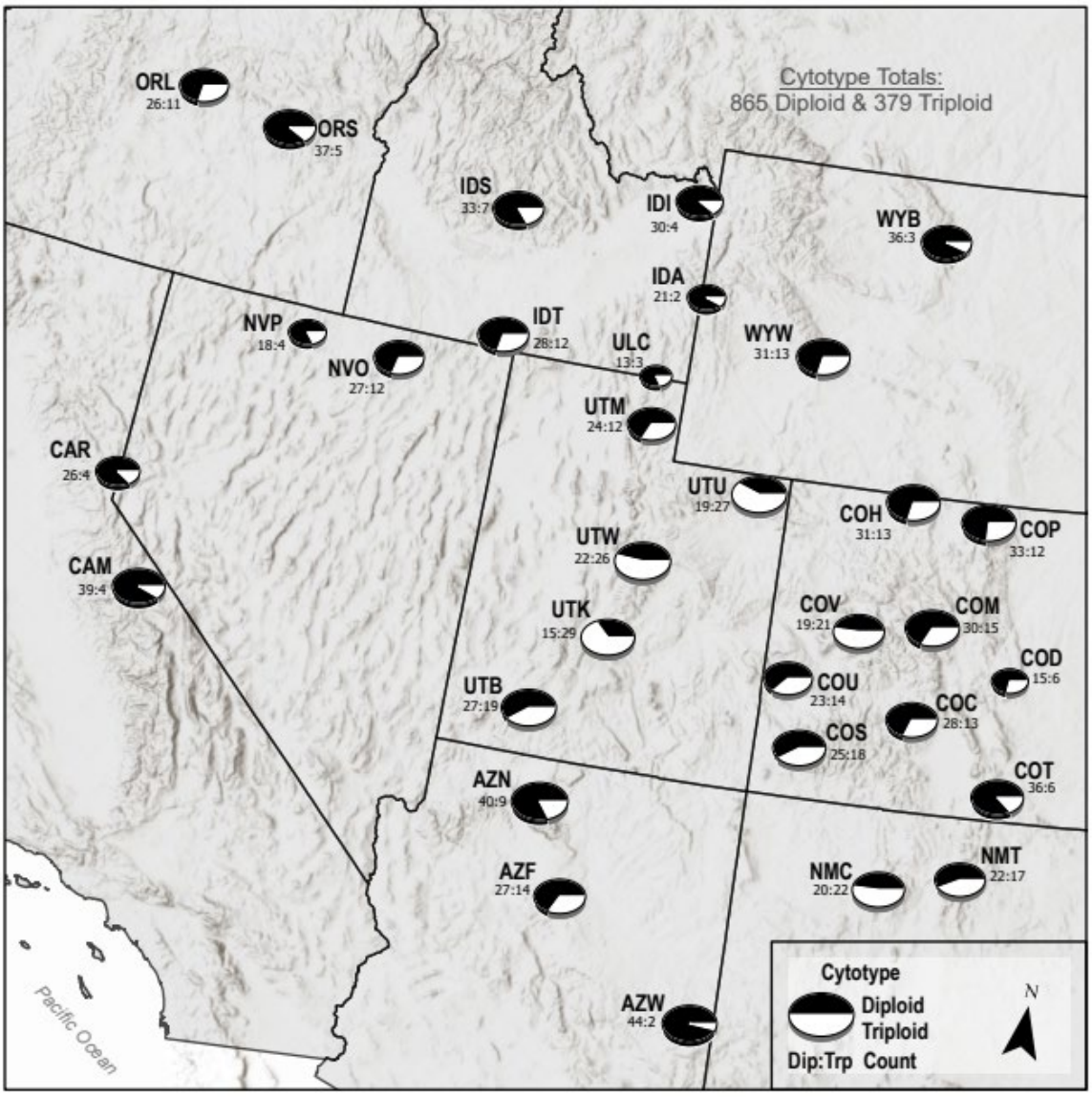

Figure 4. Assigned diploid and triploid counts by site. Each site is labeled with the site name and the counts of 'diploid/triploid' individuals within each site. The count of diploids are represented in black, and triploids in white within each pie chart. Pie charts are sized based on the number of samples within each site.

Ploidy Level Modeling - Site, Ecoregion, and Latitude Mixed Effects

Our 4-variable SITE ME model indicated that triploid aspen was more likely than diploids to occur in landscapes conducive to soil moisture content and was the most robust model tested in this study. This model incorporated site as a mixed effect, with adjusted elevation, slope, PRATIO, and growing season soil moisture covariates. (Table 9). The 4-variable SITE ME model had a conditional R^2(R^2_{con}) value of 0.403 with the majority of the model variation attributed to mixed effects rather than fixed effects (R^2_{mar}=0.123). Across all ME models (models 1 - 4), there is an increase in predicted triploidy as AELEV and SMOGS increase and SLOPE and PRATIO decrease (Figures 5-7). In the 6-variable LATGRP and ECOGRP ME models, there is an increase in predicted triploidy as Tmax increases and Trng decreases (Figures 6-7).

Similar patterns emerged when latitude and ecoregion groups (LATGRP and ECOGRP) were implemented as mixed effects in the 6-variable LATGRP ME and 6-variable ECOGRP ME models. Mixed effects within each model were predominant in explaining the variance of the model (R^2_{con} of 0.207 & 0.33 and R^2_{mar} of 0.152 & 0.164, respectively), and the 6-variable LATGRP ME model had less total variance in the model overall (Table 9). These two models shared common covariates with the 4-variable SITE ME model including AELEV, SLOPE, PRATIO, and SMOGS, but also added Tmax and Trng to the modeled covariates (Table 6 & Table 9). PRATIO and SMOGS were not significant in the 6-variable LATGRP ME model but contributed to the fit of the model. AELEV and SLOPE were significant across all three mixed effects models.

Table 9. Cytotype probability of variables for each 4- or 6-variable ME (mixed effect) models including intercept (β) and 95% confidence interval (95% CI). Variable significance is also listed at the following levels <.001(***), <0.01 (**), <0.05(*), and <0.1 (**..**).

	4-Variable SITE ME Model β_0 = -0.97517		**6-Variable LATGRP ME Model** β_0 = -0.90162		**6-Variable ECOGRP ME Model** β_0 = -0.90915	
	β	**95% CI**	***β***	**95% CI**	***β***	**95% CI**
AELEV	0.4008	0.1108 to 0.6948**	0.3904	0.1183 to 0.6145**	0.4814	0.1700 to 0.8206**
SLOPE	-0.1487	-0.2895 to -0.0116*	-0.1481	-0.2816 to -0.0183*	-0.1567	-0.2950 to -0.0223*
PRATIO	-0.1732	-0.3452 to -0.0045*	-0.1102	-0.2512 to 0.0306	-0.1583	-0.3228 to 0.0021..
SMOGS	0.2286	0.0553 to 0.40681*	0.1066	-0.0335 to 0.2475	0.1523	0.0025 to 0.3030*
Trng	-	-	-0.2703	-0.4077 to -0.1342***	-0.1809	-0.3317 to -0.0300*
Tmax	-	-	0.1695	0.02459 to 0.3172*	0.1336	-0.0254 to 0.2934..

Examining the relationships between covariates and the predicted probability of triploidy revealed consistent trends across modeled geographic groups. As the predicted probability of triploidy increases, elevation and soil moisture increased while slope and PRATIO decreased, this is true in all versions of the models presented. In the 6-variable LATGRP ME and ECOGRP ME models the predicted probability of triploidy increased when Trng decreased and Tmax increased (Figures 5, 6, and 7).

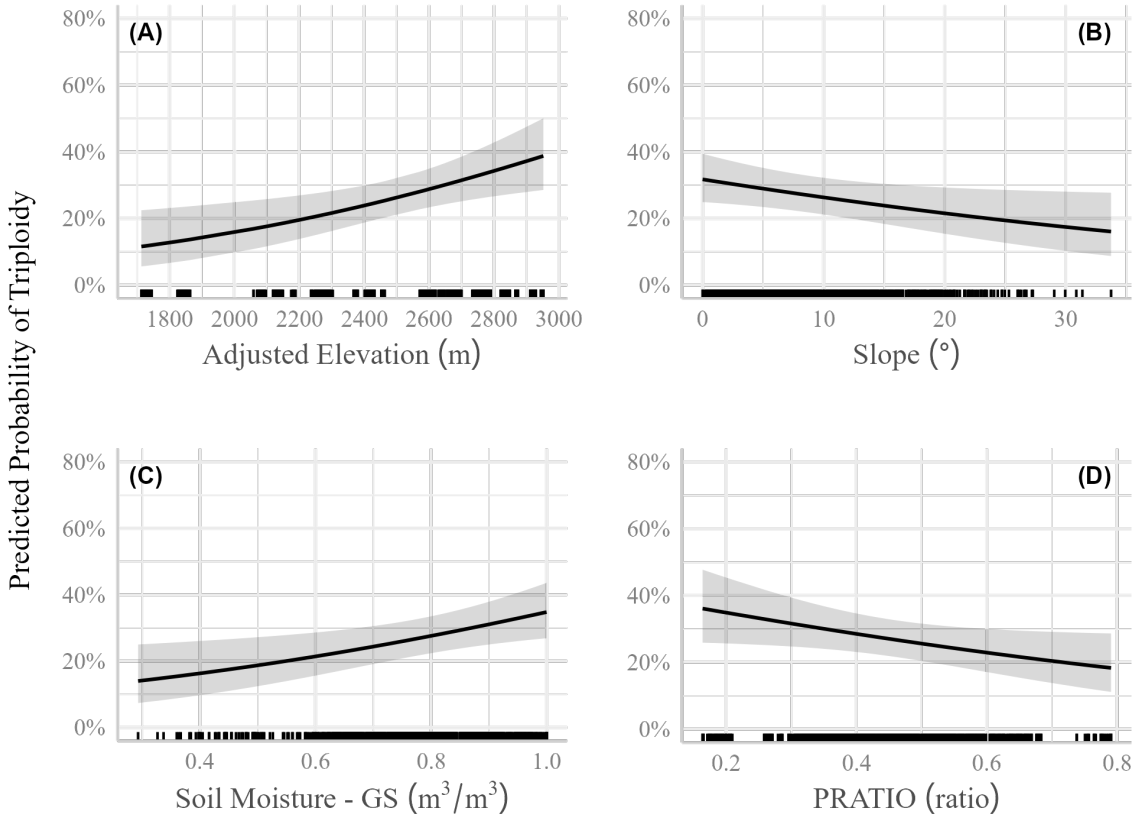

Figure 5. Predicted probability of triploidy in 4-variable SITE ME model. Including adjusted elevation (A; AELEV), slope (B; SLOPE), growing season soil moisture (C; SMOGS), and annual precipitation divided by growing season precipitation (D; PRATIO).

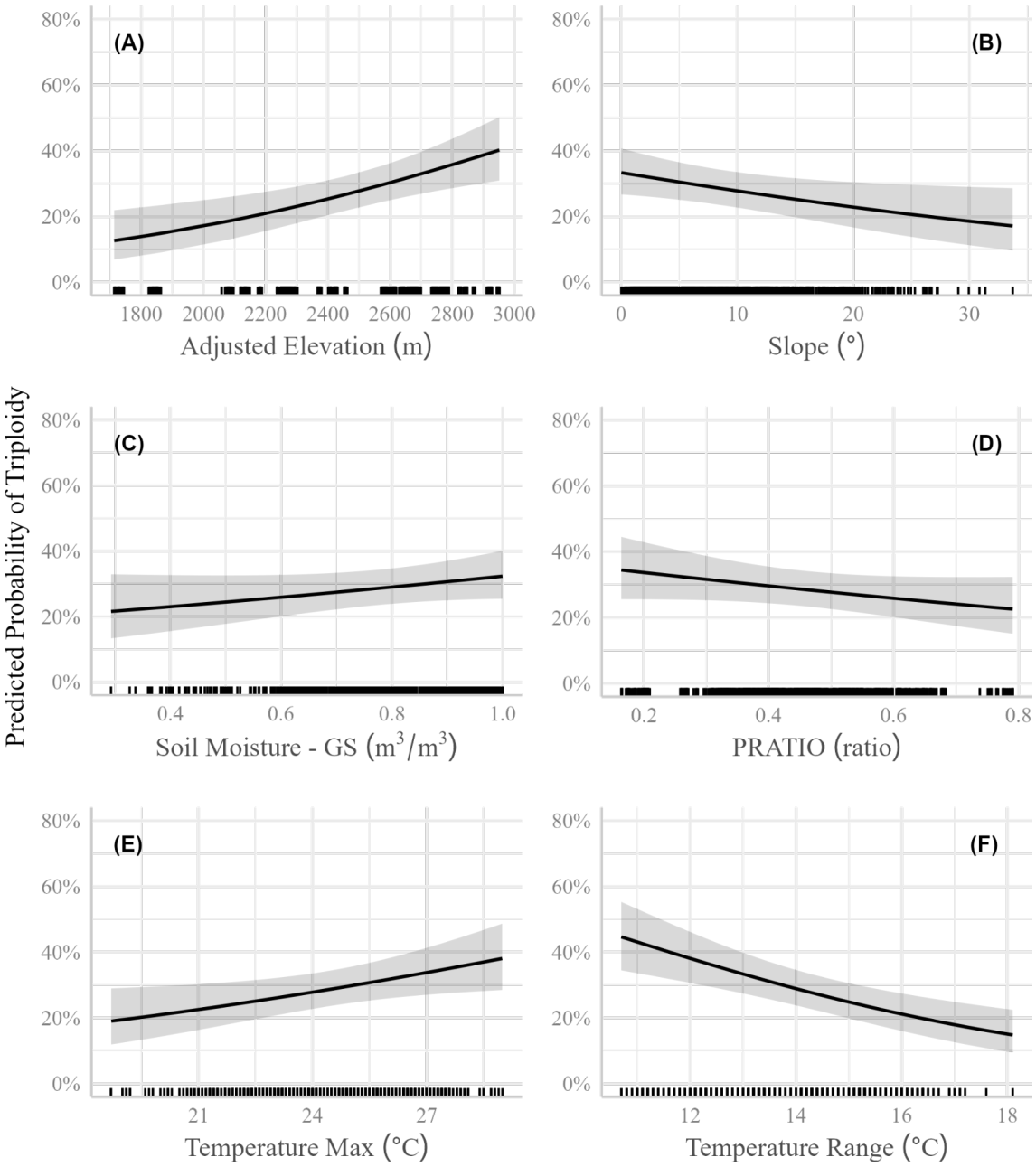

Figure 6. Predicted probability of triploidy in 6-variable LATGRP ME model with latitude group (LATGRP) mixed effects, including adjusted elevation (A; AELEV), slope (B; SLOPE), growing season soil moisture (C; SMOGS), annual precipitation divided by growing season precipitation (D; PRATIO), temperature maximum (E; Tmax), and temperature range (F; Trng) covariates.

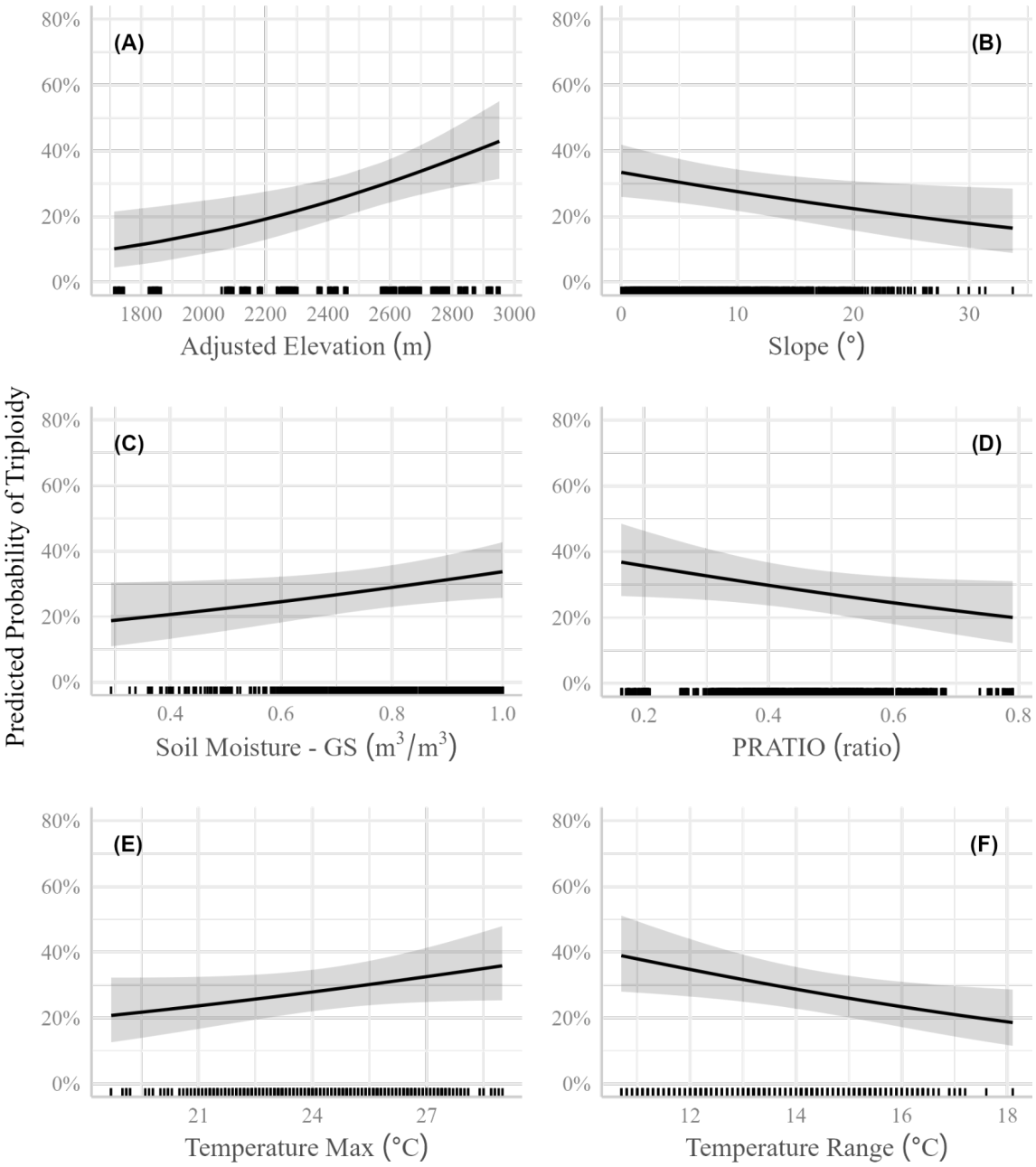

Figure 7. Predicted probability of triploidy in 6-variable ECOGRP ME model with ecoregion (ECOGRP) mixed effects. Including adjusted elevation (A; AELEV), slope (B; SLOPE), growing season soil moisture (C; SMOGS), annual precipitation divided by growing season precipitation (D; PRATIO), temperature maximum (E; Tmax), and temperature range (F; Trng) covariates.

Ploidy Level Modeling - Ecoregion, and latitude as covariates

When introducing LATGRP or ECOGRP as covariates in models we identified SLOPE, Trng, and SMOGS as contributing to the 3-variable LATGRP covariate model, while SLOPE, PRATIO, and SMOGS contributed to the fit in the 3-variable ECOGRP covariate model. Trng was included in the 3-variable LATGRP covariate model but was not significant (β_{Trng}=0.2391, 95% CI: -0.2811 to 0.0274). The persistent significance of SLOPE and SMOGS across both latitude and ecoregion groups underscore their importance to the model fit across spatial scales. (Table 10).

In the 3-variable LATGRP covariate model, the predicted probability of triploidy peaked in latitudes central to the contiguous Western United States (39° to 41°) and decreased with distance from this central point (Figure 8A). Mixed effects again predominantly explained most of the variation in the 3-variable LATGRP covariate model (R^2_{con}=0.396; R^2_{mar}=0.16; SLOPE, SMOGS, Trng).

The 3-variable ECOGRP covariate model predicted probability of triploidy and echoed results of the latitude group patterns, with the probability of triploidy decreasing from central latitudes, but also added a longitudinal dimension (Figure 10(A)), indicating an increase in triploidy as longitude decreased. Notably, specific ecoregions, such as Wasatch South, exhibited higher probabilities of triploidy at latitudes similar to their more eastward Colorado counterparts, e.g. the Southern Rockies Groups B and C (Figures 3, 4 & 10). The Sierra Nevada group showed the lowest predicted probability of triploidy. In comparison to the 3-variable LATGRP covariate model, the fixed effects (SLOPE, PRATIO, SMOGS) in the 3-variable ECOGRP covariate model explained greater variance in the model (R^2_{con}=0.398 and R^2_{mar}=0.31).

All covariates within the 3-variable LATGRP and ECOGRP covariate models demonstrated patterns similar to the SITE, LATGRP, and ECOGRP ME models in terms of increases or decreases in predicted probability of triploidy (Figure 9). Additionally, individual covariate trends conformed to patterns of increasing or decreasing predicted triploid probability across latitudinal and longitudinal gradients in the ECOGRP and LATGRP covariate models (Figures 8A & 9, Figures 10A & 11). Further, the LATGRP and ECOGRP covariate models also found Trng and Tmax to be significant predictors of triploidy, implying that the effects of environmental factors are not uniform across regions in the Intermountain West (Table 10).

Table 10. Cytotype probability of variables for each 3-variable LATGRP and 3-variable ECOGRP covariate model including intercept (β) and 95% confidence interval (95% CI). Variable significance is also listed at the following levels <.001(***), <0.01 (**), <0.05(*), and <0.1 (**.**).

	3-Variable LATGRP Covariate Model β_0 = -0.73580		**3-Variable ECOGRP Covariate Model** β_0 = -1.56037	
	β	**95% CI**	***β***	**95% CI**
SLOPE	-0.1515	-0.2892 to -0.0138*	-0.15557	-0.2947 to -0.0165*
PRATIO	-	-	-0.20635	-0.3801 to -0.0326*
SMOGS	0.2391	0.0635 to 0.4147**	0.19002	0.0111 to 0.3689*
Trng	-0.1268	-0.2811 to 0.02741	-	-

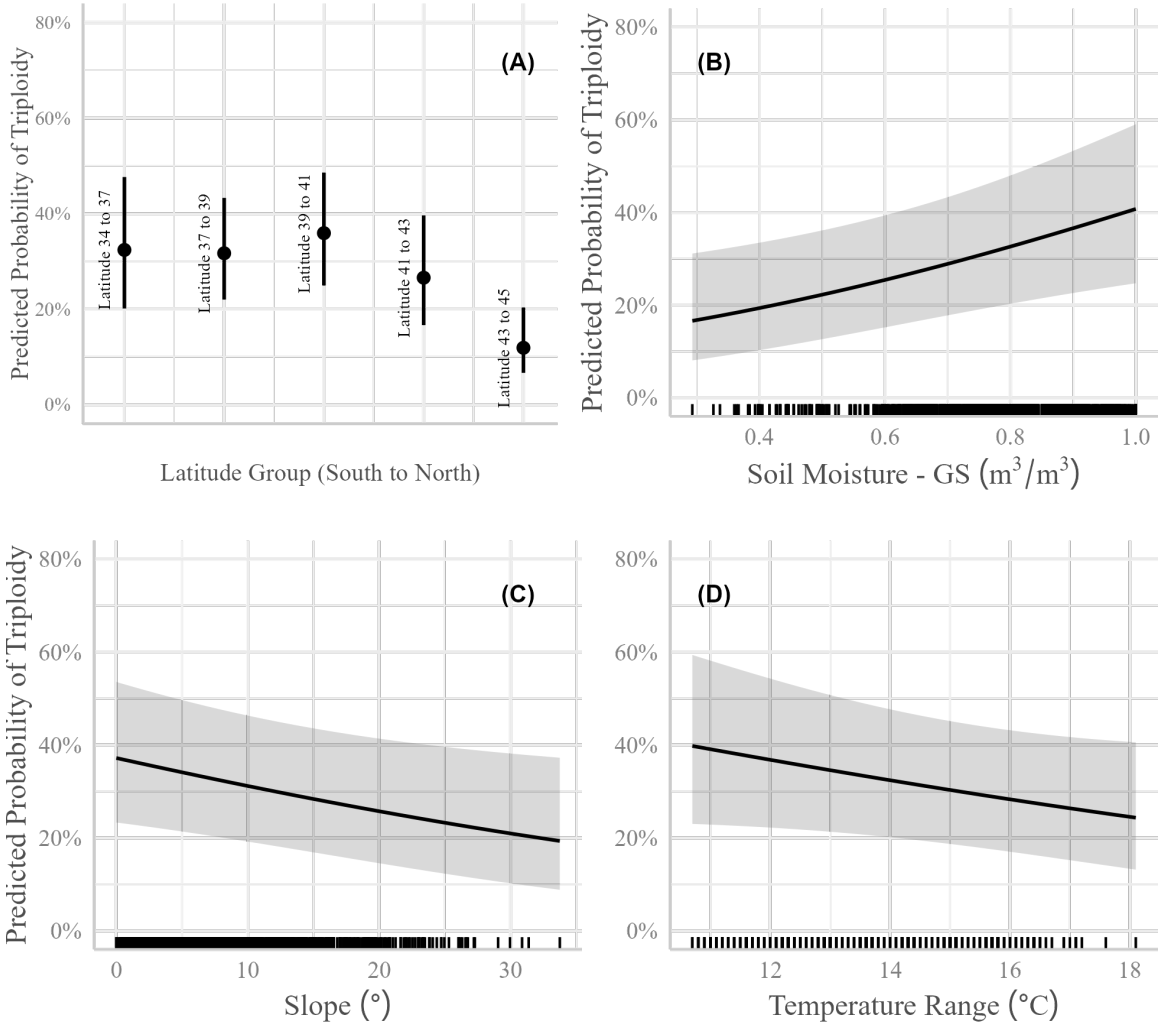

Figure 8. Predicted probability of triploidy in 3-variable LATGRP covariate model with site mixed effects. (A) is the predicted probability of triploidy by latitude group. (B), (C), and (D) show the predicted probability of each variable in the model; growing season soil moisture (SMOGS), slope (SLOPE), and temperature range (Trng).

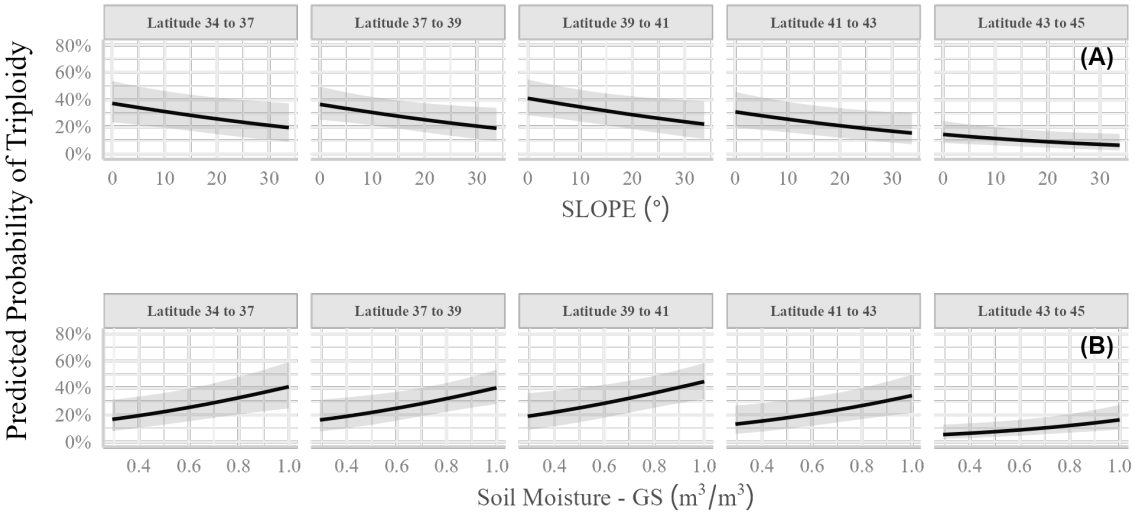

Figure 9. Predicted probability of triploidy in 3-variable LATGRP covariate model with site mixed effects for each significant covariate broken into latitude groups. (A) slope (SLOPE) and (B) growing season soil moisture (SMOGS).

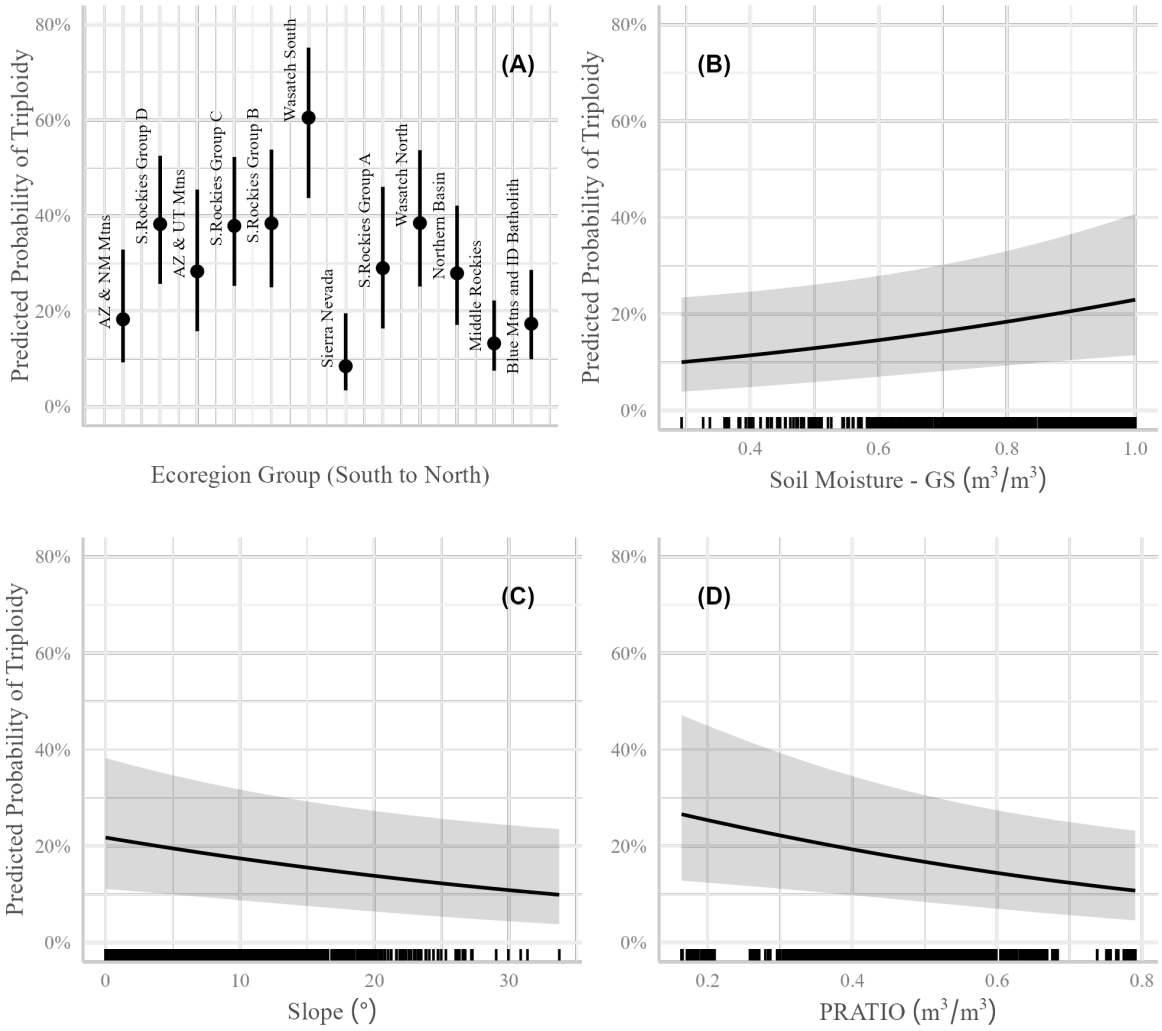

Figure 10. Predicted probability of triploidy in the 3-variable ECOGRP covariate model with site mixed effects. (A) is the predicted probability of triploidy by ecoregion group. (B), (C), and (D) show the predicted probability of each variable in the model; growing season soil moisture (SMOGS), slope (SLOPE), and the annual precipitation divided by the growing season precipitation (PRATIO).

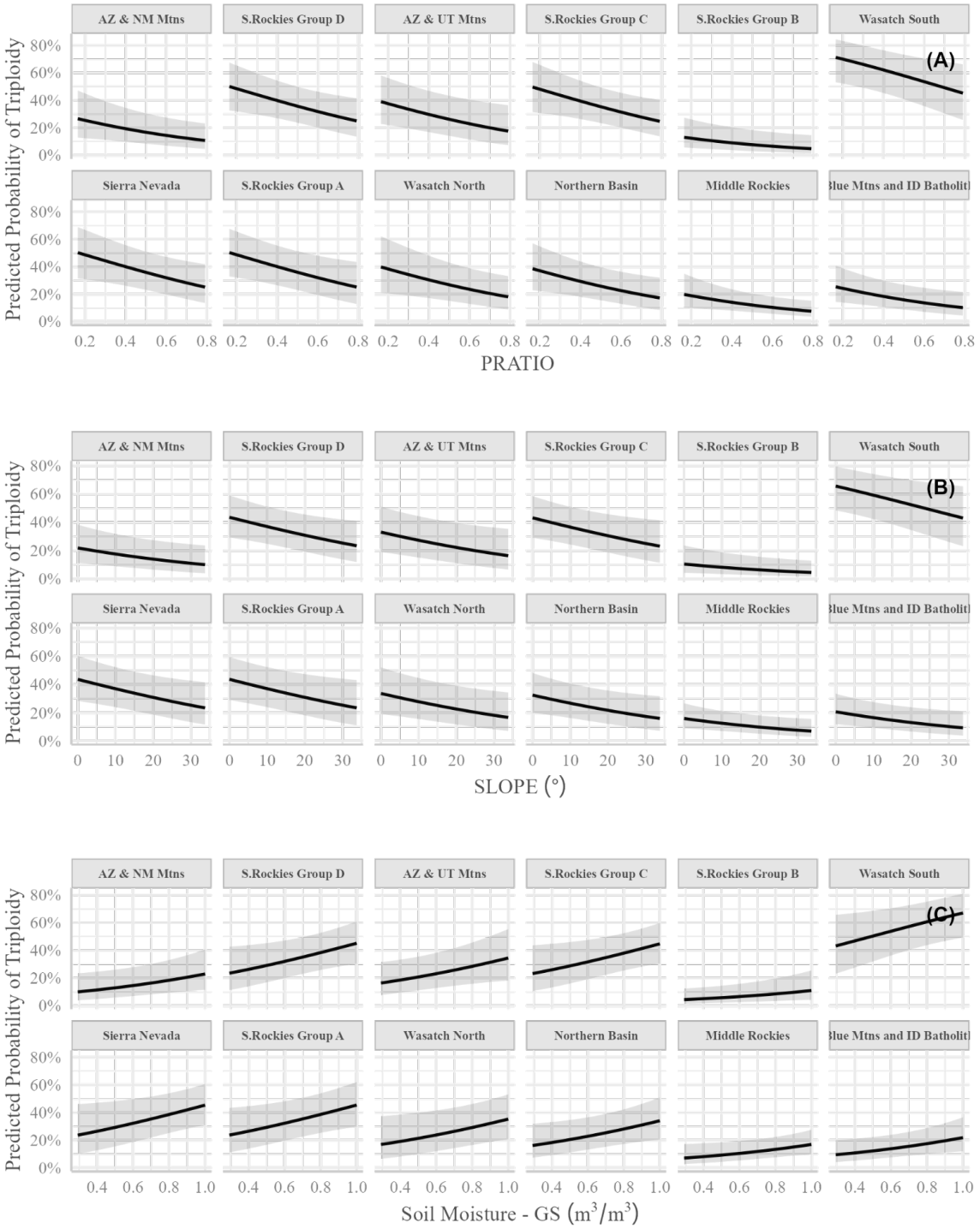

Figure 11. Predicted probability of triploidy in 3-variable ECOGRP covariate model with site mixed effects for each significant variable broken into ecoregion groups (ECOGRP). (A) annual precipitation divided by the growing season precipitation (PRATIO), (B) slope (SLOPE), and (C) growing season soil moisture (SMOGS).

Sex Ratios and Modeling

Sex ratios varied among cytotypes and across landscapes. Among cytotypes, we observed a strong male bias in sex ratios among triploids, ranging from 1:1 to 12:1 across sites, and 1:1 to 4:1 for diploids across sites. Triploid aspen has an approximate 3:1 higher proportion of males to females over their diploid counterparts (Figure 12; Table 11). Final sex identification calls can be found in the supplementary material (Appendix A, Supplementary File 2).

Environmental modeling of genetic sex identification across the landscape initially incorporated 9 variables (AELEV, SLOPE, SL, Trng, PRATIO, SMOGS, TmaxGS, and and RH); of those, only AELEV exhibited significance and contributed to the model fit. All other variables were found to diminish the overall fit of the model and were consequently excluded from the final model. Variability explained by the Sex ID ME model was low for both conditional and marginal effects (R^2_{con}=0.035 and R^2_{mar}=0.014). The significance of AELEV (β_{AELEV}=0.1202, 95% CI: -0.0094 to 0.2528 (β_0 = 0.5542)) along with the low variability found in the Sex ID ME model suggests that the observed relationships between environmental cytotype models, specifically the 4-variable SITE ME model, are not confounded by sexes possible stressors placed on aspen due to reproductive burden.

Table 11. Confusion matrix by cytotype (diploid and triploid) and sex identification (male and female).

	Diploid	**Triploid**	
Female	360	80	**440 (37%)**
Male	473	291	**764 (63%)**
	833 (69%)	**371 (31%)**	**1204 (100%)**

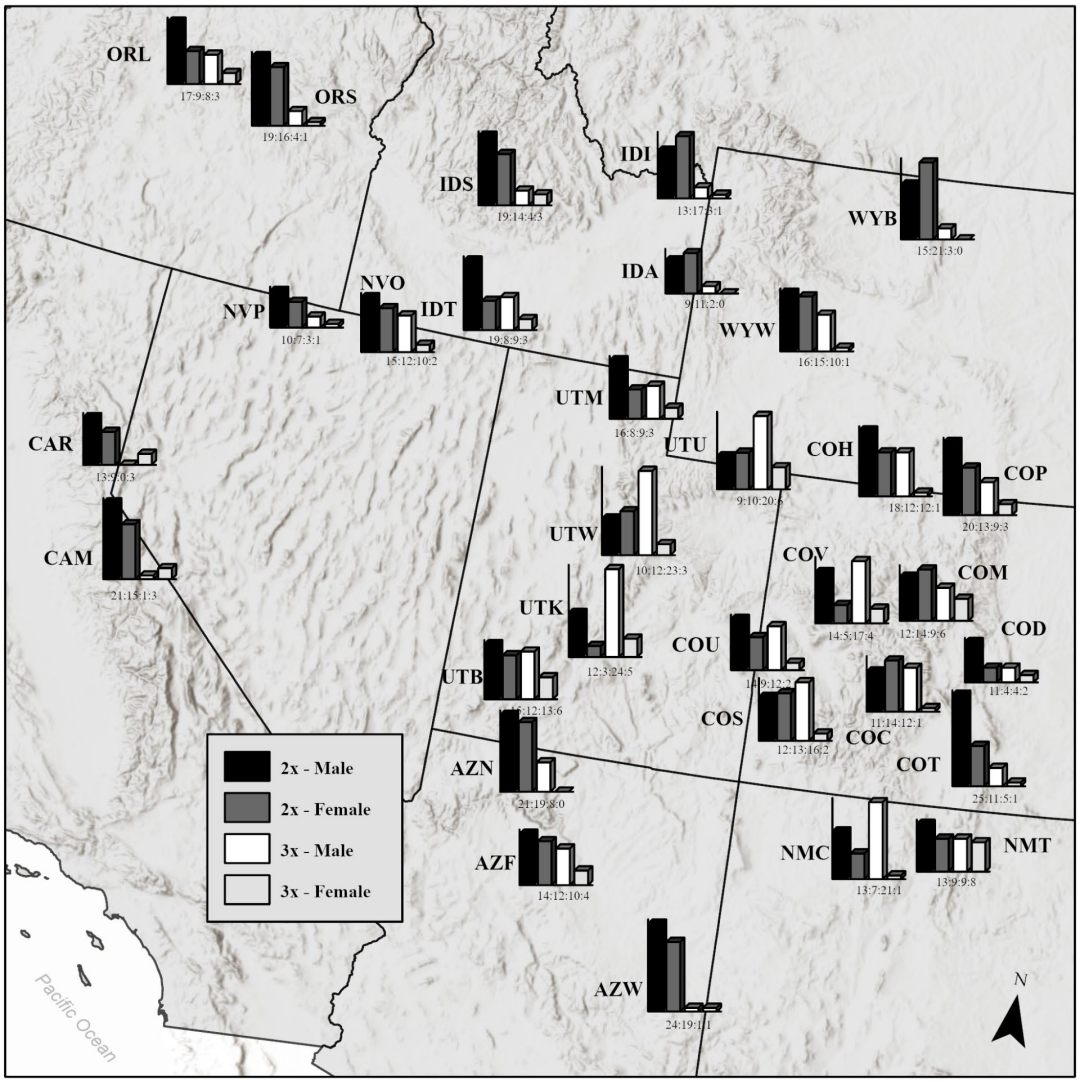

Figure 12. Genetic Sex ID counts by cytotype at each 30 km x 30 km sampling site. Sample counts below each bar graph indicate the counts for each of the groups; Diploid-Male:Diploid-Female:Triploid-Male:Triploid-Female.

CHAPTER IV

DISCUSSION

Our study aimed to better understand the environmental factors influencing the distribution of triploid aspen populations in the Intermountain West. Specifically, we examined the occurrence of diploid and triploid individuals at various spatial scales in relation to environmental variables associated with soil moisture content. We hypothesized that triploid aspen may require greater water resources than diploids due to biophysical differences, we also assessed the role of sex in aspen distributions, hypothesizing that the persistence of males vs. females may also be influenced by soil moisture.

Cytotype-Environment Relationships

Findings from this study collectively lend support to our hypothesis that the persistence and distribution of triploids on the landscape are linked with environmental variables that favor soil moisture content. Specifically, our models indicate that the predicted probability of triploidy increases as growing season soil moisture (SMOGS) increases. This finding is consistent with the findings and conclusions of previous experiments and field observations involving triploid aspen (Benson & Einspahr, 1967; Einspahr et al., 1963; Greer et al., 2018). Other variables generally favoring soil moisture content were also associated with increased triploid probability: higher PRATIO (the proportion of precipitation occurring in the growing season), Trng (annual temperature range), elevation (adjusted for latitude effects), and slope steepness (SLOPE).

PRATIO has been described as a measure of the annual periodicity of precipitation (Ledig et al., 2010; Rehfeldt et al., 2009) and associated with dryness or aridity of a site (Heiderman & Kimsey, 2021). It has also been suggested that PRATIO may be linked to regulating moisture stress by mitigating imbalances between temperature and precipitation (Rehfeldt et al., 2009). Elevated mean PRATIO values were observed at sites in the high desert regions of western Colorado and the Northern Basin Range of Nevada. These values diminish in sites nearer to the coast and in more northern locations.

Increased annual temperature range (Trng) is characteristic of regions with lower ambient humidity, and in general Trng is expected to be associated with lower soil moisture. Our models indicated that the probability triploidy increases with reduced Trng, consistent with our hypothesis. Similarly, we found that triploid aspen is predicted to be more common in areas of less-steep topography, which is consistent with other research findings (B. Blonder et al., 2020). We expect that these lower gradient landscapes generally retain more soil moisture than higher gradient landscapes

The probability of triploidy also increased at higher adjusted elevations, where higher soil moisture would be expected. This finding is contrary to a previous smaller scale study in which triploids were identified more commonly at lower elevations at a Colorado site (B. Blonder et al., 2020). This contradiction is likely an artifact of scale differences in the two studies, as the Blonder et al. (2020) study was located in a high-elevation site where landscape concavity and monsoonal effects may have dominated elevation effects. Surprisingly, our models also predicted that triploids were more likely to occur with increased maximum temperatures. This does not support our hypothesis

about soil moisture favoring triploidy, as we would expect high temperatures to decrease soil moisture.

The significance of environmental factors varied across spatial scales and regions, as would be expected over such a large landscape. Annual temperature range and maximum temperature had significant impacts at the ECOGRP and LATGRP ME modeling levels but not the SITE ME modeling level, indicating that the effects of these variables become more apparent when considering larger geographic and climate categories. PRATIO, on the other hand, was significant in the 4-variable SITE ME model ($p < 0.05$) but decreased as geographic analysis encompassed more sites in the 6-variable ECOGRP ME ($p < 0.1$) and LATGRP ME (no significance) models. This could imply that PRATIO has a more site-specific impact that may not generalize well across broader spatial scales or environmental contexts.

Our study was focused on soil moisture as a driver of triploid frequency. However, it is possible that other factors are impacting the occurrence of triploids over such a large landscape. For example, we note that across models and environmental variables, predicted probability of triploidy generally diminished closer to coastal areas and extreme latitudes, increasing toward central latitudes and eastward longitudes. This distribution suggests that factors such as phylogeographic and evolutionary history may also influence the occurrence of triploidy. Aspen in the Intermountain West are composed of distinct phylogeographic clusters (Bagley et al., 2020; Goessen et al., 2022; Mock et al., 2012), and more eastern populations within the Western United States, such as those in Colorado, may have separate evolutionary histories than more western populations in the region. Phylogeographic analysis of our sites was beyond the scope of

this study, but if there is a genetic component to triploid seed formation, this could complicate interpretation of our environmental variables.

Another factor potentially driving the occurrence of triploids may be the generation of unreduced gametes, which itself can be stress-induced (Comai, 2005). Our study framework regarding soil moisture presumes that the rate of triploid seed production is uniform across the continent. At latitude and longitude extremes, where greater availability of climate moisture (northern latitudes) or summer monsoonal precipitation patterns exist (southern latitudes), unreduced gamete production may be reduced, potentially resulting in lower densities of triploid aspen. Latitude group and ecoregion group covariate models (LATGRP & ECOGRP covariate models) in this study lend support to this hypothesis (Figures 8 & 10). Further, Ally et al. (2010) note that longevity in aspen clones may be related to accumulation of deleterious mutations and reduced pollen viability. Such accumulated deleterious mutations may also increase the rate of unreduced gametes. Thus, the occurrence of triploids may be related not only to their survival relative to diploids, but also to the frequency of their occurrence in seeds. The factors influencing production of unreduced gametes and triploid seed production have not been explored in aspen, but are likely to be many of the same factors that could influence differential cytotype survival.

Sex-Environment Relationships

Aspen is a dioecious species, and because flowering does not occur every year, the distribution of sexes is somewhat cryptic. Research suggests that sex-specific differences in dioecious plants can result in a higher energetic burden associated with seed production in females. (Barrett & Hough, 2013; Cipollini & Whigham, 1994;

Dawson & Ehleringer, 1993; Espirito-Santo, 2003; Freeman et al., 1980; Gross & Soule, 1981; C. Li et al., 2007). Sex biases in energetic burden could increase survival rates for males in areas of low soil moisture (Dawson & Ehleringer, 1993; Hultine et al., 2007), and females' ability to outcompete males with faster growth rates and canopy size in areas of high soil moisture (Dawson & Ehleringer, 1993; Ward et al., 2002) may create spatial segregation between sexes and decreased chances of successful fertilization (Hultine et al., 2007; Iszkuło & Boratyński, 2011; Nuñez et al., 2008). Disparities in stress vulnerability between males and females may amplify or counteract cytotype-specific responses to environmental factors.

Previous studies have noted a strong male bias in aspen distributions, particularly at higher elevations (Bidner, 2021; B. W. Blonder et al., 2023; Grant & Mitton, 1979). In an unpublished study, 100 seeds were collected from two maternal trees in northern Utah and were raised in a greenhouse to determine whether sex biases were present (Mock, unpublished data). Progeny from the first maternal clone consisted of 51 males (55%) out of 92 successfully sexed individuals. Progeny from the second maternal clone consisted of 46 males (56%) out of 82 successfully sexed individuals. Though this study was limited in number and geographical extent, it demonstrated, as expected, that sex ratios in these seed crops were not significantly different from 1:1. This finding suggests the absence of sex biases prior to seed formation, despite a male bias on the landscape, further suggesting that environmental factors post-seed production may be influencing the survival advantage of male trees over females within the landscape.

Female reproductive burden may share many of the same environmental factors that drive aspen distribution on the landscape as ploidy level, mainly those that have to do

with resource or soil moisture availability. Our results indicate an overall bias towards males in aspen across its range in the West, but this male bias was especially pronounced in triploids. This suggests that the distribution of triploids may be influenced by factors driving the distribution of males. However, our analysis, which included sex as a response variable only identified adjusted elevation as significant, but also showed low model fit and did not lend support to this link. This suggests that some other process or measure of selection may be driving aspen sex determination post-seed production.

Conclusion

Overall, we found evidence that the occurrence of triploidy was related to several environmental variables which could be associated with low soil moisture availability, including seasonality of precipitation, slope, elevation, and temperature range. To better understand aspen-environment-ploidy dynamics, future studies could consider the assessment of triploidy among seed crops at a continental scale and exploring the potential association between triploidy and geographic genetic structure in aspen populations. By evaluating the association between cytotype diversity and the environmental factors driving that diversity, we hope to better inform conservation, restoration, assisted migration, or other silviculture practices.

REFERENCES

Ally, D., Ritland, K., & Otto, S. P. (2010). Aging in a long-lived clonal tree. *PLoS Biology*, *8*(8), e1000454. https://doi.org/10.1371/journal.pbio.1000454

Altieri, A. H., Silliman, B. R., & Bertness, M. D. (2007). Hierarchical organization via a facilitation cascade in intertidal cordgrass bed communities. *The American Naturalist*, *169*(2), 195–206. https://doi.org/10.1086/510603

Anderegg, L. D. L., Anderegg, W. R. L., Abatzoglou, J., Hausladen, A. M., & Berry, J. A. (2013). Drought characteristics' role in widespread aspen forest mortality across Colorado, USA. *Global Change Biology*, *19*(5), 1526–1537. https://doi.org/10.1111/gcb.12146

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data* [Computer software]. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Angelini, C., Altieri, A. H., Silliman, B. R., & Bertness, M. D. (2011). Interactions among foundation species and their consequences for community organization, biodiversity, and conservation. *BioScience*, *61*(10), 782–789. https://doi.org/10.1525/bio.2011.61.10.8

Bagley, J. C., Heming, N. M., Gutiérrez, E. E., Devisetty, U. K., Mock, K. E., Eckert, A. J., & Strauss, S. H. (2020). Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (*Populus tremuloides*). *Ecology and Evolution*, *10*(11), 4609–4629. https://doi.org/10.1002/ece3.6214

Barbour, M. G., & Billings, W. D. (2000). *North American Terrestrial Vegetation*. Cambridge University Press. https://books.google.com/books?id=IZZZd6qdwEcC

Barrett, S. C. H., & Hough, J. (2013). Sexual dimorphism in flowering plants. *Journal of Experimental Botany*, *64*(1), 67–82. https://doi.org/10.1093/jxb/ers308

Bartos, D. L., & Campbell, R. B., Jr. (1998). Decline of quaking aspen in the Interior West—Examples from Utah. *Rangelands*, *20*(1), 17–24.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beaulieu, J. M., Leitch, I. J., Patel, S., Pendharkar, A., & Knight, C. A. (2008). Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist*, *179*(4), 975–986. https://doi.org/10.1111/j.1469-8137.2008.02528.x

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Benson, M. K., & Einspahr, D. W. (1967). Early growth of diploid, triploid and triploid hybrid aspen. *Forest Science*, *13*(2), 150–155.

Bidner, R. (2021). *Methods to improve Understanding of aspen regeneration and aspen distribution across the Intermountain West* [MS, Utah State University]. All Graduate Theses and Dissertations. https://digitalcommons.usu.edu/etd/8336

Blonder, B., Brodrick, P. G., Walton, J. A., Chadwick, K. D., Breckheimer, I. K., Marchetti, S., Ray, C. A., & Mock, K. E. (2022). Remote sensing of cytotype and its consequences for canopy damage in quaking aspen. *Global Change Biology*, *28*(7), 2491–2504. https://doi.org/10.1111/gcb.16064

Blonder, B., Graae, B. J., Greer, B., Haagsma, M., Helsen, K., Kapás, R. E., Pai, H.,

Rieksta, J., Sapena, D., Still, C. J., & Strimbeck, R. (2020). Remote sensing of ploidy level in quaking aspen (*Populus tremuloides* Michx.). *Journal of Ecology*, *108*(1), 175–188. https://doi.org/10.1111/1365-2745.13296

Blonder, B. W., Brodrick, P. G., Chadwick, K. D., Carroll, E., Cruz-de Hoyos, R. M., Expósito-Alonso, M., Hateley, S., Moon, M., Ray, C. A., Tran, H., & Walton, J. A. (2023). Climate lags and genetics determine phenology in quaking aspen (*Populus tremuloides*). *New Phytologist*, *238*(6), 2313–2328. https://doi.org/10.1111/nph.18850

Burrill, E. A., DiTommaso, A. M., Turner, J. A., Pugh, S. A., Menlove, J., Christiansen, G., Carol, J. P., & Conkling, B. L. (2021). *The Forest Inventory and Analysis Database: Database description and user guide version 9.0.1 for Phase 2.* (Version 1026) [dataset]. U.S. Department of Agriculture, Forest Service. http://www.fia.fs.fed.us/library/database-documentation/

Cai, J. (2019). *Humidity: Calculate water vapor measures from temperature and dew point* (R package version 0.1.5) [Computer software]. https://github.com/caijun/humidity

Chong, D., Yang, R.-C., & Yeh, F. C. (1994). Nucleotide divergence between populations of trembling aspen (*Populus tremuloides*) estimated with RAPDs. *Current Genetics*, *26*(4), 374–376. https://doi.org/10.1007/BF00310504

Chong, G. W., Simonson, S. E., Stohlgren, T. J., & Kalkhan, M. A. (2001). Biodiversity: Aspen stands have the lead, but will nonnative species take over? *USDA Forest Service Proceedings*, *18*, 261.

Cipollini, M. L., & Whigham, D. F. (1994). Sexual dimorphism and cost of reproduction

in the dioecious shrub *Lindera benzoin* (Lauraceae). *American Journal of Botany*, *81*(1), 65–75. https://doi.org/10.1002/j.1537-2197.1994.tb15410.x

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, *6*(11), 836–846. https://doi.org/10.1038/nrg1711

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

David, A. J., Zasada, J. C., Gilmore, D. W., & Landhäusser, S. M. (2001). Current trends in the management of aspen and mixed aspen forests for sustainable production. *The Forestry Chronicle*, *77*(3), 525–532. https://doi.org/10.5558/tfc77525-3

Dawson, T. E., & Ehleringer, J. R. (1993). Gender-specific physiology, carbon Isotope discrimination, and habitat distribution in boxelder, *Acer Negundo*. *Ecology*, *74*(3), 798–815. https://doi.org/10.2307/1940807

DeRose, R. J., Mock, K. E., & Long, J. N. (2015). Cytotype differences in radial increment provide novel insight into aspen reproductive ecology and stand dynamics. *Canadian Journal of Forest Research*, *45*(1), 1–8. https://doi.org/10.1139/cjfr-2014-0382

Des Roches, S., Pendleton, L. H., Shapiro, B., & Palkovacs, E. P. (2021). Conserving intraspecific variation for nature's contributions to people. *Nature Ecology & Evolution*, *5*(5), 574–582. https://doi.org/10.1038/s41559-021-01403-5

Dixon, G. B., & DeWald, L. E. (2015). Microsatellite survey reveals possible link between triploidy and mortality of quaking aspen in Kaibab National Forest,

Arizona. *Canadian Journal of Forest Research*, *45*(10), 1369–1375. https://doi.org/10.1139/cjfr-2014-0566

Dudley, M. M., Burns, K. S., & Jacobi, W. R. (2015). Aspen mortality in the Colorado and southern Wyoming Rocky Mountains: Extent, severity, and causal factors. *Forest Ecology and Management*, *353*, 240–259. https://doi.org/10.1016/j.foreco.2015.06.002

Dynesius, M., & Jansson, R. (2000). Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences*, *97*(16), 9115–9120. https://doi.org/10.1073/pnas.97.16.9115

Edgeloe, J. M., Severn-Ellis, A. A., Bayer, P. E., Mehravi, S., Breed, M. F., Krauss, S. L., Batley, J., Kendrick, G. A., & Sinclair, E. A. (2022). Extensive polyploid clonality was a successful strategy for seagrass to expand into a newly submerged environment. *Proceedings of the Royal Society of Biological Sciences*, *289*(1976), 20220538. https://doi.org/10.1098/rspb.2022.0538

Einspahr, D. W., Van Buijtenen, J., & Peckham, J. (1963). Natural variation and heritability in triploid aspen. *Silvae Genetica*, *12*(2), 51–58.

Eisenring, M., Lindroth, R. L., Flansburg, A., Giezendanner, N., Mock, K. E., & Kruger, E. L. (2023). Genotypic variation rather than ploidy level determines functional trait expression in a foundation tree species in the presence and absence of environmental stress. *Annals of Botany*, *131*(1), 229–242. https://doi.org/10.1093/aob/mcac071

Ellison, A. M., Bank, M. S., Clinton, B. D., Colburn, E. A., Elliott, K., Ford, C. R.,

Foster, D. R., Kloeppel, B. D., Knoepp, J. D., Lovett, G. M., Mohan, J., Orwig, D. A., Rodenhouse, N. L., Sobczak, W. V., Stinson, K. A., Stone, J. K., Swan, C. M., Thompson, J., Holle, B. V., & Webster, J. R. (2005). Loss of foundation species: Consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment*, *3*(9), 479–486. https://doi.org/10.1890/1540-9295(2005)003[0479:LOFSCF]2.0.CO;2

Espirito-Santo, M. M. (2003). Sexual differences in reproductive phenology and their consequences for the demography of *Baccharis dracunculifolia* (Asteraceae), a dioecious tropical shrub. *Annals of Botany*, *91*(1), 13–19. https://doi.org/10.1093/aob/mcg001

ESRI. (2022). *ArcGIS Pro 3.0.2* [Computer software]. Envronmental Systems Research Institute. https://www.esri.com/en-us/home

FAIRsharing Team. (2018). *FAIRsharing record for: Hardwood Genomics Project* [dataset]. FAIRsharing. https://doi.org/10.25504/FAIRSHARING.SRGKAF

Fechner, G. H., & Barrows, J. S. (1976). Aspen stands as wildfire fuel breaks. *Eisenhower Consortium Bulletin*, *4*, 26.

Fife, H. (1994). *Warriors and guardians: Native Highland trees*. Argyll Publishing.

Freeman, D. C., Harper, K. T., & Charnov, E. L. (1980). Sex change in plants: Old and new observations and new hypotheses. *Oecologia*, *47*(2), 222–232. https://doi.org/10.1007/BF00346825

Goessen, R., Isabel, N., Wehenkel, C., Pavy, N., Tischenko, L., Touchette, L., Giguère, I., Gros-Louis, M., Laroche, J., Boyle, B., Soolanayakanahally, R., Mock, K., Hernández-Velasco, J., Simental-Rodriguez, S. L., Bousquet, J., & Porth, I. M.

(2022). Coping with environmental constraints: Geographically divergent adaptive evolution and germination plasticity in the transcontinental *Populus tremuloides*. *Plants, People, Planet*, *4*(6), ppp3.10297. https://doi.org/10.1002/ppp3.10297

Gompert, Z., & Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Molecular Ecology Resources*, *17*(6), 1156–1167. https://doi.org/10.1111/1755-0998.12657

Grant, M. C., & Mitton, J. B. (1979). Elevational gradients in adult sex ratios and sexual differentiation in vegetative growth rates of *Populus tremuloides* Michx. *Evolution*, *33*(3), 914–918. https://doi.org/10.1111/j.1558-5646.1979.tb04744.x

Greer, B. T., Still, C., Cullinan, G. L., Brooks, J. R., & Meinzer, F. C. (2018). Polyploidy influences plant–environment interactions in quaking aspen (*Populus tremuloides* Michx.). *Tree Physiology*, *38*(4), 630–640. https://doi.org/10.1093/treephys/tpx120

Greer, B. T., Still, C., Howe, G. T., Tague, C., & Roberts, D. A. (2016). Populations of aspen (*Populus tremuloides* Michx.) with different evolutionary histories differ in their climate occupancy. *Ecology and Evolution*, *6*(9), 3032–3039. https://doi.org/10.1002/ece3.2102

Gross, K. L., & Soule, J. D. (1981). Differences in biomass allocation to reproductive and vegetative structures of male and female plants of dioecious, perennial herb *Silene alba* (Miller) Krause. *American Journal of Botany*, *68*(6), 801–807. https://doi.org/10.1002/j.1537-2197.1981.tb12414.x

Heiderman, R. R., & Kimsey, M. J. (2021). A species-specific, site-sensitive maximum

stand density index model for Pacific Northwest conifer forests. *Canadian Journal of Forest Research*, *51*(8), 1166–1177. https://doi.org/10.1139/cjfr-2020-0426

Heilborn, O. (1934). On the origin and preservation of polyploidy. *Hereditas*, *19*(1–2), 233–242. https://doi.org/10.1111/j.1601-5223.1934.tb02624.x

Hodgson, J. G., Sharafi, M., Jalili, A., Díaz, S., Montserrat-Martí, G., Palmer, C., Cerabolini, B., Pierce, S., Hamzehee, B., Asri, Y., Jamzad, Z., Wilson, P., Raven, J. A., Band, S. R., Basconcelo, S., Bogard, A., Carter, G., Charles, M., Castro-Díez, P., … Simmons, E. (2010). Stomatal vs. genome size in angiosperms: The somatic tail wagging the genomic dog? *Annals of Botany*, *105*(4), 573–584. https://doi.org/10.1093/aob/mcq011

Hultine, K. R., Bush, S. E., West, A. G., & Ehleringer, J. R. (2007). Population structure, physiology and ecohydrological impacts of dioecious riparian tree species of western North America. *Oecologia*, *154*(1), 85–93. https://doi.org/10.1007/s00442-007-0813-0

Iszkuło, G., & Boratyński, A. (2011). Initial period of sexual maturity determines the greater growth rate of male over female in the dioecious tree *Juniperus communis* subsp. *Communis*. *Acta Oecologica*, *37*(2), 99–102. https://doi.org/10.1016/j.actao.2011.01.001

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*,

473(7345), 97–100. https://doi.org/10.1038/nature09916

Jones, B. E., Rickman, T. H., Vazquez, A., Sado, Y., & Tate, K. W. (2005). Removal of encroaching conifers to regenerate degraded aspen stands in the Sierra Nevada. *Restoration Ecology*, *13*(2), 373–379. https://doi.org/10.1111/j.1526-100X.2005.00046.x

Kira, T. (1991). Forest ecosystems of east and southeast Asia in a global perspective. *Ecological Research*, *6*, 185–200.

Krasnow, K. D., & Stephens, S. L. (2015). Evolving paradigms of aspen ecology and management: Impacts of stand condition and fire severity on vegetation dynamics. *Ecosphere*, *6*(1), 1–16. https://doi.org/10.1890/ES14-00354.1

Kreider, M. R., & Yocom, L. L. (2021). Aspen seedling establishment, survival, and growth following a high-severity wildfire. *Forest Ecology and Management*, *493*, 119248. https://doi.org/10.1016/j.foreco.2021.119248

Landhäusser, S. M., Pinno, B. D., & Mock, K. E. (2019). Tamm Review: Seedling-based ecology, management, and restoration in aspen (*Populus tremuloides*). *Forest Ecology and Management*, *432*, 231–245. https://doi.org/10.1016/j.foreco.2018.09.024

Latva-Karjanmaa, T., Suvanto, L., Leinonen, K., & Rita, H. (2003). Emergence and survival of *Populus tremula* seedlings under varying moisture conditions. *Canadian Journal of Forest Research*, *33*(11), 2081–2088. https://doi.org/10.1139/x03-129

Ledig, F. T., Rehfeldt, G. E., Sáenz-Romero, C., & Flores-López, C. (2010). Projections of suitable habitat for rare species under global warming scenarios. *American*

Journal of Botany, *97*(6), 970–987. https://doi.org/10.3732/ajb.0900329

Levin, D. A. (1983). Polyploidy and novelty in flowering plants. *The American Naturalist*, *122*(1), 1–25. https://doi.org/10.1086/284115

Li, C., Xu, G., Zang, R., Korpelainen, H., & Berninger, F. (2007). Sex-related differences in leaf morphological and physiological responses in *Hippophae rhamnoides* along an altitudinal gradient. *Tree Physiology*, *27*(3), 399–406. https://doi.org/10.1093/treephys/27.3.399

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lin, Y.-C., Wang, J., Delhomme, N., Schiffthaler, B., Sundström, G., Zuccolo, A., Nystedt, B., Hvidsten, T. R., de la Torre, A., Cossu, R. M., Hoeppner, M. P., Lantz, H., Scofield, D. G., Zamani, N., Johansson, A., Mannapperuma, C., Robinson, K. M., Mähler, N., Leitch, I. J., … Street, N. R. (2018). Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proceedings of the National Academy of Sciences*, *115*(46), E10970–E10978. https://doi.org/10.1073/pnas.1801437115

Little, E. L. (1971). *Atlas of United States trees* (Vol. 1). U.S. Dept. of Agriculture, Forest Service. https://doi.org/10.5962/bhl.title.130546

Love, D. M., Venturas, M. D., Sperry, J. S., Brooks, P. D., Pettit, J. L., Wang, Y., Anderegg, W. R. L., Tai, X., & Mackay, D. S. (2019). Dependence of aspen stands on a subsurface water subsidy: Implications for climate change impacts.

Water Resources Research, *55*(3), 1833–1848. https://doi.org/10.1029/2018WR023468

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Mason, A. S., Nelson, M. N., Yan, G., & Cowling, W. A. (2011). Production of viable male unreduced gametes in *Brassica* interspecific hybrids is genotype specific and stimulated by cold temperatures. *BMC Plant Biology*, *11*(1), 103. https://doi.org/10.1186/1471-2229-11-103

McIntyre, P. J. (2012). Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *American Journal of Botany*, *99*(4), 655–662. https://doi.org/10.3732/ajb.1100466

Mills, T. R., Rumble, M. A., & Flake, L. D. (2000). Habitat of birds in ponderosa pine and aspen/birch forest in the Black Hills, South Dakota. *Journal of Field Ornithology*, *71*(2), 187–206. https://doi.org/10.1648/0273-8570-71.2.187

Mitton, J. B., & Grant, M. C. (1996). Genetic variation and the natural history of Quaking Aspen. *BioScience*, *46*(1), 25–31. https://doi.org/10.2307/1312652

Mock, K. E., Callahan, C. M., Islam-Faridi, M. N., Shaw, J. D., Rai, H. S., Sanderson, S. C., Rowe, C. A., Ryel, R. J., Madritch, M. D., Gardner, R. S., & Wolf, P. G. (2012). Widespread triploidy in Western North American aspen (*Populus tremuloides*). *PLoS ONE*, *7*(10), e48406. https://doi.org/10.1371/journal.pone.0048406

Mock, K. E., Rowe, C. A., Hooten, M. B., Dewoody, J., & Hipkins, V. D. (2008). Clonal

dynamics in western North American aspen (*Populus tremuloides*). *Molecular Ecology*, *17*(22), 4827–4844. https://doi.org/10.1111/j.1365-294X.2008.03963.x

Mueggler, W. F. (1985). In *Aspen: Ecology and management in the Western United States* (pp. 129–134). U.S. Dept. of Agriculture, Forest Service, Rocky Mountain Forest and Range.

Nuñez, C. I., Nuñez, M. A., & Kitzberger, T. (2008). Sex-related spatial segregation and growth in a dioecious conifer along environmental gradients in northwestern Patagonia. *Ecoscience*, *15*(1), 73–80. https://doi.org/10.2980/1195-6860(2008)15[73:SSSAGI]2.0.CO;2

Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, *34*(1), 401–437. https://doi.org/10.1146/annurev.genet.34.1.401

Pakull, B., Groppe, K., Mecucci, F., Gaudet, M., Sabatti, M., & Fladung, M. (2011). Genetic mapping of linkage group XIX and identification of sex-linked SSR markers in a *Populus tremula* × *Populus tremuloides* cross. *Canadian Journal of Forest Research*, *41*(2), 245–253. https://doi.org/10.1139/X10-206

Pakull, B., Kersten, B., Lüneburg, J., & Fladung, M. (2015). A simple PCR-based marker to determine sex in aspen. *Plant Biology*, *17*(1), 256–261. https://doi.org/10.1111/plb.12217

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine: Association mapping of serotiny. *Molecular Ecology*, *21*(12), 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Parisod, C., Holderegger, R., & Brochmann, C. (2010). Evolutionary consequences of

autopolyploidy: Research review. *New Phytologist*, *186*(1), 5–17. https://doi.org/10.1111/j.1469-8137.2009.03142.x

PRISM Climate Group. (2022). *PRISM Climate Group, OSU*. https://www.prism.oregonstate.edu/

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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

Rehfeldt, G. E., Ferguson, D. E., & Crookston, N. L. (2009). Aspen, climate, and sudden decline in western USA. *Forest Ecology and Management*, *258*(11), 2353–2364. https://doi.org/10.1016/j.foreco.2009.06.005

Rowe, C. (2019). *vcf2Jaccard* [HTML]. https://github.com/carol-rowe666/vcf2Jaccard (Original work published 2019)

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., de Pamphilis, C. W., Wall, P. K., & Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, *96*(1), 336–348. https://doi.org/10.3732/ajb.0800079

Stevens-Rumann, C. S., Kemp, K. B., Higuera, P. E., Harvey, B. J., Rother, M. T., Donato, D. C., Morgan, P., & Veblen, T. T. (2018). Evidence for declining forest resilience to wildfires under climate change. *Ecology Letters*, *21*(2), 243–252. https://doi.org/10.1111/ele.12889

Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E.,

Brown, J. W., Sessa, E. B., & 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

U.S. Geological Survey. (2021). [dataset]. Digital Elevation Models. 1/3 Arc Second.

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

Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., & Wood, E. F. (2020). Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sensing of Environment*, *242*, 111740. https://doi.org/10.1016/j.rse.2020.111740

Wang, X., Cheng, Z.-M., Zhi, S., & Xu, F. (2016). Breeding triploid plants: A review. *Czech Journal of Genetics and Plant Breeding*, *52*(2), 41–54. https://doi.org/10.17221/151/2015-CJGPB

Ward, J. K., Dawson, T. E., & Ehleringer, J. R. (2002). Responses of *Acer negundo* genders to interannual differences in water availability determined from carbon isotope ratios of tree ring cellulose. *Tree Physiology*, *22*(5), 339–346. https://doi.org/10.1093/treephys/22.5.339

Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., & Burbano, H. A. (2018). nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*, *19*(1), 122. https://doi.org/10.1186/s12859-018-2128-z

Worrall, J. J., Rehfeldt, G. E., Hamann, A., Hogg, E. H., Marchetti, S. B., Michaelian,

M., & Gray, L. K. (2013). Recent declines of *Populus tremuloides* in North America linked to climate. *Forest Ecology and Management*, *299*, 35–51. https://doi.org/10.1016/j.foreco.2012.12.033

APPENDICES

Appendix A. Raw Sequence Data & Supplementary File Accessibility

Raw Sequence Data

Raw Sequence data can be found at the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/), BioProject 'PRJNA1063753', Sequence Read Archive (SRA) accessions 'SRR27495551 to SRR27496309' & 'SRR27500821 to SRR27501579', BioSample accessions 'SAMN39407291 to SAMN39408049' & 'SAMN39413581 to SAMN39414339'. Sequencing data will be released January 1, 2025. Access to data via reviewer link can be made available upon request until date of release.

Additional scripts related to the processing of this data can be made available upon request to the Author.

Supplementary Files

Supplementary File 1 - Barcodes and Sequencing Runs

Microsoft Office Excel Document with a list of samples included in this project, a list of the libraries sequenced in each lane of the Illumina™ HiSeq 2500 Platform lanes, and barcode indices used for each unique sequenced sample.

Spreadsheet 'ProjectSampleList':

This spreadsheet contains a list of aspen samples included in this project prior to quality control filtering.

Table A1 Supplementary File 1, spreadsheet 'ProjectSampleList' field definitions

Field	Definition
Unique ID	Unique laboratory accession number assigned to field collected samples as maintained by the Utah State University Molecular Ecology Laboratory, Wildland Resources Department. Also known as 'LABID'.
Sequencing ID	Unique sample identifiers as run in ddRAD-Seq libraries.
Site	Identifies which 30 km x 30 km site the sample belongs to
Other ID	Additional identifying information, useful for laboratory lookup when 'Tag' is not available.
Tag	Unique aluminum tag identifier placed on tree during leaf tissue sample collection.

Spreadsheet 'Sequencing Runs':

This spreadsheet contains a list of the Illumina HiSeq2500 100bp single-end read runs performed to sequence this project. Nine runs were completed. Four plates, per Illumina run, containing uniquely barcoded samples, were run in two separate sequencing runs, i.e., samples from plates 1-4 were run two times 'BP0104a' and 'BP0104b' and contained identical sample names. Plates 17 and 18 were run in a single Illumina run due to the presence of two plates of samples instead of four.

Sequencing runs can be coordinated with index spreadsheets 'BP0104a&b', 'BP0508a&b', 'BP0912a&b', 'BP1316a&b', and 'BP1718a' to identify barcodes used for each sample. Note that Illumina runs contained additional samples that were not part of this project and the 'ProjectSampleList' spreadsheet should be consulted. These spreadsheets are informational in purpose as raw sequence data available at NCBI have been demultiplexed and adaptor bases removed (See 'Raw Sequence Data heading in Appendix A).

Table A2. Supplementary File 1, spreadsheet 'Sequencing Runs' field definitions.

Field	Definition
Library	Index of libraries sequenced.
Library Plate Numbers	Identifies the plate numbers sequenced in each ddRAD-Seq library.
Sequencing Run	Unique identifier for each individual sequencing run.
Sequencing Run Filename	Individual sequencing run filename.

Spreadsheet BP0104a&b', 'BP0508a&b', 'BP0912a&b', 'BP1316a&b', and 'BP1718a':

This series of spreadsheets lists the unique 8-10 base-pair barcodes that differed by 4 base pairs between barcodes in each sequencing run.

Table A3. Supplementary File 1, spreadsheets *BP0104a&b', 'BP0508a&b', 'BP0912a&b', 'BP1316a&b', and 'BP1718a'* field definitions.

Field	Definition
Sequencing ID	Unique sample identifiers as run in ddRAD-Seq libraries.
Barcode	Unique 8-10 base-pair barcodes associated with each sequenced sample.

This file is accessible online at https://doi.org/10.26078/k7hg-st08. Filename: Supplementary_File_1.xlsx

Supplementary File 2 - Clonal and Cytotype Assignment Results and Sample Information

Microsoft Office Excel Document results of cytotype assignment, clonal assignment, sample grouping information (SITE, LATGRP, ECOGRP), genetic sex identification, and fuzzed geographic coordinates.

Table A4. Supplementary File 2 field definitions**.**

Field	Definition
LABID	Unique laboratory accession number assigned to field collected samples as maintained by the Utah State University Molecular Ecology Laboratory, Wildland Resources Department.
SITE	Identifies which 30 km x 30 km site the sample belongs to.
CYTO	Ploidy level or cytotype the sample was assigned as part of this study.
SEXID	Genetic sex ID assigned as part of this work and Bidner (2021).
ECOGRP	Assigned ecoregion group of each sample as described in Methods, Figure 3.
LATGRP	Assigned latitude group of each sample as described in Methods, Figure 2.
x_fuz	Longitudinal coordinates fuzzed to approximately 1km.
y_fuz	Latitudinal coordinates fuzzed to approximately 1km.

This file is accessible online at https://doi.org/10.26078/k7hg-st08. Filename: Supplementary_File_2.xlsx

Supplementary File 3 - Environmental Variable Correlation Plots

Adobe Acrobat PDF document containing pairwise correlation plots for each environmental covariate considered in this study and listed in Table 3 of the main text. Correlation plots for each sample include Pearson's correlation coefficient (r) and p-value. For further visual representation, stronger negative correlations are a darker blue, while stronger positive correlations are represented in darker red, both of which fade to white as correlations decrease.

This file is accessible online at https://doi.org/10.26078/k7hg-st08. Filename: Supplementary_File_3.pdf

Supplementary File 4 - Jaccard Correlation Coefficient Histograms by Site

Adobe Acrobat PDF document containing histograms of Jaccard Correlation coefficients for each site in the Intermountain West study area. Histograms include pairwise comparisons within and between the identified site. An additional companion

histogram showing the location of the Jaccard Correlation coefficient cutoff at values greater than the 0.8975 cutoff is also included. See Figure C4 for a histogram containing all pairwise comparisons. The histogram on the left is coded by cytotype pairs, i.e., both samples in the pair are diploid (2x) cytotype, both samples in the pair are triploid (3x) cytotype, and samples pairs that contain one of each diploid and triploid (2x & 3x) sample types.

This file is accessible online at https://doi.org/10.26078/k7hg-st08. Filename: Supplementary_File_4.pdf

Appendix B. Ploidy Level Classification Comparison

After initially using gbs2ploidy (Gompert & Mock, 2017), it was brought to our attention that gbs2ploidy was not assigning ploidy level accurately to datasets which covered North America from Mexico to Canada. As gbs2ploidy was developed using samples from a limited geographic extent (Utah and Colorado) that a bias may be present when it is applied to a larger or more geographically diverse sample set in which heterozygosity estimates are more variable. Further, it was found that Fastploidy (Goessen et al., 2022) calls showed more accuracy with microsatellites and flow cytometry data when the entire geographic range was included over gbs2ploidy, but that gbs2ploidy calls improved when certain samples groups, i.e. Mexico populations were left out. This may be a concern because of gbs2ploidy's inclusion of genome-wide heterozygosity estimates. To alleviate this concern, we compared and analyzed three separate ploidy classifiers; nQuire (Weiß et al. 2018), FastPloidy (Goessen et al. 2022), and gbs2ploidy. Due to the concern of gbs2ploidy's effect on geographic range or sample size, it was also analyzed based on geographic subsets from individual sites, states, and the Intermountain West.

Each of the ploidy level callers uses different approaches when classifying chromosome counts. gbs2ploidy employs a Bayesian framework to infer allelic ratios and allelic proportions. Further, it includes an estimate of genome-wide heterozygosity, a measure we adjust in this analysis. Fastploidy extracts the allelic depth and calculates their proportions, fitting them into defined ranges (0.273-0.939, 0.440-0.560, and 0.607-0.727), assigning triploids to the highest and lowest ranges and diploids to the middle range. Beyond allelic ratios, it does not, however, apply a statistical inference approach.

Both gbs2ploidoy and Fastploidy are based on using a filtered variant calling file (VCF) to extract allelic information. The third ploidy level classifier used was nQuire, which uses a Gaussian Mixture Model to model base frequencies of variable sites and maximum likelihood to select the most plausible ploidy level model. nQuire bases its calling structure on sorted BAM files for analysis instead of VCF outputs.

To better get an understanding of the differences between models and their ability to call ploidy levels across geographic ranges we took many approaches. First, we selected 35 samples, approximately 2 from each of 31 sites where gbs2ploidy and Fastploidy disagreed in calls. These variables were then run at 12 variable microsatellite loci (Mock et al., 2012). If a sample had 2 or more loci with 3 alleles it was assigned as a triploid. Samples with 1 or less loci with 2 or less alleles were assigned diploid. Second, gbs2ploidy was run in the following variations a) with all Intermountain West samples, b) with samples grouped by state, c) with samples grouped by site, d) with heterozygosity set the same across samples (0.5), and finally, e) with no heterozygosity. Third, Fastploidy and nQuire were run using the entire dataset presented in this study.

When comparing the results from the samples run with microsatellites gbs2ploidy matched with 28 of the 35 samples that were analyzed. nQuire and Fastploidy matched with 15 and 11 microsatellite calls, respectively. When comparing the best-performing model from each of the ploidy level callers on the sample-set from the Intermountain West, 969 samples matched across all 3 ploidy level callers, 49 samples matched between gbs2ploidy and nQuire, no additional samples matched with Fastploidy. Of the remaining 398 samples examined, gbs2ploidy made an additional 290 diploid or triploid calls, all remaining (n=108) were called either as ambiguous (no clear diploid or triploid

proportion) or were dropped due to low number of heterozygous alleles. Calls for each model can be found in the Supplementary File 2, Ploidy Level Identification, Ploidy Model Comparison.

Overall, there is good consensus between ploidy level callers and a small set of samples may be causing the differences in calls between them. Generally, those that had matching calls between 2 or 3 models had higher heterozygous allele counts (n > 1000 SNPs, Figure B1). This gives confidence that samples matching between all 3 callers have less false positive calls.

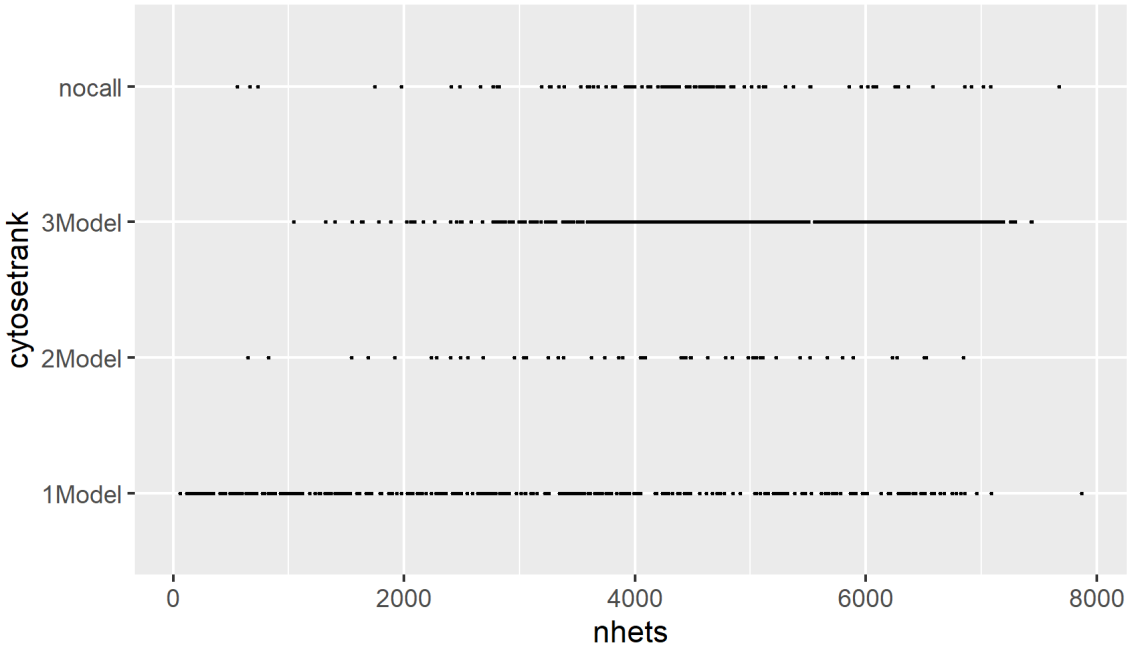

Figure B1. Number of heterozygous alleles by the number of models in consensus across all samples in the Intermountain West presented in this study.

Final calls for this project were ultimately made using gbs2ploidy and removing samples with <1000 heterozygous alleles. We believe this reduced any error in ploidy level calls to acceptable levels associated with this particular dataset. While nQuire and

Fastploidy caller did not perform as well on this dataset we believe this to be a characteristic of this data set in particular, where gbs2ploidy may be better suited to make calls based on as little as ~3x SNP depth while nQuire and Fastploidy may require as much as ~10x SNP depth to increase accuracy of calls.

Appendix C. Supporting Tables and Figures

Table C1. Variable Importance R^2(Adj) and modeled AIC value. Results from GLMM modeling of individual variables using ploidy level as a binomial response and site as a random intercept in order to aid in selection of appropriate variables in the final model.

Variable	AIC	Conditional R^2(Adj)	Marginal R^2(Adj)
ELEV	1437.82	0.40745	0.13153
SLOPE	1452.05	0.38030	0.01185
SL	1455.91	0.36525	0.00037
Trng	1454.81	0.36148	0.00482
TmaxGS	1456.00	0.36510	0.00016
PPTGS	1455.50	0.37439	0.00299
VPDmax	1455.98	0.36533	0.00021
RHGS	1455.91	0.36354	0.00059
DD5	1455.97	0.36575	0.00027
GSPDD5	1456.01	0.36663	0.00016
ADI	1455.45	0.36425	0.00198
PRATIO	1453.54	0.38049	0.01131
SMOGS	1449.98	0.39446	0.02071
Tmax	1456.04	0.36548	0.00000
Tmin	1455.13	0.36995	0.00379
Tmean	1455.79	0.36872	0.00112
DEW	1455.65	0.37036	0.00180
MAP	1454.12	0.36423	0.00602

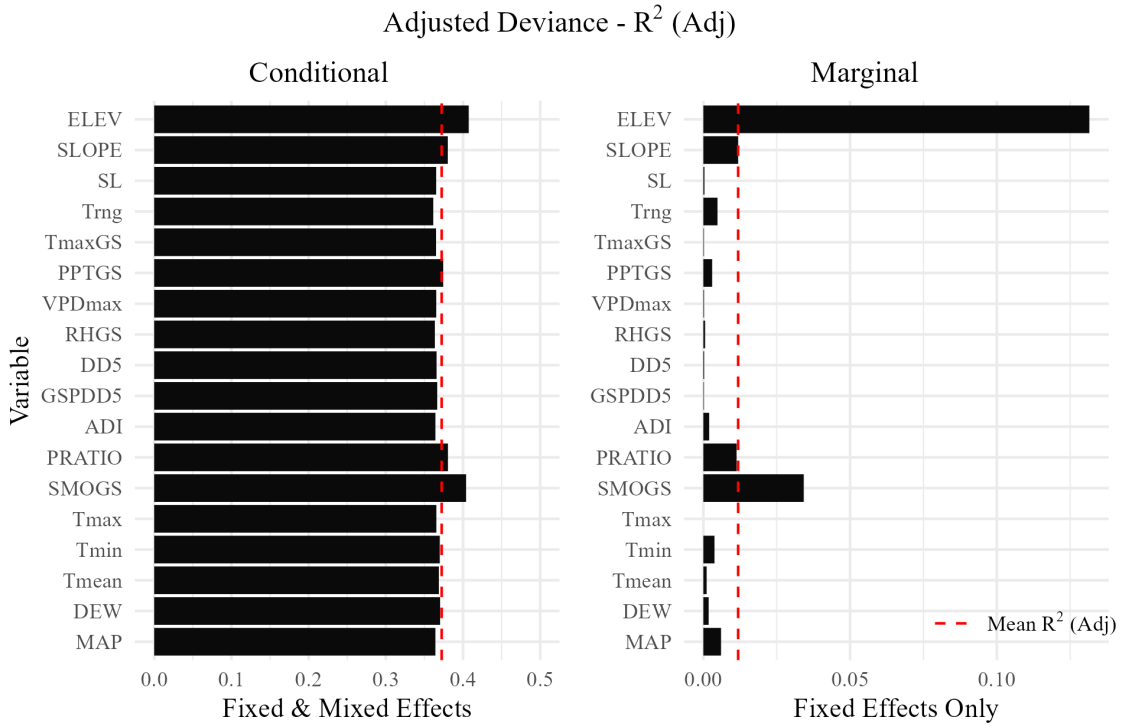

Figure C1. Variable importance by conditional and marginal R^2(Adj). Mean R^2(Adj) shown with dashed red line for reference. The marginal R^2(Adj) considers only the variance of the fixed effects (without the random effects), while the conditional R^2(Adj) takes both fixed and random effects into account.

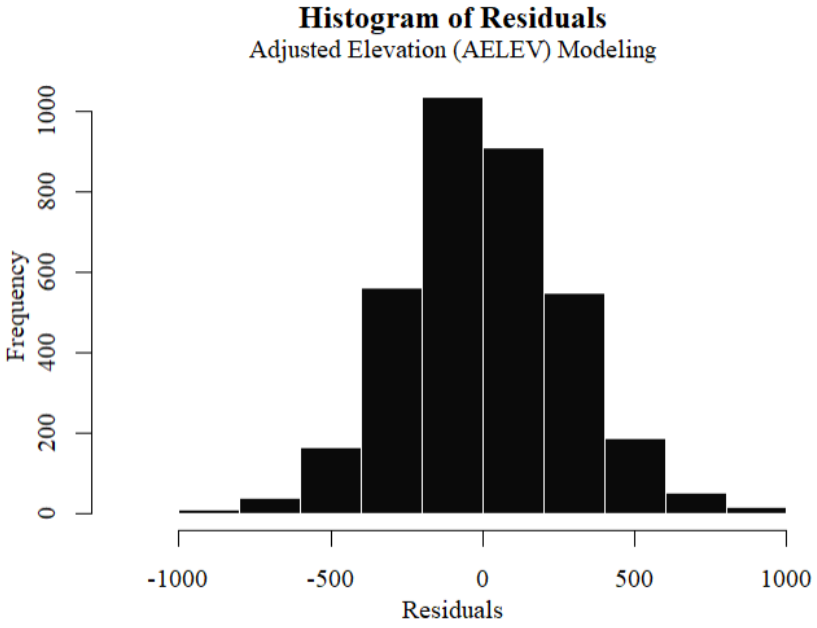

Figure C2. Modeling residuals resulting from elevation adjustment (AELEV) based on linear regression model 2 (Table 4).

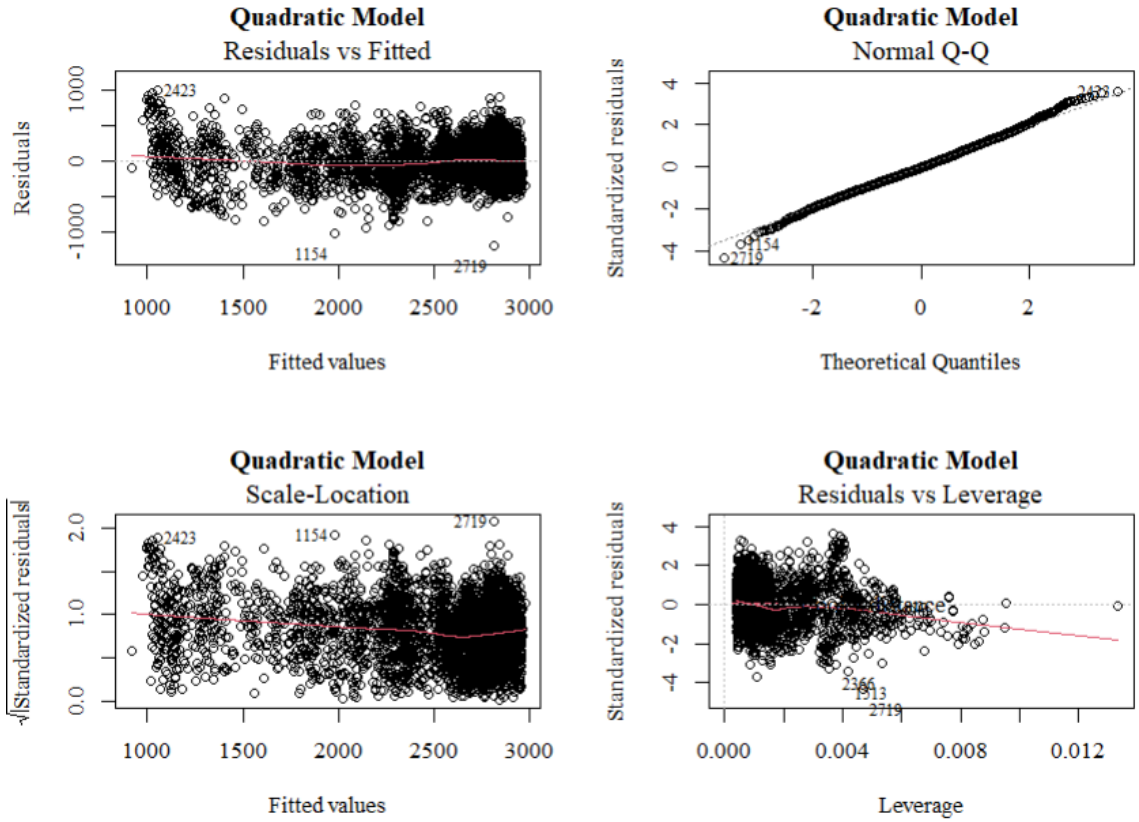

Figure C3. Modeling residuals resulting from elevation adjustment (AELEV) based on linear regression model 3 (Table 4).

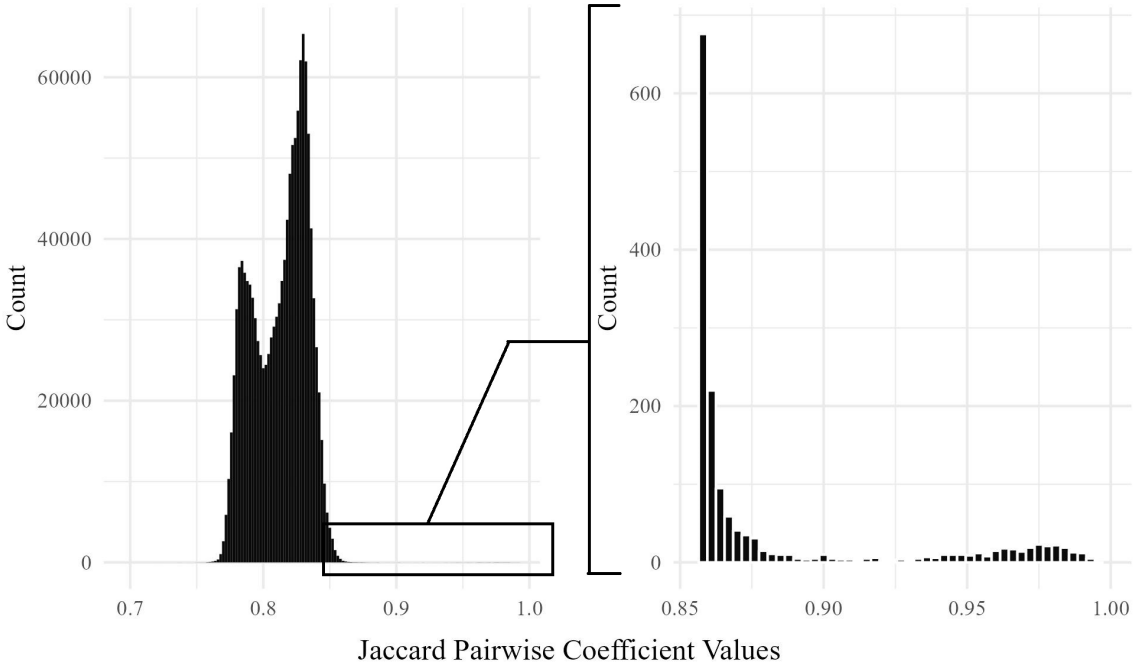

Figure C4. Pairwise Jaccard Coefficient Values. All pairwise Jaccard values are shown in the left graph. Typically, two distinct peaks are present, one shown centered around 0.8 (as seen here) and another at approximately 0.95 or greater. As samples in this study were collected to avoid clonal replicates the second peak is very small, as shown in the graph on the right, though some clones were identified.

Appendix D. Methods for Environmental Variable Generation

Climate Variables

Climate Variables (Table 3) were either used directly from PRISM Climate Group (PRISM Climate Group, 2022) or calculated from the monthly climate measures provided. Data provided are 30-year normals (1991-2020) at a spatial resolution of 800 m. PRISM normals are based on geographic location using a digital elevation model as a predictor grid. With an average nearest-neighbor distance of 700 m between tree samples we found the 800m climate data to be an acceptable resolution to minimize sampling location from falling within identical pixel values. If samples fell within the same pixel samples were randomly removed until a single sample remained in the pixel.

Climate variables calculated from this data include the following:

Maximum Temperature (Tmax) and **Minimum Temperature** (Tmin) were calculated by averaging either the maximum or minimum temperature value across all months of the year. The **Growing Season Maximum Temperature** (TmaxGS) for the months of April to September was calculated by finding the maximum value in those months. **Mean Temperature** (Tmean) was calculated by taking the mean temperature value for all months. All temperature measures are in degrees °C.

The **Annual Temperature Range** (Trng) is calculated from the yearly average maximum (Tmax) and minimum (Tmin) data values. May also be referred to as TrngAnn.

Mean Annual Precipitation (MAP or PPTAnn) was calculated by summing precipitation values over all months of the year while **Growing Season Precipitation**

(PPTGS or GSP) was calculated by summing monthly precipitation values during the growing season (April to September).

For **Maximum Vapor Pressure Deficit** (VPDmax) and **Mean Dewpoint** (DEW), the mean value as provided in the climate data set, was taken across all months of the year.

Relative Humidity during the growing season (RHGS) was calculated using the R (R Core Team, 2021) 'humidity' package (Cai, 2019) based on Tmean and DEW values for the months of April to September.

Growing degree days (DD5) represents yearly cumulative warmth as an index, representing a collection location over 5°C for the year and is used as a minimum threshold of plant productivity. Calculation of DD5 was completed as in Greer ((Greer et al., 2016), Supporting Information) and modified from Kira (1991) otherwise known as the 'warmth index'.

Variables expressing interactions between precipitation and growing degree days such as **GSPDD5** and an **Annual Dryness Index** (ADI) were also included.

$$GSPDD5 = (GSP \times DD5) / 1000$$

$$ADI = (DD5)^{0.5} / MAP$$

Additionally, a variable that reflects the periodicity of precipitation was also included. Aspen occurrence has been documented to be most common between a PRATIO of 0.4 and 0.6, and it is hypothesized that it may also be a measure related to moisture stress (Rehfeldt et al., 2009).

$$PRATIO = GSP / MAP$$

Topographic Variables

Topographic variables were based on digital elevation models procured from the U.S. Geological Survey 3D Elevation Program (*U.S. Geological Survey*, 2021). 1/3 Arc Second (Approximately 10 m resolution) digital elevation model (DEM) tiles (1 Degree x 1 Degree) were downloaded for each geographic area containing at least 1 sampling location. Tiles were then projected to match sampling locations and merged using ArcGIS Pro (ESRI, 2022). ArcGIS Pro tools were used to extract **elevation** (ELEV) and calculate **slope** (SLOPE, degrees) and aspect (degrees). From aspect, **southerlyness** (SL) was calculated as

$$sin((((aspect\ +\ 180)\ *\pi)\ /\ 180)).$$

Soil Moisture Variable

Soil moisture (SMO) values were derived from SMAP HydroBlocks, a 30 m resolution dataset for the conterminous U.S. which uses scalable cluster-based merging scheme that combines high-resolution land surface modeling, radiative transfer modeling, machine learning, SMAP microwave data, and in situ observation (Vergopolan et al., 2020). Monthly soil moisture values were extracted from 2015 - 2019 and represent the first 5 cm of soil depth. The Soil Moisture Growing Season (SMOGS) variable was calculated by averaging soil moisture values across years by month and then a mean value being calculated from months of April to September.