

DEEP SEQUENCING OF AMPLICONS REVEALS WIDESPREAD INTRASPECIFIC HYBRIDIZATION AND MULTIPLE ORIGINS OF POLYPLOIDY IN BIG SAGEBRUSH (*ARTEMISIA TRIDENTATA*; ASTERACEAE)¹

BRYCE A. RICHARDSON^{2,5}, JUSTIN T. PAGE³, PRABIN BAJGAIN^{3,4}, STEWART C. SANDERSON²,
AND JOSHUA A. UDALL³

²Rocky Mountain Research Station, USDA Forest Service, Provo, Utah 84606 USA; ³Plant and Wildlife Science Department, Brigham Young University, Provo, Utah 84602 USA; and ⁴Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota 55108 USA

- **Premise of the study:** Hybridization has played an important role in the evolution and ecological adaptation of diploid and polyploid plants. *Artemisia tridentata* (Asteraceae) tetraploids are extremely widespread and of great ecological importance. These tetraploids are often taxonomically identified as *A. tridentata* subsp. *wyomingensis* or as autotetraploids of diploid subspecies *tridentata* and *vaseyana*. Few details are available as to how these tetraploids are formed or how they are related to diploid subspecies.
- **Methods:** We used amplicon sequencing to assess phylogenetic relationships among three recognized subspecies: *tridentata*, *vaseyana*, and *wyomingensis*. DNA sequence data from putative genes were pyrosequenced and assembled from 329 samples. Nucleotide diversity and putative haplotypes were estimated from the high-read coverage. Phylogenies were constructed from Bayesian coalescence and neighbor-net network analyses.
- **Key results:** Analyses support distinct diploid subspecies of *tridentata* and *vaseyana* in spite of known hybridization in ecotones. Nucleotide diversity estimates of populations compared to the total diversity indicate the relationships are predominately driven by a small proportion of the amplicons. Tetraploids, including subspecies *wyomingensis*, are polyphyletic occurring within and between diploid subspecies groups.
- **Conclusions:** *Artemisia tridentata* is a species comprising phylogenetically distinct diploid progenitors and a tetraploid complex with varying degrees of phylogenetic and morphological affinities to the diploid subspecies. These analyses suggest tetraploids are formed locally or regionally from diploid *tridentata* and *vaseyana* populations via autotetraploidy, followed by introgression between tetraploid groups. Understanding the phylogenetic vs. ecological relationships of *A. tridentata* subspecies will have bearing on how to restore these desert ecosystems.

Key words: *Artemisia tridentata*; Asteraceae; gene discordance; haplotype clustering; networks; phylogeny; pyrosequencing; reticulate evolution.

Hybridization is a pervasive evolutionary mechanism in ancient and recent plant lineages. Hybridization of different taxa can result in an increase of chromosome number (allopolyploidy) or in a chromosome number identical to the parents (homoploidy). The former, allopolyploidy, has been shown to be most predominant in plant speciation because the offspring can often maintain their genetic identity due to genetic incompatibility from the differences in chromosome numbers (Rieseberg and Willis, 2007). However, genome doubling within a taxon (autopolyploidy) followed by interspecific hybridization has largely gone unrecognized as an important speciation mechanism until recently (Soltis et al., 2007). Among desert and semidesert

shrubs the rate of polyploidization is likely much higher. For many desert shrub species, polyploids often dominate the more extreme environments, which can be a large portion of the geographic distribution (McArthur et al., 1981). However, an understanding of the evolutionary processes involved in the formation of polyploid lineages is often lacking. For many of these shrub taxa, allopolyploidy, autopolyploidy, or a combination of both is believed to play a major role in polyploid formation (McArthur and Sanderson, 1999; Sanderson and Stutz, 2001).

Artemisia is a large genus in the Asteraceae, tribe Anthemideae, in which polyploidy is found in over half of the species (Vallès et al., 2011). The earliest members of the genus are believed to have origins in central Asia during the late Eocene based on pollen data (Yunfa et al., 2011). *Artemisia* has subsequently spread to Europe and North America. Speciation in *Artemisia* has resulted in approximately 600 species of woody shrubs and perennial and annual herbs, which are found on all continents except Antarctica (Vallès et al., 2011). They are most abundant in the circumpolar Arctic and cold deserts of Asia and North America (Vallès and McArthur, 2001). Phylogenetic assessments of *Artemisia* have largely focused on broad taxonomic relationships (i.e., species to subgenera). These studies have shown divisions between clades corresponding to continental

¹Manuscript received 18 July 2012; revision accepted 22 October 2012.

The authors thank the numerous volunteers from several state and federal agencies that assisted in collection of plant materials and Drs. E. D. McArthur and K. E. Mock for their thoughtful review of this manuscript. This research was funded by USDA Forest Service, National Fire Plan (#2012-NFP-GSD-1), BYU Mentored Environment Grant, and the Great Basin Native Plant Selection and Increase Project (GBNPSIP).

⁵Author for correspondence (e-mail: brichardson02@fs.fed.us)

barriers, especially between North America and Eurasia (Watson et al., 2002; Tkach et al., 2008; Garcia et al., 2011). For example, subgenus *Tridentatae*, the sagebrushes, is found solely in North America (McArthur et al., 1981). It is hypothesized that this subgenus formed after Asian *Artemisia* species colonized North America via the Beringia land bridge (McArthur et al., 1981; Garcia et al., 2011). However, for some *Artemisia* species, Beringia may have provided multiple episodes of colonization between Asia and North America (Riggins and Seigler, 2012).

The sagebrushes of western North America are most prevalent in the cold, arid intermountain regions of the western United States where the most abundant and widespread species is big sagebrush (*A. tridentata* Nutt.; Asteraceae). This species extends from the northern Baja California of Mexico to southern British Columbia of Canada. Over much of this range *A. tridentata* is a landscape-dominant species fostering plant and animal biodiversity. However, a large portion of these ecosystems have been lost to urbanization and conversion to agriculture or severely degraded from increased frequencies of disturbances such as wildfire, followed by displacement by exotic annual grasses. This species provides important forage and vegetative cover for wildlife (Wambolt, 1996; Gregg et al., 1994) including the threatened greater sage-grouse (*Centrocercus urophasianus*) (Connelly et al., 2000). Thus, *A. tridentata* is a critical component in many of the seed mixes for ecological restoration in these desert ecosystems. In years with large and frequent wildfires, state and federal land management agencies request large amounts of native seeds to meet the demand for postfire restoration projects. In 2012, *A. tridentata* seed requests will exceed 1 million pounds (P. Krabacher, Bureau of Land Management, personal communication). There are numerous challenges associated with restoration of *A. tridentata* ecosystems (Shaw et al., 2005; Pyke, 2011); however, the first step is ensuring the seed is correctly identified to the subspecies and ploidy levels as discussed later. The foundation for diagnosing subspecies lies in understanding their evolutionary relationships.

Artemisia tridentata is highly variable in morphology and has been divided into three major subspecies. While some disagreement exists on the taxonomy of relatively geographically restricted varieties or subspecies [e.g., *Artemisia tridentata* subsp. *parishii* (A. Gray) H. M. Hall & Clements and others], three predominant subspecies are widely recognized: *Artemisia tridentata tridentata*, *Artemisia tridentata vaseyana* (Rydb.) Beetle, and *Artemisia tridentata wyomingensis* Beetle and Young (Goodrich et al., 1985; Shultz, 2009). These subspecies are the focus of this study. With the exception of subsp. *wyomingensis*, which is exclusively tetraploid, subsp. *tridentata* and *vaseyana* have both diploid and tetraploid cytotypes. Each of the three subspecies occupies a distinct environment. Subspecies *tridentata* (known as the basin ecotype) occupies alluvial flats at elevations typically below 1800 m a.s.l. Subspecies *vaseyana* (the mountain ecotype) occupies elevations from 1660 m a.s.l. to timberline. Subspecies *wyomingensis* occupies elevations suitable for either subsp. *tridentata* or subsp. *vaseyana*, but the environments tend to have relatively lower annual precipitation and shallow soils (McArthur et al., 1988; Kolb and Sperry, 1999). However, in zones or ecotones between mountain and basin habitats, homoploid hybrid swarms form between subsp. *tridentata* and *vaseyana* (McArthur et al., 1988; Freeman et al., 1991). Subsequent reciprocal transplant studies along the ecotonal gradients have shown that hybrids between

mountain and basin ecotypes display higher fitness in germination and survivorship within the ecotone, but lower fitness in the parental habitats (Wang et al., 1997; Rodriguez et al., 2009). Beyond the fitness of F_1 hybrids, it is unknown how hybridization within ecotones influences the genetic structure between *A. tridentata* subspecies.

In addition to homoploid hybridization between subspecies, tetraploidy has played a critical role in the landscape dominance of big sagebrush. Entire landscapes of big sagebrush populations are composed of tetraploid plants. The genetic mechanism(s) of polyploid formation in big sagebrush is unknown, although it has been speculated that unreduced gametes within or between diploids of subsp. *vaseyana* and *tridentata* could produce autotetraploid and allotetraploid progeny (McArthur et al., 1981; Ramsey and Schemske, 1998; Van de Peer et al., 2009). Approximately half of all big sagebrush samples collected throughout the geographic range ($N = 1103$) were identified as tetraploid (McArthur et al., 1981; McArthur and Sanderson, 1999), emphasizing the important role polyploids play in this species' geographic distribution and landscape dominance, as has been investigated in other species (Meimberg et al., 2009). It is hypothesized that the frequency of tetraploids in the overall sagebrush population has increased during the Holocene warming as aridity has increased (McArthur et al., 1981; Vallès et al., 2011). Similarly, differential selection of ploidy levels due to the contemporary climate or other environmental variables has been shown in several plant species (Burton and Husband, 1999; Nuismer and Cunningham, 2005; Ramsey, 2011).

Knowledge of the evolutionary relationships between diploid and tetraploid populations in big sagebrush is lacking. Cytogeographic and morphological analyses suggest tetraploids could form from within subsp. *tridentata* or *vaseyana* or between subsp. *tridentata* and *vaseyana* (McArthur and Sanderson, 1999). Previous phylogenetic studies of interspecific relationships between big sagebrush and sister taxa (e.g., *A. arbuscula* Nutt., *A. cana* Pursh, and *A. nova* A. Nelson) or between big sagebrush subspecies have not resolved any consistent pattern with the morphology-based taxonomy (Stanton et al., 2002; Garcia et al., 2011). Because of the phylogenetic and morphological incongruence, some authors have proposed interspecific hybridization could be the cause for a lack of phylogenetic signal between species (Garcia et al., 2011). However, given the putatively young age of subgenus *Tridentatae*, incomplete lineage sorting could also play a role in the lack of resolution, especially for high copy number genes like the internal transcribed spacer (Álvarez and Wendel, 2003). Another study, employing randomly amplified polymorphic DNAs (RAPDs), showed genetic differences among big sagebrush subspecies (McArthur et al., 1998). However, the variability and reproducibility of RAPDs has since been questioned (Perez et al., 1998), and since it is a dominant marker system, the effect of differential genetic dosage from polyploidy could influence the interpretation of presence-absence scoring.

High sequence depth across multiple genetic loci using next-generation sequencing holds the greatest prospect for unraveling evolutionary relationships between closely related taxa and between levels of polyploidy. Amplicon sequencing has been shown to be an effective approach for discerning the relationships of polyploids (Griffin et al., 2011). In this study, we focus on the intraspecific relationships of big sagebrush from high-coverage pyrosequencing of transcriptome-based amplicons. The amplicons were chosen based on sequence and SNP quality

and putative sequence annotation associated with secondary metabolite pathways (Bajgain et al., 2011). Our objectives were to (1) elucidate the phylogenetic relationships among subspecies of *A. tridentata*, (2) assess the nucleotide diversity within and between subspecies and ploidy levels, (3) infer the modes of polyploidy (i.e., autopolyploidy and allopolyploidy), and (4) evaluate *A. tridentata* subspecies morphological and biochemical characters in a phylogenetic context.

MATERIALS AND METHODS

Plant materials and DNA extractions—In total, 329 *A. tridentata* samples were collected from 48 collection sites, hereafter referred to as populations. These populations were distributed across nine states in the western United States (Fig. 1). One outgroup species, *A. arbuscula*, was also included. Collections consisted of seed and leaf tissue. For most populations, DNA was extracted from onsite-collected leaf tissue, but for several populations leaf tissue was obtained from greenhouse-grown seedlings. Subspecies of *A. tridentata* were determined by (1) height, leaf morphology, and branching morphology of plants at the population, (2) the degree of blue fluorescence emitted from ground leaves submerged in water under UV light (Stevens and McArthur, 1974; Goodrich et al., 1985), and (3) genome size based on flow cytometry, described below. The morphology and level of blue fluorescence was estimated by sampling leaves from three individuals per population and averaging fluorescence (0–5 scale) with 0 indicating no fluorescence. Based on previous studies

(e.g., McArthur and Sanderson, 1999), sympatry at the subspecies and ploidy levels is rare and occurs in narrow ecotones, and therefore, assessing three individuals per population was sufficient. The geographic and taxonomic information for these collections are provided in Table 1.

Collected leaves were dried in Thermo Savant ModulyoD (Fisher Scientific, Pittsburgh, Pennsylvania, USA) at -50°C in <100 mbar pressure for 48 h and then stored at -20°C with silica beads until used for DNA extraction. Genomic DNA was extracted from approximately 0.05 g of dehydrated leaf tissue using a modified cetyltrimethylammonium bromide (CTAB) protocol (Kidwell et al., 1992). The extracted DNA was quantified using a Nanodrop (ND 1000 Spectrophotometer, NanoDrop Technologies, Montchanin, Delaware, USA).

Flow cytometry—Flow cytometry was used for genome size estimation and ploidy determination in three individuals from each population and for the outgroup species. Analyses were conducted on a Partec PAII flow cytometer under UV fluorescence using known *A. tridentata* diploid and tetraploid plants as standards. The standard and experimental samples were chopped finely from fresh samples with a razor blade from a portion (0.5 to 1 cm^2) of each plant leaf. The CyStain UV precise P assay procedure, supplied by the manufacturer (Partec, Münster, Germany) was followed for nuclei extraction and staining. Plants suspected of being diploid were run with the tetraploid standards and vice versa. Diploid samples were approximately one-half the value of the standard and tetraploid samples were twice the value of the standard.

Gene selection, target amplification, and sequencing—Amplicon selection was based upon SNP quality between subsp. *tridentata* and subsp. *vaseyana* and associated secondary metabolite pathways (e.g., terpenoid and coumarin biosynthetic pathway), potentially important in response to abiotic or biotic

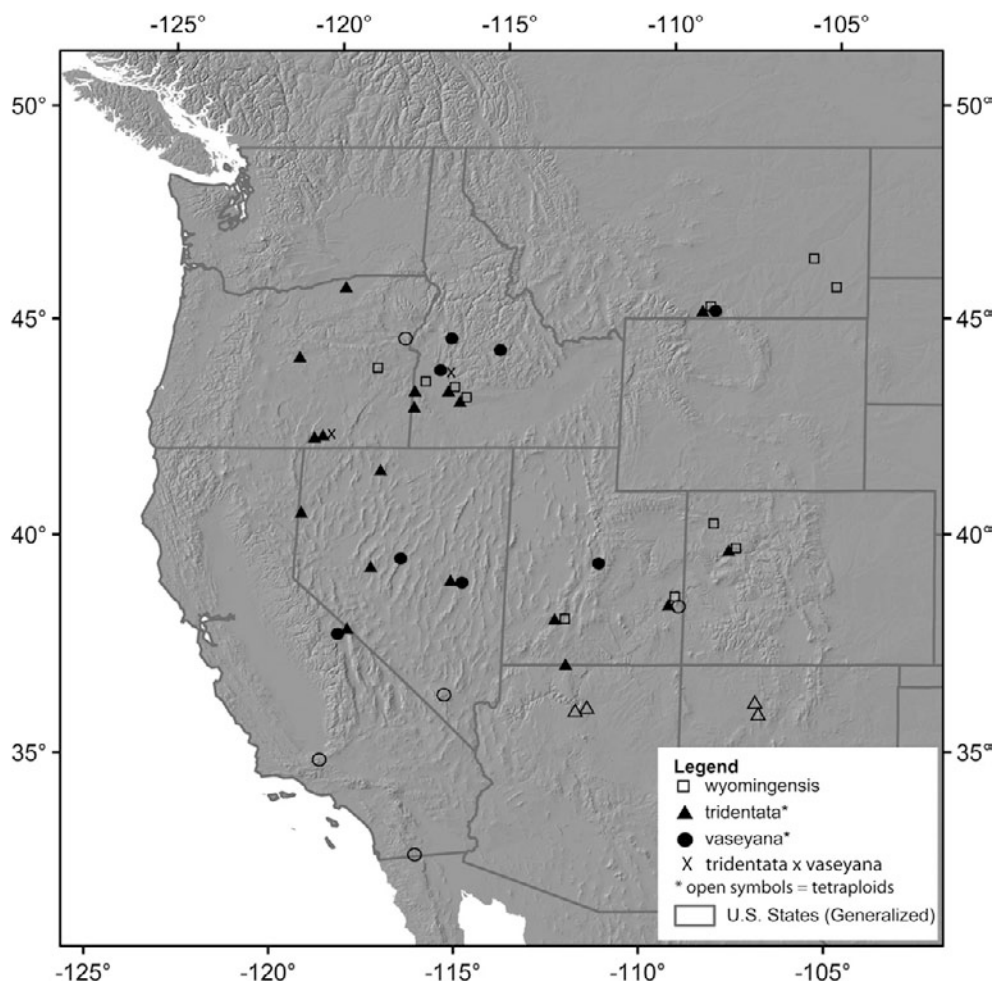


Fig. 1. Map of the western United States showing the locations of the sampled populations of *Artemisia tridentata*. Symbols designate subspecies as defined in the legend. Solid symbols = diploids, open symbols = tetraploids.

TABLE 1. List of *Artemisia tridentata* collections and an outgroup species showing geographical information, morphological subspecies (Taxon), number of samples (*N*) and ultraviolet (UV) fluorescence test of an average of three samples (0 = no blue iridescent fluorescence, 5 = high fluorescence).

Collection	Latitude	Longitude	Elevation (m a.s.l.)	Taxon	Ploidy	<i>N</i>	UV
CAV1	40.505	−120.562	1535	outgroup (<i>A. arbuscula</i>)	2x	7	5
CAT1	40.541	−119.919	1213	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	5	0
CAT2	37.797	−118.570	1974	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
IDT1	43.096	−115.657	949	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
IDT2	43.337	−116.008	939	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
IDT3	43.337	−116.964	1358	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
MTT1	45.206	−108.826	1240	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
NVT1	39.292	−117.852	1629	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
NVT2	38.797	−115.392	1797	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
NVW1	41.515	−117.775	1349	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
ORT1	42.975	−117.185	1349	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0.3
ORT2	45.759	−119.208	223	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
ORT3	42.326	−119.367	1782	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
ORW2	42.317	−119.393	1707	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
ORW3	44.134	−120.356	1048	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	3	0
UTT1	37.993	−112.509	2106	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
UTT2	38.306	−109.388	1820	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
UTT3	37.040	−112.264	1618	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
AZT1	35.917	−111.773	1898	<i>A. tridentata</i> subsp. <i>tridentata</i>	4x	7	0.7
AZW1	35.825	−112.131	1914	<i>A. tridentata</i> subsp. <i>tridentata</i>	4x	7	0.7
NMT1	35.980	−107.226	2122	<i>A. tridentata</i> subsp. <i>tridentata</i>	4x	7	0
NMT2	35.757	−107.148	1979	<i>A. tridentata</i> subsp. <i>tridentata</i>	4x	7	0.3
COT1	39.613	−107.810	1790	<i>A. tridentata</i> subsp. <i>tridentata</i>	2x	7	0
ORV2	42.321	−119.354	1828	<i>A. tridentata</i> subsp. <i>tridentata</i> x <i>vaseyana</i>	2x	7	2.7
IDV2	43.678	−115.974	1003	<i>A. tridentata</i> subsp. <i>tridentata</i> x <i>vaseyana</i>	2x	7	4
CAV2	37.723	−118.593	2330	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
IDV3	44.259	−114.640	1766	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	4.3
IDV4	44.508	−116.024	1438	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
IDV5	43.839	−116.256	1309	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
MTV1	45.172	−108.455	2156	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
NVV1	39.473	−117.050	2412	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
NVV2	38.818	−115.284	2132	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
UTV1	39.341	−111.522	2097	<i>A. tridentata</i> subsp. <i>vaseyana</i>	2x	7	5
CAV3	32.646	−116.186	827	<i>A. tridentata</i> subsp. <i>vaseyana</i>	4x	7	4
CAV4	34.825	−118.872	1120	<i>A. tridentata</i> subsp. <i>vaseyana</i>	4x	5	3
NVV3	36.305	−115.617	2410	<i>A. tridentata</i> subsp. <i>vaseyana</i>	4x	7	2.7
ORV1	44.488	−117.323	986	<i>A. tridentata</i> subsp. <i>vaseyana</i>	4x	7	3.3
UTV3	38.341	−109.217	2326	<i>A. tridentata</i> subsp. <i>vaseyana</i>	4x	7	3.6
COW1	39.613	−107.810	1790	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
COW2	40.181	−108.456	1788	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0.3
IDW1	43.093	−115.653	963	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
IDW2	43.327	−116.004	977	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
IDW3	43.466	−116.853	870	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
MTW1	45.655	−105.155	972	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
MTW2	46.322	−105.826	757	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
MTW3	45.207	−108.783	1458	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	0
ORW1	43.785	−118.259	1196	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	1
UTV2	37.993	−112.509	2118	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	7	1.7
UTW1	38.328	−109.435	1797	<i>A. tridentata</i> subsp. <i>wyomingensis</i>	4x	1	0.7

stress. This selection was guided by our previous transcriptome analysis (Bajgain et al., 2011). Initial primers for 60 target loci were designed using the Primer3 program (Rozen and Skaletsky, 1999), with an average target size of 385 base pairs (bp). During the design process, a “common oligo” was added to the end of each designed primer to enable multiplex identifier (MID) assignment of each sample through nested PCR amplification (Cronn et al., 2012). Two sets of primers (locus-specific primers and MID-barcoded primers) were synthesized according to the Access Array System User Guide (Fluidigm Corporation, San Francisco, California [CA], USA) at IDT (Coralville, Iowa, USA). Each primer set was first evaluated on a single sagebrush DNA source to (1) validate amplification of a single locus and to (2) validate that the actual amplicon size of genomic DNA matched the expected size as designed using an EST template. A total of 48 loci that met these criteria were selected for the Access Array system. Primers (Appendix S1) were validated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Foster City, CA). The sample loading, mixture, and amplification were performed on the Fluidigm 48.48 Access

Array IFC using the Fluidigm FC1 cycler. The amplified targets were pooled together (2 μ L of each sample) and size selected using Agencourt Ampure beads (Beckman Coulter Genomics, Morrisville, North Carolina, USA). The resulting DNA fragments were sequenced using the 454 Genome Sequencer reagents according to published specifications (Roche, Branford, Connecticut, USA). The amplified targets were sequenced in three phases: (1) one set of amplicons (48 individuals) was sequenced on a quarter plate of 454 titanium plate; (2) six sets of amplicons ($48 \times 3 = 144$ MID or individuals) were pooled and each sequenced on a half PicoTiterPlate (PTP); (3) the remaining 34 individuals (including the outgroup) were sequenced on a quarter PTP plate.

Sequence quality filtering and assembly—Sequencing errors were minimized using the following steps: (1) the program QUALITY_TRIM (CLC Bio) was used to remove low quality reads (nucleotides with quality less than 20 were rejected), (2) only reads that contained the entire amplicon sequence were used, (3) sequences were trimmed at the internal ends of the forward and

reverse primers, (4) homopolymer errors were reduced by eliminating gaps that followed homopolymer repeats >3 bases, (5) unique reads were identified with FASTAUNIQUE script (Quince et al., 2011), and (6) chimeras were identified for removal using the program UCHIME (Edgar et al., 2011). The filtered reads were then mapped to reference sequences (Bajgain et al., 2011) with a minimum identity of 95% nucleotide sequence similarity using CLC Genomic Workbench to prevent analysis of paralogous loci. For each of the contigs, a majority-rule consensus sequence was created for each of the 48 populations and one outgroup (*A. arbuscula*). Prior to coalescent Bayesian phylogenetic analyses (discussed below), the amplicon consensus sequences and haplotypes of individuals were evaluated for intragenic recombination with RECCO v0.93 with 1000 permutations (Maydt and Lengauer, 2006). Population-level consensus sequences that had significant recombination ($P > 0.05$) were replaced with an “N” between the recombination breakpoints.

Haplotype identification and nucleotide diversity—Putative haplotypes within amplicons were identified using hapHunt (J. T. Page and J. A. Udall, unpublished program), a K -means clustering algorithm (MacQueen, 1967). This algorithm assigns each read to the closest cluster (K), then recalculates the center of each cluster and repeats the process until converging on a solution. We set $K = 30$ to identify more clusters than were actually present, then omitted clusters containing a single read (the seed read) or no read. The remaining clusters were considered putative haplotypes. K -means clustering was replicated 10 times for each amplicon to judge for consistency in haplotype assignment. Clustering results were compared on the basis of Dunn’s cluster validity index (Dunn, 1973) and Chord distances (Cavalli-Sforza and Edwards, 1967). In-frame alignments of sequence reads were generated by remapping individual reads to cluster consensus sequences using CLC Bio Genomics Workbench (Aarhus, Denmark).

To assess amplicon diversity within and between populations, nucleotide diversity (π) was calculated by the BioPerl module Bio::PopGen::Statistics as an average of all pairwise comparisons (Stajich and Hahn, 2005). For each amplicon, we calculated each statistic both between individuals and within each population, according to the most common haplotype for each individual. For each amplicon population genetic structure, Φ_{ST} , was estimated as

$$\Phi_{ST} = \frac{\pi_T - \pi_S}{\pi_T} = \frac{\pi_B}{\pi_T},$$

where π_S is nucleotide diversity within a subpopulation, π_T is nucleotide diversity within the total population, and π_B is nucleotide diversity between subpopulations. To infer any selective pressure on the selected amplicon, we calculated K_a/K_s , the ratio of nonsynonymous to synonymous substitutions, using the program PAML (Yang, 2007) using the in-frame alignments.

Phylogenetic analyses—Phylogenetic analyses were conducted on a concatenated, diploid-only sequence data set and a combined diploid and tetraploid data set, deposited in the Dryad repository (website <http://dx.doi.org/10.5061/dryad.cs947>). Both data sets were constructed from the most common haplotype at each population for each amplicon. Two approaches were implemented to obtain species trees: (1) a coalescent-based Bayesian approach using the program BEAST v1.4 (Drummond and Rambaut, 2007) and (2) a phylogenetic network using the program SplitsTree v4.1 (Huson, 2006). For the Bayesian analysis, alignments of 25 genes (for diploids only) or 24 genes (for diploids and tetraploids combined) were used as input to BEAUti, which generated input files for BEAST. The program jModelTest v0.1.1 (Posada, 2008; Guindon and Gascuel, 2003) was used to select the most appropriate model for each gene. These tests revealed that the best models for all amplicons were the HKY or variations of this model. To simplify BEAST runs, the substitution model HKY was used for all contigs with gamma and invariant sites. The clock model chosen for all sequence sets was a relaxed clock with an uncorrelated log-normal model with a constant size coalescent tree prior. Each run consisted of 10,000,000 generations, starting from a random tree and sampling one in every 1000 generations. Likelihood scores were analyzed using the program Tracer v1.5 (Rambaut and Drummond, 2009), and the first 1000 trees from each run were determined as burn-in and discarded from the analysis, using the program TreeAnnotator. The consensus trees were constructed with TreeAnnotator and visualized with the program FigTree v1.3.1 (Rambaut, 2009) with posterior probabilities reported at each node.

To accommodate for ambiguous and conflicting phylogenetic signal potentially caused by reticulation events, we analyzed phylogenetic networks using

the program SplitsTree. This analysis was run using the default parameters, uncorrected p -distances with a neighbor-net network. Networks were drawn with the EqualAngle method. Lento plots (Lento et al., 1995) were constructed from p -distances in the program Spectronet v1.27 (Huber et al., 2002) to further analyze the support and conflict of splits on the diploid data set.

Haplotype networks—The alignments from individual amplicons of putative haplotypes were also used to generate haplotype networks using the program TCS v1.21 (Clement et al., 2000) to assess hybridization between diploids of subsp. *vaseyana* and subsp. *tridentata*. TCS was run with a 95% connection limit, and gaps were set as a 5th state. Node membership is based on subspecies frequencies of diploid individuals.

RESULTS

Amplicon assembly—The total number of reads generated from sequencing of all amplicons (48) in all individuals (329) was 1,864,777. Cognizant that targeted amplicons of individual genes belong to unique evolutionary trajectories, a conservative sequence identity threshold was used (95%). Lowering the threshold for sequence identity allowed for a greater percentage of mapped reads, but also increased the likelihood that amplified, nontarget loci would confound the sequencing results. Considering other forms of target-enrichment (Cronn et al., 2012), amplicon sequencing with the AccessArray System provided a high level of target specificity because a majority of sequenced reads corresponded to the targeted loci. Of the reads generated for this experiment, 66% of the reads mapped to the reference target with 95% sequence identity. Of the 17,338 total base pairs (bp) that were targeted by amplicon primers, an average of 79.5% were sequenced for each individual. The percentage sensitivity was lower than expected, mainly due to the use of ESTs for the design template of unproven amplicon sequences. The coverage of the reference-mapped amplicon sequences among all haplotypes ranged from 0× to 1334× with an average of 76× coverage of each MID-target amplicon.

Quality control (QC) steps were used for each read (trimming at the internal primer edges, filtering out homopolymer, low-quality, and chimeric reads) to generate an alignment for sequence diversity analysis. After QC, the mean length of all alignments in all individuals was 343.3 bp (median length of 360 bp). Some of the selected amplicon targets were not amenable to amplification due to a length bias (preferential amplification of shorter fragments), GC content, or primer binding efficiency (Appendix S1, see Supplemental Data with the online version of this article). A postsequencing analysis found little correlation between coverage depth and amplicon length or between coverage depth and GC content (Appendix S1). This suggests that localized sequence polymorphism or genomic background may have caused the reduced amplification of certain amplicons in certain populations. Thus, 24 or 25 amplicon targets (GenBank JX489390 to JX489414, Appendix S1) with deep, complete coverage from all populations were used to analyze the diploids and diploid-tetraploid data sets, respectively (average coverage: 109, median: 106, mode: 106).

Nucleotide diversity—Nucleotide diversity (π) was calculated for each population and amplicon. The overall levels of π averaged 0.0062 with a standard deviation of 0.0048 (min: 0.0012; max: 0.0192; Table 2). Within each population π averaged 0.0054 with a standard deviation of 0.0047 (min: 0.0009; max: 0.0191; Table 2). Of the 1152 population × amplicon combinations, 66 had 0 nucleotide diversity. Population IDW2

TABLE 2. The total nucleotide diversity (Total π), the average for within populations (Pop π), the percentage of π found within populations (Pop %) and population genetic structure (Φ_{ST}) for each amplicon. Below are the average, minimum (min), maximum (max) and standard deviation (SD) for all amplicons.

Contig	Total π	Pop π	Pop %	Φ_{ST}
20213	0.0192	0.0191	0.9917	0.0083
18511	0.0171	0.0165	0.9657	0.0343
23535	0.0128	0.0080	0.6296	0.3704
28528	0.0104	0.0098	0.9369	0.0631
20734	0.0088	0.0079	0.8935	0.1065
6046	0.0084	0.0071	0.8417	0.1583
26223	0.0083	0.0072	0.8683	0.1317
10446	0.0078	0.0076	0.9821	0.0179
19198	0.0071	0.0069	0.9728	0.0272
1647	0.0059	0.0058	0.9793	0.0207
1455	0.0057	0.0054	0.9491	0.0509
18504	0.0052	0.0047	0.8995	0.1005
15321	0.0051	0.0034	0.6722	0.3278
28210	0.0039	0.0032	0.8163	0.1837
16389	0.0035	0.0025	0.6935	0.3065
15020	0.0030	0.0027	0.9298	0.0702
1135	0.0029	0.0021	0.7282	0.2718
9907	0.0026	0.0019	0.7252	0.2748
15745	0.0025	0.0023	0.9083	0.0917
1936	0.0023	0.0016	0.6852	0.3148
15067	0.0020	0.0012	0.6055	0.3945
16964	0.0020	0.0019	0.9502	0.0498
7128	0.0014	0.0010	0.7201	0.2799
153	0.0012	0.0009	0.7744	0.2256
Average	0.0062	0.0054	0.8383	0.1617
Min	0.0012	0.0009	0.6055	0.0083
Max	0.0192	0.0191	0.9917	0.3945
SD	0.0048	0.0047	0.1263	0.1263

had the largest amount of diversity in any single amplicon (contig 20213, 0.0293; Appendix S2). The average population nucleotide diversity was 0.0054 (SD 0.0005). Populations IDW1 and CAV2 had the least (0.0042) and greatest (0.0064) amounts of nucleotide diversity averaged across the 25 amplicons, respectively. No single population (diploid or tetraploid) contained exceptionally high levels of nucleotide diversity compared to the other populations. When the π of individual amplicons was considered between populations, some amplicons had more nucleotide divergence than others. In particular, contig 18511 (Mg protoporphyrin IX chelatase [predicted]), contig 20213 (cc-nbs-lrr resistance protein [predicted]), and contig 23535 (NAC domain protein [predicted]) had exceptionally high levels of nucleotide diversity between populations (ranging from 2- to 5-fold more diversity) compared to the other amplicons (Table 2).

For each amplicon, six comparisons were made of nucleotide diversity between subspecies and ploidy level groups using Student's t test (Appendix S2). Of the 144 comparisons, only seven tests were significant after multiple-hypothesis correction ($P < 0.0021$; Dunn, 1961), and all tests indicating statistical significance were of comparisons with the diploid subsp. *tridentata* indicating that for some loci subsp. *tridentata* has unique levels of nucleotide diversity. Five tests were significant between diploid spp. *tridentata* and diploid subsp. *vaseyana* (contigs 23535, 6046, 1647, 9907, 153). Two tests were significant between diploid spp. *tridentata* and tetraploid subsp. *tridentata* (contigs 28210 and 9907). No significant differences were found between ploidy in subsp. *vaseyana* or between diploid subsp. *tridentata* and subsp. *wyomingensis*. Overall, the corrected t tests suggest that the nucleotide diversity varies between amplicon and

that the majority of significant differences occur between diploid subspecies and secondarily between different ploidy levels.

The estimated population structure (Φ_{ST}) for each amplicon represented the proportional amount of nucleotide diversity between populations compared to the total nucleotide diversity. In general, the proportion of total nucleotide diversity due to within population polymorphisms was high (min: 59.31% max: 97.93%) compared to the amount attributed to between populations (min: 2.07% max: 17.58%). The average Φ_{ST} was 0.16 and extremely variable between amplicons (SD 0.12). Two of the three amplicons with the highest levels of overall π (contigs 20213 and 18511) had some of the lowest amounts of within population nucleotide diversity. In addition to the magnitude difference within and between population π , these exceptional cases emphasized the general trend that most of the genetic variation detected in this study resided within the populations. Two amplicons, contig 23535 and 15067, had the highest levels of Φ_{ST} compared to overall π . Contig 23535 had relatively high overall levels of π and had the second highest amount of nucleotide diversity between populations. In contrast, amplicon contig 15067 (6-xylosyltransferase [predicted]) had the highest amount of within population nucleotide diversity with one of the lowest amounts of overall π suggesting that most of its nucleotide diversity was found within populations. In summary, these results indicated distinct evolutionary histories for the loci targeted in this study.

The target reference sequences were selected by their high-quality blastx hits (Appendix S1), and their long open reading frames (ORF) that encoded putative proteins. Using the long ORF, we identified synonymous and nonsynonymous nucleotide substitutions in the target amplicons. In general, amplicons with the high levels of K_a/K_s (i.e., intrasite average $K_a/K_s >> 1$; contigs 26223, 20734, and 1455) were also amplicons with higher than average π . We conducted the same 144 Student's t tests for levels of K_a/K_s . There were two significant comparisons of K_a/K_s levels (contig 15475 between diploid subsp. *vaseyana* and diploid subsp. *tridentata*; contig 28210 between diploid subsp. *tridentata* and tetraploid subsp. *tridentata*). Other amplicons have relatively high levels of K_a/K_s variation between the populations (e.g., contigs 153, 1647, and 26223), although this variation did not coincide with either subspecies or ploidy level. These results suggest that diversifying selection was generally not population dependent.

Phylogenetics—Intragenic recombination—The alignment of consensus sequences from each amplicon was analyzed for intragenic recombination. For the diploid data set, recombinants were detected in 13 of the 25 amplicons ($P > 0.05$). For the combined diploid and tetraploid data set, recombinants were detected in 11 of the 24 amplicons (Appendix S3). For both data sets, detected breakpoints averaged 90 bp in length (0.7% of the total sequence data). These data were used to build recombination breakpoints for Bayesian analysis, described below.

Relationships of diploids—Phylogenetic tree and neighbor-net network analyses were conducted with a total of 12 367 bp of concatenated consensus sequences data from 25 amplicons. A preliminary analysis of the diploid data set with Bayesian coalescence without removing intragenic recombination produced a polyphyletic clade that was inconsistent with the morphological assessments and the neighbor-net network. This preliminary Bayesian tree with a poor effective sample size (ESS) of 3.2 failed to segregate *tridentata* and *vaseyana* into

groups (data not shown). Replacing the intragenic recombination between detected breakpoints with missing data resolved this polyphyly. The resulting tree supports distinct *tridentata* and *vaseyana* groups with high posterior probability support (0.98) and an ESS of 203 (Fig. 2A). While replacement of intragenic recombinants greatly improved the Bayesian coalescent tree, the topology of the neighbor-net tree did not change significantly between the original and recombinant modified data sets (data not shown). In this case, the original consensus sequences were used for the neighbor-net network.

Results from the neighbor-net network were congruent with the Bayesian coalescence tree, supporting separate groups of *vaseyana* and *tridentata*. Collections of putative *tridentata* × *vaseyana* hybrids flank the *vaseyana* group (Fig. 3A). The net-like structure between and within subspecies groups is indicative of conflicts between splits. Of the 85 splits, 65 contribute to the net-like structure. The Lento plot shows that the majority of splits segregating *vaseyana* and *tridentata* are not bifurcating (Fig. 4). Only two of the 85 splits, 58 and 62, were exclusive to *vaseyana* (Figs. 3A, 4). While separating subspecies, these two splits show discordance with the outgroup, *A. arbuscula*, leading to placement of the outgroup in the middle of the network. While many of other splits support the groupings, they do not entirely bifurcate subspecies. Most splits are derived from SNPs that have high membership differences between subspecies, but are not exclusive to one subspecies (Fig. 4). Examples of these SNP frequency differences are illustrated by the haplotype network analysis discussed below.

Relationships of polyploids—Like the diploid analysis, Bayesian coalescence failed to produce a species tree with an acceptable ESS value without the replacement of intragenic recombination between breakpoints with missing data (Appendix S3). With the modified data set, relationships between diploids resolved into separate groups. Tetraploid collections were placed within both *tridentata* and *vaseyana* groups (Fig. 2B, ESS = 312). Tetraploids identified as subsp. *wyomingensis* were predominately interspersed within the *tridentata* clade. However, one collection, COW2, was placed among diploid *vaseyana* lineages. Other tetraploid collections displaying *vaseyana* or *tridentata* phenotypic characteristics were also polyphyletic, having members associated with both the *vaseyana* and *tridentata* groups. Posterior probability support was reduced (0.20) for *vaseyana* and *tridentata* groups with the inclusion of tetraploids, compared to the diploid tree (0.98). Placement of the outgroup, CAV1, with *tridentata* is discordant with the Bayesian diploid tree.

The neighbor-net network and Bayesian tree have similar topologies. Most tetraploid collections associated with the *tridentata* and *vaseyana* groups in the Bayesian tree are found in the similar positions in the neighbor-net network. The differences between phylogenetic approaches are most apparent for collections with a high degree of introgression between subspecies. Tetraploid collections NVV3 and AZW1 are examples where the neighbor-net network can illustrate collections with putative admixture (Fig. 3B).

Haplotype networks—Haplotype networks provide a different perspective to the consensus-based approach of the Bayesian and neighbor-net species trees. The haplotype networks depict relationships between haplotype and nodal membership of diploid subsp. *tridentata* ($N = 148$) and *vaseyana* ($N = 89$) individuals. Two examples are provided: contig 16389, a partial

gene sequence associated with phenylalanine ammonia-lyase (Fig. 5A) and contig 15321, a partial gene sequence associated with phosphofructokinase (Fig. 5B). These contigs are components of the coumarin and glycolysis cellular pathways, respectively. Both examples illustrate a general association between haplotype relationships and frequency subspecies membership. Contig 16389 contains 14 haplotypes of which the majority is shared by subsp. *tridentata* and *vaseyana* and four closely related haplotypes that are exclusive to *tridentata* (Fig. 5A). Contig 15321 is more complex, with 34 haplotypes, of which seven have exclusive membership to *tridentata* and three to *vaseyana* (Fig. 5B). However, the pattern is similar to contig 16389 in which haplotypes with higher membership have a tendency to be an admixture of both subspecies. When tetraploids are included in the network analysis of the two examples (i.e., contigs 15321 and 16389), all the nodes, including those that segregate diploid species, have tetraploid members (figure not shown), illustrating the breadth of haplotype diversity of tetraploids, which is consistent with the results of π and Φ_{ST} .

DISCUSSION

A major goal of this study was to assess the phylogenetic relationships between the taxonomically challenging *A. tridentata* subspecies. Transcriptome-based amplicon sequencing was chosen based upon (1) the known hybridization and polyploidy in this species (McArthur et al., 1988), which can confound traditional population genetic approaches, (2) the ability to target genes with putative association to functional traits that are diagnostic to subspecies identification (i.e., coumarin content detected by UV light and other plant volatiles), (3) the capacity to assess nucleotide and haplotype variation among subspecies and ploidy levels and (4) the ability to develop functional and ecological genomic tools for future research that are applicable to other related species (Wheat, 2010). Amplicon sequencing provided high coverage depth, high-throughput, and was capable of specifically targeting genes of interest. This depth of sampling provided a detailed view of nucleotide variation within coding regions of *A. tridentata* and the inference of putative haplotypes.

Subspecies relationships—Phylogenetic analyses of 24 secondary metabolite-related sequences in *A. tridentata* support distinct groups of diploid *tridentata* and *vaseyana* subspecies. However, the results also suggest hybridization between subspecies. This hypothesis is supported by (1) Bayesian coalescence analyses that were greatly improved by replacing putatively intragenic recombination (Appendix S3) with missing data, (2) the majority of splits in the neighbor-net analyses have some degree of conflict due to shared SNPs between subspecies (Fig. 4), forming the net-like appearance (Fig. 3A), and (3) haplotype networks of individual amplicons that show the most common haplotypes share membership between diploid subspecies (Fig. 5). These results support previous studies that have shown that the formation of homoploid hybrids occurs in narrow ecotones between subspecies *tridentata* and *vaseyana* ecotypes (McArthur et al., 1988). Wang et al. (1997) and Rodriguez et al. (2009) demonstrated in a reciprocal transplant study that homoploid hybrids were limited to ecotones due to lower fitness in parental habitats. Although it is difficult to differentiate hybridization from incomplete lineage sorting, especially in closely related taxa (Degnan and Rosenberg,

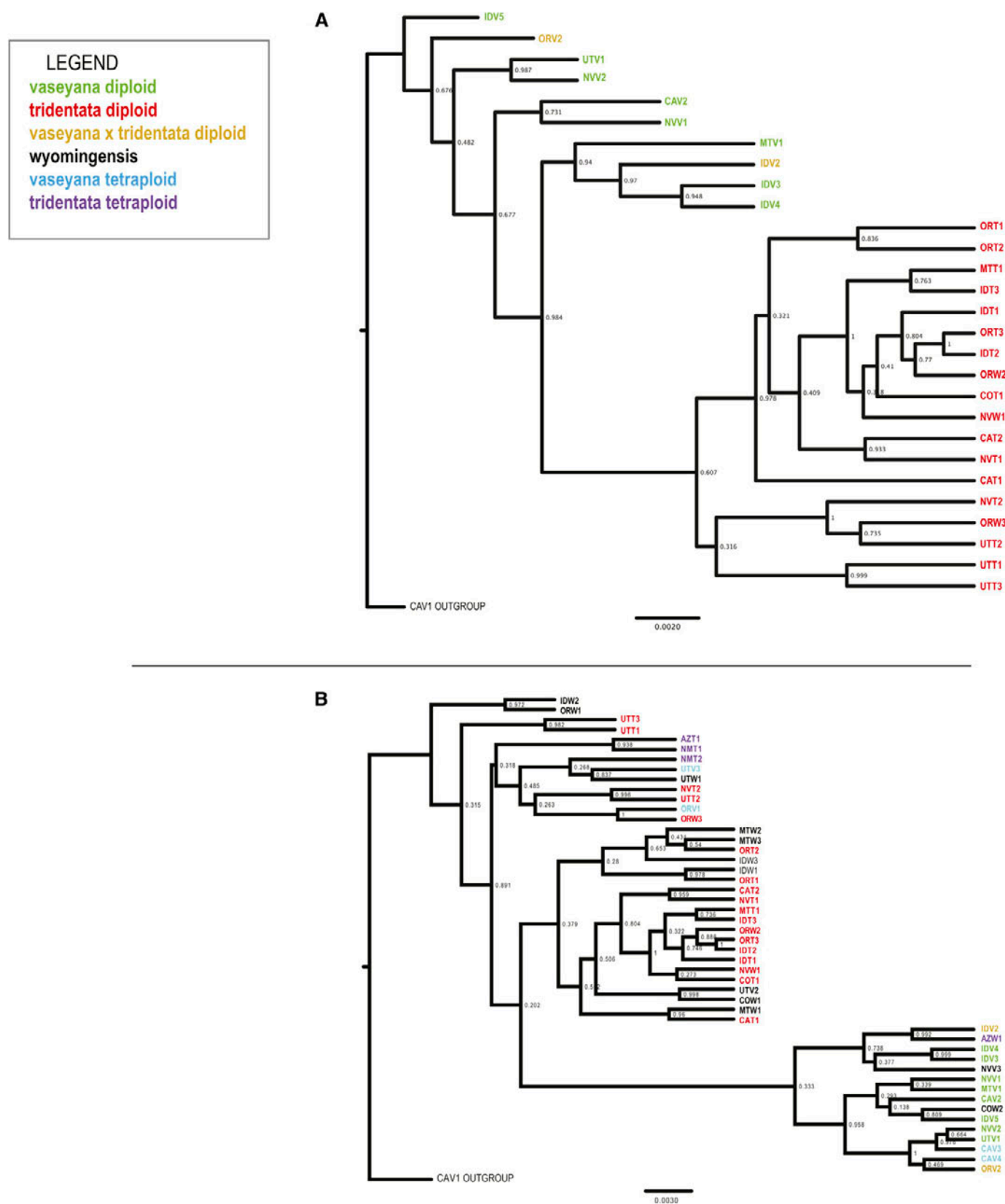


Fig. 2. Bayesian coalescent tree of *Artemisia tridentata* diploids (A) and diploids and tetraploids (B) constructed from the consensus sequence of individuals at each population. Subspecies *tridentata* and *vaseyana* groups are indicated and posterior probabilities are reported for each node.

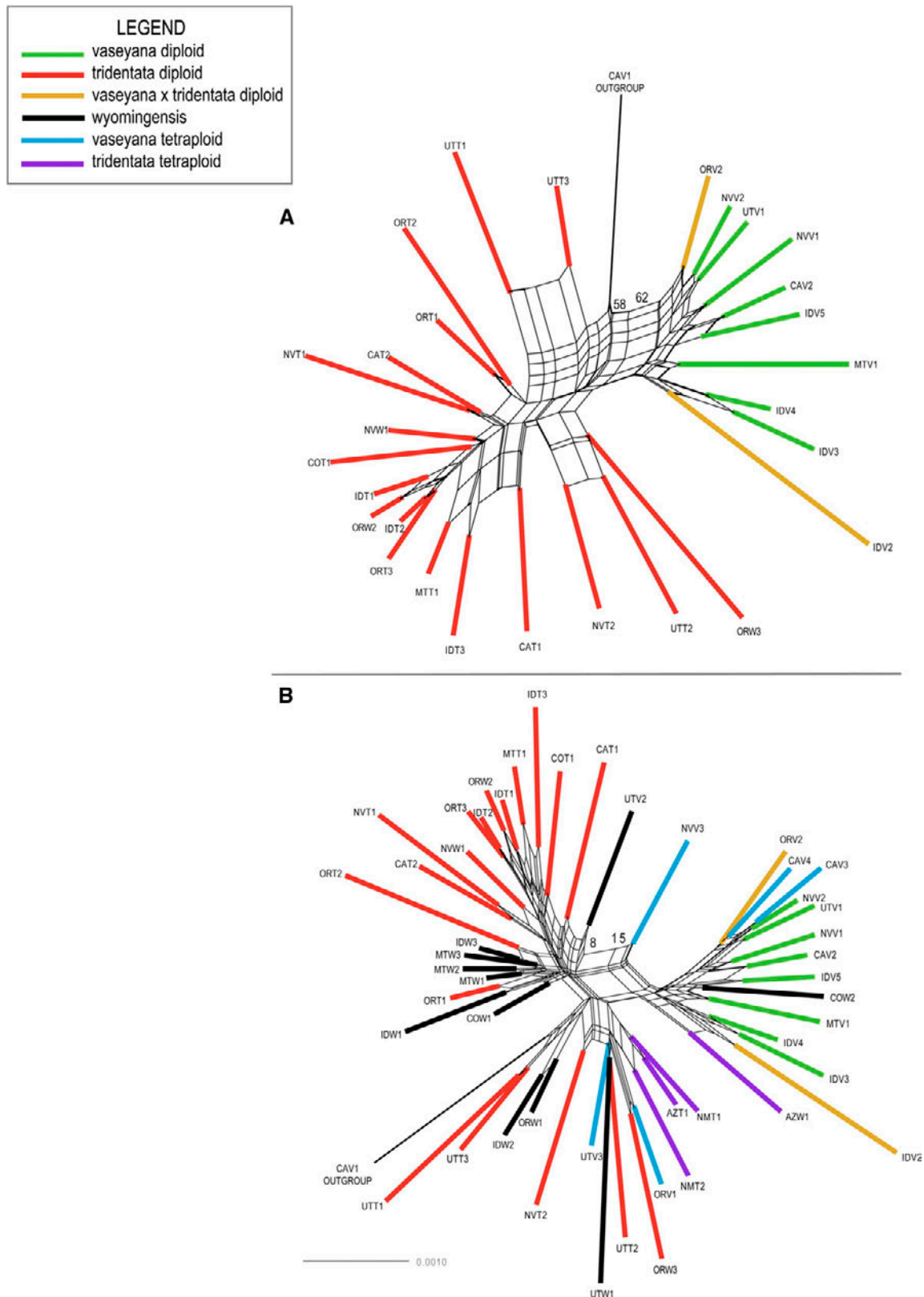


Fig. 3. Unrooted, phylogenetic networks of *Artemisia tridentata* using the neighbor-net algorithm. The networks were constructed from uncorrected p -distances based on 24 contigs of consensus sequences of individuals at each population. (A) Illustrates relationships between 28 diploid collections, and (B) includes the diploid collections and 20 tetraploid collections. Taxonomic differences are represented by colored edges, and the scale bar indicates expected changes per site. The numbered splits highlight nucleotide substitutions that bifurcated *A. tridentata* and *A. vaseyana* subspecies.

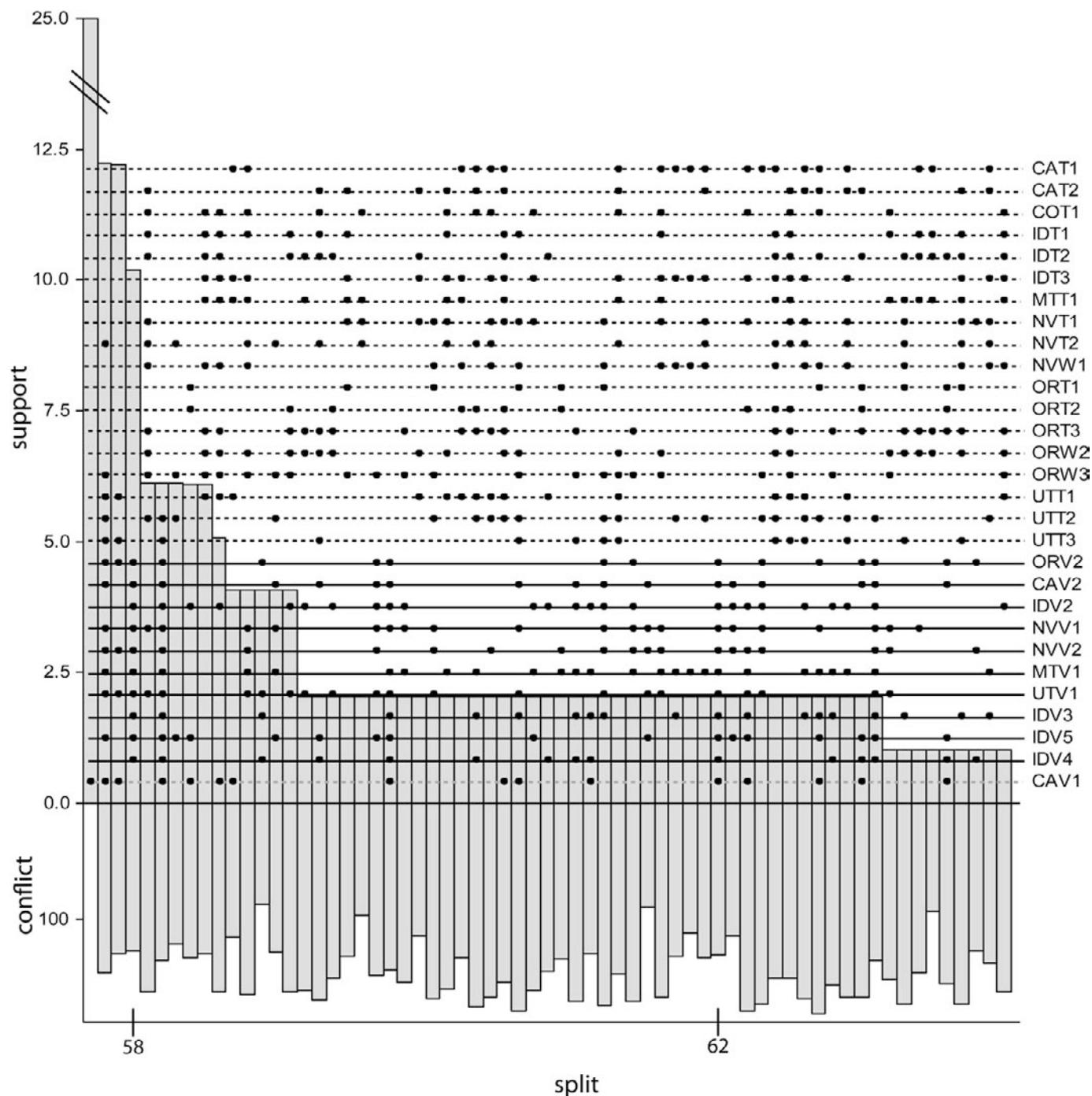


Fig. 4. Lento plot of 18 *Artemisia tridentata* subsp. *tridentata* and 10 *A. tridentata* subsp. *vaseyana* diploid populations (including two putative hybrids) derived from consensus sequences of 24 contigs. The plot shows the population membership (x-axis) represented by dots at 65 splits (y-axis) associated with Fig 3A. Support and conflict is depicted with the gray bars. The horizontal lines (solid = *vaseyana*, dashed = *tridentata*, gray dashed = *A. arbuscula*) relate to the collection names on the right. The highlighted splits 58 and 62 are discussed in the text.

2009), our results suggest that ecotonal hybrids may serve as a conduit for gene flow between the diploid subspecies *tridentata* and *vaseyana*.

Phylogenetic analyses that included both diploid and tetraploid populations supported a topology that placed tetraploid populations within and between diploid subsp. *tridentata* and *vaseyana* groups (Figs. 2B, 3B). Inclusion of tetraploids

in Bayesian coalescence decreased posterior probability support between *tridentata* and *vaseyana* groups from 0.98 to 0.2 (Fig. 2A, B), suggesting that at least some of tetraploid consensus sequences greatly increased the phylogenetic conflict. The apparent admixture of some tetraploid populations could be created by homoploid hybridization followed by autopolyploidy, by allotetraploidy, or by autopolyploidy followed by



Fig. 5. Haplotype network analysis of two contigs from partial gene sequences: (A) 16389 (phenylalanine ammonia-lyase, [predicted]) and (B) 15321 (phosphofructokinase, [predicted]). Node memberships show the relative frequencies of *Artemisia tridentata* subsp. *tridentata* (black) and subsp. *vaseyana* (white) diploids. Labels above the node indicate the haplotype cluster number assigned by *K*-means algorithm.

tetraploid hybridization. The latter, autopolyploidy followed by tetraploid hybridization, seems the most likely given the predominant parapatric distribution of diploid subsp. *tridentata* and *vaseyana*. Postpolyploid hybridization has also been found to be the most likely mode of introgression, providing genetic diversity to autopolyploids of *Capsella bursa-pastoris* (L.) Medik. (Slotte et al., 2008; St. Onge et al., 2012), *Rorippa amphibia* (L.) Besser (Stift et al., 2010), and *Arabidopsis* species (Jørgensen et al., 2011). Hybridization between tetraploid

A. tridentata subspecies is also supported by pairwise comparisons of π between subspecies and ploidy levels. Five amplicons had significant differences between diploid subspecies, whereas no differences were found between tetraploid subspecies (Appendix S2). These nucleotide diversity patterns are supported by haplotype networks (e.g., contigs 15321, 16389) where all nodes, even those that segregate *vaseyana* and *tridentata* diploids, have tetraploid membership (data not shown).

In *A. tridentata*, two types of hybridization, homoploid and polyploid, are important factors in its evolution and landscape dominance. Previous studies have shown that the latter appears to have the most ecological significance (Wang et al., 1997; McArthur and Sanderson, 1999). Based on the phylogenetic relationships, polyploidy is a dynamic process in *A. tridentata*, stemming from multiple events, as has been found in other plant species (Soltis and Soltis, 2000; Rebernig et al., 2012). McArthur and Sanderson (1999) found a tightly parapatric distribution between diploid subsp. *tridentata* and *wyomingensis*. Our results suggest that in some instances diploid *tridentata* and tetraploid lineages can be closely related at a local scale. For example, collections UTV3, UTW1, and UTT2 are taxonomically distinct (tetraploid *vaseyana*, *wyomingensis*, and 2× *tridentata*, respectively), yet genetically closely related (Fig. 3B). All three collections are distributed along an elevational gradient geographically separated by less than 20 km. In contrast, close phylogenetic relationships were not seen between *tridentata*–*wyomingensis* parapatric collections (e.g., COT1 and COW1, UTT1 and UTV2). We postulate that tetraploid lineages are likely derived from local or regional diploid populations followed by some degree of postpolyploid hybridization that is likely determined by environmental factors such as climate and soils. A better understanding of the interaction between tetraploid formation and the environment is a critical need for restoration, but further study at finer geographic scales is required to address these hypotheses.

Phylogenetic signal—Gene discordance is often a consequence of multilocus data sets, and there are various approaches to assess discordance (Carstens and Knowles, 2007; Yang and Rannala, 2012). In our amplicon data set, three analyses suggest some degree of gene discordance. First, high read depth provided an opportunity to estimate phylogenetic signal for each gene using Φ_{ST} calculated from π . This data shows a highly variable Φ_{ST} , ranging from 0.008 to 0.39 (Table 2). Second, the Lento plot supports varying degrees of conflict/support for each SNP (Fig. 4), and SNPs with higher support were often associated with a few genes (data not shown). Third, intragenic recombination was detected within nearly half of the genes and 34.7% and 61.8% of diploid and tetraploid individuals, respectively (Appendix S3). This gene discordance is likely a consequence of historic and ongoing introgression between subspecies, incomplete lineage sorting, or both. If a similar study was conducted with only a few genes, it is likely no phylogenetic signal would be detected between subspecies just by chance. This data supports the use of larger multilocus data sets with deep sequence coverage to discern patterns between closely related taxa.

Phylogenetic relationships and taxonomic characters—It has long been recognized that formation of polyploids from diploid taxa creates a highly variable range of morphological characters (Stebbins, 1950; Grant, 1971; de Wet, 1971). *Artemisia tridentata* is an excellent example. Diploid collections in this study demonstrate a concordant pattern between taxonomic characters (i.e., height, branching, and UV fluorescence from coumarin content) and phylogenetic relationships. However, as discussed above, tetraploid collections are often phylogenetically intermediate between diploid taxa. The phylogenetic placement (Figs. 2B, 3B) is typically complemented by the morphology. For example, UV fluorescence is a bright blue for diploid *vaseyana* and absent for diploid *tridentata*, whereas

tetraploids vary in UV fluorescence (Table 1). McArthur and Sanderson (1999) found similar results. This character typically corresponds to the phylogenetic placement; however, one exception is COW2 (data not shown). This population displayed low UV fluorescence, but was placed within the *vaseyana* group. Other observed morphologically intermediate characteristics include height and branching angle. The congruence between taxonomic characters and phylogeny could be influenced in part because we targeted three genes putatively associated with the coumarin pathway (contigs 1647, 1936, 16389), which is the source for the UV fluorescence characteristic (Shafizadeh and Melnikoff, 1970). Two of the three contigs (1936 and 16389) had high Φ_{ST} (>0.30), and contig 16389 showed partial segregation of haplotypes between diploid subspecies (Fig. 5A).

Conclusions—Amplicon sequencing of secondary metabolite genes in *A. tridentata* revealed a pattern of phylogenetic relationships that supports distinct diploid subspecies. However, tetraploid populations, including subsp. *wyomingensis*, are polyphyletic. The ability to form autopolyploids followed by postpolyploid hybridization of local or regional diploid progenitors (i.e., subsp. *tridentata* and *vaseyana*) appears to be the predominant mode of polyploidy and hybridization. This process may be the most important mechanism that enabled the widespread dominance of *A. tridentata* tetraploids. Results of the analysis of nucleotide diversity and the high variability between amplicons support a degree of gene discordance and suggests hybridization between subspecies. Hybridization may also explain intermediate morphological characters, especially UV fluorescence, among tetraploids. Greater understanding of the distribution of tetraploid lineages and their associations with environmental factors will be an important step in developing restoration guidelines for big sagebrush ecosystems.

LITERATURE CITED

- ÁLVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.
- BAJGAIN, P., B. A. RICHARDSON, J. C. PRICE, R. C. CRONN, AND J. A. UDALL. 2011. Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). *BMC Genomics* 12: 370.
- BURTON, T. L., AND B. C. HUSBAND. 1999. Population cytotype structure in the polyploid *Galax urceolata* (Diapensiaceae). *Heredity* 82: 381–390.
- CARSTENS, B. C., AND L. L. KNOWLES. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *Systematic Biology* 56: 400–411.
- CAVALLI-SFORZA, L., AND A. EDWARDS. 1967. Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics* 19: 233.
- CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.
- CONNELLY, J. W., M. A. SCHROEDER, A. R. SANDS, AND C. E. BRAUN. 2000. Guidelines to manage sage grouse populations and their habitats. *Wildlife Society Bulletin* 28: 967–985.
- CRONN, R., B. J. KNAUS, A. LISTON, P. J. MAUGHAN, M. PARKS, J. V. SYRING, AND J. UDALL. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.
- DEGNAN, J. H., AND N. ROSENBERG. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.
- DE WET, J. M. J. 1971. Polyploidy and evolution in plants. *Taxon* 20: 29–35.

- DRUMMOND, A. J., AND A. RAMBAUT. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.
- DUNN, J. 1973. A relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3: 32–57.
- DUNN, J. O. 1961. Multiple comparisons among means. *Journal of the American Statistical Association* 56: 52–64.
- EDGAR, R. C., B. J. HAAS, J. C. CLEMENTE, C. QUINCE, AND R. KNIGHT. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194–2200.
- FREEMAN, D. C., W. A. TURNER, E. D. MCARTHUR, AND J. H. GRAHAM. 1991. Characterization of a narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). *American Journal of Botany* 78: 805–815.
- GARCIA, S., E. MCARTHUR, J. PELLICER, S. SANDERSON, J. VALLÈS, AND T. GARNATJE. 2011. A molecular phylogenetic approach to western North America endemic *Artemisia* and allies (Asteraceae): Untangling the sagebrushes. *American Journal of Botany* 98: 638–653.
- GOODRICH, S., E. D. MCARTHUR, AND A. H. WINWARD. 1985. A new combination and a new variety in *Artemisia tridentata*. *Western North American Naturalist* 45: 99–104.
- GRANT, V. 1971. Plant speciation. Columbia University Press, New York, New York, USA.
- GREGG, M., J. CRAWFORD, M. DRUT, AND A. DELONG. 1994. Vegetational cover and predation of sage grouse nests in Oregon. *Journal of Wildlife Management* 58: 162–166.
- GRIFFIN, P. C., C. ROBIN, AND A. A. HOFFMANN. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology* 9: 19.
- GUINDON, S., AND O. GASCUEL. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.
- HUBER, K., M. LANGTON, D. PENNY, AND V. MOULTON. 2002. Spectronet: A package for computing spectra and median networks. *Applied Bioinformatics* 1: 159–161.
- HUSON, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.
- JØRGENSEN, M. H., D. EHRLICH, R. SCHMICKL, M. A. KOCH, AND A. K. BRYSTING. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evolutionary Biology* 11: 346.
- KIDWELL, K. K., T. C. OSBORN, AND J. S. BECKMAN. 1992. Simple plant DNA isolation procedures. In J. S. Beckmann and T. C. Osborn [eds.], *Plant genomes: Methods for genetic and physical mapping*, 1–13. Kluwer, Dordrecht, Netherlands.
- KOLB, K. J., AND J. S. SPERRY. 1999. Differences in drought adaptation between subspecies of sagebrush (*Artemisia tridentata*). *Ecology* 80: 2373–2384.
- LENTO, G. M., R. E. HICKSON, G. K. CHAMBERS, AND D. PENNY. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. *Molecular Biology and Evolution* 12: 28–52.
- MACQUEEN, J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, 281–297. University of California Press, Berkeley, California, USA.
- MAYDT, J., AND T. LENGHAUER. 2006. Recco: Recombination analysis using cost optimization. *Bioinformatics* 22: 1064–1071.
- MCARTHUR, E. D., J. MUDGE, R. VAN BUREN, W. R. ANDERSEN, S. C. SANDERSON, AND D. G. BABEL. 1998. Randomly amplified polymorphic DNA analysis (RAPD) of *Artemisia* subgenus *Tridentatae* species and hybrids. *The Great Basin Naturalist* 58: 12–27.
- MCARTHUR, E. D., L. C. POPE, AND D. C. FREEMAN. 1981. Chromosomal studies of subgenus *Tridentatae* of *Artemisia*: evidence for autopolyploidy. *American Journal of Botany* 68: 589–605.
- MCARTHUR, E. D., AND S. C. SANDERSON. 1999. Cytogeography and chromosome evolution of subgenus *Tridentatae* of *Artemisia* (Asteraceae). *American Journal of Botany* 86: 1754–1775.
- MCARTHUR, E. D., B. L. WELCH, AND S. C. SANDERSON. 1988. Natural and artificial hybridization between big sagebrush (*Artemisia tridentata*) subspecies. *Heredity* 79: 268–276.
- MEIMBERG, H., K. J. RICE, N. F. MILAN, C. C. NJOKU, AND J. K. MCKAY. 2009. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* 96: 1262–1273.
- NUISMER, S. L., AND B. M. CUNNINGHAM. 2005. Selection for phenotypic divergence between diploid and autotetraploid *Heuchera grossulariifolia*. *Evolution* 59: 1928–1935.
- PEREZ, T., J. ALBORNOZ, AND A. DOMINGUEZ. 1998. An evaluation of RAPD fragment reproducibility and nature. *Molecular Ecology* 7: 1347–1357.
- POSADA, D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.
- PYKE, D. A. 2011. Restoring and rehabilitating sagebrush habitats. In S. T. Knick and J. W. Connelly [eds.], *Greater sage-grouse: Ecology and conservation of a landscape species and its habitats*. Studies in Avian Biology, vol. 38, 531–548. University of California Press, Berkeley, California, USA.
- QUINCE, C., A. LANZEN, R. J. DAVENPORT, AND P. J. TURNBAUGH. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 12: 38.
- RAMBAUT, A. 2009. FigTree, version 1.31 [computer program]. Available at <http://tree.bio.ed.ac.uk/software/figtree/>.
- RAMBAUT, A., AND A. DRUMMOND. 2009. Tracer version 1.5 [computer program]. Available at http://beast.bio.ed.ac.uk/Main_Page.
- RAMSEY, J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences, USA* 108: 7096–7101.
- RAMSEY, J., AND D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.
- REBERNIG, C. A., H. WEISS-SCHNEEWEISS, C. BLÖCH, B. TURNER, T. F. STUESSY, R. OBERMAYER, J. L. VILLASEÑOR, AND G. M. SCHNEEWEISS. 2012. The evolutionary history of the white-rayed species of *Melampodium* (Asteraceae) involved multiple cycles of hybridization and polyploidization. *American Journal of Botany* 99: 1043–1057.
- RIESEBERG, L. H., AND J. H. WILLIS. 2007. Plant speciation. *Science* 317: 910–914.
- RIGGINS, C. W., AND D. S. SEIGLER. 2012. The genus *Artemisia* (Asteraceae: Anthemideae) at a continental crossroads: Molecular insights into migrations, disjunctions, and reticulations among Old and New World species from a Beringian perspective. *Molecular Phylogenetics and Evolution* 64: 471–490.
- RODRIGUEZ, R. J., D. C. FREEMAN, E. D. MCARTHUR, Y. O. KIM, AND R. S. REDMAN. 2009. Symbiotic regulation of plant growth, development and reproduction. *Communicative & Integrative Biology* 2: 141–143.
- ROZEN, S., AND H. SKALETSKY. 1999. Primer3 on the WWW for general users and for biologist programmers. In S. Misener, S. A. Krawetz, S. Rozen, and H. Skaletsky [eds.], *Bioinformatics methods and protocols*, 365–386. Humana Press, Totowa, New Jersey, USA.
- SANDERSON, S. C., AND H. C. STUTZ. 2001. Chromosome races of fourwing saltbush (*Atriplex canescens*), Chenopodiaceae. In E. D. McArthur and D. J. Fairbanks [compilers], *Shrubland ecosystem genetics and biodiversity: proceedings*, 75–88. U. S. Department of Agriculture, Forest Service Rocky Mountain Research Station Proceedings, RMRS-P-21, Fort Collins, Colorado, USA.
- SHAFIZADEH, F., AND A. B. MELNIKOFF. 1970. Coumarins of *Artemisia tridentata* ssp. *vaseyana*. *Phytochemistry* 9: 1311–1316.
- SHAW, N. L., A. M. DEBOLT, AND R. ROSENTERER. 2005. Reseeding big sagebrush: Techniques and issues. In N. L. Shaw, M. Pellant, and S. B. Monsen [compilers], *Sage-grouse habitat restoration symposium proceedings*, 99–108. U. S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Proceedings RMRS-P-38, Fort Collins, Colorado, USA.
- SHULTZ, L. M. 2009. Monograph of *Artemisia* subgenus *Tridentatae* (Asteraceae-Anthemideae). *Systematic Botany Monographs* 89: 1–131.
- SLOTTE, T., H. HUANG, M. LASCoux, AND A. CEPLITIS. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Molecular Biology and Evolution* 25: 1472–1481.

- SOLTIS, D. E., P. S. SOLTIS, D. W. SCHEMSKE, J. F. HANCOCK, J. N. THOMPSON, B. C. HUSBAND, AND W. S. JUDD. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.
- SOLTIS, P. S., AND D. E. SOLTIS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences, USA* 97: 7051–7057.
- STAJICH, J. E., AND M. W. HAHN. 2005. Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution* 22: 63–73.
- STANTON, D. J., E. D. MCARTHUR, D. C. FREEMAN, AND E. M. GOLENBERG. 2002. No genetic substructuring in *Artemisia* subgenus *Tridentatae* despite strong ecotypic subspecies selection. *Biochemical Systematics and Ecology* 30: 579–593.
- STEBBINS, L. J. 1950. Variation and evolution in plants. Columbia University Press, New York, New York, USA.
- STEVENS, R., AND E. MCARTHUR. 1974. A simple field technique for identification of some sagebrush taxa. *Journal of Range Management* 27: 325–326.
- ST. ONGE, K. R., J. P. FOXE, J. LI, H. LI, K. HOLM, P. CORCORAN, T. SLOTTE, M. LASCoux, AND S. I. WRIGHT. 2012. Coalescent-based analysis distinguishes between allo- and autopolyploid origin in shepherd's purse (*Capsella bursa-pastoris*). *Molecular Biology and Evolution* 29: 1721–1733.
- STIFT, M., R. BREGMAN, J. G. B. OOSTERMEIJER, AND P. H. VAN TIENDEREN. 2010. Other tetraploid species and conspecific diploids as sources of genetic variation for an autotetraploid. *American Journal of Botany* 97: 1858–1866.
- TKACH, N., M. H. HOFFMANN, M. RÖSER, A. A. KOROBKOV, AND K. B. VON HAGEN. 2008. Parallel evolutionary patterns in multiple lineages of arctic *Artemisia* L. (Asteraceae). *Evolution* 62: 184–198.
- VALLÈS, J., S. GARCIA, O. HIDALGO, J. MARTIN, J. PELLICER, M. SANZ, AND T. GARNATJE. 2011. Biology, genome evolution, biotechnological issues and research including applied perspectives in *Artemisia* (Asteraceae). In J. Kader and M. Delseny [eds.], *Advances in botanical research*, vol. 60, 349–419. Academic Press, Burlington, Massachusetts, USA.
- VALLÈS, J., AND E. MCARTHUR. 2001. *Artemisia* systematics and phylogeny: Cytogenetic and molecular insights. In E. D. McArthur and D. J. Fairbanks [compilers], *Shrubland ecosystem genetics and biodiversity: Proceedings*, 67–74. U. S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Proceedings RMRS-P-21, Fort Collins, Colorado, USA.
- VAN DE PEER, Y., S. MAERE, AND A. MEYER. 2009. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* 10: 725–732.
- WAMBOLT, C. L. 1996. Mule deer and elk foraging preference for 4 sagebrush taxa. *Journal of Range Management* 49: 499–503.
- WANG, H., E. D. MCARTHUR, S. C. SANDERSON, J. H. GRAHAM, AND D. C. FREEMAN. 1997. Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* 51: 95–102.
- WATSON, L. E., P. L. BATES, T. M. EVANS, M. M. UNWIN, AND J. R. ESTES. 2002. Molecular phylogeny of Subtribe Artemisiinae (Asteraceae), including *Artemisia* and its allied and segregate genera. *BMC Evolutionary Biology* 2: 17.
- WHEAT, C. W. 2010. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* 138: 433–451.
- YANG, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.
- YANG, Z., AND B. RANNALA. 2012. Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics* 13: 303–314.
- YUNFA, M., M. QINGQUAN, F. XIAOMIN, Y. XIAOLI, W. FULI, AND S. CHUNHUI. 2011. Origin and development of *Artemisia* (Asteraceae) in Asia and its implications for the uplift history of the Tibetan Plateau: A review. *Quaternary International* 236: 3–12.