Article Type: Research Article

RESEARCH ARTICLE

Running title: Klein et al.—Evolutionary relationships among North American *Vitis*

High-throughput sequencing data clarify evolutionary relationships among North American *Vitis* species and improve identification in USDA *Vitis* germplasm collections

Laura L. Klein[1], Allison J. Miller[1,5], Claudia Ciotir[1], Katie Hyma[2], Simon Uribe-Convers[3], and Jason Londo[4,5]

Manuscript received 4 November 2017; revision accepted 4 January 2018.

[1] Department of Biology, Saint Louis University, St. Louis, MO 63110, USA

[2] Cornell University, Institute for Biotechnology, Cornell University, Ithaca, NY 14853, USA. (Current address: Syracuse University College of Law, 950 Irving Ave., Syracuse, NY 13244, USA)

[3] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.

[4] United States Department of Agriculture, Agricultural Research Service, Grape Genetics Research Unit, Geneva, NY 14425, USA.

[5] Authors for correspondence (e-mail: allison.j.miller@slu.edu, jason.londo@ars.usda.gov)

Citation: Klein, L. K., A. J. Miller, C. Ciotir, K. Hyma, S. Uribe-Convers[3], and J. Londo. 2018. High-throughput sequencing data clarify evolutionary relationships among North American *Vitis* species and improve identification in USDA *Vitis* germplasm collections. *American Journal of Botany* 105(2): XXX.

DOI: XXXX

PREMISE OF THE STUDY: Grapes are one of the most economically important berry crops worldwide, with the vast majority of production derived from the domesticated Eurasian species *Vitis vinifera*. Expansion of production into new areas, development of new cultivars, and concerns about adapting grapevines for changing climates necessitate the use of wild grapevine species in breeding programs. Diversity within *Vitis* has long been a topic of study; however, questions remain regarding relationships between species. Furthermore, the identity of some living accessions is unclear.

METHODS: This study generated 11,020 single nucleotide polymorphism (SNP) markers for more than 300 accessions in the USDA-ARS grape germplasm repository using genotyping-by-sequencing. Resulting data sets were used to reconstruct evolutionary relationships among several North American and Eurasian *Vitis* species, and to suggest taxonomic labels for previously unidentified and misidentified germplasm accessions based on genetic distance.

KEY RESULTS: Maximum likelihood analyses of SNP data support the monophyly of *Vitis*, subg. *Vitis*, a Eurasian subg. *Vitis* clade, and a North American subg. *Vitis* clade. Data delineate species groups within North America. In addition, analysis of genetic distance suggested taxonomic identities for 20 previously unidentified *Vitis* accessions and for 28 putatively misidentified accessions.

CONCLUSIONS: This work advances understanding of *Vitis* evolutionary relationships and provides the foundation for ongoing germplasm enhancement. It supports conservation and breeding efforts by contributing to a growing genetic framework for identifying novel genetic variation and for incorporating new, unsampled populations into the germplasm repository system.

KEY WORDS: genotyping-by-sequencing; germplasm; grapevine; phylogenomics; Vitaceae; *Vitis*

Understanding evolutionary relationships among crops, their wild progenitors, and close relatives provides the requisite framework for conserving and using crop genetic diversity (Fielder et al., 2015; Dempewolf et al., 2017; Migicovsky and Myles, 2017). While the evolutionary histories of many annual crop species have been reconstructed, well-resolved phylogenies remain elusive for many crop genera, in particular those that include woody perennials (Barakat et al., 2012).

Long-lived plants such as woody vines and trees have several basic biological attributes that complicate phylogenetic reconstruction: they are often obligate outcrossers that are highly heterozygous, undergo extensive interspecific hybridization, exhibit little among-population variation, and commonly share haplotypes among species (Petit and Hampe, 2006). Traditional approaches to molecular phylogenetics, including the sequencing of chloroplast and nuclear genes, have contributed to the resolution of relationships in some groups (Soltis et al., 1999; Rokas et al., 2003). The advent of high-throughput sequencing and analysis has greatly enhanced our capacity to analyze hundreds of thousands of sites from across the genome and offers great potential to advance resolution of relationships in groups that have posed challenges to traditional phylogenetic approaches (e.g., Cavender-Bares et al., 2015; Hipp et al., 2014, Uribe-Convers et al., 2016).

Approximately 75% of woody perennial crops are clonally propagated, including most fruit and nut trees (Zohary and Spiegel-Roy, 1975; McKey et al., 2010; Miller and Gross, 2011; Warschefsky et al., 2016). Clonal propagation allows growers to select desirable individuals and replicate them in future plantings. While the scientific community relies on seed banks to preserve variation in plants that are grown from seed, clonally propagated plants are often maintained in living collections (e.g., U.S. Department of Agriculture, National Tropical Breadfruit Institute), which cultivated varieties as well as a wide range of wild-collected, closely related accessions critical for breeding programs and scientific study. A primary goal of managing these collections is to ensure accurate identification of clonal varieties as well as closely related species (National Research Council, 1991).

The genus *Vitis* L. includes one of the most economically important berry crops in the world, the domesticated European grapevine (*V. vinifera* L.), which comprises the vast majority of

cultivated grapevines worldwide. *Vitis vinifera* is grown for its berries, which are used to produce fresh grapes, raisins, grape seed oil, grape juice, and wine. However, native North American grapevine species have been of interest for centuries. Early American viticulturalists recognized their potential to produce wine grapes in regions where *V. vinifera* succumbs to pest and pathogen pressure (Kilman, 2010). Several North American grapevines are cultivated for their berries (e.g., *V. labrusca* L.). Many of these native North American species have been used in the generation of both early French-American and modern day hybrid cultivars for wine grape production (e.g., *V. aestivalis* Michx., *V. cinerea* (Engelm.) Engelm. ex Millardet, *V. riparia* Michx., *V. rupestris* Scheele, etc.). Furthermore, when the root-feeding North American aphid, phylloxera (*Daktulosphaira vitifoliae* Fitch), was accidentally introduced into Europe in the mid-nineteenth century, phylloxera-tolerant North American subg. *Vitis* species (e.g., *V. riparia*, *V. rupestris*) began to be used as rootstocks for the global grape industry.

In North America, breeding efforts have helped drive growth of grapevine production in many regions outside of the Mediterranean-like climate of California (e.g., Missouri, New York, Texas, and Virginia, among others). One component of this growth is the development of grapevines that can withstand abiotic and biotic stress in areas not traditionally used for viticulture. Beyond their economic contributions, North American grapevines are a charismatic component of the North American flora, occurring throughout the eastern, central, and southwestern United States, southern Canada, and northern Mexico.

Recent taxonomic revisions indicate that there are approximately 70 *Vitis* species distributed in the North Temperate Zone in Europe, Asia, and North America (e.g., Ren and Wen, 2007; Moore and Wen, 2016). *Vitis* species are climbing or sprawling lianas with alternate leaves, tendrils, or clusters (flowers or fruits) opposing the leaves at each node, paniculate inflorescences, berries, and seeds with a distinct chalaza (Zhang et al., 2015; Moore and Wen, 2016). Phylogenetic analyses support the monophyly of *Vitis* based on chloroplast sequence data (Terral et al., 2009; Tröndle et al., 2010; Zecca et al., 2012; Liu et al., 2016) and nuclear sequence data (Zecca et al., 2012; Wan et al., 2013; Liu et al., 2016). These data indicate that New World members of the Vitaceae genus *Ampelocissus* Planch. are sister to *Vitis*, and place New World *Vitis* as sister to the remaining species within the genus (Liu et al., 2016).

Within *Vitis* two subgenera have been defined: subg. *Muscadinia* (Planch.) Rehder (2N = 40; two species in Southeastern and South Central North America) and subg. *Vitis* (2N = 38; ~60+ species in Asia, Europe, and North America). Phylogenetic analyses place members of subg. *Muscadinia* (*V. rotundifolia* Michx. and *V. poponoi* J.L. Fennell) together, distinct from all members of a monophyletic subg. *Vitis* (Péros et al., 2011; Tröndle et al., 2010; Zecca et al., 2012; Miller et al., 2013; Wan et al., 2013; Liu et al., 2016). Species of subg. *Muscadinia* occur primarily in southeastern United States and central America and have simple tendrils and continuous pith through the nodes—traits that are lacking in subg. *Vitis* (Wen et al., 2007). Subgenus *Vitis* taxa share the same chromosome number and the presence of bifid or trifid tendrils and nodal diaphragms (Liu et al., 2016).

Vitis subg. *Vitis* has been the subject of numerous investigations aimed at clarifying evolutionary relationships in the group. As the larger of the two subgenera, subg. *Vitis* includes approximately 19 species in North America and 38 species in Eurasia (Wan et al., 2013). Relationships among Asian and North American subg. *Vitis* species have been of particular interest. Early analyses suggested that Asian subg. *Vitis* is paraphyletic with North American subg. *Vitis* nested within it (Tröndle et al., 2010; Péros et al., 2011). Wan et al. (2013) suggest the opposite scenario, with a monophyletic Asian subg. *Vitis* clade nested within a paraphyletic North American subg. *Vitis* group. Mounting evidence for a third possibility identifies two large clades within subg. *Vitis*: a clade with North American species and a clade with European and Asian species (Zecca et al., 2012; Miller et al., 2013; Liu et al., 2016).

Current work in *Vitis* focuses on elucidating relationships within the Eurasian subg. *Vitis* clade and within the North American subg. *Vitis* clade. Among the Eurasian group, genetic analyses have confirmed that the cultivated grapevine (*V. vinifera* ssp. *vinifera*) is derived from native European grapevine populations (*V. vinifera* ssp. *sylvestris* (C.C. Gmel.) Hegi; Myles et al., 2011), and that these two taxa are distinct from the remaining wild Asian *Vitis* species (Tröndle et al., 2010; Zecca et al., 2012; Miller et al., 2013; Wan et al., 2013; Liu et al., 2016). Liu et al. (2016) identified at least five clades of Asian *Vitis* species, but the relationships among these groups and the species within them are still under investigation.

Here, we mainly focus on reconstructing relationships among North American subg. *Vitis* species. Evidence presented to date recognizes a group of primarily (but not exclusively) southwestern and southcentral species including *V. acerifolia* Raf., *V. arizonica* Engelm., *V. girdiana* Munson, *V. riparia*, and *V. rupestris* (Zecca et al. 2012; Miller et al., 2013; Wan et al., 2013). Some evidence suggests a second group includes *V. aestivalis, V. cinerea, V. labrusca*, and *V. vulpina* L. (Miller et al., 2013; Liu et al., 2016). However, support for these groups is not strong and the relationships of other taxa within North American subg. *Vitis* remain unclear.

With the goal of contributing to current understanding of evolutionary relationships among North American members of subg. *Vitis*, we leveraged data generated for the United States Department of Agriculture—Agricultural Research Service (USDA-ARS) collections of wild grapevines housed at the Plant Genetic Resource Unit (Geneva, New York) and National Clonal Germplasm Repository (Davis, California). These valuable living collections comprise nearly 5000 grapevine accessions, including extensive collections of *V. vinifera* cultivars and hybrids and mapping populations. The USDA-ARS collections include other temperate Vitaceae species like *Ampelopsis spp.*, a widely distributed relative of *Vitis* (Liu et al., 2016). Within these collections exist hundreds of wild-collected accessions from North America and Asia. Living collections of perennial crops and their wild relatives are valuable resources for breeding, but also offer unprecedented opportunities to address basic questions in plant biology and evolution (e.g., Chitwood et al., 2014, 2016a,b). Ongoing genetic analyses of the wild *Vitis* germplasm housed at the USDA-ARS repositories include the generation of single nucleotide polymorphisms (SNPs) by genotyping-by-sequencing (Elshire et al., 2011) for the USDA National Institute of Food and Agriculture's Specialty Crops Research Initiative VitisGen project (www.Vitisgen.org; Hyma et al., 2015).

Here, we address outstanding questions in *Vitis* evolutionary biology through novel phylogenomic analyses of genotyping-by-sequencing data. The goals of this study were to (1) reconstruct evolutionary relationships among North American subg. *Vitis* species and outgroups and to identify major clades of North American grapevines, and (2) use genetic distance data to resolve unidentified and misidentified accessions in the USDA-ARS collection. This study

extends current understanding of *Vitis* species relationships and enhances valuable USDA-ARS collections through improved accession identification.

<h1>MATERIALS AND METHODS

<h2>*Sampling and DNA Extraction*

Individuals used in this study (*n* = 359, Appendix S1; see Supplemental Data with this article) represent wild grapevine germplasm preserved in the USDA germplasm repository system. DNA from these samples were extracted, libraries constructed, and sequenced in coordination with the VitisGen project (www.Vitisgen.org; Hyma et al., 2015) and thus represent a subset of samples from a data set of 8353 *Ampelopsis* and *Vitis* samples. Coordination with this larger data set allowed for the development of the bioinformatic pipeline used to call SNPs from such a diverse species collection. We selected samples that occur primarily in Eastern North America, because this represents the majority of collections housed at the USDA-ARS Plant Genetic Resources Unit (Geneva, New York). All samples came from living germplasm collections maintained by the USDA-ARS Plant Genetic Resources Unit and the National Clonal Germplasm Repository (Davis, California; Table 1). In total, 24 *Vitis* species (12 North American species, seven Eurasian species) and four *Ampelopsis* species are included in this study, representing approximately one third of the known diversity in *Vitis*. While not exhaustive, this sampling scheme offers species-level representation on par with similar studies and provides a solid framework for the investigation of phylogenetic relationships and accession identification.

Tissue collection, DNA extraction, quantification, and sequencing follow methods outlined in Hyma et al. (2015). A young leaf (less than 1 cm diameter) was collected from each individual and placed in a tube of a 96-well cluster tube collection plate. Tissues were ground using a Geno/Grinder 2000 (OPS Diagnostics LLC, Lebanon, New Jersey) following the addition of two stainless steel beads to each tube and freezing the plate at –80˚C. DNA was extracted using DNeasy 96-well DNA extraction kits (Qiagen, Valencia, California) with the addition of PVP-40 (2% w/v) to the AP1 lysis buffer to ensure DNA quantity and quality. DNA quantification was performed using the QuantiFlor dsDNA System (Promega, Madison, Wisconsin).

<h2>*Genotyping-by-Sequencing*

Samples containing ≥10 ng/μl DNA were processed by Cornell University's Genomic Diversity Facility (Ithaca, New York) for genotyping-by-sequencing (GBS; Elshire et al., 2011). Relative to other reduced representation library sequencing protocols, this approach has fewer PCR steps and no fragment size selection, which results in thousands of SNPs from throughout the genome at low coverage (Davey et al., 2011). GBS utilizes restriction enzymes to digest DNA and ligate adaptor sequences in a single well. DNA was digested using the *Ape*KI restriction enzyme, and DNA fragments were then ligated with unique barcode and common adaptors. Next, all samples were combined for PCR purification in preparation for sequencing. Sequencing was performed on an Illumina HiSEq. 2000 (Illumina Inc., San Diego, California). These data were then filtered to ensure retention of quality, high-coverage SNP sites. Raw sequence fastq files can be found at the Sequence Read Archive under bioproject accessions SAMN07808873-SAMN07809231.

Sequence tags were aligned to the 12XV2 *V. vinifera* reference genome PN40024 (Jaillon et al., 2007; Adam-Blondon et al., 2011) using the Burrows-Wheeler Alignment (BWA) version 0.6.2-r126 (Li and Durbin, 2009), and SNPs were called using the TASSEL-GBS pipeline version 3.0.139 (Glaubitz et al., 2014) at the Genomic Diversity Facility. The original VitisGen data set included wild, cultivated, hybrid, and mapping population samples (Hyma et al. 2015). The SNP database generated for this phylogenomic study includes wild, cultivated, and hybrid accessions of *Vitis* and close relatives, but excludes mapping populations. Sequencing of 359 individuals used here resulted in 1,660,674 SNPs (Appendix S1). Using VCFtools v01.11 (http://vcftools.sourceforge.net/; Danecek et al., 2011), SNPs were filtered to retain only biallelic sites with a minimum allele frequency of 0.01, and a minimum mean depth of 10x. Only sites with <20% missing data and individuals with <20% missing data were retained in final analyses. Following filtering, the data set included 359 individuals with 11,020 SNPs.

To prepare the data for phylogenetic analyses, variant SNPs were concatenated and reformatted. The vcf files were converted into a tab-delimited text (.txt) file using the 'vcf-to-tab' command in VCFTools. Next, each biallelic SNP site was converted to a consensus single allele SNP state using the International Union of Pure and Applied Chemistry (IUPAC) coding system (custom scripts available at https://github.com/uribe-convers/Vitis_Phylogenomics). Finally, file format

conversions were performed using the programs Phyutility v.2.2 (Smith and Dunn, 2008) and NCLconverter v2.1 (David et al., 2012).

<h2>*Phylogenetic analysis*

To explore relationships among *Vitis* species, we constructed three data sets (Table 1): (1) a combined *Vitis* and *Ampelopsis* Michx. data set—359 accessions used to confirm the monophyly of genus *Vitis*; (2) a *Vitis*-only data set in which *Ampelopsis* individuals were removed—304 individuals used to maximize SNP variation within *Vitis* to evaluate unidentified/misidentified individuals; and (3) a reduced *Vitis* data set—87 individuals (1–5 individuals for each *Vitis* species sampled) used for phylogenetic analysis with SVDquartets (Chifman and Kubatko, 2014), a program designed specifically for large data sets of unlinked loci and small sample sizes.

<h2>*Monophyly of* Vitis

The full data set of 359 individuals including *Vitis* and *Ampelopsis* (11,020 SNPs) was used to test the monophyly of *Vitis* using RAxML v8.2.9 (Stamatakis, 2014). The data set was run on CIPRES Science Gateway V.3.3 (Miller et al., 2010) with parameters set to rapid bootstrapping, a maximum likelihood convergence criterion, a GTRCAT model of nucleotide evolution, and 1000 replicates of nonparametric bootstrapping. A majority rule consensus was produced from the 1000 bootstrap trees, and branches with bootstrap support of less than 50 were collapsed.

<h2>*Phylogenetic relationships within* Vitis

To conduct phylogenetic analyses, we used SVDquartets (Chifman and Kubatko, 2014, 2015). GBS data consist of variant sites, which can reflect acquisition bias (Leaché et al., 2015). These biases can cause branch length overestimation when using traditional programs such as RAxML or BEAST to reconstruct phylogeny (Drummond et al., 2012). SVDquartets is a method of phylogenetic inference that attempts to correct for acquisition bias. This method evaluates the optimal relationship between four taxa in the data set by randomly sampling quartets from the data matrix. For each quartet, a singular value decomposition (SVD) score is generated for each of the three possible splits among the four taxa. The split with the best (lowest) score is selected (Chifman and Kubatko, 2014). Following quartet sampling and evaluation, the program then

reconstructs the quartets into a phylogenetic tree. Because of the intractable number of quartets that are possible for 304 species, we generated a reduced data set (87 individuals; Table 1). We used SVDquartets to sample all possible quartets (1,466,127) for these taxa and ran nonparametric bootstrapping with 100 replicates.

<h2>*Unidentified and misidentified accessions*

To determine unidentified and putatively misidentified accessions, we used a combination of two analyses to suggest the most likely taxonomic identification of a given accession. First, we generated a genetic distance tree using neighbor joining (NJ) based on Nei's genetic distance model implemented in Geneious v10.1.2 (http://www.geneious.com; Biomatters Inc., Auckland, New Zealand; Kearse et al., 2012). These analyses were carried out for the *Vitis*-only data set (304 individuals), which was generated to maximize the number of accessions analyzed and the number of SNPs identified within *Vitis* (Table 1). The consensus tree was rooted with *V. rotundifolia* (subg. *Muscadinia*) as the outgroup to subg. *Vitis*. One thousand bootstrap replicates were run to assess support, and branches with less than 50% support values were collapsed. The resulting tree was inspected manually. Taxon names were suggested for previously unidentified accessions based on genetic similarity and the identity of adjacent accessions in the tree. Putatively misidentified taxa were described from situations where a single accession clustered in a group of accessions with a different taxonomic name.

Second, we examined the first four coordinates of a multidimensional scaling plot generated from the subg. *Vitis* accessions to observe the congruence of labeled accessions with species clusters. The SNP data were converted from VCF into Plink format (http://zzz.bwh.harvard.edu/plink/; Purcell et al., 2007) to conduct multi-dimensional scaling analysis (MDS). *Vitis rotundifolia* accessions were removed to reduce ordination space and improve visualization of genetic distance among subg. *Vitis* samples (n = 291, Appendix S2). The MDS analysis was conducted with the --noweb, --cluster, and --mds-plot 4 commands, and MDS coordinates were visualized using the R package 'ggplot2' (Wickham, 2009; R Core Team, 2013).

Historical phenotype data stored at the USDA Germplasm Resoursces Information Network (GRIN) database (https://www.ars-grin.gov/) for unidentified and putatively misidentified individuals were examined to cross-verify results of the genetic analysis. Descriptions and photos of unidentified accessions, as well as accessions that were placed together with individuals labeled as something different than the identification of that particular accession, were examined for morphological traits that might provide additional evidence to support phylogenetic results. For example, the trait of flower sex (bisexual in domesticated *V. vinifera*, unisexual in wild *Vitis* species) was used to help identify putative crop-wild hybrids, and leaf shape was used as additional evidence of species identification (Chitwood et al., 2014;2016a).

<h1>RESULTS

<h2>*SNP identification*

Genotyping-by-sequencing of 359 individuals were called from 206 million reads, identifying 1,660,674 SNPs from 64 bp tags before filtering (Table 1). Three data sets including sites with ≥ 10x coverage resulted (Table 1): (1) the combined *Vitis* and *Ampelopsis* data set (359 individuals) included 11,020 SNPs; (2) the *Vitis*-only data set in which *Ampelopsis* individuals were removed (304 accessions) included 10,565 SNPs; and (3) the reduced *Vitis* data set (87 individuals) included 8617 SNPs.

<h2>*Reconstruction of evolutionary relationships among North American subg.* Vitis *species and outgroups*

Maximum likelihood (ML) phylogenetic analyses using the full data set (359 individuals with 11,020 SNPs) performed in RAxML confirm a monophyletic *Vitis* (Appendix S3). This data set, as well as the reduced data set, support a monophyletic subg. *Vitis* (Fig. 1 and Appendix S3). *Vitis* subg. *Muscadinia* (*V. rotundifolia*) is distinct from subg. *Vitis*, supported by 97% bootstrap support (Appendix S3). Subgenus *Vitis*, which includes the vast majority of species in the genus, is strongly supported as a monophyletic group in all data sets.

The reduced *Vitis* data set (87 *Vitis* accessions, 8617 SNPs) used in SVDquartets analyses supports two subgenera and geographically representative clades within subg. *Vitis* (Fig. 1). Within subg. *Vitis,* two clades are resolved: a Eurasian clade and a North American clade. The

Eurasian clade is further subdivided to include a clade of *V. vinifera* accessions (vars. 'Gewurztraminer', 'Fruhburgunder', 'Riesling', 'Ruby Cabernet', and 'Syrah') and a clade of Asian subg. *Vitis* species (*V. amurensis* Rupr. + *V. coignetiae* Pulliat ex Planch. and *V. piasezkii* Maxim. + *V. romanetii* Rom. Caill.). These data illustrate that European and Asian grapevines are evolutionarily distinct lineages that are more closely related to each other than either is to the North American grapevines.

Within North American subg. *Vitis*, analyses of the reduced *Vitis* data set recovered two major clades (Fig. 1; NA Clade I, NA Clade II). North American Clade I includes *V. acerifolia, V. arizonica, V. monticola* Buckley, *V. riparia,* and *V. rupestris*. Within NA Clade I, accessions of the same species generally cluster together with a few exceptions. Accessions of *V. arizonica* form a clade with the exception of *V. arizonica* DVIT 1872, which is recovered outside of the group that includes *V. acerifolia, V. arizonica, V. riparia,* and *V. rupestris*. There is one *V. monticola* accession included in this group. Also intriguing is the placement of *V. labrusca* 483164 in this clade; other *V. labrusca* samples are recovered in a separate clade (see below), suggesting this individual may be misidentified or perhaps is of hybrid descent.

A second clade of North American subg. *Vitis* species consists of *V. aestivalis, V. cinerea, V. labrusca, V. mustangensis* Buckley, *V. palmata* Vahl, *V. shuttleworthii* House, and *V. vulpina* (Fig. 1; NA Clade II). Data presented here support further division of this group into two subclades. North American Clade IIa includes *V. mustangensis, V. palmata,* and *V. shuttleworthii,* where *V. mustangensis* and *V. shuttleworthii* are sister taxa. North American Clade IIb includes *V. aestivalis, V. cinerea, V. labrusca,* and *V. vulpina*. In NA Clade IIb, four *V. aestivalis* accessions cluster together, while a fifth *V. aestivalis* accession (483185) does not cluster with the other *V. aestivalis* accessions; rather, it is nested within a group of *V. vulpina* accessions. It is worth noting that the previous entry in the USDA-GRIN system denoted this accession as *V. vulpina.* Data presented here suggest that the original classification of this accession was correct (https://www.ars-grin.gov/).

<h2>*Resolution of unidentified and misidentified accessions in the USDA-ARS collection*

A second application of the GBS-derived SNP data is to use genetic data to improve identification of vines in the USDA collection. As described above, the tree generated using the reduced *Vitis* data set (87 accessions; Fig. 1) identified two individuals as potentially misidentified (*V. labrusca* 483164 in NA Clade 1 and *V. aestivalis* 483185 in NA Clade IIb). To further explore identification within the USDA *Vitis* collection we used the *Vitis*-only data set (*n* = 304; Table 1).

The NJ analysis of accessions—based on the 10,565 SNPs generated for the 304 accessions in the *Vitis*-only data set—offered a powerful tool with which to examine taxonomic identity of USDA accessions. Genetic similarity among accessions revealed that several previously unidentified accessions were nested within clusters of accessions of known identity (Fig. 2; Appendix S2). Species identifications for 20 previously unidentified accessions in the USDA collection were suggested based on this analysis (Appendix S4). These accessions fall within clades that included multiple accessions of a single species, such as *V. acerifolia, V. labrusca, V. mustangensis, V. palmata, V. piasezkii, V. riparia,* and *V. yenshanensis* J.-X.Chen (Fig. 2, black arrows/bolded text). Some of these accessions form small subclades within accessions of known species identities such as *V. riparia* (grouping together with other accessions labeled as *V. acerifolia*), *V. piasezkii,* and *V. yenshanensis*. These accessions were cross checked with USDA-GRIN data records on the geographic origins of the acquisition, and were also assessed for diagnostic traits including flower sex and leaf shape to confirm the suggested identification based on GBS data (Appendix S4).

Additionally, the NJ tree of the *Vitis*-only data set identified 28 accessions whose original taxonomic identification did not correspond to their placement within the clade, suggesting they may be misidentified within the collection. Individuals *V. aestivalis* DVIT2203_6, 483185, and 588626, *V. amurensis* DVIT2006_1, *V. cinerea* GVIT171, *V. coignetiae* 588451e, *V. labrusca* 597104 and 483130, *V. palmata* DVIT2227_1, and *V. yenshanensis* 588422a were placed among species other than their labeled accession identities (Fig. 2, red arrowheads/text).

To examine identification discrepancies within the collection further, we used MDS. We compared MDS coordinates for species groups to identify individuals whose placement in MDS

space was inconsistent with their original taxonomic assignment (Appendices S2 and S3). The MDS plots clarified the placement of unidentified or misidentified subg. *Vitis* accessions (*n* = 291; Fig. 3). In Fig. 3, accessions of the same species clustered together with some previously unidentified (black triangles) and misidentified (red triangles) individuals (e.g., *Vitis* spp. 588501 is placed among *V. labrusca* accessions, *V. cinerea* GVIT 171 is placed among *V. vulpina* accessions, etc.). Some unidentified and misidentified accessions appeared as intermediate to species clusters, suggesting they are likely hybrids. These putative hybrid individuals formed distinct clusters in Fig. 2 (red label bars). For example, eight individuals (red triangles) fall between a cluster of *V. labrusca* and the Eurasian subg. *Vitis* species (Fig. 3A), which are sister to the core *V. labrusca* group in Fig. 2 ("*V. labrusca* hybrids," red arrowheads and text). Similarly, three individuals that were previously identified as either *V. aestivalis* 'bicolor' or *V. labrusca* (Fig. 2 "*V. aestivalis* hybrids," red arrowheads and text) are between *V. aestivalis* and *V. labrusca* in MDS ordination space (Fig. 3A). The eight individuals that are labeled "*V. riparia* hybrids" in the NJ tree (Fig. 2) also fall intermediate between the cluster of *V. riparia* accessions and Eurasian subg. *Vitis* accessions in the MDS plot (Fig. 3A, red triangles). By visualizing genetic distance with NJ and MDS, we increase identification accuracy for unidentified and misidentified accessions within the USDA collection.

<h1>DISCUSSION

Data presented here contribute to a growing body of literature using next-generation sequencing to clarify phylogenetic relationships and to confirm accession identity in valuable living collections. Genotyping-by-sequencing has been a cost-effective method for SNP generation, which has been used in a wide range of crops and other species (Elshire et al., 2011; Davey et al., 2011; Poland and Rife, 2012; He et al., 2014; McAllister and Miller, 2016; Migicovsky and Myles, 2017). In this study, GBS data were generated for grapevines held in the USDA-ARS germplasm collections in Geneva, NY and Davis, CA. Results offer novel insights into evolutionary relationships within *Vitis* and refine identification for some accessions in the USDA collection.

<h2>*Improved resolution of relationships among North American grapevines*

Consistent with previous analyses, this study confirms that *Vitis* and subg. *Vitis* are monophyletic (Zecca et al., 2012; Miller et al., 2013; Wan et al., 2013; Liu et al., 2016), and provides additional support for separate Eurasian and North American clades in subg. *Vitis* (Tröndle et al., 2010; Zecca et al., 2012; Miller et al., 2013; Zhang et al., 2015). Within North American subg. *Vitis*, this study supports two major clades (NA Clades I and II; Fig. 1). Our NJ analysis (Fig. 2) disagrees with the species tree produced with the reduced data set (Fig. 1), placing the Eurasian subg. *Vitis* clade within the North American clade. While genetic distance matrices like NJ methods are useful for identifying genetic similarity among taxa, NJ is not designed to assess species relationships with large SNP data sets where ascertainment bias may be present (Clark et al., 2005; Ollitrault et al., 2015; but see Miller et al., 2013). Therefore, we consider the species tree produced by SVDquartets to be a more representative depiction of species relationships among *Vitis* species in this study. These data contribute additional resolution to a growing body of literature describing major groups of *Vitis* species in North America, facilitating enhanced understanding of morphology and biogeography.

Within subg. *Vitis*, North American Clade I includes *V. acerifolia*, *V. arizonica*, *V. monticola*, *V. riparia*, and *V. rupestris*, a group that has been recognized by some other studies (Zecca et al., 2012; Miller et al., 2013; Liu et al., 2016). This group may also include *V. bloodworthiana* Comeaux, *V. blancoi* Munson, *V. flexuosa* Thunb., *V. girdiana*, and *V. treleasei* Munson ex L.H. Bailey (Miller et al., 2013; Wan et al., 2013). North American Clade I includes and expands on taxa included by Moore (1991) in his series *Ripariae*, members of which share several morphological traits. For example, Moore (1991) and Moore and Wen (2016) report that *V. acerifolia*, *V. riparia*, and *V. rupestris* have a pith that is interrupted by nodal diaphragms, which are usually < 1 mm in diameter and not typically red-banded, branches that are mostly terete, glabrous, with branchlet tips enveloped by unfolding leaves. They have large stipules (> 3 mm long) and abaxial leaf surfaces that are not glaucous and usually glabrous to slightly arachnoid pubescent (Moore, 1991; Moore and Wen, 2016). Additionally, *V. acerifolia* and *V. rupestris* often have shrubby, low-climbing growth habits compared to the more typical moderate-to-high climbing habit of typical *Vitis* species (Moore, 1991).

North American Clade I expands on series *Ripariae* with the inclusion of *V. arizonica* and *V. monticola*. While Moore (1991) did not include *V. arizonica* in his treatment, he listed *V. monticola* among series *Cordifoliae*. This series shares the majority of morphological traits that unite series *Ripariae*; differences include nodal diaphragms usually > 1 mm in diameter, branchlet growing tips mostly not enveloped by unfolding leaves, stipules < 3 mm long (Moore, 1991). *Vitis arizonica* also shares morphological traits with species in NA Clade I, including a shrubby growth habit, terete branches at maturity, non-red-banded nodal diaphragms, and an abaxial leaf surfaces moderately to thinly arachnoid pubescent (Moore and Wen, 2016).

Data presented here and elsewhere (Miller et al., 2013; Wan et al., 2013) suggest that the southwestern/central southern species *V. arizonica* and *V. rupestris* are sister to the rest of the species in the NA Clade I. This pattern may reflect diversification in response to North American Quaternary glacial cycles (Mullins et al., 1992). As temperatures shifted and refugial populations cycled, expansion, retraction, adaptation, and subsequent speciation likely contributed to contemporary patterns of North American *Vitis* biogeography (Péros et al., 2010; Wan et al., 2013). Environmental niche modeling demonstrates that contemporary distributions of North American Clade I species occupy divergent environmental niches (Callen et al., 2016). For example, the first diverging taxon *V. arizonica* inhabits the warmest and driest niches of all North American *Vitis* species east of the Rocky Mountains; *V. rupestris*, sister to the rest of the group, is found on rocky gravel bars. *Vitis acerifolia* extends to the west in drier environments, and is sister to *V. riparia,* which occupies the coldest and driest niche in the clade (Callen et al., 2016). Phylogenetic relationships described in this manuscript and elsewhere, in combination with environmental niche modeling, suggest diverse environmental conditions may have driven the diversification of *Vitis* species in North American Clade I.

The second clade of North American subg. *Vitis* species (NA Clade II) consists of *V. aestivalis, V. cinerea, V. labrusca, V. mustangensis, V. palmata, V. shuttleworthii,* and *V. vulpina*. This group includes species from Moore's (1991) Series *Aestivales* (includes *V. aestivalis*), Series *Cinercentes* (includes *V. cinerea* and subspecies), Series *Cordifoliae* (includes *V. monticola, V. palmata, V. vulpina*), and Series *Labruscae* (includes *V. labrusca, V. mustangensis,* and *V. shuttleworthii*). This classification groups species in series based on similarities in morphology,

phenology, and habitat preferences. Species within North American Clade II are clustered into two subclades based on GBS data (North American Clades IIa and IIb). These subclades do not conform to the series described by Moore (1991). However, they are largely congruent with species relationships identified by Miller et al. (2013) and Wan et al. (2013).

Species within North American Clade IIa include *V. mustangensis* (series *Labruscae*), *V. palmata* (series *Cordifoliae*), and *V. shuttleworthii* (series *Labruscae*; Fig. 1). Sister taxa *V. mustangensis* and *V. shuttleworthii* share unique morphological traits of abaxial leaf surfaces that are densely tomentose and berries ≥ 12 mm in diameter (Moore and Wen, 2016). *Vitis palmata,* on the other hand, has abaxial leaf surfaces that are glabrous or sparsely pubescent and smaller berries. These three species occur in southern regions of the United States. Their climatic niches are characterized by warm temperature gradients and low ranges of diurnal temperatures (Callen et al., 2016). Despite similarities in some environmental variables, the geographic the distributions of these three species have little to no overlap, perhaps reflecting speciation in allopatry.

North American Clade IIb comprises *V. aestivalis, V. cinerea, V. labrusca,* and *V. vulpina* (Fig. 1). These species clustered together in previous works including Zecca et al. (2012), Miller et al. (2013), and Liu et al. (2016). Wan et al. (2013) also recovered a clade that included *V. aestivalis*, *V. labrusca*, and *V. vulpina*; however, in this analysis *V. cinerea* clustered with other North American species *V. biformis* Rose and *V. palmata*. Neither Munson (1909), Bailey (1934), Galet (1988), nor Moore (1991) identified *V. aestivalis, V. cinerea, V. labrusca,* and *V. vulpina* as a group. These species share core subg. *Vitis* morphological synapomorphies including branched tendrils, exfoliating bark, inconspicuous or absent lenticels, and pith interrupted by nodal diaphragms (Moore and Wen, 2016); however, we do not know of specific traits unique among the four taxa in this group. Within NA Clade IIb, the species pair *V. aestivalis* and *V. labrusca* share the presence of globose berries with skin that separates from the pulp (Moore and Wen, 2016). Furthermore, sister taxa *V. cinerea* and *V. vulpina* are united in the presence of moderate to high climbing, sparsely branched habits, nodal diaphragms 1–4 mm thick, persistent tendrils, branchlet growing tips not enveloped by unfolding leaves, leaf abaxial surfaces that are glabrous

or sparsely to densely arachnoid or hirtellous, and berries 4–12 mm in diameter (Moore and Wen, 2016).

In a contrast to the patterns observed in Clades I and IIa, all species within NA Clade IIb have similar climatic niches and partial geographic overlap throughout the eastern United States (Callen et al., 2016). Despite their climatic similarity and geographic proximity, many of these species have disparate phenological traits like divergence in budburst, flowering, and fruiting, or unique morphological traits such as variable leaf and branchlet shape and pubescence, differences in berry size, and variation in nodal diaphragm thickness and coloration (Moore and Wen, 2016). Species-specific differences in phenological and morphological traits may explain how these species co-occur in similar climates.

<h2>*Accession identification in living collections*

For clonally propagated, long-lived plants, living collections are the primary way in which diversity is preserved ex situ and provide the requisite foundation for breeding programs. Major collections in the US include apples and grapes (Geneva, NY), citrus and dates (Riverside, CA), tree fruit, nut crops, and grapes (Davis, CA), pecans and hickories (College Station, TX), among others (https://ars.usda.gov/). These collections facilitate screening of natural variation for traits of ecological and agricultural importance (Cadle-Davidson, 2008; Cadle-Davidson et al., 2011; Gross et al., 2013).

Passport data including provenance of original collection accompanies the majority of living accessions; however, in some cases these data are incomplete (species identifications are missing) or incorrect. In the past, genetic variation and identification of grapevines in germplasm repositories have relied on morphological keys by collectors and breeders, as well as genetic data derived from a set of SSR markers (https://www.ars-grin.gov/). In large part, these methods have proven effective at classifying wild species material. However, naturally occurring *Vitis* hybrids can present unique challenges. This can result in misidentification of F1, BC1, and later hybrid generations. With the utility of whole genome SNP data, like those presented here, we can use genetic signatures to help place genotypes within species, or to designate accession identity as suspect. Beyond identification, genetic data have been used to quantify variation in living

collections (Hyma et al., 2015; Bielenberg et al., 2015; Migicovsky et al., 2017), to determine the percentage of wild variation housed in living collections (Sawler et al., 2015; Migicovsky et al., 2016), and to reconstruct evolutionary relationships among cultivars and wild relatives (Myles et al., 2011; Miller et al., 2013).

In this study, GBS data were used to suggest taxon names for 20 previously unidentified accessions and for 28 putatively misidentified accessions in the USDA grape collection. For example, *V. aestivalis* accessions 483185 and 588626, as well as *V. cinerea* GVIT 171, all clustered within the *V. vulpina* clade (Fig. 2). Cross referencing these accessions with MDS coordinates demonstrate genomic signatures matching that of other *V. vulpina*. Based on the combination of these results, we can recommend that these three accessions be reidentified as *V. vulpina*. In addition to verifying morphology, this result has implications for any grape breeder or researcher looking to evaluate trait aspects of a given species. This method can be further extended by looking at two major clusters of accessions in the NJ tree; *V. riparia* hybrids which represent a subclade within the *V. riparia/ V. acerifolia/ V. rupestris* clade, and *V. labrusca* hybrids, which represents a subclade of the *V. aestivalis/ V. labrusca* clade. The NJ tree analysis cannot resolve these accessions within the clades because of low bootstrap support, but nevertheless associates them closely with these species groups. However, MDS coordinates demonstrate deviation of these accessions from the species patterns in a way that suggests hybrid ancestry (Fig. 3; Appendix S2, S3).

Vitis diversity included in this study represents approximately one third of the species diversity in the genus, with multiple accessions per taxon. The backbone of phylogenetic results are consistent with previous studies (Zecca et al., 2012; Miller et al., 2013; Liu et al., 2016) and suggestions for unlabeled or previously mislabeled taxa come from their placement amidst several accessions of the same identity. However, identification suggestions presented here are based on a subset of *Vitis* species. It is possible that with more comprehensive sampling, phylogenetic relationships might change and/or that identifications suggested here might shift. Unidentified and putatively misidentified taxa might represent hybrid derivatives whose parents are not be represented in the current sampling scheme. Despite incomplete sampling, we contend that phylogenetic relationships and taxon placement presented in this manuscript offer a

substantial advance in current understanding of *Vitis* phylogeny and USDA accession identification, and that these data contribute to an important foundation for future work.

Living grapevine collections are used in the improvement of existing cultivars and in the development of new ones. Future climate change scenarios suggest climate variation and climate related shifts in biotic pressure are expected to increase pressure on *V. vinifera* production (Hannah et al., 2013; Cook and Wolkovich, 2016). Wild North American *Vitis* species are uniquely adapted to a suite of abiotic and biotic environmental conditions and have been utilized to improve cultivated grapevines. For example, *V. riparia* is extremely pest resistant and winter hardy, *V. labrusca* has unique metabolites that can contribute to new enological profiles, and *V. cinerea* has unique soil adaptations and pest resistance. Hybrid derivatives of *V. cinerea* var. *helleri* (L.H. Bailey) M.O. Moore, *V. riparia*, and *V. rupestris* have yielded some of the most widely-used rootstocks in the viticulture industry (e.g., 1103-Paulsen, 3309-Couderc, and S04; Galet, 1979). Unknown adaptive properties of *Vitis* species may hold additional traits necessary for maintaining future grapevine sustainability.

Living collections housed at the USDA Germplasm Repositories represent a precious source of breeding material for ongoing and future breeding efforts. However, these collections house a subset of known grapevine species; furthermore, accessions represent a relatively small portion of the species distributions and thus a small portion of the genetic and phenotypic diversity of this genus. Further work is needed to expand living collections of contemporary and emerging crops and their wild relatives, and to use genomic and phenomic approaches to characterize diversity in these taxa.

<h1>CONCLUSIONS

Vitis represents one of the most economically important fruit crops in the world. The study presented here adds clarity to the evolutionary relationships among grapevine species. With projected shifts in climatic stability, grapevine producers are expected to be faced with changing abiotic and biotic stresses to which *V. vinifera* may not be well adapted. The results of this study have helped identify 28 misidentified accessions and to better clarify 20 unknown accessions. Perhaps more importantly, this data set demonstrates the power of SNP markers in optimizing

the germplasm collection. We can now use these tools to better sample wild grapevine species distributions and compare genetic signatures with what is already preserved. The future of grapevine breeding and of grapevine sustainability depends on the use of elite germplasm, much of which remains untapped across North America.

<h1>ACKNOWLEDGEMENTS

The authors acknowledge Dr. Prins and Dr. Chao for supplying leaf tissue and access to USDA germplasm repositories, and to all contributors to the USDA grape germplasm collection and those who are contributing to the maintenance of this important collection. We thank Dr. Migicovsky for valuable comments on a previous version of this manuscript. The authors thank two anonymous reviewers who provided helpful comments on a previous version of the manuscript. We acknowledge USDA 2011-51181-30635 Vitisgen to JPL, NSF 1546869 to AJM and JPL, and Saint Louis University for a Graduate Research Assistantship that supported LLK during a portion of this work.

<h1>AUTHOR CONTRIBUTIONS

A.J.M., L.L.K., and J.P.L. conceived of the project; J.P.L. and K.H. generated the data and constructed data sets; and L.L.K., C.C., and S.U.-C. conducted phylogenetic analyses. J.P.L., A.J.M., and L.L.K. developed the manuscript framework, and A.J.M., L.L.K., J.P.L., S.U.-C., and C.C. contributed to manuscript writing and editing.

<h1>LITERATURE CITED

Adam-Blondon, A.F., O. Jaillon, S. Vezzulli, A. Zharkikh, M. Troggio, R. Velasco, and J. Martinez-Zapater. 2011. Genome sequence initiatives. *In* A.F. Adam-Blondon, J.M. Martinez-Zapater, and C. Kole [eds.], Genetics, Genomics, and Breeding of Grapes, 211–234. CRC Press, Boca Raton, Florida, USA.

Bailey, L.H. 1934. The species of grapes peculiar to North America. *Gentes Herbarum* 3: 151–244.

Barakat, A., M. Staton, C.H. Cheng, J. Park, N.M.B. Yassin, S. Ficklin, C.C. Yeh, et al. 2012. Chestnut resistance to the blight disease: insights from transcriptome analysis. *BMC Plant Biology* 12: 38.

Bielenberg, D.G., B. Rauh, S. Fan, K. Gasic, A.G. Abbott, G.L. Reighard, W.R. Okie, and C.E. Wells. 2015. Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. *PLoS One* 10(10): e0139406.

Cadle-Davidson, L., 2008. Variation Within and between *Vitis* spp. for Foliar Resistance to the Downy Mildew Pathogen *Plasmopara viticola*. *Plant Disease* 92(11): 1577–1584

Cadle-Davidson, L., D.R. Chicoine, N.H. Consolie. 2011. Variation within and among *Vitis* spp. for foliar resistance to the powdery mildew pathogen *Erysiphe necator*. *Plant Disease* 95(2): 202-211.

Callen, S.T., L.L. Klein, and A.J. Miller. 2016. Climatic niche characterization of 13 North American *Vitis* species. *American Journal of Enology and Viticulture* 67: 339–349.

Cavender-Bares, J., A. González-Rodríguez, D.A.R. Eaton, A.A.L. Hipp, A. Beulke, and P.S. Manos. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): A genomic and population genetics approach. *Molecular Ecology* 24(14): 3668–3687.

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30(23): 3317–3324.

Chifman, J., and L. Kubatko. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology* 374: 35–47.

Chitwood, D.H., L.L. Klein, R. O'Hanlon, S. Chacko, M. Greg, C. Kitchen, A.J. Miller, and J.P. Londo. 2016a. Latent developmental and evolutionary shapes embedded within the grapevine leaf. *New Phytologist* 210(1): 343–355.

Chitwood, D.H., A. Ranjan, C.C. Martinez, L.R. Headland, T. Thiem, R. Kumar, M.F. Covington, et al. 2014. A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant Physiology* 164(1): 259–272.

Chitwood, D.H., S.M. Rundell, D.Y. Li, Q.L. Woodford, T.T. Yu, J.R. Lopez, D. Greenblatt, et al. 2016b. Climate and developmental plasticity: interannual variability in grapevine leaf morphology. *Plant Physiology* 170(3): 1825–2015.

Clark, A.G., M.J. Hubisz, C.D. Bustamante, S.H. Williamson, and R. Nielsen. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research* 15(11): 1496–1502.

Cook, B.I., and E.M. Wolkovich. 2016. Climate change decouples drought from early wine grape harvests in France. *Nature Climate Change* 6(7): 715–719.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R. Handsaker, et al. 2011. The Variant Call Format and VCFtools. *Bioinformatics* 27(15): 2156–2158.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12(7): 499–510.

David, B., R. Bouckaert J. Felsenstein, N.A. Rosenberg, and A.R. Choudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* 29(8):1917–1932.

Dempewolf, H., G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino. 2017. Past and future use of wild relatives in crop breeding. *Crop Science* 57(3): 1070–1082.

Drummond, A.J., M.A. Suchard, D. Xie and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* 6(5): e19379.

Galet, P. 1979. *A Practical Ampelography*. Cornell University Press, Ithaca, New York, USA.

Galet, P. 1988. Cépages et vignobles de France. Tome 1. Les vignes Américaines, 2nd ed. Pierre Galet, Montpellier, France.

Glaubitz, J., T. Casstevens, F. Lu, J. Harriman, R. Elshire, Q. Sun, and E. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PloS One* 9(2): e90346.

Gross, B.L., C.M. Richards, P.A. Reeves, A.D. Henk, P.L. Forsline, A. Szewc-McFadden, G. Fazio, and C.T. Chao. 2013. Diversity Captured in the USDA-ARS National Plant Germplasm System Apple Core Collection. *Journal of the American Society for Horticultural Science* 138(5): 375–381.

Hannah, L., P.R. Roehrdanz, M. Ikegami, A.V. Shepard, M.R. Shaw, G. Tabor, G., L. Zhi, et al. 2013. Climate change, wine, and conservation. *Proceedings of the National Academy of Sciences* 110(17): 6907–6912.

He, J., X. Zhao, A. Laroche, Z.X. Lu, H. Liu, H., and Z. Li. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* 5: 1–8.

Hipp, A.L., D.A.R. Eaton, J. Cavender-Bares, E. Fitzek, R. Nipper, and P.S. Manos. 2014. A framework phylogeny of the American Oak clade based on sequenced RAD data. *PLoS One* 9(4): e93975.

Hyma, K., P. Barba, M. Wang, J. Londo, C. Acharya, and S. Mitchell, Q. Sun, et al. 2015. Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: A case study in grapevine. *PLoS One* 10(8): e0134880.

Jaillon, O., J.-M. Aury, B. Noel, N. Choisne, C. Jubin, C. Dasilva, J. Poulain, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463–467.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647–1649.

Kliman, T. 2010. *The Wild Vine: A Forgotten Grape and the Untold Story of American Wwine*. Broadway Books, New York, New York, USA.

Leaché, A.D., B.L. Banbury, J. Felsenstein, A.N.M. de Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64(6): 1032–1047.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754–1760.

Liu, X.Q., S.M. Ickert-Bond, Z.L. Nie, Z. Zhou, L.Q. Chen, and J. Wen. 2016. Phylogeny of the *Ampelocissus-Vitis* clade in Vitaceae supports the New World origin of the grape genus. *Molecular Phylogenetics and Evolution* 95: 217–228.

McAllister, C.A., and A. J. Miller. 2016. SNP discovery via genotyping-by-sequencing for assessment of population genetic structure and recurrent polyploidization in big bluestem (*Andropogon gerardii*). *American Journal of Botany*. 103:1326–1335.

McKey, D., M. Elias, B. Pujol, A. Duputié. 2010. The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* 186: 318–332.

Migicovsky, Z., and S. Myles. 2017. Exploiting wild relatives for genomics-assisted breeding of perennial crops. *Frontiers in Plant Science* 8: 460.

Migicovsky, Z., J. Sawler, K.M. Gardner, M.K. Aradhya, B.H. Prins, H.R. Schwaninger, C.D. Bustamante et al. 2017. Patterns of genomic and phenomic diversity in wine and table grapes. *Horticulture Research* 4: 17035.

Migicovsky, Z., J. Sawler, D. Money, R. Eibach, A.J. Miller, J.J. Luby, A.R. Jamieson, et al. 2016. Genomic ancestry estimation quantifies use of wild species in grape breeding. *BMC Genomics* 17(1): 478.

Miller, A.J., and B.L. Gross. 2011. From forest to field: perennial fruit crop domestication. *American Journal of Botany* 98(9): 1389–1414.

Miller, A.J., N, Matasci, H. Schwaninger, M.K. Aradhya, B. Prins, B., G.-Y. Zhong, C. Simon, et al. 2013. *Vitis* phylogenomics: hybridization intensities from a SNP array outperform genotype calls. *PloS one* 8(11): e78680.

Miller, M.A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. New Orleans, Louisiana, USA.

Moore, M.O., 1991. Classification and systematics of eastern North American *Vitis* L. (Vitaceae) north of Mexico. *SIDA Contributions to Botany* 14(3): 339–367.

Moore, M.O., and J. Wen. 2016. Vitaceae. *In* Flora of North America Editioral Committee (eds.), Flora of North America North of Mexico, *Vol. 12*. Oxford University Press, New York, New York, USA.

Mullins, M.G., A. Bouquet, and L.E. Williams. 1992. The grapevine and its wild relatives. *In* M.G. Mullins, A. Bouquet, and L.E. Williams (eds.) Biology of the grapevine. Cambridge University Press, Cambridge, United Kingdom.

Munson, T.V. 1909. *Foundations of American Grape Culture*. Orange Judd Company.

Myles, S., A.R. Boyko, C.L. Owens, P.J. Brown, F. Grassi, M.K. Aradhya, B. Prins, et al. 2011. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences* 108(9): 3530–3535.

National Research Council. 1991. US Committee on Managing Global Genetic Resources: Agricultural Imperatives. Managing Global Genetic Resources: The U.S. National Plant Germplasm System. Washington (DC): National Academies Press, USA. Available from: https://www.ncbi.nlm.nih.gov/books/NBK235638/.

Ollitrault, P., A., Garcia-Lor, J. Terol, F. Curk, F. Ollitrault, M. Talon, and L. Navarro. 2015. Comparative values of SSRs, SNPs and InDels for citrus genetic diversity analysis. *Acta Horticulturae*, 1065: 457–466.

Péros, J.P., G. Berger, A. Portemont, J.M. Boursiquot, and T. Lacombe. 2011. Genetic variation and biogeography of the disjunct *Vitis* subg. *Vitis* (Vitaceae). *Journal of Biogeography* 38(3): 471–486.

Petit, R.J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* 37(1): 187–214.

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* 5(3): 92–102.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, et al. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 81: 559–575.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: http://www.R-project.org/.

Ren, H., and J. Wen. 2007. *Vitis*. *In* Z.Y. Wu, P.H. Raven, D.Y. Hong (eds.), Flora of China, vol. 12. Missouri Botanical Garden Press, St Louis, Missouri, USA.

Rokas, A., B.L. Williams, N. King, and S.B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425(6960): 798.

Sawler, J., J.M. Stout., K.M. Gardner, D. Hudson, J. Vidmar, L. Butler, J.E. Page and S. Myles. 2015. The genetic structure of marijuana and hemp. *PloS One* 10(8): e0133292.

Smith, S.A. and C. Dunn. 2008. Phyutility: a phyloinformatics utility for trees, alignments, and molecular data. *Bioinformatics* 24: 715–716.

Soltis, P.S., D.E. Soltis, and M.W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402(6760): 402.

Stamatakis, A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312–1313.

Terral, J.F., E. Tabard, L. Bouby, S. Ivorra, T. Pastor, I. Figueiral, S. Picq, et al. 2009. Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Annals of Botany* 105(3): 443–455.

Tröndle, D., S. Schröder, H.H. Kassemeyer, C. Kiefer, M.A. Koch, and P. Nick. 2010. Molecular phylogeny of the genus *Vitis* (Vitaceae) based on plastid markers. *American Journal of Botany* 97(7): 1168–1178.

Uribe-Convers, S., M.L. Settles, and D.C. Tank. 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: Resolving the diversity within the South American species of *Bartsia* L. (Orobanchaceae). *PLoS One* 11(2): e0148203.

Wan, Y., H.R. Schwaninger, A.M. Baldo, J.A. Labate, G.-Y. Zhong, and C.J. Simon. 2013. A phylogenetic analysis of the grape genus (*Vitis* L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. *BMC Evolutionary Biology* 13: 1–20.

Warschefsky, E.J., L.L. Klein, M.H. Frank, D.H. Chitwood, J.P. Londo, J. P., E.J.B. von Wettberg, and A.J. Miller. 2016. Rootstocks: diversity, domestication, and impacts on shoot phenotypes. *Trends in Plant Science* 21(5): 418–437.

Wen, J., Z.L. Nie, A. Soejima, and Y. Meng. 2007. Phylogeny of Vitaceae based on the nuclear GAI1 gene sequences. *Canadian Journal of Botany* 85(8): 731–745.

Wickham, H. 2009. Ggplot2: elegant graphics for data analysis. Berlin: Springer Science & Business Media. https://doi.org/10.1007/978-0-387-98141-3

Zecca, G., J.R. Abbott, W.B. Sun, A. Spada, F. Sala, and F. Grassi. 2012. The timing and the mode of evolution of wild grapes (*Vitis*). *Molecular Phylogenetics and Evolution* 62(2): 736–747.

Zhang, N., J. Wen, and E.A. Zimmer. 2015. Expression patterns of AP1, FUL, FT and LEAFY orthologs in Vitaceae support the homology of tendrils and inflorescences throughout the grape family. *Journal of Systematics and Evolution* 53(5): 469–476.

Zohary, D., and P. Spiegel-Roy. 1975. Beginnings of fruit growing in the old world. *Science* 187: 319–327.

TABLE 1. Number of accessions representing each taxon and its geographic origin for phylogenomic and genetic distance analyses of *Ampelopsis* and *Vitis* taxa.

	N per Data Set			
	Full (11,020 SNPs)	***Vitis*-only (10,565 SNPs)**	**Reduced *Vitis* (8617 SNPs)**	**Geographic Distribution**
Analysis Performed	RAxML	NJ	SVD-quartets	
Ampelopsis				
A. cordata	2	0	0	SE United States
A. delavayana var. *glabra*	3	0	0	E Asia
A. glandulosa	2	0	0	S Asia
A. glandulosa var. *brevipedunculata*	6	0	0	NE Asia
Vitis* subg. *Muscadinia				
V. rotundifolia	13	9	13	SE United States
Vitis* subg. *Vitis				
V. acerifolia	19	18	5	E United States
V. aestivalis	19	18	5	E United States
V. amurensis	12	8	4	E Asia
V. arizonica	3	4	4	SW United States
V. cinerea	42	39	5	E United States
V. coignetiae	4	4	5	E Asia
V. labrusca	37	31	5	E United States
V. monticola	10	1	1	SW United States
V. mustangensis	6	4	4	SW United States
V. nesbittiana	2	4	0	S Mexico
V. palmata	9	9	5	SE United States
V. piasezkii	11	10	5	E Asia

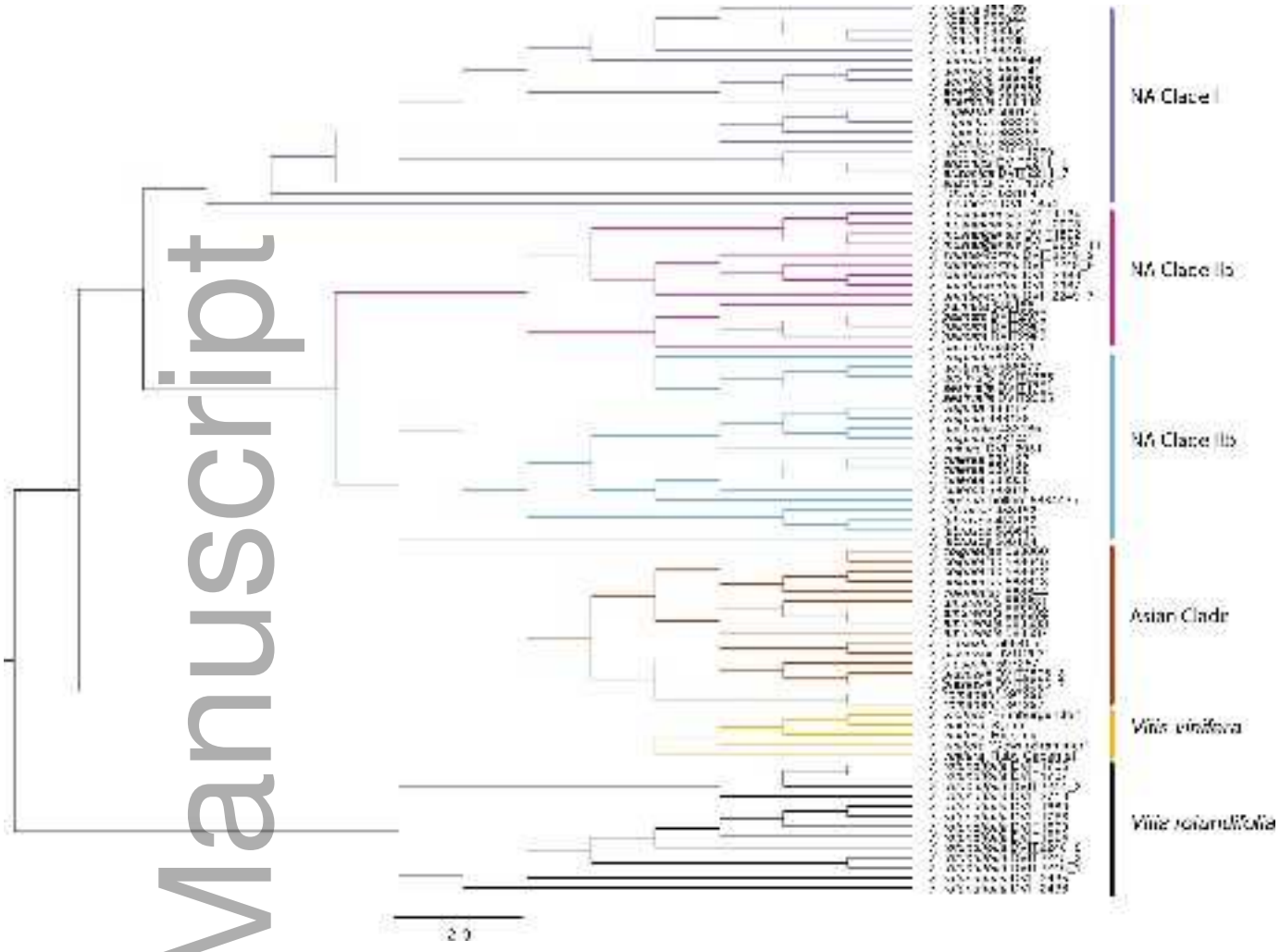
V. riparia	80	77	5	NE United States
V. romanetii	0	0	2	E Asia
V. rupestris	27	25	4	E United States
V. shuttleworthii	5	5	5	SE United States
V. vinifera	13	6	5	Europe
V. vulpina	12	10	5	E United States
V. yenshanensis	3	3	0	E Asia
unidentified *Vitis*	19	19	0	-
Total:	**359**	**304**	**87**	

FIGURE 1. Species tree generated in SVDquartets using the reduced *Vitis* data set (*n* = 87) to represent 18 *Vitis* species. *Vitis rotundifolia* represents subg. *Muscadinia*. North American and Eurasian *Vitis* species form two clades within subg. *Vitis*. Within the North American *Vitis* clade, two subclades are present: NA Clades I (*V. acerifolia/ V. arizonica/ V. monticola/ V. riparia/ V. rupestris*) and II (*V. aestivalis/ V. cinerea/ V. labrusca/ V. mustangensis/ V. palmata/ V. shuttleworthii/ V. vulpina*). NA Clade II is further divided by subclades 'a' and 'b.'

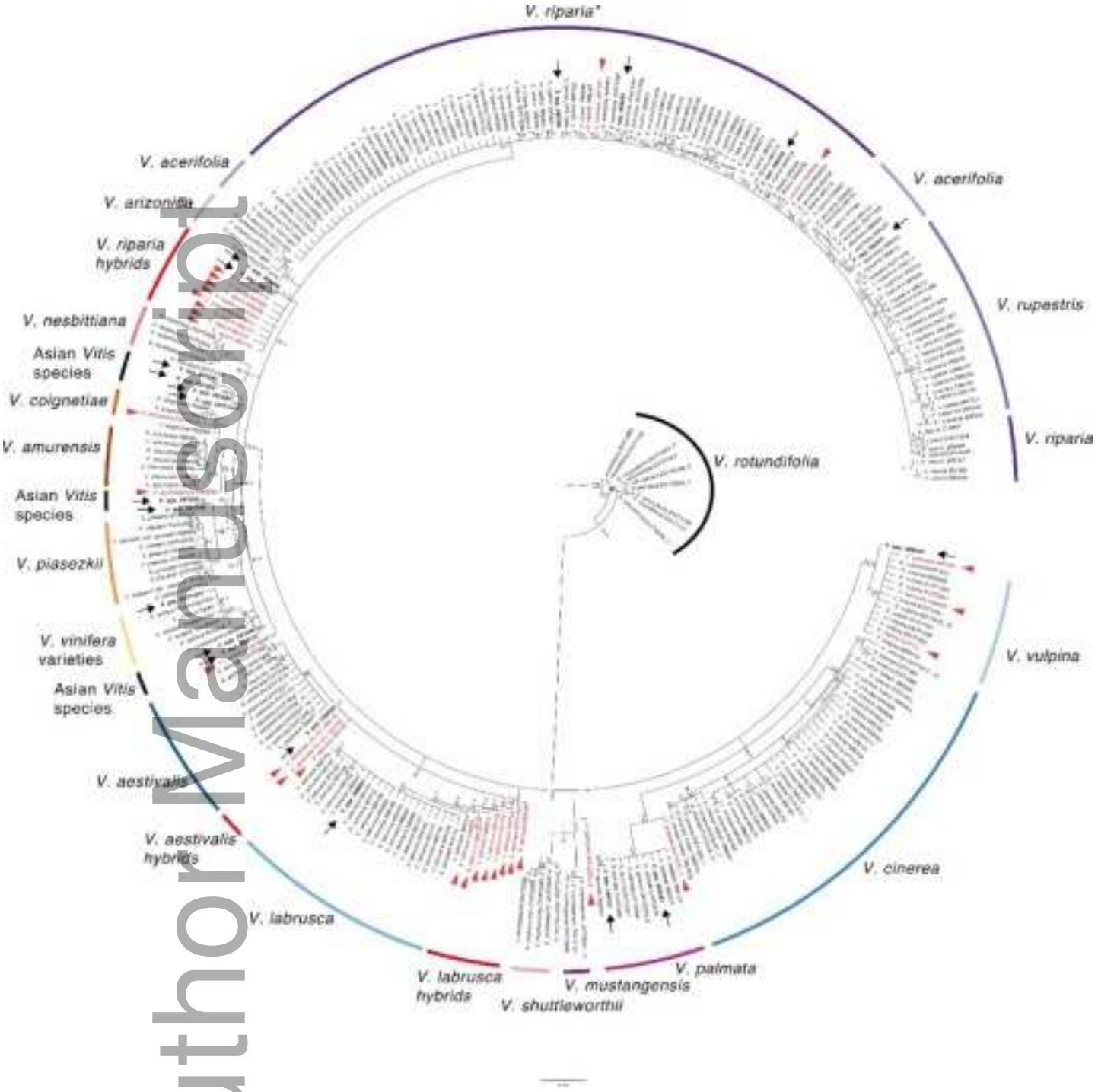
FIGURE 2. NJ tree of the *Vitis*-only data set (*n* = 304). Node values denote bootstrap support. Black arrows point to previously unidentified taxa (*Vitis* spp.) in bold text. Misidentified accessions are denoted with red arrowheads and red text. The core *V. riparia* accessions are denoted with an asterisk, though some closely related species are grouped among these accessions due to genetic similarity.

FIGURE 3. MDS plots of subg. *Vitis* accessions (*n* = 291), where species are represented by colored points, unidentified accessions are black triangles, and misidentified accessions are red triangles (see legend). Those triangles that are intermediate between species clusters likely represent hybrid individuals. (A) Dimensions one and two; (B) dimensions two and three; (C) dimensions one and three; and (D) dimensions three and four.

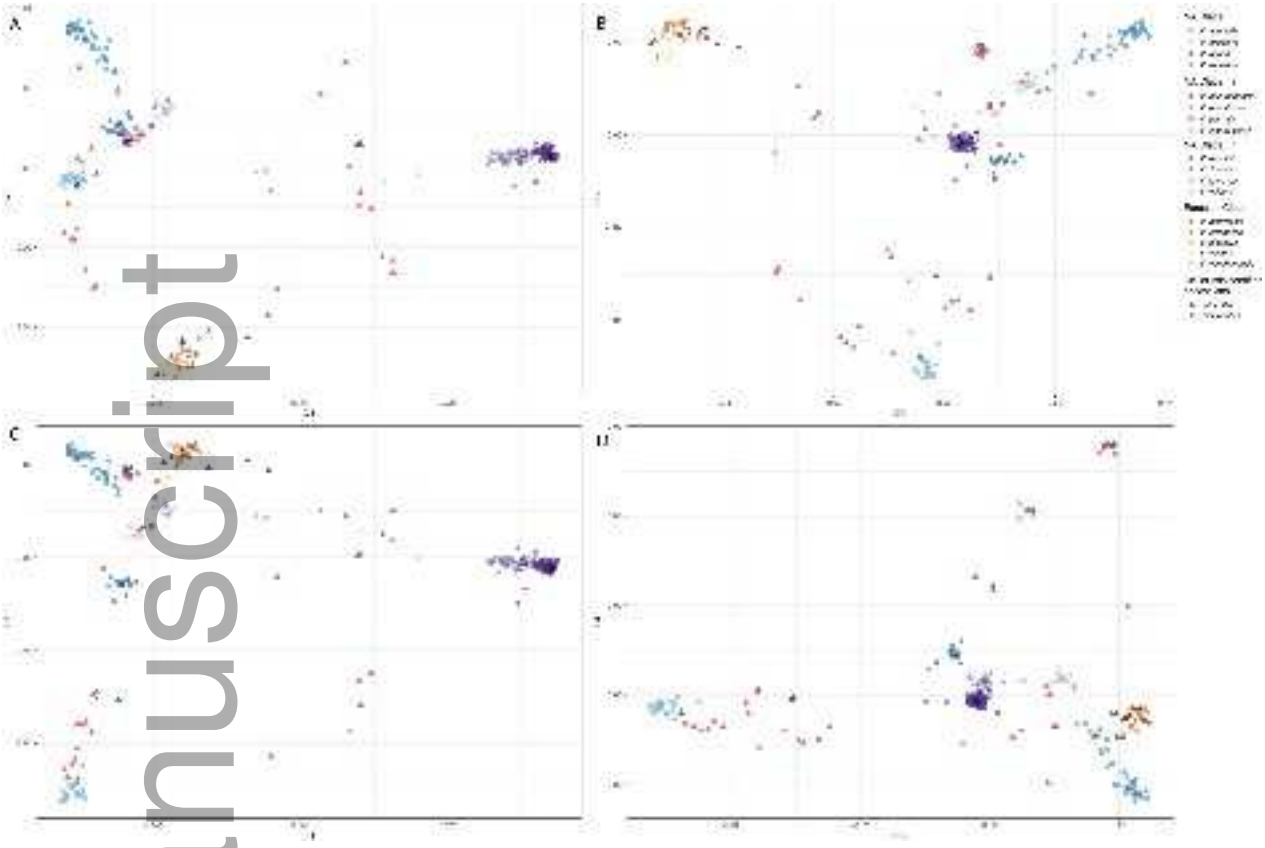
Appendix S3. Maximum likelihood 50% majority rule consensus tree generated in RAxML using the full data set (n = 359). Node values denote bootstrap support. Branch lengths are proportional to the number of substitutions per site as measured by the scale bar.



ajb2_1033_f1.tif



ajb2_1033_f2.tif



ajb2_1033_f3.tif