bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Characterizing grapevine (*Vitis* spp.) inflorescence architecture using X-ray imaging: implications for understanding cluster density

Mao Li[1], Laura L. Klein[1,2], Keith E. Duncan[1], Ni Jiang[1], Jason P. Londo[3], Allison J. Miller[1,2], Christopher N. Topp[1,*]

[1]Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132-2918, USA
[2]Department of Biology, Saint Louis University, 3507 Laclede Avenue, St. Louis, MO 63103-2010, USA
[3]United States Department of Agriculture, Agricultural Research Service: Grape Genetics Research Unit, 630 West North Street, Geneva, NY 14456-1371, USA

[*]Corresponding author:
Christopher N. Topp
Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132-2918, USA
Tel:: (314) 587-1609
Email: ctopp@danforthcenter.org

Summary

We characterized grapevine inflorescence architecture (the rachis and all branches without berries) to describe variation among 10 wild *Vitis* species, assess phylogenetic signals underlying inflorescence architecture traits, and interpret this variation in the context of breeding objectives.

Three-dimensional X-ray tomography scans of grapevine inflorescences were used to measure geometric traits and inflorescence topology using persistent homology, a mathematical approach that can comprehensively measure and compare shapes. We simulated potential space available for berry growth within a given inflorescence architecture by evaluating expanding spheres attached to pedicels, referred to as "berry potential." Lastly, we performed phylogenetic analysis and mapped trait variation.

We detected wide variation in inflorescence architecture features among *Vitis* species. Hierarchical clustering and correlation analyses revealed relationships among traits. Multivariate analyses identify traits contributing the most to variation and distinguish between species with high accuracy. Phylogenetic analyses revealed 12 morphological traits with strong phylogenetic signal.

Morphometric analysis uncovered novel differences in inflorescence architecture among clades and between *Vitis* species. Cluster density is an important trait for assessing crop quality and forecasting yield; analyses presented here can be used to tease apart subtle, heritable features and environmental influences on this major agronomic trait.

Key words: berry potential; inflorescence; morphology; persistent homology; phylogenetic analysis; topological data analysis; *Vitis* spp.; X-ray tomography

Introduction

Inflorescences are major adaptations of the angiosperm lineage whose architectural variation affects fertilization, fruit development, dispersal, and crop yield (Wyatt, 1982;

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Hake, 2008; de Ribou *et al.*, 2013; Kirchoff & Claßen-Bockhoff, 2013; Périlleux *et al.*, 2014; Chanderbali *et al.*, 2016). These branched reproductive structures with multiple flowers reflect the extraordinary diversity across angiosperm species, from an ear of corn to palms with inflorescences measuring five meters long (Hodel *et al.*, 2015). Yet seemingly simple processes give rise to these vastly different shapes - during development reproductive meristems may either switch to floral identity or proliferate additional inflorescence meristems and branches (Prusinkiewicz *et al.*, 2007). Complex topologies reflect the evolution of this functional diversity, but have proven difficult to quantify with conventional tools.

Detailed descriptions of inflorescences by trained experts are often unique to specific research communities or groups of taxa, and are not always readily transferable, hindering meaningful comparative analysis (Endress, 2010). Inflorescences are sometimes described typologically: indeterminate or determinate, simple or compound, as a raceme, cyme, panicle or spike, etc. (Wyatt, 1982; Weberling, 1992). Other approaches describe qualitative attributes of inflorescences such as the presence or absence of certain structures (Weberling, 1992; Doebley *et al.*, 1997; Feng *et al.*, 2011; Hertweck & Pires, 2014). A third method for characterizing inflorescences is through quantification of component structures (e.g., branch length, inflorescence length and width, angular traits; Kuijt, 1981; Marguerit *et al.*, 2009; Landrein *et al.*, 2012; Le *et al.*, 2018). Although these classical quantitative approaches facilitate comparative statistical analyses, the three-dimensional (3D) complexity of inflorescences is largely undescribed. Furthermore, descriptions may be confounded by developmental stage at the time of measurement, and distinguishing between vegetative and reproductive branching structures can be difficult (Wyatt, 1982; Weberling, 1992; Guédon *et al.*, 2001). Thus, new technological and analytical approaches that can represent comprehensive, multi-dimensional information about inflorescence diversity are needed to normalize and enrich analysis of these structures.

One promising approach for capturing 3D shapes of inflorescences and other plant structures is X-ray tomography (XRT). XRT generates high quality reconstructions of the internal and external shapes of plants, preserving nearly complete geometric and topological information in 3D. These 3D digital models then can be used to extract quantitative data (features) from plant structures. X-rays have been used to quantify wheat and rice seed and inflorescence traits from intact samples for non-destructive yield calculations (Hughes *et al.*, 2017; Jhala & Thaker, 2015), internal anatomy of willow trees (Brereton *et al.*, 2015), stem morphology and anatomy in sorghum (Gomez *et al.*, 2018), root structure of barley seedlings (Pfeifer *et al.*, 2015), leaf anatomy in monocots and dicots (Mathers *et al.*, 2018) and dynamic starch accumulation in living grapevine stems (Earles *et al.*, 2018), among others. Most critically, whereas manual measurements can be laborious and destructive, non-destructive sampling for XRT analysis facilitates comprehensive quantification of complex morphological traits.

Quantifying complex shapes with XRT requires appropriate analytical approaches. Topological modeling, a mathematical field concerned with the connectedness of branching structures, can quantify inflorescence architecture by parsing geometric 3D structures into

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

distinct, yet connected, components (Godin & Caraglio, 1998). Topological modeling has yielded important insights into inflorescence development, functional analysis, and crop improvement in a variety of plant species (e.g., *Arabidopsis thaliana*, *Capsicum annuum, Malus pumila*, and *Triticum*; Godin *et al.*, 1999; Letort *et al.*, 2006; Kang *et al.*, 2009). While powerful, these reductionist approaches rely on an a priori understanding of the mechanisms that contribute to complexity (e.g., branching patterns), and lose power when shapes vary drastically from one another (e.g., comparing a corn tassel to a grape cluster). Approaches that capture emergent properties of complex structures without presupposing the importance of individual structural components are complementary to traditional topological models (Bucksch *et al.*, 2017).

An emerging mathematical approach to interpret topological models is persistent homology (PH). PH extracts morphological features from two- or three-dimensional representations and can be used to compare very different shapes. PH has been applied to explain a wide range of features including atomic structures, urban and forested areas, cancers, cell shapes, and jaw shape, among others (Edelsbrunner & Morozov, 2013). In plants, PH has been used to estimate shapes that are otherwise difficult to measure including leaves, leaflet serration, spikelet shape, stomatal patterning, and root architecture (Li *et al.*, 2018a,b; Haus *et al.*, 2018; McAllister *et al.*, 2019; Migicovsky *et al.* 2018). Previous work showed that PH could capture more quantitative variation than traditional plant morphological measures (described above) resulting in the identification of otherwise latent quantitative trait loci (Li *et al.*, 2018b). PH is especially well-suited for quantifying branching topology as it can quantitatively summarize complex variation with a single measure (Li *et al.*, 2017; Delory *et al.*, 2018). Rachis, pedicel, and branches include inherently topological features that can be especially well-analyzed with PH-based methods.

Grape clusters (or bunches) are branched structures supporting berries produced by grapevines (*Vitis* spp.) and are an ideal system in which to apply XRT and PH. Grape infructescences are historically, culturally, and economically important and vary extensively in nature and in cultivation (Iland *et al.*, 2011). Cluster architecture determines bunch density, and is defined as "arrangement of berries in a cluster and the distribution of free space" (Richter *et al.*, 2018). The density of berries in a cluster is an important breeding feature because it determines yield, wine character, and disease resistance (amount of air flow between berries is a primary determinant of pests and pathogens on the fruit). Cluster density is a characteristic identified by the Organization Internationale de la Vigne et du Vin, and varies from "berries clearly separated" (loose clusters) to "berries deformed by compression" (very dense clusters; OIV, 2001). As one of the primary determinants of yield, end-product characteristics, and disease resistance cluster architecture has been studied extensively in grapevine (reviewed in Tello & Ibáñez, 2018). These studies have shown that wine grape cultivars (*Vitis vinifera*) display distinct bunch densities (Shavrukov *et al.*, 2004). However, less is known about cluster architecture in wild *Vitis* species, an important source of natural variation used by breeders in the development of hybrid grapevine varieties.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Historically, researchers have focused on a suite of cluster traits such as cluster size, shape, weight, and density/compactness to characterize bunch density quantified in grapevines (Rovasenda, 1881; Pulliat, 1888; Bioletti, 1938; Galet, 1979; Bettiga, 2003). Measurements are made primarily using traditional tools including rulers, digital calipers, volume displacement, and/or through human judging panels. More recently, automated image-based approaches have been implemented to capture aspects of cluster architecture in the lab and field (Ivorra *et al.*, 2015; Aquino *et al.*, 2017, 2018; Rist *et al.*, 2018). However, these image-based methods cannot penetrate the internal inflorescence structure. Therefore resulting models are based only the visible surface and the underlying topology cannot be fully captured, limiting an understanding of how inflorescence architecture and berry features co-vary. XRT and PH applications offer an important opportunity to understand grapevine bunch density through detailed analyses of inflorescence architecture. This work will deepen our understanding of natural variation of inflorescence structure, identify priority targets for breeding, and permit connecting 3D structure to underlying processes and genetics of inflorescence development.

We use X-ray tomography, geometric measurements, persistent homology, and computational simulation to characterize wild grapevine inflorescence architecture. We target the branching architecture of the mature inflorescence: the rachis and all branches that remain following the removal of ripe berries (Fig. 1). Specifically, we aim to: 1) characterize variation in component traits of inflorescence architecture within and among *Vitis* species; 2) assess phylogenetic signals underlying inflorescence architecture traits; and 3) interpret inflorescence trait variation in the context of breeding objectives. This work represents an important advance for the characterization of 3D plant architecture using a powerful combined imaging and computational approach.

Materials and Methods

Plant Material

In this study, we sampled grapevine bunches from 136 unique genotypes representing 10 wild *Vitis* species living in the USDA germplasm repository system (Geneva, NY; Table 1, Fig. S1). Grapevines have a paniculate inflorescence that consists of a rachis with several primary and secondary branches, tapering towards the terminus of the organ (Iland *et al.*, 2011). Wild grapevines are dioecious; consequently, unbalanced sample sizes for different species reflect numbers of female genotypes available in the germplasm collection. Each unique genotype is represented in the germplasm collection by two clonally replicated vines. For most of the 136 genotypes, we collected a total of three clusters from the two clonal replicates combined, representing average cluster morphology. We avoided clusters that were visibly damaged or indirectly altered (e.g., tendril or trellis interference). For each vine, clusters were removed from separate canes at the point of peduncle attachment (Fig. 1a). In total, 392 clusters were collected in September 2016 when berries were soft, equivalent to EL38 developmental stage (Coombe, 1995; Fig. 1b). Berries were manually removed from clusters in the field, and the remaining inflorescence stalks (including rachis,

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

branches, and pedicels; hereafter referred to as inflorescence or inflorescence architecture) were used to assess inflorescence architecture.

X-ray tomography and data preprocessing

Grapevine inflorescences were scanned at the Donald Danforth Plant Science Center (St. Louis, MO) using a North Star Imaging X5000 X-ray tomography instrument (NSI; Rogers, MN) equipped with a 16-bit Varian flat panel detector (1536 x 1920 pixels with 127um pixel pitch) and 225kV microfocus reflection target X-ray source. Each inflorescence was held between two pieces of construction-grade expanded polystyrene, clamped in a panavise, and positioned on the X-ray turntable in one of two configurations (Fig. 1c): 725mm from the source, generating 1.26x magnification and 101um voxel resolution, or 766mm from the source, generating 1.19x magnification and 107um voxel resolution. Each scan used X-ray wattage set to 60kV and 1200uA at 10 frames per second, collecting 1200 16-bit TIFF projections over 360 degrees of rotation during a 2min continuous standard scan. Projections for each scan (Fig. 1d) were combined into a single 3D volume using NSI efX-CT software, converted to a density-based surface rendering Polygon file (PLY), and exported for analysis (Fig. 1e). The full PLY data set for this work is 7.85GB, and can be downloaded from: https://www.danforthcenter.org/scientists-research/principal-investigators/chris-topp/resources.

We exported the surface mesh data (.ply files) into Meshlab (v1.3.3, (Cignoni *et al.*, 2008) and performed the following processing steps to remove topological noise: 1) deleted the vertices where branches touch using "Select Vertexes" and "Delete Selected vertices" filters; 2) removed duplicates and isolated vertices and faces using the filters "Remove Duplicated Vertex," "Remove Duplicate Faces," "Remove Isolated pieces (wrt Diameter)," and "Remove Unreferenced Vertex."

Geometric inflorescence architecture traits

We extracted 15 geometric traits from scanned inflorescences (Fig. 2, Fig. S2). Detailed trait descriptions and calculations are explained in Table S1. Trait illustrations, including examples of low and high values for each trait, are available in Fig. 2 and Fig. S2. Traits were organized in one of three trait groups: global-size features, local-branching features, and size-invariant features (Table 2). PedicelDiameter and PedicelBranchAngle were measured using the software DynamicRoots (Symonova et al. 2015) on a subset of detected pedicels from the raw 3D volume data. All other traits were derived from Matlab algorithms. Branch length traits (i.e., TotalBranchLength, RachisLength, PedicelLength, and AvgBranchLength) were derived from the persistence barcode (see next subsection).

Quantifying branching topology using persistent homology, a topological data analysis method

Persistent homology measures shapes based on a tailored mathematical function, such as geodesic distance, which we used here to capture both curved length and topology of the branches (Fig. 3, Video S1). The geodesic distance of a point is the length of the shortest

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

curve connecting the point and the base (e.g. purple curves, Fig. 3a), where the tailored base can be set as the first node or ground level (the brown line in Fig. 3a). For each branch, the tip always has the largest geodesic distance from the base (Fig. 3b). A level represents the collection of points whose geodesic distances are the same (e.g. geodesic distance=90, pink curve in Fig. 3a). A superlevel set, for example, at 90, is all the points whose geodesic distances are greater than 90 (black branch tips, Fig. 3a). Changing the level value from largest to smallest (x axis, Fig. 3c), the sequence of nesting superlevel sets can be formed, which is named superlevel set filtration (top panel, Fig. 3c). During the change of the level value, bars record the connected components for each of the superlevel sets. When a new component arises, a new bar starts (e.g. at level 112, purple branch, Fig. 3c). When two components merge (e.g. at level 65, orange branch merges into purple branch, Fig. 3c), the shorter bar stops (e.g. the orange bar stops at level 65, Fig. 3c). This bar graph, called the persistence barcode, summarizes topological information such as branching hierarchy, branch arrangement, and branch lengths. In our study, we set the base as the junction between peduncle and rachis (the lowermost node, indicated by a brown line in Fig. 1e, Fig. 3d, f) and use this base to compute the persistence barcode for the inflorescence architecture (Fig. 3e, g).

The persistence barcode can be used to compare topological similarity between any two inflorescences. To compute pairwise distance among persistence barcodes for the entire inflorescence population, we used the bottleneck distance (Cohen-Steiner *et al.*, 2007). Bottleneck distance is a robust metric that calculates the minimal cost to move bars from one persistence barcode to resemble another (Li *et al.*, 2017). We performed multidimensional scaling (MDS) on the pairwise bottleneck distance matrix and projected the data into lower dimensional Euclidean space by preserving the pairwise distance as well as possible. The Matlab (R2017a) MDS function cmdscale() projects the data so that MD1 acts as PC1 representing the most variation. The first three PCs (MDs) explained about 80% of the total variation and were included as traits: PersistentHomology_PC1 (PH_PC1, explained about 54% variation), PersistentHomology_PC2 (PH_PC2, explained about 20% variation), and PersistentHomology_PC3 (PH_PC3, explained about 6% variation). Those traits not only measure the topological structure, but also relate to geometric variation (e.g. global size) as the data were not normalized (Fig. 2, Table S1).

Next, we normalized the persistence barcode by the TotalBranchLength (summation of the bar lengths) so that the TotalBranchLength was 1. By a similar procedure, we derived the first three PCs named PersistentHomologyNormalizedByTotalBranchLength_PC1 (PHn_PC1, explained about 45% variation), PersistentHomologyNormalizedByTotalBranchLength_PC2, (PHn_PC2, explained about 21% variation), and PersistentHomologyNormalizedByTotalBranchLength_PC3 (PHn_PC3, explained about 7% variation) for the normalized inflorescence topological structure (Fig. 2, Table S1).

Berry potential, an approach to indirectly explore the space limited by inflorescence architecture

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

An ongoing question in grapevine cluster architecture is the relationship between inflorescence architecture and berry number and size. Inflorescence architecture is one of several factors determining the number of berries that can form, due to the number of pedicels and the available space for berry development. In this study, berries were removed because of concerns about berry integrity during transport from New York to Missouri, and the time between harvest and scanning. Instead of looking directly at berries on the cluster, we used inflorescence architecture as a starting point to simulate potential space available for berry growth by evaluating expanding spheres attached to pedicels. The extent of sphere expansion allowed by each pedicel is referred to as "berry potential" (Fig. 4, Video S2).

We first determined the growth direction for each berry potential based on the pedicel orientation. When spheres expand, the center moves along the pedicel direction (Fig. 4a). This step can be achieved by performing principal component analysis (PCA) on the near-berry segment of the pedicel. The first principal axis is the pedicel direction. We adjusted the arrow of the direction to make sure berry potential increases outward along the pedicel orientation. Then the berry potential increases until one of three situations is encountered (Fig. 4b): 1) if two berry potentials touch to each other, both berry potentials will stop increasing; 2) if a berry potential touches any part of the inflorescence, it will stop increasing; 3) if the diameter of the berry potential reaches the maximum size known for that species (Table 1), it will stop increasing. For each species, the maximum size is defined as the maximum berry diameter, a number estimated from known ranges of berry sizes for each species, based on values obtained from (Galet, 1988; Moore & Wen, 2016).

Berry potential does not reflect true berry growth; rather, berry potential is a derived attribute of inflorescence architecture, an indirect estimate of the space potentially available for berry growth. It also does not account for the possibility of branches bending or otherwise becoming re-oriented due to pressure from growing berries. Berry potential is based on the number of neighbor pedicels, neighbor pedicel lengths, and neighbor pedicel mutual angles. Larger values for berry potential are associated with fewer neighbor pedicels, and/or longer pedicel lengths, and/or larger mutual angles. From the berry potential simulation, we calculated three features, TotalBerryPotentialVolume, AvgBerryPotentialDiameter, and BerryPotentialTouchingDensity, which is the berry potential touching number (i.e., touching either another berry potential or any part of the inflorescence) divided by the number of berry potential (Fig. 2, Table S1).

Phylogenetic analysis

Phylogenetic analyses were conducted to understand evolutionary trends in inflorescence architecture in *Vitis*. Single nucleotide polymorphism (SNP) markers were generated as part of a separate study of the USDA Grapevine Germplasm Reserve in Geneva, NY (Klein *et al.*, 2018). The original dataset consisted of 304 individuals representing 19 species that were sequenced using genotyping-by-sequencing (GBS; Elshire *et al.*, 2011). Briefly, Klein *et al.* (2018) filtered data to retain biallelic sites with a minimum allele frequency of 0.01, a minimum mean depth of coverage of 10x, and only sites with <20% missing data and individuals with <20% missing data. SNP data for 99 individuals from this study that were

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

also genotyped in (Klein *et al.*, 2018); Table 1) were extracted using custom scripts. We performed phylogenetic analysis on the sequence data extracted for 99 individuals using SVDquartets (Chifman & Kubatko, 2014), a maximum likelihood approach designed to address ascertainment bias associated with reduced representation sequencing techniques like GBS. We analyzed all possible quartets and carried out 100 bootstrap support runs (Fig. S1) using PAUP* version 4.0a (Swofford, 2003). The three main clades recovered in the tree were consistent with previous phylogenetic work in *Vitis*: 1) an Asian Clade (*V. amurensis and V. coignetiae*), 2) North American Clade I (*V. riparia, V. acerifolia, and V. rupestris*), and 3) North American Clade II (*V. vulpina, V. cinerea, V. aestivalis, V. labrusca, and V. palmata*) (Tröndle *et al.*, 2010; Zecca *et al.*, 2012; Miller *et al.*, 2013; Zhang *et al.*, 2015; Klein *et al.*, 2018).

To visualize trait distributions on a phylogenetic tree using branch lengths, we used Mega X (Kumar *et al.*, 2018) to generate a neighbor joining tree with 2000 bootstrap replicates. All measurements were averaged across the three replicates per genotype to produce an average value for each trait for each genotype. We computed Pagel's lambda to estimate phylogenetic signal for each morphological trait and mapped each trait onto the phylogeny (Fig. S3a-x) using the R package phytools (v. 0.6-44; Revell, 2012). We calculated variation of each morphological trait for each clade based on the mean value for each species (Fig. S4).

Statistical analysis

PCA, MDS, and hierarchical cluster analysis generating a hierarchical tree were performed in Matlab using functions pca(), cmdscale(), and clustergram(). The R function cor.mtest() and package corrplot (Wei & Simko, 2017) were used for significance tests and correlation matrix visualization. The function lda() in R package MASS (Venables & Ripley, 2002) was used for the linear discriminant analysis (LDA) with a jackknifed 'leave one out' cross validation method.

Code availability

All Matlab functions used to calculate persistence barcodes, bottleneck distances, simulation for berry potential, other geometric features used in this study, and the script for extracting phylogenetic information can be found at the following GitHub repository: https://github.com/Topp-Roots-Lab/Grapevine-inflorescence-architecture.

Results

Inflorescence morphological variation and trait correlation within *Vitis* species

We investigated 24 morphological traits (15 geometric traits, six PH traits, and three berry potential traits) of inflorescence architecture in 10 wild *Vitis* species (136 genotypes, 392 samples) and detected wide variation in morphological features within and between species (Fig. 2, Fig. S2, and Table S2). In particular, of all the species examined, *V. aestivalis*

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

has the largest variance for TotalBerryPotentialVolume. *V. labrusca* has the largest variance for ten traits (i.e., pedicel features, Sphericity, AvgBranchDiameter, AvgBerryPotentialDiameter, and normalized topological traits). *V. cinerea* has the largest variance for six traits (i.e., most global-size features, PH_PC2, and PH_PC3). In comparison, *V. palmata* has smallest variance for eight traits (i.e. pedicel features, Sphericity, AvgBranchDiameter, TotalBerryPotentialVolume, PH_PC3, and PHn_PC3), as does *V. amurensis* (global-size features, RachisLength, PH_PC1, and PH_PC2).

All traits were hierarchically clustered based on the mean trait values for each species, classifying traits into two main categories: mostly size-invariant + local-branching features (PHn_PC3 to PedicelLength), versus global-size features (AvgBranchLength to BerryPotentialTouchingDensity) (Fig. 5a). Hierarchical clustering (Fig. 5a) and pairwise correlation for morphological traits (Fig. 5b) show that global-size features (ConvexHullVolume, SurfaceArea, Volume, NumberOfPedicel, and TotalBranchLength), PH_PC1, and RachisLength are all highly positively correlated. We refer to these seven traits as size-associated features. Size-associated features are negatively correlated with PedicelLength/RachisLength, Solidity, Sphericity, and PHn_PC1. Some traits are relatively independent such as 2nd/LongestBranchLength, PedicelLength, PedicelBranchAngle, PH_PC2, PHn_PC2, and PHn_PC3 (Fig. 5b). PH_PC3 has some negative relation with size-invariant features. PHn_PC1 positively correlates with Sphericity, Solidity, and AvgeBerryPotentialDiameter (Fig. 5b). Pairwise correlations of morphological features (allometric relationships) for each of the species vary widely (Fig. 5c; for all traits see Fig. S5a-x). For example, more pedicels typically result in smaller berry potential diameters, except for *V. aestivalis*. Longer branches tend to be thinner, except for *V. coignetiae,* and correlate with larger inflorescences, except in *V. acerifolia*.

Hierarchical clustering of 10 *Vitis* species based on the 24 morphological traits resolved four groups: 1) *V. cinerea*, 2) *V. aestivalis,* 3) *V. coignetiae/ V. vulpina/ V. palmata/ V. acerifolia/ V. riparia/ V. rupestris,* and 4) *V. amurensis/ V. labrusca* (Fig. 5a). Among the 10 *Vitis* species examined in this study, the largest variance in mean trait values are seen in *V. cinerea* (Fig. 5a). *V. cinerea* samples are generally larger than those from the other species, as reflected in size-associated traits. Topology traits such as PHn_PC3 and size-invariant traits like Sphericity and Solidity are lower in the mean trait value for *V. cinerea* than for other species. Similarly, mean trait values are larger for size-associated traits in *V. aestivalis* (Fig. 5a). Compared to other species, topology and berry potential traits are larger in *V. aestivalis*. Mean trait values of the third group (*V. coignetiae/ V. vulpina/ V. palmata/ V. acerifolia/ V. riparia/ V. rupestris*, Fig. 5a) tend to be nearer to middle values compared to the other species. Within this group, *V. acerifolia/ V. riparia/ V. rupestris* typically are larger in the mean trait value for berry potential touching (i.e., denser berry potentials). These three species and *V. palmata* tend to have large, first primary branches (i.e., wings; Fig. 1e). *V. coignetiae* has thicker branches and *V. vulpina* has longer pedicels compared to other species in this group. The final group, *V. amurensis* and *V. labrusca,* have relatively smaller inflorescences with thicker branches compared to the other species sampled here. These general features are reflected in larger mean values for several size-invariant and local-branching features and smaller mean values for many branch length dependent and size-associated features, respectively (Fig. 5a).

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Multivariate, discriminant analysis of *Vitis* species based on inflorescence architecture

In order to understand how overall inflorescence architecture varies among *Vitis* species, we performed PCA using all 24 morphological features and all samples. PC1 explained 37.12% of the total variation in the measured architecture (Fig. 6a). The traits with the largest values for PC1 loadings, indicating that they contributed most to variation, are size-associated features, Solidity and Sphericity. PC2 explained 15.4% of the total variation in the measured inflorescence architecture, with variation primarily explained by local-branching features such as PedicalDiameter, PedicelLength, PedicelLength/RachisLength, AvgBranchLength, BranchDiameter, three berry potential traits, and PHn_PC1 (Fig. 6a). Although inflorescences from each species occupy different regions of morphospace, these regions overlap considerably.

LDA performed on the first 18 PCs, explaining 99.5% of the variation, distinguished between species with a classification accuracy rate of 78.32%. A confusion matrix (Fig. 6b) shows the proportion of samples correctly predicted for each species. LD1 primarily separates *V. cinerea*, *V. labrusca*, and *V. amurensis* from the other species while LD2 primarily separates *V. vulpina* and *V. coignetiae*. The traits that are most important for distinguishing these species, as indicated by LD loadings, are TotalBerryPotentialVolume and PHn_PC1 for LD1, and AvgBranchLength and AvgBerryPotentialDiameter for LD2 (Fig. 6b). The most important predictors for correctly separating any two species are shown as the grey scaled boxes in Fig. S6, and Table S3. For example, BranchDiameter and PedicelDiameter are key when contrasting *V. coignetiae* and *V. vulpina*, suggesting that different branch thickness easily distinguishes these two species. This method correctly determined species classifications with 100% accuracy when contrasting *V. aestivalis* and *V.cinerea, V. aestivalis* and *V.palmata*, *V. aestivalis* and *V. vulpina, V. amurensis* and *V. cinerea*, *V. amurensis* and *V. palmata*, *V. cinerea* and *V. coignetiae*. Other combinations of species are harder to distinguish on the basis of inflorescence characters. For example, the classification accuracy rate was only 80% when distinguishing between *V. amurensis* and *V. labrusca* and 82% for *V. aestivalis* and *V. coignetiae*.

Phylogenetic signal of inflorescence architecture within clades

The phylogeny dataset (N=99) is generally well-supported at the species level and correlates well with current taxonomy. Using average trait values per individual, Pagel's lambda shows 12 morphological traits (seven size-associated features along with PedicelDiameter, TotalBerryPotentialVolume, Sphericity, PH_PC2, PHn_PC1) have strong phylogenetic signal (lambda>0.8, Fig. 7, Table S4). While most species sampled tend to have small values for the seven size-associated features, *V. aestivalis, V. cinerea,* and *V. vulpina* tend to have values that are either close to median, or larger. On average, *V. labrusca* has larger values for Sphericity and PHn_PC1 compared to other species sampled, while *V. cinerea* generally has some of the smallest values for these traits. Only two morphological traits (2nd/LongBranchLength, lambda=0.06 and BerryPotentialTouchingDensity, lambda=0.25) lack phylogenetic signal (Fig.7, Table S4).

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We observe differences in *Vitis* inflorescence architecture among clades and between species. For North American (NA) clade I (*V. acerifolia, V. riparia, V. rupestris*), variation in the 24 morphological traits measured have similarly small values among species, particularly for several size-associated traits, although there is relatively large variation for PH_PC3 and BerryPotentialTouchingDensity (Fig. 7). Within NA Clade I, we observe differences among clade members for traits such as Sphericity and PHn_PC1 (larger in *V. rupestris* compared to other clade members) and PedicelDiameter and BranchDiameter (slightly larger in *V. acerifolia* compared to other clade members; Fig. 7). NA Clade II appears to be more variable among clade members. *V. cinerea* has larger values for size-associated traits compared to clade members *V. labrusca*, *V. palmata,* and *V. vulpina*. Meanwhile, *V. labrusca* typically has larger values for local features (e.g., Sphericity, PedicelDiameter, AvgBerryPotentialDiameter, PedicelBranchAngle) compared to the other clade members (Fig. 7).

We calculated the mean value for each species of each morphological trait to study variation within the three clades and detect subtle signatures (Fig. 7). We computed the variance for the multivariate trait (combining all the 24 traits), and each of these 24 traits for each clade (Fig. S4, Table S5). Overall, based on the samples used in this analysis, variance of the multivariate trait for the NA Clade I (variation=0.14) is much smaller than the NA Clade II (variation=0.64), while the variation for Asian Clade is 0.39. Some traits have almost no variance in Asian Clade such as PedicelDiameter, PHn_PC2, PH_PC3, and 2nd/LongestBranchLength. However, North American species (8/~19 taxa) in this study are better represented than Asian species (2/~37 taxa), so we are cautious not to overinterpret this finding. Traits with the greatest variance in the Asian Clade included PedicelLength/RachisLength, RachisLength, and PH_PC1, while NA Clade I has greatest variance in PHn_PC2. All the other traits have greatest variance in the NA Clade II (Fig. S4, Table S5). Traits with the smallest variance in the Asian Clade included PHn_PC3, PHn_PC1, PedicelDiameter, BranchDiameter, NumberOfPedicel, 2nd/LongestBranchLength, PH_PC3, and BerryPotentialTouchingDensity. The other traits had small variance in NA Clades I (Fig. S4, Table S5). Our results highlight clade-specific variation in inflorescence architecture for previously undescribed traits.

Discussion

Inflorescence architecture provides the scaffold on which flowers and fruits develop, and consequently is a primary trait under investigation in many crop systems. Studies extend into interspecific variation, pollen dispersal, genetic architecture, evolution, regulation, and development of inflorescence structures (e.g., Bradley *et al.*, 1996; Friedman & Harder, 2004; Kellogg, 2007; Morris *et al.*, 2013; Han *et al.*, 2014; Hodge & Kellogg, 2015; Whipple, 2017; Stitzer & Ross-Ibarra, 2018; Ta *et al.*, 2018; Richter *et al.*, 2018). Yet the challenge remains to analyze these complex 3D branching structures with appropriate tools. High resolution data sets are required to represent the actual structure and comprehensive analysis of both the geometric and topological features relevant to phenotypic variation and to clarify evolutionary and developmental inflorescence patterns.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Our results demonstrate the power and potential of X-ray imaging and advanced morphometric analysis for investigating complex 3D phenotypic features. We analyzed the phenotypic variation in inflorescence architecture of 10 wild *Vitis* species using computer vision and an emerging biological shape analysis method, persistent homology, which allowed comprehensive comparisons of shape. Although samples analyzed here represent only a subset of the known variation in *Vitis*, which includes an estimated 60 species, our analyses demonstrate significant variation within and among *Vitis* species and among clades. Correlation analysis (Fig. 5b) showed that PH is a complementary feature, as it is relatively independent from most geometric features. We were able to assign widely differing architectures to biological species with high accuracy (Fig. 6) from the 24 different morphometric traits surveyed in this study. PH provides an important contribution to this discriminatory power, as does berry potential (Fig. 6b). We observed that traits such as the rachis length, the sum of all branches, the space encompassing the inflorescence architecture (ConvexHullVolume), and PH can be indicative of species and clade (Fig. 7). Our results suggest meaningful, comprehensive information about the inflorescence structure was captured with a single measure (i.e., the persistence barcode) and that PH is a valuable method for quantifying and summarizing topological information.

Persistent homology analysis has led to a deeper understanding of trait genetic variation and architecture in plants. Li *et al.* (2018a) used PH to analyze two-dimensional (2D) leaf shape and predicted family identity with accuracy greater than expected by chance in over 140 plant families, outperforming other widely-used methods of digital shape analysis. Li *et al.* (2018b) showed that PH-based, topological data analysis distinguished between genotypes and identified many new quantitative trait loci (QTL) with 2D tomato leaf shape and root architecture data. This work sets a precedent for measuring observable, yet previously undescribed, phenotypes. In grapevine, QTL analysis indicates a genetic basis to inflorescence architecture and berry compactness (Correa *et al.*, 2014; Richter *et al.*, 2018). Deploying PH-based, topological modeling to grapevine mapping populations could lead to the rapid identification of additional inflorescence trait QTL for breeding. For example, we observed total branch length (a proxy for bigger or smaller clusters) correlates with number of pedicels (a proxy for berry number; Fig. 5), an informative relationship to assess potential yield. However, selecting for total branch length might lead to a negative correlation with the average berry potential diameter (i.e., smaller berries). Although this correlation may be desirable for wine grapes, it is not for table grapes.

Grapevine cluster architecture is a composite feature that reflects multiple subtraits including stalk traits (inflorescence architecture) and berry features (Richter *et al.*, 2018). OIV 204 uses “bunch: density” to describe variation in clusters, ranging from (1) berries clearly separated with many visible pedicels to (9) berries deformed by compression (OIV, 2001; Rombough, 2002). Other authors have deconstructed traits contributing to cluster architecture primarily through individual measurements taken by hand (e.g., Shavrukov *et al.*, 2004; Tello *et al.*, 2015; Zdunić *et al.*, 2015; Tello & Ibáñez, 2018) and more recently, with image-based technologies (Cubero *et al.*, 2014; Roscher *et al.*, 2014; Ivorra *et al.*, 2015; Aquino *et al.*, 2017, 2018; Rist *et al.*, 2018). Here, we are able to describe traits of interest that contribute greatly to the morphological features captured by the OIV scale (e.g.,

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

548 NumberOfPedicel, PedicelLength, PedicelBranchAngle, RachisLength, overall shape using
549 PH; Fig 2, Fig. S2). This method could facilitate precision breeding for both whole
550 inflorescence structure topology and specific desirable geometric traits.
551
552 While several studies have quantified cluster structure in cultivated grapevines, similar
553 studies of wild *Vitis* inflorescence architecture are lacking. Munson (1909) and Galet
554 (1979) describe North American *Vitis* cluster structure qualitatively, commenting on
555 compactness, size, shape, and the presence of large first primary branches
556 (wings/shoulders). Taxonomic descriptions typically do not examine inflorescence
557 architecture beyond categorical type, position on the vine, and the average number of
558 berries per cluster (Comeaux *et al.*, 1987; Moore, 1991; Moore & Wen, 2016). Descriptions
559 of the position of the inflorescence are useful for identification and are included in
560 dichotomous keys; however, to our knowledge, other inflorescence architecture traits have
561 not been rigorously quantified among wild *Vitis* species. Although qualitative descriptions
562 are valuable and accessible, powerful phenotyping tools are required to associate complex
563 phenotypes with evolutionary and developmental patterns.
564
565 Using 3D imaging and PH with a topological modeling approach, we identified attributes of
566 inflorescence architecture that vary within and among *Vitis* species that, to our knowledge,
567 have not been previously described. Differences in inflorescence architecture among clades
568 mirror other phenotypic differences among members of North American *Vitis*. For example,
569 members of NA Clade I (*V. acerifolia*, *V. riparia*, and *V. rupestris*) have small values for size-
570 associated features (e.g., RachisLength, ConvexHullVolume, NumberOfPedicel,
571 TotalBranchLength, SurfaceArea, Volume) and relatively large values for PH_PC3 and
572 BerryPotentialTouchingDensity (Fig. 7). These species share suites of other morphological
573 characters (nodal diaphragm, branch, and leaf surface traits, and large stipules; Moore
574 1991, Moore and Wen 2016, Klein *et al.*, 2018). It is possible that among closely related
575 species conserved pathways generate vegetative and reproductive similarities.
576
577 Sample size is low for the Asian Clade and most of NA Clade II, limiting our ability to assess
578 variation in these species; however, members of NA Clade II do not have suites of shared
579 inflorescence traits (*V. aestivalis, V. cinerea, V. labrusca, V. vulpina*; Klein *et al.*, 2018).
580 Rather, *V. labrusca* has very small values for size-associated traits and larger values for
581 local features compared to the other clade members, whereas *V. cinerea* has larger values
582 for size-associated features and smaller values for local features (Fig. 7). This is consistent
583 with the observation that aside from core phenotypic synapomorphies in the genus
584 (tendril, bark, lenticel, and nodal diaphragm characters), members of NA Clade IIb (*V.*
585 *aestivalis, V. cinerea, V. labrusca,* and *V. vulpina*) do not share morphological traits unique to
586 the clade (Klein *et al.*, 2018). These species mostly co-occur across their distributions
587 (Callen *et al.*, 2016) and additional sampling of *Vitis* taxa is necessary to further explore
588 these complex evolutionary patterns. We observe *V. amurensis* grouping with *V. labrusca*
589 and *V. coignetiae* grouping with North American species in hierarchical cluster analysis
590 (Fig. 5a). The former two species have relatively smaller inflorescence architectures with
591 thicker branches compared to the other species sampled here. Taxonomic relationships
592 among North American and Asian *Vitis* species have been historically challenging, with
593 clades comprised of species with disjunct distributions (Mullins *et al.*, 1992). Since current

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

taxonomy resolves separate Asian and North American clades (Klein *et al.*, 2018), morphological similarity between these species likely reflects convergent evolution.

Future Directions

Three-dimensional imaging through XRT and advanced mathematical approaches like persistent homology provide new ways to visualize and interpret complex biological structures including inflorescences, and to understand the genetic and environmental factors underlying variation in their architecture. In grapevines, cluster density is an important trait that is used to assess grapevine crop quality and to forecast yield, in part because of the association between bunch density and fungal infestations such as *Botrytis* (Hed *et al.*, 2009; Iland *et al.*, 2011; Molitor & Beyer, 2014; Molitor *et al.*, 2018). This study expands on previous work identifying variation in inflorescence architecture among cultivars (Shavrukov *et al.*, 2004), finding notable differences in cluster architecture among species. A logical next step may be to use 3D images and PH with topological modeling to trace the development of inflorescences across multiple growing seasons in a mapping population. Methods presented here are also amenable to scanning with berries, provided some noteworthy technical challenges are first addressed (e.g. minimizing berry damage and rotting during transportation, cluster stabilization during scanning, and segmentation of 3D volumes with features that vary widely in their X-ray absorbance). This work would provide a more complete representation of cluster structure, as well as inform our berry potential simulation with genotype-specific empirical data. We plan to develop predictive structural models of grapevine cluster development using these techniques.

Imaging and shape analysis approaches presented here can also be used to tease apart subtle environmental influences on inflorescence architecture, and the major agronomic trait of bunch density. Identifying environmental effects on phenotypic variation has important implications both for vineyard management and the assessment of intra-clone variation across geographic space. Cluster compactness can be manipulated through a variety of agronomic practices (Molitor et al. 2012; Gil et al. 2013; Frioni et al. 2017; Gourieroux et al. 2017; Poni et al. 2018; Reeve et al. 2018). Techniques described here can be used to quantify influences of specific treatments on cluster architecture. In addition, because grapevines are clonally propagated, clusters from the same widespread clones can be collected from different geographic locations, scanned and analyzed for variation. High resolution assessment of inflorescence architecture offers important insights into natural variation in bunch density and the genetic and environmental factors that influence it. The capacity to capture 3D variation in this complex trait over space and time represents a promising advance for a valuable potential target of selection in one of the most economically important berry crops in the world.

Acknowledgements

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The authors would like to acknowledge Elizabeth A. Kellogg (DDPSC) for valuable comments, particularly on phylogenetic analysis and inflorescence anatomy. We thank Noah Fahlgren (DDPSC) for computational assistance and Kari Miller (Washington University) for scanning assistance. We thank Zoë Migicovsky (Dalhousie University) for valuable comments.

Author contributions

CNT and JL designed the research; JL collected the samples and consulted on the biology; KD generated the X-ray data; LLK and AJM provided phylogenetic data and consulted for the biology; NJ and ML extracted pedicel diameter and angle; ML developed and extracted all the traits and conducted all the analysis and figures; ML, LLK, KD, JL, AJM, and CNT wrote the manuscript.

Tables

Table 1 Number of samples/individuals each species and berry information used in the study

	Number (N)			Berry information (Galet (1988); Moore and Wen (2016))		
	Samples	Individuals	Individuals used in phylogenetic analysis	low diameter (mm)	High diameter (mm)	Berries per bunch
V. acerifolia	32	11	9	8	12	>25
V. aestivalis	5	2	1	8	20	>25
V. amurensis	13	5	2	8	15	NA
V. cinerea	45	15	13	4	8	>25
V. coigneti ae	6	2	1	NA	8	NA
V. labrusca	62	22	12	12	23	<25
V. palmata	3	1	1	8	10	>25

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

V. riparia	158	53	48	8	12	>25
V. rupestris	41	16	10	8	12	<25
V. vulpina	27	9	2	8	12	>25
Total	392	136	99			

Table 2. Fifteen geometric traits were organized into three categories based on the type of shape information captured by the trait. See STable 1. for a more detailed description of each trait.

Global-size features	Local-branching features	Size-invariant features
Volume*	RachisLength*	Solidity
ConvexHullVolume*	PedicelLength	Sphericity
SurfaceArea*	AvgBranchLength	2nd/LongestBranchLength
TotalBranchLength*	BranchDiameter	PedicelLength/RachisLength
NumberOfPedicel*	PedicelDiameter	
	PedicelBranchAngle	

Size-associated features (traits with * +PH_PC1)

Supporting Information

Fig. S1 A maximum likelihood phylogenetic tree for ten *Vitis* species.
Fig. S2 Summary of inflorescence geometric and topological traits and the distribution for ten *Vitis* species.
Fig. S3 Morphological traits mapped on the phylogenetic tree.
Fig. S4. Variation for each clade.
Fig. S5 Pairwise correlations of morphological traits (allometric relationships) showing linear regression lines for each species.
Fig. S6 Pairwise species classification.
Table S1. Trait description and calculation.
Table S2. Trait variance for each species.
Table S3. Trait loadings for two species classification.
Table S4. Trait Pagel's lambda for phylogenetic analysis.
Table S5. Trait variation for each clade.
Video S1 Illustration of quantifying branching topology using persistent homology.
Video S2 Berry potential simulation

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

Aquino A, Diago MP, Millán B, Tardáguila J. 2017. A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosystems Engineering* **156**: 80–95.

Aquino A, Barrio I, Diago M-P, Millan B, Tardaguila J. 2018. vitisBerry: An Android-smartphone application to early evaluate the number of grapevine berries by means of image analysis. *Computers and Electronics in Agriculture* **148**: 19–28.

Bettiga LJ. 2003. *Wine grape varieties in California*. UCANR Publications.

Bioletti F. 1938. Outline of ampelography for the vinifera grapes in California. *Hilgardia* **11**: 227–293.

Bradley D, Carpenter R, Copsey L, Vincent C, Rothstein S, Coen E. 1996. Control of inflorescence architecture in *Antirrhinum*. *Nature* **379**: 791–797.

Brereton NJB, Ahmed F, Sykes D, Ray MJ, Shield I, Karp A, Murphy RJ. 2015. X-ray micro-computed tomography in willow reveals tissue patterning of reaction wood and delay in programmed cell death. *BMC plant biology* **15**: 83.

Bucksch A, Atta-Boateng A, Azihou AF, Battogtokh D, Baumgartner A, Binder BM, Braybrook SA, Chang C, Coneva V, DeWitt TJ, *et al.* 2017. Morphological Plant Modeling: Unleashing Geometric and Topological Potential within the Plant Sciences. *Frontiers in plant science* **8**: 900.

Callen ST, Klein LL, Miller AJ. 2016. Climatic Niche Characterization of 13 North American *Vitis* Species. *American journal of enology and viticulture* **67**: 339-349

Chanderbali AS, Berger BA, Howarth DG, Soltis PS, Soltis DE. 2016. Evolving ideas on the origin and evolution of flowers: new perspectives in the genomic era. *Genetics* **202**: 1255–1265.

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

Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. 2008. Meshlab: an open-source mesh processing tool. In: *Eurographics Italian chapter conference*. 129–136.

Cohen-Steiner D, Edelsbrunner H, Harer J. 2007. Stability of Persistence Diagrams. *Discrete & computational geometry* **37**: 103–120.

Comeaux BL, Nesbitt WB, Fantz PR. 1987. Taxonomy of the Native Grapes of North Carolina. *Castanea* **52**: 197–215.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Coombe BG. 1995. Adoption of a system for identifying grapevine growth stages. *Growth stages of the grapevine Australian Journal of Grape and Wine Research* **1**: 100–110.

Correa J, Mamani M, Muñoz-Espinoza C, Laborie D, Muñoz C, Pinto M, Hinrichsen P. 2014. Heritability and identification of QTLs and underlying candidate genes associated with the architecture of the grapevine cluster (*Vitis vinifera* L.). *Theoretical and applied genetics.* **127**: 1143–1162.

Cubero S, Diago MP, Blasco J, Tardáguila J, Millán B, Aleixos N. 2014. A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis. *Biosystems Engineering* **117**: 62–72.

Delory BM, Li M, Topp CN, Lobet G. 2018. archiDART v3.0: A new data analysis pipeline allowing the topological analysis of plant root systems. *F1000Research* **7**: 22.

Doebley J, Stec A, Hubbard L. 1997. The evolution of apical dominance in maize. *Nature* **386**: 485–488.

Earles JM, Knipfer T, Tixier A, Orozco J, Reyes C, Zwieniecki MA, Brodersen CR, McElrone AJ. 2018. In vivo quantification of plant starch reserves at micrometer resolution using X-ray microCT imaging and machine learning. *The New phytologist* **218:** 1260-1269.

Edelsbrunner H, Morozov D. 2013. Persistent homology: theory and practice.In: *Proceedings of the European congress of mathematics*, 31-50.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* **6**: e19379.

Endress PK. 2010. Disentangling confusions in inflorescence morphology: patterns and diversity of reproductive shoot ramification in angiosperms. *Journal of systematics and evolution* **48**: 225–239.

Feng C-M, Xiang Q-YJ, Franks RG. 2011. Phylogeny-based developmental analyses illuminate evolution of inflorescence architectures in dogwoods (*Cornus* s. l., Cornaceae). *The New phytologist* **191**: 850–869.

Friedman J, Harder LD. 2004. Inflorescence architecture and wind pollination in six grass species. *Functional ecology* **18**: 851–860.

Frioni T, Zhuang S, Palliotti A, Sivilotti P, Falchi R, Sabbatini P. 2017. Leaf Removal and Cluster Thinning Efficiencies Are Highly Modulated by Environmental Conditions in Cool Climate Viticulture. *American journal of enology and viticulture*: ajev.2017.16098.

Galet P. 1979. *A practical ampelography*. Cornell University Press.

Galet P. 1988. *Cépages et Vignobles de France. Tome I: Les Vignes Américaines*. Imprimérie

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

747 Déhan, Montpellier.

748 **Gil M, Esteruelas M, González E, Kontoudakis N, Jiménez J, Fort F, Canals JM,**
749 **Hermosín-Gutiérrez I, Zamora F. 2013**. Effect of two different treatments for reducing
750 grape yield in *Vitis vinifera* cv Syrah on wine composition and quality: berry thinning
751 versus cluster thinning. *Journal of agricultural and food chemistry* **61**: 4968–4978.

752 **Godin C, Caraglio Y. 1998**. A Multiscale Model of Plant Topological Structures. *Journal of*
753 *theoretical biology* **191**: 1–46.

754 **Godin C, Costes E, Sinoquet H. 1999**. A Method for Describing Plant Architecture which
755 Integrates Topology and Geometry. *Annals of botany* **84**: 343–357.

756 **Gomez FE, Carvalho G Jr, Shi F, Muliana AH, Rooney WL. 2018**. High throughput
757 phenotyping of morpho-anatomical stem properties using X-ray computed tomography in
758 sorghum. *Plant methods* **14**: 59.

759 **Gourieroux AM, Holzapfel BP, McCully ME, Scollary GR, Rogiers SY. 2017**. Vascular
760 development of the grapevine (*Vitis vinifera* L.) inflorescence rachis in response to flower
761 number, plant growth regulators and defoliation. *Journal of plant research* **130**: 873–883.

762 **Guédon Y, Barthélémy D, Caraglio Y, Costes E. 2001**. Pattern Analysis in Branching and
763 Axillary Flowering Sequences. *Journal of theoretical biology* **212**: 481–520.

764 **Hake S. 2008**. Inflorescence Architecture: The Transition from Branches to Flowers.
765 *Current biology: CB* **18**: R1106–R1108.

766 **Han Y, Yang H, Jiao Y. 2014**. Regulation of inflorescence architecture by cytokinins.
767 *Frontiers in plant science* **5**: 669.

768 **Haus MJ, Li M, Chitwood DH, Jacobs TW. 2018**. Long-Distance and Trans-Generational
769 Stomatal Patterning by CO_2 Across Arabidopsis Organs. *Frontiers in plant science* **9**: 1714.

770 **Hed B, Ngugi HK, Travis JW. 2009**. Relationship Between Cluster Compactness and Bunch
771 Rot in Vignoles Grapes. *Plant disease* **93**: 1195–1201.

772 **Hertweck KL, Pires JC. 2014**. Systematics and Evolution of Inflorescence Structure in the
773 *Tradescantia* Alliance (Commelinaceae). *Systematic botany* **39**: 105–116.

774 **Hodel DR Greby K, Ohara LM, Ohara ET. 2015**. Infructescence and fruit characteristics of
775 Washingtonia (Arecaceae: Coryphoideae). *PalmArbor* **2**: 1–7.

776 **Hodge JG, Kellogg EA. 2015**. Patterns of Inflorescence Development of Three Prairie
777 Grasses (Andropogoneae, Poaceae). *International journal of plant sciences* **175**(9): 963-974.

778 **Hughes N, Askew K, Scotson CP, Williams K, Sauze C, Corke F, Doonan JH, Nibau C.**
779 **2017**. Non-destructive, high-content analysis of wheat grain traits using X-ray micro
780 computed tomography. *Plant methods* **13**: 76.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

781 **Iland P, Dry P, Proffitt T, Tyerman S. 2011**. *The grapevine: from the science to the practice*
782 *of growing vines for wine*. Patrick Iland Wine Promotions Adelaide.

783 **Ivorra E, Sánchez AJ, Camarasa JG, Diago MP, Tardaguila J. 2015**. Assessment of grape
784 cluster yield components based on 3D descriptors using stereo vision. *Food control* **50**:
785 273–282.

786 **Jhala VM, Thaker VS. 2015**. X-ray computed tomography to study rice (Oryza sativa L.)
787 panicle development. *Journal of experimental botany* **66**: 6819–6825.

788 **Kang M-Z, Cournède P-H, Mathieu A, Letort V, Qi R, Zhan Z-G. 2009**. A Functional-
789 Structural Plant Model—Theories and Its Applications in Agronomy. In: *Crop Modeling and*
790 *Decision Support*. Springer Berlin Heidelberg, 148–160.

791 **Kellogg EA. 2007**. Floral displays: genetic control of grass inflorescences. *Current opinion*
792 *in plant biology* **10**: 26–31.

793 **Kirchoff BK, Claßen-Bockhoff R. 2013**. Inflorescences: concepts, function, development
794 and evolution. *Annals of botany* **112**: 1471–1476.

795 **Klein LL, Miller AJ, Ciotir C, Hyma K, Uribe-Convers S, Londo J. 2018**. High-throughput
796 sequencing data clarify evolutionary relationships among North American Vitis species and
797 improve identification in USDA Vitis germplasm collections. *American journal of botany*
798 **105**: 215–226.

799 **Kuijt J. 1981**. Inflorescence morphology of Loranthaceae. *Blumea* **27:** 1-73.

800 **Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018**. MEGA X: Molecular Evolutionary
801 Genetics Analysis across Computing Platforms. *Molecular biology and evolution* **35**: 1547–
802 1549.

803 **Landrein S, Prenner G, Chase MW, Clarkson JJ. 2012**. Abelia and relatives: phylogenetics
804 of Linnaeeae (Dipsacales–Caprifoliaceae s.l.) and a new interpretation of their inflorescence
805 morphology. *Botanical journal of the Linnean Society* **169**: 692–713.

806 **Le C-T, Liu B, Barrett RL, Lu L-M, Wen J, Chen Z-D. 2018**. Phylogeny and a new tribal
807 classification of Opiliaceae (Santalales) based on molecular and morphological evidence:
808 Phylogeny and classification of Opiliaceae. *Journal of Systematics and Evolution* **56**: 56–66.

809 **Letort V, Cournede P, Lecoeur J, Hummel I, Reffye PD, Christophe A. 2006**. Effect of
810 Topological and Phenological Changes on Biomass Partitioning in Arabidopsis thaliana
811 Inflorescence: A Preliminary Model-Based Study. In: *Plant Growth Modeling, Simulation,*
812 *Visualization and Application*, eds Thierry F and Zhang XP, IEEE Computer Society: 65–69.

813 **Li M, Duncan K, Topp CN, Chitwood DH. 2017**. Persistent homology and the branching
814 topologies of plants. *American journal of botany* **104**: 349–353.

815 **Li M, An H, Angelovici R, Bagaza C, Batushansky A, Clark L, Coneva V, Donoghue MJ,**

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Edwards E, Fajardo D, *et al.* 2018a. Topological Data Analysis as a Morphometric Method: Using Persistent Homology to Demarcate a Leaf Morphospace. *Frontiers in plant science* **9**: 553.

Li M, Frank MH, Coneva V, Mio W, Chitwood DH, Topp CN. 2018b. The Persistent Homology Mathematical Framework Provides Enhanced Genotype-to-Phenotype Associations for Plant Morphology. *Plant physiology* **177**: 1382–1395.

Marguerit E, Boury C, Manicki A, Donnart M, Butterlin G, Némorin A, Wiedemann-Merdinoglu S, Merdinoglu D, Ollat N, Decroocq S. 2009. Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **118**: 1261–1278.

Mathers AW, Hepworth C, Baillie AL, Sloan J, Jones H, Lundgren M, Fleming AJ, Mooney SJ, Sturrock CJ. 2018. Investigating the microstructure of plant leaves in 3D with lab-based X-ray computed tomography. *Plant methods* **14**: 99.

McAllister CA, McKain MR, Li M, Bookout B, Kellogg EA 2019. Specimen-based analysis of morphology and the environment in ecologically dominant grasses: the power of the herbarium. *Philosophical transactions of the Royal Society B: Biological sciences* **374**: 20170403.

Migicovsky Z, Li M, Chitwood DH, Myles S. 2018. Morphometrics reveals complex and heritable apple leaf shapes. *Front. Plant Sci.* **8**: 2185.

Miller AJ, Matasci N, Schwaninger H, Aradhya MK, Prins B, Zhong G-Y, Simon C, Buckler ES, Myles S. 2013. *Vitis* phylogenomics: hybridization intensities from a SNP array outperform genotype calls. *PloS one* **8**: e78680.

Molitor D, Behr M, Hoffmann L, Evers D. 2012. Impact of Grape Cluster Division on Cluster Morphology and Bunch Rot Epidemic. *American journal of enology and viticulture* **63**: 508–514.

Molitor D, Beyer M. 2014. Epidemiology, identification and disease management of grape black rot and potentially useful metabolites of black rot pathogens for industrial applications - a review. *The Annals of applied biology* **165**: 305–317.

Molitor D, Biewers B, Junglen M, Schultz M, Clementi P, Permesang G, Regnery D, Porten M, Herzog K, Hoffmann L, *et al.* 2018. Multi-annual comparisons demonstrate differences in the bunch rot susceptibility of nine *Vitis vinifera* L. 'Riesling' clones. *Vitis* **57**: 17-25.

Moore MO. 1991. Classification and systematics of eastern North American *Vitis* L. (vitaceae) north of Mexico. *SIDA, contributions to botany* **14**: 339–367.

Moore M, Wen J. 2016. Vitaceae. In: *Flora of North America North of Mexico*, ed. Flora of North America Editorial Committee. 3–23.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, *et al.* 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 453–458.

Mullins MG, Bouquet A, Williams LE. 1992. *Biology of the grapevine*. Cambridge University Press.

Munson TV. 1909. *Foundations of American Grape Culture*. Orange Judd Company.

OIV. 2001. *2nd Edition of the Oiv Descriptor List for Grape Varieties and Vitis species*.

Périlleux C, Lobet G, Tocquin P. 2014. Inflorescence development in tomato: gene functions within a zigzag model. *Frontiers in plant science* **5**: 121.

Pfeifer J, Kirchgessner N, Colombi T, Walter A. 2015. Rapid phenotyping of crop root systems in undisturbed field soils using X-ray computed tomography. *Plant methods* **11**: 41.

Poni S, Gatti M, Palliotti A, Dai Z, Duchêne E, Truong T-T, Ferrara G, Matarrese AMS, Gallotta A, Bellincontro A, *et al.* 2018. Grapevine quality: A multiple choice issue. *Scientia horticulturae* **234**: 445–462.

Prusinkiewicz P, Erasmus Y, Lane B, Harder LD, Coen E. 2007. Evolution and development of inflorescence architectures. *Science* **316**: 1452–1456.

Pulliat V. 1888. Mille variétés de vignes, description et synonymies, Paris Montpellier.

Reeve AL, Skinkis PA, Vance AJ, McLaughlin KR, Tomasino E, Lee J, Tarara JM. 2018. Vineyard Floor Management and Cluster Thinning Inconsistently Affect 'Pinot noir' Crop Load, Berry Composition, and Wine Quality. *HortScience: a publication of the American Society for Horticultural Science* **53**: 318–328.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in ecology and evolution* **3**: 217–223.

de Ribou S de B, Douam F, Hamant O, Frohlich MW, Negrutiu I. 2013. Plant science and agricultural productivity: Why are we hitting the yield ceiling? *Plant science: an international journal of experimental plant biology* **210**: 159–176.

Richter R, Gabriel D, Rist F, Töpfer R, Zyprian E. 2018. Identification of co-located QTLs and genomic regions affecting grapevine cluster architecture. *Theoretical and applied genetics*. doi.org/10.1007/s00122-018-3269-1

Rist F, Herzog K, Mack J, Richter R, Steinhage V, Töpfer R. 2018. High-Precision Phenotyping of Grape Bunch Architecture Using Fast 3D Sensor and Automation. *Sensors* **18**: 763.

Rombough L. 2002. *The Grape Grower: A Guide to Organic Viticulture*. Chelsea Green Publishing.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

887 **Roscher R, Herzog K, Kunkel A, Kicherer A, Töpfer R, Förstner W. 2014**. Automated
888 image analysis framework for high-throughput determination of grapevine berry sizes
889 using conditional random fields. *Computers and Electronics in Agriculture* **100**: 148–158.

890 **Rovasenda J. 1881**. Essai d'une ampélographie universelle. traduit de l'Italien par
891 F.Cazalis, G. Foëx et al., Paris Montpellier.

892 **Shavrukov YN, Dry IB, Thomas MR. 2004**. Inflorescence and bunch architecture
893 development in *Vitis vinifera* L. *Australian journal of grape and wine research* **10**: 116–124.

894 **Stitzer MC, Ross-Ibarra J. 2018**. Maize domestication and gene interaction. *New*
895 *phytologist* **220**: 395–408.

896 **Swofford DL. 2003**. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.

897 **Ta KN, Khong NG, Ha TL, Nguyen DT, Mai DC, Hoang TG, Phung TPN, Bourrie I,**
898 **Courtois B, Tran TTH, *et al.* 2018**. A genome-wide association study using a Vietnamese
899 landrace panel of rice (*Oryza sativa*) reveals new QTLs controlling panicle morphological
900 traits. *BMC plant biology* **18**: 282.

901 **Tello J, Aguirrezábal R, Hernáiz S, Larreina B, Montemayor MI, Vaquero E, Ibáñez J.**
902 **2015**. Multicultivar and multivariate study of the natural variation for grapevine bunch
903 compactness: Multicultivar study of grapevine bunch compactness. *Australian journal of*
904 *grape and wine research* **21**: 277–289.

905 **Tello J, Ibáñez J. 2018**. What do we know about grapevine bunch compactness? A state-of-
906 the-art review: Review on bunch compactness. *Australian journal of grape and wine*
907 *research* **24**: 6–23.

908 **Tröndle D, Schroder S, Kassemeyer H-H, Kiefer C, Koch MA, Nick P. 2010**. Molecular
909 phylogeny of the genus *Vitis* (Vitaceae) based on plastid markers. *American journal of*
910 *botany* **97**: 1168–1178.

911 **Venables WN, Ripley BD. 2002**. *Modern Applied Statistics with S. Fourth Edition*. Springer.

912 **Weberling F. 1992**. *Morphology of Flowers and Inflorescences*. Cambridge University Press.

913 **Wei T, Simko V. 2017**. R package 'corrplot': visualization of a correlation matrix (version
914 0.84). URL https://github. com/taiyun/corrplot.

915 **Whipple CJ. 2017**. Grass inflorescence architecture and evolution: the origin of novel
916 signaling centers. *New phytologist* **216**: 367–372.

917 **Wyatt R. 1982**. Inflorescence architecture: how flower number, arrangement, and
918 phenology affect pollination and fruit-set. *American journal of botany* **69**: 585–594.

919 **Zdunić G, Mucalo A, Budić-Leto I, Humar I, Pejić I, Maletić E**. Cluster architecture of old,
920 neglected Croatian grapevine varieties (*Vitis vinifera* L.). *Vitis* **54**: 177-180.

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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

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

Figure Legends

Fig. 1 Sample preparation and imaging. (a) The ten *Vitis* species sampled for this study display diverse grape bunch morphology. (b) Inflorescence architectures after berry removal. (c) Inside the X-ray tomography instrument; the inflorescence is clamped in a panavise between two pieces of polystyrene on the X-ray turntable. (d) Two dimensional radiogram of grape inflorescence; X-rays, absorbed or passing through the inflorescence, are detected to create a silhouette. (e) Three dimensional reconstruction and the structure of the same inflorescence shown in (d) by taking radiograms at successive different angles and then computationally combining the images.

Fig. 2 Examples of inflorescence geometric and topological traits and their distribution for ten *Vitis* species. Each panel shows one of the three traits categories (geometric traits, topological traits, and berry potential traits). Geometric traits are organized as global size features, local branching features, and size-invariant features. Each trait is listed at the top of the column and two inflorescence examples demonstrating low and high trait values listed to the left. At the bottom of each column is a boxplot indicating the distribution and variance within the ten *Vitis* species, represented in different colors. On each box, each dot indicates an outlier if it is more than 1.5 interquartile ranges; the central vertical line indicates the median; the left and right edges of the box represent the 25th and 75th percentiles; and the whiskers extend to the most extreme nonoutlier data. The label for each species is listed in the boxplot y axis of the leftmost plot, with the number of individuals sampled for each species shown in parentheses. For a more complete example and detailed description of each trait, see Fig. S2 and Table S1.

Fig. 3 Persistent homology with geodesic distance comprehensively quantifies branching structures. (a) A level (pink solid line) defined by the same geodesic distance (length of any of the purple curves, in this case, set to 90) to the base of the inflorescence. The super level set is the pixels (in black) having greater geodesic distance than the pink level. (b) Pixels on a branching structure are colored by their geodesic distance to the base. They are colored with red
representing the most distant through to blue for the closest ones. (c) A persistence barcode for each branching structure records the connected components for each level set at each geodesic distance value. The “birth” and “death” values for each bar represent the level where each branch starts and gets merged. Colored bars correspond to colored branches. (d) Above: example inflorescence. The stem is digitally cut at the base (brown

bioRxiv preprint doi: https://doi.org/10.1101/557819; this version posted February 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

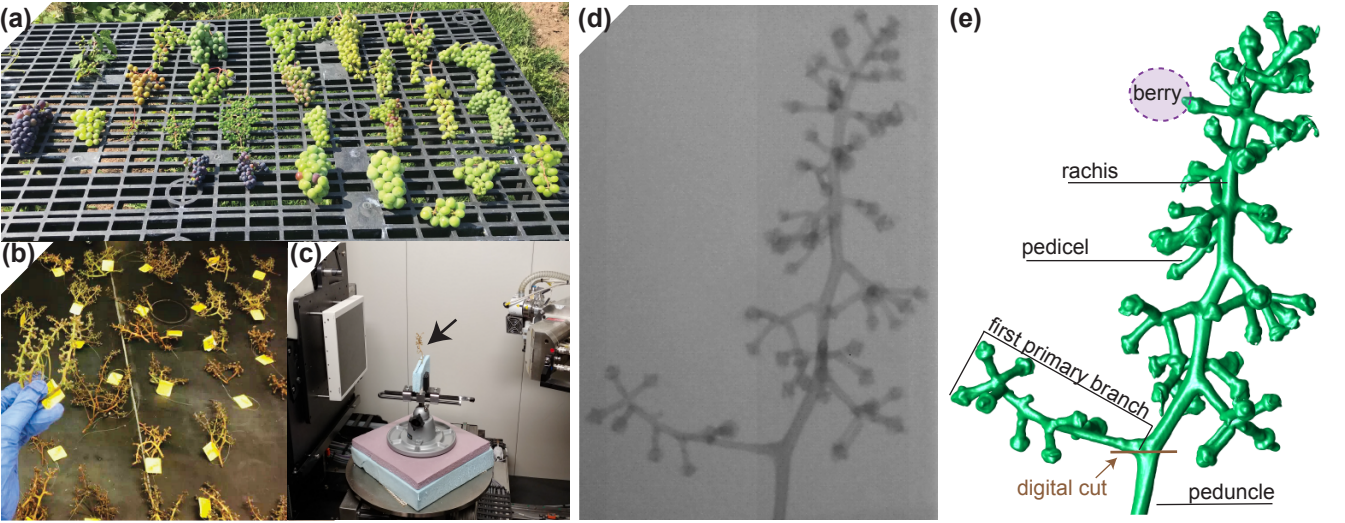
line) where it meets the first branch. Below: 3D surface on the example inflorescence as in (b). (e) Persistence barcode for the inflorescence in (d). (f) and (g), similar to (d) and (e), show a different inflorescence architecture.

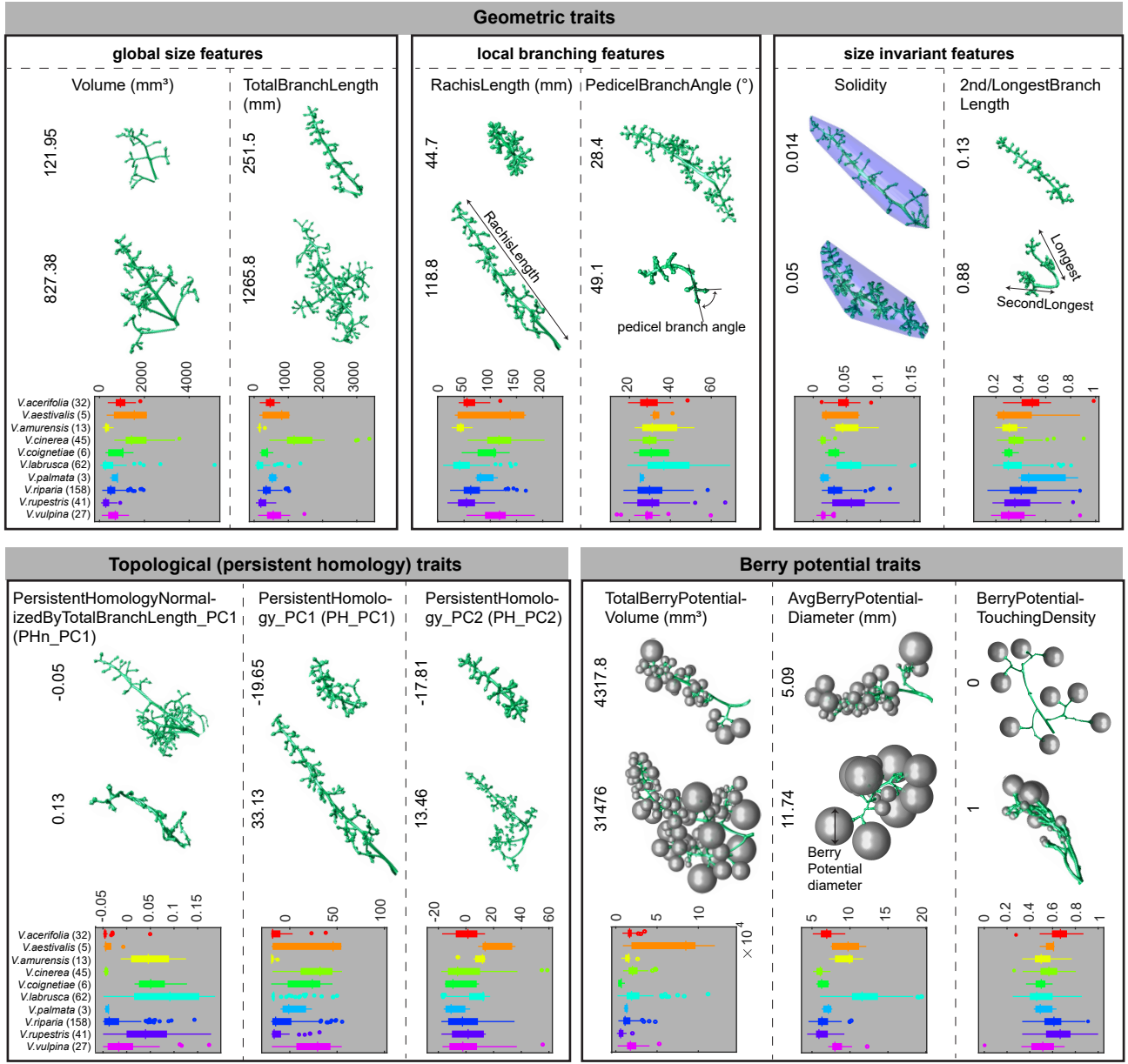
Fig. 4 Berry potential simulation to explore the space determined by inflorescence architecture. (a) Determine the growth direction for each berry potential. (b) Expand berry potential by increasing the size and moving the center along the growth direction until it meets any of these three cases: 1) two berry potentials touch each other; 2) a berry potential touches any part of the inflorescence; 3) the diameter of the berry potential reaches the maximum for the species..

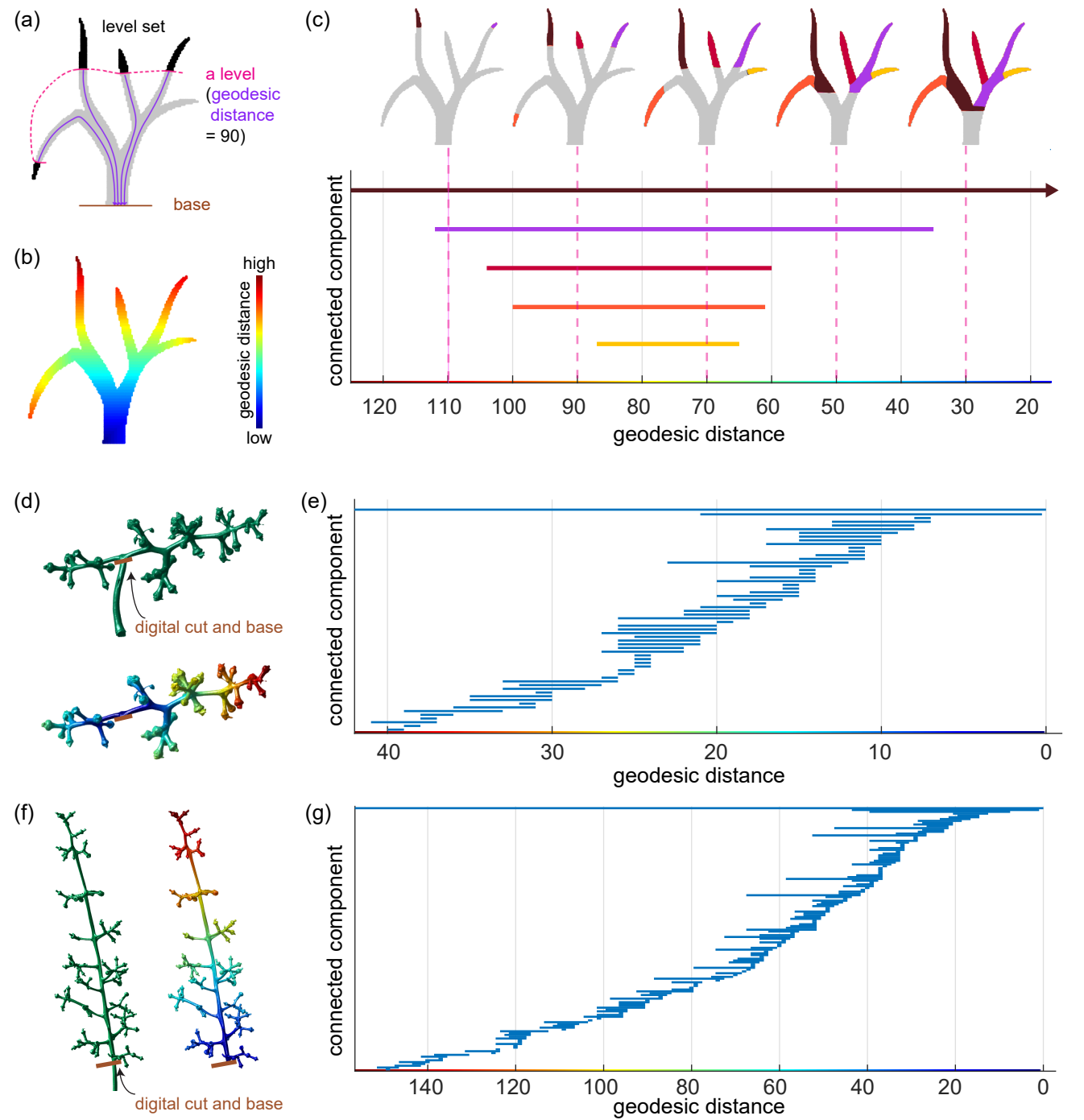
Fig. 5 Hierarchical cluster analysis and correlation analysis. (a) Cluster analysis based the mean value for each trait of 10 *Vitis* species. The heatmap shows values above (red) or below (blue) the mean for each trait. The morphological traits (rows) are clustered hierarchically with the name shown on the right and hierarchical tree listed on the left. The species (columns) are also clustered hierarchically with the name and hierarchical tree shown at the top. (b) Correlation matrix plot shows pairwise positively stronger correlation (green and larger circle) or negatively stronger correlation (purple and larger circle). Non-significant correlations ($p>0.05$) are crossed out. The traits are ordered in the same way as (a). (c) Selected pairs of traits showing linear regression lines for each species.

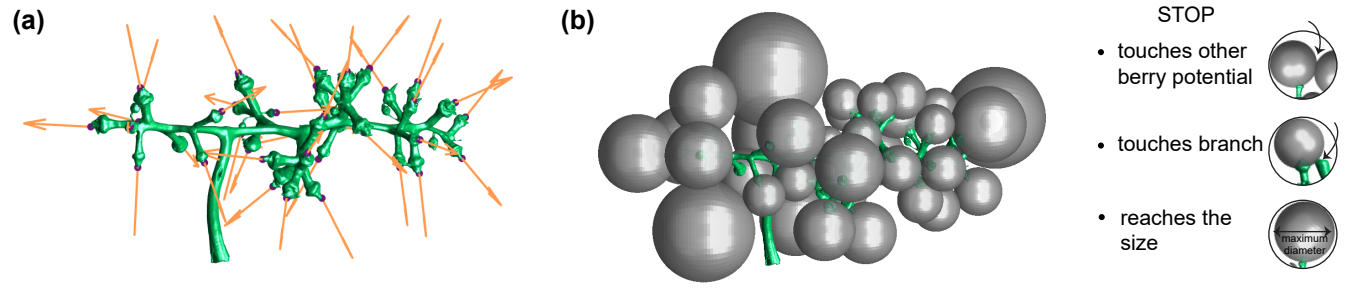
Fig. 6 Classification for ten *Vitis* species based on inflorescence architecture. (a) Left: Principal component analysis (PCA) plot on 24 morphological traits. The percent variance for each PC explained is shown in parentheses. Species are shown in different colors. Right: The loadings for the traits that contribute to the variance are shown. (b) Left: Linear discriminant analysis (LDA) plot on the first 18 PCs (99.5% variance). Species are shown in different colors. The confusion matrix for predicted species is shown in the upper right corner. Right: The loadings for the traits that best distinguish species from each other are shown. Using a jacknifed 'leave one out' cross validation, we obtain a 78.32% classification accuracy rate.

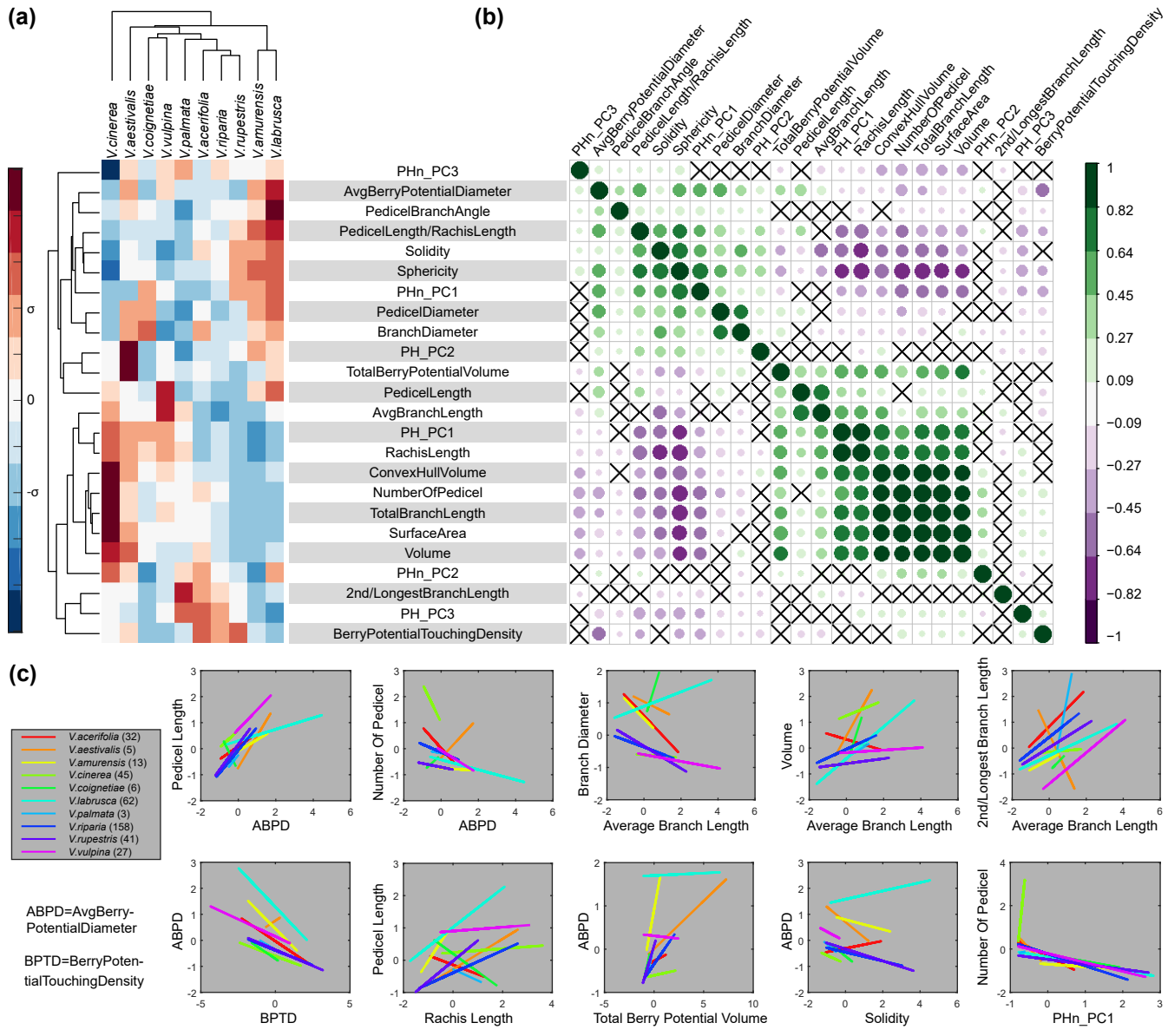
Fig. 7 Phylogenetic analysis. A Neighbor Joining phylogenetic tree for a subset of the *Vitis* data set (n=99). Node values denote bootstrap support for values greater than or equal to 50. Ten *Vitis* species are highlighted in different colored backgrounds. Three clades (Asian Clade, NA Clade I, and NA Clade II) are labeled and marked by vertical bars. The barplot showing values of Pagel's lambda, an estimate of phylogenetic signal, overlaps with the trait name on the right top panel. Below each trait, a rainbow colormap shows the values for individuals (small values in red to large values in blue). Rectangles surround the trait value map for species with more than five individuals. One trait (PHn_PC1) was randomly selected to be projected onto the phylogenetic tree branches, and indicates trait variation (red, lower values; blue, higher values) within individuals and among clades.

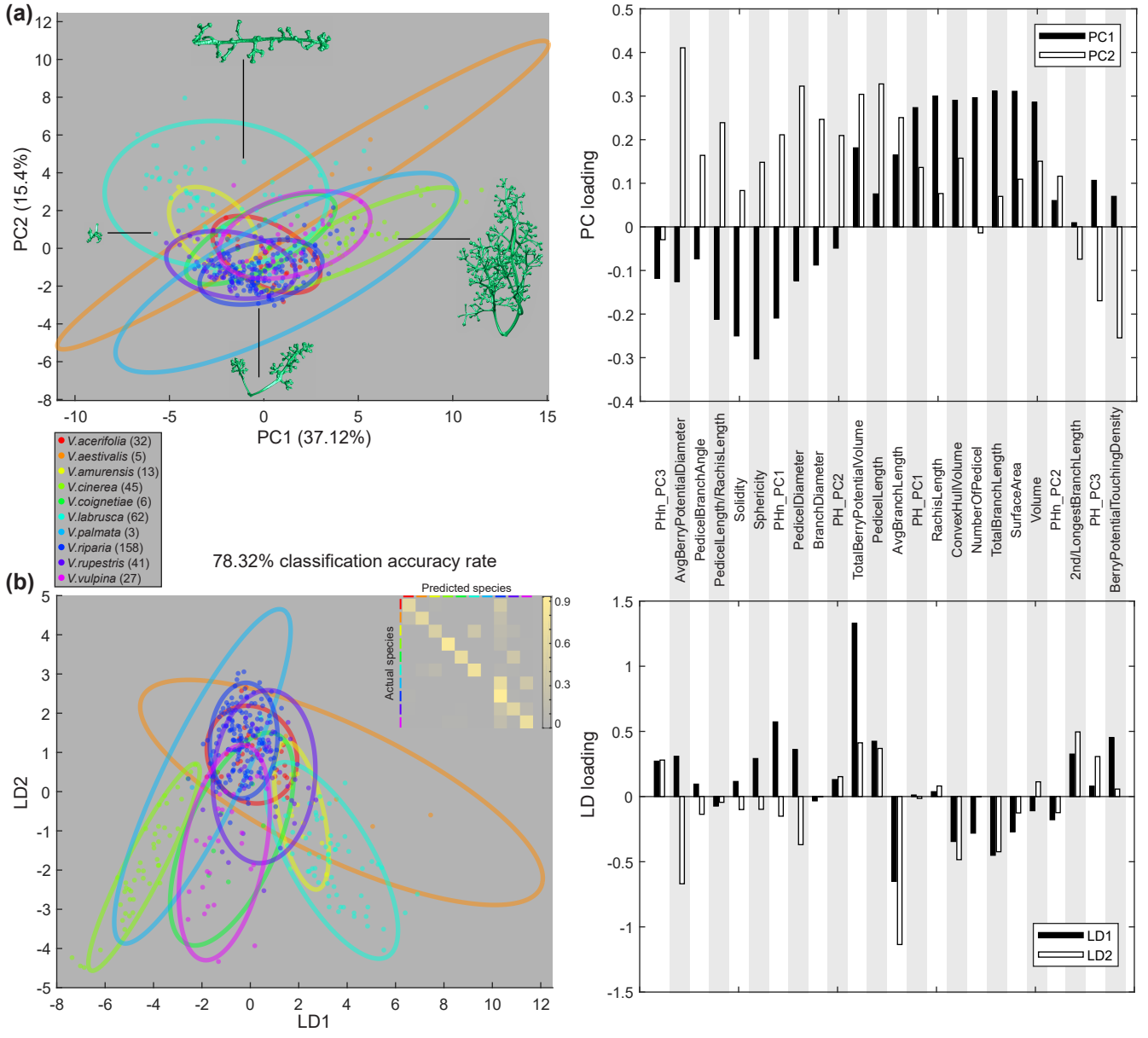
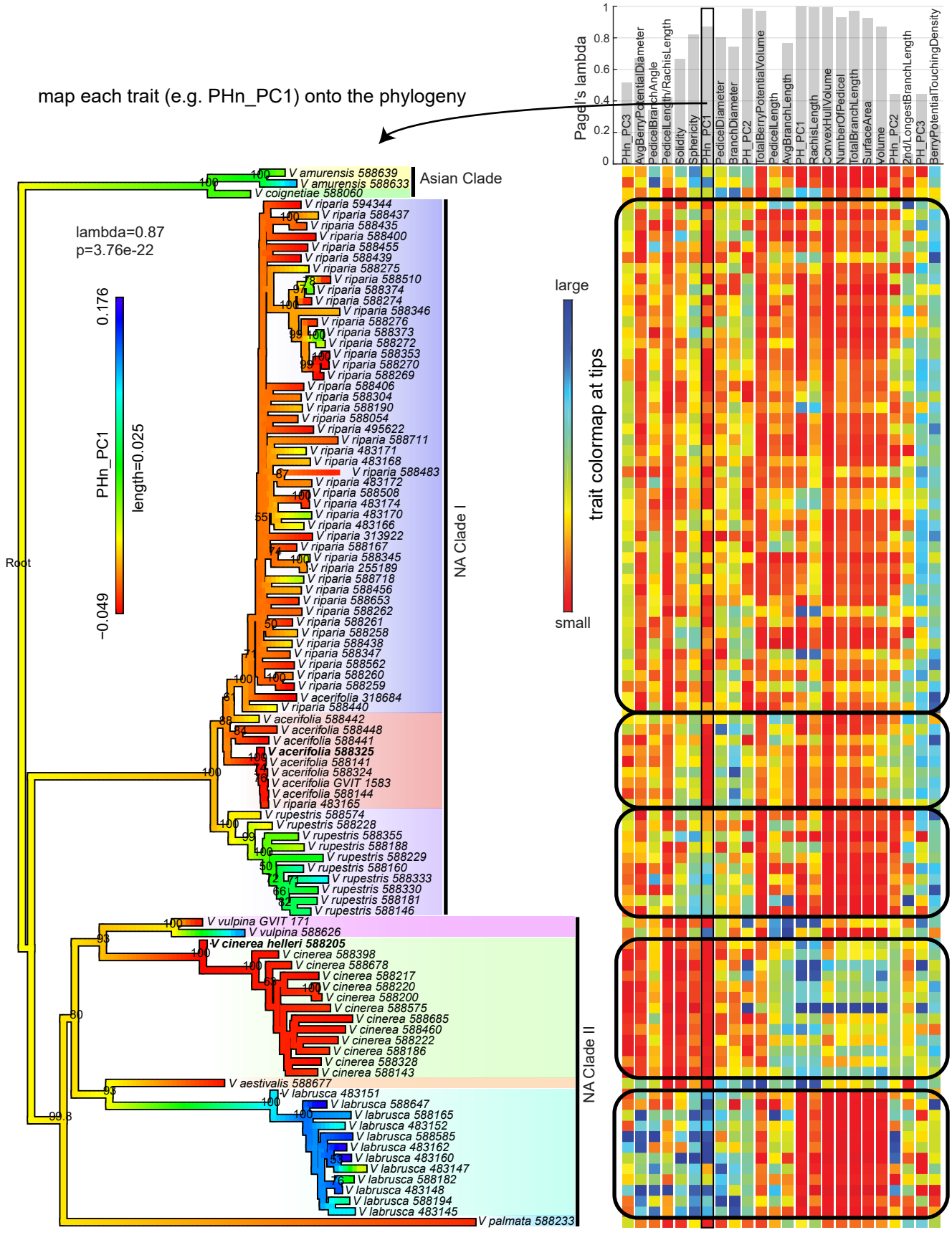

(a)
(b)
(c)
(d)
(e)
berry
rachis
pedicel
first primary branch
digital cut
peduncle


Geometric traits
global size features
local branching features
size invariant features
Volume (mm³)
TotalBranchLength (mm)
RachisLength (mm)
PedicelBranchAngle (°)
Solidity
2nd/LongestBranch Length
Examples with low and high trait values
Distribution
121.95
827.38
251.5
1265.8
44.7
118.8
28.4
49.1
0.014
0.05
0.13
0.88
RachisLength
pedicel branch angle
Longest
SecondLongest
V.acerifolia (32)
V.aestivalis (5)
V.amurensis (13)
V.cinerea (45)
V.coignetiae (6)
V.labrusca (62)
V.palmata (3)
V.riparia (158)
V.rupestris (41)
V.vulpina (27)
Topological (persistent homology) traits
Berry potential traits
PersistentHomologyNormalizedByTotalBranchLength_PC1 (PHn_PC1)
PersistentHomology_PC1 (PH_PC1)
PersistentHomology_PC2 (PH_PC2)
TotalBerryPotentialVolume (mm³)
AvgBerryPotentialDiameter (mm)
BerryPotentialTouchingDensity
-0.05
0.13
-19.65
33.13
-17.81
13.46
4317.8
31476
5.09
11.74
0
1
Berry Potential diameter
×10⁴


(a)
level set
a level
(geodesic
distance
= 90)
base
(b)
high
geodesic distance
low
(c)
connected component
120 110 100 90 80 70 60 50 40 30 20
geodesic distance
(d)
digital cut and base
(e)
connected component
40 30 20 10 0
geodesic distance
(f)
digital cut and base
(g)
connected component
140 120 100 80 60 40 20 0
geodesic distance


(a)
(b)
STOP
• touches other berry potential
• touches branch
• reaches the size
maximum diameter


(a)
V.cinerea
V.aestivalis
V.coignetiae
V.vulpina
V.palmata
V.acerifolia
V.riparia
V.rupestris
V.amurensis
V.labrusca
σ
0
-σ
PHn_PC3
AvgBerryPotentialDiameter
PedicelBranchAngle
PedicelLength/RachisLength
Solidity
Sphericity
PHn_PC1
PedicelDiameter
BranchDiameter
PH_PC2
TotalBerryPotentialVolume
PedicelLength
AvgBranchLength
PH_PC1
RachisLength
ConvexHullVolume
NumberOfPedicel
TotalBranchLength
SurfaceArea
Volume
PHn_PC2
2nd/LongestBranchLength
PH_PC3
BerryPotentialTouchingDensity
(b)
1
0.82
0.64
0.45
0.27
0.09
−0.09
−0.27
−0.45
−0.64
−0.82
−1
(c)
V.acerifolia (32)
V.aestivalis (5)
V.amurensis (13)
V.cinerea (45)
V.coignetiae (6)
V.labrusca (62)
V.palmata (3)
V.riparia (158)
V.rupestris (41)
V.vulpina (27)
ABPD=AvgBerry-PotentialDiameter
BPTD=BerryPoten-tialTouchingDensity
Pedicel Length
ABPD
Number Of Pedicel
ABPD
Branch Diameter
Average Branch Length
Volume
Average Branch Length
2nd/Longest Branch Length
Average Branch Length
ABPD
BPTD
Pedicel Length
Rachis Length
ABPD
Total Berry Potential Volume
ABPD
Solidity
Number Of Pedicel
PHn_PC1


(a)
PC2 (15.4%)
PC1 (37.12%)
V.acerifolia (32)
V.aestivalis (5)
V.amurensis (13)
V.cinerea (45)
V.coignetiae (6)
V.labrusca (62)
V.palmata (3)
V.riparia (158)
V.rupestris (41)
V.vulpina (27)
PC loading
PC1
PC2
PHn_PC3
AvgBerryPotentialDiameter
PedicelBranchAngle
PedicelLength/RachisLength
Solidity
Sphericity
PHn_PC1
PedicelDiameter
BranchDiameter
PH_PC2
TotalBerryPotentialVolume
PedicelLength
AvgBranchLength
PH_PC1
RachisLength
ConvexHullVolume
NumberOfPedicel
TotalBranchLength
SurfaceArea
Volume
PHn_PC2
2nd/LongestBranchLength
PH_PC3
BerryPotentialTouchingDensity
(b)
78.32% classification accuracy rate
Predicted species
Actual species
LD2
LD1
LD loading
LD1
LD2

map each trait (e.g. PHn_PC1) onto the phylogeny