The grapevine leaves that are (or were) not: constraints on leaf development and choosing to see what seemingly is not

Daniel H. Chitwood[1,2,*], Joey Mullins[1]
[1]Dept. Horticulture, [2]Dept. Computational Mathematics, Science & Engineering, Michigan State University, East Lansing, MI 48823 USA

*Correspondence: Daniel H. Chitwood, chitwoo9@msu.edu

ABSTRACT

The shapes of leaves have been extensively measured and modeled. Using conventional statistical approaches there exist powerful methods to discern between groups and classify shapes. Embedded in calculated morphospaces is information that allows us to visualize theoretical leaves. These unmeasured leaves, that may or may not exist, are never considered in their relation to empirically observable leaves nor how the negative morphospace can inform us about the forces responsible for shaping leaf morphology. Using landmark data for 8465 grapevine leaves, we model leaf shape using an allometric indicator of leaf size, the ratio of vein to blade areas. The borders of the observable morphospace are restricted by the constraint of the subregions of the leaf reducing to nothing. Assuming that self-intersecting polygonal regions in leaves are a developmental constraint, we define a region of morphospace that represents leaf shapes that are possible but not observed. We end with a discussion of how we choose not to observe what seemingly does not exist and the ramifications of this perspective to survivorship bias in the sciences in general.

INTRODUCTION

Leaf shape across plants is diverse and spectacular, but it is not random. Development, evolution, and the environment sculpt leaf shape in specific ways (Chitwood and Sinha, 2016). Leaves allometrically expand, first shown by Stephen Hales through pin pricks on developing fig leaves that were displaced differentially along the length versus the width of the leaf (1727). The developmental programing of leaves changes from node-to-node resulting in changing leaf shapes. Goethe described this process as “metamorphosis” and in terms of the mutable, changing internal state of leaves (1817). Environment modulates leaf size and serrations, as observed by Bailey and Sinnott (1915) who used the distribution of entire leaves across latitudes to estimate the temperatures of paleoclimates. If we measure leaf shape across the seed plants, clear demarcations between phylogenetic groups are observed (Li et al., 2018). We have measured enough leaf shapes to know the borders and demarcations of what exists and the processes that shape leaves in specific ways.

The shapes of grapevine leaves have been measured under intense scrutiny and with purpose. Originally through morphometric techniques developed by Louis Ravaz (1902), the field of ampelography ("vine" + "process of measuring") sought to discern, using leaves and other features of the vine, American *Vitis* species that were new to Europeans and would eventually be used as rootstocks against Phylloxera. Eventually the techniques would be famously applied to wine grape varieties by Pierre Galet (1979; 1985; 1988; 1990; 2000; Chitwood, 2020). Morphometric techniques have been used to genetically study the basis of leaf shape in grapevines (Chitwood et al., 2014; Demmings et al., 2019), how grapevine leaves develop (Chitwood et al., 2016a), as well as the effects of environment (Chitwood et al., 2016b; Baumgartner et al., 2020). Modeling has been used in several ways, including calculating average shapes of grapevine varieties while preserving features (Martínez et al., 1995; 1997a; 1997b; 1999), modeling development across grapevine shoots (Bryson et al., 2020), and using leaf allometry, specifically the ratio of vein to blade areas, as a proxy of leaf size and to measure the effects of year-to-year variation in leaf shape (Chitwood et al., 2021). For grapevines, as for many other types of leaves, we have extensively measured and modeled leaf shape, allowing us to discern genetic, developmental, and environmental effects with great power.

But what about leaves that are not available for us to measure? Or what about leaves that never existed? Using what we know about the underlying structure of leaf morphospaces across genotypic, developmental, end environmental effects, and making modeling assumptions about what is and is not possible, could we compare what we have measured and observed against the boundaries of what we know is possible to reconstruct leaves that *might have been* or *could be*?

Here, we measure the shapes of 8465 grapevine leaves and model them against an allometric indicator of leaf size, vein-to-blade ratio, across *Vitis* species. The expansion of blade area at the expense of that for veins is found to be a principal determinant of the resulting morphospace. Using an inverse transform of the Principal Component Analysis (PCA) space, theoretical leaves around the border of empirically observable leaves are visualized. These border leaves are defined by the reduction of blade or vein area (or both) to the point that the areas of subregions in the leaf are reduced to nothing. Assuming that self-intersecting polygonal regions represent theoretical leaves that cannot exist, a region of morphospace containing shapes that are *possible* but not *observed* is defined using this developmental constraint. We end with a discussion of how we choose not to observe what seemingly does not exist and the ramifications of this perspective to survivorship bias in the sciences.

MATERIALS AND METHODS

As described in Bryson et al., 2020 (and copied verbatim here for convenience), a total of 8465 leaves were collected from 209 vines at the USDA germplasm repository vineyard in Geneva, New York, USA. Samples were taken from the same vines during the second week of June, annually, in 2013 and 2015–2017. The vines sampled represent 11 species (*Ampelopsis glandulosa* (Wall.) Momiy. var. *brevipedunculata* (Maxim.) Momiy., *V. acerifolia* Raf., *V. aestivalis* Michx., *V. amurensis* Rupr., *V. cinerea* (Engelm.) Millardet, *V. coignetiae* Pulliat ex Planch., *V. labrusca* L., *V. palmata* Vahl, *V. riparia* Michx., *V. rupestris* Scheele, and *V. vulpina* L.), four hybrids (*V.* ×*andersonii* Rehder, *V.* ×*champinii* Planch., *V.* ×*doaniana* Munson ex Viala, and *V.* ×*novae-angliae* Fernald), and 13 *Vitis* vines with an unassigned identity. Starting at the shoot tip (with shoot order noted for each leaf), leaves greater than ~1 cm in length were collected in stacks and stored in a cooler in labeled plastic bags with ventilation holes. Within two days of collection, the leaves were arranged on a large-format Epson Workforce DS-50000 scanner (Tokyo, Japan) in the order they were collected, with a small number near each leaf indicating which node it came from and a ruler for scale within the image file. The image files were named with the vine identification number, followed by a sequential lowercase letter if multiple scans were needed. The original scans are available on Dryad (Chitwood et al., 2020). 21 landmarks were placed on one half of each leaf outlining the midvein, distal vein, proximal vein, and the most proximal branching vein of each of these major veins as well as distal and proximal lobe sinuses using ImageJ (Abràmoff et al., 2004). Two landmarks are placed at the base of each vein to measure the width. The exact placement and correspondence of landmark number to leaf features can be found in Bryson et al., 2020. Landmarks were superimposed through scaling, translation, rotation, and reflection using Generalized Procrustes Analysis with the shapes (Dryden and Mardia, 2016) package in R. It is the Procrustes-adjusted landmark coordinates from Bryson et al., 2020 that are used as the starting point for the analysis presented here.

Data was analyzed using Python and Jupyter notebooks (Kluyver et al., 2016). Code to reproduce the analysis in this manuscript can be found at the *Github* repository DanChitwood/leaves_that_are_or_were_not: https://github.com/DanChitwood/leaves_that_are_or_were_not). The Jupyter notebook (.ipynb) comments on the code and also contains a narrative to guide the reader through the analysis. Calculation of the natural log of the ratio of vein to blade area, $\ln(vein\ to\ blade\ ratio)$, is as described in Chitwood et al. (2021) using the shoelace algorithm, also known as Gauss' are formula, to calculate polygon areas as originally described by Meister (1769), where n is the number of polygon vertices defined by x and y coordinates:

$$\frac{1}{2}|x_1y_2 + x_2y_3 + \ldots + x_{n-1}y_n + x_ny_1 - x_2y_1 - x_3y_2 - \ldots - x_ny_{n-1} - x_1y_n|$$

Principal Component Analysis (PCA) (and calculation of its inverse) was performed using the scikit learn decomposition PCA module (Pedregosa et al., 2011). Modeling of $\ln(vein\ to\ blade\ ratio)$, principal components, and landmarks as polynomial functions of each other and shoot position was performed using the np.polyfit and np.poly1d functions from NumPy (Oliphant, 2006). An implementation of the Bentley Ottman sweep-line algorithm (Bentley and Ottman, 1979) was used to detect self-intersecting polygons in leaf shapes (Kiers and Barton, 2010). Pandas (McKinney, 2010) and Matplotlib (Hunter, 2007) were used for data analysis and visualization.

RESULTS

Allometry

Previously, we modeled leaf shape continuously across grapevine shoots as a polynomial function of each Procrustes landmark coordinate value as a function of relative node position. Relative node position is the node number counting from the shoot tip divided the total number of leaves in a shoot, such that node number is converted to a 0 to 1 scale, from tip to base (Bryson et al., 2020). We also previously described the natural log of the ratio of vein to blade area, $\ln(vein\ to\ blade\ ratio)$, which is more sensitive to leaf area than size itself due to the exponential increases in blade relative to vein area during development (Chitwood et al., 2021).

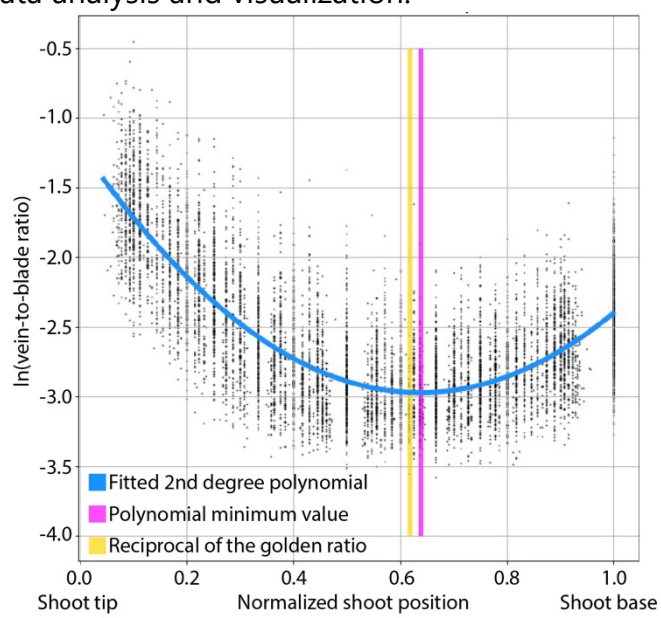

Figure 1: Modeling *ln(vein to blade ratio)* as a function of normalized shoot position. The natural log of the ratio of vein-to-blade area, *ln(vein to blade ratio)*, is modeled as a 2nd degree polynomial of normalized shoot position (where 0 is the shoot tip and 1 is the shoot base). The normalized shoot position value corresponding to the minimum *ln(vein to blade ratio)* value is 0.6379505834434538 (magenta vertical line). This value has a 3.22% difference from the value of the inverse of the golden ratio, 0.61803398875 (gold vertical line).

Before we explore the limits of the grapevine leaf morphospace (and beyond), we must first model shape across development to understand how continuous developmental trajectories change between species during evolution. We plotted $\ln(vein\ to\ blade\ ratio)$ versus relative shoot position (**Figure 1**), which can be modeled as a second-degree polynomial. $\ln(vein\ to\ blade\ ratio)$ is highest at the shoot tip and reaches its minimum in the middle of the shoot. At the shoot tip, veins dominate young,

small leaves before blade area starts to exponentially expand pushing the ratio value down. $\ln(vein\ to\ blade\ ratio)$ increases again at the shoot base where the first leaves that emerge are smaller than those that emerge subsequently. A curiosity that is perhaps coincidental, we note that the corresponding relative node position to the minimum $\ln(vein\ to\ blade\ ratio)$ value is within 3.22% of the inverse of the golden ratio (**Figure 1**). Although this may arise as a developmental phenomenon, it could also be spurious and warrants further investigation.

From previous work we know that allometric changes during grapevine leaf expansion dominate the morphospace (Chitwood et al., 2016a; Chitwood et al., 2016b; Bryson et al., 2020). We therefore performed a Principal Component Analysis (PCA) on Procrustes-adjusted landmark coordinates and modeled this space as a function of $\ln(vein\ to\ blade\ ratio)$ (**Figure 2**). PCA decomposes multivariate data, in essence rotating and projecting it onto orthogonal axes (principal components) that more efficiently explain variation in the data than the original measurements (in this case, Procrustes-adjusted coordinate values). The inverse of this transformation can be used to reverse calculate original data, which we will later use to visualize theoretical leaves in the morphospace. PC1 and PC2 explain 90.2% and 9.6% of the variance in the data, respectively, and can modeled as simply second-degree polynomial functions of $\ln(vein\ to\ blade\ ratio)$ that for some species can almost be linear (**Figures 2A,B**). These models can subsequently be combined in a visualization of PC2 vs. PC1, showing the developmental trajectory of different species in the first two principal components that explain 99.8% of all variance (**Figure 2C**). We can then visualize all the data, colored by $\ln(vein\ to\ blade\ ratio)$, and superimpose on this the developmental trajectories of different *Vitis* and *Ampelopsis* species as modeled by $\ln(vein\ to\ blade\ ratio)$ (**Figure 2D**). Note that $\ln(vein\ to\ blade\ ratio)$ is the major source of variance in the morphospace, varying strongly and continuously by both PC1 and PC2. $\ln(vein\ to\ blade\ ratio)$ also makes a strong, straight border in this space along its lowest values (white in **Figure 2D**) that correspond to large, mature leaves with a small ratio of vein-to-blade areas. The modeled trajectories of different species each traverse through higher (young, small leaves) and lower (mature, large leaves) $\ln(vein\ to\ blade\ ratio)$ values, but the different species trajectories separate orthogonally to this direction. As such, development as measured by $\ln(vein\ to\ blade\ ratio)$ and genetic variation as measured through different species define the morphospace.

We can visualize the continuous allometric changes in leaf shape across the modeled developmental trajectories (**Figure 3**). We can both visualize modeled leaf shapes themselves and keep track of the proportion of the leaf occupied by each region and

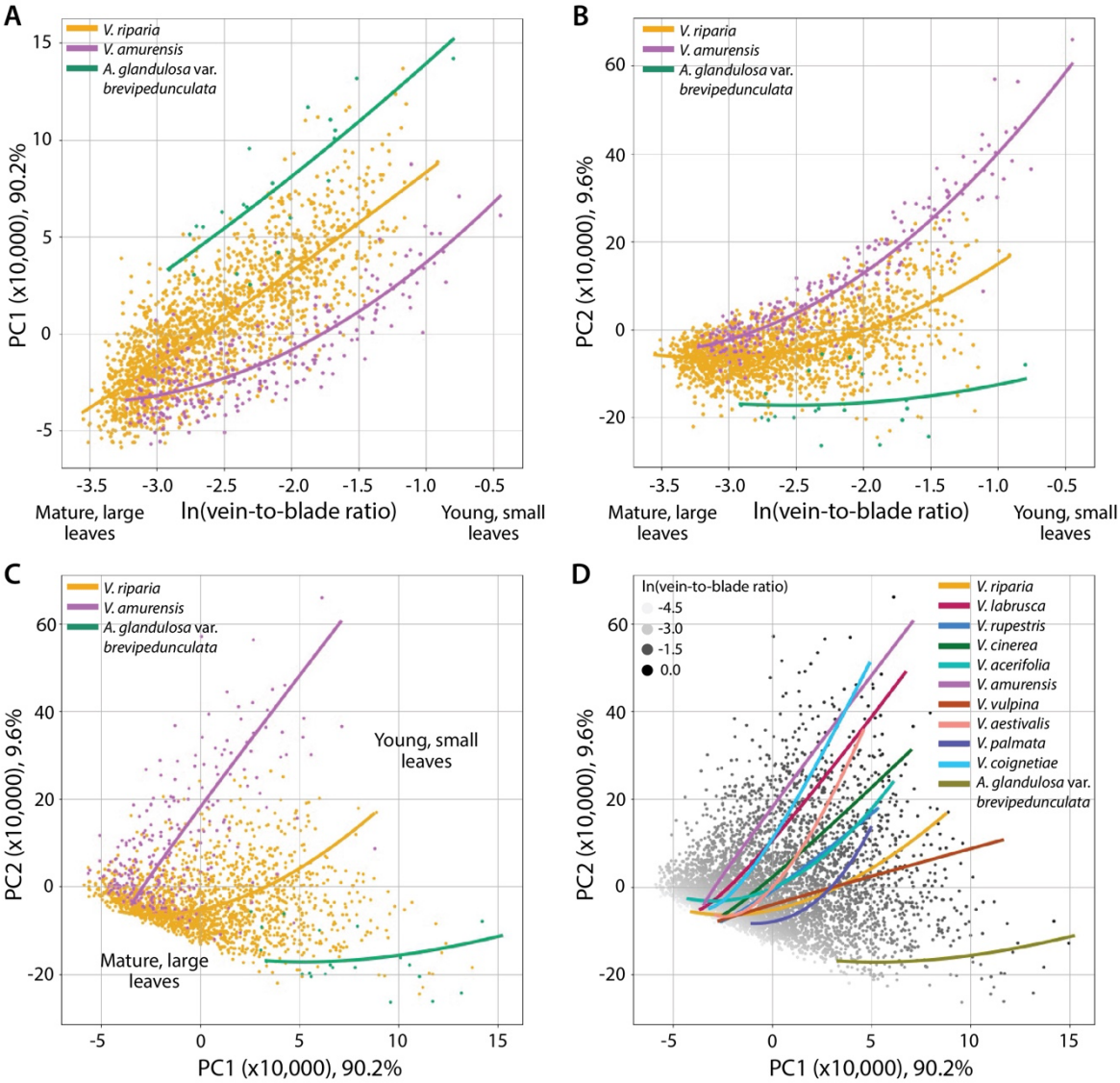

Figure 2: Developmental trajectories of leaf development through Principal Component Analysis (PCA) space by species. A-B) 2nd degree polynomials modeling PC1 (A) and PC2 (B) values by *ln(vein to blade ratio)* for three species (*Vitis riparia*, gold; *V. amurensis*, lavender; *Ampelopsis glandulosa* var. *brevipedunculata*, green). **C)** Data points and models plotted as PC2 vs. PC1. Regions corresponding to mature, old leaves and young, small leaves are indicated. **D)** Models for 11 species (indicated by color) and data points colored by *ln(vein to blade ratio)* (a white-to-black gradient, scale provided). Percent variance explained by each PC are indicated in axis labels.

clearly see how $\ln(vein\ to\ blade\ ratio)$ corresponds to the vein and blade regions it measures. Young and small leaves have relatively high ratios of vein area (yellow in **Figure 3**). As leaves expand in size and mature, blade regions expand at an exponential rate before plateauing, and reduce the relative proportion of vein area in the leaf. Note that in *A. glandulosa* var. *brevipedunculata*, which is deeply lobed, the ratios of vein and blade area are more stable across $\ln(vein\ to\ blade\ ratio)$ values compared to more entire species. In one of the most entire species, *V. amurensis*, the decreases in vein area as the leaf expands are more dramatic, perhaps corresponding to the fact that there is indeed more blade that expands.

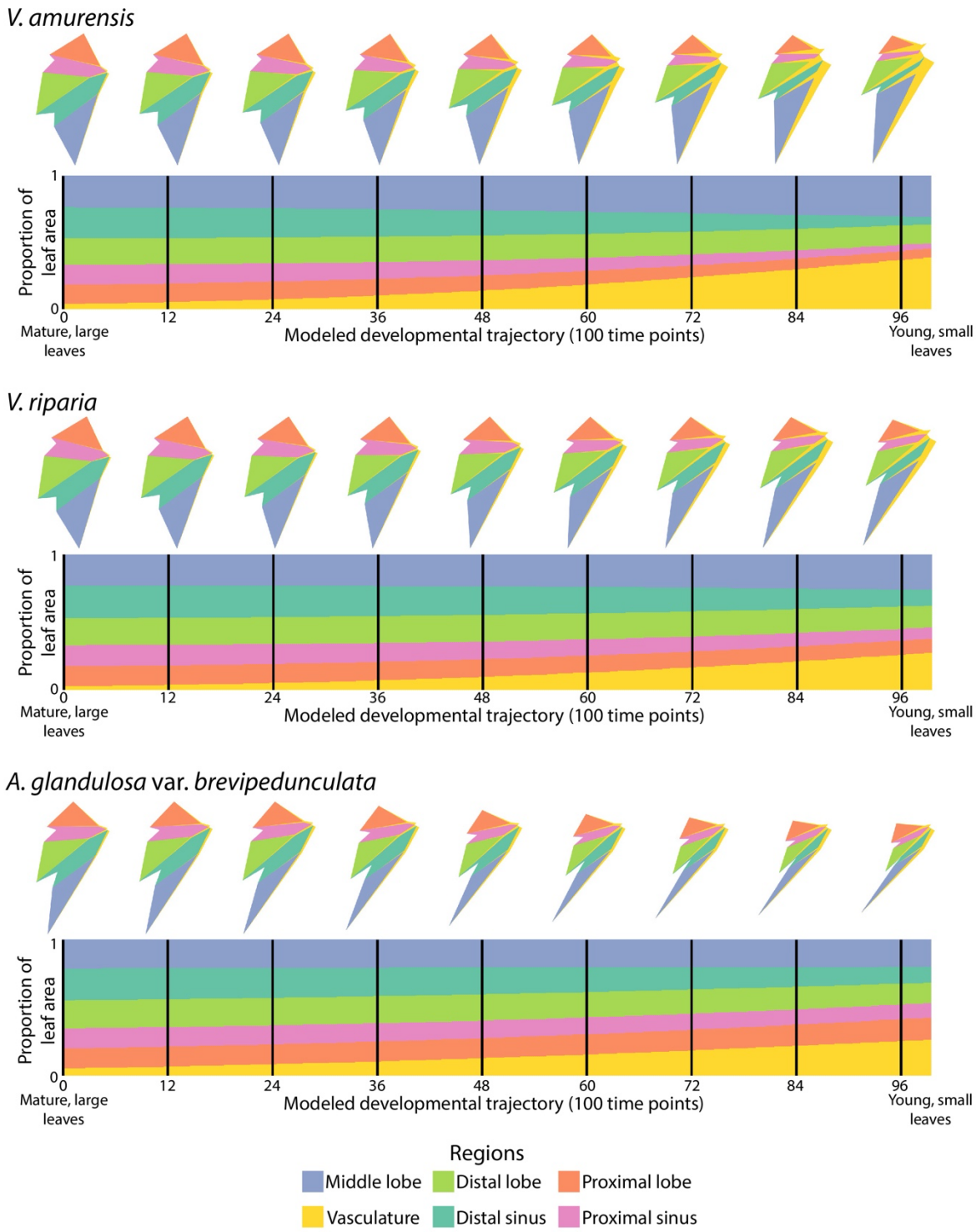

Figure 3: Allometric changes in leaf shape during development. Changes in the proportional area different regions of the leaf occupy (middle lobe, lavender blue; vasculature, yellow; distal lobe, light green; distal sinus, dark green; proximal lobe, orange; proximal sinus, magenta) are modeled as developmental trajectories across 100 theoretical time points for three different species: *V. riparia*, *V. amurensis*, and *A. glandulosa* var. *brevipedunculata*. Theoretical grapevine leaves are reconstructed for timepoints 0, 12, 24, 36, 48, 60, 72, 84, and 96 for each species. Note that the trajectories run from mature, large leaves on the left to young, small leaves on the right.

209

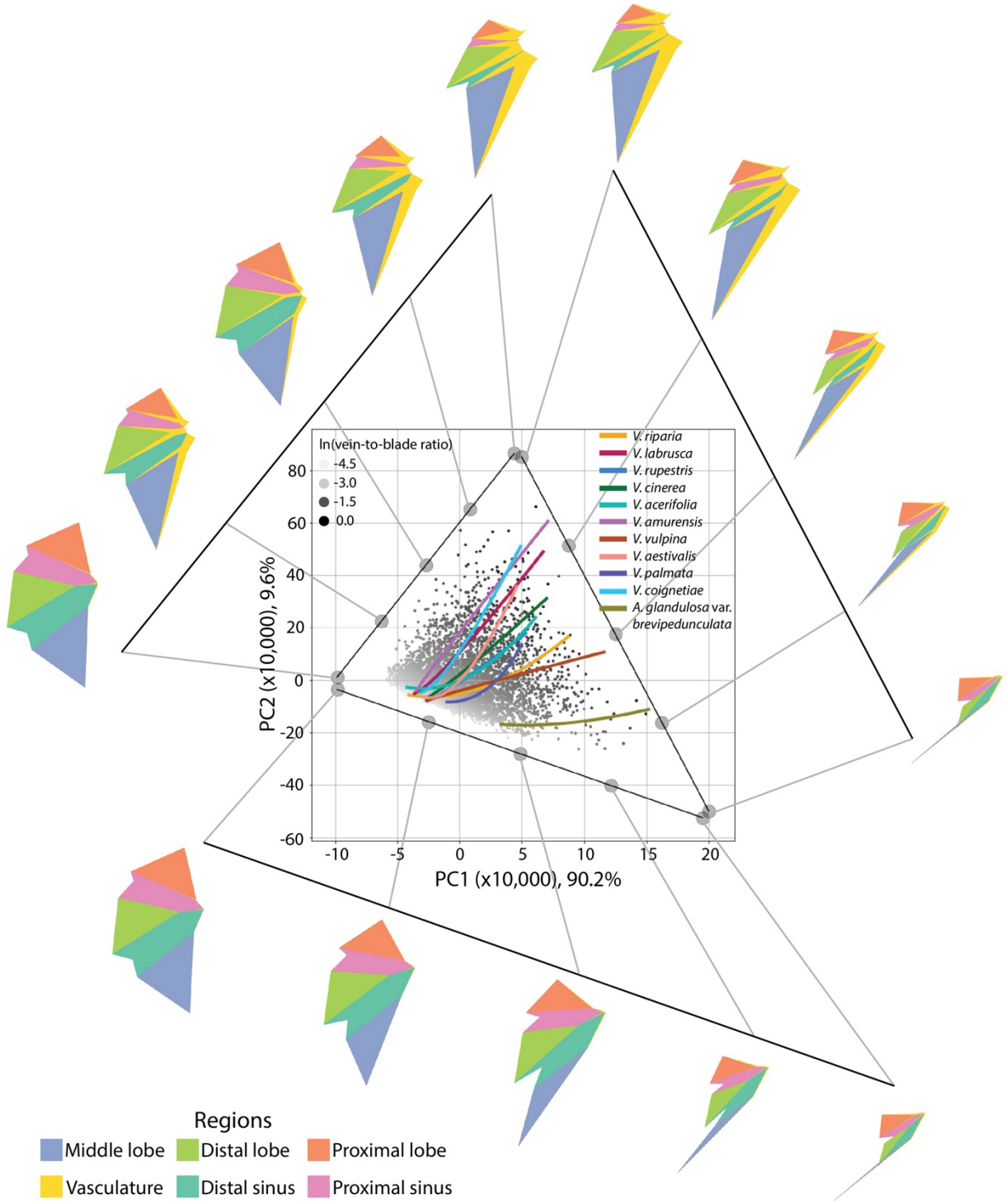

Figure 4: Theoretical grapevine leaves at the limits of what is empirically observed. Principal Component Analysis (PCA) space plotting individual leaves colored by *ln(vein to blade ratio)* (a white-to-black gradient, scale provided) and modeled developmental trajectories of different species (indicated by color). Three lines are drawn that encompass the shape of all empirically measured leaves. Theoretical grapevine leaves are reconstructed for five equidistant points along each line (large gray dots), pointed to by lines outside of the plot. Different regions of the leaf are indicated by color: middle lobe, lavender blue; vasculature, yellow; distal lobe, light green; distal sinus, dark green; proximal lobe, orange; proximal sinus, magenta.

Morphospace

The morphospace is clearly defined (**Figure 2D**): for the leaves we have measured there is a strong, continuous gradient of $\ln(vein\ to\ blade\ ratio)$ values and for the highest values that correspond to mature, large leaves in which blade has expanded at the expense of veins, there is a clearly delineated border that the developmental trajectories of species terminate at. For the leaves that we have measured, what do leaves at the border of the morphospace look like, the most "extreme" leaves? Drawing three, arbitrary lines around the data and taking five equidistant points in morphospace along each, theoretical leaves were reconstructed using PCA loadings and the inverse transformation (**Figure 4**). The most obvious varying feature of these border leaves is that which most strongly affects the structure of the morphospace itself: $\ln(vein\ to\ blade\ ratio)$. Along the defined line of border leaves with low $\ln(vein\ to\ blade\ ratio)$ values (white in **Figure 4**), the vein area is almost non-existent relative to blade in reconstructed leaves. The line opposite it in the border triangle, the leaves with high $\ln(vein\ to\ blade\ ratio)$ values (black in **Figure 4**), the opposite is true, leaves have thick veins and reduced blade area. These two lines intersect at a point with high PC1 and low PC2 values that combines the reductions in blade and vein into an unsustainable crisis: an imploding leaf with no blade or vein area. Opposite this vertex, the line connecting the spaces of low and high $\ln(vein\ to\ blade\ ratio)$ values shows what one would expect: a gradual transition between a leaf with almost no vein area to one where it predominates.

If we take Procrustes-adjusted landmarks to be a model of leaf shape, then we can consider instances that violate assumptions of reality on which that model is based. In this specific case, if any of the polygonal areas that comprise a reconstructed leaf self-intersect and cross themselves, this is a leaf that cannot exist because of developmental constraint: it simply is not possible. This does not imply that the condition of non-intersecting polygonal areas accurately represents the actual boundaries of constraint, only that for our model of leaf shape, that it is clearly a violation of what is possible. Expanding beyond the PCA morphospace and reconstructing theoretical leaves, we can use the Bentley Ottman sweep-line algorithm to test if leaves contain any self-intersecting polygonal regions (**Figure 5**). We can also measure $\ln(vein\ to\ blade\ ratio)$ values of these theoretical leaves to help us understand why such constraints arise as well as visualize the leaves themselves (**Figure 6**).

On at least two sides and a point of intersection of the empirically observed distribution of the grapevine leaves, the border of leaves that are self-intersecting comes close to that of those that we measured. These borders are along the sharply defined line of leaves with low $\ln(vein\ to\ blade\ ratio)$ values in which blade is greatly expanded at the expense of vein regions and along the opposite border with high $\ln(vein\ to\ blade\ ratio)$

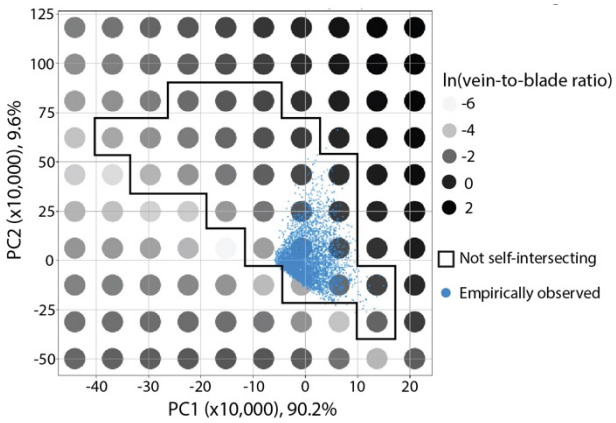

Figure 5: Self-intersecting leaves. 100 theoretical leaves, equidistantly spaced along PC1 and PC2, were reconstructed. The range was selected to find the limits of theoretical leaves with any self-intersection in any sub-region of the leaf. For each reconstructed leaf, *ln(vein to blade ratio)* was calculated (a white-to-black gradient, scale provided). The boundary that encompasses theoretical leaves with no self-intersection of any region is indicated by a solid, black line. Empirically measured leaves are indicated by points (dodger blue).

values in which vein areas predominate at the expense of blade (**Figures 5-6**). Also, the "point of implosion" where these two axes intersect, where the reduction of both vein and blade area simultaneously reduces the areas of all regions to the point of non-existence, corresponds closely to that observed in measured leaves. If we take this apex and we extend the two lines of low and high $\ln(vein\ to\ blade\ ratio)$ values outward, we find that there is a substantial region of morphospace of leaves *without* self-intersecting regions that extends beyond the region of empirically measured leaves in this study. These are leaves, given the caveats of our model, that potentially *could have*, *do*, or *can* exist but have not been measured.

DISCUSSION

In exploring the leaves beyond the boundaries that have been empirically observed and are described in this manuscript, it is important that we distinguish between different types. First (1), the leaves that exist or existed: there are (1a) the leaves that we have measured here and (1b) the leaves that we have not measured but *do* or *have* existed. Although relatively comprehensive across *Vitis* spp. excluding varieties of *V. vinifera* (discussed in the following paragraphs below), the leaves empirically observed in this manuscript do not come close to measuring all *Vitis* leaves. There are (1bi) extant species and genotypes known or unknown that are not represented, (1bii) all the environments and circumstances those extant genotypes have existed in, (1biii) extinct species and genotypes, and (1biv) the environments and circumstances that extinct grapevine leaves existed in. Then there are (2) the leaves that have not existed or maybe never will. There are (2a) leaves that *could* have existed or *can*, that represent (2ai) evolutionary, genetic, and environmental possibilities that remain unsampled or never were sampled or (2aii) leaves that *cannot* exist and never have and never will. This last category is what is referred to as *developmental constraint* here and, given the limitations of the model of grapevine leaf shape, is being measured by observing self-intersecting polygonal areas.

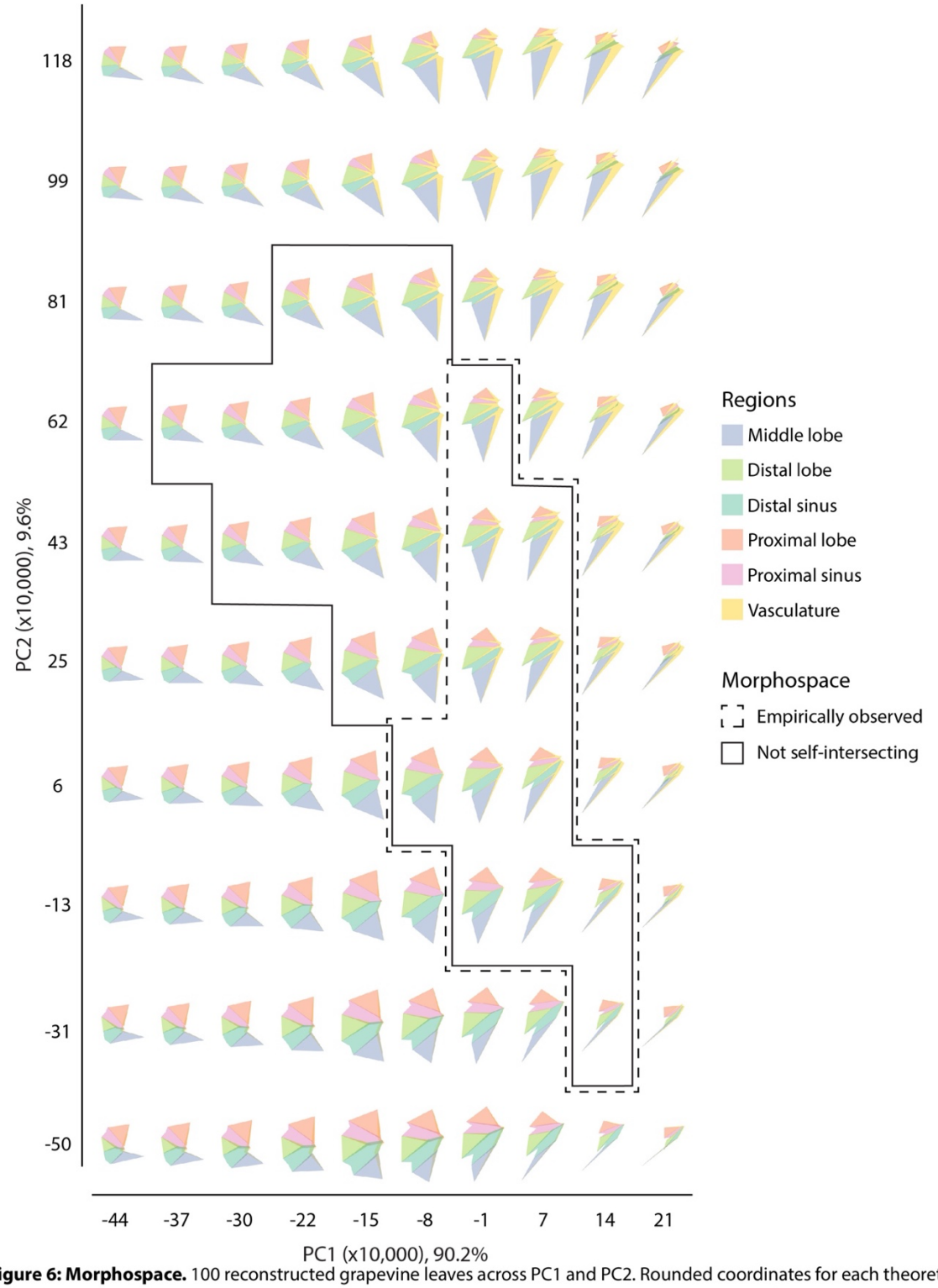

Figure 6: Morphospace. 100 reconstructed grapevine leaves across PC1 and PC2. Rounded coordinates for each theoretical leaf are indicated along the axes. The boundary that encompasses theoretical leaves with no self-intersection of any region is indicated by a solid black line. The boundary that encompasses nearly all empirically measured leaves is indicated by a dashed black line. Different regions of the leaf are indicated by color: middle lobe, lavender blue; vasculature, yellow; distal lobe, light green; distal sinus, dark green; proximal lobe, orange; proximal sinus, magenta. Alpha transparency is used so that overlapping regions in leaves can be seen.

291

In describing the features of leaves that define the morphospace, $\ln(vein\ to\ blade\ ratio)$ is the strongest, but a detail needs to be clarified. High PC1 and PC2 values correspond to a region with high $\ln(vein\ to\ blade\ ratio)$ values (black in **Figure 5**). The border of leaves with self-intersecting polygonal regions opposite of this space has the lowest observed $\ln(vein\ to\ blade\ ratio)$ values (the band of white in **Figure 5**), but then it appears that $\ln(vein\ to\ blade\ ratio)$ increases again in the area defined by low PC1 and PC2 values. The shoelace algorithm can be used to calculate the areas of self-intersecting polygons, and these are non-negative values. The area defined by low PC1 and PC2 values has vein areas so reduced that they have "imploded" and self-intersected on themselves (observe reconstructed leaves from this area in **Figure 6**), increasing $\ln(vein\ to\ blade\ ratio)$ values in a misleading way. Unlike **Figure 5** shows, the morphospace is divided into regions of strongly high and low $\ln(vein\ to\ blade\ ratio)$ values, with empirically measured leaves in the middle. With respect to developmental constraint, it likely does limit observable grapevine leaf shapes. The most mature, largest leaves encounter a constraint in which vein widths cannot be further reduced and blade area is maximized (low PC1 and PC2 values); the youngest, smallest leaves cannot eliminate any more blade region at the expense of vasculature (high PC1 and PC2 values); and then there is the "point of implosion" (high PC1, low PC2 values), in which reductions in both blade and vein area threaten the existence of any leaf area at all (**Figure 6**).

The most interesting region of the morphospace identified in this study is that in which no self-intersection is detected but there are no empirically measured leaves (**Figure 6**). A characteristic of these leaves is a "croissant" shape: it appears as if the leaf is bending over upon itself. A first thought is that no empirically measured leaves occupy this space because of the fact we measured leaf halves, that there is no way these shapes can make a whole leaf and it is an artifact of the analysis. However, these half leaves can be put together into whole leaves, it will just be that the petiolar veins and sinus (that are comprised of the "proximal lobe" and "proximal sinus" regions in this study) will overlap. An overlapping petiolar sinus is a historical ampelographic feature, especially in *V. vinifera* varieties (Ravaz, 1902; Galet, 1979; 1985; 1988, 1990, 2000). We do have scans and have performed landmark analysis in *V. vinifera* varieties previously (Chitwood et al., 2014; Chitwood, 2020), but we used a fewer number of landmarks that do not measure the widths of veins (and therefore do not measure vein area). Additionally, the *V. vinifera* leaves that have been measured do not include young leaves, which are mostly contributing to leaves with high $\ln(vein\ to\ blade\ ratio)$ values. It is outside of the scope of this manuscript to measure these leaves, but it is an important next step. Landmarking and projecting these leaves onto the morphospace will reveal how much of the space without developmental constraints but not currently occupied by the leaves measured in this study is occupied and if it is truly empty. It might be possible that

developmental constraint, defined by self-intersecting regions of the leaf, is sufficient to explain the boundaries of extant grapevine leaves and likewise that grapevine leaves fully occupy the space available to them.

Although reconstructing leaves from a PCA morphospace is routine statistically, this work focuses on interpretation and how we can use morphometrics to see shape and natural phenomena through different lenses. Embedded in the morphology of morphospaces we measure are the constraints by which development and evolution are modulating natural forms. Measured in sufficient quantities and making reasonable assumptions about the limits of our models, we can begin to deduce and quantify constraint, drawing a line between that which is impossible and that which is possible but may not exist (or once existed). Both in past and present, morphometrics has been used for nefarious purposes to distinguish groups of people and justify exclusion, discrimination, and erasure (Clayton, 2020), and it is from this past that the current scientific focus on classification and parametric statistics is based. But we can *choose* to explore our data more fully and focus on more than only what exists (or that we choose to acknowledge). Failing to do so is a type of survivorship bias: by only considering that which exists, is present, survived, or that we selectively acknowledge, we are led to false conclusions and commit logical errors (Wald, 1943; Wallis, 1980). This philosophy can be extended to how we conduct ourselves and relate to others in science as well, from how we address and prevent demographic biases in the sciences (Madzima and MacIntosh, 2021; Field and Rajewski, 2021) to the response of the academy to racial injustices and the pandemic (Montgomery, 2020). By choosing to see what seemingly is not—who we exclude around us but nonetheless exist, who have made contributions we neglect, and who are part of the community and its history but have not been properly acknowledged—we can see a bigger picture and better understand the dynamics that underlie the spaces that we study and inhabit.

DATA AVAILABILITY STATEMENT

The code to reproduce this analysis can be found at the *Github* repository DanChitwood/leaves_that_are_or_were_not: https://github.com/DanChitwood/leaves_that_are_or_were_not. The Procrustes landmark data was originally used in Bryson et al., 2020 and is posted at the *Github* repository DanChitwood/grapevine_shoots: https://github.com/DanChitwood/grapevine_shoots. The original leaf scans used to produce the landmarks are archived on Dryad (Chitwood et al., 2020).

AUTHOR CONTRIBUTIONS

JM developed methods and tools, performed data analysis, and reviewed the manuscript. DHC conceived of the project, performed data analysis, and wrote the manuscript.

FUNDING

This work is supported by the National Science Foundation Plant Genome Research Program award number 1546869. This project was supported by the USDA National Institute of Food and Agriculture, and by Michigan State University AgBioResearch.

REFERENCES

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International* 11, 36-42.

Bailey, I. W., and Sinnott, E. W. (1915). A botanical index of Cretaceous and Tertiary climates. *Science* 41 (1066), 831-4.

Baumgartner, A., Donahoo, M., Chitwood, D. H., and Peppe, D. J. (2020). The influences of environmental change and development on leaf shape in *Vitis*. *American Journal of Botany* 107, 1-13.

Bentley, J. L., and Ottmann, T. A. (1979). Algorithms for reporting and counting geometric intersections. *IEEE Transactions on Computers* C-28 (9), 643–647.

Bryson, A. E., Brown, M. W., Mullins, J., Dong, W., Bahmani, K., Bornowski, N., Chiu, C., Engelgau, P., Gettings, B., Gomezcano, F., Gregory, L. M., Haber, A. C., Hoh, D., Jennings, E. E., Ji, Z., Kaur, P., Kenchanmane Raju, S. K., Long, Y., Lotreck, S., Mathieu, D. T., Ranaweera, T., Ritter, E. J., Sadohara, R., Shrote, R. Z., Smith, K. E., Teresi, S. J., Venegas, J., Wang, H., Wilson, M. L., Tarrant, A. R., Frank, M. H., Migicovsky, Z., Kumar, J., VanBuren, R., Londo, J. P., and Chitwood, D. H. (2020). Composite modeling of leaf shape along shoots discriminates *Vitis* species better than individual leaves. *Applications in Plant Sciences* 8 (12), e11404.

Chitwood, D. H., and Sinha, N. R. (2016). Evolutionary and environmental forces sculpting leaf development. *Current Biology* 26 (7), R297-306.

Chitwood, D. H, Ranjan, A., Martinez, C. C., Headland, L. R., Thiem, T., Kumar, R., Covington, M. F., Hatcher, T., Naylor, D. T., Zimmerman, S., et al. (2014). A modern

411 ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant*
412 *Physiology* 164, 259-272.
413
414 Chitwood, D. H., Klein, L. L., O'Hanlon, R., Chacko, S., Greg, M., Kitchen, C., Miller, A. J.,
415 and Londo, J. P. (2016a). Latent developmental and evolutionary shapes embedded
416 within the grapevine leaf. *New Phytologist* 210, 343-355.
417
418 Chitwood, D. H., Rundell, S. M., Li, D. Y., Woodford, Q. L., Tommy, T. Y., Lopez, J. R.,
419 Greenblatt, D., Kang, J., and Londo, J. P. (2016b). Climate and developmental plasticity:
420 interannual variability in grapevine leaf morphology. *Plant Physiology* 170, 1480-1491.
421
422 Chitwood, D. H., VanBuren, R., Migicovsky, Z., Frank, M., and Londo, J. (2020). Data from:
423 Latent developmental and evolutionary shapes embedded within the grapevine
424 leaf. *Dryad*. Dataset, doi.org/10.5061/dryad.zkh189377.
425
426 Chitwood, D. H. (2020). The shapes of wine and table grape leaves: an ampelometric
427 study inspired by the methods of Pierre Galet. *Plants, People, Planet* 00, 1-16.
428
429 Chitwood, D. H., Mullins, J., Migicovsky, Z., Frank, M., VanBuren, R., and Londo, J.P.
430 (2021). Vein-to-blade ratio is an allometric indicator of climate-induced changes in
431 grapevine leaf size and shape. *American Journal of Botany* 108 (4), 571-579.
432
433 Clayton, A. (2020). How eugenics shaped statistics. *Nautilus* 28 October. Retrieved 24
434 January 2022 https://nautil.us/how-eugenics-shaped-statistics-9365/
435
436 Demmings, E. M., Williams, B. R., Lee, C. R., Barba, P., Yang, S., Hwang, C. F., Resich, B. I.,
437 Chitwood, D. H., and Londo, J. P. (2019). Quantitative Trait Locus Analysis of Leaf
438 Morphology Indicates Conserved Shape Loci in Grapevine. *Frontiers in Plant Science* 10,
439 1373.
440
441 Dryden, I. L, and Mardia, K. V. (2016). Statistical shape analysis: with applications in
442 R. Vol. 995. Hoboken, New Jersey, USA: John Wiley & Sons.
443
444 Field, S., and Rajewski, A. (2021). Challenges facing LGBTQ+ early-career scientists and
445 how to engage in changing the status quo. *Plant Cell* 33, 1859-1862.
446
447 Galet, P. (1979). *A practical ampelography* (L.T. Morton, Trans.). Ithaca, USA: Cornell
448 University Press.
449
450 Galet, P. (1985). *Précis d'ampélographie pratique*, 5 ed., Montpellier, France: Déhan.

Galet, P. (1988). *Cépages et vignobles de France*, vol. I, *Les vignes américaines*. Montpellier, France: Déhan.

Galet, P. (1990). *Cépages et vignobles de France*, vol. II. *L'ampélographie française*. Montpellier, France: Déhan.

Galet, P. (2000). *Dictionnaire encyclopédique des cépages*. Paris, France: Hachette.

Goethe, J. W. (1817). *Goethe's Werk, Italienische Reise*. Stuttgart, Germany: Dreizehnter Band.

Hales, S. (1727). *Vegetable Staticks: Or, an Account of Some Statical Experiments on the Sap in Vegetables: Being an Essay Towards a Natural History of Vegetation. Also, a Specimen of an Attempt to Analyse the Air, by a Great Variety of Chymio-statical Experiments; which Were Read at Several Meetings Before the Royal Society* (Vol. 1). p. 344, Figs. 43-45. London, England: W. and J. Innys and T. Woodward.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9, 90-95.

Kiers, B., and Barton, C. (copyrights 2010 and 2015, respectively). Poly Point Intersections. *Github*. https://github.com/ideasman42/isect_segments-bentley_ottmann. Retrieved 27 January 2022.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., et al., and Willing, C. (2016). Jupyter Notebooks-a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* pp. 87-90.

Li, M., An, H., Angelovici, R., Bagaza, C., Batushansky, A., Clark, L., Coneva, V., Donoghue, M. J., Edwards, E., Fajardo, D., et al. (2018). Topological data analysis as a morphometric method: using persistent homology to demarcate a leaf morphospace. *Frontiers in Plant Science* 9, 553.

Madzima, T. F., and MacIntosh, G. C. (2021). Equity, diversity, and inclusion efforts in professional societies: intention versus reaction. *Plant Cell* 33, 3189-319.

Martínez, M. C., Loureiro, M. D., and Mantilla, J. L. G. (1995). Importancia y validez de distintos parámetros ampelométricos de hoja adulta para la diferenciación de cepas de Vitis vinifera L., de distintos cultivares. *Invest. Agr. Prod. Prot. Veg*. 9, 377-389.

Martínez, M. C., Boursiquot, J. M., Grenan, S., and Boidron, R. (1997a). *Étude ampelométrique de feuilles adultes de somaclones du cv. Grenache* N (Vitis vinifera L.). *Can. J. Bot.* 75, 333-345.

Martínez, M. C., Grenan, S., Boursiquot, J. M. (1997b). Variabilidad de algunos caracteres ampelográficos y de producción, en somaclones del cultivar Grenache N (Vitis vinifera L.). *Acta Hortic*. 18, 271-280.

Martínez, M., and Grenan, S. S. (1999). A graphic reconstruction method of an average vine leaf. *Agronomie, EDP Sciences* 19, 491-507.

McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*. Vol. 445, pp. 51-56.

Meister, A. L. F. (1769), *Generalia de genesi figurarum planarum et inde pendentibus earum affectionibus*, Nov. Com. Gött. (in Latin) 1, 144.

Montgomery, B. L. (2020). Planting equity: Using what we know to cultivate growth as a plant biology community. *Plant Cell* 32, 3372-3375.

Oliphant, T. E. (2006). A guide to NumPy. USA: Trelgol Publishing.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825-2830.

Ravaz, L. (1902). *Les vignes américaines: Porte-greffes et producteurs directs*. Montpellier and Paris, France: Goulet. Digitized by Google Books from Cornell University. Public domain.

Wald, A. (1943). A Method of Estimating Plane Vulnerability Based on Damage of Survivors. Statistical Research Group, Columbia University. CRC 432 — reprint from July 1980.

Wallis, W. A. (1980). The Statistical Research Group, 1942-1945: Rejoinder. *Journal of the American Statistical Association*. 75 (370), 334–335.

FIGURE LEGENDS

Figure 1: Modeling $\ln(vein\ to\ blade\ ratio)$ as a function of normalized shoot position. The natural log of the ratio of vein-to-blade area, $\ln(vein\ to\ blade\ ratio)$, is modeled as a 2nd degree polynomial of normalized shoot position (where 0 is the shoot tip and 1 is the shoot base). The normalized shoot position value corresponding to the minimum $\ln(vein\ to\ blade\ ratio)$ value is 0.6379505834434538 (magenta vertical line). This value has a 3.22% difference from the value of the inverse of the golden ratio, 0.61803398875 (gold vertical line).

Figure 2: Developmental trajectories of leaf development through Principal Component Analysis (PCA) space by species. A-B) 2nd degree polynomials modeling PC1 (A) and PC2 (B) values by $\ln(vein\ to\ blade\ ratio)$ for three species (*Vitis riparia*, gold; *V. amurensis*, lavender; *Ampelopsis glandulosa* var. *brevipedunculata*, green). **C)** Data points and models plotted as PC2 vs. PC1. Regions corresponding to mature, old leaves and young, small leaves are indicated. **D)** Models for 11 species (indicated by color) and data points colored by $\ln(vein\ to\ blade\ ratio)$ (a white-to-black gradient, scale provided). Percent variance explained by each PC are indicated in axis labels.

Figure 3: Allometric changes in leaf shape during development. Changes in the proportional area different regions of the leaf occupy (middle lobe, lavender blue; vasculature, yellow; distal lobe, light green; distal sinus, dark green; proximal lobe, orange; proximal sinus, magenta) are modeled as developmental trajectories across 100 theoretical time points for three different species: *V. riparia*, *V. amurensis*, and *A. glandulosa* var. *brevipedunculata*. Theoretical grapevine leaves are reconstructed for timepoints 0, 12, 24, 36, 48, 60, 72, 84, and 96 for each species. Note that the trajectories run from mature, large leaves on the left to young, small leaves on the right.

Figure 4: Theoretical grapevine leaves at the limits of what is empirically observed. Principal Component Analysis (PCA) space plotting individual leaves colored by $\ln(vein\ to\ blade\ ratio)$ (a white-to-black gradient, scale provided) and modeled developmental trajectories of different species (indicated by color). Three lines are drawn that encompass the shape of all empirically measured leaves. Theoretical grapevine leaves are reconstructed for five equidistant points along each line (large gray dots), pointed to by lines outside of the plot. Different regions of the leaf are indicated by color: middle lobe, lavender blue; vasculature, yellow; distal lobe, light green; distal sinus, dark green; proximal lobe, orange; proximal sinus, magenta.

Figure 5: Self-intersecting leaves. 100 theoretical leaves, equidistantly spaced along PC1 and PC2, were reconstructed. The range was selected to find the limits of theoretical leaves with any self-intersection in any sub-region of the leaf. For each reconstructed leaf, $\ln(vein\ to\ blade\ ratio)$ was calculated (a white-to-black gradient, scale provided). The boundary that encompasses theoretical leaves with no self-intersection of any region is indicated by a solid, black line. Empirically measured leaves are indicated by points (dodger blue).

Figure 6: Morphospace. 100 reconstructed grapevine leaves across PC1 and PC2. Rounded coordinates for each theoretical leaf are indicated along the axes. The boundary that encompasses theoretical leaves with no self-intersection of any region is indicated by a solid black line. The boundary that encompasses nearly all empirically measured leaves is indicated by a dashed black line. Different regions of the leaf are indicated by color: middle lobe, lavender blue; vasculature, yellow; distal lobe, light green; distal sinus, dark green; proximal lobe, orange; proximal sinus, magenta. Alpha transparency is used so that overlapping regions in leaves can be seen.