Persistent homology and the branching topologies of plants

Mao Li, Keith Duncan, Christopher N. Topp, and Daniel H. Chitwood

And from the turf would leap a branching tree—
Wonders unheard of; for, by Nature, each
Slowly increases from its lawful seed...”
—from Titus Lucretius Carus, “Substance is eternal” in
On the Nature of Things, Book I (translated, in verse,
by W. E. Leonard)

In On the Nature of Things, Lucretius speculates on the necessity of plant development: Branching trees simply do not “leap” from the turf, rather, the branching patterns of shoots and roots develop over time, “slowly increas[ing] from [their] lawful seed” (Leonard, 2015). Over 2000 years ago, the essence of the plant phenotype was written in a poem; plants are four-dimensional beings, branching structures that emerge through time. This qualitative realization of the nature of plant phenotype is self-apparent, but quantitative models of plant morphology are less forthcoming.

Plant morphology should be quantified comprehensively. Conventional analyses of phenotypic traits consider specific plant features and assess only a small proportion of overall morphological variation. There is a critical need for methods to quantify the complete morphology of a plant, including the growing branching structures of both the root and shoot. Plant morphology can be considered across scales; for example, the main trunks of a tree define its coarse architecture, but local branching patterns of twigs distributed throughout the tree also contribute to the overall morphology.

The branching patterns of plants can be considered from the perspective of topology. Topology is a field of mathematics concerned with the connectedness, or contiguousness, of structures. Smaller branches split from larger branches, creating a hierarchy of connectedness appropriate for topological analysis. The ability to quantify and compare the total branching structures of plants across scales has implications for studies of plant genetics, development, evolution, and environmental response, all of which currently rely on traits that measure facets of overall plant morphology.

Persistent homology, a mathematical method that captures topological features across scales, is well suited to quantify the growing branching architectures of plants. We begin by considering existing models and morphometric methods used to quantify and compare plant morphology before introducing persistent homology and its applications.

METRICS AND MODELS OF PLANT MORPHOLOGY

Morphometrics, as the name implies, is concerned with the measurement of shape and size. Morphometric analyses can measure linear features as well as shapes and three-dimensional (3D) structures. Leaves, floral organs, seeds, and cell shapes are some plant structures amenable to shape analysis. Landmarks, homologous points found in every sample (Chitwood et al., 2016), or pseudo-landmarks, equidistant points placed between landmarks (Langlade et al., 2005), are a simple way to represent shape as a multivariate object. Shapes can also be viewed as waves that form a closed contour, to which a Fourier-based decomposition technique, elliptical Fourier descriptors (EFDs), can be applied (Kuhl and Giardina, 1982). Both landmarks and EFDs are multivariate representations of shapes that are descriptive. They can be used to identify the main sources of shape variance or distinguish different groups—such as species, organs, or developmental stages.

Although leaves, flowers, seeds, cells, and other parts of plants can be described as shapes, the overall architecture of a plant is not a shape. Rather, plants—all the shoots and the roots—are branching structures. Many analyses quantifying branching patterns have been applied to plants previously. Leonardo da Vinci described relationships in the diameters and lengths of branch hierarchies in trees (Long, 1994), but this fails to capture branching architecture itself. The parameters underlying branch patterns, and their potential adaptive significance, have been modeled and can be quantified using fractal-based methods (Zeide and Pfeifer, 1991). Fractal methods, however, measure complexity and self-similarity rather than properties of topological spaces. L-systems (Lindenmayer syst-
tems) are recursive systems that expand strings of symbols into larger strings based on a set of rules (Prusinkiewicz and Hanan, 2013). The iterated results of these systems can produce intricate branching patterns, with self-similarity, reminiscent of diverse plant morphologies. However, L-systems are generative models and cannot descriptively measure topological properties.

Although each powerful in their own way, morphometrics, fractal-based methods, branch hierarchies, L-systems, and generative models cannot comprehensively measure the topologies of branching architectures in plants.

**PERSISTENT HOMOLOGY: A METHOD WELL SUITED FOR CAPTURING PLANT TOPOLOGIES**

Persistent homology is a mathematical theory of topology that has much potential if applied to plants. “Homology” in persistent homology is not the same as in biology, that is, features that correspond between organisms based on descent from a common ancestor. Rather, mathematical homology refers to homology groups recording the connectedness of components. For example, $H_0$ (zero order homology) describes path-connected components (that is, contiguous features). A solid cylinder and a branching structure are both a single, connected component. But the details of a branching structure can be revealed by studying how homology persists across the scales of a mathematical function (Edelsbrunner and Harer, 2008; Weinberger, 2011). For example, consider the height, as measured by the vertical distance to the ground, of any point on a tree. We might create a simple function of “height”, and then traverse the structure of the tree, starting at the highest tips and proceeding to the trunk and the ground. As we traverse the tree and the function, starting at the highest points there would be many isolated branches that are not connected. As we proceed through the function, some branches merge. We can record the “birth” and “death” of homology group components (path-connected components, in this example branches) as a persistence barcode (Fig. 1A, B). The $x$-axis of the persistence barcode is the scale of the function (“height”, in this case) and the $y$-axis distinct, connected components (in the jargon of topology, referred to as $H_0$ bars). For example, the “birth” of an $H_0$ bar is due to a new connected component, and the “death” of an $H_0$ bar is because two components merged. When two components merge, the shortest “dies” and the longer “persists”. Each bar in a persistence barcode therefore corresponds to a branch and records where a branch begins and ends with respect to the scales of a function.

Persistence barcodes can be compared against each other as a pairwise distance matrix using a bottleneck distance method, providing a useful tool to compare the similarity of any branching structure to another. Briefly, bottleneck distance calculates the minimal cost to move a branch from one branching structure to resemble another based on

![FIGURE 1 Persistence barcode of the topology of a grape cluster rachis. (A) Lower panel, surface voxels (like pixels, but 3D) of a grape rachis are colored by their geodesic distance (the curved distance along the rachis) to the base. The most distant to closest voxels are colored from red to blue. Panels, left to right, traverse the geodesic distance function, and portions of the rachis are colored gray as the function progresses. Upper panel, a persistence barcode, in which each bar corresponds to a branch. $H_0$ (zero order homology, $y$-axis) branches are “born” and “die” along the distance function ($x$-axis). When two branches merge, the longest persists in the barcode. Vertical lines in the barcode indicate the corresponding position along the geodesic distance function indicated in the panels below, and the number of connected branches corresponds to the number of bars. (B) $H_0$ persistence barcode for a simple branching structure, to demonstrate the relationship between connected components along the scale of a geodesic distance function and the “birth” and “death” of bars in the barcode. (C) Bottleneck distance is a robust metric to compare the overall distance between persistence barcodes. It can be used with traditional statistical techniques used in biology to quantify overall morphological differences between plant structures. Shown are three different branching structures that have been analyzed using principal component analysis (PCA) and the corresponding bottleneck distance between the structures.](image-url)
permuting the persistence barcodes of two structures against each other. The bottleneck distance is a robust metric of similarity between two branching structures (Edelsbrunner and Harer, 2008) that can be used to perform principal component analysis, discriminant analysis, hierarchical clustering, or other statistical methods commonly used in biological studies (Fig. 1C). Persistence barcodes and bottleneck distances can be calculated using software packages written for a variety of programming languages, such as phom (Tausz, 2011), Dionysus (Morozov, 2012), Perseus (Nanda, 2012), PHAT (Bauer et al., 2014), Gudhi (Maria et al., 2014), TDA (Fasy et al., 2014), or javaPlex (Adams et al., 2014). The computational intensiveness of persistent homology methods depends on the complexity and size of the data being analyzed.

It is useful to explain the application of persistent homology to plant morphology using actual examples from plants. Take for example, the branching architecture of the rachis of a grape cluster. If successive two-dimensional (2D) radiograms (Fig. 2A) created by the detection of X-rays that pass through (rather than being absorbed by) the rachis are taken at different angles, a tomographic 3D reconstruction can be computed (X-ray computed tomography [CT]) (Fig. 2B). Now, apply a geodesic distance function to every surface voxel (a voxel is a 3D pixel), measuring the voxel distance to the base. Geodesic distance is calculated as the shortest curved distance of each voxel to the rachis base; it is different from simple “vertical height to the ground” because it records the distance of any point in a structure to its base as if driving along the curves of the structure itself, as if it were a road, to the base (Fig. 1A). If we traverse the geodesic distance function, we start at the most distal termini of the rachis branches, farthest from the base. As we get closer to the base, these termini will fuse with each other, such that where there were two branches there is only one; or, as the distance function is traversed, a new branch will be detected. As the geodesic distance function is traversed across scales, the “birth” and “death” of the connected components (the branches) are recorded as bars in a persistence barcode. Each bar represents an individual connected component. The “birth” and “death” of each bar, in this instance, records the geodesic length of each branch. If two branches fuse, the longest persists, and ultimately only a single bar persists in the barcode. If many different rachises were measured similarly, a pairwise distance matrix between their persistence barcodes, quantifying the overall differences in their topological spaces, could be calculated using the bottleneck distance.

Persistent homology is flexible enough to accommodate more than strict branching topologies. The overall morphology of shoots and roots, including lateral organs like leaves, also possess a topological space. The shoots and roots of a tomato seedling, for example, can be modeled as surface voxels from a 3D X-ray CT scan reconstruction. A number of distance functions can be calculated relative to soil level where the shoot and root meet. Height distance is a vertical straight line from any surface voxel to the soil (Fig. 3A). Geodesic distance, as explained earlier, is the shortest curved path along the seedling to the soil (Fig. 3B). Functions combining different distance functions can measure novel features. For example, the arccos(height distance/geodesic distance) of any voxel is an approximate measure of the angle relative to soil level, which is sensitive to organ bending and branch angles (Fig. 3C). From these two functions (the geodesic function and the height function), persistence barcodes, capturing the respective topological spaces, can be calculated and compared with each other (Fig. 3D, E). Each bar in the barcodes in Fig. 3D and E, for example, corresponds to a branch that arises across the scale of the respective function. Using a bottleneck distance method, a comparison of the topological distance between any two barcodes—any two shoots, any two roots, or a shoot and a root (Fig. 1C)—can quantify branching architecture and phenotypic variation.

### APPLICATIONS OF PERSISTENT HOMOLOGY: MORE THAN JUST BRANCHING

Persistent homology opens new vistas into ways to capture the exquisite features of plants comprehensively. Persistent homology is an adaptable solution that can provide a common framework to interpret innumerable types of phenotypic data. Importantly, persistent homology can be used with any function that scales topological spaces of an object. That persistent homology can be used with functions tailored to specific questions lies at the heart of its versatility.

Consider a set of points—for example, stomata on a leaf, locations of trees imaged by satellite across large swaths of land (Mander et al., 2017), or point cloud data from an agricultural field as measured by lidar (light detection and ranging). Apply to these point data a function increasing the radii of balls around the points and recording the number of connected components as a persistence barcode. As the points with larger and larger circumferences intersect with each other,
they form connected components that have a “birth” and a “death”. The resulting persistence barcode captures the unique topological patternning of the distances of the points to each other.

Shapes, too, can be considered as a collection of 2D points. A density function, measuring the density of nearby pixels for any given pixel can be calculated. Then, thresholds of the spatial distribution and connectedness of different density levels results in a persistence barcode, effectively measuring shapes as a topological space (Li et al., 2017). The density function can be selected to be orientation invariant or robust to disparate shape features, excelling where traditional morphometric methods often fail. Textures can also be analyzed using a persistent homology approach, classifying grass pollen based on surface ornamentation, for example (Mander et al., 2013).

Ultimately, any topological space in plant morphology manifests over time. If plant morphology is simplified to a branching structure, then plants are four-dimensional beings, topologies that grow through time. It is tempting to simply measure the topology of the plants we see before our eyes, on our timescale. Homology groups, through time. It is tempting to simply measure the topology of the plant forms. The versatility of persistent homology to describe diverse topological spaces across scales, and in any number of dimensions, promises to reveal previously unnoticed facets of the plant form and to perhaps bring us closer to their true underlying nature.

**LITERATURE CITED**


**FIGURE 3** Persistent homology applied to the shoot and root architecture of a tomato seedling. (A–C) Colormaps of distance functions of surface voxels to soil level for shoots and roots of a seedling of Solanum lycopersicum cv. M82. (A) A height distance function, which is the vertical distance of each surface voxel to soil level. (B) A geodesic distance function, which is the shortest curved distance of any surface voxel along the surface of the plant to soil level. (C) An angle function, which is arccos(height/geodesic), resulting in the angle of each surface voxel relative to soil level. (D, E) Persistence barcodes for the geodesic distance functions of the (D) shoot and (E) root.


