

Quantifying barley morphology using the Euler Characteristic Transform

Erik J. Amézquita¹, Michelle Y. Quigley², Tim Ophelders⁴, Jacob B. Landis^{5,6},
Daniel Koenig⁶, Daniel H. Chitwood^{1,2}, and Elizabeth Munch^{1,3}

¹*Dept. of Computational Mathematics, Science, & Engineering, Michigan State University*

²*Dept. of Horticulture, Michigan State University*

³*Dept. of Mathematics, Michigan State University*

⁴*Dept. of Mathematics and Computer Science, TU Eindhoven*

⁵*School of Integrative Plant Science, Cornell University*

⁶*Dept. of Botany & Plant Sciences, University of California—Riverside*

Abstract

Observing and documenting shape has fueled biological understanding as the shape of biomolecules, cells, tissues, and organisms arise from the effects of genetics, development, and the environment. The vision of Topological Data Analysis (TDA), that data is shape and shape is data, will be relevant as biology transitions into a data-driven era where meaningful interpretation of large datasets is a limiting factor. We focus first on quantifying the morphology of X-ray CT scans of barley spikes and seeds using topological descriptors based on the Euler Characteristic Transform. We then successfully train a support vector machine to distinguish and classify 28 different varieties of barley based solely on the 3D shape of their grains. This shape characterization will allow us later to link genotype with phenotype, furthering our understanding on how the physical shape is genetically coded in DNA. (This work has been accepted in *inSilico Plants*.)

1 Introduction

Biologists are accustomed to thinking about how the shape of biomolecules, cells, tissues, and organisms arise from the effects of genetics, development, and the environment. Traditionally, biologists use morphometrics to compare and describe shapes. The shape of leaves and fruits is quantified based on homologous landmarks—similar features due to shared ancestry from a common ancestor—or harmonic series from a Fourier decomposition of their closed contour. While these methods are useful for comparing many shapes in nature, they cannot always be used: there may not be homologous points between samples or a harmonic decomposition of a shape is not appropriate. Topological data analysis (TDA) offers a more comprehensive, versatile way to quantify plant morphology. In particular, Euler characteristic curves (ECC) serve as a succinct, computationally feasible topological signature that allows downstream statistical analyses [12]. For example, ECCs have been used to determine a morphospace for all leaves to then predict plant family and location [5]. Further analysis has determined the genetic basis of 2D leaf shape in apple [9], tomato [6], and cranberry [3]. Here, we show the use of the Euler Characteristic to comprehensively describe the shape of 3D voxel-based X-ray CT scans of barley seeds as a proof of concept.

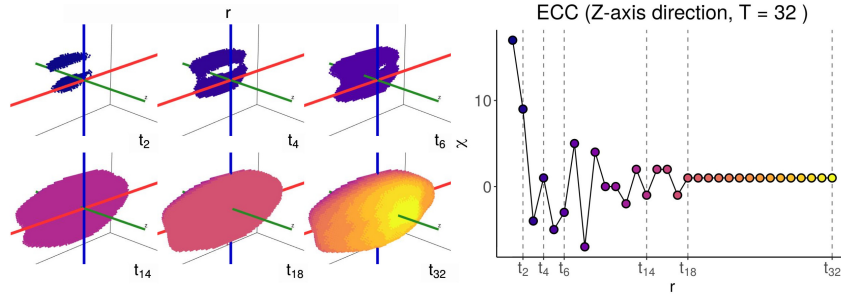


Figure 1: Filtration of a barley seed along the z -axis with 32 thresholds and its corresponding Euler Characteristic Curve.

Table 1: SVM classification accuracy of barley seeds from 28 different accessions after 100 randomized training and testing sets. Classification scores was computed for each accession; the weighted average for each score was taken afterwards, where the weight depended on the number of test seeds used. The use of topological outperforms the use of exclusively traditional descriptors.

Shape descriptors	Dimension reduction	No. of dims	Scores (weighted average \pm standard deviation)		
			Precision	Recall	F_1
Traditional	*	11	0.58 ± 0.050	0.58 ± 0.016	0.57 ± 0.016
Topological	KPCA	2	0.88 ± 0.031	0.87 ± 0.010	0.87 ± 0.011
Topological	UMAP	12	0.75 ± 0.047	0.75 ± 0.016	0.74 ± 0.016
Combined	KPCA	13	0.73 ± 0.052	0.72 ± 0.017	0.71 ± 0.017
Combined	UMAP	23	0.89 ± 0.028	0.89 ± 0.010	0.89 ± 0.010

2 Methods

Consider a cubical complex X of dimension d . For a fixed direction $\nu \in S^{d-1}$, and a height value $h \in \mathbb{R}$, we define

$$X(\nu)_h = \{\Delta \in X : \langle x, \nu \rangle \leq h \text{ for all } x \in \Delta\},$$

to be the subcomplex containing all cubical cells below height h in the direction ν . The Euler characteristic at height h is $\chi(X(\nu)_h)$, the alternating sum of counts of cells in the subcomplex $X(\nu)_h$. The Euler Characteristic Curve (ECC) of direction ν is defined as $\{\chi(X(\nu)_h)\}_{h \in \mathbb{R}}$, exemplified in Figure 1. The Euler Characteristic Transform (ECT) is defined as the collection of all ECCs corresponding to all possible directions. To be more precise, the ECT of complex X is defined as the function

$$\begin{aligned} ECT(X) : S^{d-1} &\rightarrow \mathbb{Z}^{\mathbb{R}} \\ \nu &\mapsto \{\chi(X(\nu)_h)\}_{h \in \mathbb{R}}. \end{aligned}$$

We focus on the shape of barley seeds from a collection of 28 different barley accessions from diverse regions across the Eurasian continent. Using X-ray CT—computed tomography—scanning technology, we have created voxel-based 3D reconstructions of over 875 spikes, from which we have isolated 3121 parental seeds. Since the seeds are oblong in shape, we aligned them according to their three main principal components.

On one hand, we computed 11 traditional quantifiable shape descriptors for each seed, such as length, width, and volume. On the other hand, we computed topological shape descriptors using the ECT. For topological purposes, we treated each voxel-based image as a dual cubical complex where each nonzero voxel is treated as a vertex [13].

We favor the use of the ECT for two reasons. First, the ECT is computationally inexpensive; since it is based on successive alternating sums of counts of cells, a single ECC

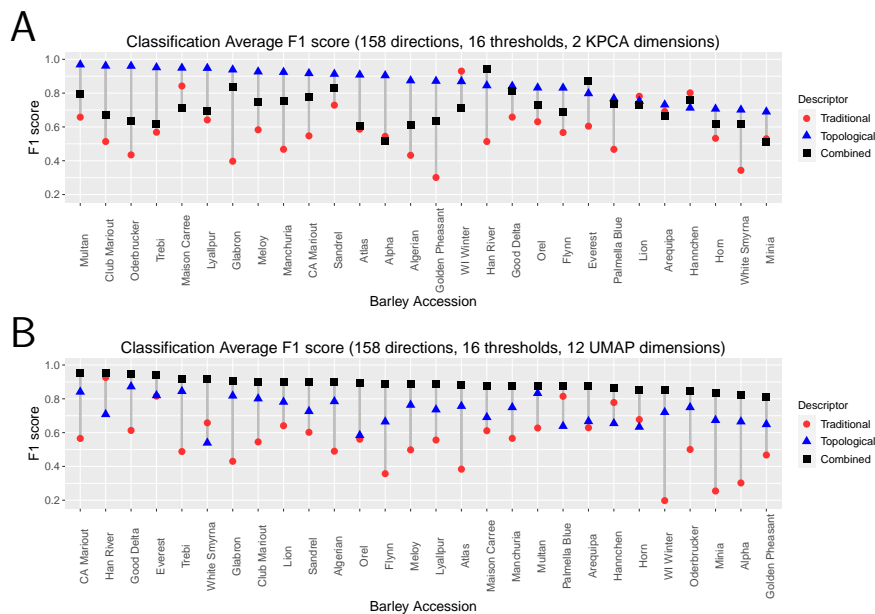


Figure 2: SVM classification results for individual barley accessions. A. Results when using a KPCA 2-dimension reduced topological vector. Accessions ordered according to their classification accuracy determined by the topological shape descriptors. **B.** Results when using a UMAP 12-dimension reduced topological vector. Accessions ordered according to their classification accuracy determined by the combined shape descriptors.

can be computed in linear time with respect to the number of voxels in the image [10]. Second, the ECT effectively summarizes all the morphological features of any 3D complex as it encodes sufficient information to reconstruct the initial complex [12], a result later extended to the n -dimensional case [2]. Nonetheless, the idea of efficiently reconstructing an arbitrary 3D object solely from its ECT remains elusive [1, 4, 8].

In our case, we used 158 different directions with 16 uniformly spaced thresholds. We emphasized directions toward the crease of the seeds. This yielded a 2528-dimensional vector for every seed. These high-dimensional vectors were later reduced in dimension separately using non-linear KPCA [11] and UMAP [7].

The descriptiveness of both traditional and topological measures was tested by training five non-linear support vector machines (SVM). These characterized and predicted the seeds from the 28 distinct barley accessions based on three different collection of descriptors: traditional, topological, and combined. We also varied the dimension-reduction method. In every case, all the descriptors were centered and scaled to variance 1 prior to classification. Given that SVM is a supervised learning method, we first randomly sampled 75% of the seeds from every founder as our training data set. The remaining 25% was used to test the accuracy of our prediction model. This setup was repeated 100 times. Average scores were considered for the overall data set (Table 1), as well as for individual accessions (Fig. 2).

3 Results and Conclusions

The majority of the barley accessions studied are more easily distinguished with the topological lens but not with traditional measures, with few exceptions (Fig. 2). Exceptions like Hannchen, Han River and Palmella Blue have slightly distinctive traditional trait distributions, so seed size does matter and it is important to take it into account. At the same time,

we observe accessions like Alpha, Glabron, Minia, and Wisconsin Winter, that are poorly differentiated with traditional information but report considerably higher classification accuracies whenever using topological information. When looking at a more robust dimension reduction technique like UMAP, classification accuracy is increased when combined with size-related information.

The Euler characteristic is a simple yet powerful way to reveal features not readily visible to the naked eye. There is “hidden” morphological information that traditional and geometric morphometric methods are missing. The Euler characteristic, and Topological Data Analysis in general, can be readily computed from any given image data, which makes it a versatile tool to use in a vast number of biology-related applications. TDA provides a comprehensive framework to detect and compare morphological nuances, nuances that traditional measures fail to capture and that remain unexplored using simple geometric methods. In the specific case of barley seeds presented here, these “hidden” shape nuances provides enough information to not only characterize specific accessions, but the individual spikes from which seeds are derived. Our results suggest a new exciting path, driven by morphological information alone, to explore further the phenotype-genotype relationship.

4 Literature cited

- [1] **Betthausen LM.** *Topological reconstruction of grayscale images.* PhD thesis, University of Florida, Gainesville, Florida (2018).
- [2] **Curry J, Mukherjee S, Turner K** (2018). How many directions determine a shape and other sufficiency results for two topological transforms. [arXiv:1805.09782](https://arxiv.org/abs/1805.09782).
- [3] **Diaz-García L, Covarrubias-Pazaran G, Schlautman B, Grygleski E, Zalapa J** (2018). Image-based phenotyping for identification of QTL determining fruit shape and size in american cranberry (*Vaccinium macrocarpon L.*). *PeerJ* **6**. [doi:10.7717/peerj.5461](https://doi.org/10.7717/peerj.5461).
- [4] **Fasy BT, Micka S, Millman DL, Schenfisch A, Williams L** (2019). The first algorithm for reconstructing simplicial complexes of arbitrary dimension from persistence diagrams. [arXiv:1912.12759](https://arxiv.org/abs/1912.12759).
- [5] **Li M, An H, Angelovici R, Bagaza C, Batushansky A, Clark L, Coneva V, Donoghue MJ, 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. [doi:10.3389/fpls.2018.00553](https://doi.org/10.3389/fpls.2018.00553).
- [6] **Li M, Frank MH, Coneva V, Mio W, Chitwood DH, Topp CN** (2018). The persistent homology mathematical framework provides enhanced genotype-to-phenotype associations for plant morphology. *Plant Physiology* **177**: 1382–1395. [doi:10.1104/pp.18.00104](https://doi.org/10.1104/pp.18.00104).
- [7] **McInnes L, Healy J, Melville J** (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. [arXiv:1802.03426](https://arxiv.org/abs/1802.03426).
- [8] **Micka SA.** *Searching and Reconstruction: Algorithms with Topological Descriptors.* PhD thesis, Montana State University, Bozeman, Montana (2020).
- [9] **Migicovsky Z, Li M, Chitwood DH, Myles S** (2018). Morphometrics reveals complex and heritable apple leaf shapes. *Frontiers in Plant Science* **8**: 2185. [doi:10.3389/fpls.2017.02185](https://doi.org/10.3389/fpls.2017.02185).
- [10] **Richardson E, Werman M** (2014). Efficient classification using the Euler characteristic. *Pattern Recognition Letters* **49**: 99 – 106. [doi:10.1016/j.patrec.2014.07.001](https://doi.org/10.1016/j.patrec.2014.07.001).
- [11] **Schölkopf B, Smola A, Müller KR** (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* **10**: 1299–1319. [doi:10.1162/089976698300017467](https://doi.org/10.1162/089976698300017467).
- [12] **Turner K, Mukherjee S, Boyer DM** (2014). Persistent homology transform for modeling shapes and surfaces. *Information and Inference* **3**: 310–344. [doi:10.1093/imaiai/iau011](https://doi.org/10.1093/imaiai/iau011).
- [13] **Wagner H, Chen C, Vućini E,** Wagner, Hubert and Chen, Chao and Vućini, Erald. Efficient computation of persistent homology for cubical data. In Peikert R, Hauser H, Carr H, Fuchs R, editors, *Topological Methods in Data Analysis and Visualization II: Theory, Algorithms, and Applications*, pages 91–106. publisher. [doi:10.1007/978-3-642-23175-9_7](https://doi.org/10.1007/978-3-642-23175-9_7).