archiDART v3.0
Delory, Benjamin; Li, Mao; Topp, Christopher; Lobet, Guillaume

Citation for pulished version (APA):
Delory, B., Li, M., Topp, C., & Lobet, G. (2018). archiDART v3.0: A new data analysis pipeline allowing the topological analysis of plant root systems. *Faculty of 1000 Research*, *7*(22), Article 22.
https://doi.org/10.12688/f1000research.13541.1

SOFTWARE TOOL ARTICLE

archiDART v3.0: A new data analysis pipeline allowing the topological analysis of plant root systems [version 1; referees: 2 approved, 1 approved with reservations]

Benjamin M. Delory [iD] [1], Mao Li[2], Christopher N. Topp[2], Guillaume Lobet[3,4]

[1]Ecosystem Functioning and Services, Institute of Ecology, Leuphana University, Lüneburg, 21335, Germany
[2]Donald Danforth Plant Science Center, St. Louis, MO, 63132, USA
[3]Agrosphäre (IBG-3), Forschungszentrum Jülich GmbH, Jülich, 52428, Germany
[4]Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, 1348, Belgium

Abstract

Quantifying plant morphology is a very challenging task that requires methods able to capture the geometry and topology of plant organs at various spatial scales. Recently, the use of persistent homology as a mathematical framework to quantify plant morphology has been successfully demonstrated for leaves, shoots, and root systems. In this paper, we present a new data analysis pipeline implemented in the R package archiDART to analyse root system architectures using persistent homology. In addition, we also show that both geometric and topological descriptors are necessary to accurately compare root systems and assess their natural complexity.

Keywords

archiDART, plant root systems, topology, persistent homology, Fitter indices, Data Analysis of Root Tracings (DART), Root System Markup Language (RSML)

GODAN This article is included in the Global Open Data for Agriculture and Nutrition gateway.

This article is included in the RPackage gateway.
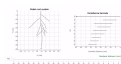
This article is included in the Interactive Figures collection.

Open Peer Review

Referee Status: ✓ ? ✓

	Invited Referees		
	1	**2**	**3**
version 1 published 08 Jan 2018	✓ report	? report	✓ report

1 **Magdalena M. Julkowska** [iD] , King Abdullah University of Science and Technology, Saudi Arabia

2 **Alexander Bucksch** [iD] , University of Georgia, USA

3 **Nathan D. Miller**, University of Wisconsin-Madison, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Benjamin M. Delory (benjamin.delory@leuphana.de)

Author roles: Delory BM: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Li M**: Methodology, Resources, Writing – Review & Editing; **Topp CN**: Methodology, Resources, Writing – Review & Editing; **Lobet G**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

How to cite this article: Delory BM, Li M, Topp CN and Lobet G. **archiDART v3.0: A new data analysis pipeline allowing the topological analysis of plant root systems [version 1; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2018, **7**:22 (https://doi.org/10.12688/f1000research.13541.1)

First published: 08 Jan 2018, **7**:22 (https://doi.org/10.12688/f1000research.13541.1)

Introduction

The quantification of plant root systems is central in many research areas, ranging from developmental studies[1] to crop phenotyping[2]. Root systems are often characterised using geometric descriptors, such as the total root length[3], total root surface[4], or number of root tips[5]. However, such descriptors often fail to describe the full complexity of root systems. Additional descriptors able to describe the topology of root systems are often missing, despite the fact that they can help understanding the feedback between plant morphology and root system functions[6,7].

The topology of a root system is an important component of its architecture and refers to how individual roots are connected to each other through branching[8]. Studying the topology of branching structures as complex as root systems is challenging and requires quantitative methods allowing the description and comparison of plant morphologies[7]. In the 1980's, A.H. Fitter introduced a method to describe branching structures and classify root systems into topologically distinct networks[9]. His method relies on the calculation of three indices to describe the topology of a root system, namely the magnitude, altitude, and external path length. A detailed description of this method can be found in 9,10. More recently, the use of persistent homology to quantify plant morphologies was introduced in the plant sciences community. This method was successfully used to quantify leaf shapes, leaf serrations, and root system architectures[11]. Persistent homology is a mathematical framework allowing the quantification of plant morphologies at different scales (from organs to organisms). Because plant roots can be represented as a succession of nodes connected by straight lines in a tree graph, they are referred to as zero-order homology groups (H_0, path-connected component) in mathematics. The goal of a persistent homology analysis applied to a root system is to study how H_0 features persist across the scales of a continuous mathematical function. A common mathematical function used to capture the topology of branching structures, such as plant shoots and root systems, is the geodesic distance (i.e., the distance measured along the roots between the root system base and any point of the root system). A nice explanation of how persistent homology can be applied to capture plant topologies is provided in 12. The main output of a persistent homology analysis is a persistence barcode recording the birth (apparition of a new connected component) and death (fusion of two connected components) of each H_0 branch when a distance function traverses the branching structure (Figure 1). The degree of similarity between different root system topologies can be assessed by computing a pairwise distance matrix using a bottleneck distance method to compare persistence barcodes. Multivariate statistical tools, such as multidimensional scaling, can then be used to visualize topological differences between root systems[12].

In the past decade, many tools were developed to analyse root systems from digital images (for an extensive list, see the plant-image-analysis.org website[13]) or model root system architectures[14,15]. Several of these tools are able to extract the full root system architecture from the images, including the topology[16–19]. A common format for the storage of root architecture data, the Root System Markup Language (RSML)[20], was also created to facilitate the exchange of information between researchers.

Building on this new format, several tools were created to analyse root architecture data[20,21]. Among these tools, the R package archiDART offers a wide range of functionalities to analyse root system architectures in a free, open-source, and popular data analysis environment[21].

In this paper, we present a new version (v.3.0) of the R package archiDART. In comparison with the version described earlier[21], this version now includes several topological analysis methods, including, but not limited to, persistent homology. Our main objective is to demonstrate how the functions of the archiDART package can be used to analyse and compare the topology of plant root systems using persistent homology. In addition, we also aim to show that the topological analysis of plant root systems is highly complementary to the more classical approach that uses a set of geometric descriptors to compare root systems.

Methods

Implementation

archiDART is an R package developed for the automated analysis of plant root system architectures using Data Analysis of Root Tracings (DART)[17] and Root System Markup Language files (RSML)[20]. The version 3.0 of archiDART can be downloaded from the CRAN repository. An overview of the functions available in the package is presented in Table 1. Among the 10 functions developed for the package, 5 were already presented elsewhere[21] and will not be further discussed in this paper.

In comparison with the version presented earlier, the version 3.0 of the package supports the analysis of 3D root systems. In addition, time series data in RSML files can be analysed if the root system age is stored as a continuous function along the root segments. Finally, we developed a set of 5 new functions and updated the architect function to allow the topological
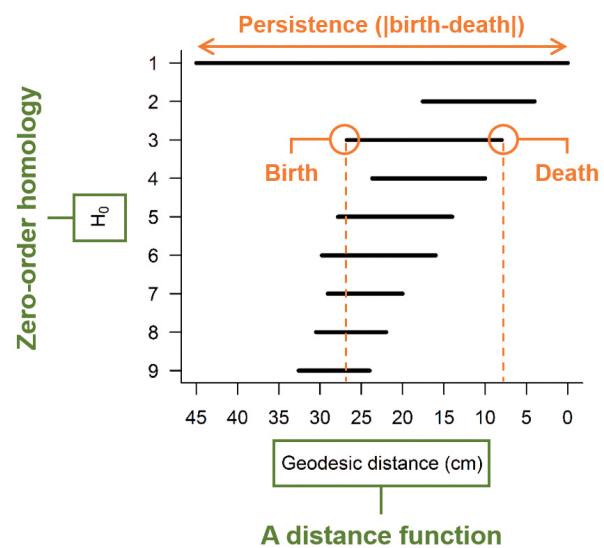

Figure 1. Persistence barcode capturing the topology of a plant root system. When two H_0 branches merge, the longest one persists and the shortest one dies.

Table 1. Summary of the functions available in the archiDART package version 3.0. The italicized functions were already presented in 21. DART, Data Analysis of Root Tracings; RSML, Root System Markup Language.

R functions	Description	Input data	Returned R objects	Version
architect	Computing traits describing the global root system architecture	DART and RSML files dartToTable and rsmlToTable objects	Data frame	1.0
archidraw	Plotting vectorized root systems	DART and RSML files	Plot window	1.0
archigrow	Computing growth rates and plotting vectorized root systems	DART and RSML files	List and plot window	1.0
latdist	Computing lateral root length and density distribution	DART and RSML files	List	1.0
trajectory	Computing root growth directions and trajectories	DART and RSML files	List and plot window	1.0
rsmlToTable	Import RSML files into a single data frame	RSML files	rsmlToTable object (data frame)	3.0
dartToTable	Import DART files into a single data frame	DART files	dartToTable object (data frame)	3.0
perhomology	Topological analysis using persistent homology	dartToTable and rsmlToTable objects	perhomology object (list). Each element of the list is a barcode object (matrix).	3.0
plot.barcode	Plot the persistence barcode (S3 method)	barcode object	Plot window	3.0
bottleneckdist	Computing a pairwise bottleneck distance matrix	perhomology object	Matrix	3.0

analysis of plant root systems. The architect function is now able to calculate the topological indices introduced by Fitter[10], and the 5 new functions presented in this paper are devoted to the topological analysis of root systems using persistent homology[12].

Operation
All functions of archiDART were coded using the R programming language. The package is compatible with Windows, Mac OS X, and major Linux operating systems. A detailed documentation file listing the package dependencies and describing all the functions listed in Table 1 can be downloaded from the CRAN package area. The bottleneckdist function of archiDART relies on the bottleneck function of the TDA package[22] to compute the bottleneck distance between two persistence diagrams.

Root system library
The root system library used in this paper has already been presented elsewhere[23]. Briefly, this library contains a total of 10,464 simulated root systems created using the root architecture model ArchiSimple[14]. The result of each simulation was stored as an RSML file. The library consists of two categories of root systems: tap-rooted (5212) and fibrous (5252). For the use cases presented in this paper, 50 tap-rooted and 50 fibrous root systems were selected from the RSML library. All root systems used in this paper had a total root length comprised between 17 and 23 m (20 m ± 15%). Summary statistics describing the root system library used in this study are presented in Table S1.

archiShiny: A web application demonstrating the capabilities of archiDART
In order to demonstrate and illustrate the capabilities of archiDART, we developed a web application (archiShiny) using the Shiny library[24]. This application is freely available here: https://plantmodelling.shinyapps.io/archidart. We developed archiShiny with the following aims in mind: (1) demonstrating how multivariate statistical tools (such as principal component analysis) can be used on the aggregated metrics computed by the architect function to differentiate root systems; (2) showing how root systems can be plotted using the advanced graphical functions of the ggplot2 library[25]; and (3) comparing the topology of root systems using persistent homology. The web application uses a library of 70 RSML files created using the root architecture model ArchiSimple[14]. Based on the initial values of the parameters of the model, the root systems were classified into seven genotypes (mock, dense, sparse, steep, shallow, slow, and fast). Each genotype was represented by 10 simulations. The different genotypes were based on a standard parameter set (mock) and had one parameter changed: growth rate (slow vs. fast), inter-lateral distance (dense vs sparse) or gravitropism (steep vs shallow).

Use cases
After package installation, the topological analysis of plant root systems (RSML files) using persistent homology comprises four main steps: (1) creating an rsmlToTable object (2), computing persistence barcodes, (3) computing a pairwise bottleneck distance matrix, and (4) visualizing topological differences between

root systems using non-metric multidimensional scaling (NMDS). The main steps of the analysis performed in this section of the paper are summarized in Figure 2. Although we only present the analysis pipeline developed for RSML files, root systems vectorized with DART can be analysed using exactly the same approach (see Figure 2).

Creating an rsmlToTable object

The first step of the analysis is to import the RSML files into R with the rsmlToTable function. If root systems were vectorized with DART, the dartToTable function should be used instead. The rsmlToTable function creates a data frame (table) containing at least 23 columns (spatial coordinates, length, diameter, surface, volume, growth rate, orientation, geodesic distance, etc.) and as many lines as root segments. Here, a root segment is defined as the straight line between two nodes in the data file. The table is an rsmlToTable object that can directly be used as an input to compute the persistence barcodes using the perhomology function. It is worth noting that rsmlToTable objects can also be used as an input for the architect function of this new version of the package to compute a set of aggregated metrics describing the global architecture of plant root systems.

Computing the persistence barcodes

The perhomology function computes the persistence barcode of each root system stored in an rsmlToTable or a dartToTable object. Each persistence barcode is computed using a geodesic distance

function (Figure 3). For each root system, the results are stored as a barcode object in a list that contains as many elements as root systems. A barcode object is a matrix with 3 columns (dimension, birth, and death) and has as many lines as zero-order homology bars in the persistence barcode. An S3 method (plot.barcode) was developed for plotting persistence barcodes. A code example to compute and plot persistence barcodes from RSML files is provided below.

```
path <- "PATH_TO_FOLDER_WITH_RSML_FILES"
table <- rsmlToTable(path, fitter=TRUE)
ph <- perhomology(table)
plot(ph$RSML_NAME)
```

Computing a pairwise bottleneck distance matrix

To compare persistence barcodes against each other, a pairwise distance matrix is needed and the bottleneck distance is one possible option. The bottleneck distance is considered as a robust dissimilarity metric between two persistence barcodes, and its interpretation is quite straightforward: the greater the distance between two persistence barcodes, the greater will be the dissimilarity between them[12]. Such pairwise bottleneck distance matrix can be calculated with the bottleneckdist function of the package. This function only requires a perhomology object as an input. It has to be noted that the computation time required to compute a bottleneck distance matrix is highly dependent on the number and
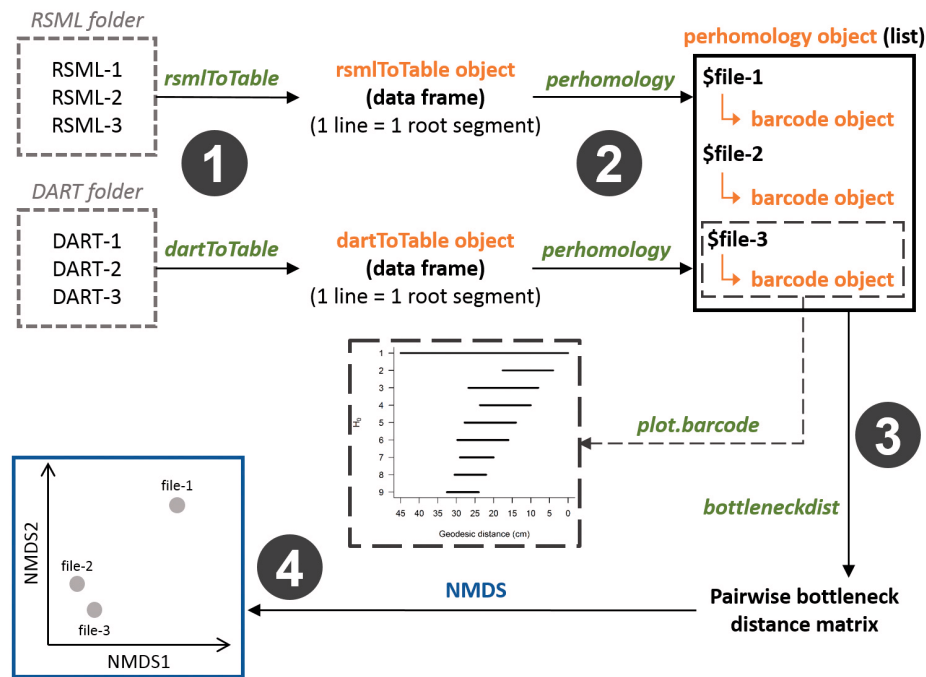

Figure 2. Analysis pipeline used in archiDART to compare the topology of plant root systems using persistent homology. The archiDART functions are italicized and written in green. archiDART objects are written in orange. DART, Data Analysis of Root Tracings; RSML, Root System Markup Language; NMDS, non-metric multidimensional scaling.
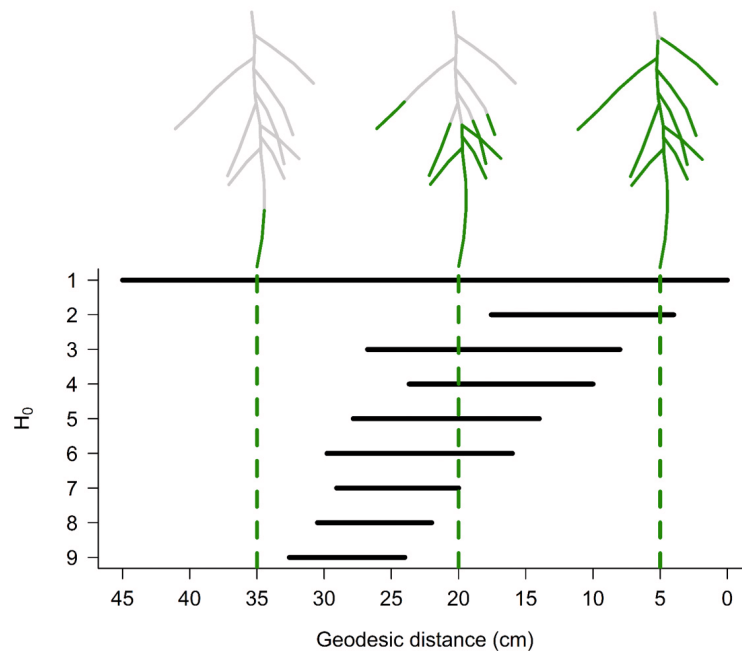
Figure 3. Persistence barcode of the topology of a plant root system computed using a geodesic distance function. Vertical lines indicate the position along the geodesic distance function (from left to right). The version of this figure in the online article is interactive and was produced with the plotly library[26].

complexity of root systems being compared. A code example to compute a bottleneck distance matrix from persistence barcodes is provided below.

```
path <- "PATH_TO_FOLDER_WITH_RSML_FILES"
table <- rsmlToTable(path, fitter=TRUE)
ph <- perhomology(table)
dist <- bottleneckdist(ph)
```

Persistent homology: An efficient method allowing the topological analysis of plant root systems

A large variety of morphological, architectural, and topological traits can be measured on plant root systems (e.g., total root length, diameter, number of lateral roots per branching order, lateral root density, Fitter indices, etc.). When working with root architecture models and image analysis tools supporting the RSML format, such traits can be easily extracted from RSML files using the architect function of the archiDART package[21] or the ImageJ plugin RSML Reader[20]. Using multivariate statistical tools, such as principal component analysis (PCA), one can then determine the key traits differentiating the root systems being compared[23]. In the next section, we would like to show that the information gained with this approach can be nicely complemented by a topological analysis of root systems using persistent homology.

After selecting 100 root systems from a large RSML library, we first used the architect function to compute a set of 20 traits for each root system (Table S1). Then, we performed a PCA to visualize

differences between root systems and find the most interesting morphological, architectural, and topological variables to differentiate them. On the score plot constructed with the two first principal components, a good separation between fibrous and taproot root systems can be observed on the first principal component (Figure 4A). On average, fibrous root systems were characterized by a greater number/length of first-order roots, while taproot systems had a greater lateral root length and a greater secondary root density (Figure 4B, Table S2). Interestingly, two topological indices (altitude and external path length) were on average greater for taproot systems. On the second principal component, however, a separation between dicotyledonous root systems can be observed (Figure 4A). Negative PC2 scores were mainly associated with taproot systems having a greater number/length of tertiary roots and a greater magnitude, while root systems with positive PC2 scores had on average greater root diameters, surface, and volume (Figure 4B, Table S2). Although this approach is very useful to assess root system diversity and derive a functional classification of root systems[27], it poorly takes into account topological differences that might exist between root systems sharing similar trait values.

To illustrate this, we plotted four representative root systems from the RSML library used in this study (Figure 5). Although the global architecture and topology of these root systems clearly differ, fibrous 1 and fibrous 2, as well as taproot 1 and taproot 2, were poorly separated by the PCA (Figure 4A). Therefore, we performed a topological analysis of the root systems in our library using the persistent homology analysis pipeline described
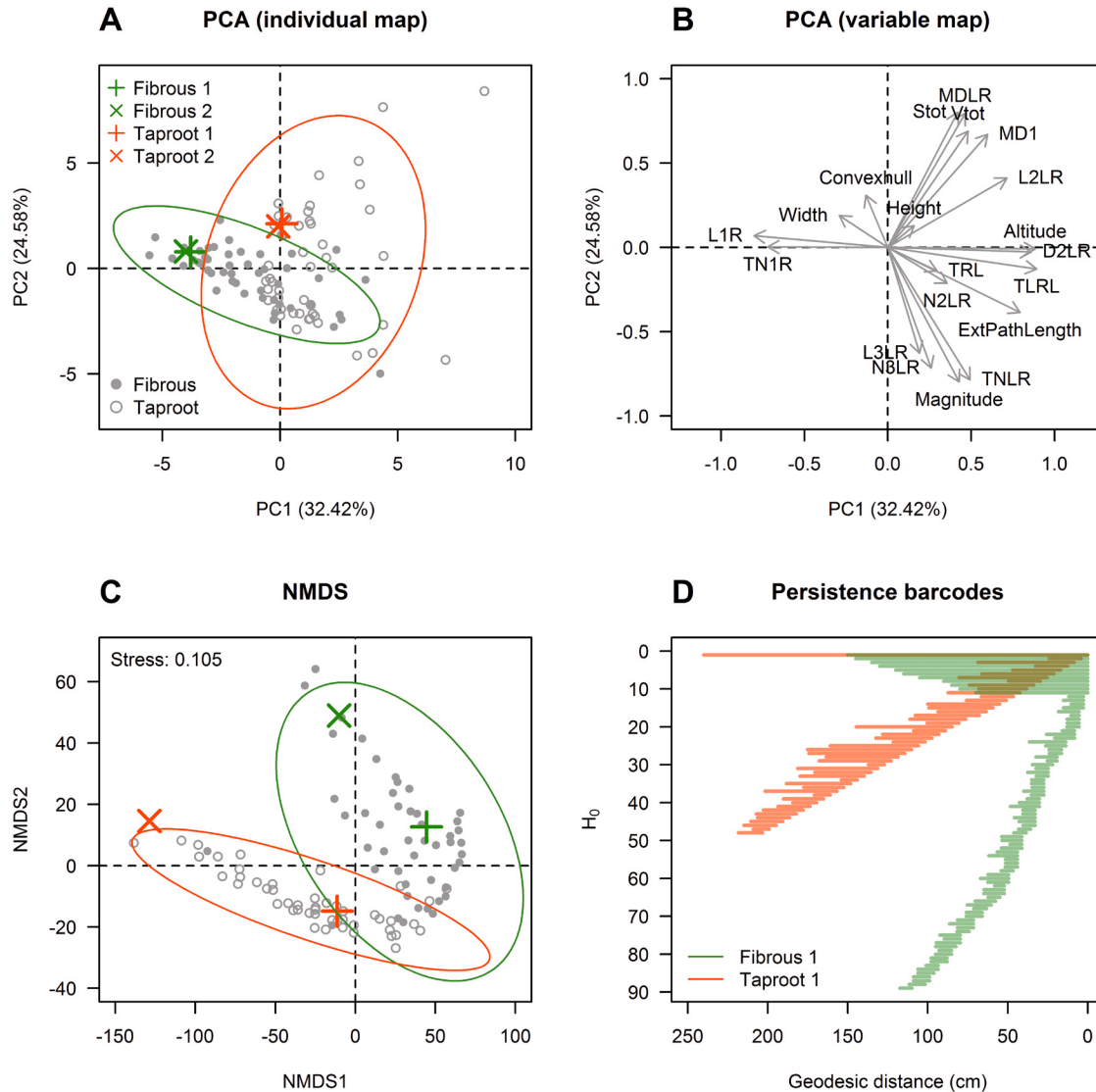
Figure 4. Two complementary approaches for comparing root system architectures. In total, 100 root systems were considered for the analysis (50 fibrous and 50 tap-rooted). In the first approach, root systems were compared using a set of 20 traits computed by the architect function of archiDART. A PCA was then performed to visualize differences between root systems and find the most interesting traits to differentiate them (panels **A** and **B**). The PCA was performed on a correlation matrix constructed from scaled variables using the PCA function of the FactoMineR package[28]. In the second approach, we used persistent homology to compare the topology of root systems. Topological differences between root systems were visualized using non-metric multidimensional scaling (NMDS, panel **C**). The NMDS was performed on a pairwise bottleneck distance matrix with the metaMDS function of the vegan library[29]. In panel **D**, two persistence barcodes are compared. In panels **A** and **C**, each dot is a branching structure and four root system of interests are spotted using orange (taproot) and green (fibrous) crosses. Abbreviations used in panel **B**: TRL, total root length; L1R, total first-order root length; TN1R, number of first-order roots; TNLR, total number of lateral roots; TLRL, total lateral root length; N2LR, number of second-order roots; N3LR, number of third-order roots; L2LR, total second-order root length; L3LR, total third-order root length; MD1, mean first-order root diameter; MDLR, mean lateral root diameter; D2LR, second-order root density; Convexhull, convex hull area; Stot, total root surface area; Vtot, root system volume; ExtPathLength, external path length.
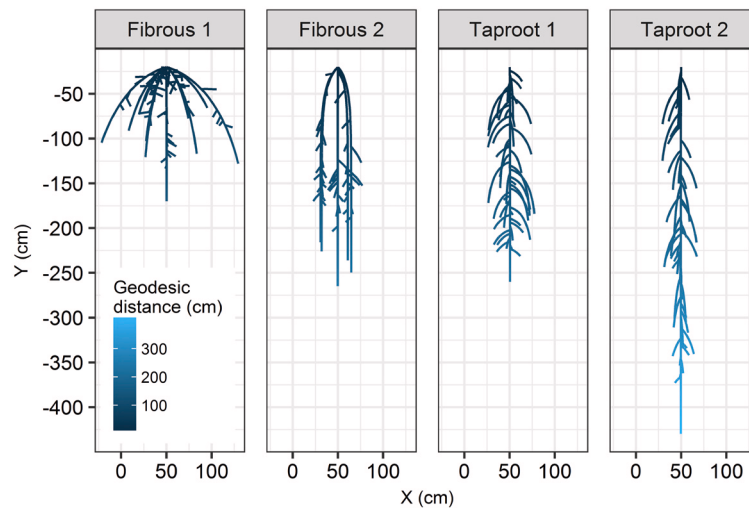
Figure 5. Representative root systems highlighted in Figure 4 (colour crosses). The colour code refers to the geodesic distance (cm).

previously. Non-metric multidimensional scaling (NMDS) was used to visualize dissimilarities between persistence barcodes (Figure 4C and D). Results showed that (1) fibrous and taproot root systems can be clearly separated using persistent homology, and (2) strong topological differences exist between fibrous 1 and fibrous 2, as well as between taproot 1 and taproot 2, despite the fact that these root systems were not separated by the PCA. Altogether, these results showed that persistent homology is highly complementary to the more traditional approach consisting at using a set of aggregated metrics to compare root systems.

Conclusions

In this paper, we presented a new analysis pipeline implemented in the R package archiDART to perform topological analysis of plant root systems using root architectural data (DART or RSML files). Using root architecture models, we showed that persistent homology is an efficient tool to capture and compare the topology of a large diversity of root systems. In addition, our results showed that the use of both geometric and topological descriptors are necessary to capture the natural complexity of plant root systems. Because topology is independent of transformation and deformation, the analysis pipeline described in this paper is highly flexible and can be used on data describing the architecture of 3D (e.g., root architecture models) and 2D (e.g., excavated root systems) root systems. Altogether, we believe that this great flexibility in root architecture data, the ease of use of the functions developed for the analysis pipeline presented in this paper, as well as the open-source nature of archiDART, make topological analysis of root systems widely accessible to the scientific community.

Data and software availability

The data and R codes used for the use cases presented in this manuscript are available: https://doi.org/10.5281/zenodo.1117836[30]

archiDART

The latest stable version of archiDART is available on the CRAN repository: http://cran.r-project.org/package=archiDART

Source code available from: https://github.com/archidart/archidart

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.1117864[31]

Software license: GNU GPL v2.0

More information about archiDART can be found on this website: https://archidart.github.io/

Web application

The web application illustrating the capabilities of archiDART is accessible at: https://plantmodelling.shinyapps.io/archidart

The data and codes used to make the web application are available: https://github.com/archidart/archishiny

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.1133435[32]

License for web application: GNU GPL v3.0

Supplementary material

Table S1. Descriptive statistics of the root system library used in this study. The RSML library consisted of 50 fibrous and 50 taproot root systems created using the ArchiSimple model[14].

Click here to access the data.

Table S2. Principal component analysis: Correlation between each root system variable and the two first principal components. Correlation coefficients written in bold contributed significantly to a principal component (PC). We considered that a variable contributed significantly to a PC if its contribution (in %) was greater than the contribution that would have been expected if all variables contributed equally to a PC (threshold value equal to 5%). The topological indices used in the PCA were calculated following[9,10].

Click here to access the data.

References

1. Julkowska MM, Hoefsloot HC, Mol S, *et al.*: **Capturing Arabidopsis root architecture dynamics with ROOT-FIT reveals diversity in responses to salinity.** *Plant Physiol.* 2014; **166**(3): 1387–402.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Das A, Schneider H, Burridge J, *et al.*: **Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics.** *Plant Methods.* 2015; **11**: 51.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Delaplace P, Delory BM, Baudson C, *et al.*: **Influence of rhizobacterial volatiles on the root system architecture and the production and allocation of biomass in the model grass *Brachypodium distachyon* (L.) P. Beauv.** *BMC Plant Biol.* 2015; **15**: 195.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Fang S, Clark RT, Zheng Y, *et al.*: **Genotypic recognition and spatial responses by rice roots.** *Proc Natl Acad Sci U S A.* 2013; **110**(7): 2670–5.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Pound MP, Atkinson JA, Townsend AJ, *et al.*: **Deep machine learning provides state-of-the-art performance in image-based plant phenotyping.** *Gigascience.* 2017; **6**(10): 1–10.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Bucksch A, Atta-Boateng A, Azihou AF, *et al.*: **Morphological Plant Modeling: Unleashing Geometric and Topological Potential within the Plant Sciences.** *Front Plant Sci.* 2017; **8**: 900.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Balduzzi M, Binder BM, Bucksch A, *et al.*: **Reshaping Plant Biology: Qualitative and Quantitative Descriptors for Plant Morphology.** *Front Plant Sci.* 2017; **8**: 117.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Lynch J: **Root Architecture and Plant Productivity.** *Plant Physiol.* 1995; **109**(1): 7–13.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Fitter AH: **The topology and geometry of plant root systems: influence of watering rate on root system topology in *Trifolium pratense*.** *Ann Bot.* 1986; **58**(1): 91–101.
   Publisher Full Text

10. Fitter AH: **An architectural approach to the comparative ecology of plant root systems.** *New Phytol.* 1987; **106**(s1): 61–77.
    Publisher Full Text

11. Li M, Frank MH, Coneva V, *et al.*: **Persistent homology: a tool to universally measure plant morphologies across organs and scales.** *bioRxiv.* 2017.
    Publisher Full Text

12. Li M, Duncan K, Topp CN, *et al.*: **Persistent homology and the branching topologies of plants.** *Am J Bot.* 2017; **104**(3): 349–353.
    PubMed Abstract | **Publisher Full Text**

13. Lobet G, Draye X, Périlleux C: **An online database for plant image analysis software tools.** *Plant Methods.* 2013; **9**(1): 38.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Pagès L, Bécel C, Boukcim H, *et al.*: **Calibration and evaluation of ArchiSimple, a simple model of root system architecture.** *Ecol Modell.* 2014; **290**: 76–84.
    Publisher Full Text

15. Schnepf A, Leitner D, Landl M, *et al.*: **CRootBox: A structural-functional modelling framework for root systems.** *bioRxiv.* 2017.
    Publisher Full Text

16. Pound MP, French AP, Atkinson JA, *et al.*: **RootNav: navigating images of complex root architectures.** *Plant Physiol.* 2013; **162**(4): 1802–14.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Le Bot J, Serra V, Fabre J, *et al.*: **DART: a software to analyse root system architecture and development from captured images.** *Plant Soil.* 2010; **326**(1–2): 261–73.
    Publisher Full Text

18. Lobet G, Pagès L, Draye X: **A novel image-analysis toolbox enabling quantitative analysis of root system architecture.** *Plant Physiol.* 2011; **157**: 29–39.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Leitner D, Felderer B, Vontobel P, *et al.*: **Recovering root system traits using image analysis exemplified by two-dimensional neutron radiography images of lupine.** *Plant Physiol.* 2014; **164**: 24–35.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Lobet G, Pound MP, Diener J, *et al.*: **Root system markup language: toward a unified root architecture description language.** *Plant Physiol.* 2015; **167**: 617–27.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Delory BM, Baudson C, Brostaux Y, *et al.*: **archiDART: an R package for the automated computation of plant root architectural traits.** *Plant Soil.* 2016; **398**(1–2): 351–65.
    Publisher Full Text

22. Fasy BT, Kim J, Lecci F, *et al.*: **TDA: Statistical Tools for Topological Data Analysis.** R package version 1.5.1. 2017.
    Reference Source

23. Lobet G, Koevoets IT, Noll M, *et al.*: **Using a Structural Root System Model to Evaluate and Improve the Accuracy of Root Image Analysis Pipelines.** *Front Plant Sci.* 2017; **8**: 447.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Chang W, Cheng J, Allaire J, *et al.*: **shiny: Web Application Framework for R.** R package version 1.0.5. 2017.
    Reference Source

25. Wickham H: **ggplot2: Elegant Graphics for Data Analysis.** Springer-Verlag New York; 2016.
    Publisher Full Text

26. Sievert C, Parmer C, Hocking T, *et al.*: **plotly: Create Interactive Web Graphics via "plotly.js".** R package version 4.7.1. 2017.
    Reference Source

27. Bodner G, Leitner D, Nakhforoosh A, *et al.*: **A statistical approach to root system classification.** *Front Plant Sci.* 2013; **4**: 292.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Husson F, Josse J, Le S, *et al.*: **FactoMineR: Multivariate Exploratory Data Analysis and Data Mining.** R package version 1.39. 2017.
    Reference Source

29. Oksanen J, Blanchet FG, Friendly M, *et al.*: **vegan: Community Ecology Package.** R package version 2.4-4. 2017.
    Reference Source

30. Delory BM, Li M, Topp CN, *et al.*: **Data and R codes used in Delory *et al* (2017) F1000Research (Version v1.1) [Data set].** *Zenodo.* 2017.
    Data Source

31. Delory BM, Li M, Topp CN, *et al.*: **archiDART 3.0 (Version v3.0).** *Zenodo.* 2017.
    Data Source

32. Delory BM, Li M, Topp CN, *et al.*: **archiShiny: a web app to demonstrate the capabilities of archiDART 3.0 (Version 1.0).** *Zenodo.* 2017.
    Data Source

Open Peer Review

Current Referee Status: ✓ ? ✓

Version 1

Referee Report 26 March 2018

✓

Nathan D. Miller
Department of Botany, University of Wisconsin-Madison, Madison, WI, USA

"archiDART v3.0: A new data analysis pipeline allowing the topological analysis of plant root systems", presents new analysis methods available in with the archiDART platform.

Constructive Criticisms:
As is often the case with multivariate analysis, we reach for our favorite or most comfortable analysis from the myriad of linear alphabet soup. In this case, the authors choose principal component analysis (PCA) to "… determine the key traits differentiating the root systems being compared". When using linear algebra tools for finding separation of groups based on multiple traits, the authors should consider using a discriminant style analysis such as Linear Discriminant Analysis (LDA), Multi-class LDA (mLDA), or Partial Least Squares Discriminant Analysis (PLS-DA). Unlike PCA, these methods include group structure in the objective function resulting in an answer which is closer to the authors goal.

The authors need to take care of the scale or units of the traits when using PCA. For example, if some traits yield numbers on the order of 10^3 and others yield numbers on the order of 10^{-3}, how would the authors expect this to affect the presented analysis? Similar care needs to be taken when using LDA/mLDA, perhaps a normalization such as z-score or min-max would overcome the problem. As the objective function for PLS-DA has z-score "built" in, it might most easily yield reasonable results.

The authors chose to use non-metric multidimensional scaling (NMDS), on the pairwise distance matrix for the persistent homology analysis and PCA for the trait-style analysis. If there a way to present a common approach between these two methodologies?

Complements:
Simply and most importantly, the authors meet their goal. They present a free, open source, and usable R package for persistent homology analysis and useful accompanying graphical abilities.

Conclusions:
While a more thorough compare/contrast workup could be presented between the multivariate and persistent homology approaches, this paper does not claim to handle the task. The authors should consider softening the expectations/phrasing of PCA as a discrimination method or use a different approach.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 12 March 2018

? **Alexander Bucksch** 🆔

Department of Plant Biology, Warnell School of Forestry and Natural Resources, Institute of Bioinformatics, University of Georgia, Athens, GA, USA

Delory et al. present an extension package for ArchiDart that allows the user to compute topological descriptors of branching plant shapes. The contribution is valuable for the community of researchers studying plant morphology. I do recognize that this was a lot of work to implement.

I have several suggestions to improve the paper for the readers and the user community.

1. The paper gives overall a good intuition of how topological indices of branching structures are computed. Yet, the detail for further exploration is missing. Looking at the code and the paper, it was unclear to me how a user can change the function used to expand over the surface of the branching structure. Currently only the geodesic distance is used to extract a homology of a given branching structure. Does the code provide an interface to choose the function? In my opinion, It would be good to have a pseudo code to the paper that allows to see into the many functions used in the R implementation.

2. The capabilities of the software were demonstrated nicely on simulated branching structures. While the results are convincing for "perfect data" I missed a validation on noisy and incomplete data. In other words, I couldn't find evidence on the robustness of the method and if it would translate to branching structures extracted from imaging data. It would be good to know for the reader which quality criteria apply to the input data and what are the limitations if data is partial.

That way the code becomes useful for many more applications.

3. I do appreciate the contribution a lot and know the previous papers referenced in the manuscript. However, for a new reader it is very difficult to follow all the cross-references in the paper. I would like to see a one or two sentence summary of concepts for each referenced paper that contributes to the content of the presented software paper. For example: "A nice explanation of how persistent homology can be applied to capture plant topologies is provided in [12]" -> Until here it is not clear what persistent homology is, nor the concept of homology groups is explained sufficiently. The reader needs to read to more papers in order to understand H_0 groups etc. In part this can be resolved by being clear in the preceding description of a graph and how the geodesic distance is measured on it. It is left to the reader to assume an unknown embedding of the graph in to a metric space. Currently, the graph is only introduced by having nodes and edges. Intuitively I would expect the distance between two nodes to be the number of edges between to nodes. Yet later on it is talked about the geodesic distance on the graph. I assume that the graph is embedded in an Euclidean space and the "straight edges" might be curves for which the curve length is taken to calculate the geodesic distance between any two nodes. Or did I got this wrong? More precision in the mathematical language would be helpful. Similar issues are present elsewhere in the manuscript whenever the explanation is referenced to another paper, e.g. to Archidart 1.0.

4. A confusing aspect of the paper is that Archidart 1.0 has Archidart 3.0 as a follow up. Did I miss something? You state that Archidart 3.0 implements 5 new functions compared to Archidart 1.0. None of the functions presented in table 1 refer to Archidart 2.0. Should the presented version be Archidart 2.0?

5. A very general comment: It is nothing new that geometry and topology are needed to quantify shapes. Work of Biassotti, Spagnuolo, Veltkamp etc. show that since many years. May be most closely related to the paper here is the paper Biasotti, Silvia et al. (2008)[1] that discusses the interplay of geometry and topology for skeletal shape analysis. Many other papers show the same.

References

1. Biasotti S, Attali D, Boissonnat J, Edelsbrunner H, Elber G, Mortara M, di Baja G, Spagnuolo M, Tanase M, Veltkamp R: Skeletal Structures. 2008. 145-183 Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 26 January 2018

https://doi.org/10.5256/f1000research.14706.r29709

✔ **Magdalena M. Julkowska** (iD)

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

The paper by Delory et al., "archiDART v3.0: A new data analysis pipeline allowing the topological analysis of plant root systems" describes an R package archiDART that includes methods allowing analysis root topology using persistent homology. The paper describes the novelty of the package really well and it is providing new and exciting ways for the community to differentiate between various root architecture types. The Shiny app developed is illustrating the possibilities of the new package very neatly.

I do have few questions/suggestions:

1. When computing a pairwise bottleneck distance matrix between individual files, what kind of statistical analysis can be used to decide whether a distance between two genotypes/groups is statistically significant? The authors could at least put a suggestion of the follow-up analysis. That would be extremely useful for the part of a scientific community that works with mutants/introgression lines and examines root architecture to answer hypothesis driven questions.

2. The authors compare the classical root traits with persistent homology using PCA and non-metric multidimensional scaling and showing that the later is resulting in clear differences between fibrous and tap-root root types. I think that this difference could result in using PCA vs NMDS rather than the input. For the fair comparison, the value of two different inputs in distinguishing between two different root types, should be compared using the same method for reduction in dimensionality.

3. Although the persistent homology is an alternative way of differentiating between the root types, the method is still having limited capability in distinguishing between "mock", "shallow" and "steep" root types (as shown in the archiDart shiny app). The same types are also not being differentiated using PCA for the "classical" root phenotypes. It would be very helpful if the authors state the limitations of using the homology (like - problems with implementing the root angle) and how exactly is the persistent homology complementary to the more classical analysis of root system architecture.

4. There is a glitch in the app when user removes the genotypes to be plotted in the tabs "archiDRAW" and archiHOMOLOGY".

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com