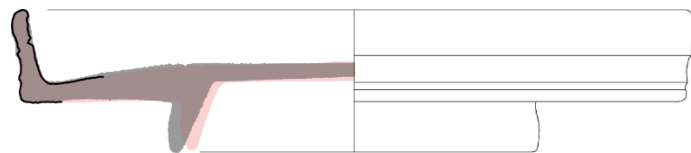


RACORD

Computer-Assisted Shape Classification of Archaeological Pottery Fragments



R A C O R D

Computer-Assisted Shape Classification of Archaeological Pottery Fragments

Credits:

Josef Wilczek, KA UHK Hradec Králové / ARTEHiS Dijon (josef.wilczek@hotmail.com)

Fabrice Monna, ARTEHiS Dijon (Fabrice.Monna@u-bourgogne.fr)

Nicolas Navarro, EPHE Dijon (nicolas.navarro@u-bourgogne.fr)

Carmela Chateau Smith, CPTC Dijon (carmela.chateau@laposte.net)

Version: 2021/02/08

Installation

1. Support program installation

Download and install following applications, depending on your operational system, from official developer web sites:

- 'R program' (<https://www.r-project.org/>)
- 'RStudio Desktop' (<https://www.rstudio.com/>)

2. Application installation

Download and unpack the **RACORD.zip** file. Copy the unpacked RACORD folder on your computer (for example into **C:/RACORD**).

3. Installing packages

Run 'RStudio' program and open **server.R** file (Fig. 1). RACORD uses several libraries for computation and visualisation. In further steps, all necessary libraries must be installed on your computer. Internet connection is required. Two library installation options are possible:

- a) Simple click on the 'Install' link (Fig. 1)
- b) Copy the following code lines to the R console and press enter (Fig. 2).

```
install.packages(c("DT", "kmlShape", "MASS", "shiny", "sp"))
```

4. Preparing and running the RACORD application

The path to the application must be defined in its source code. Specify the path to the application folder by replacing the code at the beginning of the **server.R** file by (Fig. 3):

```
WD <- c("C:/RACORD")
```

Note that instead of **C:/DACORD** in the code, you can specify the path to the folder containing the application.

You can also specify the folder containing '.txt' files including profile coordinates (Fig. 3).

```
dir_profiles <- c("C:/RACORD/profiles/")
```

The RACORD application is run by clicking on the **Run App** icon in the RStudio panel (Fig. 4; Fig. 5). The graphical layout of the application can be adjusted at the beginning of the **ui.R** file, depending on the screen resolution (Fig. 6).

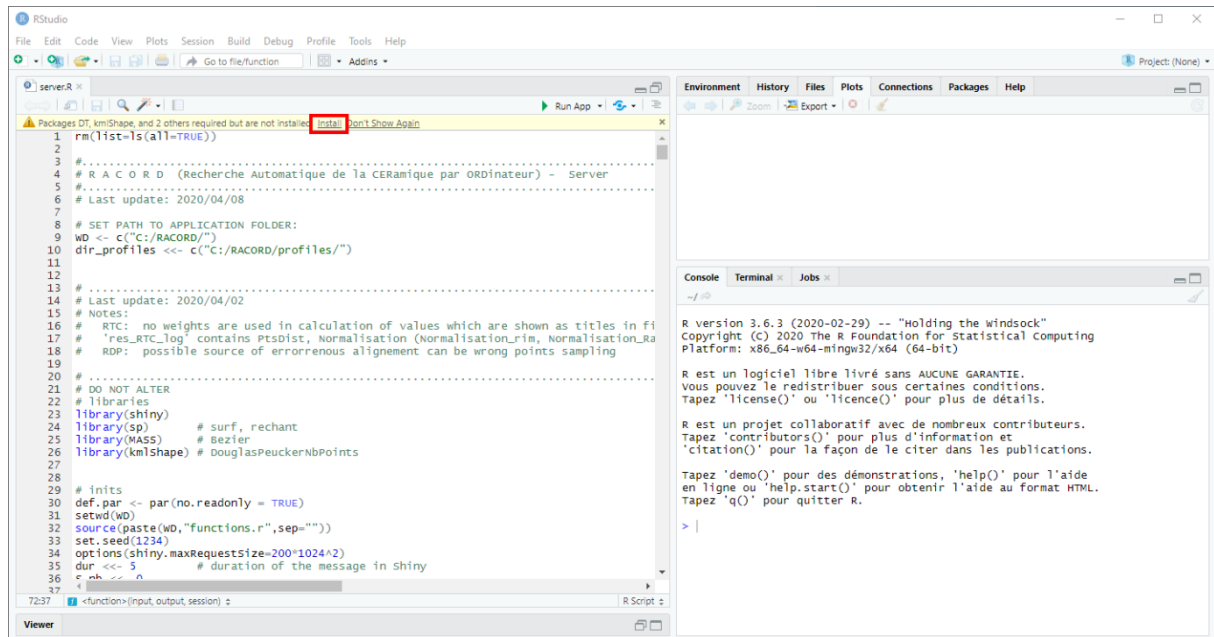


Fig. 1. Packages installation.

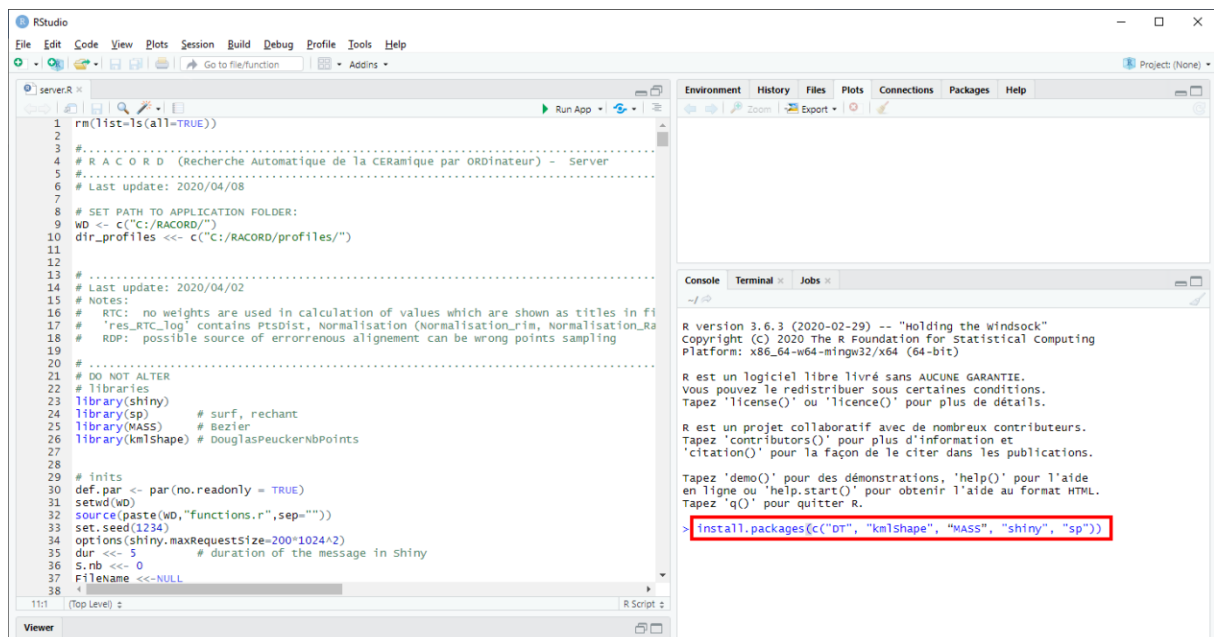


Fig. 2. Packages installation.

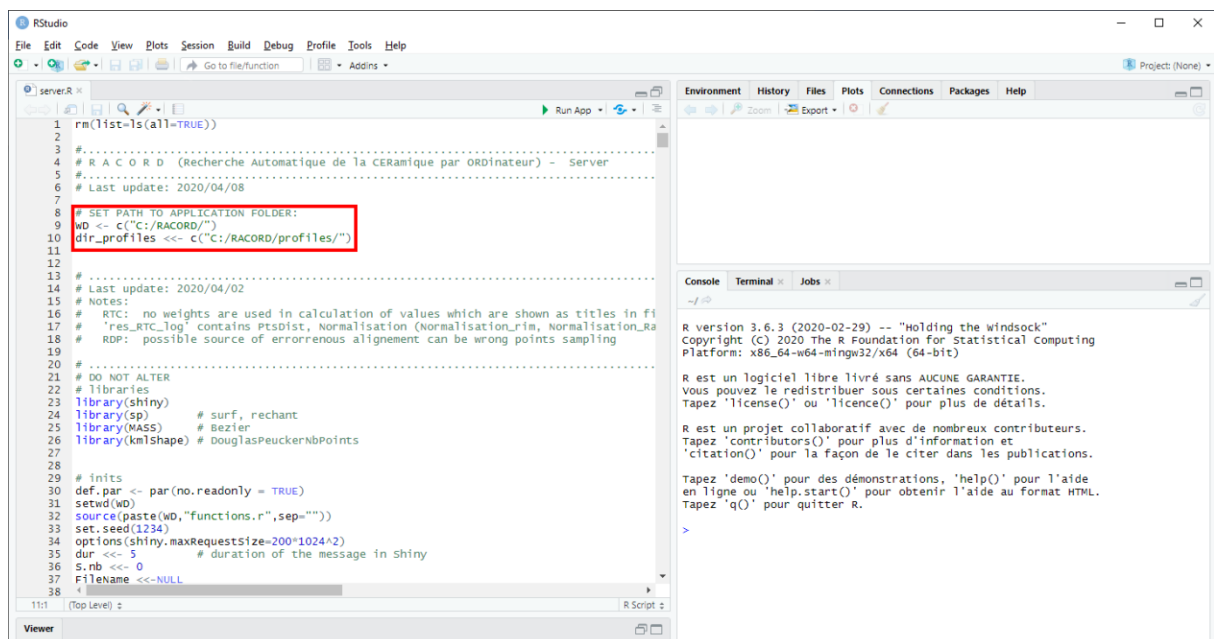


Fig. 3. Setting path to the source code and to the folder containing 'txt' files with profile coordinates.

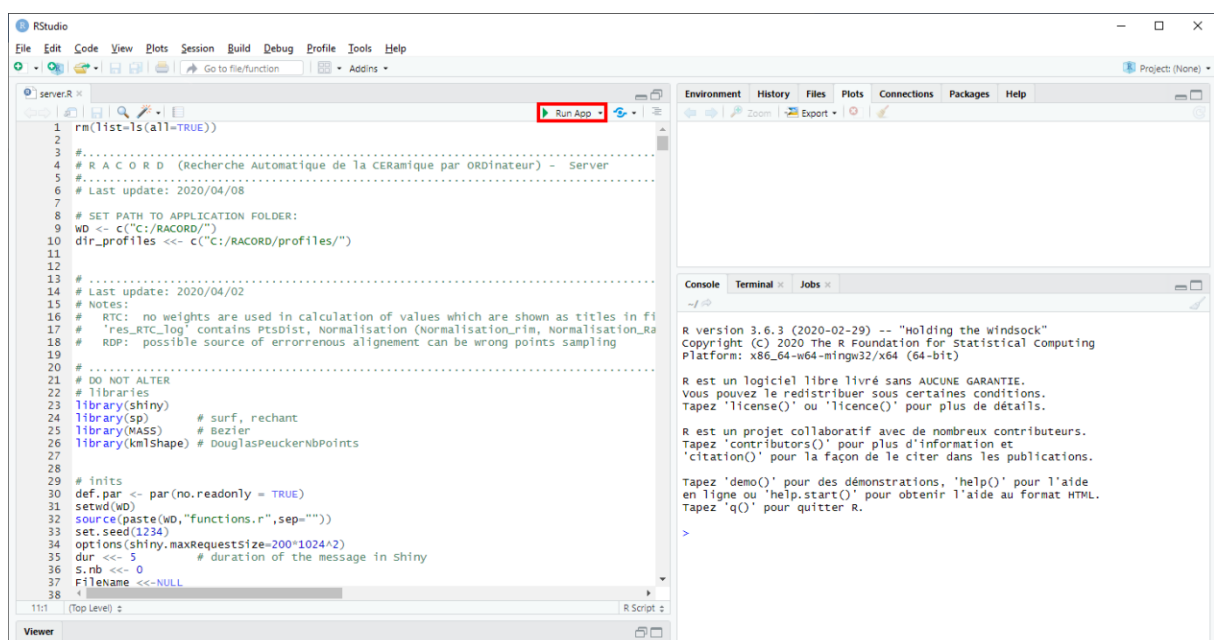


Fig. 4. Running the application.

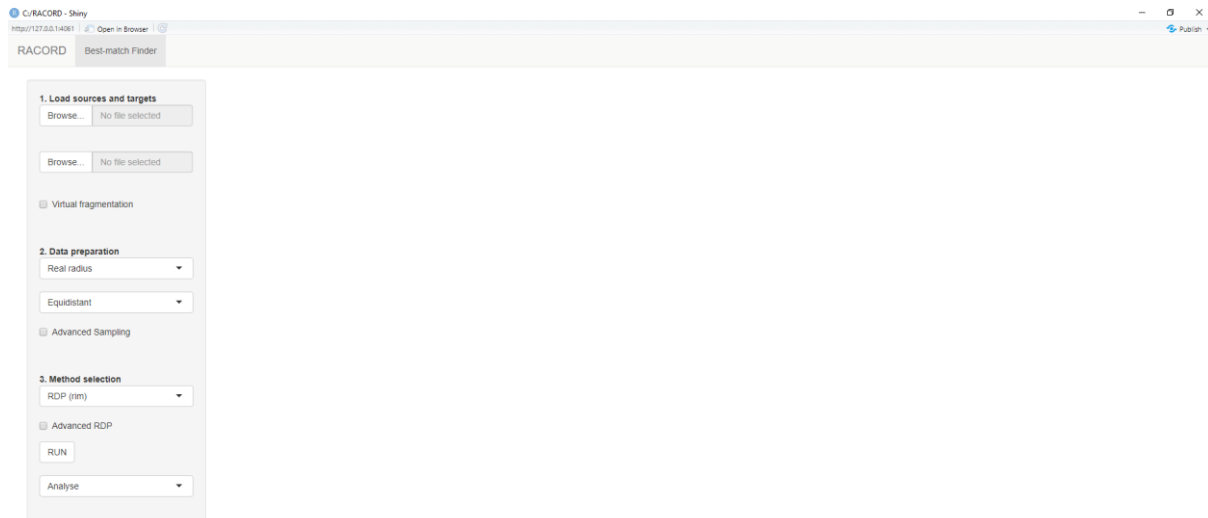


Fig. 5. Running the application.

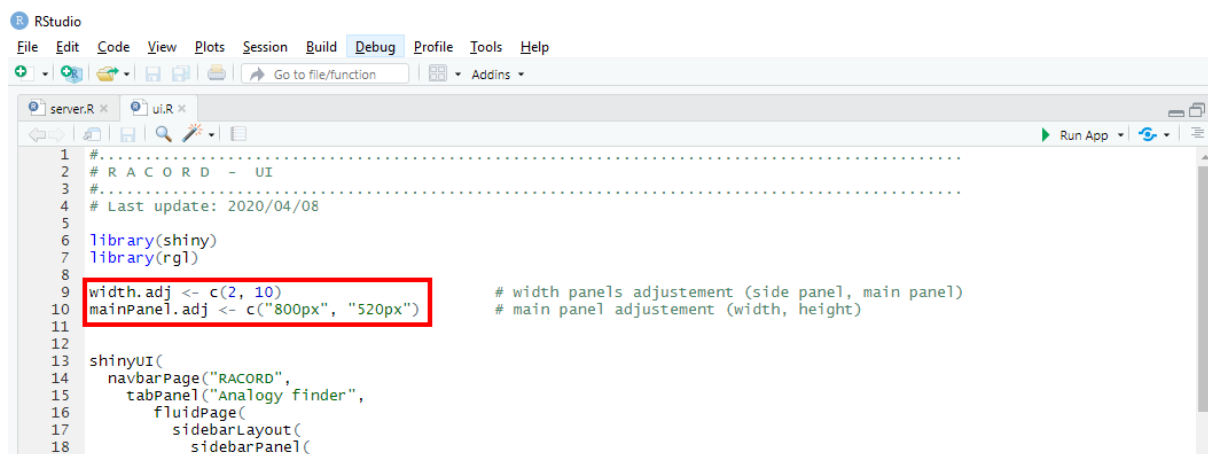


Fig. 6. Adjusting graphical layout of the application.

Manual

The goal of the application is the retrieval of the best match to a given **source** fragment from a dataset containing **target** complete vessel profiles. The procedure consists of three steps: 1) Load sources and targets, 2) Prepare data, 3) Select a method and view results.

The application therefore contains three sections dedicated to the corresponding tasks (Fig. 5).

The Grafeusénque dataset used in the test study is provided as an example.

1. Load sources and targets

Two [Browse](#) buttons serve to load files containing information about sources and targets stored in the form of tables. Each table has a header (Fig. 7:A). Table rows of the table correspond to fragments to be classified (for sources) and to complete vessels (for targets), table columns correspond to three descriptors (ID, diameter, class) separated by a semicolon symbol (Fig. 7:B):

- ID: fragment/vessel identification.
- Diameter: rim diameter of the fragment/vessel (in cm); comma is used for decimals
- Class: morphological class of the fragment/vessel.

Note that the ID of each fragment/vessel must be the same as 'txt' file containing its profile coordinates (Fig. 7:C). This 'txt' file needs to be stored in the 'profile' folder specified earlier (Fig. 7:D; see also Figure 3 for setting the path to this folder).

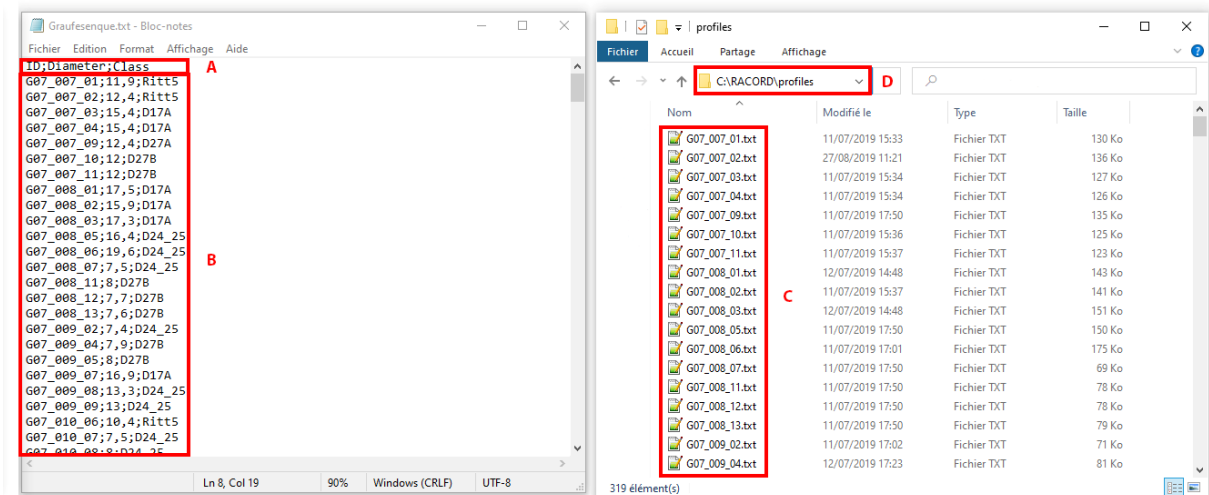


Fig. 7. Structure of source/target file. Note that the ID of vessels must refer to txt files containing corresponding profile coordinates.

Once loaded, both tables appear on the screen. Sources are on the left, targets on the right. Items selected from these tables by left mouse click can be visualised together on the right of the screen (Fig. 8).

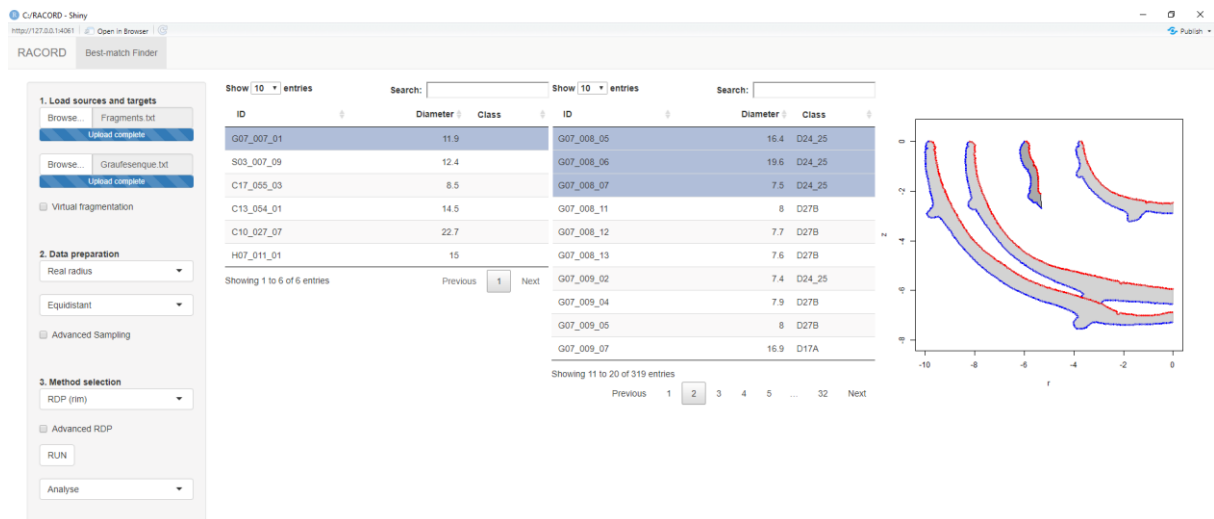


Fig. 8. One selected source fragment (G07_007_01; dark grey polygon) is plotted with three selected target complete vessels (G07_008_05, G07_008_06, and G07_008_07; light grey polygon).

Virtual fragmentation generates source fragments by virtually breaking (i.e. segmenting) complete vessels. Fragments are generated by selection of a complete source vessel (Fig. 9:A), checking **Virtual fragmentation** option (Fig. 9:B), and clicking on **Break profile** button (Fig. 9:C). Newly created fragments then appear in the source table (Fig. 9:D).

Four parameters can be additionally specified:

- **Number of fragments** specifies number of fragments generated.
- **Minimum % size of one fragment** sets a minimal size of one fragment.
- **Set seed** specifies pseudorandom number used in fragment generation.
- **Extract only rim** generates only one rim fragment.

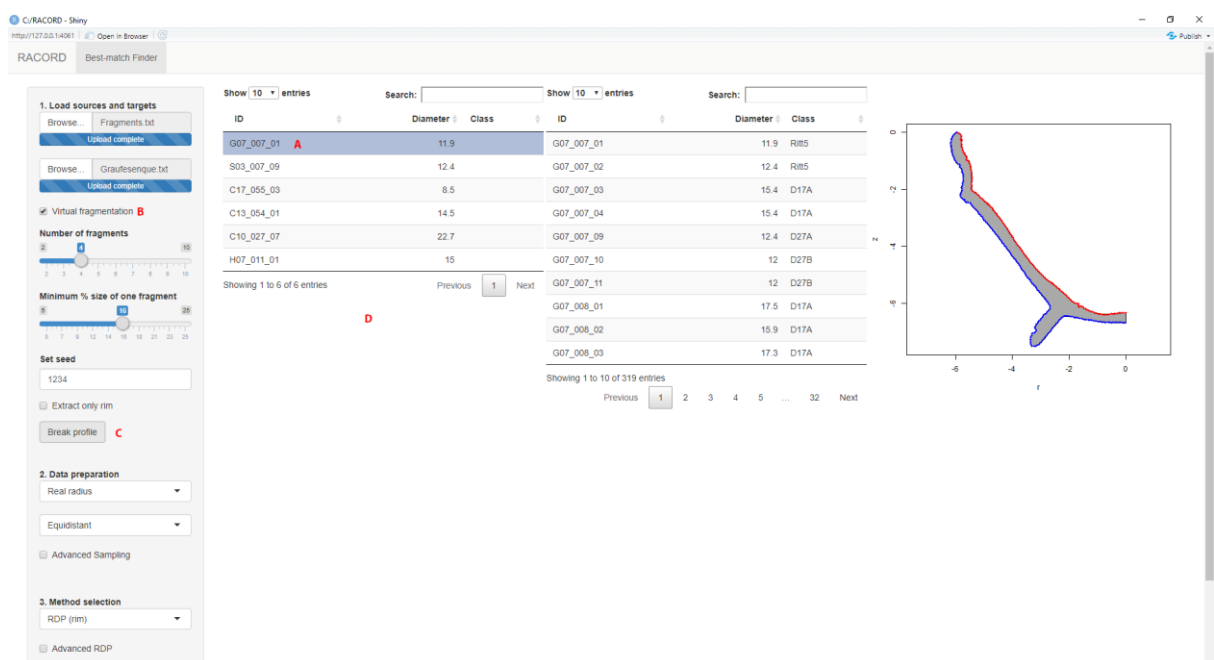


Fig. 9. Generating source fragments by virtually breaking (i.e. segmenting) complete vessel.

2. Data preparation

Data preparation is used to normalise the size of profiles and to choose the strategy used for sampling points on profile outlines.

Size normalisation

- **Real radius** uses real metrics (i.e. cm) in matching.
- **Rim units** standardises outlines to the same rim unit. If selected, only the shape (and not size) of profiles is used in profile comparison.

Sampling specification

- **Segments** splits a profile into a specified number of segments. This method should be used only for ICP algorithm. Number of segments can be specified by checking **Advanced sampling** option.
- **Equidistant** generates equally spaced points along the outline, starting from the uppermost point. It assumes the homology of segments sampled along the profile. **Distance between points** can be specified by checking **Advanced sampling** option.

3. Method selection and Viewing results

Once the data prepared, five different methods and their variants can be used for fragment matching:

- (1) **DCT (rim)** – Discrete Cosine Transform
- (2) **RDP (rim)** – Ramer-Douglas-Peucker algorithm
- (3) **RTC (rim)** – Radius, Tangent and Curvature
- (4) **Bezier (rim)** – Bezier Polynomials
- (5) **ICP (rim)** – Iterative Closes Point. No translation, size, or rotation allowed. Function thus calculates solely Euclidean distances between profiles.
- (6) **ICPz (all)** – Iterative Closes Point. Only translation along the rotational axis is allowed.
- (7) **ICP (all)** – Iterative Closes Point – translation, size, and rotation are allowed.
- (8) **All Methods (rim)** – All rim methods (1, 2, 3, 4, 5) executed in one batch.

The choice of the most suitable method depends on the fragment part (rim vs no matter which part), and the feature the user chooses to focus on.

The calculation parameters used in matching can be fine-tuned in **Advanced options** (numbers in parentheses correspond to matching methods):

Harmonics number (1) indicates the number of harmonics obtained with Discrete Cosine Transform (20, by default).

Number of points (2) indicates number of points generated by Ramer-Douglas-Peucker algorithm (20, by default).

Function weights (Rs, Ts, Ks) (3) indicate weights of Radius, Tangent and Curvature functions as proposed by Karasik and Smilansky (2011).

Translation parameter (z) (6) indicates the translation step along the rotational axis used in searching.

Optimisation parameters (r, theta, size) (7) indicates the limits of r-translation (in cm or radius units), rotation (in degrees), and size (magnitude) parameters used in optimisation.

SAN parameters (MaxIter, InitTemp, MaxTemp) (7) indicates the maximum number of iterations, the initial temperature, and the number of evaluations at each temperature used in Simulated Annealing optimisation calculation (2000, 100, and 40, by default).

The matching process is started by selecting source fragments to be searched and by pressing the **Run** button.

Once finished, the source table changes its appearance (**Fig. 10**). Columns in the table now indicate which fragment(s) was/were analysed, by which method.

The screenshot shows the RACORD web application interface. On the left is a sidebar with three sections: '1. Load sources and targets' (with 'Browse...' and 'Load complete' buttons), '2. Data preparation' (with 'Real radius' and 'Equidistant' dropdowns), and '3. Method selection' (with 'RDP (rim)' dropdown and 'RUN' button). The main area displays two tables. The left table, titled 'Showing 1 to 6 of 6 entries', has columns: ID, Diameter, BEZ, DCT, RDP, and RTC. The right table, titled 'Showing 1 to 10 of 318 entries', has columns: ID, Diameter, and Class. The 'RDP' column in the left table contains the value 'O' for the first entry.

ID	Diameter	BEZ	DCT	RDP	RTC
1 G07_007_01	11.9			O	
2 S03_007_09	12.4				
3 C17_005_03	8.5				
4 C19_054_01	14.5				
5 C10_027_07	22.7				
6 H07_011_01	15				

ID	Diameter	Class
G07_007_02	12.4	R05
G07_007_03	15.4	D17A
G07_007_04	15.4	D17A
G07_007_09	12.4	D27A
G07_007_10	12	D27B
G07_007_11	12	D27B
G07_008_01	17.5	D17A
G07_008_02	15.9	D17A
G07_008_03	17.3	D17A
G07_008_05	16.4	D24_25

Fig. 10. Example of the screen with finished matching. The source fragment ‘G07_007_01’ was analysed by ‘RDP’ method.

The best match or matches is/are then visualised by selecting analysed source fragment (**Fig. 11:A**), switching **Analyse** selection to **Results** (**Fig. 11:B**), sorting target complete vessels increasingly according to the RMSD value (**Fig. 11:C**), and selecting target vessel or vessels for which the RMSD value is/are smallest (**Fig. 11:D**).

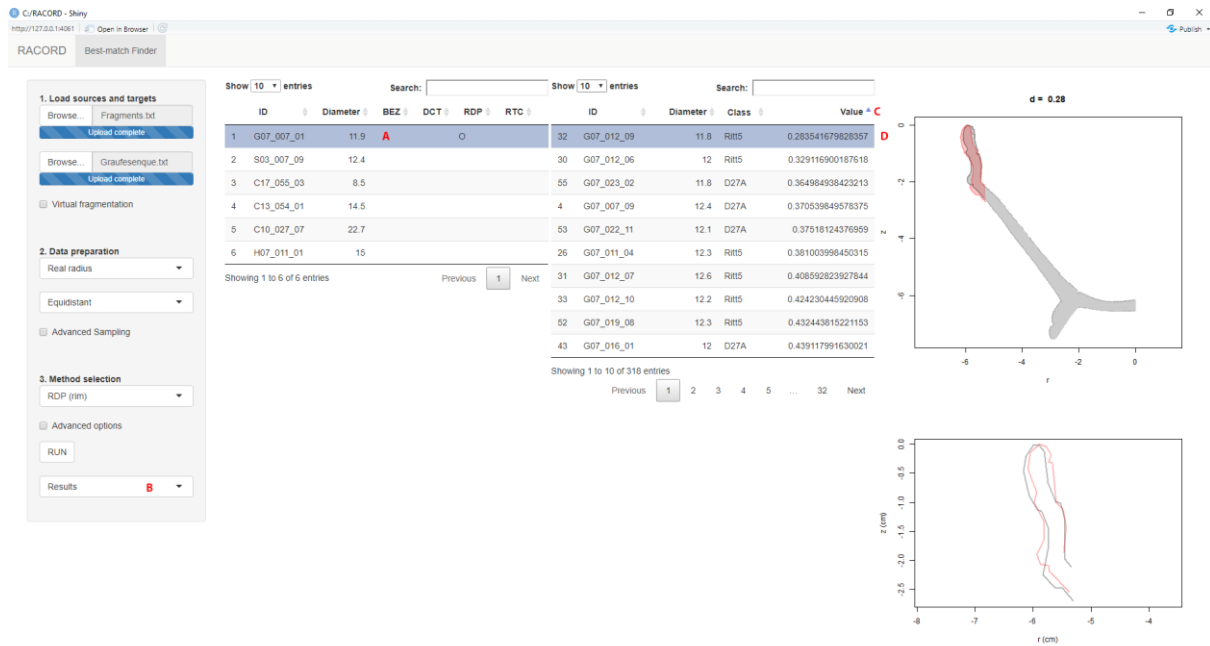


Fig. 11. Example of viewing the best match for the 'G07_007_01' fragment analysed by 'RDP' method.

4. Attribution (Supervised classification)

In the new version (February 2021), it is possible to label fragments based on morphological classes of the selected number of best candidates (k).

Attribution is automatic after running the matching process (see Section 3).

The classification parameters used in attribution can be adjusted in [Labelling options](#).

- **Number of best matches** (k) specifies the number of best matches (best candidates) used for classification (1, by default).
- Switching between **Majority voting** and **All the same class** selects the classification strategy. **Majority voting** option attributes the most numerous class of k best candidates to the fragment. **All the same class** option is similar to Majority voting; the difference is that the fragment is classified **only** if all k best candidates possess the same class.

The classification parameters for the fragment analysed are validated by selecting fragment (Fig. 12:A) and by pressing the [Refresh](#) button (Fig. 12:B). Once classification is finished, the source table changes its appearance (Fig. 12:C), and shows the proposed labelling, based on the criteria selected.

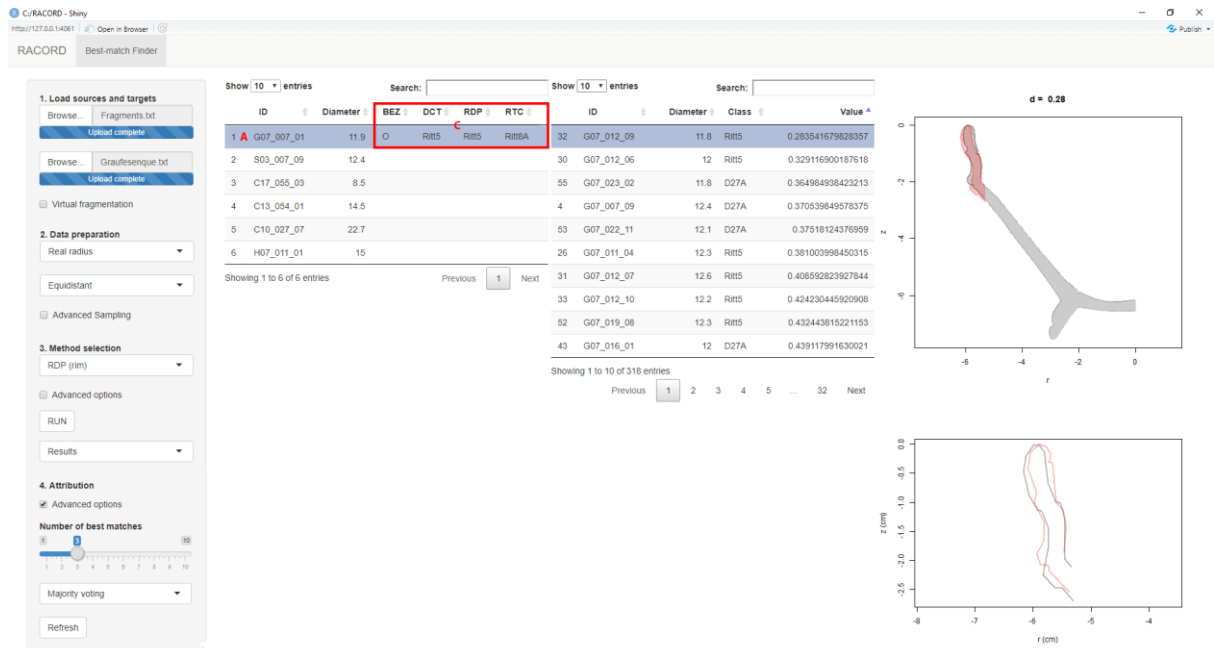


Fig. 12. Example of supervised classification for the 'G07_007_01' fragment analysed by 'BEZ', 'DCT', 'RDP', and 'RTC' methods. Classification was made using majority voting with 3 best matches. Figure 12 shows that the fragment was classified as 'Ritt5' by the 'DCT' and 'RDP' methods, and as 'Ritt8A' by the 'RTC' method. The 'BEZ' method could not label the fragment based on the selected criteria.

Note that morphological labels provided by RACORD are only indicative. The final decision as to the category in which the fragment will be classified is left to the operator.