qfingerprinting analysis.r[1] manual

by Alban Ramette
October 13, 2008

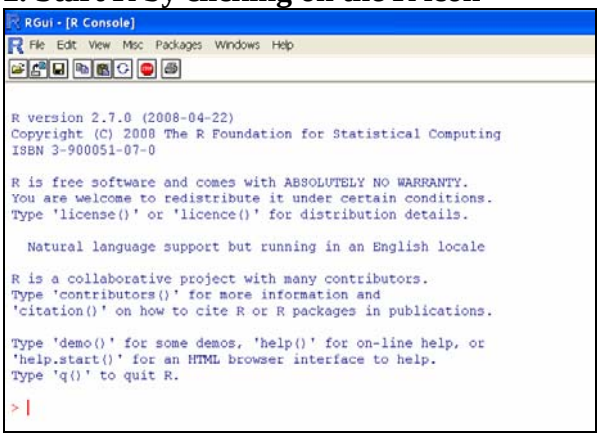
1. Preparation of the data table

See the Excel table provided as example (D_for_qfingerprinting_analysis.xls). Make sure while preparing the data that the row and column labels do not contain spaces.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	bins	01_cm_A	01_cm_A1	01_cm_A2	01_cm_A3	01_cm_A4	01_cm_A6	01_cm_B	01_cm_B1	01_cm_B2	01_cm_B3	01_cm_B4	01_cm_B5	01_cm_B6	01_cm_C	01_cm_C1	01_cr
2	101	0	0	0.23	0	0	0	0	0	0	0	0	0	0	0	0	
3	103	0	0	0	0	2.56	9.56	0	0	0	2.28	0	0	0	0	0.26	
4	105	0	0	0	0	0	0	0	0	0	0	0	0	0.81	0	0	
435																	
436	991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
437	993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
438	995	0.19	0.45	0.75	0	0	0	0	0.55	0.53	0	0	0	0	0.21	0.29	
439	997	0	0	0	0	0	0	0.19	0	0	0	0	0	0	0	0	
440	Dilution	0	1	2	3	4	6	0	1	2	3	4	5	6	0	1	2
441	Sample	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
442																	

In this example, the first row contains the sample names, the first column contains the bin names and the last two rows contain the dilution level (0 to 6) and the sample identification number (1 to 9). The whole table (including the row and column labels) is then copied to a text file (e.g. using Notepad). An example is given in the Dforqfingerprint.txt file.

2. Start R by clicking on the R icon



3. Load the data into the R workspace

At the prompt (>), indicate in which directory you want to read and write the data (i.e. where you also put your .txt file. The directory should be created beforehand) and press enter. Note that quotes and \\ are used to indicate the path to the directory.

```
>setwd("c:\\R\\ARISA")
```

Then, load the data into the object D by writing (make sure to exactly type the dots, commas, and punctuation signs, as indicated and use " instead of ”):

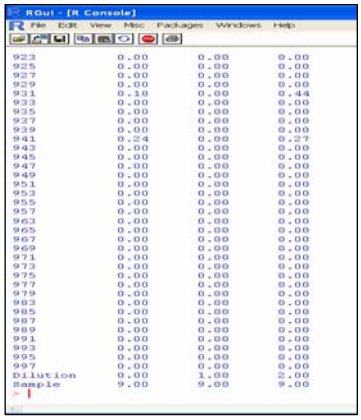
---

[1]This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.
This program is distributed in the hope that it will be useful, but **without any warranty**; without even the implied warranty of **merchantability** or **fitness for a particular purpose**. See the GNU General Public License for more details (Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307, USA)

```
> D=read.table("Dforqfingerprint.txt",h=TRUE,row.names=1)
```

If you now type:
```
>D
```

You should see your data table appear in the R console (if it is a big table, you will not see the first rows, but just the end of the table):



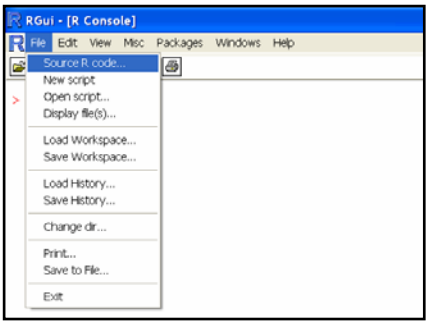
In our example, the table was produced in Excel in a OTUs-by-samples format (because Excel is limited to 256 columns). We need to transpose the data to obtain a samples-by-OTUs table and save it under the same name D:
```
>D=t(D)
```

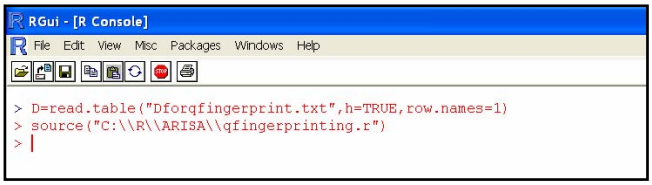
If the table was initially in the right format, there is no need to transpose the table. An error message may appear if the sample direction in the table is not correct. It is therefore always recommended to double-check the table orientation to avoid wrong calculations.

We are now ready to run the R qfingerprinting script on the data.

4. Running the qfingerprinting script

In the menu bar, go to Source R code…

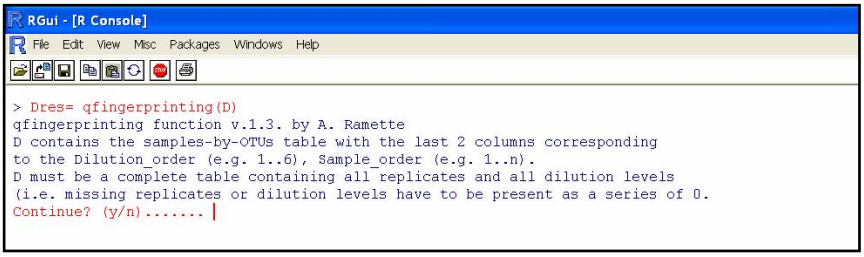




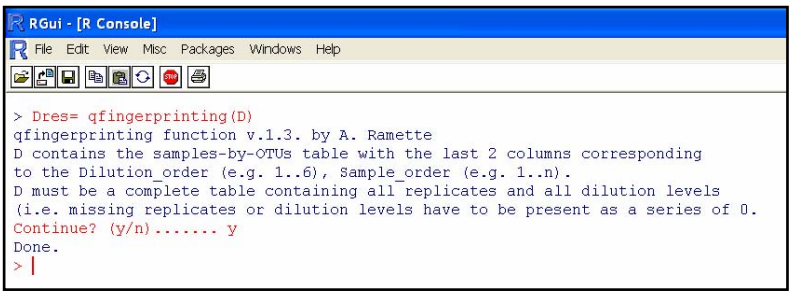
And indicate the location of your saved version of the qfingerprinting.r script.

Apply the function to the data `D` and store the results in the object `Dres`:
```
>Dres= qfingerprinting(D)
```

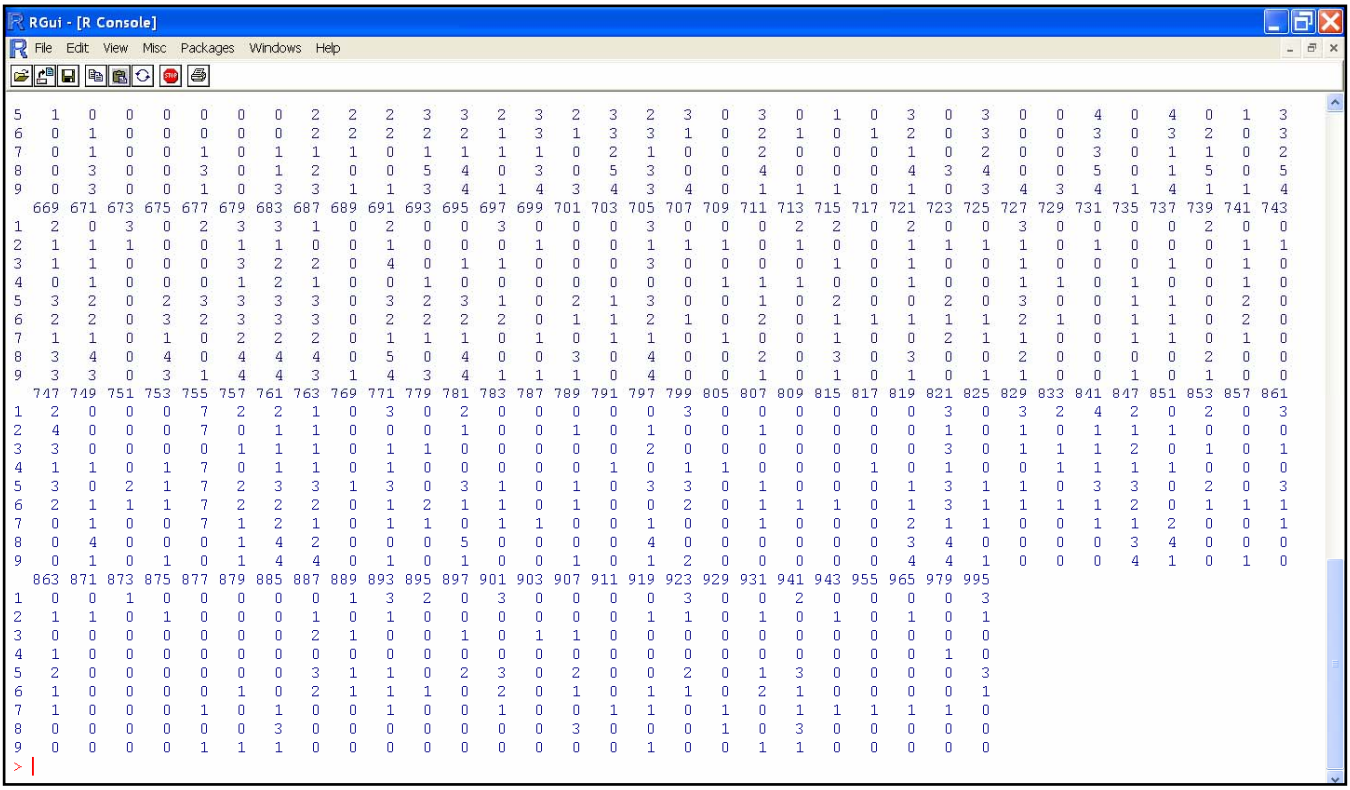
The script starts by indicating some basic information about the version, expected data format and ask if you want to proceed. Type "`y`" (without the quotes), if the data table corresponds to the description provided, otherwise type "`n`" and see point 3) above.



After few seconds, the following message should appear in the R console, indicating that the calculations are finished:

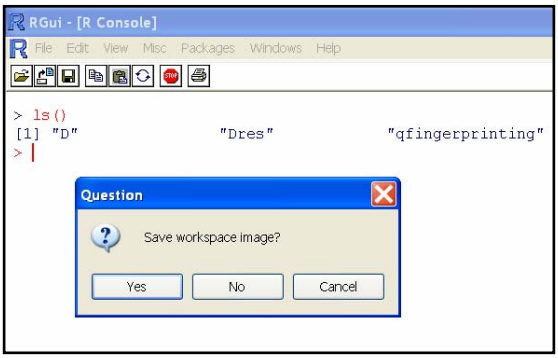


By typing:
>`Dres`
You will get the whole table converted into $\log_{10}$ abundance values (see reference for detailed explanation about how the consensus and continuity rules are applied to convert peak presence into a $\log_{10}$ abundance table).



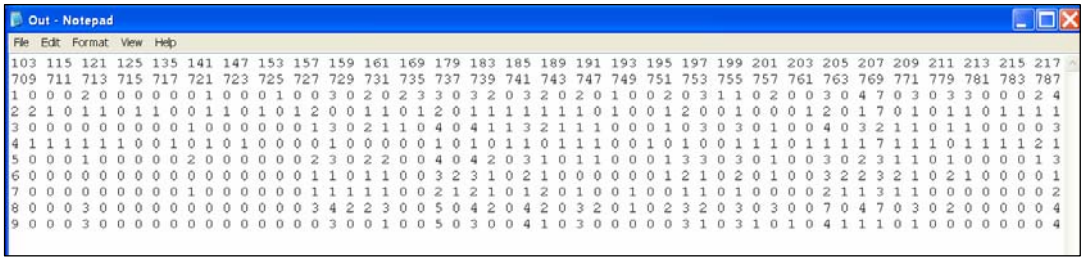
The script automatically saves and exports the result table to Out.txt file of the working directory (e.g. to C:\R\ARISA).

6. Exiting from R



The data are stored in the current workspace and you can save them via the File\save workspace option in the menu bar. Note that typing ls() lists all objects currently available in your session. You can choose to save the objects or not for future work before closing the R console.

The Out.txt file (below) contains OTUs indicated in the first row as column names and the sample number indicated in the first column (here 1 to 9). The data can be further copied to Excel and processed.



How to cite the script?
Ramette, A. (2008) Quantitative molecular community fingerprinting for estimating the abundance of operational taxonomic units in natural microbial communities. *submitted*