Guide to Turning an RDP File into a Data Set

Berkley Shands, Patricio S. La Rosa, Elena Deych, William Shannon

October 30, 2024

Below we will define the basic steps required to generate a data set from an RDP file

1. Take an RDP file such as this example:
;Root:1.0;Bacteria:1.0;Firmicutes:1.0;Bacilli:1.0;Bacillales:1.0;Staphylococcaceae:1.0;
;Root:1.0;Bacteria:1.0;Firmicutes:1.0;Bacilli:1.0;Lactobacillales:1.0;Carnobacteriaceae:0.8;
;Root:1.0;Bacteria:1.0;Bacteroidetes:1.0;Bacteroidia:1.0;Bacteroidales:1.0;Prevotellaceae:1.0;
;Root:1.0;Bacteria:1.0;Bacteroidetes:1.0;Bacteroidia:1.0;Bacteroidales:1.0;Prevotellaceae:1.0;
;Root:1.0;Bacteria:1.0;Bacteroidetes:1.0;Bacteroidia:1.0;Unclassified:0.5;Unclassified:0.5;

2. Take the first entry and seperate each level into its own row, while seperating levels by a period:
Root, 1
Root.Bacteria, 1
Root.Bacteria.Firmicutes, 1
Root.Bacteria.Firmicutes.Bacilli, 1
Root.Bacteria.Firmicutes.Bacilli.Bacillales, 1
Root.Bacteria.Firmicutes.Bacilli.Bacillales.Staphylococcaceae, 1

3. Do the same with each following row, adding to the number at the end if it is the same:
Root, 5
Root.Bacteria, 5
Root.Bacteria.Firmicutes, 2
Root.Bacteria.Firmicutes.Bacilli, 2
Root.Bacteria.Firmicutes.Bacilli.Bacillales, 1
Root.Bacteria.Firmicutes.Bacilli.Bacillales.Staphylococcaceae, 1
Root.Bacteria.Firmicutes.Bacilli.Lactobacillales, 1
Root.Bacteria.Firmicutes.Bacilli.Lactobacillales.Carnobacteriaceae, .8
Root.Bacteria.Bacteroidetes, 3
Root.Bacteria.Bacteroidetes.Bacteroidia, 3
Root.Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales, 2
Root.Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae, 2
Root.Bacteria.Bacteroidetes.Bacteroidia.Unclassified, .5
Root.Bacteria.Bacteroidetes.Bacteroidia.Unclassified.Unclassified, .5

4. Change any unclassifieds or anything else that could appear under mul-

tiple different parents (we prefer adding a 'U' to the parents name):
Root.Bacteria.Bacteroidetes.Bacteroidia.BacteroidiaU, .5
Root.Bacteria.Bacteroidetes.Bacteroidia.BacteroidiaU.BacteroidiaUU, .5

5. Remove the 'Root' level since we are only looking at Bacteria in this example:
Bacteria, 5
Bacteria.Firmicutes, 2
Bacteria.Firmicutes.Bacilli, 2
Bacteria.Firmicutes.Bacilli.Bacillales, 1
Bacteria.Firmicutes.Bacilli.Bacillales.Staphylococcaceae, 1
Bacteria.Firmicutes.Bacilli.Lactobacillales, 1
Bacteria.Firmicutes.Bacilli.Lactobacillales.Carnobacteriaceae, 1
Bacteria.Bacteroidetes, 3
Bacteria.Bacteroidetes.Bacteroidia, 3
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales, 2
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae, 2
Bacteria.Bacteroidetes.Bacteroidia.BacteroidiaU, .5
Bacteria.Bacteroidetes.Bacteroidia.BacteroidiaU.BacteroidiaUU, .5

6. Repeat for any additonal RDP files and add them as new columns (and rows as needed) in our data set:
Taxa, Sample 1, Sample 2
Bacteria, 5, 5
Bacteria.Firmicutes, 2, 5
Bacteria.Firmicutes.Bacilli, 2, 5
Bacteria.Firmicutes.Bacilli.Bacillales, 1, 5
Bacteria.Firmicutes.Bacilli.Bacillales.Staphylococcaceae, 1, 5
Bacteria.Firmicutes.Bacilli.Lactobacillales, 1, 0
Bacteria.Firmicutes.Bacilli.Lactobacillales.Carnobacteriaceae, 1, 0
Bacteria.Bacteroidetes, 3, 0
Bacteria.Bacteroidetes.Bacteroidia, 3, 0
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales, 2, 0
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae, 2, 0
Bacteria.Bacteroidetes.Bacteroidia.BacteroidiaU, .5, 0
Bacteria.Bacteroidetes.Bacteroidia.BacteroidiaU.BacteroidiaUU, .5, 0
Bacteria.Firmicutes.Bacilli.Lactobacillales.Streptococcaceae, 0, 5

7. Save the file as a .csv (or anything other file readable by R) and load it in R:

Notes:

- Any symbol can be used as a taxa level seperator except colons.
- There can only be one top level node, if we were looking at Archaea too we would need to keep the Root level in.