

How to make a file of “your favorite genes” for the Tree Constructor

Brian White

Brian.white@umb.edu

June 17, 2004 9:00 PM

1. Open a file in Word
2. Save it as “Text Only” and call it “sequences.txt”
3. click the toolbar button to show paragraph marks (it looks like a backwards P) – you will need to see these marks to be sure the file is formatted correctly.
4. Set up your file as shown below.
5. The sequences do not have to be in any particular order.
6. Burn it onto a CD.

File format:

- Each sequence needs to have three lines with a single return at the end of each. The three lines are:
 - >Name – the name of the organism as it will appear on the web page. It should be something descriptive & not too long. Some other features:
 - The line must start with a > sign; no spaces between it and the name.
 - It can **only** have the digits 0-9, letters A-Z (upper or lower case), and spaces.
 - The first 10 characters (including spaces) must be unique.
 - they will be listed in alphabetical order (capital letters are alphabetized separately from lower-case, so be consistent)
 - #Group – the name of the group of organisms you want this organism to be included in. In the original, they were Animals, Plants, and Other. Some notes:
 - Each organism must be in a group.
 - The line must start with a # sign; no spaces between it and the group.
 - It can **only** have the digits 0-9, letters A-Z (upper or lower case), and spaces.
 - There can be as few as one group, but probably not more than 4 (there isn't enough room).
 - They will be listed in alphabetical order from left to right. So, if you want to put them in different order, you can number them “1 Zoo”, “2 Ocean”, etc.
 - !Sequence – the DNA or protein sequence for that organism
 - The line must start with a !; no spaces between it and the sequence.
 - There can only be one ‘return’ at the end of the line. Be sure that the sequence does not have a paragraph mark in the middle of the sequence. The program will stop reading at the first ‘return’.
 - The program can detect whether or not you are using DNA or protein, but a given set of sequences must be only DNA or protein, NOT BOTH.

Here is part of a sample file:

```
>dog
#Animal
!GDVEKGKKIFVQKCAQCHTVEKGGKHKHTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKK
YIPGTKMIFAGIKKTGERADLIAYLKKATKE

>southern elephant seal
#Animal
!GDVEKGKKIFVQKCAQCHTVEKGGKHKHTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKK
YIPGTKMIFAGIKKTGERADLIAYLK IATKE

>Schreibers long-fingered bat
#Animal
!GDVEKGKKIFVQKCAQCHTVEKGGKHKHTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEATLMEYLENPKK
YIPGTKMIFAGIKKSAERADLIAYLKKATKE

>eastern gray kangaroo
#Animal
!GDVEKGKKIFVQKCAQCHTVEKGGKHKHTGPNLNGIFGRKTGQAPGFTYTDANKNKGIIWGEDTLMEYLENPKK
YIPGTKMIFAGIKKKGERADLIAYLKKATNE
```

If you show the paragraph marks, here is how one set should look on the screen:

```
>chimpanzee¶
#Animal¶
!GDVEKGKKIFIMKCSQCHTVEKGGKHKHTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKKYIIPGTKMI
FVGIKKKEERADLIAYLKKATNE¶
¶
¶
```

Paragraph mark

Note no paragraph mark here