

Feedback — Quiz 2: Consensus Trees (computer)

[Help Center](#)

You submitted this quiz on **Sun 14 Jul 2013 11:10 AM PDT**. You got a score of **7.00** out of **7.00**.

Getting started

1. Start Terminal window

2. Construct working directory:

```
mkdir condist
```

```
cd condist
```

3. Copy data file:

Note: in the following command you are copying a file that you were supposed to create as part of the week 2 quiz. Specifically, this is the alignment of HCV sequences in Nexus format. If you haven't finished that step, go back and do so now. Also note that your file name may be different than `hcv.nexus` - if so, then substitute your own file name in the following commands.

```
cp /home/student/parsimony/hcv.nexus hcv.nexus
```

```
nedit hcv.nexus &
```

This file contains an alignment (in nexus format) of 41 Hepatitis C virus (HCV) sequences isolated from 5 different patients. Sequences are named in the following way:

```
Patient_Time_Clone
```

For instance, the sequence labeled 1_1_5 was isolated from patient number 1 at time point 1 and is clone number 5 from that patient and that time point.

Question 1

- **Start the paup program and load the data file:**

```
paup hcv.nexus
```

This command opens the PAUP* program and automatically executes the nexus file at the same time.

- **Define the outgroup:**

```
outgroup 2_1_1 2_1_2 2_1_3 2_1_4 2_1_5 2_1_7 2_1_8 2_1_9 2_1_10
```

```
set root=outgroup outroot=monophyl
```

This puts all nine sequences from patient 2 in the outgroup, and ensures that the outgroup is printed as a monophyletic sister group to the ingroup. This will help make the tree-plots clearer.

- **Enable PAUP* to store an unlimited number of trees:**

```
set increase=auto
```

Normally PAUP* will only store up to "maxtrees" trees in memory. This command allows maxtrees to be increased automatically (without prompting for user confirmation) if the need arises during the heuristic search.

- **Perform a heuristic search using TBR:**

```
hsearch start=stepwise addseq=random nreps=20 rseed=98367 swap=TBR
```

This command starts a heuristic search with tree-rearrangements of the TBR type, where the initial tree is constructed using sequential addition where sequences are added in random order, and 20 different starting trees are tried.

After a brief processing time you will be back where you ended last Wednesday. Among a total of approximately 10^{60} possible trees, PAUP* has found about 200 equally parsimonious best trees. Two hundred may sound like a depressingly large number of alternative reconstructions but as you will now see, these trees do in fact have a lot in common.

- **Question:** What is the length of the best trees?

You entered:

Your Answer

Score

Explanation

355



1.00

Total

1.00 / 1.00

Question 2

- **Convert trees to rooted form:**

```
roottrees
```

Above we have specified an outgroup and requested that trees be plotted with a root determined by this outgroup. However, the trees that we found by heuristic searching are still unrooted, and we need to explicitly specify that we want them to be rooted. Placement of the root is of course done on the basis of the outgroup.

- **Inspect resulting trees individually:**

```
describetrees 37/plot=cladogram label=no
```

This shows you one randomly picked tree (tree number 37) among the >200 best trees that were found by the heuristic search. (The option `label=no` turns off labeling of the internal nodes in the tree). Make sure that your Terminal window is wide enough that the tree plot fits (if it is not: Widen the window and run the command again). Notice how the viral sequences from each individual patient group together. This shows that while there is considerable diversity in the viral population within any single patient, those viruses are nevertheless more closely related to each other than to viruses from other patients. This is of course a result of the viruses in one patient all having descended from the virus that originally infected that patient. Plotting the tree with branch lengths may make this clustering more apparent:

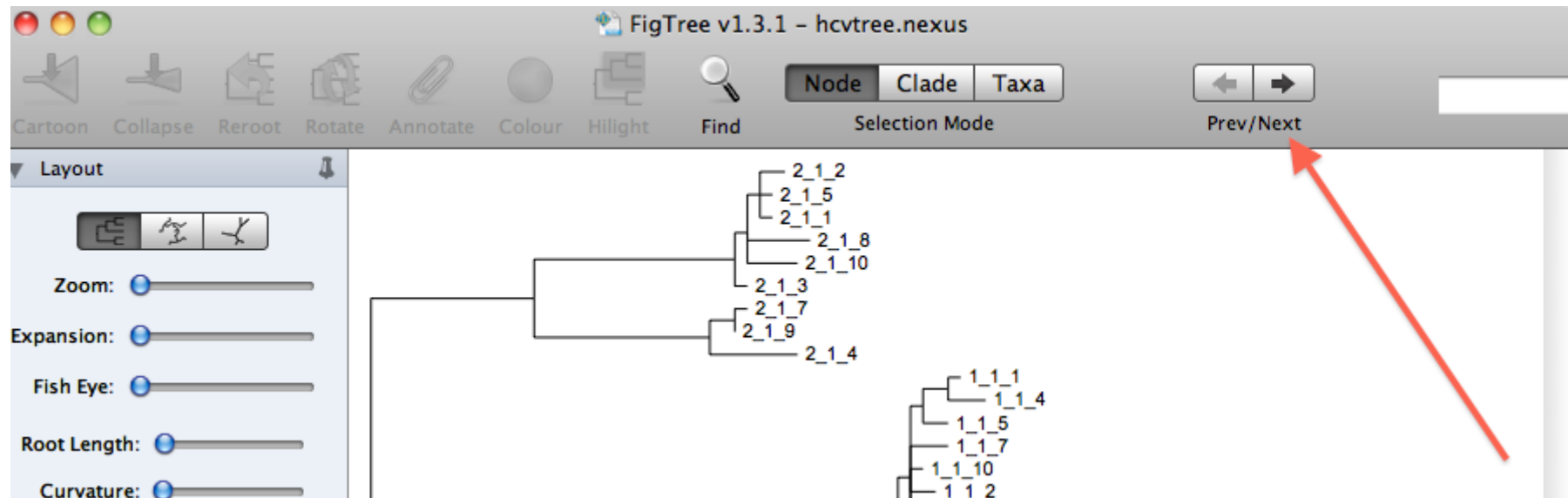
```
describetrees 37/plot=phylogram label=no
```

Remember: you also have the option of saving one or more trees to file and then viewing the tree using FigTree. For instance, you save tree number 37 by the following command:

```
savetrees file=hcvtree.nexus brlens=yes from=37 to=37
```

You can also save a range of trees of course.

To see whether this phenomenon is limited to the tree we selected first, save a range of 10 trees to a file and then inspect them in FigTree. Notice that when more than one tree is opened in FigTree you can use the small arrows labeled "Prev/Next" to move between trees:



- **Construct a consensus tree :**

You should now be convinced that the more than 200 equally good trees do in fact have quite a lot in common. Importantly it seems that all trees have viruses from individual patients grouped separately (forming five monophyletic groups). In order to investigate this question we will now construct a majority rule consensus tree summarizing the branching patterns in all the >200 trees:

```
contree all /strict=no majrule=yes percent=50
```

This constructs a consensus tree showing monophyletic groups occurring in more than 50% of all trees. Scroll back to see the tree. At each internal node it is indicated how often the corresponding group (meaning all taxa descending from that internal node) was found in the set of all trees. (Numbers are percentages). The option `percent=50` specifies that we want to see only groups occurring at least 50% of the time (i.e., we are requesting a "majority rule consensus"). You can increase this value (not lower it) if you want to set a different cutoff.

You will note that there are some sub-trees where the branching order is now unresolved, meaning that three or more taxa all split out from the same internal node. These multifurcations show that while more than 50% of the individual trees had those taxa together as a group (the precise number is indicated at the internal node), different trees nevertheless disagreed on the exact branching order *within* that group.

As you can see, consensus trees are a handy way of summarizing the evidence shared in a set of trees, and they are therefore useful when a search identifies several good reconstructed phylogenies.

- **Question:** Are all the sequences for patient 1 grouped in the consensus tree?

Your Answer		Score	Explanation
No			
Yes	✓	1.00	
Total		1.00 / 1.00	

Question 3

- In what fraction of the original trees was this the case for patient 1 (this is the percentage written at the internal node at the basis of that patient's group of sequences)?

Your Answer		Score	Explanation
<input type="radio"/> 50 %			
<input type="radio"/> 25 %			
<input checked="" type="radio"/> 100 %	✓	1.00	
<input type="radio"/> None			
<input type="radio"/> 67 %			
Total		1.00 / 1.00	

Question 4

- Are all the sequences for patient 5 grouped in the consensus tree?

Your Answer		Score	Explanation
No			
Yes	✓	1.00	
Total		1.00 / 1.00	

Question 5

- In what fraction of the original trees was this the case for patient 5 (this is the percentage written at the internal node at the basis of that patient's group of sequences)?

Your Answer		Score	Explanation
<input type="radio"/> 60 %			
<input type="radio"/> 25 %			
<input type="radio"/> 10 %			
<input checked="" type="radio"/> 100 %	✓	1.00	
<input type="radio"/> None			

Total

1.00 / 1.00

Question 6

- Are all the sequences for patient 7 grouped in the consensus tree?

Your Answer**Score****Explanation**

No

Yes



1.00

Total

1.00 / 1.00

Question 7

- In what fraction of the original trees was this the case for patient 7 (this is the percentage written at the internal node at the basis of that patient's group of sequences)?

Your Answer**Score****Explanation**☐ None☐ 25 %☐ 50 %

☒ 100 %

1.00

☐ 10 %

Total

1.00 / 1.00