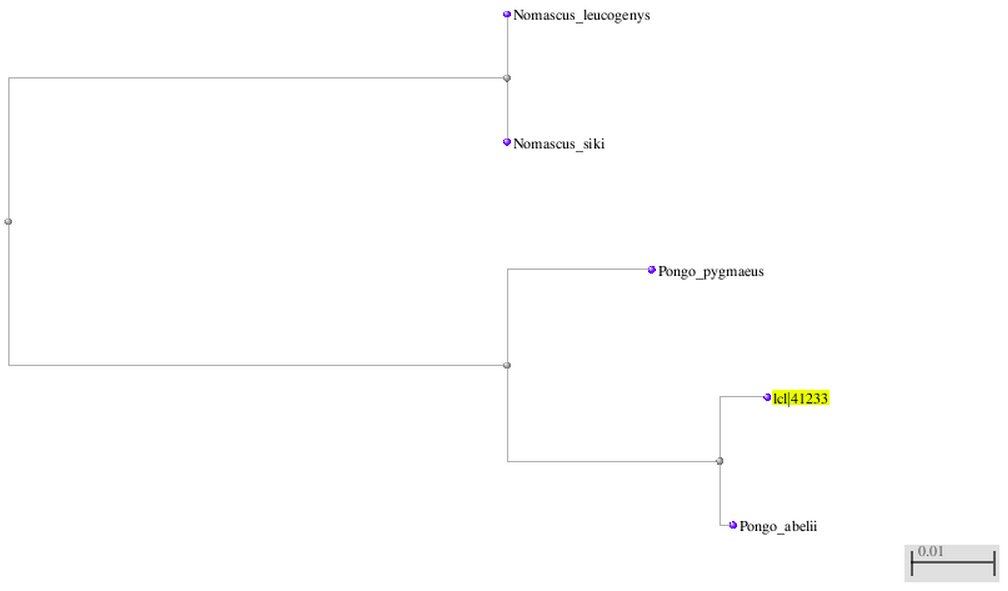
Karan Sikka 1/31/2014

02-250 Homework 2

1. Top five matches:
   1. Pongo abelii (Sumatran orangutan)
   2. Pongo pygmaeus (Bornean orangutan)
   3. Homo heidelbergensis (Extinct species of Homo)
   4. Nomascus leucogenys (Northern white-cheeked gibbon)
   5. Nomascus siki (Southern white-cheeked gibbon)
2. All of the top ten hits provide statistically significant alignments, because their E-values are extremely small (approximately zero).
3. It would increase the E-values, since the search space would be much larger without the mitochondrion keyword, and E-value is proportional to the size of the search space.
4. Leaving out the mitochondrion keyword would increase the amount of time we need to wait. The reason is that more seeds alignments would be found without the keyword, and it would require more work to complete their alignments than if those seeds filtered out.
5. Let,  
   E be the E-value  
   P be the P value   
   m be the length of the query  
   n be the sum of the lengths of the sequences in the database  
   then,  
   E = mn \* P  
   so to obtain P,  
   P = E / (mn),  
   Since E is 0.0, P is 0.0 regardless of the quantity (mn).
6. Note that the *Homo heidelbergensis* is not shown since its max seq distance > 0.75, which is the max that the tree making tool allows.
7. The organism is an unknown species in the genus *Pongo*, aka it is an Orangutan. Notably, the species is more similar to the Sumatran variety, *Pongo abelii*, than the Bornean variety, *Pongo pigmaeus*.
8. The suspect is an Orangutan. It may be upto ~1.4m tall, with orange-ish hair/fur, and may weigh up to 90 kg.
9. cytochrome c oxidase subunit II [Pongo abelii]  
   The associated gene is called COX2  
   score = 267  
   ident = 92%  
   E-value = 2 x 10^-86
10. The function of this gene is to produce cytochrome c oxidase subunit II, which is part of cytochrome c oxidase, which is an enzyme involved in the electron transport chain process of aerobic respiration.