

# EEOB Final Project: Phylogenetic Analysis of BRCA1: Exon 11 in Certain Primates

Cameron Vannoy, Kaitlyn Barnes, and Rachel Rompala

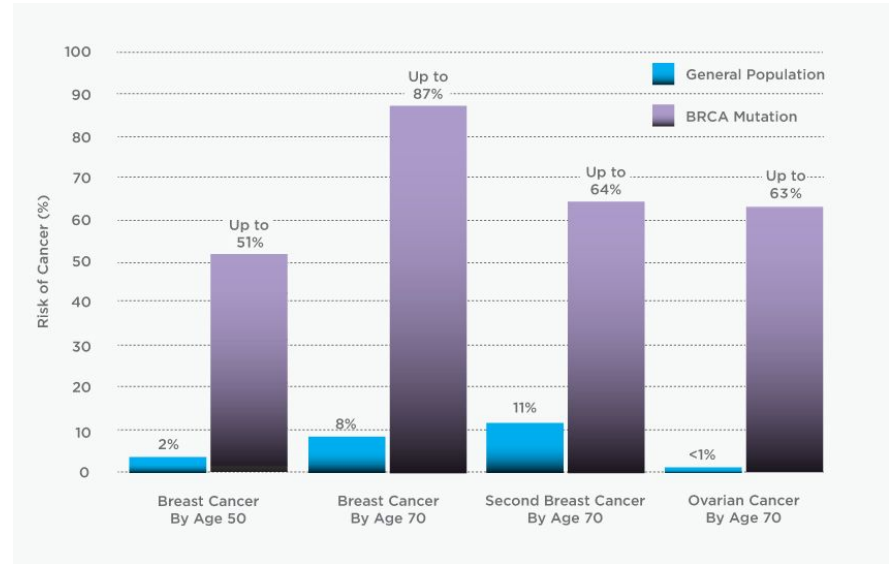




# Introduction: The BRCA1 Gene and Importance of Exon 11

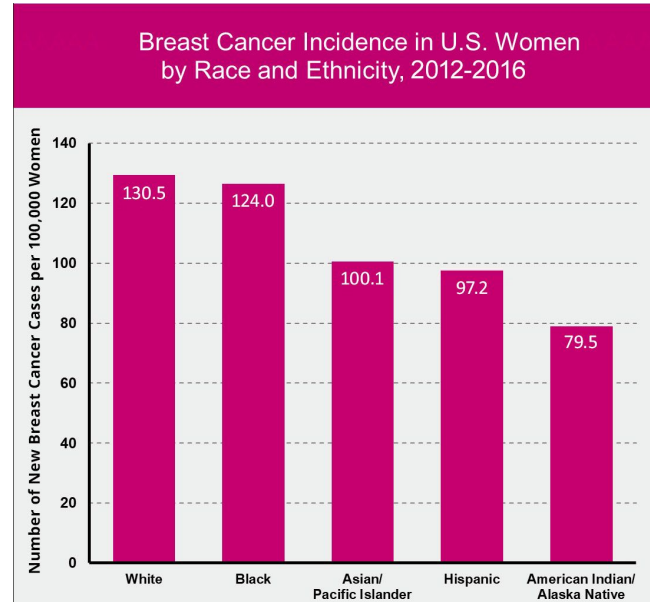
# BRCA1 and Breast Cancer

- BRCA1 is a tumor suppressing gene on chromosome 17 that functions to repair damaged DNA
- BRCA1 variants create an increased risk of the development of breast and ovarian cancers
- Increased risk of triple negative breast cancer (more aggressive) with BRCA1 mutant
- Likelihood of developing breast cancer from a harmful BRCA1 variant increases with age



# Prevalence of Breast Cancer in the US

- 1 in 8 women will develop breast cancer throughout her lifetime
- 1 in 833 men will develop breast cancer throughout his lifetime
- Breast cancer accounts for 30% of cancers diagnosed in women
- Sex (female) and age (older) are the two biggest risk factors



# BRCA 1: Exon 11

- Encodes over 60% of BRCA1 coding sequence
- Exon 11 is highly variable
- Exon 11 variants within BRCA1 are found in most familial breast cancer



# Research Question

# Research Question

Which primate is most closely related to humans when looking at exon 11 in BRCA1 tumor suppressor gene?

If the gene relationship follows the species relationship trend, then we expect chimpanzee BRCA1: Exon 11 to show the closest phylogenetic relationship to human BRCA1: Exon 11.

# Phylogenetic Analysis





# Mafft Alignment

# Sequence Alignment Using MAFFT

- We used an online resource to complete the protein multiple sequence alignment.
- <https://www.ebi.ac.uk/Tools/msa/mafft/>
- This alignment was needed to proceed with the analysis and for the creation of a distance matrix
- The alignment was performed using the BLOSUM62 matrix with a 1.53 gap open penalty, 0.123 gap extension, maxiterate of 2, no performed FFTS, and a tree rebuilding number of 2.



Tools > Multiple Sequence Alignment > MAFFT

## Multiple Sequence Alignment

MAFFT (**M**ultiple **A**lignment using **F**ast **F**ourier **T**ransform) is a high speed multiple sequence alignment program.



# Distance Analysis using PAUP

# Distance Analysis

To complete a distance analysis we first needed to use the previously made alignment file to create a distance matrix.

This was done using PAUP.

- Execute alignment file
- showdist

```
paup> showdist
```

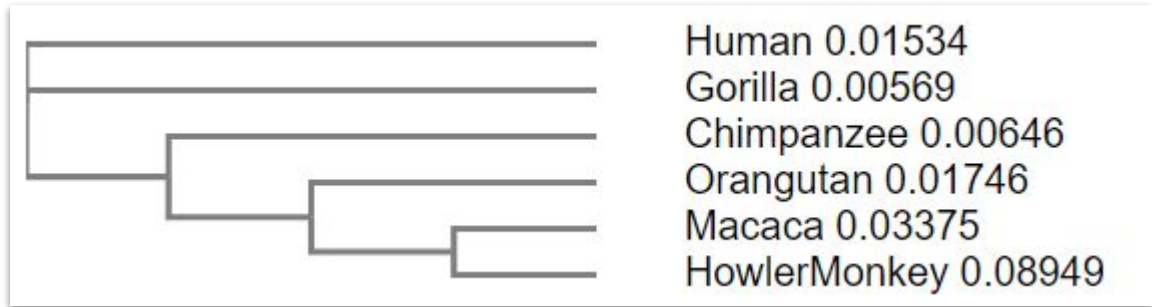
```
Uncorrected ("p") distance matrix  
1801 characters are included  
All characters have equal weight
```

		1	2	3	4	5	6	
1	AAB61673	1	-					
2	AAC39584	1	0.02103	-				
3	AAC39583	1	0.02279	0.01227	-			
4	AAC39585	1	0.03856	0.02892	0.02892	-		
5	AAC39586	1	0.07375	0.06497	0.06497	0.07024	-	
6	AAC39587	1	0.12841	0.12137	0.12049	0.12665	0.12324	-

# Neighbor-Joining Analysis: Unrooted Consensus Tree

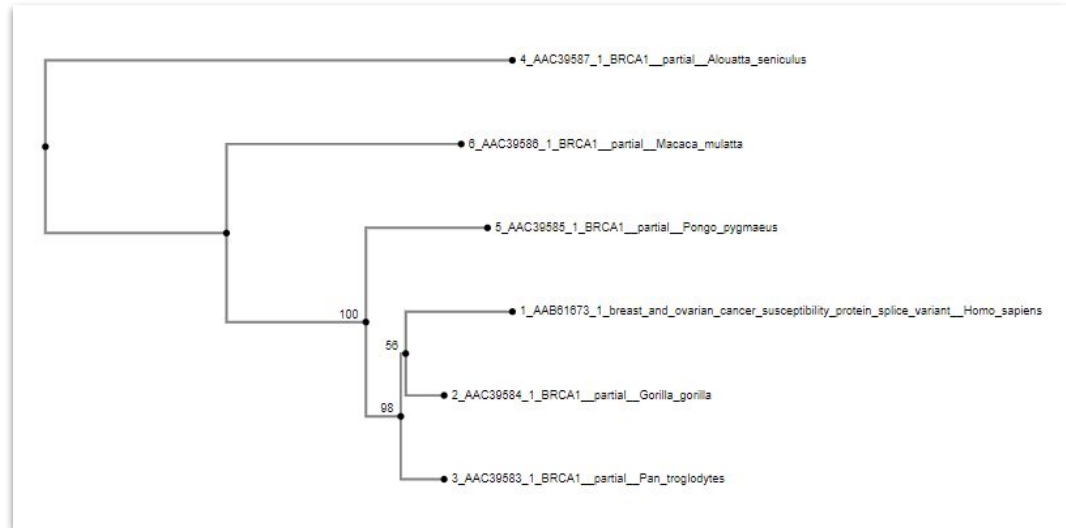
For the NJ Analysis, the human exon 11 sequence was used as the outgroup.

This tree shows the distances between primates when using the previously computed distance matrix and no bootstrapping.



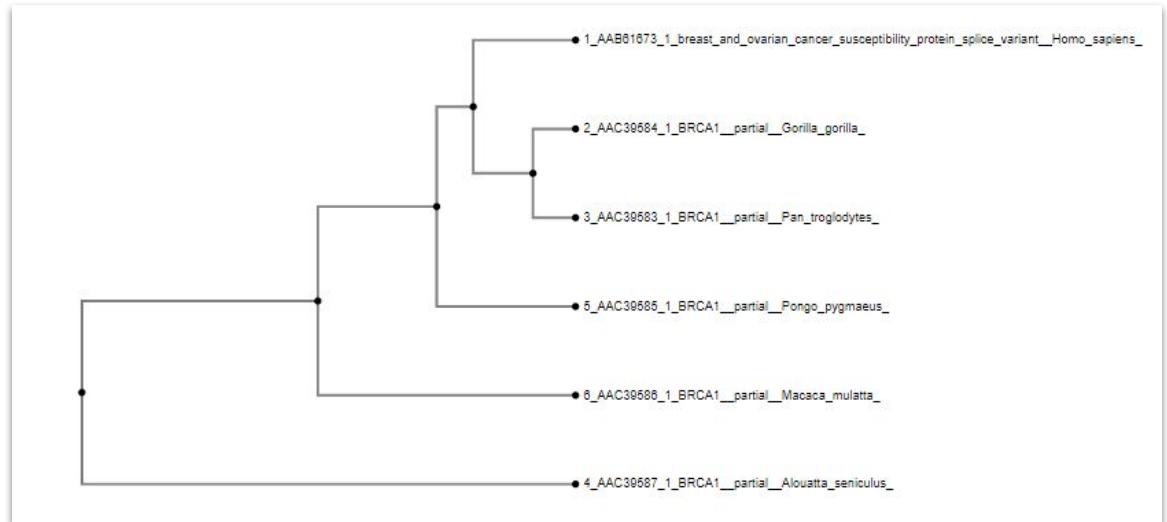
# Neighbor-Joining Analysis: Minimum Evolution, 1000 Bootstrap Replicates

We repeated the NJ analysis, but with this replicate we used 1000 bootstrap replicates and were looking for minimum evolution.



# UPGMA Tree

Using the same distance matrix we completed a UPGMA analysis and built a UPGMA tree in PAUP.



# NJ (Bootstrapping) Minimum Evolution Tree vs UPGMA tree:

These rooted trees differ at one node. In the bootstrapping tree, gorillas and humans are more closely related and share a common ancestor with chimpanzees.

In the UPGMA tree, gorillas and chimpanzees are more closely related and share a common ancestor with humans.

The UPGMA is most likely incorrect as it assumes the rate of evolution to be constant for each taxon which we know to be a false assumption.





# Bayesian Analysis

# Run Information

Likelihood of best state for "cold" chain of run 1 was -7014.75

Likelihood of best state for "cold" chain of run 2 was -7014.75

Chain swap information for run 1:

	1	2	3	4
1		0.88	0.78	0.68
2	166760		0.89	0.79
3	166224	166472		0.90
4	166389	166566	167589	

Chain swap information for run 2:

	1	2	3	4
1		0.88	0.78	0.68
2	166826		0.89	0.79
3	167105	166675		0.90
4	166153	166938	166303	

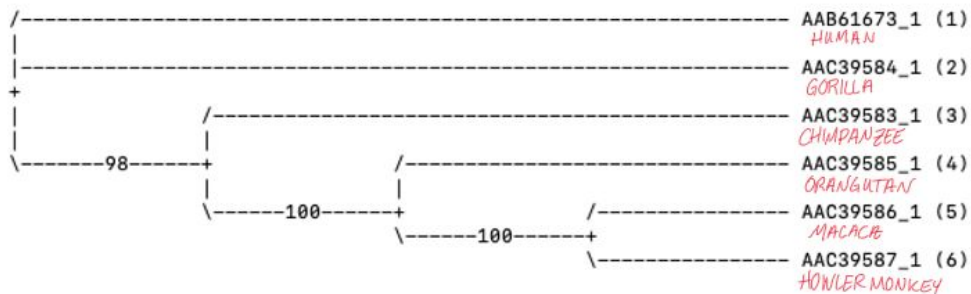
Upper diagonal: Proportion of successful state exchanges between chains  
Lower diagonal: Number of attempted state exchanges between chains

# Bayes Trees

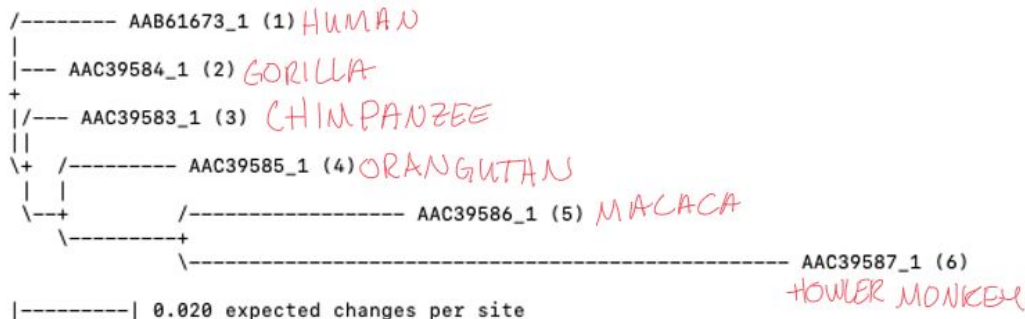
Bayesian analysis looks for the highest posterior probability between clades being analyzed.

Analysis done using MrBayes.

Clade credibility values:



Phylogram (based on average branch lengths):



0.020 expected changes per site

Calculating tree probabilities...

Credible sets of trees (3 trees sampled):

99 % credible set contains 2 trees

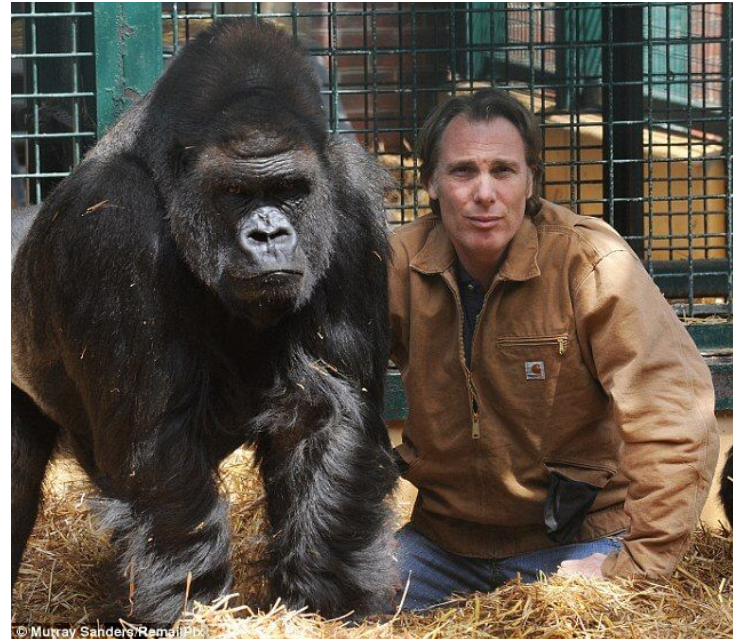
# Conclusion

# Findings

Humans were most closely related to gorillas -- not chimpanzees! @ BRCA1: exon 11

Low support (56) for humans and gorillas

Potential coalescence OR gene tree/species tree differences



# Error and Potential Changes

- Throughout this analysis there was the potential to create error through incorrect or improper models, accidental distortion of data through file damage, or human error in interpretation.
- If we were to do this project again, we would like to use corrected p-distance values as this increases the validity of our trees (and thus our analysis).
- Rooting the tree from the Howler Monkey would give more accurate results



Questions?



# Thank you!

If you have further questions please contact us:

Cameron Vannoy, [vannoyscm@iastate.edu](mailto:vannoyscm@iastate.edu)

Rachel Rompala, [rrompala@iastate.edu](mailto:rrompala@iastate.edu)

Kaitlyn Barnes, [kgbarnes@iastate.edu](mailto:kgbarnes@iastate.edu)

