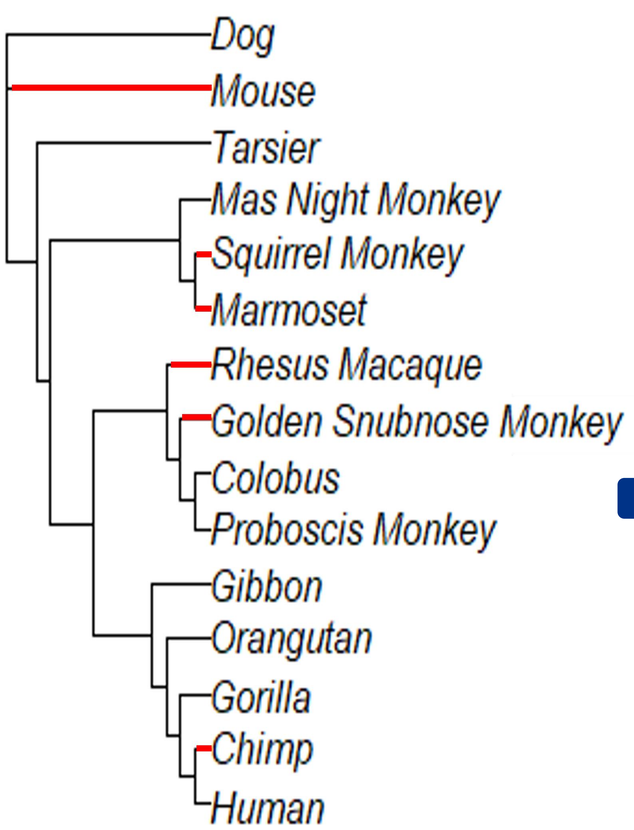
Genome-wide search for genes influenced by sexual selection in primates

Brianna Ports and Michael Jensen-Seaman
Dept. of Biological Sciences, Duquesne University, Pittsburgh, PA

Background

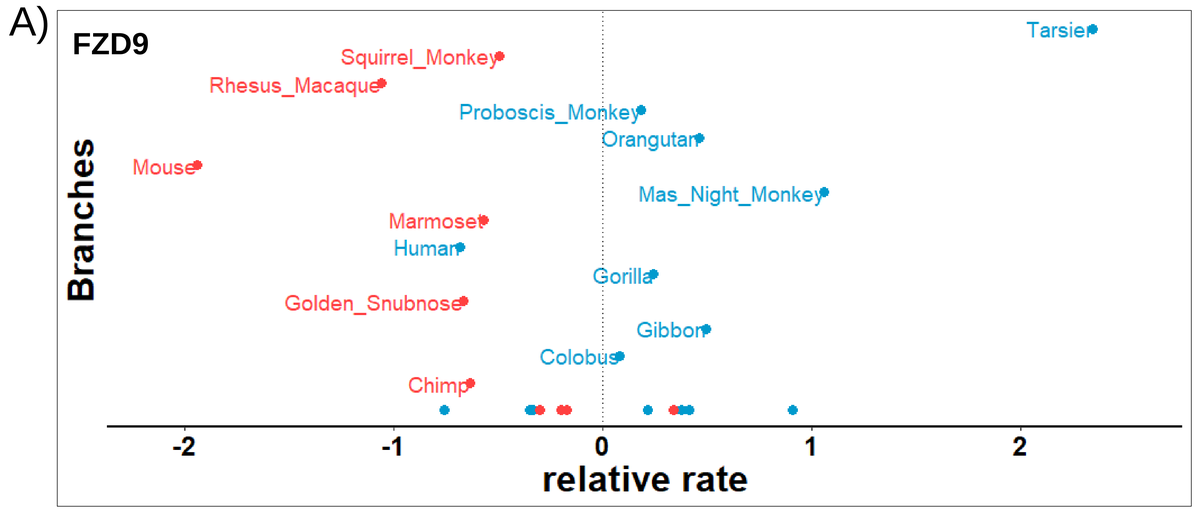
- Sexual selection occurs when a trait increases the ability of an organism to mate and produce offspring.[1]
- Sperm competition is a form of sexual selection where the ejaculates of multiple males compete to fertilize an individual ovum.[2]
- Primates are an exciting model because of the wide variety of mating types and levels of sperm competition between closely related species.[3]
- An often used bioindicator of sperm competition is the relationship between testes and body weights, with larger testes to body weight ratios suggesting higher levels of postcopulatory competition.[4]

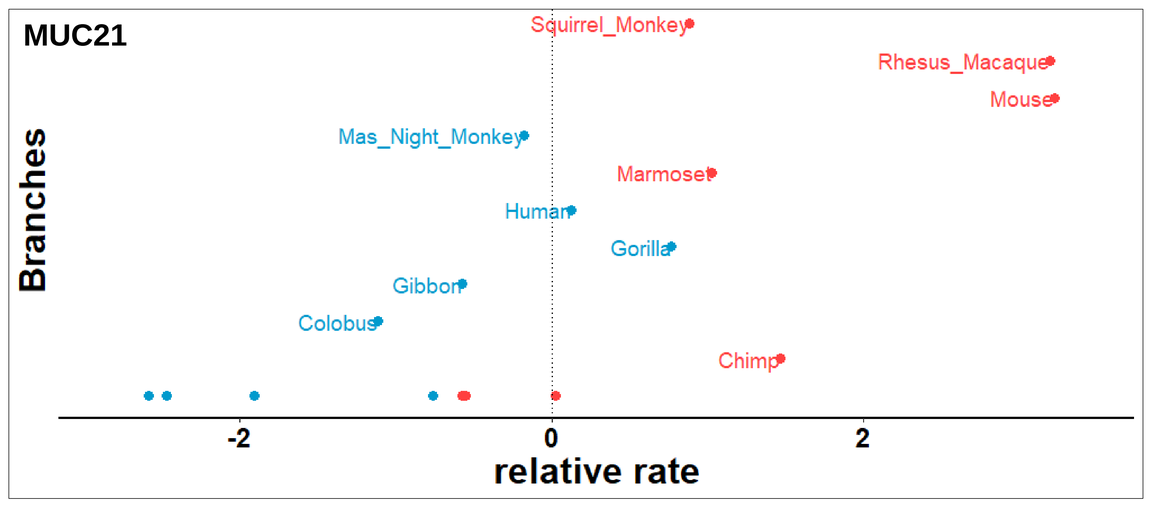
Proteome-Wide Average Tree



Phylogenetic tree representing the average evolutionary rate for the amino acid sequences of 18,412 protein-coding genes generated using the PAML program. Branches of species with combined testes weights greater than .15% of the male body mass are shown in red.

Relative Evolutionary Rate Analysis





Relative evolutionary rate analysis allows for the identification of genes which are evolving due to selective pressures induced by a specific phenotype. Here we detected 49 genes which were evolving significantly different in species with high (red) and low (blue) testes to body weight ratios. The graphs above represent the relative evolutionary rates of the most conserved (A) and most accelerated gene (B) in the high testes to body weight grouping.

Phylogenetic Models of Codon Evolution

	Branch Site Models			Site Models	
Gene	dN/dS for foreground	-2ΔlnL	p-value	-2ΔlnL	p-value
MUC21	3.70999	49.756582	**<0.00001**	116.451914	**<0.00001**
RPGRIP1	2.01936	4.848276	**0.027675**	2.080462	0.149201
TMX1	1.61352	1.992962	0.158038	0	1
PGLYRP4	1.37249	0.354004	0.551858	0.287718	0.591698
CD96	1.03827	0.019122	0.66	4.227364	**0.039779**
CXorf40B	1.03824	0.020388	0.88	0	1
C6orf15	1.84618	2.633044	0.89	2.670256	0.102243
MSN	1.16397	0.014534	0.904055	2E-06	1
TOMM70	1.07298	0.012708	0.910273	2.2E-05	1
SEC24B	1	-2E-06	1	4.00339	**0.04541**

Results of Branch-Site and Site models using the Phylogenetic Analysis of Maximum Likelihood (PAML) program. Branch-Site Models are used to identify genes which are evolving under positive selection. While site models identiy genes where individual sites are evolving differently. Genes with significant p-values are represented in bold.

Conclusions

- 49 genes are evolving differentially in species with high levels of sperm competition
- 15 genes are evolving under purifying selection
- 34 genes which are rapidly evolving in high sperm competition species
  - 2 genes are evolving under *positive selection* (MUC21 and RPGRIP1)

Here we have identified a number of genes which have evolved in response to the selective pressures induced by sperm competition

Future Directions

- Use Gene Set Enrichment Analysis to identify pathways which are enriched within this dataset
- Use previously published expression data in humans and mice to look for patterns of expression in significant genes
- Identify genes which have become pseudogenized, either through complete gene loss or premature stop codons, in species with high testes to body weights

Acknowledgements

- NIH grant 1 R15 GM123447-01 to M. Jensen-Seaman
- Duquesne University Biology Department
- The Bayer School for Natural and Environmental Sciences

Technical help in Python and R:
Dr. Nathan Clark and Colin Kessler

Literature Cited

1. Kingan, S. B., Tatar, M. & Rand, D. M. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *J. Mol. Evol.* (2003).
2. Dixson, A. F. Copulatory and Postcopulatory Sexual Selection in Primates. *Folia Primatol. (Basel).* **89,** 258–286 (2018).
3. Dixson, A. F. *Primate Sexuality*. (Oxford University Press, 2012).
4. Soulsbury, C. D. Genetic Patterns of Paternity and Testes Size in Mammals. *PLoS One* **5,** e9581 (2010).
5. Clark, N. L. & Aquadro, C. F. A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. *Mol. Biol. Evol.* **27,** 1152–61 (2010).
6.Chikina, M., Robinson, J. D. & Clark, N. L. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. *Mol. Biol. Evol.* **33,** 2182–92 (2016).
7. Yang, Z. & Dos Reis, M. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* (2011).

Methods

Dataset: Amino acid sequences of full protein-coding genomes were downloaded from the UCSC 30 species Mammals Multiz Alignment & Conservation with 30 mammalian species, including 27 primates. 15 species were removed from the dataset to eliminate redundant species (ex: crab-eating, pigtailed, and rhesus macaques), species with poor alignments, and species with large sequencing gaps. Additionally, for each gene, species with less than 80% of the gene intact were removed. The RER analysis was completed on 18,412 protein-coding genes where 13,850 included all 15 species. Individual gene trees with the same overall topology were generated using the PAML program (described below). Testes to body weight ratios were identified for each species using published data and were used in the RER analysis.
Relative Evolutionary Rate Analyses: The analysis were be performed in R using the RERconverge package. This package, when given a series of gene trees, will estimate a master tree with an estimated average rate of change for all genomic elements present. RERconverge then estimates the relative evolutionary rates (RERs) for each branch in each gene, meaning that each branch is normalized by its species on the master tree as well as the rate for that gene across species. RERs are generated using linear regression weighted by the testes to body weight ratio of each species and p-values were corrected using a FDR of 15%.[5,6]
PAML: After RER analysis, I aimed to differentiate between positive selection and pseudogenization in high testes to body weight ratio species. This was detected using the CODEML program in the PAML package on the corresponding DNA sequence of each protein-coding gene. Additional testing identified if individual sites were under increased levels of selection. Chi2 testing was used to identify significance at a p-value of 0.05.[7] **For more detail on methods, please contact Brianna Ports at portsb@duq.edu**

DUQUESNE UNIVERSITY

Department of Biological Sciences