Relating Selection of p53 to Longevity and Sexual Maturity in Mammals

Joshua Schaaf

Abstract:

Cancer rates increase with age. In mammal species with longer maximum lifespans, the selective pressure of cancer should increase. p53 is a well-known tumor suppressor protein in humans and has different homologues in many species of animals. Here, I hypothesized that changes in mammalian lifespan positively correlate with positive selection of p53. Utilizing open-source sequence data from NCBI and Ensembl, I used dN/dS and maximum likelihood tree building to determine different p53 evolution rates. I then compared these rates to the maximum lifespan, sexual maturity times, and the ratio of these two of the different species of mammals. I believed signatures of positive selection in p53 with changes in maximum lifespan would be consistent with p53's role as a tumor suppressor protein in mammals that undergo senescence. However, dN/dS values indicate a lack of positive selection at the whole-protein scale. Surprisingly, there is a negative correlation between leaf dN/dS values and longevity, possibly indicating more positive selection in mammals with a short lifespan and faster sexual maturity, opposing the original hypothesis. This may be due to p53 being important in signaling for cellular self-sacrifice, or apoptosis, and the drawback for being the guardian against cancer is having increased rates of aging.

Keywords:

- dN/dS
- positive selection
- longevity
- sexual maturity

Background:

Tumor protein 53, tp53, also known as p53, is a protein that regulates the cell cycle, as well as apoptosis (23). p53 has a known network of genes that work together to keep cells from becoming cancerous. This network, including p53, modulates cellular autophagy, which involves the degradation of a cell's own components through lysosomal activity (23). p53 is an important tumor suppressor, and its loss in mice has been shown to cause a significant amount of tumor development (23).

There have been studies that show mutations deleterious to p53 tumor suppressor function also increase organism longevity (22, 24). This seems to align with findings here, however, there have been conflicting publications about the role of p53 in the aging process and longevity, as well as inconsistent results on aging and longevity after manipulating p53 through different strategies (22). For instance, transgenic mice with constitutively higher expressions of p53 with different approaches show increased cancer resistance but premature aging phenotype (22). 'Super p53' mice with an extra copy of the Trp53 gene showed increased cancer resistance, however, no sign of premature aging (22). These conflicting studies only increase the importance of researching p53.

To further our understanding of p53, I used computational phylogenetic analysis on p53 orthologs and their species' longevity and sexual maturity. There is a large range of species-

specific longevity in Mammalia, with there being a 32-fold difference between longevity in just my dataset alone. I decided looking at the relationship between p53 and sexual maturity was beneficial, despite sexual maturity and longevity being so strongly correlated. Sexual maturity gives more insight onto when genes are passed to offspring, so I thought it may have had a different relationship to p53 than longevity. Finally, taking the ratio between longevity and sexual maturity gives insight to how early in life the species is sexually mature.

The goal of this research was to analyze hypothesized connections between longevity and sexual maturity to signatures of evolution in mammalian p53. Connections like these can inspire and inform ongoing research on p53's importance in cellular, organism, and population levels.

Results:

No significant positive selection was found in any of the twelve mammalian species analyzed here. However, dN/dS correlated with both longevity and maturity. Both factors correlated negatively with their leaf dN/dS ratios, indicating species that lived longer had more signatures of p53 purifying selection. The ratio of longevity to sexual maturity was found to correlate positively with leaf dN/dS ratios, indicating species that have offspring earlier in their lives had more signatures of p53 positive selection.

Discussion:

These results contradict the hypothesis, specifically, longevity and signatures of positive selection correlate negatively **(Figure 5, a)**. On top of this, the connection of dN/dS to the previously described longevity/sexual maturity ratio shows that species that reproduce earlier in life have more signatures of positive selection **(Figure 5, c)**. Maximum likelihood alignment of p53 orthologs in 12 mammalian species produced an unrooted phylogenetic tree **(Figure 1, b)**. Both nucleotide and protein alignments were accurate, and provided with dN/dS values, which were then placed onto the phylogenetic tree **(Figure 1, c)**. At first, this seemed backwards, but it seems to follow some previous research about p53 codon 72 Pro allele in humans (3). It is discussed that p53 plays an important role in preventing cancer, however, may actively suppress longevity during old age. The correlation between both longevity and the longevity/sexual maturity ratio to dN/dS seems to also fit with the "grandmother hypothesis" of humans, whales, and elephants, as mothers at older ages may maximize their fitness by investing resources into their living children rather than continuing to reproduce (9). This would explain the increased purifying selection of p53 in species with greater longevity, as p53 would be deleterious for grandmothers trying to help their kin.

Conclusion:

p53 shows signatures of purifying selection in mammals, specifically heightened purifying selection in mammals with increased longevity, as well as a later offspring production period. This seems to fit the "grandmother hypothesis" in anthropology, as well as a few publications describing the effects of p53 polymorphisms on aging phenotypes. Future analyses may include population-level selection studies, p53-network gene selection studies, or possibly analyzing changes in p53 over human evolution employing ancient DNA. Relationships

discovered in this research can be used in the continuation of understanding p53 function on cellular, organism, and population levels.

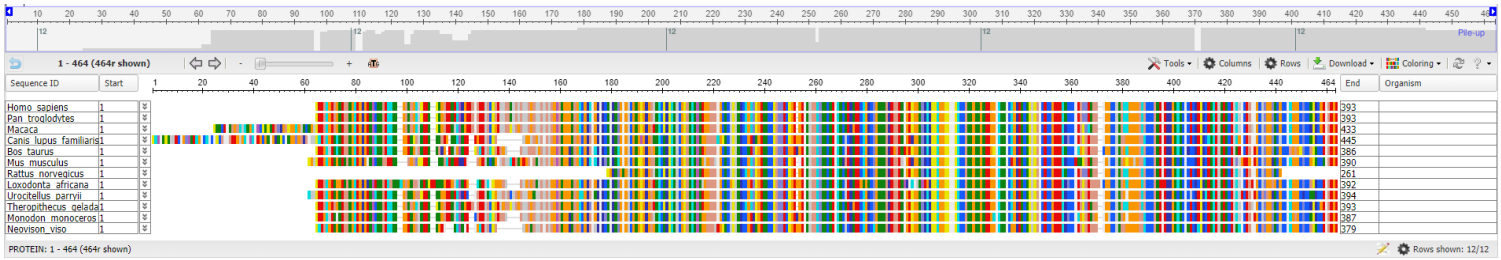
Methods:

Ensembl was used to find orthologous genes between the mammalian species (19). The gene code for TP53 is ENSG00000141510. Once the protein and nucleotide data were downloaded and placed into separate fasta files, each file was aligned by maximum likelihood method with ClustalW in python (19, 12) **(Figure 1, a)**. With the aligned files, I used PhyML to create the phylogenetic tree (4). A sequence file created by pal2nal containing the codon alignment of p53 orthologs was then used (16). With the tree file and pal2nal file, CodeML was used to obtain dN/dS values using model 1:b, amino acid distance as 0 (equal), and codon frequency at 1 (F1X4) (20) **(Figure 2, a+b)**. This model was decided upon after performing a log-likelihood ratio test between this model and model 0:one **(Figure 2, c)**. The longevity and female sexual maturity was acquired from AnAge for each species analyzed (17) **(Figure 3)**. Using the Newick string created earlier and the longevity/sexual maturity, I wrote a few R scripts using 'phytools' to obtain trees with continuous values mapped onto them **(Figure 4, a-d)**, as well as a phylogenetic tree and characteristic heatmap **(Figure 4, e)** to give some way to graphically grasp the information (13, 14). I then used R and libraries 'ape', 'geiger', 'nlme', and 'phytools' to obtain phylomorphospaces and correlations between dN/dS and longevity, maturity, and their ratio (8, 10, 11, 14) **(Figure 5)**. The GLS analysis to achieve a line of best fit was specifically from the 'nlme' library, with the Brownian correlation structure being from the 'ape' library **(Figure 5b)**. Finally, I was looking at estimating ancestral character states to try and check to see if the distances on the tree could accurately describe the relationships of longevity and sexual maturity between species. I'm pretty sure this is wrong, especially when making the 'true' values… Anyways, restricted maximum likelihood (REML) was used to predict ancestral character states, and then I plotted the 'true' states to the estimated states (8) **(Figure 6)**.

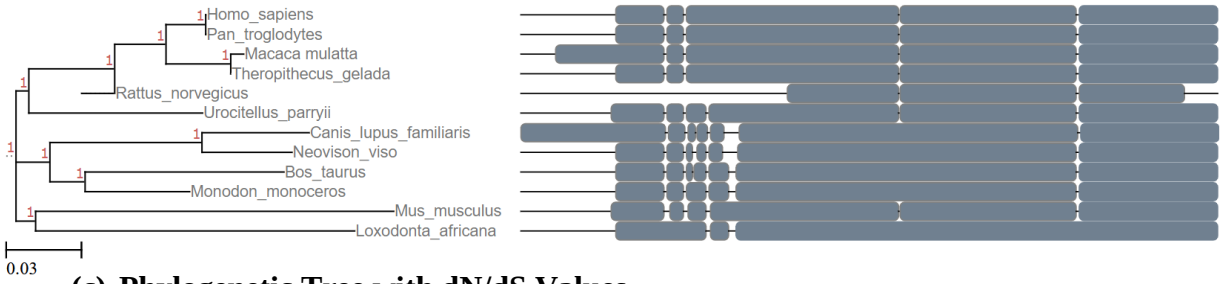
Figures/Tables:

Figure 1: **Alignment and dN/dS**

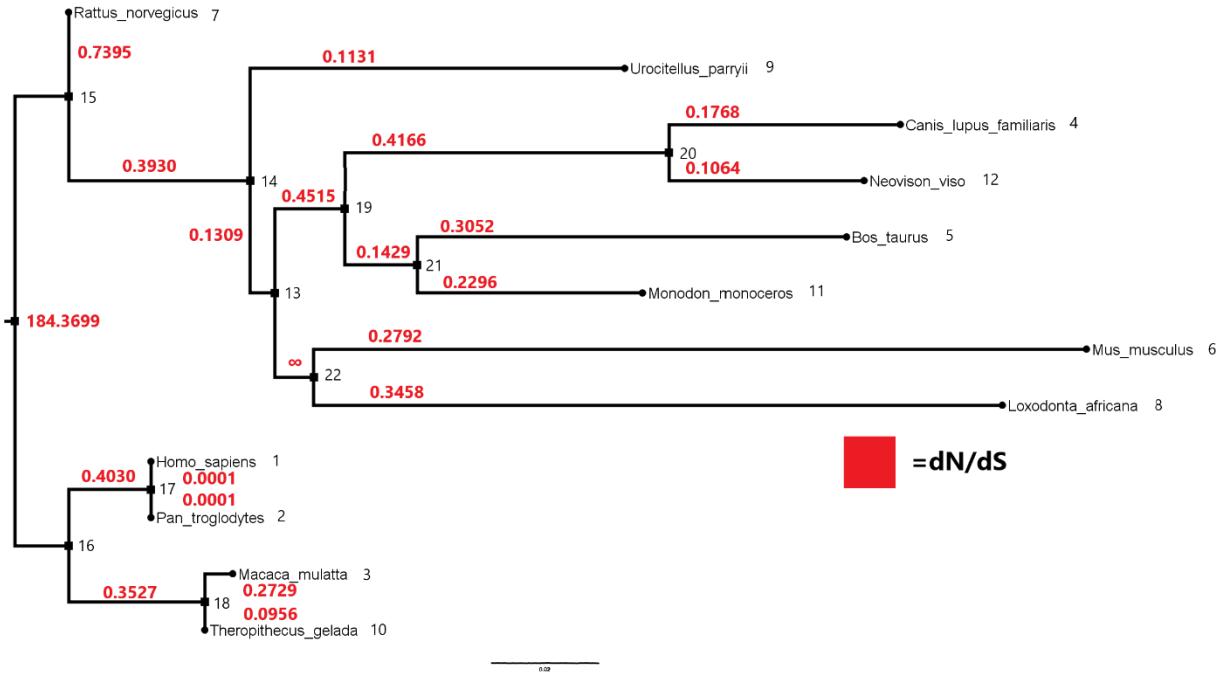
(a) NCBI MSA Viewer of p53 Orthologs



(b) Tree and Protein Alignment



(c) Phylogenetic Tree with dN/dS Values



(a) Alignment of p53 orthologous genes using NCBI's multiple sequence alignment viewer. Amino acids were aligned by the maximum likelihood method with ClustalW. **(b)** Tree and protein alignment using the same ML alignment in a. **(c)** A phylogenic tree made with distances between orthologs. dN/dS from CodeML for each edge is displayed in red. Sequences were obtained from Ensembl's ortholog gene-based display (ENSG00000141510).

Figure 2: CodeML

(a) CodeML model 0:one

branch	t	N	S	dN/dS	dN	dS	N*dN	S*dS
13..14	0.029	969.5	422.5	0.2591	0.0051	0.0196	4.9	8.3
14..15	0.233	969.5	422.5	0.2591	0.0417	0.1608	40.4	67.9
15..16	0	969.5	422.5	0.2591	0	0	0	0
16..17	0.029	969.5	422.5	0.2591	0.0052	0.0202	5.1	8.5
17..1	0	969.5	422.5	0.2591	0	0	0	0
17..2	0.008	969.5	422.5	0.2591	0.0014	0.0054	1.4	2.3
16..18	0.091	969.5	422.5	0.2591	0.0162	0.0625	15.7	26.4
18..3	0.009	969.5	422.5	0.2591	0.0017	0.0064	1.6	2.7
18..10	0.013	969.5	422.5	0.2591	0.0024	0.0091	2.3	3.8
15..7	0.039	969.5	422.5	0.2591	0.0069	0.0267	6.7	11.3
14..9	0.347	969.5	422.5	0.2591	0.0619	0.2391	60.1	101
13..19	0.094	969.5	422.5	0.2591	0.0168	0.0648	16.3	27.4
19..20	0.312	969.5	422.5	0.2591	0.0556	0.2146	53.9	90.7
20..4	0.206	969.5	422.5	0.2591	0.0367	0.1418	35.6	59.9
20..12	0.244	969.5	422.5	0.2591	0.0435	0.1678	42.2	70.9
19..21	0.108	969.5	422.5	0.2591	0.0192	0.0743	18.7	31.4
21..5	0.276	969.5	422.5	0.2591	0.0492	0.1899	47.7	80.2
21..11	0.077	969.5	422.5	0.2591	0.0137	0.0527	13.2	22.3
13..22	0.049	969.5	422.5	0.2591	0.0087	0.0336	8.4	14.2
22..6	0.593	969.5	422.5	0.2591	0.1059	0.4086	102.6	172.6
22..8	0.394	969.5	422.5	0.2591	0.0702	0.2711	68.1	114.5

(b) CodeML model 1:b

branch	t	N	S	dN/dS	dN	dS	N*dN	S*dS
13..14	0.026	969.3	422.7	0.1309	0.0029	0.0218	2.8	9.2
14..15	0.228	969.3	422.7	0.393	0.0518	0.1319	50.3	55.8
15..16	0	969.3	422.7	184.3699	0	0	0	0
16..17	0.029	969.3	422.7	0.403	0.0066	0.0163	6.4	6.9
17..1	0	969.3	422.7	0.0001	0	0	0	0
17..2	0.008	969.3	422.7	0.0001	0	0.0089	0	3.8
16..18	0.092	969.3	422.7	0.3527	0.0196	0.0556	19	23.5
18..3	0.009	969.3	422.7	0.2729	0.0017	0.0062	1.6	2.6
18..10	0.014	969.3	422.7	0.0956	0.0012	0.0122	1.1	5.2
15..7	0.037	969.3	422.7	0.7395	0.0113	0.0153	10.9	6.5
14..9	0.377	969.3	422.7	0.1131	0.0371	0.3283	36	138.8
13..19	0.09	969.3	422.7	0.4515	0.022	0.0487	21.3	20.6
19..20	0.29	969.3	422.7	0.4166	0.0679	0.163	65.8	68.9
20..4	0.209	969.3	422.7	0.1768	0.0288	0.1629	27.9	68.8
20..12	0.269	969.3	422.7	0.1064	0.0253	0.2375	24.5	100.4
19..21	0.118	969.3	422.7	0.1429	0.014	0.0979	13.6	41.4
21..5	0.267	969.3	422.7	0.3052	0.0527	0.1726	51.1	73
21..11	0.081	969.3	422.7	0.2296	0.0134	0.0585	13	24.7
13..22	0.029	969.3	422.7	999	0.0139	0	13.5	0
22..6	0.596	969.3	422.7	0.2792	0.1114	0.3989	107.9	168.6
22..8	0.396	969.3	422.7	0.3458	0.0838	0.2423	81.2	102.4

$$\lambda_{LR} = -2\left(\ell(\theta_0) - \ell(\hat{\theta})\right)$$

$$\lambda_{LR} = -2(-6134.90 + 6110.36)$$

$$\lambda_{LR} = 49.08$$

$$p = 2.995e-04$$

(c)

Figure 2 **(a)**: CodeML model using model 0 (one), with a fixed w (dN/dS). A sequence file created by pal2nal containing the codon alignment of p53 gene orthologs was used. A newick string of ML aligned p53 orthologous protein sequences was submitted as the tree file. The a value is fixed at 0, and the sequence type was codons. The amino acid distance was set to 0 (equal), and the codon frequency was set to 1 (F1X4). **(b)**: CodeML model using model 1 (b), with a variable w (dN/dS). The same sequence file and tree file was used. The a value is fixed at 0, and the sequence type was codons. The amino acid distance was set to 0 (equal), and the codon frequency was set to 1 (F1X4). **(c)**: Likelihood-ratio test for log likelihood of the nested models, giving a p value < 0.05 (variable w statistically accurate).

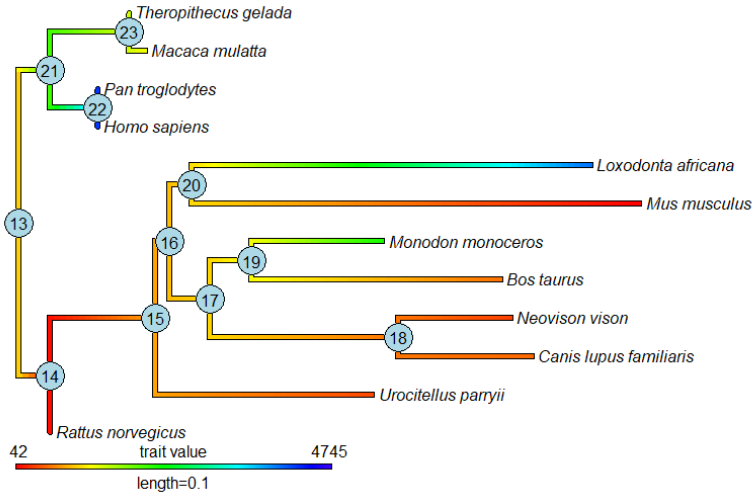
Figure 3: Longevity and Female Sexual Maturity

	Female Sexual Maturity (days)	Longevity (Years)	Longevity/Sexual Maturity
Homo sapiens	4745	122.5	9.423076923
Pan troglodytes	2920	59.4	7.425
Macaca mulatta	1231	40	11.8602762
Theropithecus gelada	1391	36	9.446441409
Rattus norvegicus	90	3.8	15.41111111
Urocitellus parryii	365	10	10
Canis lupus familiaris	510	24	17.17647059
Neovision viso	334	11.4	12.45808383
Bos taurus	548	20	13.32116788
Mondon monoceros	2191	50	8.329529895
Mus musculus	42	4	34.76190476
Loxodonta africana	4018	65	5.904678945

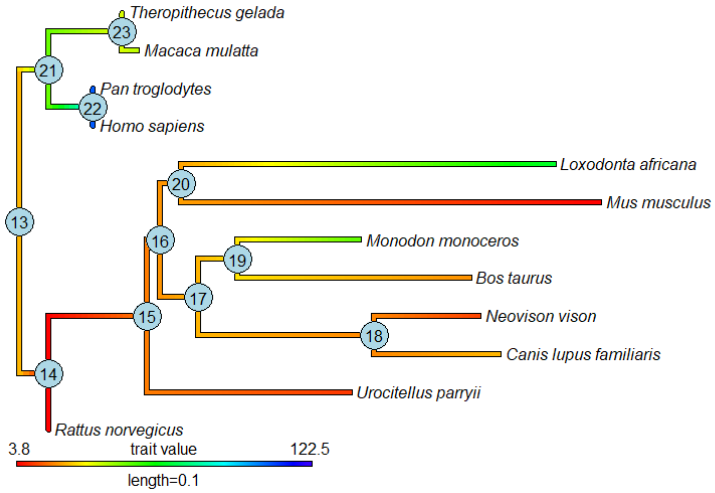
Each species and the female's sexual maturity, maximum longevity, and a ratio of their longevity and sexual maturity. The ratio of longevity to sexual maturity may be more useful than just sexual maturity, as it describes, in a sense, how early the sexual maturity is in respect to the maximum longevity of the animal. This may allow for increased insight into the evolutionary relationship between p53, reproduction, and longevity.

Figure 4: Continuous Variables Mapped to Phylogenetic Trees

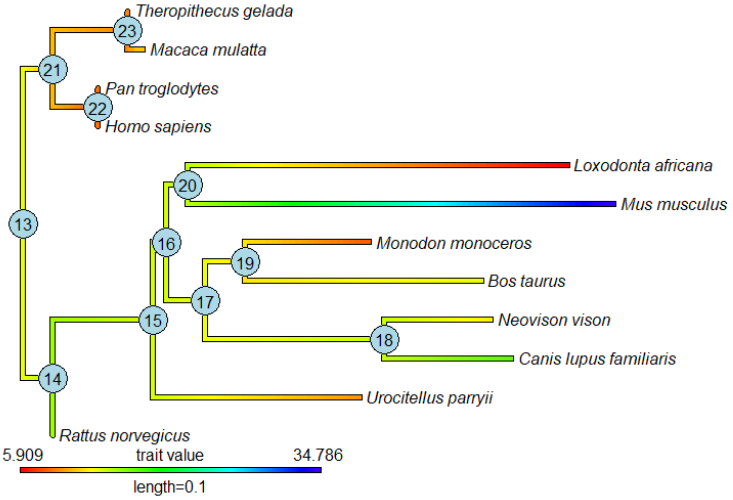
(a) Female Sexual Maturity (Days)



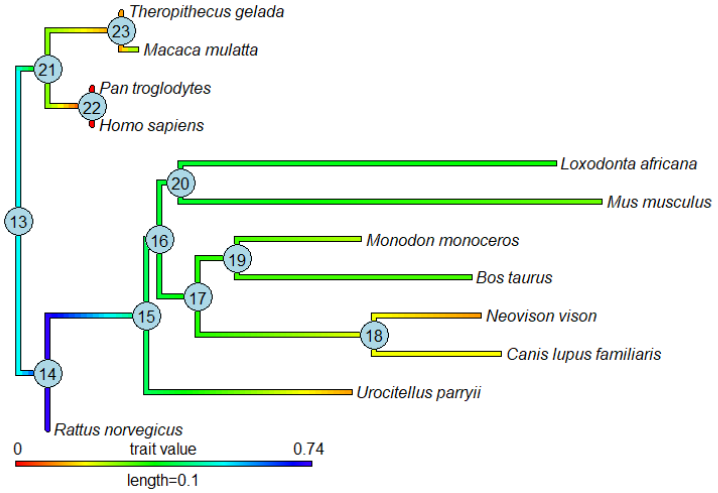
(b) Longevity (Years)



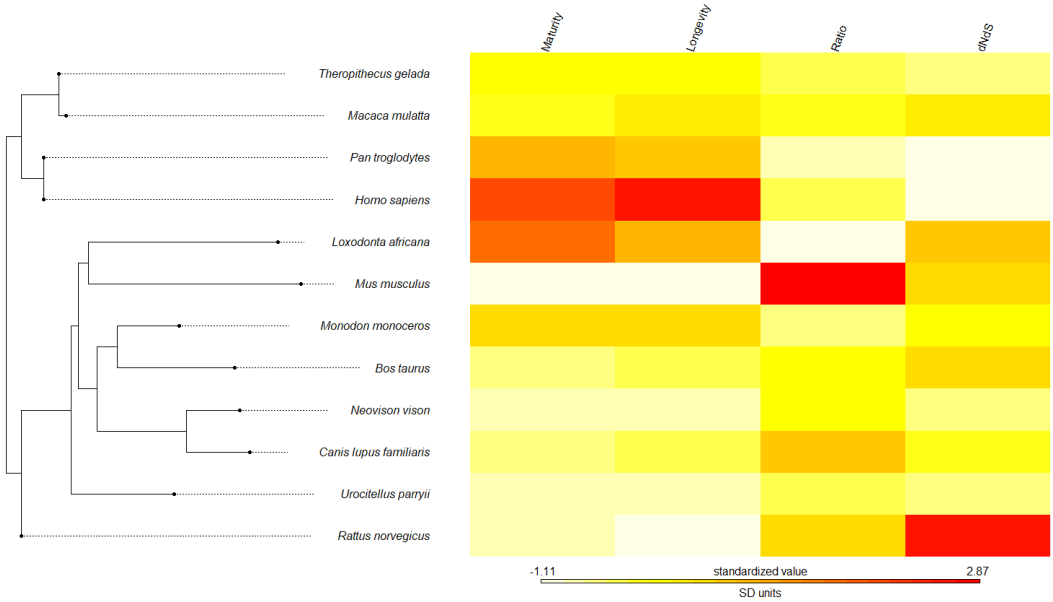
(c) Ratio (L/SM)



(d) dN/dS (Leaves)



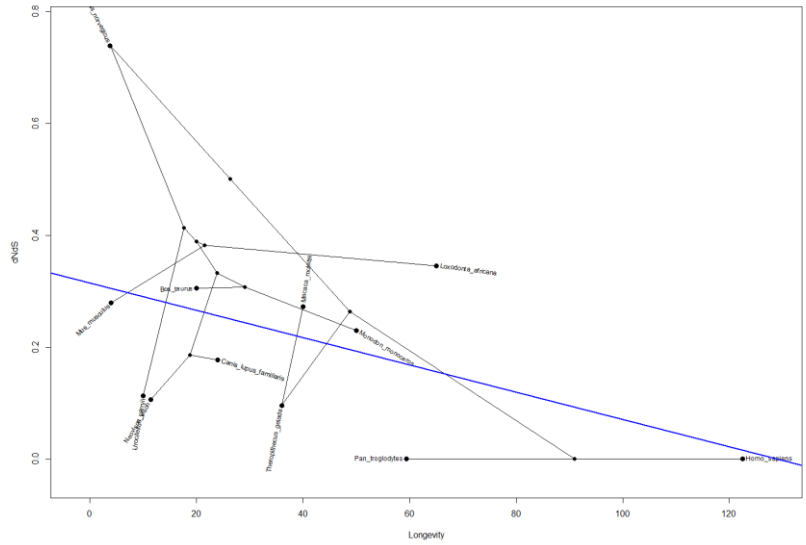
(e) Phylogenetic Tree and Characteristic Heatmap



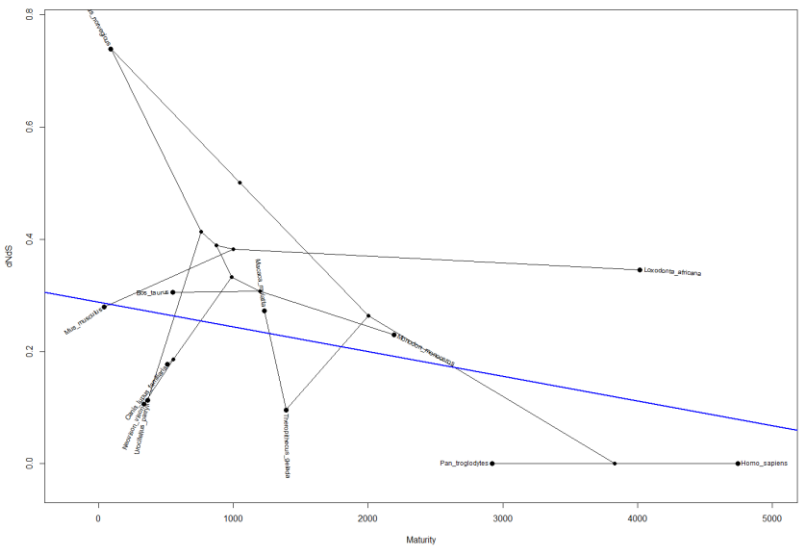
(a), (b), (c), and **(d)** are the p53 protein phylogenetic tree with their respective continuous values mapped (female sexual maturity (a), longevity (b), longevity and sexual maturity ratio (c), and leaf dN/dS (d)). The numbers are for internal nodes to refer to for **figure 6**, and the trait values are mapped in color, with cooler colors such as blue being larger numbers, and warmer colors such as red being smaller numbers. **(e)** shows the standardized values for each characteristic of each species alongside another phylogenetic tree. All of figure 4 was created in R with the use of the package "phytools".

Figure 5: Phylomorphospaces and Correlation

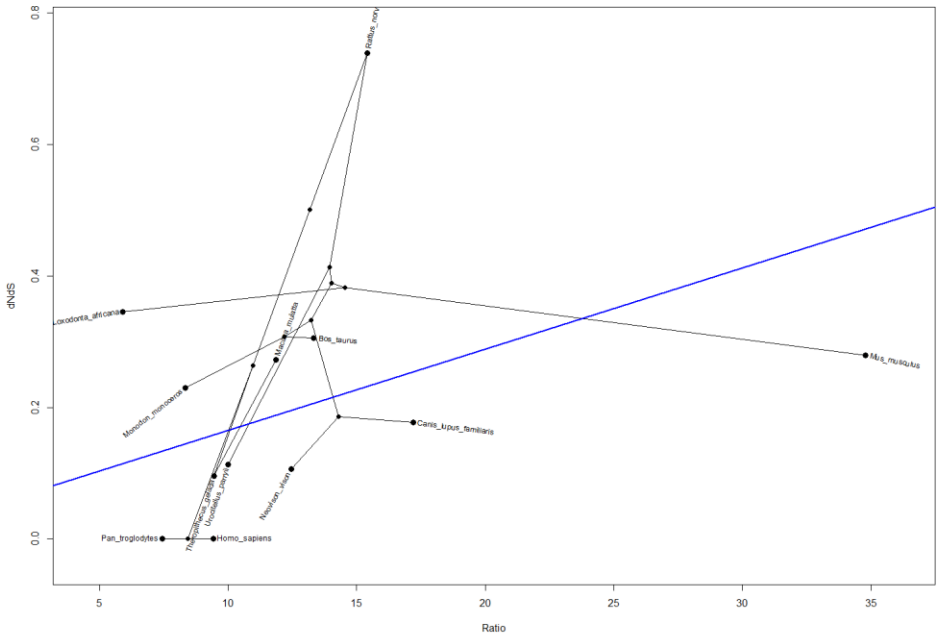
(a) Relating dN/dS and Longevity



(b) Relating dN/dS and Maturity



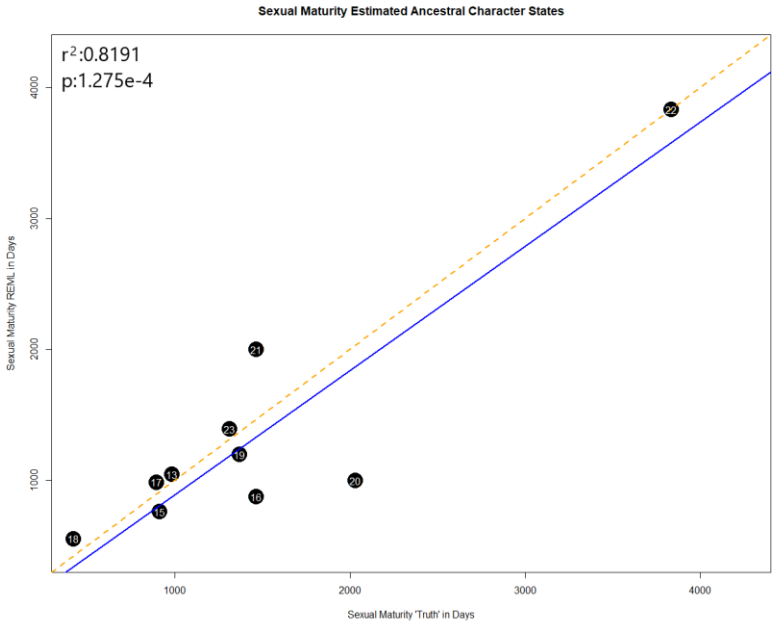
(c) Relating dN/dS and Ratio (L/SM)



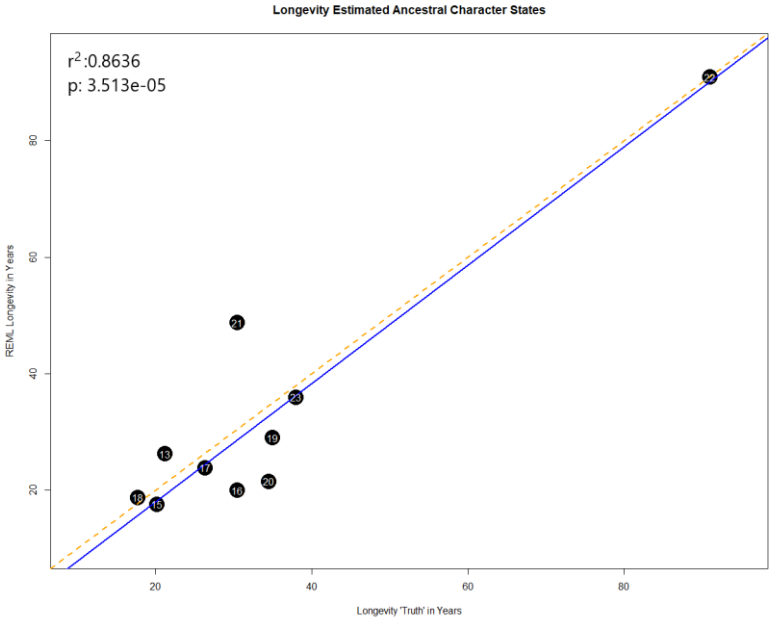
(a), (b), and **(c)** are Phylomorphospaces comparing tip dN/dS values to longevity, sexual maturity, and the ratio of the two (L/SM), respectively. All three datasets gave $p < 0.001$ for a phylogenetic generalized least square fit by maximum likelihood, represented by the blue line. This fit used a Brownian correlation structure across the tree. **(a)** shows a negative correlation between dN/dS and longevity of the species. **(b)** shows a negative correlation between tip dN/dS and sexual maturity of the species. Finally, **(c)** shows a positive correlation between tip dN/dS values and the ratio of sexual maturity to longevity of the species. Graphs were generated with an array of libraries in R.

Figure 6: Estimated Ancestral Character States

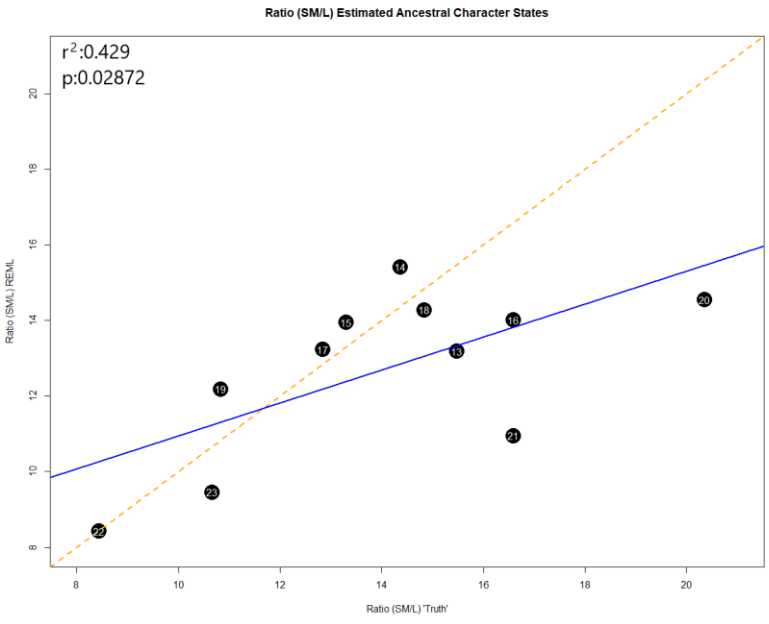
(a) Female Sexual Maturity



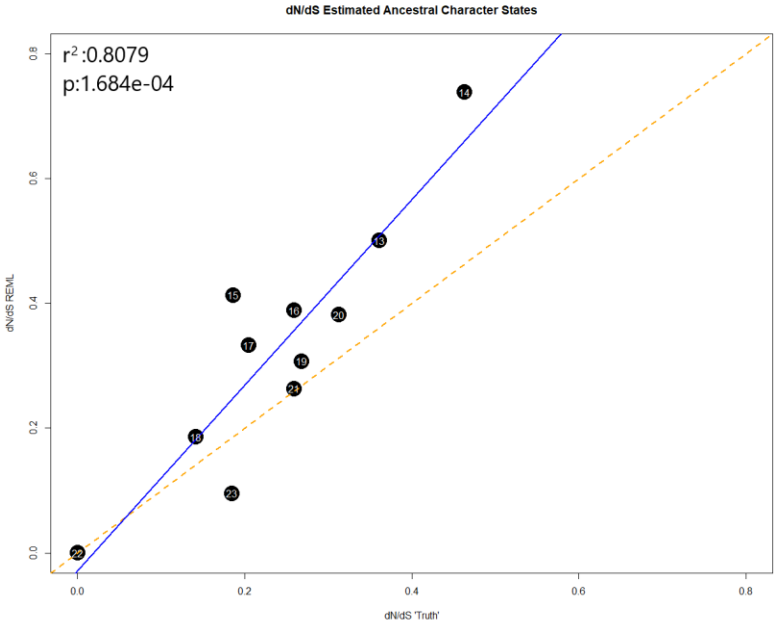
(b) Longevity



(c) Ratio (L/SM)



(d) dN/dS



(a), (b), (c), and **(d)** show the estimated ancestral character states by Restricted Maximum Likelihood (REML) on the y-axis, and the 'true' values on the x-axis. The dashed orange line is what a perfect estimate would look like, where the solid blue line is a simple linear regression of the points. The r^2 and p values are for the solid blue linear regression line. The 'true' values were achieved by taking the average value of the two 'child' nodes coming from each internal node. Model and graphs were used and generated in R with the use of the package "ape".

Figure 5b Statistical Summaries

(a) dN/dS and Longevity

```
Generalized least squares fit by maximum likelihood
  Model: dnds ~ longevity
  Data: NULL
       AIC        BIC    logLik
  -1.850637 -0.3959173 3.925319

Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
                Value  Std.Error   t-value p-value
(Intercept)  0.3146300 0.15448949  2.036579   0.069
longevity   -0.0024411 0.00028493 -8.567300   0.000

 Correlation:
          (Intr)
longevity -0.073

Standardized residuals:
       Min         Q1        Med         Q3        Max
-0.63803274 -0.49777805 -0.07278744  0.15392785  1.53546182

Residual standard error: 0.2827463
Degrees of freedom: 12 total; 10 residual
```

(b) dN/dS and SM

```
Generalized least squares fit by maximum likelihood
  Model: dnds ~ sexualMaturity
  Data: NULL
       AIC       BIC    logLik
  -0.6060229 0.848697 3.303011

Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
                     Value  Std.Error   t-value p-value
(Intercept)     0.28818271 0.16251171  1.773304  0.1066
sexualMaturity -0.00004412 0.00000546 -8.073547  0.0000

 Correlation:
               (Intr)
sexualMaturity -0.053

Standardized residuals:
       Min         Q1        Med         Q3        Max
-0.5609437 -0.4639227 -0.1441699  0.1328753  1.5288581

Residual standard error: 0.2977961
Degrees of freedom: 12 total; 10 residual
```

(c) dN/dS and Ratio

```
Generalized least squares fit by maximum likelihood
  Model: dnds ~ ratio
  Data: NULL
       AIC      BIC     logLik
  7.90636 9.36108 -0.9531798

Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
               Value  Std.Error  t-value p-value
(Intercept) 0.04094188 0.23387456 0.175059  0.8645
ratio       0.01238218 0.00238355 5.194838  0.0004

 Correlation:
       (Intr)
ratio -0.146

Standardized residuals:
       Min         Q1        Med         Q3        Max
-0.4533101 -0.2352697 -0.1344139  0.2093554  1.1955550

Residual standard error: 0.4245763
Degrees of freedom: 12 total; 10 residual
```

(a), (b), and **(c)** show the output of the Phylogenetic Generalized Least Squares (PGLS) analysis. Correlation structure was from the "ape" library in R. GLS function was from the "nlme" library in R.

Citations:

1. Belyi, V. A., et al. “The Origins and Evolution of the p53 Family of Genes.” *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 6, 2009, doi:10.1101/cshperspect.a001198.

2. Cock, P. J. A., et al. “Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics.” Bioinformatics, vol. 25, no. 11, 2009, pp. 1422–1423., doi:10.1093/bioinformatics/btp163.

3. Donehower, Lawrence A. “p53: Guardian AND Suppressor of Longevity?” *Experimental Gerontology*, vol. 40, no. 1-2, 2005, pp. 7–9., doi:10.1016/j.exger.2004.10.007.

4. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." Systematic Biology, 59(3):307-21, 2010.

5. Jaime Huerta-Cepas, Francois Serra, and Peer Bork. ETE 3: Reconstruction, analysis and visualization of phylogenomic data. Mol Biol Evol 2016; doi: 10.1093/molbev/msw046

6. “NCBI Multiple Sequence Alignment Viewer 1.18.1.” National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/projects/msaviewer/.

7. NCBI Team. “Database Resources of the National Center for Biotechnology Information.” *Nucleic Acids Research*, vol. 44, no. D1, 2015, doi:10.1093/nar/gkv1290.

8. Paradis E, Schliep K (2019). “ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.” Bioinformatics, 35, 526-528.

9. Pavard, Samuel, et al. “Senescence of Reproduction May Explain Adaptive Menopause in Humans: A Test of the ‘Mother’ Hypothesis.” American Journal of Physical Anthropology, vol. 136, no. 2, 2008, pp. 194–203., doi:10.1002/ajpa.20794.

10. Pennell M, Eastman J, Slater G, Brown J, Uyeda J, Fitzjohn R, Alfaro M, Harmon L (2014). “geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees.” Bioinformatics, 30, 2216-2218.

11. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2020). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-151, https://CRAN.R-project.org/package=nlme.

12. “Python Programming Language.” *Python.org*, www.python.org/.

13. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

14. Revell LJ (2012). “phytools: An R package for phylogenetic comparative biology (and other things).” Methods in Ecology and Evolution, 3, 217-223.

15. Sahm, Arne, et al. “Long-Lived Rodents Reveal Signatures of Positive Selection in Genes Associated with Lifespan.” *PLOS Genetics*, vol. 14, no. 3, 2018, doi:10.1371/journal.pgen.1007272.

16. Suyama, M., et al. “PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments.” *Nucleic Acids Research*, vol. 34, no. Web Server, 2006, doi:10.1093/nar/gkl315.

17. Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., de Magalhaes, J. P. (2018) "Human Ageing Genomic Resources: new and updated databases." Nucleic Acids Research 46(D1):D1083-D1090.

18. Tavernarakis, Nektarios, et al. “The Effects of p53 on Whole Organism Longevity Are Mediated by Autophagy.” *Autophagy*, vol. 4, no. 7, 2008, pp. 870–873., doi:10.4161/auto.6730.

19. Thompson, Julie D., et al. “CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice.” *Nucleic Acids Research*, vol. 22, no. 22, 1994, pp. 4673–4680., doi:10.1093/nar/22.22.4673.

20. Yang, Z. “PAML 4: Phylogenetic Analysis by Maximum Likelihood.” Molecular Biology and Evolution, vol. 24, no. 8, 2007, pp. 1586–1591., doi:10.1093/molbev/msm088.

21. Yates, Andrew D, et al. “Ensembl 2020.” *Nucleic Acids Research*, 2019, doi:10.1093/nar/gkz966.

22. Zhao, Yuhan, et al. “A Polymorphism in the Tumor Suppressor p53 Affects Aging and Longevity in Mouse Models.” ELife, Mar. 2018, doi:10.7554/elife.34701.001.

23. Zilfou, J. T., and S. W. Lowe. “Tumor Suppressive Functions of p53.” *Cold Spring Harbor Perspectives in Biology*, vol. 1, no. 5, 2009, doi:10.1101/cshperspect.a001883.

24. Ørsted, David Dynnes, et al. “Tumor Suppressor p53 Arg72Pro Polymorphism and Longevity, Cancer Survival, and Risk of Cancer in the General Population.” Journal of Experimental Medicine, vol. 204, no. 6, 2007, pp. 1295–1301., doi:10.1084/jem.20062476.