Patterns of Substitutions in Viral Interacting Proteins

SJ McGeady, Lawrence Uricchio
Department of Biology, Tufts University

Introduction

Over long evolutionary timescales, weakly deleterious alleles can accumulate and ultimately fix due to genetic drift. Other evolutionary mechanisms could also contribute to the fixation of deleterious alleles, such as transient changes in the sign or strength of selection. Here, we use an evolutionary SIR model to examine how epidemics may alter deleterious allele frequencies and fixation rates when these alleles confer a transient fitness benefit, resulting in changes in selection strength and sign. The model suggests that weakly deleterious alleles can fix at elevated rates due to transient epidemics, but strongly conserved alleles should rarely (if ever) fix. To investigate the plausibility of the model, we compare patterns of substitutions in Viral Interacting Proteins (VIPs; Enard 2016) to non-VIPs. In particular, we compare rates of substitution in VIPs to non-VIPs as a function of predicted conservation strength (phyloP score). We find mixed support for the model in a preliminary analysis.

Research Goal

VIPs are more conserved than average proteins in the human genome, but also harbor evidence for higher-than-average adaptation rates (Enard 2016). Our research asks whether it is possible for transient selection pressures (epidemics) to result increased rates of substitution at conserved sites in VIPs. This research is part of a broader initiative to integrate ecological modeling into evolutionary methods for inferring adaptation rates.

Model

We use a classic SIR (Susceptible (S), Infectious (I), Recovered (R)) and add allele frequencies for a resistance allele. The 0,1, and 2 superscripts represent the number of copies of a resistance allele carried by an individual. **D** represents the number of individuals that have succumbed to the epidemic. Given the change in frequency of the allele, we can calculate its expected fixation probability.

$$S^0(t)+S^1(t)+S^2(t)+I^0(t)+I^1(t)+I^2(t)+R^0(t)+R^1(t)+R^2(t)+D(t)=N$$

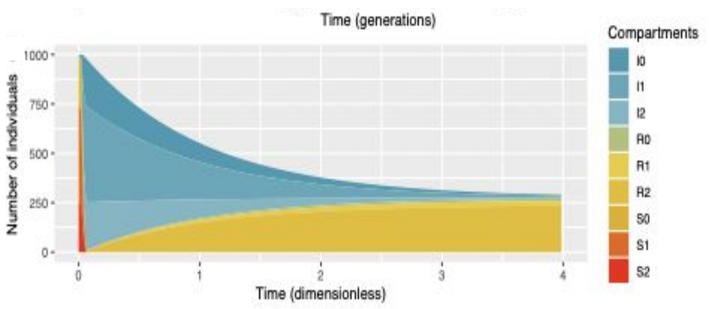

Figure 1: Predicted change in distribution of deleterious allele frequencies in a population after an epidemic beginning at Time = 0.

Data and results

We analyzed a previously published set of VIPs, substitutions along the human branch since our divergence from chimpanzees (Enard 2016), and a predicted measure of conservation (phyloP scores; Pollard 2010). We compared relative rates of substitution at VIPs to non-VIPs as a function of conservation strength, with negative phyloP representing non-conserved sites and positive phyloP representing conserved sites. **Figure 2** (below) describes the dataset. VIPs are slightly more conserved than non-VIPs.

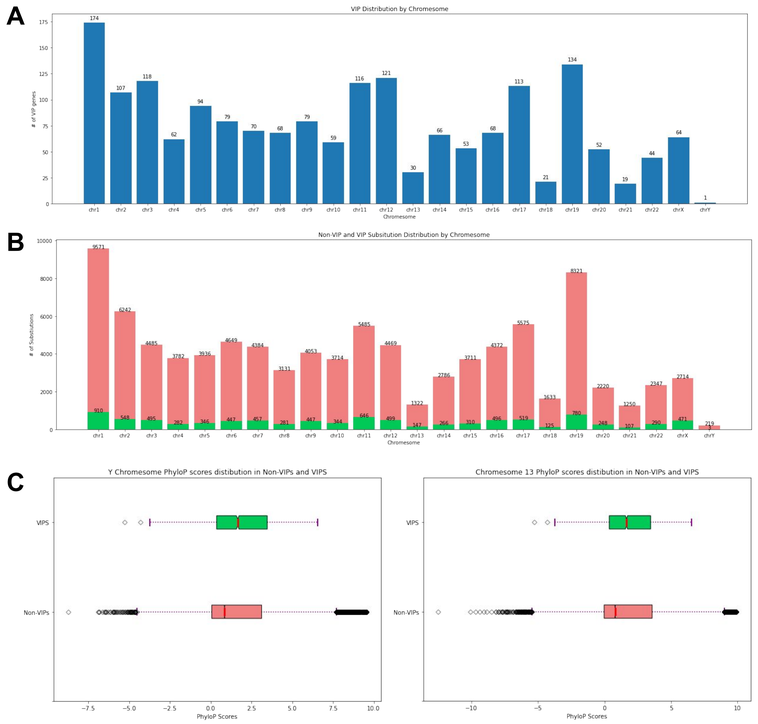

Figure 2: Number of VIPs per chromosome (**A**), number of substitutions in VIPs and non-VIPs per chromosome (**B**), and distribution of phyloP (chr13 and Y as representative chromosomes; **C**).

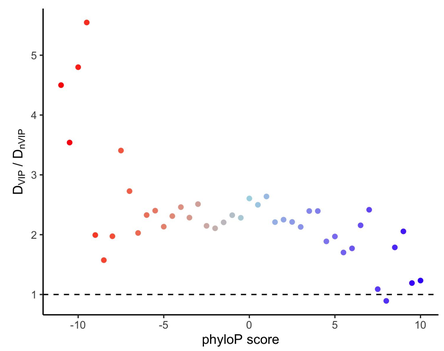

Figure 3: Relative number of substitutions in VIPs (D_{VIP}) as compared to non-VIPs (D_{nVIP}) as a function of phyloP. More conserved sites are not enriched for substitutions in VIPs, while moderately conserved sites and non-conserved sites have an excess of substitutions in VIPs.

Conclusion

Our preliminary analysis aligns with previous work suggesting that VIPs are highly conserved. However, along the human lineage they have an excess of substitutions relative to non-VIPs at all but the most conserved sites. This is somewhat consistent with our model, which suggests that density-dependent pathogens can cause an increase in substitution rate at all but the most strongly conserved sites.

Our analyses are preliminary, and further work must be done to account for potential differences in mutation rates between VIPs and non-VIPs, to account for functional predictions (nonsynonymous vs synonymous sites), and differentiate between different viruses that interact with VIPs. We hope to integrate the substitution patterns predicted by the SIR model to demonstrate how the epidemics may have influenced substitution rates in VIPs, and therefore deleterious allele frequency changes in human populations.

Acknowledgments

I would like to thank my advisor Lawrence Uricchio and everyone in the lab at Tufts university especially Alejandro Calderon and Kasturi Lele. In addition I want to thank everyone in Tufts University Department of Biology.

References

Boots, Michael, and Yoshihiro Haraguchi. “The Evolution of Costly Resistance in Host-Parasite Systems.” *The American Naturalist* 153, no. 4 (1999): 359–70.

Enard, David, Le Cai, Carina Gwennap, and Dmitri A Petrov. “Viruses Are a Dominant Driver of Protein Adaptation in Mammals.” *Elife* 5 (2016): e12469.

Fu, Wenqing, and Joshua M Akey. “Selection and Adaptation in the Human Genome.” *Annual Review of Genomics and Human Genetics* 14 (2013): 467–89.

Hejase, Hussein A, Noah Dukler, and Adam Siepel. “From Summary Statistics to Gene Trees: Methods for Inferring Positive Selection.” *Trends in Genetics* 36, no. 4 (2020): 243–58.

Lachance, Joseph, and Sarah A Tishkoff. “Population Genomics of Human Adaptation.” *Annual Review of Ecology, Evolution, and Systematics* 44 (2013): 123–43.

Nielsen, Rasmus. “Molecular Signatures of Natural Selection.” *Annu. Rev. Genet.* 39 (2005): 197–218.

Pollard K.S, Hubisz M.J., Rosenbloom K.R. , Siepel A. “Detection of nonneutral substitution rates on mammalian phylogenies.” *Genome Res.* 20 (2010): 110-121

Rousselle, Marjolaine, Maeva Mollion, Benoit Nabholz, Thomas Bataillon, and Nicolas Galtier. “Overestimation of the Adaptive Substitution Rate in Fluctuating Populations.” *Biology Letters* 14, no. 5 (2018): 20180055.

Rousselle, Marjolaine, Paul Simion, Marie-Ka Tilak, Emeric Figuet, Benoit Nabholz, and Nicolas Galtier. “Is Adaptation Limited by Mutation? A Timescale-Dependent Effect of Genetic Diversity on the Adaptive Substitution Rate in Animals.” *PLoS Genetics* 16, no. 4 (2020): e1008668.

Visher, Elisa, and Mike Boots. “The Problem of Mediocre Generalists: Population Genetics and Eco-Evolutionary Perspectives on Host Breadth Evolution in Pathogens.” *Proceedings of the Royal Society B* 287, no. 1933 (2020): 20201230.