The strength of natural selection on imprinted genes

M. Wyatt Toure

Abstract

Motivation: Studies have shown that the X chromosome genes exhibit stronger signatures natural selection than autosomal genes. Whether this is due to the hemizygosity of the X in males is unclear. Here I test the hypothesis that hemizygous loci experience stronger selection by searching for reduced nucleotide diversity, a classic signature of selection, at imprinted genes which are functionally hemizygous due to parent-of-origin based epigenetic silencing.

Results: Findings were inconclusive. I find that imprinted genes have a 15% difference in nucleotide diversity compared to autosomal genes that is not statistically significant ($p = 0.058$). Further studies should perform this study with taxa that have been able to identify more imprinted genes such as *Mus* to increase power.

Availability: All code are deposited on GitHub at https://github.com/mw-toure/nat_selection_imprinted_genes

Contact: mt3215@columbia.edu

Introduction

Regions of the genome that are under strong natural selection can provide insights into the genes that may have disproportionate impacts on evolutionary phenomena such as lineage divergence and adaptation. When deleterious mutations that arise are purged by natural selection the haplotype they arose on is also purged a phenomenon known as background selection. This has the effect of reducing diversity around this locus. Hernandez et al. 2011 find that nucleotide diversity decreases around exons, linked non-conserved non-coding sites, and conserved non-coding sites. To explain these results the authors suggested that background selection likely produced these patterns. Background selection can produce reductions in diversity around conserved non-coding regions which may reflect selection on linked genic regions.

Interestingly, in this study the X chromosome required twice the genetic distance from exons to recover nucleotide diversity and had greater declines in nucleotide diversity near exons compared to the autosomes. The increased reductions in diversity on the X is interesting because it suggests that the X experiences stronger background selection in agreement with theory on "the large X effect" (Coyne 1992). In males, the X chromosome is only present in one copy (hemizygous) so deleterious recessive variants are immediately exposed to selection and quickly purged, reducing linked diversity, and reducing the accumulation of polymorphisms. These results raise the possibility that natural selection is stronger for hemizygous loci than other diploid regions of the genome. However, several confounds specific to the X chromosome prevent concluding that hemizygous loci experience stronger selection. At any given time, 2/3s of X chromosomes in a population are in females and thus in a diploid state. Males only ever receive maternal X chromosomes. The X also has different recombination dynamics and mutation rates than the autosomes.

Certain autosomal loci behave as if they were hemizygous due to an epigenetic phenomenon called genomic imprinting (Reik and Walter 2001). For imprinted loci, the transcription of a gene depends on the sex of the parent from which the gene was inherited. If a gene is paternally imprinted, then only the maternal allele is expressed. Parent-of-origin based expression renders imprinted genes functionally hemizygous. Thus, deleterious recessive mutations on the expressed allele of an imprinted gene should be immediately exposed to selection, just like deleterious recessive mutations on male X chromosomes.

Here I test the hypothesis that functionally hemizygous imprinted loci are less likely to accumulate polymorphisms than diploid sites due strong background selection reducing nucleotide diversity. Imprinted autosomal loci having lower nucleotide diversity when compared to non-imprinted autosomal loci would be consistent with the hypothesis that natural selection is stronger on imprinted loci and have implications for the genetic basis of evolution.

Methods

Data set

I used the 1000 genomes project phased SNV/INDEL/SV calls generated by the New York Genome Center (Byrska-Bishop et al. 2021). These data were generated from 3202 individual whole genome sequences that were sequenced to 30x coverage. These sequences were aligned to human genome assembly GRCh38. The variant calls were provided as .vcf files. I only considered single-nucleotide polymorphisms for all my analyses.

Functional elements in the human genome

I obtained the start and end position in base pairs of all genes in the human genome by querying release 106 of the Ensembl database via BioMart for the Ensembl gene ID, gene name, as well as start and end positions of all genes in human genome assembly GRCh38.

The imprinted status of a gene was obtained via the GeneImprint database. Only those genes that were confirmed imprinted (*i.e.*, not of predicted imprinted status) were included in the final analysis as being imprinted.

Nucleotide diversity calculations

I calculated nucleotide diversity (π) per-site as the average pairwise difference from the SNP calls using vcftools (Danecek et al. 2011) for all 22 autosomes across the 3202 individual samples. I then obtained the average nucleotide diversity across an entire gene by calculating the average value of π bounded by the gene start position and the gene end position providing a value of nucleotide diversity for the exons and introns of a gene.

Comparing imprinted and normal gene diversity

To statistically compare the genic diversity of imprinted genes versus normal genes I fit a linear model with a fixed effect of imprinted status.

Results and Conclusions

The mean value of π was 0.042 for normal genes and 0.036 for imprinted genes (Figure 1). While the percentage difference in π between imprinted genes and normal genes is 15%, this difference is not statistically significant ($p = 0.058$). There were no confirmed human imprinted genes on the X chromosome but reassuringly the diversity for X chromosome genes was lower than that of all other autosomal genes ($\pi = 0.033$), consistent with the findings of Hernandez et al. 2011.

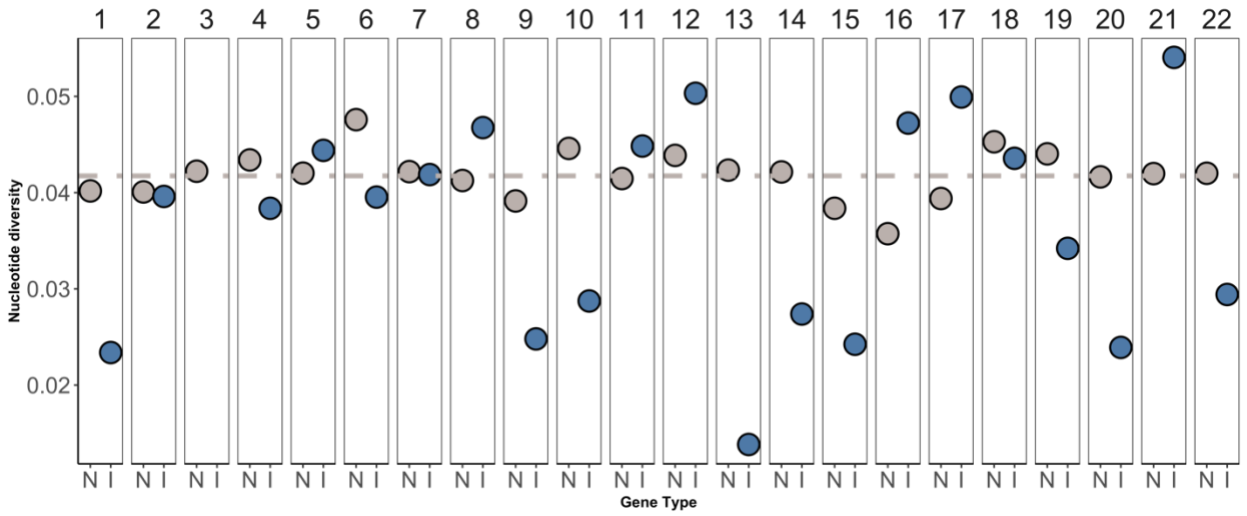

Figure 1 Genome-wide values of nucleotide diversity based on imprinting status and chromosome number. Normal genes are grey and imprinted genes are blue. Chromosome numbers are above panel boxes. Dashed line indicates mean nucleotide diversity for normal genes across all chromosomes. N = normal, I = imprinted.

The reduced diversity of imprinted genes was not consistent across all chromosomes, likely due to variability in the number of imprinted genes on any given chromosome. For example, chromosome 3, one of the largest chromosomes had no confirmed imprinted genes while chromosome 15 which is much smaller has 17 confirmed imprinted genes. Moreover, there are very few imprinted genes (n = 121) likely reducing my power to detect an effect.

If increased power supports the trend observed in this study, then the reduction in diversity of imprinted versus normal genes as well as that of the X might have evolutionary implications. The shared patterns of evolution between the X chromosome and imprinted genes might partially explain their disproportionate role in the genetics of post-zygotic isolation. If purifying selection tends to keep these regions of the genome highly homogenous within a population, then when populations diverge, they may be prime candidates for genetic incompatibilities should any polymorphisms arise when the lineages are on separate evolutionary trajectories and then reunited. If increased power *does not* support the trend observed in this study, then this raises the question of how it can be possible for hemizygous loci to accumulate polymorphisms at the same level as diploid loci.

Future work will attempt to use predicted imprinted genes and/or taxa that have more confirmed imprinted genes, incorporate genetic distances into the calculation of π rather than simply using nucleotide diversity calculated over base pairs, control for mutation rate, and incorporate diversity calculated as a function of distance to see whether sites linked to imprinted loci exhibit similar trends observed at the genic regions of the X chromosome.

Acknowledgements

I would like to thank Professors Itsik Pe'er, Guy Sella, Andres Bendesky, and peers Benjamin Bokor, Linghao Kong, James Howard II, and Alex Guo for valuable feedback.

References

Byrska-Bishop, Marta, Uday S. Evani, Xuefang Zhao, Anna O. Basile, Haley J. Abel, Allison A. Regier, André Corvelo, et al. 2021. "High Coverage Whole Genome Sequencing of the Expanded 1000 Genomes Project Cohort Including 602 Trios." **Preprint**. Genomics. doi:10.1101/2021.02.06.430068.

Coyne, Jerry A. 1992. "Genetics and Speciation." ***Nature*** 355 (6360): 511–15. doi:10.1038/355511a0.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." ***Bioinformatics*** 27 (15): 2156–58. doi:10.1093/bioinformatics/btr330.

Hernandez, Ryan D., Joanna L. Kelley, Eyal Elyashiv, S. Cord Melton, Adam Auton, Gilean McVean, 1000 Genomes Project, Guy Sella, and Molly Przeworski. 2011. "Classic Selective Sweeps Were Rare in Recent Human Evolution." ***Science*** 331 (6019): 920–24. doi:10.1126/science.1198878.

Reik, Wolf, and Jörn Walter. 2001. "Genomic Imprinting: Parental Influence on the Genome." ***Nature Reviews Genetics*** 2 (1): 21–32. doi:10.1038/35047554.