

Coalescence of Avian Lysozyme C

Introduction

There have been observations of complex traits whose distribution among species appears to conflict with known species trees. Differences between gene trees and species trees may arise from horizontal transfer (figure 5) of genes through vectors such as viruses or through hybridization across phylogeny (Maddison et al 1997). This also can be a result of a process called lineage sorting (figure 6), where “ancestral polymorphisms persist through several speciation events” (Maddison et al. 1997). Gene trees can show alternative topologies to the species tree, where genes coalesce further back than the species.

Odonophoridae and Numididae are unique from other Galliform birds in three amino acids within the lysozyme C gene, not found in the LYZ of phasianids. slight alterations to any of these three amino acids will result in creation of a highly unstable enzyme, resulting in an inviable egg. The accepted phylogeny of Galliform birds places Numididae and Odonophoridae as a paraphyletic group, with Odonophoridae being sister to the Phasianids, posing a question on the evolutionary history of Galliform birds. The literature phylogeny implies that LYZ either evolved convergently in Numididae and Odonophoridae or evolved as a synapomorphy of the three, being lost in the phasianids. The nature of the three amino acids makes both options highly unlikely. Because of this, the avian lysozyme c gene may be an example of horizontal transfer or deep coalescence.

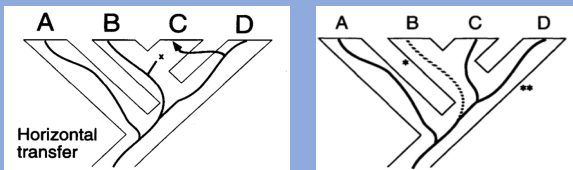


Figure 5 and 6: from Maddison et al.

Methods

We collected the sequences of the LYZ gene and its homologs in related species from NCBI by using BLAST (Camacho C et al 2009) software to extract the related gene from each of the selected genomes. This was completed by using BLAST against the known *Gallus gallus* LYZ coding region and the 2000 base pairs preceding. The gene from each of the species went through MUSCLE alignment (Edgar et al) to produce an interleaved phylip file to input into mesquite. We also removed a section at the end of LYZ for the species *Penelope pileata* because of high divergence at the 3' end of the LYZ gene. We submitted separate Mesquite files for the whole gene, ~2000 base pairs before the gene, exons, and non-coding regions through iqtree server. Site likelihood data was obtained using IA-TREE with iqtree (L. T. Nguyen et al.), which output two files, one with each base's support of the classical phylogeny and one with each bases support value of our alternative phylogeny. These values are depicted in figure 4.

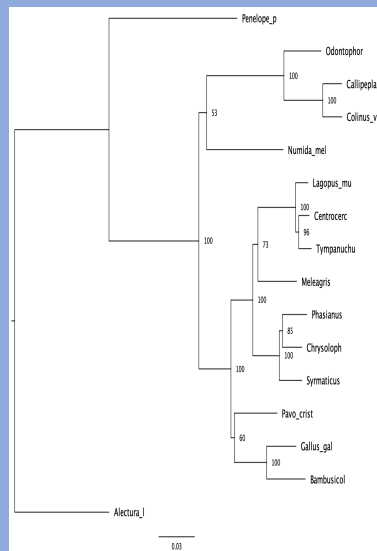


Figure 1: gene tree of the total LYZ gene and ~2000 preceding base pairs

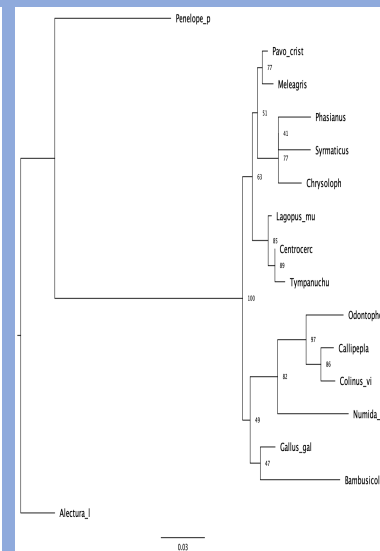


Figure 2: gene tree for exons of LYZ with bootstrap values

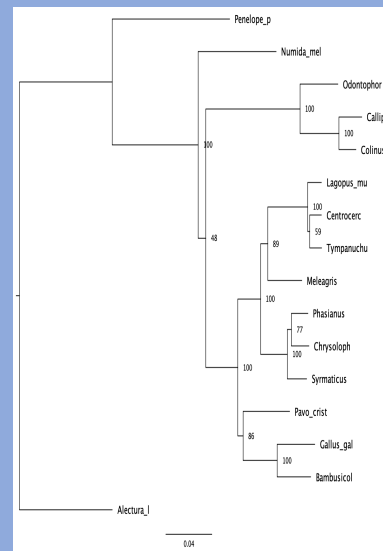


Figure 3: gene tree for introns 1 and 2 of LYZ with bootstrap values

Results

The gene trees representing the entire gene and exons of the gene (Figures 1 and 2) showed support of our alternative phylogeny, and non-coding regions showed weak support for the alternative or support for the classical phylogeny. This can be attributed to the lack of selective effects on non-coding regions. Bootstrap support of the guineafowl/new world quail node in the exon tree is. 82%, indicating strong support of the monophyletic gene hypothesis. Trees for the non-coding regions of the gene were much more likely to follow the classical phylogeny of new world quail and guineafowl. The findings of the base likelihood analysis support that the coalescence of the coding regions of LYZ follows our alternative phylogeny. Figure 4 represents the difference between likelihood values for each tree at each base, negative values indicating support for our alternative phylogeny. Negative spikes in exon 1 and 2 represent the target amino acids found in guineafowl and new world quail.

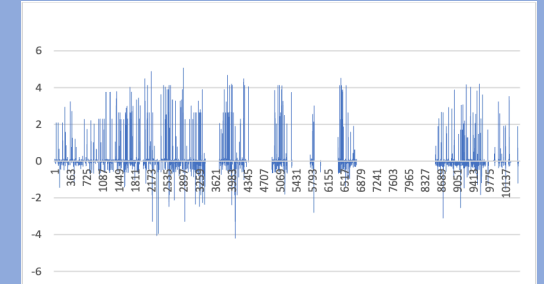


Figure 4: Differences in likelihood of each base to represent either tree.

Discussion

The results of phylogenetic and base likelihood analysis provide evidence of deep coalescence in the coding regions of LYZ in Numididae and Odonophoridae. The branches of the gene coalesce more recently than those of the species, indicating possible horizontal transfer of the specific amino acids within the two species. This would have occurred through some vector, likely a virus . Deep coalescence occurs when retained traits fail to coalesce at the same rate as the rest of the genome. This most often results in paraphyly on a gene tree where species may be monophyletic, the opposite of this case. It is possible that the rest of the genome is an example of deep coalescence, with the three target amino acids coalescing more recently. This is very unlikely, so the topology of the exons of LYZ is a result of horizontal transfer more recently in time.

References
 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421. doi: 10.1186/1471-2105-10-421. PMID: 20003500; PMCID: PMC2803557.
 Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with high accuracy and high throughput. Nucleic Acids Research 32(5):1792-1797. PMID: 15032147. DOI: 10.1093/nar/gkh340
 Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. PMID: 15318951. DOI: 10.1186/1471-2105-5-113
 Maddison, Wayne. (1997). Gene Trees in Species Trees. Systematic Biology. 46. 10.1093/sysbio/46.3.523.
 Madrin F, Park YM, Lee J, Biso N, Gur T, Mathiasodaman N, Bantkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019 Jul;47(1) W308-W311. doi:10.1093/nar/gkz208. PMID: 3097679; PMCID: PMC662279.