

ADVANTAGES OF NETWORK-BASED APPROACHES FOR THE PHYLOGENETIC ANALYSIS OF INTRAGENOMIC REPEAT REGIONS

Surya Saha^{1,2}, Susan Bridges¹ and Daniel Peterson²

¹ Department of Computer Science and Engineering, Mississippi State University

² Mississippi Genome Exploration Laboratory, Department of Plant and Soil Sciences, Mississippi State University



Abstract

Contemporary genomics research tends to focus on genes. However, the genomes of eukaryotes are dominated by non-coding, repetitive elements. While the function of most of these elements is unknown, their prevalence suggests that they confer adaptive advantages. In general, attempts at phylogenetic analysis of repeat sequence evolution have relied upon "tree-based" algorithms developed for studying genes. Such algorithms may not accurately reflect repeat sequence evolution as repetitive elements and genes are subject to different selective pressures. In this regard, we have initiated comparative analysis of the sequences from the Galluhop repeat region of chicken (*Gallus gallus*) using "tree-based" phylogenetic algorithms and newer "network-based" approaches. Our goal is to evaluate which approaches best reflect biological reality. Phylogenetic study of the evolution of repeat regions coupled with the knowledge of evolution of the gene regions may help shed light on evolution of the organism as a whole.

Objectives

- To study the evolution of repeat regions of the genome including consideration of varying copy number and parallel evolution.
- To compare the models of evolution generated by classical tree-based phylogenetic methods with those generated by network methods.

Median Joining Networks

- Proposed by Bandelt et al (1999) for reconstructing the phylogeny of regions with reticulate evolution.
- Can be used to efficiently analyze large datasets and multistate characters.
- Begins with a network of minimum spanning trees constructed using Kruskal's algorithm and adds the missing intermediates (median vectors) using the Farris maximum parsimony algorithm

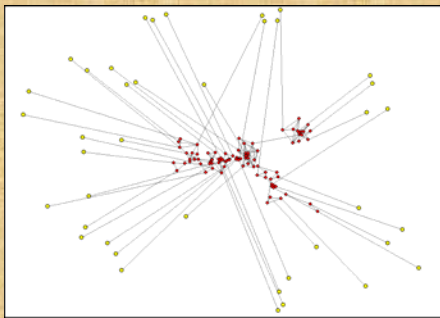


Figure 3a: Median Joining analysis of 40 sequences from the Galluhop region

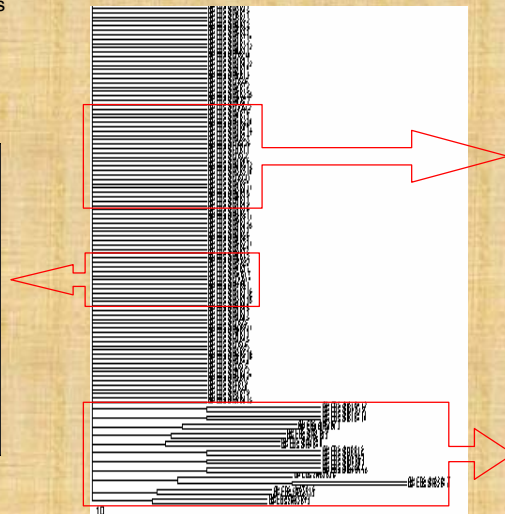


Figure 2: Maximum parsimony analysis of Galluhop region

Network Approach

Phylogenetic networks

- Represent the evolution of sequences as a directed acyclic graph (DAG).
- Construction methods usually start with computation of a minimum spanning tree.
- Utilize techniques such as median networks, median-joining networks, molecular variance parsimony, neighbornets, pyramids, weak hierarchies and split decomposition (see the review by Makarenkov et al. 2005)

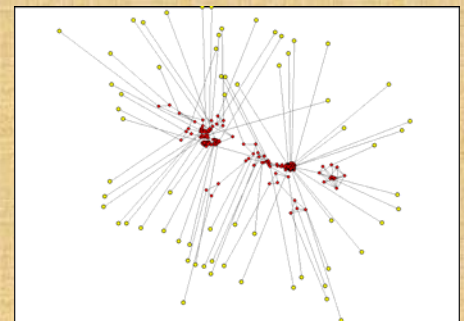


Figure 3b: Median Joining analysis of 64 sequences from the Galluhop region

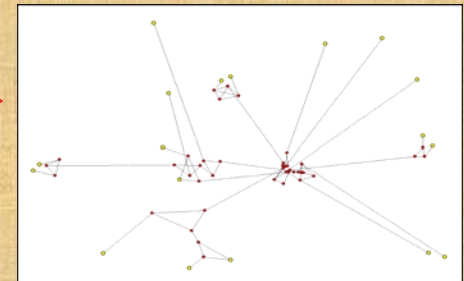


Figure 3c: Median Joining analysis of an informative sequence subset from the Galluhop region

Network Versus Tree-Based Phylogenetic Methods

Homoplasy

- Is the presence of similar features in a species that did not arise from a common ancestor.
- Is thought to be common in the evolution of repeat regions as a result of convergent evolution.
- Is problematic for classical phylogenetic algorithms.

Repeat sequences and homoplasy

- Repeat sequences may undergo parallel evolution along alternate paths in the genome
- Analysis requires consideration of a large number of sequences and complex evolutionary pathways.
- Network models can represent alternative evolutionary pathways that have the same cost for a given optimality criterion.
- The path may involve reticulations as a result of homoplasy.

Interpreting intraspecific and intragenomic relationships

- Traditional phylogenetic algorithms are designed for analysis of interspecific relationships that are best represented by bifurcating trees.
- Intraspecific and intragenomic relationships may have multifurcations along alternative pathways.
- Network methods are better suited for analysis of evolution of intraspecific and intragenomic phylogenies because the process may involve

- Low divergence among samples
- Persistence of ancestral nodes along with descendant nodes
- Multifurcating relationships

Materials and Methods

The dataset is one of the newly identified repeat regions in the chicken genome called *Galluhop* and consists of 184 sequences from the coding sequence region that is exclusive for *Galluhop* mother elements

- PAUP was used to generate trees for the *Galluhop* dataset using HKY85 distance measure matrix and also maximum parsimony.
- The NETWORK software is used for network based phylogenetic analysis using median joining of the *Galluhop* dataset. The NETWORK software is based on the implementation of the median joining algorithm by Arne Röhl.

Conclusions

- The maximum parsimony analysis of the *Galluhop* regions performed using PAUP shown in Figure 2 does not reveal any significant phylogenetic relationships. This is due to the homology among the sequences in the dataset as well as lack of sensitivity of the maximum parsimony algorithm.
- However, the network depicted in Figure 3 derived from the median joining algorithm and a subset of the *Galluhop* sequences displays significant knowledge in terms the origin and evolution of the sequences.

Next Step

Sequencing of a number of repeat regions of various organisms is underway. The evolution of these repeat regions needs to be studied with respect to the evolution of the genic regions of the organism. This can help us understand how both the genic regions and non-genic regions of the genome effect the evolution of the organism.

References and Acknowledgements

- H.J. Bandelt, P. Forster, A. Rohl, "Median-joining networks for inferring intraspecific phylogenies," *Mol Biol Evol.*, vol. 16, issue 1, Jan 1999, pp. 37-48.
- R Cordaux, D.J. Hedges, M.A. Batzer, "Retrotransposition of *Alu* elements: how many sources?," *Trends Genet.*, vol. 20, issue 10, Oct 2004, pp. 464-467.
- T Wicker, J.S. Robertson, S.R. Schulze, F.A. Feltus, V. Magrini, J.A. Morrison, E.R. Mardis, R.K. Wilson, D.G. Peterson, A.H. Paterson and R. Ivarie, "The repetitive landscape of the chicken genome," *Genome Research*, vol. 15, 2005, pp. 126-136.
- V. Makarenkov, D. Kevorkov and P. Legendre, "Phylogenetic Network Reconstruction Approaches," to appear in *Applied Mycology and Biotechnology*, v. 6, Genes, Genomics and Bioinformatics, Elsevier Science.

We would like to thank Dr Mark Fishbein for advice.