

# Finding a Final Common Path for the Gene Trees and Species Trees Congruence

Tomilowo Abijo<sup>1,2\*</sup>, Chioma Ossai<sup>2</sup>, Oluwaferanmi Bello<sup>2</sup> and Olusayo Louise-Oluwasanmi<sup>3</sup>

<sup>1</sup>Department of Anatomy, Howard University College of Medicine, Washington DC, USA

<sup>2</sup>Department of Biology, Howard University, Washington DC, USA

<sup>3</sup>Department of Genetics and Human Genetics, Howard University College of Medicine, Washington DC, USA

## Abstract

It is generally believed that since the traits and characteristics exhibited by most organisms are a result of their genetic composition, then all gene trees must be the same as the containing species tree. While both trees appear to present dissimilar results or topologies, even between closely related species, the question most phylogeneticists have been asking is what could be responsible for the lack of congruence between them. Past and recent studies have shown that gene duplication, horizontal transfer and incomplete lineage and/or deep coalescence are factors that play a major role in topological disparity. All these factors can be linked to the fluid nature of genes. To address this issue, phylogeneticists have resorted to building models that can account for gene duplication, horizontal transfer, and incomplete lineage sorting. However, as good as some of these models are, results have shown that the discord between gene and species trees cannot be fully accounted for. In this paper, we describe the factors responsible for species/gene trees disparity and some of the different model approaches that have been proposed to solve them.

**Keywords:** Gene tree; Species tree; Horizontal transfer

## Introduction

### Relationship between a species tree and a gene tree

To highlight or provide a good description of the problems associated with gene trees and species trees, it is necessary to define the phenomenon of these two concepts. It is also essential to explain the relationship that exists between them. A species tree (also known as the phylogenetic trees of species) is a diagram that depicts the evolutionary descent or relationships of a set of species. Similarly, a gene tree is a diagram that depicts the relationship between homologous genes of the same or different species. Species trees are inferred from gene trees, although it is also possible to infer them from other trees like host-parasite trees. Because morphological and behavioural characteristics of species are induced by their intrinsic genes, species trees and gene trees are always expected to be congruent. This simply explains the underlying relationship between gene trees and species trees concepts. To further explain this relationship based on congruence, a phylogenetic tree of haemoglobin gene of invertebrates is expected to present the same results as the species tree of the same group animals, presuming that the haemoglobin gene descent with regard to function and structure are evolutionarily maintained. Thus, when lineages undergo speciation, it is expected that the corresponding gene copies in these organisms are distributed in a pattern reflecting the parent species trees. Although congruence explains to a large extent the relationship between a gene and the parent species tree, the phenomenon is not universal. In fact, many occasions have arisen where two or more sister species do not have sister copies in the gene tree [1]. This lack of congruence is the basic problem of the gene trees and species tree. So far, three factors have been shown to be responsible for the lack of congruence seen in some gene trees and species trees. These problems are Horizontal transfer [2], Lineage sorting or Deep Coalescence [3] and Gene Duplication or Extinction [4]. To better appreciate the contribution of each of these factors on the problems of gene trees and species trees problems, we review the concepts separately.

### Gene tree

As stated earlier, a gene tree depicts the evolutionary descent or changes in gene copies within or among a group of species. To maintain life continuity, genes have to replicate. In addition to perfecting the

process of life continuity, genes are transcribed into messenger RNA and translated to proteins. While gene transcription and translation are vital to maintaining life processes, only gene replication will be discussed to explain the underlying process of gene tree development. Gene mutations, which occur mostly during gene replication, result from the imperfection of the replication process, as genes are constantly mutated. Although there are intrinsic systems in every organism to correct this error, it is highly unfortunate that they are not 100% efficient. The transfer of these mutations (also called polymorphisms) from parents to their progenies along evolutionary lines results in a gene tree [5]. This is more pronounced within the same species. Another factor that contributes to the development of gene tree is recombination during the process of sexual reproduction [6]. Homologous genes are constantly recombined to generate genes in offspring that are markedly different from the parents (Marcon and Moens, 2005). The accumulation of several recombined genes and their transfer could result in a gene tree. Like mutations, recombination effects on gene tree generation are more pronounced within the same species. As shown in Figure 1, species W, X, Y and Z are sister species with sister gene copies w, x, y and z. Although X and Y are sister taxa on the specie tree, they are not on the gene tree.

### Species tree

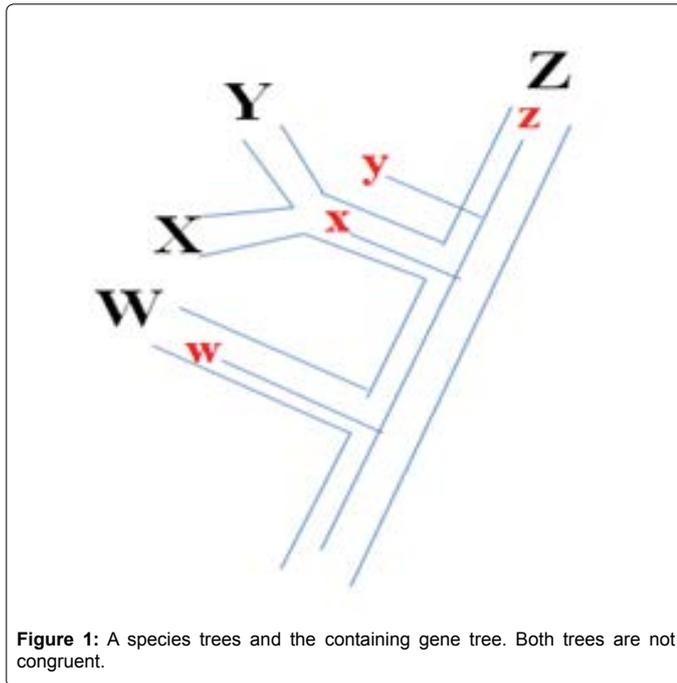
Speciation of lineages is induced by the ability of populations to adapt adequately to their natural habitat. As described by Charles Darwin in his theory of natural selection, nature selects organisms that are well adapted to survive and those that are not are subsequently eliminated [7]. This theory puts a lot of pressure on populations to continually evolve based on the prevailing environmental conditions.

**\*Corresponding author:** Tomilowo Abijo, Department of Anatomy, College of Medicine, Howard University, 520 W Street NW Washington DC 20059, Tel: 2407149226; E-mail: [tomilowo.abijo@Howard.edu](mailto:tomilowo.abijo@Howard.edu)

**Received** March 12, 2019; **Accepted** March 30, 2019; **Published** April 05, 2019

**Citation:** Abijo T, Ossai C, Bello O, Oluwasanmi OL (2019) Finding a Final Common Path for the Gene Trees and Species Trees Congruence. J Phylogenetics Evol Biol 7: 210.

**Copyright:** © 2019 Abijo T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



Thus, new species with well-suited traits continually evolve to ensure continuity. The pattern of evolutionary lineages or descent of these species is known as a species tree. As species change, it is wise to say that intrinsic genes are also likely to change. In fact, some speciation events may be triggered by gene mutations or duplications that confer new characteristics on the species and which help the species to adapt to the new environment or conditions. It is therefore reasonable to say that most species trees are likely to contain sister gene copies with similar patterns of descent along evolutionary lines.

### Gene trees and species trees problems

As stated earlier in the introduction, lack of congruence is the major problem of gene and species trees (Figure 1) and this problem is due to Horizontal Transfer, Lineage Sorting or Deep Coalescence, Gene Duplication and Extinction. Each of these factors is discussed below.

### Horizontal transfer

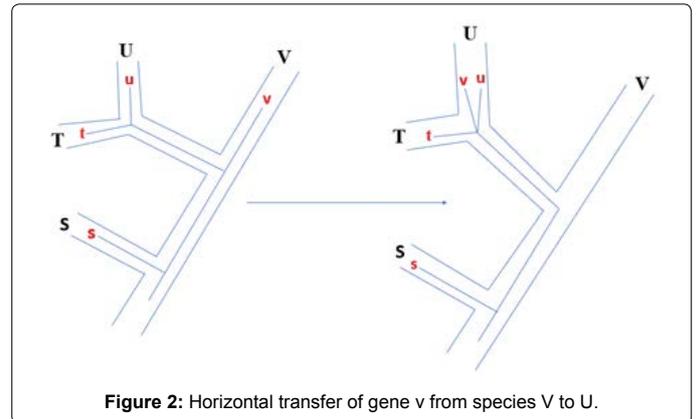
In Figure 2, the phylogenetic tree of species shows that T and U are sister taxa with a common ancestor; however, the containing gene tree does not present sister gene copies t and u as such. A possible explanation for this is that gene v might have been transferred from V to U by a mechanism other than normal sexual reproduction. For this process to be successful, it is required that the gene v is appropriately incorporated into the host genome i.e. species U in this case, and that gene u is capable of replication using the host replication machineries. Vectors, such as viruses, may be required in this case for the transfer of this gene [2]. Thus, according to the gene tree, v and u are sister taxa after the gene transfer.

### Extinction

As species are lost to extinction, a whole number of genes become unaccounted for in the species tree which may appear in the gene tree.

### Gene duplication

Duplication of genes could also cause a lack of congruence between gene trees and the species tree. Although this factor is less predominant in causing lack of congruence, its effect still cannot be ignored. When a



gene is duplicated at a different locus, it evolves independently, thereby producing a duplicate gene tree in the containing species. While this may not seem to be a big problem at the beginning, it becomes a great problem when one or both genes are lost in one of the containing species during speciation and therefore cannot be accounted for.

### Deep coalescence or independent lineage sorting

A stable mutation of a gene across the population gives rise to polymorphism. Depending on the magnitude of the polymorphism, a related copy or an entirely new copy of the gene (as in duplication) could be introduced into the species. During a speciation event, polymorphic genes may be independently sampled as two different genes, thereby appearing in two different species, a phenomenon called complete lineage sorting. However, genetic drift may cause one or more of the polymorphic genes to be incompletely sorted into a particular species during speciation. While this phenomenon is called incomplete lineage sorting, its implication is a lack of congruence between the gene tree and containing species tree.

### Model approach to species tree and gene tree problems

To solve the species tree and gene trees problems, studies are now focused on how to generate models that will account for all the factors listed above; however, the reality is that it will be difficult to come up with a single model that will account for all the factors. The parameters used in generating these models are also caveats. Models for each factor are discussed below.

### Coalescent models

Most coalescent models emphasize the effects of the population on the evolution process and its contribution to the disparity between species tree and gene trees. To better understand most coalescent models, it is very important to understand that mutations, which become stable over time, give rise to new alleles (or polymorphism), and can follow a distinct pathway in the evolutionary tree from the parent tree, thereby generating a new gene tree (or allele tree). It should also be noted that the frequency of changes of this allele is not the same as speciation time of the host species. Therefore, if the speciation time is very long compared to allele changes, then the gene tree would be markedly different from the species tree, implying that the gene trees cannot be used to infer a species tree. In recent time, models such as maximum likelihood [8] and Bayesian probability [9] have been used to infer species trees from gene trees, taking into account branch lengths, topology, divergence time, and other parameters.

### Gene duplication and loss models

Unlike the coalescent models, most gene duplication and loss

models are used for inferring gene trees from species tree. A prominent model accounting for gene duplication assumed that parent species tree contains branches with their corresponding time of divergence from their ancestor (node) and the rates of gene duplication and loss over the entire tree are assumed to be fixed. Assuming the rates of speciation and extinction are constant (i.e., Birth-Death model of diversification), the gene tree can be inferred from the species tree as the corresponding evolution rates on the species trees [10,11].

### Horizontal gene transfer model

Hordijk and Gascuel [11] proposed a model that account for gene duplication and loss. In this model, the branches are moved using Subtree Pruning and Regrafting (SPR) to generate the optimal tree. Using the Poisson distribution, the probability of a gene tree was derived from the number of the SPR branch moves. Other methods that can be employed are the Bayesian probability and MCMC.

### Conclusion

There are no models currently in place that can account for all the tree factors mentioned above to correctly reconcile a species tree and its containing gene trees. Although there are models that account for at most two of these models simultaneously, there are still questions regarding their computational and statistical powers. It is also believed that no matter how powerful a model is, it can never account for some processes that play critical role in gene tree and species tree disparity. One of these is Hybridization. Hybridization of genes increases gene diversity within and across species population. Its effect is so subtle, yet pronounced. In most models, it is always assumed that there is no hybridization, which in fact is impossible. Because of these problems and others, many phylogeneticists have proposed that gene trees and species are totally different and the belief that one can be used to infer the other should be discarded.

### References

1. Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics* 105: 437-460.
2. Cummings MP (1994) Transmission patterns of eukaryotic transposable elements: Arguments for and against horizontal transfer. *Trends Ecol Evol* 9: 141-145.
3. Avise JC, Shapiro JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Mol Biol Evolution* 1: 38-56.
4. Zhang L (2011) From gene trees to species trees II: species tree inference by minimizing deep coalescence events. *IEEE/ACM Trans Comput Biol Bioinform*. 8: 1685-1691.
5. Griffiths RC, Tavaré S (1999) The ages of mutations in gene trees. *The Annals of Applied Probability* 9: 567-590.
6. Hein J (1990) Reconstructing evolution of sequences subject to recombination using parsimony. *Math Biosci* 98: 185-200.
7. Fisher RA (1930) *The genetical theory of natural selection*. Clarendon Press, Oxford, England.
8. Kubatko LS, Carstens BC, Knowles LL (2009) STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25: 971-973.
9. Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009) Coalescent methods for estimating phylogenetic trees. *Mol Phylogenet Evol* 53: 320-328.
10. Arvestad L, Berglund ACC, Lagergren J, Sennblad B (2004) Gene tree reconstruction and orthology analysis based on an integrated model for duplications and sequence evolution. *Proceedings of the Eighth Annual International Conference on Research in Computational Molecular biology* 326-335.
11. Hordijk W, Gascuel O (2005) Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics* 21: 4338-4347.