Taken from:

https://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-6-

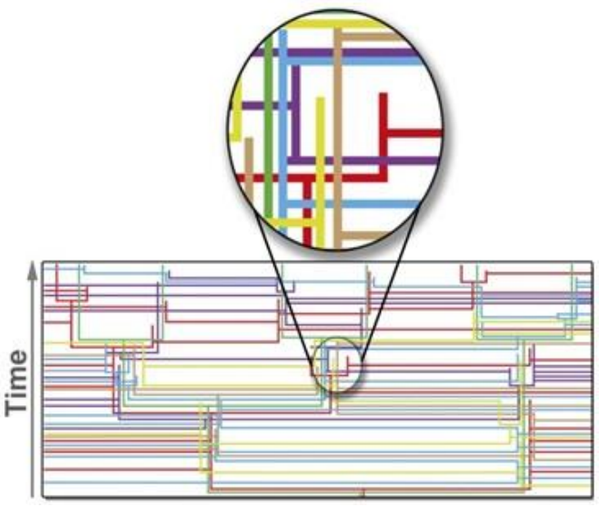
Length

LONG

ACADEMIC DIVISION:

CBS

The Public Goods Hypothesis for the Evolution of Life on Earth



The ‘Tree of Life’ hypothesis has been in existence for most of the last two centuries and is one of several hypotheses that have been put forward to explain the diversity of life on the planet. In its recent practice, this theory focuses on the vertical inheritance of genes from parent to offspring and the continuous division of lineages in the process of speciation. The theory was largely formulated in the days when the science of systematics was mostly concerned with the analysis of plant, fungi and animals and the study of evolution was essentially focused on the study of these three eukaryotic lineages. Indeed, arguably the greatest contributor to the definition of the species concept, Ernst Mayr, was so adamant in pinpointing that his views specifically applied to sexual organisms that he clearly titled his major work on speciation mechanisms "Systematics and the origin of species from the viewpoint of a Zoologist".

Microbiologists engaged in a long and largely unsatisfying search for the tree of prokaryotic life for most of the 20th Century. Woese detailed this search in his treatise on bacterial evolution, where he also wrote about the ideas and false-starts that arose from time to time in the earlier part of the century. The 1970s had resulted in significant developments in the sequencing of genes and this led Woese and others to the belief that there was a new tool available to systematists that would ultimately lead to a satisfactory and detailed resolution of the entire Tree of Life. The fine-grained picture of the tree of life was rapidly being brought into focus by the use of ribosomal RNA sequencing and analysis. Indeed, so powerful was this line of argument that today, the world's most targeted gene for sequencing is the small subunit ribosomal RNA gene. However, and remarkably so, the most sequenced and most ubiquitous kind of gene in the world is not the small subunit ribosomal RNA gene. Transposases, typically known to function by frequent movement

from one genetic element to another are the most abundantly discovered genes in metagenomic studies where sampling of genes is undirected.

This discrepancy between the most abundant marker sequenced in the traditional phylogenetic framework and the most abundant genes actually obtained in nature by chance, suggests that the scope of phylogenetic analysis is focused on the analysis of certain genes, while if a different perspective was taken - to focus on the most abundant genes - then a different interpretation of evolutionary history might be more readily obtained.

Commonly, by the 1990s, the optimism surrounding the reconstruction of a tree of life was giving way to a more realistic picture of life on the planet. Hilario and Gogarten and Martin et al. pointed out remarkable inconsistencies in molecular phylogenies derived from ATPase and Glyceraldehyde-3-phosphate dehydrogenase genes when these phylogenies were compared with ribosomal RNA phylogenies. A genome analysis showed that a substantial portion of the known E. coli genome was acquired by horizontal gene transfer since its separation from Salmonella. At the end of the century, Doolittle and Martin summarized these growing problems with the tree of life model. Soon afterwards, serious efforts were being made to identify interspecies gene transfer events and to quantify the extent of HGT in prokaryotic genomes and phylogenetic tree diagrams have been increasingly giving way to network models of genome evolution in prokaryotes. It must be pointed out, however, that the focus on HGT events has been strongly criticized and some congruence between gene trees is easily found.

It is not the case, however, that network diagrams are being universally employed and indeed they still appear in only a minority of studies that deal with the molecular systematics of prokaryotes. Tree diagrams, inferred from a subset of genes with a wide or universal distribution are still the most commonly used models for prokaryotic evolution, though it has been pointed out that usually these diagrams are only constructed from less than one percent of the genome of these organisms. When a larger portion of the genome is used, then the resulting tree diagram is highly dependent on the method used in its construction and in any case, no strongly-supported nodes are found when moving towards the base of this tree. In the middle ground are studies that try to reconstruct a tree or forest of trees in the presence of HGT events.

What is becoming increasingly obvious is the need to either improve the Tree of Life model, if that is indeed possible, or replace it with one (or several) hypotheses that better fit the data. At the moment the interpretations of this model seem to be straddling the middle ground - there is a great Tree of Life (sensu Darwin, Lamarck, etc.) but it has annotations and complications superimposed on its great frame, caused by interspecies gene transfer. The problem with this model is that it is becoming increasingly implausible. Recent estimates show, for instance, that Escherichia coli as a species uses approximately 18,000 genes, while the percent of gene families that are now known to be found in every E. coli is just 6% of the total, though a typical E. coli strain only possesses 4,000-5,500 genes. This kind of scenario is seen again and again during genome resequencing projects where multiple strains of the same species are sequenced. A minority of the genes that are found in a prokaryotic species are found in just one genome of that species. As a consequence, the Tree of Life hypothesis has been modified extensively from its original description, in order to avoid its rejection. We now know - unlike the originators of the hypothesis almost two centuries ago - that the main process of genome innovation for many of the evolving entities on this planet is not vertical descent, rather it is

recombination and gene acquisition. Genes and genomes did not form part of the original formulation of the Tree of Life hypothesis and processes such as horizontal gene transfer and mobile genetic elements such as viruses, plasmids and transposons obviously did not feature.
In order to accommodate these newly discovered, important features, the hypothesis has been stretched to fit the data, however, given our knowledge of the data, it seems that the elastic limit of the original hypothesis has been passed.

In Darwin's formulation of the Tree of Life hypothesis, he said that he attempted "[...] to show that there is a constant tendency in the forms that are increasing in number and diverging in character, to supplant and exterminate the less divergent, the less improved, and preceding forms." This particular quotation gets to the heart of tree-thinking: that evolving entities would always diverge away from one another and that evolution is a process of divergence. To put it another way, all formulations of the Tree of Life hypothesis have at their core the basic tenet that the pattern of diversity that we see on the planet is caused by a tree-like evolutionary process and the differences in these formulations is to be found in how much they allow deviation from this central idea. Almost without exception, efforts to describe the diversity of life on the planet have focused on the construction of this tree. The tree has been portrayed as a strictly bifurcating tree, as a fuzzy tree, as a tree with cobwebs hanging from it and so forth. Likewise, the number and kinds of evolving entities has changed over time, prokaryotes being largely ignored initially, mainly due to the difficulties of generating interpretable trees from the available data.

The current attempts at constructing the tree of life can be roughly divided into four approaches (though, other classifications of the approaches are easily constructed). Firstly, there is the tree as exemplified by the small subunit ribosomal RNA gene. Next, there is the multi-gene approach using widely distributed genes, usually of informational function; the third approach is to search for the biggest observable trend embedded in the data, and the fourth is to construct phylogenetic supertrees, the so-called tree-from-trees method. These approaches have different meanings and care must be taken to interpret what they say. When the ribosomal RNA approach was first advocated by Woese and co-workers, the ubiquity of the gene and its attractive properties in terms of rapidly and slowly-evolving sites, its conserved structure and its supposed recalcitrance to horizontal gene transfer meant that it was simply being used as a surrogate for the evolution of the entire organism, a position that is no longer tenable. Using multiple genes in a concatenated superalignment is designed to overcome the limitations of using a single gene, however the interpretation of this tree is somewhat similar to the interpretation of the rRNA tree and this approach has been criticised as a "Tree of 1%", not a tree of life. The Statistical Tree of Life (STOL) and phylogenetic supertrees are constructed from a much larger sample of genomes and attempt to either construct a single tree (supertree approaches) or a statistical trend that is tree-like from a large sample of a genome. The interpretation of these structures is somewhat difficult, though they may approximate a "Tree of Cells". Unfortunately, it is now necessary to carefully read each manuscript to find out what the authors are calling the "Tree of Life".

Vocabulary:

The Public Goods Hypothesis for the Evolution of Life on Earth

Engaged: Dedicarse a

Network: Interconexión, red

Point out: Puntualizar

Strain: Cepa

Rejection: Rechazo

Superalignment: Supra alineación

Offspring: Crías, descendencia

Concerned: Relacionado(a)

Fungi: Hongos

Lineages: Linajes

Adamant: Firme

Pinpointing: Subrayar