

The Tree of Life: From the Roots to the Leaves

Zane Thornburg

Literature Seminar

November 13, 2019

Since Charles Darwin published *The Origin of Species* in 1859, there has been an effort to describe evolution and make evolutionary predictions.^{1,2} Later in 1977 Carl Woese would discover the third domain of life and draw the evolutionary tree of life as we think of it today.³ In this brief review I will discuss two models from the group of Nigel Goldenfeld at the University of Illinois at Urbana-Champaign, one describing the structure of early life up to the Last Universal Common Ancestor (LUCA)⁴ and the other describing the structure of the evolutionary tree of life, or the evolution life following the LUCA.⁵ I begin with the description of early life and then will move forward in the evolutionary timeline.

It was proposed by Carl Woese that early life did not evolve or exist in the same way life does today.^{6,7} Rather than there being stable species that evolve and branch into other species evolving through inheritance and mutations during replication, evolution occurred as a communal phenomenon via alteration or displacement through injection of external genetic material from other organisms, or horizontal gene transfer. This phase of life is referred to as the progenote state as labelled in Figure 1(a).⁴ Evolution in the progenote state is horizontally, but eventually there is a point in time when life organizes itself into species in which evolution is vertically dominated known as the Darwinian threshold. Goldenfeld's model of early life begins with a population of N organisms all with genome length L that are well mixed in the genome space as in Figure 2(e). In this model the population evolves stochastically in time through replication with a probability of mutation, selection based on an assigned fitness landscape, competition and death, and horizontal gene transfer.⁴ Each of the effects included here occur with a rate similar to a chemical kinetic rate law where a rate constant is multiplied by the number of combinations that lead to the given effect, such as the number of organism pairs in the population for horizontal gene transfer. Numeric simulations of these stochastic rates were simulated stochastically using a Gillespie algorithm.

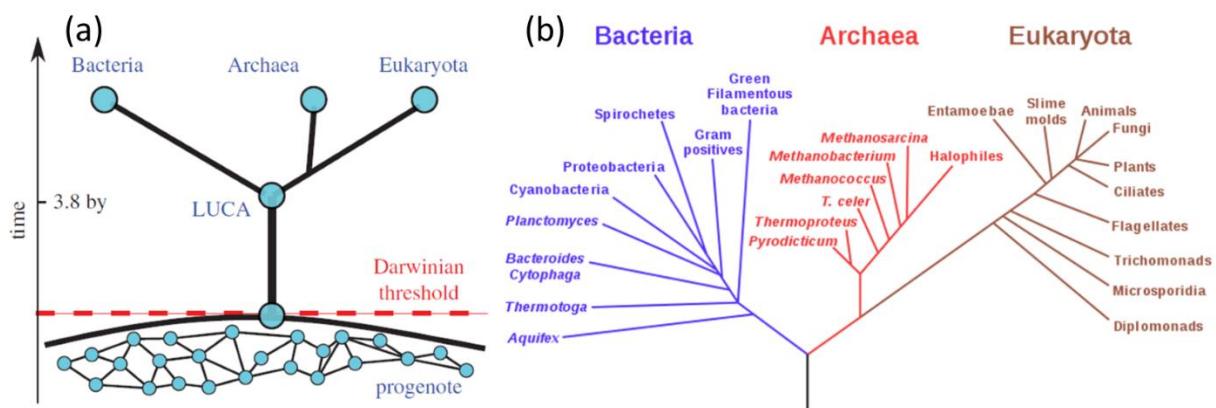


Figure 1: (a): Early life known as the progenote phase over time at which point evolution goes from being dominated by horizontal effects to dominated by vertical effects (From Goldenfeld et al 2017).⁴ (b): Once speciation and the evolutionary tree arises from the progenote, life divides into three domains identified by differences in 16S rRNA.³ (Open source image from Wikipedia)

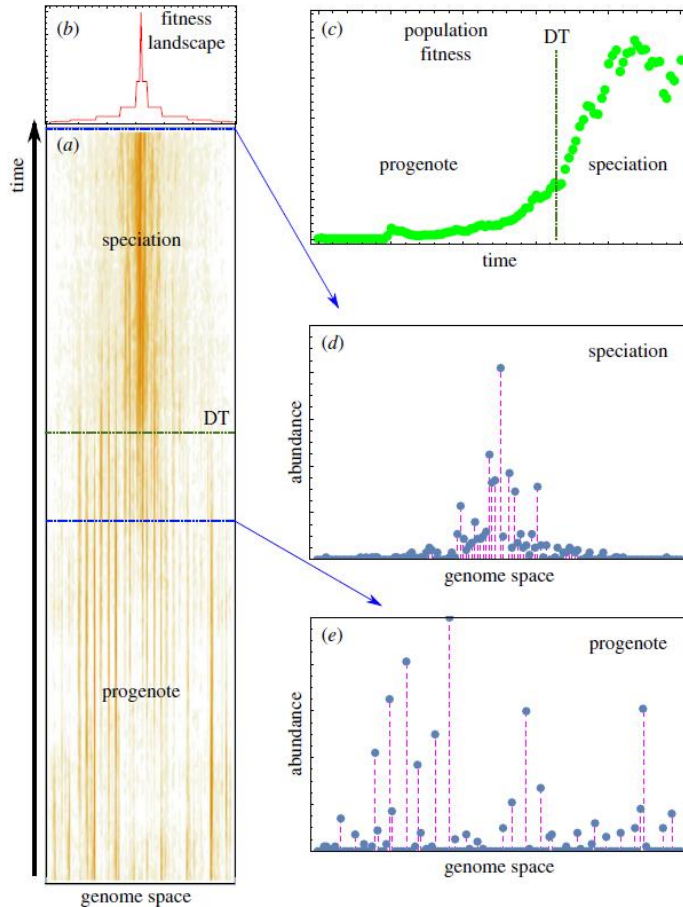


Figure 2: Results of a single simulation of Goldenfeld's early life model. 10,000 organisms with a genome length of 7 explore the genome space over time (a) beginning well stirred (e) and eventually group around the fittest genome (d) from the given fitness landscape (b). As more organisms adopt the fittest genome or ones near it, the overall population fitness increases (c).

An example of a single numeric simulation of the early life model is shown in Figure 2(a) as a plot of genome space versus time where coloration is used to show abundance of a genome in the genome space. From the time evolution of the genome space, we can see that eventually the genome progresses from well mixed in genome space as in Figure 2(e) to a high density of genomes around the fittest genome as in Figure 2(d). This corresponds to speciation and the change in evolution from being horizontally dominated to vertically dominated. The authors observed this behavior for a large range of parameter values, however they observed that the phenomenon of speciation does not occur if the fittest genome has a low fitness peak, there is a high mutation rate, or the degree of horizontal gene transfer is very high.

Now that we have a description of how to reach a LUCA, we can move up from the roots into the tree of life and its structure. To describe the shape of the evolutionary tree, we need to define three characteristics: a species, topology, and an edge length. In the context of a tree diagram like those in Figure 3, we will treat different species as nodes, denoted as points on the trees. When referring to topology for the evolutionary tree, one refers to splitting within the tree, but not when the splitting occurs. The model by Goldenfeld uses two numbers to describe the topology, A_i the number of species (nodes) rooted at species or node i and C_i the total of all A_i rooted at species or node i .⁵ If a tree is perfectly balanced, meaning it has equal splits on both sides of the tree forever as in Figure 3, C scales like $A \ln(A)$ and if a tree is perfectly imbalanced, meaning splitting only occurs on half of the tree, C scales like A^2 . With these as the limiting

cases, it has been observed that real phylogenetic trees are seen to scale with a power law of $A^{1.44}$.⁸ The last characteristic to define is the time between splitting or edge length, which is the distance in time between two splits in the tree. From these parameters they create a model to predict how phylogenetic trees will be drawn using an ecological niche model where the quantity called niche can be thought of as the evolutionary degrees of freedom available for a species to evolve. In the model species evolve with a speciation rate dependent on the amount of available niche and can also go extinct with a certain probability where both changes depend on the controlled variance of the fluctuations in niche. Using this model and varying the fluctuations in niche they observe the barriers of the perfectly balanced and unbalanced trees and conclude from their results that the presence of niche creates a significant footprint in the evolutionary process that cannot be erased.

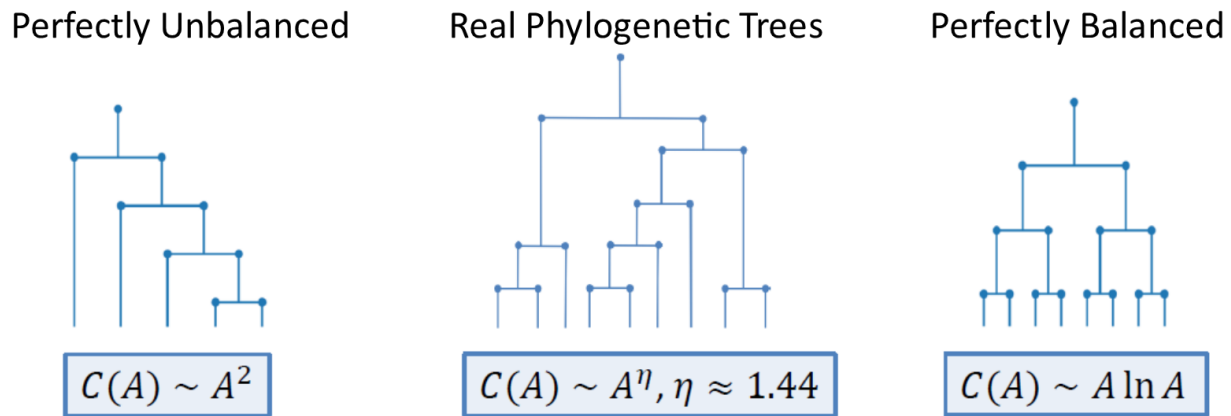


Figure 3: The scaling of the total of all node counts rooted at a given node for a perfectly unbalanced tree, a hypothetical real phylogenetic tree, and a perfectly balanced tree.

References:

- [1]: Davenport, CB. The form of evolutionary theory that modern genetical research seems to favor. *The American Naturalist*. **1916**, 50.596 449-465.
- [2]: Scriven, M. Explanation and prediction in evolutionary theory. *Science*. **1959**, 130.3374 477-482.
- [3]: Woese, CR; Fox, GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.*, **1977**, 74 (11) 5088-5090
- [4]: Goldenfeld, N.; Biancalani, T.; Jafarpour, F. Universal biology and the statistical mechanics of early life. *RSTA Phil. Trans. A*. **2017**, 375.2109
- [5]: Xue, C.; Liu, Z.; Goldenfeld, N. Scale-invariant topology and bursty branching of evolutionary trees emerge from niche construction. *Proc. Natl. Acad. Sci.* (**in review**)
- [6]: Woese, CR.; Fox, GE. The concept of cellular evolution. *J. Mol. Evol.* **1977**, 10, 1–6.
- [7]: Woese, CR. On the evolution of cells. *Proc. Natl. Acad. Sci.* **2002**, 99, 8742–8747.
- [8]: Herrada, EA. Universal scaling in the branching of the tree of life. *PLoS One*. **2008**