The emergence of individual species

Lee, Kyung Suk

December 19, 2008

Abstract

As universal phylogenetic trees can be inferred from molecular sequence comparison, the root of the trees, the early stage of evolution drew people's attention. According to studies, near (and even before) the root of phylogenetic trees, there was a community of living organisms, but they were too unstable and ephemeral to be classified as species. The communal ancestor, dominated by horizontal gene transfer, achieved miraculously huge evolution within comparatively short time period before it finally underwent "Darwinian transition" from a communal state to individual species. This paper summarizes the new evolutionary theory of the early stage of evolution, followed by my speculations on it.

Introduction

Darwin's theory of evolution has been appreciated as the primary principle of evolution so far. His theory could successfully explain how an organism can evolve to others by natural selection, which eventually aroused scientists' interest on the genealogy of life and they tried to find the "very fairly true genealogical trees of each great kingdom of nature" (1) and the root of the tree, the last universal common ancestor. As evolution is understood much better than before in pursuit for the true genealogical tree, surprisingly, it was found that (i) horizontal gene transfer (HGT) plays an important roles in evolution than it was known, and it had even greater importance than Darwinian mechanism in the early stage of life, (ii) life is not originated from a common ancestor and (iii) there were no organism stable enough to have organismal lineage in the beginning. That is to say, an early organism did not have "rigidity" required to remain as one definite organism, but "rigidity" as a stable organism emerged during the evolutionary process.

Phylogenetic Tree

The tree of life was a question that can be hardly answered in the age of Darwin, but, scientists can now draw universal phylogenetic trees based on molecular sequence comparison analysis of molecules which are universally found in almost all organisms.
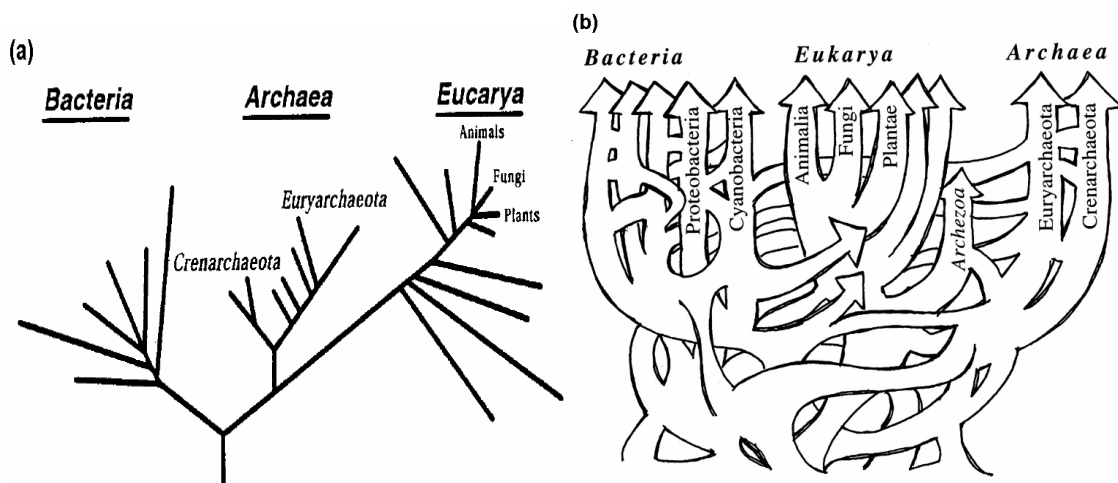

Figure 1. (a) The universal phylogenetic tree based on rRNA sequence comparison. Even though Bacteria and Archea are known to share a lot of metabolic genes, which is not the case in Eucarya (2), Bacteria separated from Archea and Eucarya, first. The shared metabolic genes are the result of HGT, which means that HGT has a very significant effect that can erase vertical inheritance. [Copied and pasted without permision from ref. 1] (b) A reticulated tree of life from a collection of phylogenetic trees, which is a lot more complex and dirtier, but closer to the history of life than (a). [Copied and pasted without permision from ref. 3]

If sequences of the molecules with the same niche in two different organisms are more similar than those in others, then the genealogical relation of the two organisms is closer than that of others. Thus, a genealogical tree can be derived by sorting out organisms in the order of sequence similarity. The reason that universally distributed molecules are used to make phylogenetic tree are evident: they want to know the genealogical relation of as many organisms as possible. It should be noted that universal phylogenetic trees inferred from different molecules show different topology; for example, rRNA based phylogenetic tree and each aminoacyl-tRNA synthetase based phylogenetric trees are different (1). Figure 1a shows a phylogenetic tree based on ribosomal RNA sequence comparison which is believed to be closest to true genealogy tree, if there is any. In Figure 1b, a reticulated phylogenetic tree which was constructed from a collection of phylogenetic trees based on different molecules is shown (3).

HGT and the communal state

HGT is a set of mechanisms that a DNA with in donor cell becomes a "functional part of a recipient cell's genome". HGT is common between bacteria and causes rapid evolution of bacteria, because they can acquire mutations from multiple parents (4). HGT was thought to be merely a secondary mechanism of evolution, but now it is clear that HGT has strong influence in terms of distribution of molecules with universal functions, because if there was no HGT at all, the distribution of molecules should be solely depends on vertical inheritance, i.e. genealogy and that would lead to a simple tree instead of a reticulated tree in Figure 1b, which is not the case. In terms of genealogy of life, HGT was considered as noise. Actually, it is so huge that we can not understand the early stage of life (near the root of phylogenetic tree) without it. Thus, we need to think of a model that treats HGT as a main factor of evolution.

Woese suggested a theory that life is originated from a communal state of ephemeral organisms which share innovations via HGT (2). This theory can explain why the evolution before the ramification into three kingdoms of life was so fast, while the universal common ancestor can not. In this theory, all organisms in the community could utilize horizontally acquired genes immediately without any "friction". Of course, these organisms in the community were ephemeral, because they could accept foreign genomes very easily, and in doing so, they kept changing their cellular design, even within their lifetime. Thus, the evolution of this communal ancestors could be extremely fast; they did not need to go through slow natural selection adaptation mechanism which takes multiple generations, instead, they evolved more like modern bacterial consortium acquires antibiotic resistance within very short time.

In this stage, organisms are thought be a lot simpler than what modern cells are. In modern cells, cellular components are tightly connected with each other, so foreign genes acquired via HGT are hardly compatible with the existing components. This makes the strength of HGT in present day a lot less than in the age of the communal ancestors.
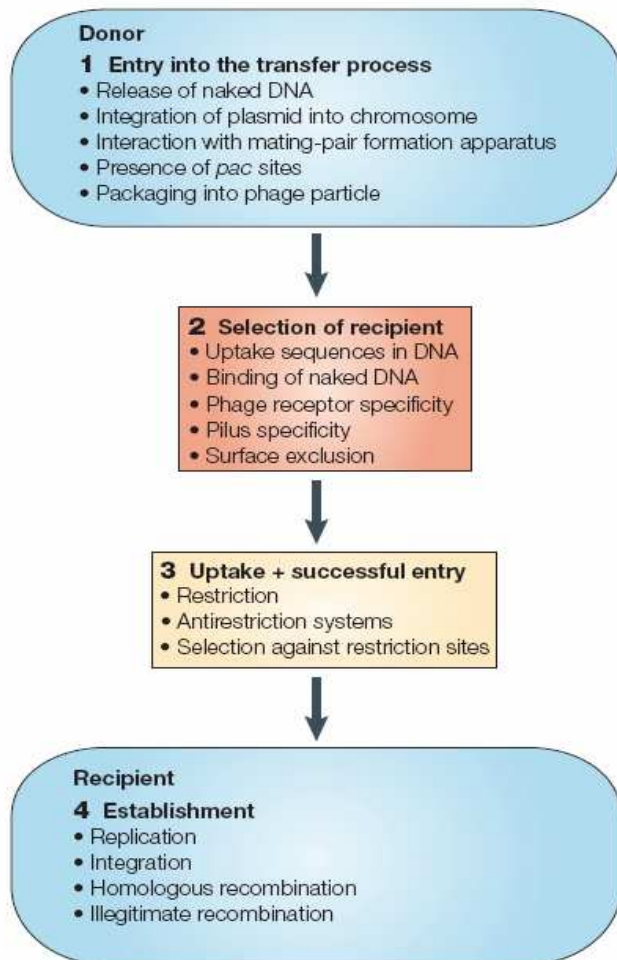
Figure 2. A flow chart that explains the process of horizontal gene transfer (HGT) [Copied and pasted without permission from ref. 4]

Universality and Optimality of the genetic code

Although the theory of Woese sounds plausible, it is hard to find direct evidence that it is really the case, because it is about what happened more than 3 billion years ago (3). However, Vetsigian did a computer simulation to show that (i) the communal ancestor model with HGT results in a universal and optimal genetic code, (ii) a genetic code from one universal common ancestor can not maintain universality without HGT (iii) and can not achieve optimality (5).

The genetic code of modern organisms has both universality and optimality: the genetic codes of all modern cells are the same and the universal code is also optimal to maximize fault tolerance in translation. Vetsigian used mean amino acid distance between codon neighbours as a measure of code optimality. Neighbours of a codon are codons which has only one different base from the codon. If mutation happens or translation apparatus misreads in a codon, in most cases, the codon would be translated as one of its neighbours.

Therefore, if distance between amino acids of codon neighbours is small, the effect of translation error is small. That is why mean amino acid distance between codon neighbours can be a good measure of code optimality. On the other hand, mean code distance is measure of code universality, which is very obvious.

The simulation tests the evolution of an ensemble of codes under Darwinian (vertical) evolution, with and without HGT. Figure 3 shows the simulation results. First of all, communal evolution leads to an optimal genetic code only when HGT is present. Secondly, even if they begin with one genetic code, they eventually lose universality without HGT. Lastly, the probability that a randomly chosen genetic code happen to be optimal is almost impossible. Even if this is a computer simulation, this result strongly supports the communal ancestor model, and gives counterevidence against every evolution models without HGT. This implies that the first organism which is refractory to foreign perturbation such as HGT emerged only after the genetic code is unified and optimized.
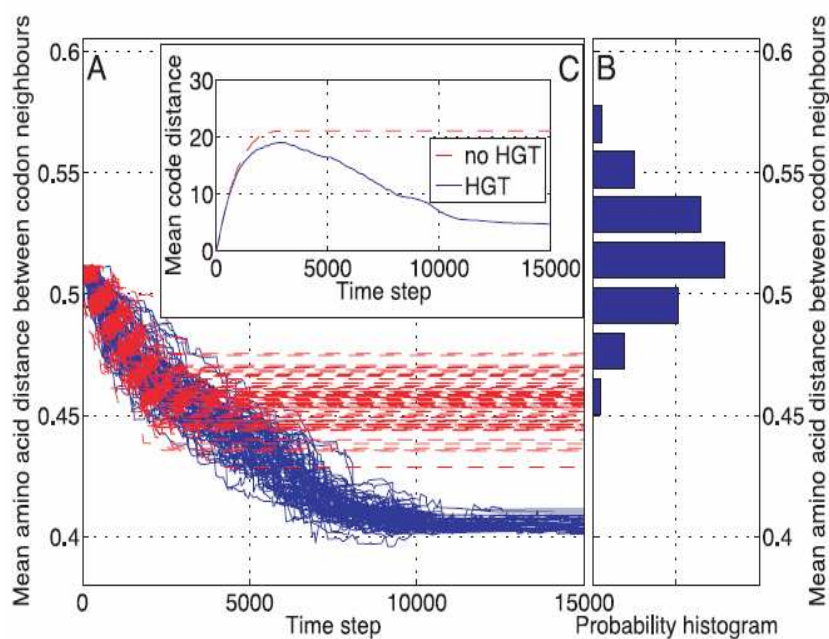

Figure 3.

(a) Communal evolution with (blue) and without (red) HGT. The vertical axis is mean amino acid distance between codon neighbours which is measure of optimality. The shorter the distance is, the more optimal the code is.
(b) Probability distribution of optimality of randomly generated codes.
(c) Time evolution of mean code distance.

[Copied and pasted without permision from ref. 5]

Darwinian threshold, Darwinian transition

In modern era of evolution, we know that HGT is negligible compared to vertical (Darwinian) inheritance. The modern cells are almost perfectly "rigid" enough to define individual species and genealogy is very clear and discrete. In contrast, the early stage of evolution, near the root of genealogical tree of life, was dominated by HGT and organisms were ephemeral, as explained in the previous part of this paper. So, at some point in the

past, the evolutionary mode has changed from horizontal mode to vertical mode, so-called Darwinian mode. At that point, the first modern cell emerged, and the genealogy started. Woese named this point the Darwinian threshold or Darwinian transition (6).

This point is where evolution can be related to the statistical mechanics. The first question would be whether this transition can be understood as one of phase transitions which have already been studied in terms of statistical mechanics. I can not answer the question, but I can say that Darwinian transition is not such that evolutionary mode flips and HGT disappears all of sudden, although it is not explicitly stated in references. One well known example case that HGT is still active is antibiotic resistance of bacterial community. Even more, the evidence of HGT among several mammals and tetrapods was found recently (7). The species in bold shown in Figure 4 have SPACE INVADERS (SPIN) sequences, a DNA transposon family. It is surprising in that the species distribution is broad (they are not closely related) and HGT is active among multicellular eukaryotes like mammals. As can be seen, HGT did not cease to happen, even after Darwinian transition. Merely, the strength of HGT gradually decreases, while that of vertical inheritance increases.
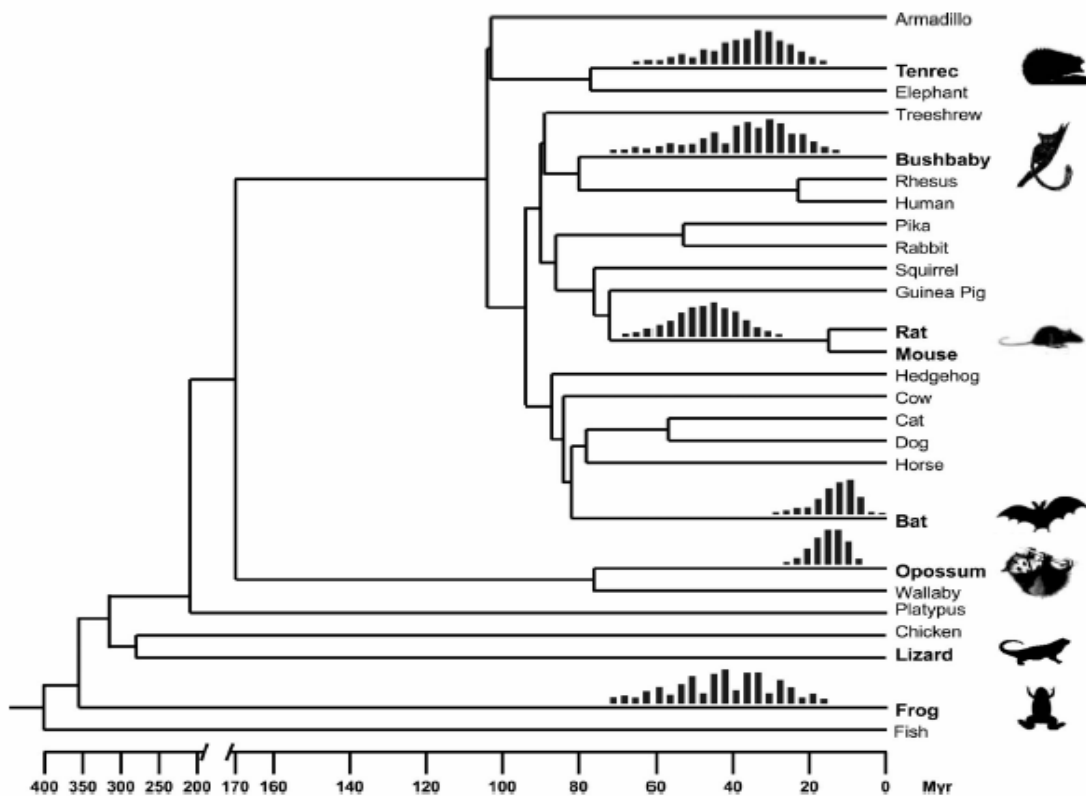

Figure 4. phylogenetic tree of species based on genomic sequence similarity. Species written in bold has SPIN transposon, and the histogram represent amplication timing of SPIN. [Copied and pasted without permision from ref. 7]

Analogy between evolution of organisms and physical annealing

In the communal ancestor, HGT could happen almost without "friction". It is because cellular design of organisms was simple and modular. In other words, cellular components were not connected or loosely connected so that each component was easy to be replaced. As cellular designs evolved in order to process more complicated tasks, connection among cellular components became tighter, and they started to work in cooperative fashion. This tendency caused the strength of HGT to decrease. After a cooperative cluster of cellular components were established, HGT of a component within the cluster would refrain, unless the component can work with the existing cooperative cluster, even if better cellular component is acquired via HGT. In this case, discarding HGT is more beneficial because the advantage from the improvement of a single component is less than the disadvantage from losing functionality of entire cluster. HGT becomes more and more suppressed in this manner, as cellular design evolves towards modern cells.

This process is similar to physical annealing. System starts from fluid phase at high enough temperature that no structure can form, and temperature drops as time goes. As temperature drops, various metastable structures form and get melted again. Eventually, at some temperature, the first and most stable structure forms and remains until the end, i.e. crystallizes. As temperature drops further and further, less stable structures form and additional structures grow on already formed structures (2). At the end, system becomes solid. In this analogy, the first crystallization event in annealing system, which remains stable until the end, is the counterpart of Darwinian transition. Then, what would be evolutionary counterpart of temperature? Why does it have to keep decreasing?

Speculations: Network representation of cellular evolution

In my opinion, the genetic annealing model is nice for illustrative purpose, but applying the knowledge of physical annealing system to understand evolution would be difficult, because we can not hope that topology of cellular networks and the annealing system would be the same. The model system used to study cellular evolution near Darwinian transition should be able to represent the topology of cellular componentry network, because HGT depends on the connectedness of cellular componentry network and Darwinian transition depends on the change of the strength of HGT, in turn.

Thus, I suggest that network representation of cellular design and study of computer simulated growth of the network could be more helpful. Cellular evolution written in terms of network representation would simplify the view very much, too. For example, a cellular component is a vertex and inter-component connection is an edge. This simple view would be acceptable, since we can forget about the evolution of genetic code which can't be included in simple network representation. Luckily, we know that Darwinian transition happens only after unification and optimization of a genetic code.

Speculations: horizontal and vertical evolutionary dynamics

According to references, the main role of horizontal evolutionary dynamics is sharing innovations among organisms, while that of vertical evolution is adaptation to environment so that the organism can be naturally selected. Horizontal dynamics is mainly general and compatible, while vertical dynamics is specific and idiosyncratic. In terms of network representation of cellular design, this would mean that (i) horizontal evolution mainly replaces or adds vertices (ii) vertical evolution mainly adds edges to networks.

In this network representation, evolutionary counterpart of entropy of a cell (a network) can be defined as the number of possible HGT, i.e. the number of possible vertex replacement and vertex addition. Therefore, vertical evolution which only adds edges always work toward the direction to decrease the entropy of a cell (a network), i.e. decrease the number of ways that a cell can be modified by HGT or the strength of HGT. Actually, this makes sense: vertical (Darwinian) evolution is a mechanism to find the fittest, stable optimum. This would answer the question why the evolutionary temperature keeps decreasing. The pressure of natural selection is driving the cellular evolution towards the state with the lowest "entropy" of a cell (and no way to perturb).

Speculations: tilting evolutionary temperature gradient

Many of phase transitions in physics depend on the temperature of systems. So, reversing the direction of temperature change, of course, reverse phase transitions can be seen. However, evolutionary temperature always decreases. Is it possible to tilt evolutionary temperature gradient?

Again, microbial consortium draws my attention, because the system did not crystallize yet, but is still going through HGT. So, this case is an example that shows tilting of evolutionary temperature gradient. In my opinion, the key difference between microbial consortium and already crystallized organisms is that microbial consortium can not survive without rapid innovations. I suppose that an organism would "crystallize" if it can survive after "crystallization", or after finding the fittest optimum of natural selection mechanism. I suppose, while it does not feel threat to its survival, it just continues on its search of fittest cellular design. On the other hand, when it comes to face threat, HGT is activated. So, I suggest that in the simple network representation of a cell, threat to extinction works as external heat, which arouse thermal fluctuations (HGT). The answer to the question of whether tilting evolutionary temperature gradient is possible would probably be yes.

Speculations: Network simulation

Since Darwinian transition and the emergence of the first organism with rigidity happened extremely long time ago, direct experimental justification is not possible. Here, the sequence comparison method would not work, either, because Darwinian transition is near HGT dominant stage of evolution, which would erase the information from the sequence.

So, the best available method is a computer simulation. That is another reason that I suggested network model, which is computer-simulation-friendly. Hopefully, simulations on the cellular network with control parameters such as "external heat" would bring us some interesting cluster formation behaviors, which is related to the evolutionary dynamics of Darwinian transition.

References

(1) C. Woese, Proc. Nat'l. Acad. Sci. 97, 8392-8396 (2000)

(2) C. Woese, Proc. Nat'l. Acad. Sci. 95, 6854-6859 (1998)

(3) W. Doolittle, Science, 284, 2124 – 2128 (1999)

(4) C. Thomas, K. Nielsen, Nat. Rev. Microbiol. 3, 711-721 (2005)

(5) K. Vetsigian, C. Woese, N. Goldenfeld, Proc. Nat'l. Acad. Sci. 103, 10696-10701 (2006)

(6) C. Woese, Proc. Nat'l. Acad. Sci. 99, 8742-8747 (2002)

(7) J. Pace et al., Proc. Nat'l. Acad. Sci. 105, 17023-17028 (2008)