

# Rethinking of Evolution

Hangjiong Chen

## 1. Introduction

Life began on earth about 3.5 billion years ago, but how the primitive form of life came into being at the first place is a forever mystery. But chemistry tells us that a chemical reaction will occur if conditions are right, regardless of whether it occurs in laboratory or in nature. The formation of early life is a mere chemistry process when chemical components for life existed on the nascent earth: lipids for cell membranes, carbohydrates for energy and structures, ribonucleotides for genetic materials and amino acids for proteins.

Early earth must be a lucky planet in the universe in which there existed an environment blessed with a mix of all those chemical components, a cozy incubator from which life could rise and develop. When the conditions are ripe, it's just a matter of time for the primeval form of life to emerge.

Life today is so rich in forms ranging from simple bacteria and archaea to highly complicated and diverse eukaryotes. Despite all this, all modern living organisms use the same set of amino acids, same set of genetic codons, same set of nucleobases, and same set of lipids, suggesting that life as we see today originates from a single ancestor in a single place on the primordial earth. Then a long journey of evolution brings early life to such an extraordinary diversity today. From the very beginning, life has strove for existence, renewal and flourishing on itself and orchestrated its own entire life cycle from inception, embryonic development, birth, maturation, reproduction, and finally to death without input of external instruction. All this starts from the genome enclosed in the nucleus in the cells. The genome is the most glorious wonder in the entire universe.

How did such a wonder arise in the ancient earth is not only intriguing, but also awe-inspiring, worth every effort to understand and explore. I have put my random thoughts about evolution of the genome in this post and come up with criticism on the theory of natural selection as a misinterpretation of evolution.

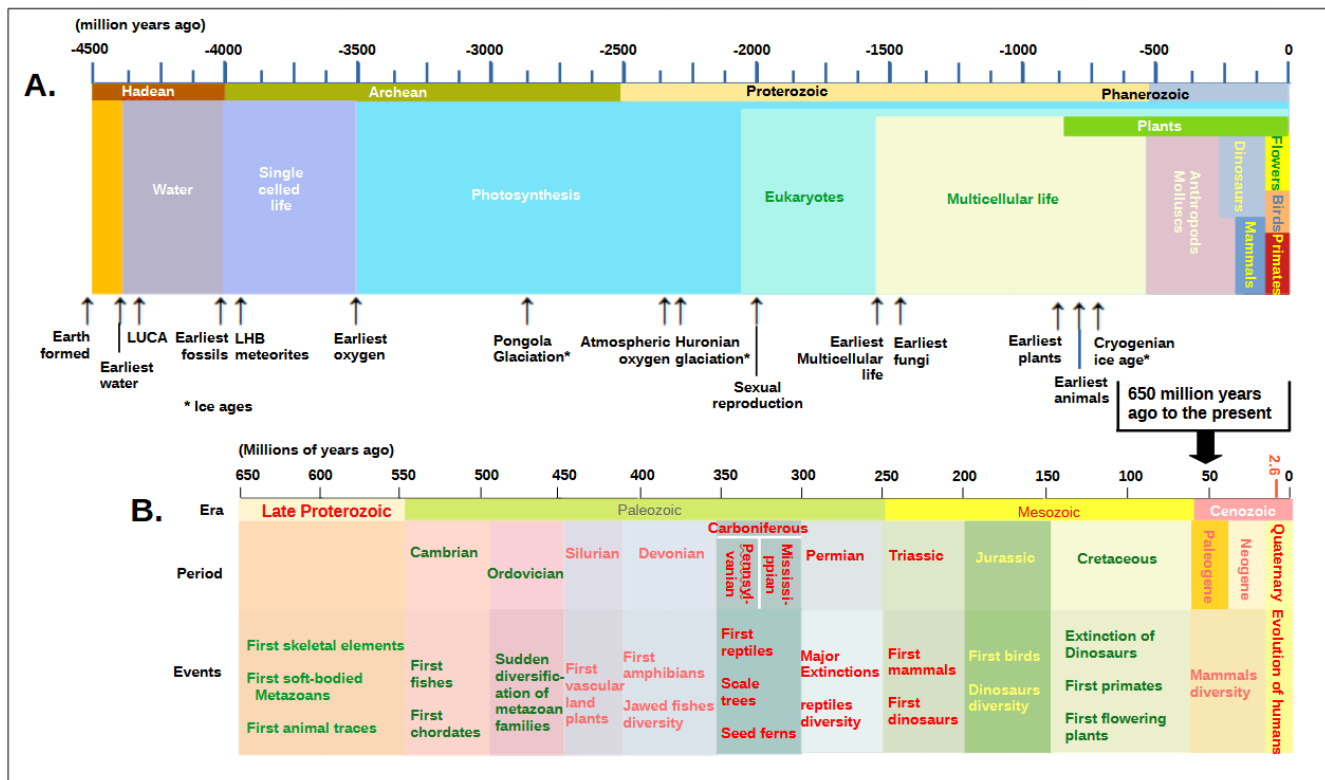
## 2. Life Timeline on Earth

Life on earth can be traced back to 3.5 billion years ago, about 1 billion years after earth was formed. Figure 1 shows the timeline of life evolving from the most primitive forms to simplest single celled forms to modern humans, although it is approximate only.

A striking characteristic of the timeline is that it dedicated a stunning long period of 2 billion years (from 3.5 to 1.5 billion years ago) to the development of life in its very simple forms, including bacteria and single celled eukaryotes. This signifies the difficulty of life arising and surviving in the primeval time. The next 1 billion years witness the rise of multicellular life like fungi and slime molds. Around 500 million years ago, an eon of accelerated evolution, living organisms began to diverge into all forms and complexities, resulting in abundant new species of plants and animals to appear and dominate the earth. This eon is divided into a few geological periods.

In the Cambrian period (about 539 to 485 million years ago) the earth endured large changes from the preceding geological period in climate, earth's biosphere, and geography that impacted life of that time with the greatest significance. The changes caused the destruction of natural environments and mass

extinction of species, but more importantly led to the emergence of many new species, some of which started to move from ocean to land. The beginning of Cambrian explosion heralded the acceleration in biotic diversity, though the species were still as low and simple as comb jellies, sponges, corals, etc.



**Figure 1.** Timeline of the evolution of life on Earth (Adopted from Evolution on Wikipedia and Britannica). Geologic period Phanerozoic comprises the Paleozoic, Mesozoic, and Cenozoic periods

Until Devonian period from 419 to 359 million years ago, arthropods insects, spiders, centipedes, etc. became part of the land ecosystem, and vertebrates started to move to the land as well. In Cretaceous period from 145 to 66 million years ago, more species of mammals, birds, and flowering plants appeared. In this period, first primates emerged and all dinosaurs went extinction. The last 66 million years of earth's history is marked by the dominance of mammals, birds, and flowering plants. More insects, butterflies, moths, fishes, amphibians, and reptiles with modern forms take over the earth long after mammals and birds emerged. Later appearance rewarded these low species with more advanced morphology and more sophisticated cellular and biochemical processes than their earlier cousins.

Diversification of primates occurred around 50 million years ago, while the apes, which were evolved from primates and gave rise to the early humans, emerged some 15–20 million years ago. Early humans called hominins diverged from the apes from 14 to 2 million years ago, a time span that is very short on the evolutionary timeline, giving the large morphological changes between apes and hominins. True modern humans are now generally believed to emerge in Africa approximately 300,000 years ago, and then migrate to other continents some 100,000 to 50,000 years ago.

Uneven distribution of evolutionary events on the evolutionary timeline is intriguing. Why did it take more than 3.5 billions of years for life to evolve to aquatic plants and animals that are very low and simple comparing with modern day life, but it took only 500 millions of years, especially the later 100

millions of years, for life to flourish with millions of species of all complexities and forms? What is hidden behind this timeline of the evolutionary history of life?

### **3. Randomness Brought Life to the Nascent Earth**

Proteins, RNA and DNA are not ordinary molecules, they are independent chemical entities that are life in its simplest forms. These molecules are so tightly interlinked that one can't be produced without the other two. A pressing question is how the initial proteins, RNA and DNA could be produced in the incubator? And how could these independent chemical entities become interlinked and assembled into the earliest form of life? What must be true is that the life incubator is an environment in which the conditions favored the chemical reactions to produce proteins, RNA and DNA, possibly facilitated with unknown non-enzymatic catalysts.

From pure chemistry point of view, amino acids are much simpler in terms of chemical composition and structure than ribonucleotides, which are compounds consisting of a nitrogenous base, a pentose sugar ribose, and a phosphate group. A dipeptide is produced when the carboxylic acid group of one amino acid reacts with the amine group of another to form a peptide bond. Dipeptides could elongate at both sides by accepting more amino acids, resulting in polypeptides. The polypeptides so produced would be random but infinite in sequence, forming a pool of polypeptides in the primeval incubator. On the other hand, it's not straight-forward to synthesize a single ribonucleotide from three totally different small molecules in the absence of some enzymatic assistance. It could be expected that ribonucleotides would be produced in low quantity, though it might be barely sufficient for RNA production. With ribonucleotides available in the incubator, it's still not straight-forward to link ribonucleotides into a polymer that is strictly ordered from 3'-5' orientation. A likely scenario is that polypeptides are more likely to be produced from amino acids than ribonucleic acids from ribonucleotides, and because of this, polypeptides are more likely to be produced earlier and in larger amount as well.

RNA synthesis was possible in the absence of enzymes. One possibility is that some unknown special surfaces in the incubator could attract ribonucleotides to adhere. If ribonucleotide molecules laid on the surface close enough, adjacent ribonucleotides could form 3'-5' phosphodiester linkage. This reaction could continue infinitely, producing RNA molecules of various lengths. Replication of RNA molecules could be achieved in similar fashion except that the complementary bases might be snapped into positions on the template through hydrogen bonding. This step could continue as a process, producing a complementary chain in the form of double stranded RNA with the template chain. Like all other polymerization reactions, RNA production must be very low efficient.

The deoxy form of ribonucleotides – deoxyribonucleotides is more stable and fits better to be genetic materials. In living organisms today production of deoxyribonucleotides from ribonucleotides is the result of an extra reaction, in which the 2' hydroxyl group of the ribose is replaced with a hydrogen atom catalyzed by ribonucleotide reductases. Reduction of ribonucleotides in the ancient time could be different, but possible if there existed some special non-enzyme catalysts in the incubator. DNA could be produced and replicated on a special surface similar to RNA synthesis, and the formation of 3'-5' phosphodiester linkage could be more specific and accurate in the absence of the 2' hydroxyl group. Despite the general consensus that DNA is the last component to join the life system because of that extra reaction, it couldn't be ruled out that DNA is a contemporary fellow of RNA and proteins. Every possibility is possible when facing a magic life incubator in that unknown world.

It seemed possible for proteins, RNA and DNA to appear independently of each other, but it makes more sense and logic and agrees more with what the nascent earth would look like to assume that it's proteins that appeared first and ignited all possibilities in the incubator. Some polypeptides could fold

into specific three dimensional structures that conferred peptides enzymatic activity or structural capability. If one peptide molecule out of 100 millions could fold on itself to form an enzyme, 100 different enzymes could emerge when the size of peptide population reached, say, 10 billions. The debut of enzymes and structural protein would have profound impact on what could happen early in the incubator, which supports the idea that enzymatic assistance were involved more or less in virtually every aspects in the early life.

The early enzymes, if available from the random peptide pool, could begin to act on substrates present in the incubator, and accelerate the chemical reactions between substrates, implying the likelihood of rudimentary RNA polymerase, DNA polymerase, and aminoacyl tRNA synthetase in action. RNA polymerase and DNA polymerase were rudimentary since their catalytic activity was more likely to add substrates one by one to the 3' end without templates in a totally random fashion. In addition, other enzymes could exist in the incubator to catalyze the synthesis of ribonucleosides, ribonucleotides and deoxyribonucleotides from basic chemical components, although specificity and efficiency must be compromised.

Life in its earliest moment could be conceived as merely a random existence. Polymerization of ribonucleotides, deoxyribonucleotides, and amino acids is a type of random production, and products are all random in terms of sequence and length. Over a long period of time, numerous things, including a variety of lipids, carbohydrates, and other compounds of unknown functions, could be produced and accumulated in the incubator and co-existed as a comprehensive pool of mixed compounds. The beauty of random production in the dark, chaotic age is that randomness could lead to the availability of biochemically significant molecules of different kinds if the random pool is large enough.

Synthesis of RNA on a special surface or by RNA polymerase was random, forming a large and ever-increasing heterogeneous RNA population. Among the population were sequences that could fold on itself into double-stranded forms to assume secondary structures similar to modern tRNA and rRNA. Like peptides, if one RNA molecule out of 100 millions could fold on itself to form tRNA or rRNA like structures, about 100 tRNA like or rRNA like molecules would emerged when the size of RNA population reached, say, 10 billions. If increase in RNA population was faster than the rates of natural degradation, the number of tRNA like or rRNA like molecule would increase. The random RNA population in the pool was the original source of all types of RNA. The RNA molecules with the secondary structures characteristic of tRNA and rRNA would become the predecessors of modern tRNA and rRNA, while those RNA products without secondary structures presumably would be the predecessors of modern mRNA.

When the things in the pool were moving around aimlessly in the dark, the right components could happen to come across and interact, forming special structural complexes. The first complex would most likely be rudimentary ribosomes or protoribosomes for protein translation. It would form when rRNA-like RNA bumped into ribosomal-like proteins with affinity for rRNA-like RNA. Such a complex would evolve slowly in size and complexity as more components joining in once they became available. Another possibility was that some random peptides could aggregate with RNA polymerase or DNA polymerase in some way to form masses that could act as a platform for the transcription of RNA and replication of DNA. Such platforms must be poor in performing its functions in terms of output and accuracy, but at least biosynthesis of RNA and DNA became possible and the incubator established itself as the common home for proteins, RNA and DNA.

Among all tRNA like molecules, at least one tRNA could carry one specific amino acid at its 3' end and bear an anticodon in the opposite site. The covalent attachment of an amino acid to the tRNA 3' end

could be catalyzed by primitive enzymes and show some specificity for anticodon. The pool contained a tRNA population large enough to represent every amino acids in the incubator. In this way the rudimentary ribosomes would serve as a platform for peptide synthesis. When rRNA-like molecules held a mRNA-like molecule, a tRNA-like molecule that bore an amino acid could align along the mRNA through anticodon matching. The complex so assembled would be the most basal form of peptide synthesis platform. RNA molecules after tRNA and rRNA were all potential mRNA and they formed an enormously large mRNA pool. A small number of mRNA-like molecules could serve as templates to produce peptides of random sequence and reasonable length on the primitive platform.

All biochemically significant peptides or RNA in the absence of DNA templates suffered a lethal deficiency. The chances to reproduce the same peptides or RNA was nearly negligible, if not zero once they were degraded. The possibility existed that their loss could be compensated by other peptides of different sequences but with similar functions, and as a result, their impact would be continuous as usual. In all likelihood biochemically active peptides in the random peptide pool would be the most critical part of the process to form nascent life because they made all other chemical reactions possible.

Life is not a random existence per se, but a remarkably ordered and consistent living entity. Nascent life must move out of randomness by establishing consistency through controlling all reactions vital to life with protein catalysts – enzymes. In this remarkable transition, nascent life would remain largely random, but randomness was DNA based. In fact, in the face of large population of random peptides, it was likely that nascent life did start with DNA based randomness. In either case, the final outcome was the same except that whatever happened in a DNA based system was meaningful and fruitful as a living entity in the long run. When DNA took the center stage in the development of primitive life, all changes would be preserved in the storage place, and life began to take shape by exploiting various products in the pool and coalescing and assembling them into DNA replication machine, RNA transcription machine and protein translation machine.

In the early life incubator there existed a single DNA molecule, which was synthesized and elongated by incorporating random deoxyribonucleotides at the 3' end through the action of template-independent DNA polymerase. In the absence of modern DNA replication complex and RNA transcription machine, the replication and transcription processes were awfully egregious. Each replication process introduced considerable amount of mutations into the sequence, quickly turning this single DNA sequence into a heterogeneous DNA population. Each single DNA sequence in the population could be transcribed into RNA molecules and formed a large and even more heterogeneous RNA population. What happened after this is what has been described earlier.

Establishing consistency and discipline for life can't rely on enzymes produced by random events, but it could benefit from and even be accelerated by infrequent random enzymes. It's all up to speculation regarding how DNA was polymerized from deoxyribonucleotides and grew in length. It isn't important in details because it must have occurred in the very early stage of life in the absence of the most primitive biochemical machineries. What's important is that early DNA molecules were random in length and sequence, and the synthesis was more efficient in the presence of random DNA polymerase.

Early RNA could be copied from random locations on random DNA templates. All RNA molecules were random in length and sequence and formed a RNA population in the pool that became the early source of all types of RNA, the predecessors to modern day tRNA, rRNA, and mRNA. Translation of mRNA predecessors into proteins on the primitive ribosomes released more random polypeptides into the pool, some of which could carry different biochemical activities. These proteins were different from the proteins that were generated solely through the templateless mechanism in one aspect. Based on the

DNA templates their amino acid sequences had been preserved in DNA and could be reproduced with reasonable probabilities in this stage. From the moment on, proteins formally and solemnly came into the primitive world of life as a reliable and durable key building blocks.

As DNA molecules grew in length infinitely in the early stage to form life, the error-prone nature of DNA replication and augmentation promoted incorporation of abundant point mutation-like errors into DNA sequences, resulting in the buildup of a colossal DNA population composed of random lengths and sequences, a treasure of all possibilities. This extremely heterogeneous DNA population increased the likelihood to produce a much larger pool of proteins that bore wider spectrum of functionalities, including more and better enzymes and various supporting protein factors. Meantime much improved protein pool made the entire life activities more accurate, more dependable, more stable, and more predictable, gradually erasing randomness and bringing the system to an orderly and consistent state. Over time the templateless peptides would be gradually overtaken by template-based peptides and diminish in quantity and roles.

Randomness has played a pivotal role in early life development. As an increasing number of functional proteins became available for DNA to execute roles as genetic materials, transition from randomness into template based consistency at the population level accelerated considerably, making life one step closer to become the system conceptualized by the self organizing principle of life. The first primitive DNA genome emerged when a DNA molecule started to show sequence loci that could be regarded as gene-like for protein synthesis. With more functional enzymes emerging, metabolism pathways started to appear in early life, a process referred to as evolution marked the completion of its first step in a journey that has lasted more than 3.5 billions of years ever since.

Randomness and consistency are incompatible and paradoxical with each other in the nascent life system. Life starts in randomness and reaches maturity in consistency. From a chronological point of view, randomness results in consistence, and consistence reduces randomness, and reduced randomness lowers the chances to generate new functional proteins, and the lowered chances for new functional proteins slow down the system to mature and advance. In other words, reduced randomness slows down evolution of life. The majority of functional proteins were first produced in the absence of DNA templates, but they could be actively engaged in all aspects of nascent life. Consequently, DNA replication would be less random, which limited the heterogeneity of the DNA population and reduced the chances to generate template-based functional proteins. Randomness is paradoxical, but it has been the most effective, albeit time consuming, approach to establish precision, consistency and disciplines that govern all biochemical and cellular operations of life. Introduction of more randomness at the early stage of life was the clever way to turn a random system into a consistent entity, is it?

The appearance of the early forms of genetic machine, albeit only a trace resemblance with their modern counterparts, improved DNA replication and RNA transcription to a great extent. As the spectrum of enzymes became broader and catalytic activity improved bit by bit, the incubator started to show primeval, but relatively functional metabolism pathways, including the pathways to produce ribonucleosides, ribonucleotides, deoxyribonucleotides and amino acids. Gradual production of structural proteins and enzymes transited to be template based, their availability somehow became reliable, repeatable, and stable, which further enhanced the consistency and accuracy of DNA replication, RNA transcription, and protein translation. More importantly, DNA, RNA and proteins started to form an interlinked and inter-dependent system for life. Establishing such a system helped to further decrease mutational rates and reduce randomness in their production, forming a reciprocally influenced development cycle.

Numerous recombination of multiple DNA molecules into single ones coupled with random point mutations accelerated the emergence of an all-potent DNA molecule – the first minimal genome – for life. The sequences of the minimal genome slowly transitioned into an array of genes that could perform the very basic functions vital for the primitive life. This minimal genome became sustainable on itself with a complete information flow from DNA to RNA to proteins and was able to encode all proteins and RNA elements required to become the most basic, but independent, self sufficient form of life. Quickly it dominated DNA population, becoming the common ancestor of many life and finally all life in the early stage. Life that descended from it shared the same set of amino acids, same set of anti-codons, same set of bases, same set of metabolism pathways, etc. This DNA molecule is eventually the common ancestor of all modern life as well.

As time went by, the peptide pool hosted more enzymes, ion transporters, proteins for cell division, protein filaments to make cytoskeleton of the cell, etc. Nascent metabolism pathways could have started to generate energy from carbohydrates and produce key chemical compounds for building basic cellular structures, especially cell membranes and cell walls. The self organizing nature of proteins of various functions allowed them, when mixed together, to perform whatever the functions individual proteins could perform and perform as single entities. For example, if some proteins could divide a cell apart into two when mixed together, they would divide the cell apart into two when mixed. If a protein could transport sugar molecules across the cell membrane, it would transport sugars across the cell membrane when embedded in the lipid bilayer. The time was finally ripe for the minimal genome to be enveloped in a lipid bilayer membrane, forming the earliest primitive cell – single celled life. This single celled life relied on a single set of genetic codons corresponding to a single set of amino acids for protein synthesis. The biological significance of the cell membrane was that it shielded the genetic machines and metabolism pathways from interference of other random peptides in the pool. As a result, this single celled life quickly became dominant in the incubator through replication and division. Today all forms of life are proud of the descendants of this grand single celled ancestor.

As early life continued to develop and evolve over enduring time, it eventually transformed into the real single celled life – mature and complex enough to be called species. These species could bear some basic capacity to survive and prosper in the face of environmental changes and attack by other species. In the meantime the genetic information flow continued to improve and preserve to present days. Establishing an orderly and consistent genome state from randomness is the very essence of evolution of the early life system at the cost of time in the form of endless trial and error.

The on-going debate of what is the earliest form of life is totally meaningless and worthless. The RNA world, in which self-replicating RNA molecules proliferated before the appearances of DNA and proteins, is improper and naive. It's like a little duckling looking for her mother. She calls mom whenever she sees a duck that looks like her mom. It created more and harder questions than the question it tried to answer.

An environment or system that is dedicated purely to RNA synthesis or protein synthesis could be created only in the laboratory. It was utterly unlikely for the nascent earth to host an environment that was rich in chemicals only for RNA production or only for protein production. Amino acids and ribonucleotides must have formed a mixed system if amino acids didn't appear first in the system. In a mixed system, nothing could prevent amino acids from linking into peptides when ribonucleotides polymerized. Purely in terms of chemistry, peptides were the things to be produced prior to RNA synthesis in such a system. If peptides were produced in the system, they couldn't be excluded in the process of RNA replication even though certain RNA had catalytic activity and could self-replicate. Is a mixed system very hard to think of? Ignoring the early earth being a mixed system is either biased or

simple minded. If we took one step back and assumed there was a RNA only world, then was there a protein only world or a DNA only world on the same planet earth? Life emergence was not directed by someone like in film production, but it happened spontaneously in a total random fashion. When and how could these three separate, independent worlds finally come together to create primitive life? Making early life isn't like making dish, each ingredient is available to you on the kitchen table and can be added to the cooking pan any time at your will. This is a puzzle bigger and harder to answer than what is the earliest form of life itself.

Not all RNA molecules can fold into double stranded forms. The catalytic activity of a RNA molecule is stringently dependent on the three dimensional structures that are built upon its nucleotide sequences, thus lacking generality to be the self-replicating genetic materials. Furthermore, the activity is weak and its substrates are limited. Research to prove that RNA can replicate itself is purely performed for the sake of proving that RNA can replicate itself, while not thinking about the relevance of RNA templates used in the experiments and the experiments themselves to the environment where life originated. RNA self-replication observed in such conditions can't ascertain that life starts from RNA. It is too forced to link RNA self-replication research to the origin of life.

Self-replicating RNA molecules are almost exclusively complicated in three dimensional structures, containing a substantial amount of self-folded double-stranded regions, which makes double-stranded RNA molecules quite stable and resistant to chain separation. Such RNA molecules don't seem suitable to carry genetic information. If they did carry genetic information, what kind of genetic information could they store when proteins are not part of the RNA world?

The premise for envisioning an early life system is that all required chemical reactions occur randomly on the nascent earth, albeit at low, even insignificant rates. A period of more than half a billion years to form the primitive life manifested a process that is driven by chances, lucks, and coincidences, all of which are characteristic of randomness to build up a fully consistent state. It also attested the utmost difficulty to establish single celled life system from ground zero purely through random processes. When we think of the origin of life, the most fascinating part isn't how protein, RNA and DNA were produced in the first place, but is in what environment in which protein, RNA, and DNA could be produced in the first place. Life incubator seems to be a plausible idea to make a point that there was such a place on the nascent earth where life originated. However, it was impossible to physically describe such a place that could have ever possibly existed on the earth. Was it as big as a puddle of water, or a small pond in a neighborhood, or a large pond near a highway, or even as large as a lake? Where could it be located if such a place did exist? In a cave or at the bottom of sea? Another mystery is that how this 500 million year long evolution timeline wasn't devastated, even disturbed by some geologic and climate upheaval considering that the nascent earth was unstable in its own right.

#### **4. Quiet 3.5 Billion Years before Cambrian Explosion**

The evolutionary process isn't evenly distributed along the timeline (Figure 1). The first forms of life appeared on Earth about 4 billion years ago, and they slowly developed to become the ancestors of the unicellular microorganisms bacteria and archaea, the dominant forms of life for about 2 billion years. Eukaryotes then emerged, likely from archaea, about 2 billion years ago. Eukaryotes evolved slowly before the earliest plants and animals appeared 800 million years ago. It's until 540 million years ago the earth experienced an explosive increase of low animals and plants, an event called Cambrian explosion. Cambrian explosion lasted for about 13 to 25 million years and resulted in the divergence of most modern metazoans. Since then, flowering plants, amphibians, reptiles, birds, and mammals emerged in astonishing speed, leading to today's mind-boggling biodiversity.



In genetics, any sequence alterations in the genome of an organism are mutations, or genetic mutations. Point mutations are completely random and refer to single base substitution, deletion or insertion, a type of replication errors. Mutations also include deletions or insertions of short pieces of DNA sequences. Large-scale mutations refer to changes that alter chromosomal structure in a considerable degree. Gene duplications are a type of sequence amplifications, while chromosomal translocations and chromosomal inversions are types of DNA rearrangement that change the orientation or location of a segment of DNA in the genome. Deletions of large chromosomal regions can lead to loss of the genes within those regions. Deletions or insertion of a segment of DNA sequence can bring together separate genes to produce functionally distinct hybrid genes. All genetic mutations can be lethal if they disrupt genes that are vital to the organisms.

Mutations that are more relevant to evolution are point mutations and gene duplications. Point mutations accounted for majority of mutations introduced by DNA polymerases during germline division and became more frequent when the fidelity of DNA polymerases is reduced. Point mutations can be lethal if they shift the reading frames for protein translation. Normally the DNA polymerases replicate DNA with high fidelity to keep mutational rates low and biotic world stable. Gene duplication is a process to make a new copy of DNA fragment that contains a gene. is a special type of DNA rearrangement. Gene duplication is a major mechanism that genome generates new genetic material for the evolution of new species. Gene duplications remain common in most species today.

From the timeline of the evolutionary process, the time taken for living organisms to evolve from the very beginning to present day can be divided into three stages (Figure 2). The primitive life system was built in Stage 1 from basic chemical components over a period of 500 million years. If Stage 1 is not considered as evolution, the evolution process can be divided into two stages, slow evolution and fast evolution. Division of evolution into slow and fast has profound implications about how evolution really occurs. The entire evolution process is a 3.5 billion year adventure. In this unthinkably long period, about 85% of the time was devoted to the slow evolution, only 15% to fast evolution. However, it's the fast evolution that has brought about today's breathtaking biodiversity.

Slow evolution can be subdivided into prokaryotic era and eukaryotic era. Transition from prokaryotes to eukaryotes was completed in about 2 billion years, while preparing eukaryotes for fast evolution cost another 1.5 billion years. Both eras were painstakingly slow and lengthy, while the organisms evolved in this period were simple in every aspect. Fast evolution started from Cambrian explosion, which brought a myriad of new species to populate the earth.

The single celled organisms from Stage 1 were far from mature and robust. They are vulnerable and defenseless against natural elements and their genomes were too small to support evolution. In the prokaryotic era, evolution brought up major changes to the single celled organisms by increasing their genome size and protein coding gene counts. Randomness from sporadic point mutations was still one of the factors that drove prokaryotes to diversify and expand, resulting in numerous species. The different species carried a certain number of genes that made them distinct from others. All species specific genes formed a huge gene pool from the standpoint of prokaryotes. In the dynamic prokaryotic era, life didn't always exist inside the cell walls, but was influenced greatly by some forms of extra-chromosomal genetic material that could co-exist inside the cell walls as an independent entities. Plasmids are extra-chromosomal genetic material that transmitted from one bacterium to another (even of another species) mostly through conjugation, and transferred genetic material from one bacterium to another. Bacteriophages are another extra-chromosomal genetic material in the forms of viruses, which played roles similar to plasmids. Bacteriophages transfer genetic material from one bacterium to another through infecting various prokaryotic species. Extra genetic materials transferred from other

species not only enriched genes coding for more functional proteins, but also increased the genome sizes. A larger genome tended to generate more novel genes through random point mutations. Whenever a random sequence became gene-like, the transcription machine would act on it to make RNA and protein from it. Transfer of random genetic material further broadened the heterogeneity of prokaryotic population.

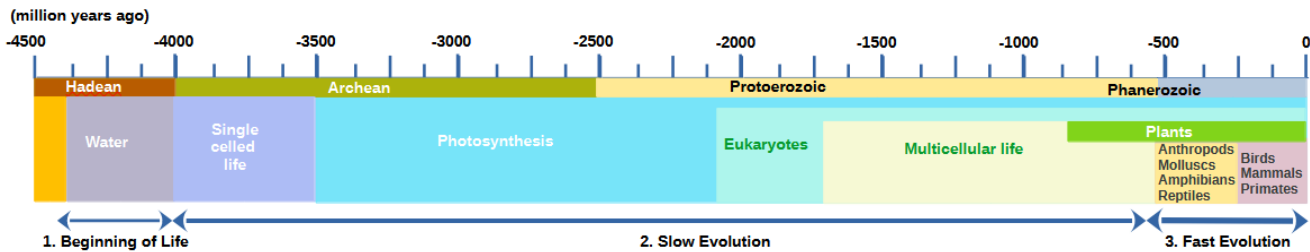


Figure 2. The evolution is a three-stage adventure, each later stage relies on the existence of early stage.

By taking the rapid cell division rates and a period of 2 billion years into consideration, even at very low rates of point mutations, single celled organisms could have accumulated an amount of mutations in intergenic DNA sequences that were sufficient to transform certain DNA sequences into functional genes. Genotype differences led to differentiation of cells into various sizes and robustness. Smaller cells could be engulfed by large cells, and their DNA integrated into the host genomes. This is a random merger process called symbiogenesis. The merged cell populations were heterogeneous immensely, in which individual organisms differed in genome size, gene set, metabolism pathways, and more importantly in physiology. All these qualities could entitle them to exist as different species. Therefore, it was randomness and time again that drove the evolution of single celled organisms to become more functionally adequate and better formed and developed morphologically.

When single celled organisms contained a relatively small number of proteins, it was much easier for them to diverge into numerous different lineages through mutations and symbiogenesis. Later divergence made lineages more different from the ancestors, but more similar among later members of the lineages. There are a billion of distinct species of prokaryotes today, including bacteria and archaea. There is a possible link between archaea and eukaryotic microorganisms as some genes and metabolic pathways found in eukaryotes are more closely related to those of archaea, especially the enzymes involved in transcription and translation.

Some merged prokaryotes became the predecessors of eukaryotes, in which biochemical processes and cellular structures started to compartmentalize into organelles. Among organelles is the nucleus, which gave genomes a designated space to grow and function, and more importantly to elude interference from other cellular activities. The appearance of histone like proteins further transformed naked genomes into tightly packed chromosomes. Formation of chromosomes and confinement of the chromosomes in the nucleus means that life has entered the eukaryotic age, a landmark in the history of life. Eukaryotic cells are full-fledged organisms at this time, boasting a genome size over 5 millions of base pairs and a variety of metabolism pathways. Possession of chloroplasts allows the organisms to capture and store unlimited energy from sunlight through photosynthesis. Having mitochondria as an energy generator, the organisms are supplied with adequate chemical energy to power a variety of biochemical processes and cellular activities. Meantime, safekeeping of the genetic machine in nucleus guarantees the high fidelity of DNA replication, further reducing randomness in DNA replication and slowing down the pace of evolution.

The appearance of eukaryotic cells didn't mean that evolution entered the fast track. In all likelihood, the genomes of nascent eukaryotes were still small and contained few protein coding genes. As a result, the eukaryotes must continue to increase the genome sizes and the number of protein coding genes. Confinement of genome in the nucleus and integration of more enzymes useful for DNA manipulation gave the organisms more freedom to bring about genetic changes to the genomes.

Genes in prokaryotic organisms are continuous without intragenic sequences, suggesting that DNA insertion is lethal and isn't a general mechanism to increase the genome size. However, majority of genes in modern eukaryotic organisms are disrupted by large amounts of intragenic DNA, called introns. This suggests that DNA insertions were random, and common in nascent eukaryotes. Foreign DNA fragments could come from internalized cells or viruses through endocytosis. Meantime duplication of DNA fragments provided another main means to increase the genome size, while random point mutations accumulated constantly on the chromosomes, generating a variety of possible genetic loci with potential biological significance. All random genetic changes diversified the population into different species and prompted organisms to differentiate into different cell types, a prelude to the emergence of multicellular life.

Slime molds are amoeba-like, typically single-celled organisms, and some of them can aggregate into loosely associated colonies. Such colonies are the infant form of multicellular organisms, and the cells in the colony were just about to differentiate into cell types. The genetic base of cell differentiation is the differential expression of genes in different cell types. In other words, gene expression must be regulated stringently according to the roles of individual genes in cell types. Therefore, it's imperative to establish rigorous regulatory mechanisms to control gene expression according to the cell types.

Immediate questions were that how to guarantee that particular genes were expressed only in cell types in which they were intended to express? Were the regulatory elements in the promoter and any other regions sufficient to confine the expression of a particular gene into a particular cell type? The answer seemed to be a no. Leak expression in the wrong cell types seemed to be a common occurrence for all genes and could ruin cell differentiation, thus ruin evolution of life.

Except some house keeping genes, overwhelming majority of genes have introns interspersed along the gene sequences. The genetic machine must remove the intron sequences from the newly transcribed RNA molecules before exporting them out of the nucleus, a process called RNA splicing. RNA from leak expression might not be able to survive the RNA splicing process due to insufficient amount, eliminating the possibility of protein synthesis in the wrong cell types. On the other hand, house keeping genes are not specific to cell types, but common to all cell types. Splicing their RNA is a waste of resources, and many of them are indeed intron-free genes. Adding non-coding DNA sequences inside genes increased the genome size considerably, and as a result, the cells would consume more energy and material to operate and maintain the huge genomes. Therefore introduction of introns isn't a cost effective, but is a viable, way to guarantee the integrity of cell types. Insertion of non-coding sequences into genes wasn't for making sure that gene expression was leak-proof, but it just happened randomly and was preserved because it served this purpose well when cells started to differentiate. Otherwise what strategy could replace introns to serve the same roles, but better?

Evolution of eukaryotes from the moment of appearance to the moment before Cambrian explosion spanned a period of staggering 1.5 billion years, during which organisms had been multicellular, but cell differentiation and tissue morphology were kept minimum. It's a stark contrast to the fast evolution that followed immediately after. What did happen in this period could be very revealing about the evolution and how the evolution itself is evolved over its entire course.

So far hundreds of genomes from species covering almost all levels of evolution have been sequenced and annotated, and data are available from several research institutions for research. Table 1 lists minimum genome information, including genome size and the number of protein coding genes, from several selected species ranging from archaea to bacteria to organisms emerged from Cambrian explosion. What we can draw from the data in the table will help us understand evolution and what is the bottleneck of evolution.

Organism	Type	Genome Size	Coding Genes
<i>Pyrobaculum aerophilum</i>	Archaea	~2,200,000	2,562
<i>Haloferax volcanii</i>	Archaea	~4,000,000	3,895
<i>Chlamydomophila pneumoniae</i>	Bacteria	1,225,935	1,113
<i>Chlorobium chlorochromatii</i>	Bacteria	2,572,079	2,002
<i>Gloeobacter violaceus</i>	Bacteria	4,659,019	4,430
<i>Bacteroides fragilis</i>	Bacteria	5,205,140	4,260
<i>Micromonas pusilla</i>	Algae	~20,900,000	10,056
<i>Guillardia theta</i>	Algae	~87,200,000	24,840
<i>Dunaliella salina</i>	Algae	~343,700,000	16,697
<i>Breviolum minutum</i>	Algae	~1,500,000,000	47,014
<i>Tieghemostelium lacteum</i>	slime molds	~23,400,000	10,225
<i>Hortaea werneckii</i>	Yeast	~49,900,000	15,974
<i>Aspergillus nidulans</i>	Fungi	29,828,291	10,534
<i>Armillaria ostoyae</i>	Fungi	60,106,801	22,299
<i>Blumeria graminis</i>	Fungi	124,489,486	7,118
<i>Puccinia graminis</i>	Fungi	88,724,376	15,800
<i>Crambe crambe</i>	sponges	~370,500,000	
<i>Chrysaora quinquecirrha</i>	jellyfish	~337,400,000	
<i>Bolinopsis microptera</i>	comb jellies	~265,400,000	14,118
<i>Lytechinus pictus</i>	urchin	~811,700,000	20,054
<i>Cydia splendana</i>	Moths	~630,600,000	14,646
<i>Penaeus indicus</i>	Shrimps	~1,900,000,000	21,824

Table 1. Genome sizes of various species on different levels of evolution. The cells that display the number of protein-coding genes are left blank if data are not available. Data are taken from NCBI, Ensembl Bacteria, and Ensembl Fungi.

The number of protein coding genes for each organism listed in the table is obtained using genome analysis software, not necessarily mirror the true protein production in the organisms. However at least it demonstrates that these genetic loci exhibit gene structures and can be considered as genes, especially as genetic material to derive new genes or duplicated genes.

Genomes of prokaryotes are small and contains fewer protein coding genes when comparing with the genomes of eukaryotes. Furthermore, the sizes and gene counts varies greatly from species to species. On average each gene takes up about 1000 bps and encodes a protein of about 250 amino acids. This

shows that prokaryotic genomes contain sparse intergenic DNA sequences. What could be inferred from this is that the genomes of the first forms of life must be far smaller and encode far fewer proteins than present day prokaryotes.

Genomes of single celled eukaryotes vary in size and gene counts greatly from species to species without exception. They are usually 5 to 20 times larger than the genomes of prokaryotes, but protein coding gene counts are only 5 to 10 times larger. On average each gene takes up about 2000 bps. This shows that eukaryotic genomes contain a large amount of intergenic and intragenic DNA sequences. What could be inferred from this is that the genome sizes and gene counts of the first forms of eukaryotes must be close to those of prokaryotes.

Genomes of multicellular eukaryotes vary in size and gene counts greatly from species to species as well. A general trend is that the genome sizes increase dramatically, but gene counts were relatively stable, as the organisms move up the evolutionary ladder. The average base pairs per gene are about 3000 bps for fungi, but dramatically increased to about 30,000 bps in pre-Cambrian organisms sponges, jellyfish and comb jellies, and to 45,000 bps in post-Cambrian organisms urchin and moths. This indicates that genome size increase is largely due to the increase in intergenic and intragenic sequences, not in protein coding sequences. Occurrence of DNA duplication could be frequent, and duplicated DNA diverged in nucleotide sequences upon long time accumulation of point mutations. Protein coding gene counts usually fluctuate around 15,000 to 20,000 regardless of genome sizes after they reached certain values. This is true for organisms like mammals. It is quite a shock that in multicellular organisms protein coding gene counts are not much correlated to the complexity of the organisms.

Prokaryotic era and eukaryotic era share some common characteristics. Increase in genome size accompanied the increase in gene counts. Starting from a smaller number of count ended up in a much larger number of count. The most significant and indicative common characteristic is that both era endured a period of more than 1.5 billion years to conclude the evolution triumph. From phenotype point of view and comparing with what happened in post Cambrian era, the advances in cell differentiation and overall organism development seemed too meager to worth 1.5 billion years dedicated to each of the eras. This is a mystery until we can divulge into it with the availability of huge amounts of genome sequence data. The whopping 1.5 billion years each for these two eras reveal the unthinkable difficulty for the biological system to create and assimilate any new proteins into the existing biochemical processes or cellular structures and function as an integral part.

Assume that a new small chemical named X arose from nowhere to regulate body temperature in the extremely cold environment. To generate a receptor for X, the X receptor, an existing receptor gene for a different small molecule Y was duplicated. To turn Y-specific receptor to X-specific receptor was a long evolutionary journey. It would be hard to imagine how many changes must be made to the Y receptor so that it could be turned into an X receptor. First the changes must be able to create an internal space to accommodate X and at the same time maintain an overall three-dimensional structure to support such an internal space. Second the X receptor, upon binding with X, must be able to undergo conformational changes into an active state. Third the X receptor must gain another structure after activation which could either interact with a downstream component involved in regulation of body temperature or act as an enzyme by itself. Fourth the X receptor gene must be subjected to regulatory control so that this receptor would be expressed only in tissues it was intended to express. High fidelity of the DNA polymerases in eukaryotes makes point mutations at very low frequency. In a likely scenario no single point mutation could ever occur in this duplicated gene in many generations. If point mutations indeed occurred and accumulated over many many generations, it's still highly unlikely that those mutations could turn Y receptor into a functional X receptor. Over a long period of time, early

effective mutations could have been nullified by later mutations. It's essentially an extremely lengthy trial and error process.

In the above hypothesized scenario, the gene for X receptor was duplicated from an existing gene. In theory it was a fully functional gene, and no excessive changes would be needed to make it fit for X structurally and functionally. If a signal transduction pathway was established from ground zero and involved four totally new proteins, then there were so much to be created, changed, and interlinked from genes to proteins to finally make them function as a novel signal transduction pathway. A more likely scenario is that all biological functions and structures aren't made by purpose, but it happens naturally when some particular components come together. Prokaryotes or eukaryotes were constantly generating random genes that encoded proteins of a variety of functions during slow evolution stage. When right proteins coexisted in the cell and happened to come across, they formed novel biochemical processes like signal transduction pathway or cellular structures. Randomness gives rise to all fortuitous events at the cost of time. From such a standpoint, the whopping 3.5 billion years seems not whopping anymore for these two eras.

At the end of slow evolution stage, species are still low and not sophisticated at all from any stand point of view, but the average gene counts are large enough to be on a par with all other higher animals, including mammals. An implication is that evolution has entered a new mode to move forward.

## **5. Protein Variants and Evolution**

Large scale genome sequencing indicates that species in the same genus, even in the same family, share extremely high percentage of identical sequences. The implication of this is that large morphological differences doesn't mean similar differences in genotypes. In fact as species move up the evolutionary ladder, the differences between genotypes diminish. For species that are classified into the same family, differences in morphology, biochemistry and cell structures generally are attributed not to all new proteins, but to the use of variants or isoforms of the same proteins that are expressed in different species. The advent of protein variants in evolution is a giant step forward towards more complex and advanced species on a number of superior benefits.

*C. elegans* is a free-living transparent nematode or worm belonging to a type of metazoan organism with 959 cells. *C. elegans* genome is relatively small, consisting of 100,286,401 bps, and contains an estimated 19,985 protein-coding genes. 83% of proteins expressed in the worm were found to have human homologous genes. Only 11% or less genes are nematode specific. Some proteins can be exchanged between *C. elegans* and humans or mammals. This means that the majority of genes that make up all genes needed by much more complex organisms like mammals are already available in animals as low as *C. elegans*. An implication of it is that development of higher organisms doesn't depend on creation of animal specific genes, but primarily on utilization and localization of genes that have existed in lower organisms, a strategy of derivation and reuse.

All protein molecules must assume unique three dimensional structures to assume their biochemical functions, while the three dimensional structures are mostly determined by the amino acid sequences. Proteins will exhibit altered biochemical or structural properties after certain amino acids are substituted. They are the variants of the same protein molecule. The protein variants impacts more on biological processes that rely on three dimensional structures. For example, a neurotransmitter receptor variant could have its affinity for the ligand increased or decreased relative to the original receptor, thus changing the behavior of the organism. Signal transduction pathway could be affected because of changed physical interaction between each protein component. Structural variants will have more visible impact. An organism could assume a different morphology if tissue orientation in normal

development skewed to a certain degree due to a structural variation in the process. If variants are engaged in embryonic development, limb could grow out of wrong place or into wrong forms.

If species can be classified into the same family, their looks are similar but not identical because the proteins working behind the morphology are not the same and not all-new, but variants. During the evolution process, the same protein could diverge into multiple variants in offspring due to the randomness of mutations, resulting in offspring to assume looks that were similar but not identical in size, morphology and behaviors, which qualified them as new species. Protein variants randomly produced in the evolution must be one fact that has shortened the time for new species to appear and made biodiversity so extraordinary today.

Protein variants in the evolution of species reveal an important biochemical properties of proteins. The sequence based three dimensional structure is not rigid, but shows great elasticity in the cells. In other words, the three dimensional structure of a variant is elastic enough to withstand certain sequence changes and remain compatible with the existing biochemical and cellular processes, allowing them to proceed as if nothing has happened.

From evolution point of view, gene duplications are the easiest and quickest way to produce new proteins for new species. The ancestor gene from which all subsequent genes coding for protein variants are derived is called master gene, and derived genes are child genes as a result of gene duplication. The child genes must be subjected to changes in DNA sequences so that they encode variants of the master protein. The differences in amino acid sequences will widen as species diverge further apart.

Opsins are protein molecules responsible for wavelength sensitivity when coupled with light sensitive chromophore 11-cis retinal. S-opsin absorbs short wavelength light, M-opsin absorbs middle wavelength light, and L-opsin absorbs long wavelength light. Primate retina consists of three types of photoreceptors, each of which contains S-opsin, M-opsin, and L-opsin, respectively, while most mammals lack L-opsin containing photoreceptor, making them insensitive to long wavelength light. When primates split from most mammals, M-opsin gene underwent a duplication event that led to an extra copy of the gene. Then random point mutations turned this extra gene into a functional gene coding for L-opsin. All this is evidenced by the fact that M-opsin and L-opsin are identical except 15 amino acids out of 364 total. This small difference in amino acid sequence changes L-opsin sensitivity to long wavelength light from middle wavelength light. L-opsin is a variant of M-opsin.

Generally speaking, a variant assumes the same biochemical roles of its master protein, but with subtle changes in biochemical properties. It's these subtle changes that empowers a variant to fit new occasion or fill the void that the master proteins can't fill, or complement the action of master proteins in the new species. L-opsin is the brilliant example in this regard. Muscarinic acetylcholine receptor has about five subtypes, all of which are variants of each other. These receptor variants show unique gene expression profiles, unique sensitivity to ligand acetylcholine and various drugs, and more importantly they elicit unique neurological effects in different target cells.

Most important impact of protein variants on evolution is not short-term, but long-term. Multicellular forms emerged after cells started to differentiate into types when the genome size and protein coding gene count increased to a certain level. Cell differentiation became the major changes in the later stage of slow evolution, as multicellular life began to take a more defined and sophisticated morphology like soft-bodied metazoans, some of which displayed a trace of skeletal elements. The appearance of new and more complex life forms in stage 2 must be supported by new protein factors, for example new

protein factors that control the development of a morphology and proteins that carry on the underlying biochemical processes. The genes that encode these early proteins will serve as master genes, from which protein variants would be derived and impact the future development of organisms in a fundamental way, making new species more complex not only in morphology but also in tissue types.

Fin development begins in the morphogenetic fin field in the fish embryo. Some fin inducing factors act on mesenchymal cells in that field and cause the outer germ layer to proliferate and bulge out, forming a fin bud. A growth factor then guides further development of the fin bud into a fin. The fin inducing factors control the exact direction and position the fin bud bulges out in the morphogenetic fin field, which determines the final morphology and location of the mature fin on the body. Assumed that one genetic locus was duplicated from the gene encoding one of the fin inducing factors. Over time under random mutations, the sequence of this locus deviated from its master gene and encoded a protein variant. This variant assumed a three dimensional structure that was slightly different from that of the master protein. Because of this slight difference it induced the fin bud to bulge out at slight different position and towards a slight different direction. The overall impact on the fin development was that the final fin was quite different from the fin on the ancestor organisms morphologically and in location. If a factor variant assumed a three dimensional structure that was too skewed to induce the normal fin development, the final fin could be in a deformed morphology. If a factor variant was produced in the wrong part of the embryo, it could induce the growth of a fin at a wrong place.

During an evolution period, genome-wide changes led to emergence of new species. Among all changes, a particular fin inducing factor variant gave rise to a new and unique fin morphology. If the ancestor fish ended up in five new species, their fin morphologies could range from identical to very different, depending on how far apart they were in the period or cycle as will be described in the next section and the amount of mutations accumulated on each duplicated locus. Closer these variants in biochemical properties and cellular structures, more similar the final fins on the body of the new species. This isn't the proper example to illustrate the Cambrian explosion, but the principle behind each explosion is the same.

Weathered by tens, even hundreds of million years, a particular duplicate sequence from an ancient master gene could have undergone extensive mutations and diverged into a large family and each locus in the family encodes one variant or isoform. For a particular variant it may be found in a particular species or a wide range of species of different complexity. Sequence homology among variants might not be apparent anymore, but each variant might still retain core functions of its master gene. In other words, some family members have become distant variants of each other after many evolution cycles.

It is apparent that a remarkable consequence of the divergence of a master gene into a variant family is the rise of myriad forms of phenotypes in which these variants are an influencing factor. Taking fin inducing factor as hypothesized example, the early variants descending from the master gene gave rise to various forms of fin morphology on different fishes. Further divergence in sequences gave rise to many more distant variants, which control the development of limbs on amphibians, limbs on reptiles, limbs on birds, limbs on mammals, limbs on primates, and finally limbs on humans. The degree of divergence from the master gene mirrors well the positions of organisms on the evolutionary ladder. This is a vertical view of roles protein variants have played in the course of evolution. If we take a horizontal view, what we see is so many distinct fins on fishes, so many distinct limbs on amphibians, so many distinct limbs on reptiles, so many distinct limbs on birds, so many distinct limbs on mammals, many distinct limbs on primates, finally one unique limb on humans. Through divergence into a large variant family, the overall impact of the fin inducing factor on the fin and later limb development has been amplified to the utmost degree throughout the past 500 million years.



Keratin genes provides us with another illustrious example of how protein variants evolve in parallel with the evolution of organisms. Keratin consists of a family of structural fibrous proteins called intermediate filaments. The master gene of keratin is present as early as in sea squirts before Cambrian explosion. Today numerous variants of keratin exist in a vast variety of species, including vertebrates and invertebrates. They form the hair, outer layer of skin, horns, nails, claws, scales, shells, feathers, beaks and hooves for sea squirts, fishes, reptiles, birds, and mammals. It's the sweeping divergence of keratin master gene in the past 600 million years that has made variety forms of such tough structures possible. And it's these variety forms of tough structures that confer animals bearing one of these structures distinct capacity to inhabit a suitable environment. A particular keratin variant can function in many species, and a particular species can have many keratin variants to fulfill different functions. For example, The human genome encodes 54 functional keratin genes, located in two clusters on chromosomes 12 and 17. It's clear that the distinct amino acid sequence of each keratin variant has been preserved over the evolution for advantage in its unique three dimensional structure that is just suitable to build beaks, or feathers, or hair, or nails, etc.

The first generation of protein variant deviates from its master protein through random changes in its amino acid sequences. The second generation of protein variant deviates from the first generation of the protein variant through random changes in its amino acid sequences. After round and round of deviation and evolution cycle and evolution cycle, it forms a large family of variants. However this is a family of variants from evolution point of view, they are unlikely a family of variants from structure and biochemical function point of view. In other words, a master gene could have diverged into a giant group of protein variants over evolution, and many of them are completely different from others in amino acid sequences, biochemical functions and three dimensional structures. They have lost qualifications to be variants of other members per se, but are all-new protein molecules on their own. Evolution placed no constraints on duplicate genes to diverge as long as a variant is not lethal to the organism. Generation of an all-new protein from a duplicate gene is obviously a far better choice than from a piece of random DNA sequence, because the duplicate is blessed with a full gene structure. A gene can be transformed into a new functional gene through random point mutations, exon and regulatory sequences shuffling, alternative splicing, and etc. on various parts of the gene.

The large gene counts of multicellular species in the pre-Cambrian period seemed to be an indicator of accumulation of duplicated genes over 1 billion years, and their roles were manifested in the Cambrian explosion and all later evolutionary events.

## **6. New Species Arise in Explosive Mode – Evolution Cycle**

It seems bewildering why all over a sudden evolution accelerated, bringing millions of complex and advanced new species into existence in a short period of time. In Stage 3, a burst of new species in general accompanied certain dramatic climate and geological changes. For example the Cambrian world differed greatly from the preceding Proterozoic Eon in terms of climate and geography. During the transition of the two periods, the earth experienced a gradual global warming, rising oxygen level, and split of a single continent into two. Climate and geological changes could make mutations occur more frequently in all species. When mutations struck DNA polymerases, DNA polymerases replicated DNA at lower fidelity, causing organisms to suffer from accelerated genetic changes. Direct consequences of this are two folds, mass extinction of old species and proliferation of new species.

Prior to Cambrian explosion, the majority of living organisms were small, unicellular, and simple, classified under kingdom protists, including slime molds and fungi, while a little more complex, multicellular organisms like sponges, jellyfish, sea anemones, corals, etc just gradually emerged in the

later stage of slow evolution. Living organisms exploded into millions of forms and complexities during Cambrian period. Many organisms we see today appeared in this period, for example, insects, flies, spiders, centipedes, shrimps, ticks, mites, scorpions, snails, shells, starfish, brittle stars, sea urchins, sand dollars and sea cucumbers. First plants and fishes appeared at the later stage of Cambrian period. From evolution point of view all these species remain very low on the evolutionary ladder despite of stunning varieties in complexities and forms.

In the following 540 million years, evolution greatly sped up the emergence of new species of higher complexities. In a nutshell, evolution of species is the evolution of genomes. The genomes became more advanced after each evolutionary event, laying down the foundation for more complex and advanced species to appear.

As described earlier, when organisms evolve to higher levels, they rely more on protein variants to build unique morphology, cellular structures and biochemical processes. Protein sequence comparisons tell a lot about those proteins that play similar biological roles in different species. A large number of those proteins can be classified into groups based on the similarities of overall amino acid sequences or sharing of certain short amino acid sequences called motifs. Proteins that shared similarities of overall amino acid sequences are essentially variants or isoforms of each other. By taking advantage of protein variants and motifs, the creation of proteins with desired functions and properties became easier, and at the same time it alleviated the challenge to assimilate new biological components into existing cellular structures and biochemical processes. Otherwise it would incur a formidable array of problems for creation and integration of all new proteins into the existing system even when organisms were low and relatively simple as it had been vividly shown and told in the slow evolution stage.

Evolution is a process of constant changing, and constant changes bring about all sorts of consequences to the organisms. It's highly unlikely that all modern day organisms, archaea, bacteria, animals and plants, are in the path of evolution. The planet earth is full of living organisms as simple as single celled life and as complex as mammals. If all organisms were in a constant state of evolution over the past billions or millions of years, we wouldn't be able to see organisms as low as archaea, bacteria, algae, fungi, even jelly fish, sea urchins, etc. This indicates that not all low organisms can be ancestors of higher organisms. Most species stay where they have been since they appeared long time ago.

The concept "ancestor" must be right because it agrees with the evolution of living things. Then what organisms can be ancestors from which higher forms of life arise? Formation of new species isn't a simple event, but involves numerous changes in the genotype that generates a phenotype that is sufficiently different from the phenotype of the old species. This level of genetic changes won't be possible in a normal organism, implying that ancestor organism was special on its own. The genome of ancestor organisms was likely to be quite elastic and blessed with a genetic machinery that could perform large-scale genetic changes – changes that will generate enough new genes or gene variants for a new phenotype that defines an organism as a new species.

What exact events could trigger large scale genome changes is a forever mystery just as how life exactly started, unless humans could be fortunate enough to experience a new round of mass extinction and mass formation and we were not part of the mass extinction. However, it's worth to think about it and envision something fictitious to get some idea.

In usual time, genomes of all organisms, including ancestor organisms, were in a disarmed state, in which the genome is consistent and stable except low random point mutations and normal DNA recombinational events during meiosis. When sudden geological and climate changes broke out,

ancestor organisms in a population suffered from more mutations. When mutations lessened the replication fidelity of DNA polymerases, genome wide accumulation of point mutations accelerated. These genetic changes acted as a perturbation that drove the genome from its disarmed state to an inconsistent and unstable state, an armed state. In the armed state the genetic machinery of an individual ancestor was activated to perform genetic changes that would reshape the genome to start the evolutionary event, a phase called genotype reshape. Genotype reshape initially reshaped the genome, but as it brought further changes to the genome, the genome gradually re-established a consistent and stable state, a new disarmed state. This was a process called genotype healing. After genomes of all individuals descending from the common ancestors went from disarmed state to armed state and then back to a new disarmed state, it completed one evolution cycle (Figure 3). All individuals that appeared in the cycle, including those dead at the embryonic stage, were intermediates of the evolution. Processes reshape and healing are quite blurry and overlapping, but are two different concepts useful to reveal what is happening in an evolution cycle.

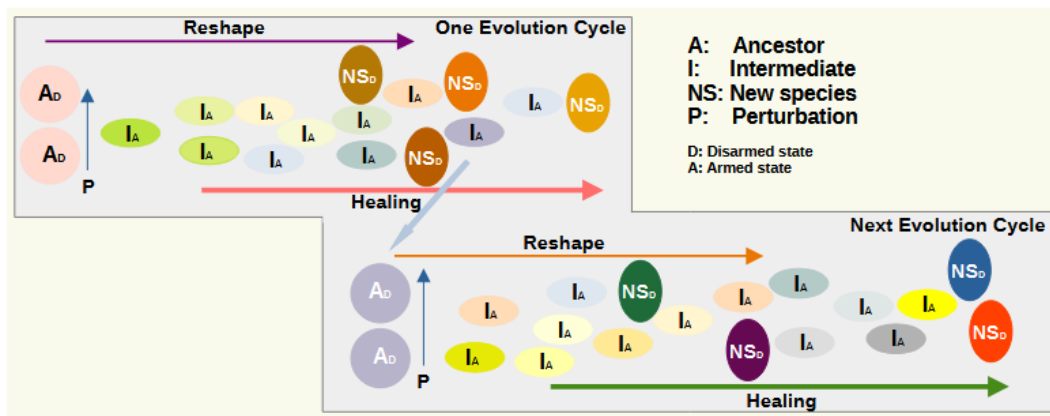


Figure 3. Evolution cycle from ancestor organisms in disarmed states to new species in new disarmed states, including two processes reshape and healing and numerous intermediates in armed states. Reshape and healing are largely overlapping, as some mutations bring changes to make genes defect-like and others to turn them back into functional variants or new genes.

Genome evolution cycle is unique because it could take up to 10 million and even 100 million years or generations to complete just one single cycle, and its cycle path is composed of a single armed state and two processes, reshape and healing. From genetic mutation point of view, all changes that occurred in the path are random and irregular. Randomness and irregularity are the key to the fascinating biodiversity that arises after each evolution cycle. Genetic changes that occurred in a cycle are deemed as large-scale, but the changes to one generation of any intermediate must be on a granular level so that some intermediates in the population would survive every change. Not all evolution cycles would lead to new species if no intermediates survived.

Assume there was a population of a single ancestor species before sudden geological and climate changes. When physical upheaval came, all genomes were hit with random mutations and entered an armed state. As a result one armed state was different from all other armed states among the population in terms of DNA sequences thanks to the randomness of mutations. The judge of a change to be good, bad, or not-good-not-bad is if the intermediates could survive the change. In an evolution cycle, all lethal changes were eliminated from the population after causing carriers to die, leaving no impact in the evolution cycle. Only changes that could be passed down to the next viable generation would have impact and cause phenotypes of the next generations to vary.

After entering the cycle the homogeneous population of the ancestor quickly became heterogeneous in terms of genotypes. As the cycle continued, more random and irregular genetic changes made the genotypes of the population more heterogeneous. In other words, the genotypes among individual intermediates became different after a few generations and the differences widened as more generations passed. Eventually at the end of healing process and arrival at a new disarmed state, survived individuals carried genotypes that were so different from one another, and they were no longer the same species, but different new species. The genomes of all intermediates were in armed state, and their fate in the cycle was unpredictable in the random world. It's all up to their luck.

Mutations generated in the early stage of reshape would perturb the normal biochemical machine and cellular structures to a greater extent, giving the cycle a bumpy start. The perturbation could lead to the death of most of early intermediates, but it would be necessary for an evolution cycle to produce new species. As process healing was slowly overtaking process reshape, any new components, including protein variants or new proteins if any, gradually reached an equilibrium with the existing system after undergoing numerous changes throughout the cycle. In addition the morphology of each intermediate underwent corresponding changes as the genotype changed. The happy ending of an evolution cycle was the emergence of new species in new disarmed states. In this state all biochemical and cellular functions were balanced again after they had been agitated by genome wide changes and returned back to work in harmony, a result of enduring long time instability and biochemical conflicts. The difficulty to achieve such a balanced or disarmed state would be unthinkable if numerous protein variants were not part of the cycle, especially in species as advanced as fishes, amphibians, etc. It is expected that overwhelming majority of intermediates couldn't reach the final disarmed state, but disappeared quietly because of failure to survive seemingly endless mutations. They are the dead ends in the cycle.

The random nature of genetic changes could result in a number of first generation intermediates, depending on the size of ancestor population. From the moment of birth, an intermediate would move along its own path in the cycle and produce its own next generation intermediates in a manner independent of other intermediates. It was unlikely to predict how many more generations were needed for a random intermediate to reach the final disarmed state if it was a lucky one. New species would resemble each other more strongly if they descended from the same intermediate fewer generations apart and differ with each other more strongly if they were more generations apart. Figure 4 illustrates an evolution cycle in its entirety starting from an ancestor population. Individual ancestors (orange solid circles) sit in the center and were surrounded by light pink solid circles that represent numerous intermediates, whose distance to the center represents the number of generations down from the ancestor. Intermediates that are dead ends in the cycle are represented by outermost solid black circles. Some lucky intermediates that end up in disarmed states – the new species arising from this evolution cycle – are indicated by outermost solid red circles. When a line with arrow connects the center and one of the outermost circles through a series of intermediates in between, we could see the complete evolution trails that start from an individual ancestor and passes through every intermediate that leads to the next intermediate before reaching the outermost circle. The picture clearly shows that new species arise in an explosive mode in an evolution cycle totally due to the random nature of genetic changes. Therefore the size of new species descending from one common ancestor is determined by the number of intermediates that survive to the disarmed states.

We can use human evolution to help illustrate an evolution cycle at work, rough but a little more intuitive. It is more appropriate to say that all mammals arose not from a single ancestor, but from distinct ancestors that shared a lot of similarities. About 60 million years ago there was one ancestor X0. X0 could be an ancestor organism or an intermediate from another ancestor organism. Regardless

of its origin, it diverged into a number of intermediates after X1 generations. One intermediate led to a variety of monkeys (disarmed states) after X2 generations with many intermediates becoming dead ends in the cycle. Another intermediate diverged into more intermediates of its own after X3 generations, among which one intermediate developed into early ape species (a disarmed state) after X4 generations, and another intermediate moved on and diverged into more intermediates of its own. One intermediate became gorilla (a disarmed state) after X5 generations, and another intermediate moved on and produced more of its own intermediates. One intermediate among them developed into different forms of chimpanzees (disarmed states) after X6 generations, and another intermediate moved on and led to more intermediates. One of these intermediates finally reached the earliest two-footed animal bipeda after X7 generations, establishing genus homo. This earliest two-footed animal bipeda wasn't a dead end, but a lucky intermediate on evolution that generated an unknown number of its own intermediates, one of which led to humans, the only species (a disarmed state) emerged from this lucky intermediate after X8 generations. Which species first reached its disarmed state in the cycle can't be speculated but must be established by research, especially through DNA sequence comparison. Some intermediates left fossils behind that allowed researchers to trace human evolution in the past 3 million years. However scarcity of fossils have limited progress in this field.

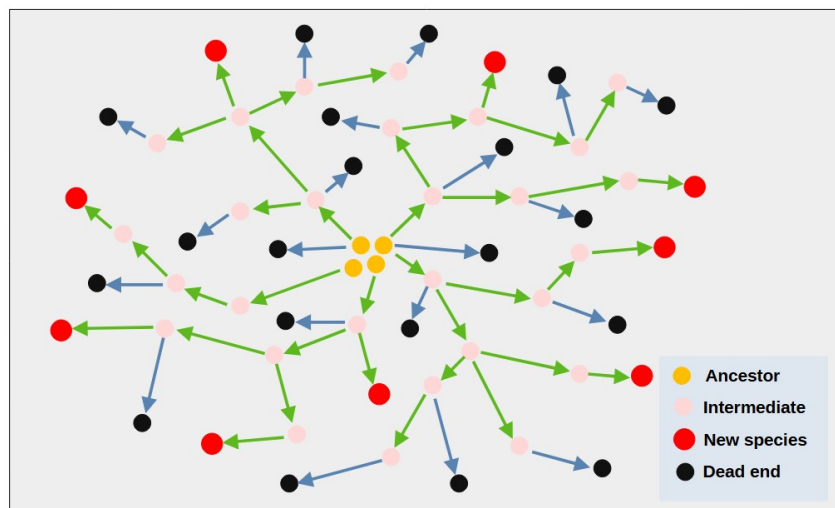


Figure 4. A simplified diagram to illustrate how new species arise in explosive mode in one evolution cycle. Distribution of new species is random relative to the ancestor organism. All new species can be classified into a single class.

In biology, there is a classification system that classifies living organisms into eight levels based on shared characteristics. The last four levels are order, family, genus, and species. Animal gorilla can be classified to the last four levels as primate order, hominidae family, gorilla genus, and gorilla species. In similar way, chimpanzee as primate order, hominidae family, pan genus, and chimpanzee species, and human as primate order, hominidae family, homo habilis genus, and homo sapiens species. It would be obvious now that gorilla, chimpanzee and human share the same ancestor and a series of common intermediates until reaching a particular intermediate, from which gorilla left the trail to human and established its own genus gorilla. Chimpanzee and human continued to share some common intermediates before chimpanzees diverged from the trail to human and established its own genus pan. However, the order of appearance can't be determined purely based on which species is more advanced physiologically and morphologically. In other words, the appearance of chimpanzees was not necessarily earlier than humans. Nevertheless, if this evolution cycle ended at the time of modern human emergence, then it is ended about 30,000 years ago.

Since the beginning of Cambrian period, life evolves through evolution cycles, and what happened in Cambrian explosion has also happened in human evolution albeit more complexity and higher richness in genetic materials.

New species are most likely to stay in a disarmed state indefinitely as long as their natural habitats are not too harsh to endanger them. Evolution has been a continuous process. While numerous new species emerged from intermediates in evolution cycles, certain intermediates would transit into new generation of ancestors – daughter ancestor – to keep evolution going. When their DNA polymerases and related enzymes lost their high replication fidelity upon sudden geological and climate changes, they would become a driving force to start new evolution cycles, making evolution of life an inevitable and sure thing to happen. It's an interesting and intriguing mystery if any potential ancestors are still crawling somewhere on the earth, waiting for a geologic event to rouse their evolutionary spirit.

Randomness has been changing its meanings since life-like activity appeared in the incubator in the nascent earth, and is tightly connected to the evolution of living organisms. On the macro level, randomness in Stage 1 and Stage 2 leads to heterogeneity of all organisms in the entire living population as each individual organism carries its own unique genome. If each of these individual organisms diverges into a few new species, new species arise in exponential mode. On the micro level, emergence of a new species is the result of establishing a new balanced biological system, in which a series of changes brought up by the expression of new functional genes would be integrated into the existing system. As species become more complex and advanced, macro level randomness no longer refers to the genome uniqueness of individual organisms, but refers to the genome uniqueness of a population of a particular species, while micro level randomness remains the same.

Randomness is the driving force of evolution throughout the entire evolutionary history. In the dark period of time, only vast randomness embedded in a mess can give rise to something good, which is what exactly had happened in the life incubator 3.5 billion years ago. Time is infinite enough to allow evolution to progress at a snail's pace.

Fossil records give approximate time species appear on the earth, but no clues could be drawn to show the time an ancestor organism began to evolve. In other words, the end of an evolution cycle is relatively clear, but its beginning isn't. Genome sequence comparison would be the only resort to determine the closeness of the species on the evolutionary ladder by establishing the degree of the homology of their genome sequences.

From evolution point of view, it doesn't make sense to consider genes that had small, even no tolerance to genetic changes in the cycle. These genes are too fundamental to life and have little margin for further changes. Mutations would occur to them as usual, but the consequences are either lethal or minuscule in derivation of new functionality. These genes should be excluded in the evolutionary process. What makes sense is to look at genes that have great tolerance to genetic changes. It's these genes that drive evolution forwards as they are able to accept random mutations and deviate into a variety of isoforms or even totally independent forms. Opsin and Keratin genes are the typical examples. It's these genes as well that make it possible to change the complex phenotype in a relatively short time.

Genome sizes and protein coding gene counts shown in Table 1 provided a hint at what happened in the slow evolution stage. Table 2 shows similar information for post-Cambrian species to reveal what evolution is really about in the fast evolution stage.

Organism	Genome Size	Coding Genes	Non-Coding Genes	Pseudogenes
<i>S. cerevisiae</i>	12,157,105	6,600	424	12
<i>C.elegans</i>	100,286,401	19,985	24,813	2,128
<i>Ciona intestinalis</i>	115,227,500	16,671	455	27
<i>D. melanogaster</i>	143,726,002	13,986	4,054	340
<i>Ciona savignyi</i>	177,003,750	11,616	340	216
Sea lamprey	885,550,958	10,415	2,652	47
Zebrafish	1,373,471,384	25,545	6,599	315
Tropical clawed frog	1,451,301,209	22,107	2,023	484
Common carp	1,460,365,368	44,807	9,749	417
Common wall lizard	1,511,002,858	22,062	1,613	267
Softshell turtle	2,202,483,752	18,189	1042	97
Painted turtle	2,365,766,571	21,186	5,523	108
Coelacanth	2,860,591,921	19,569	2918	141
White-throated sparrow	1,052,600,561	13,862	4,340	50
Chicken	1,053,332,251	17,007	13,040	61
Small tree finch	1,064,016,346	16,191	883	93
Dog	2,396,858,295	20,567	9,944	610
Mouse	2,728,222,451	21,955	18,008	13,755
Black snub-nosed monkey	2,977,074,946	20,966	8,376	563
Gorilla	3,063,362,794	21,794	7,768	522
Human	3,099,750,718	19,830	26,462	15,222

Table 2. Genome sizes of various species on different levels of evolution. Data are taken from Ensembl.

Data are largely identical to what is shown in Table 1, but genome sizes increase significantly as species move up the evolutionary ladder. Invertebrates *Ciona intestinalis* and *Ciona savignyi* are low species from Cambrian explosion. Their genomes are relatively small, only 100 to 200 millions base pairs, but contain around 10,00 to 20,000 protein coding genes, about 50% to 100% of the protein coding gene counts of mammals. Tropical clawed frog genome contains about 22,000 protein-coding genes, which are comparable with the numbers from humans while its genome is only about half the size of humans. Exact number are virtually impossible to obtain just by analyzing the entire genome sequences, but it gives us a rough idea that gene counts and genome sizes are not proportional. Protein-coding gene counts from fishes to mammals are largely similar, ranging approximately from 15,000 to 25,000. An important implication of these numbers is that each evolution cycle, for example from fishes to amphibians or reptiles to mammals, seems to need few all-new proteins to become new species. An implication of this is that protein variants seem to play more significant roles than we thought.

If the total number of genes, on average, was assumed to be 30,000, including non-coding genes and pseudogenes, at the onset of an evolution cycle. The genetic changes, including gene duplication, must happen to these 30,000 genetic loci to be effective for evolution. If one evolution cycle took 10 million years, all genetic changes that finally brought about new species must complete in this 10 million years. DNA recombination is largely independent of point mutations, and the occurrence of one would not interfere with the other. A likely scenario is that DNA recombination during meiosis might occur more frequently in the early phase of the cycle, adding more gene-like loci into the gene pool. Overall DNA recombinational events would be less frequent than point mutations. Throughout the entire cycle, gene pool could reach slowly a final count of up to 40,000. The pro-evolution DNA polymerases would introduce more random point mutations at higher frequencies into the entire genome with certain percentage falling onto the genes in the pool. It's these random point mutations that slowly but steadily

transformed the grand old landscape of biochemical processes and cellular structures into a new species. But it's more appropriate to look at evolution from mutations that occur to the entire genome.

In a hypothesized scenario for the purpose of illustration, suppose there was a single individual ancestor organism with a genome size of 1 billion bps and one evolution cycle resulted in one new species. In other words, there was only one evolution trail in the cycle. On average many animals produce offspring one year after birth, meaning one generation per year. If pro-evolution DNA polymerases incorporates 5 random point mutations in one meiosis per 1 billion bps, equal to  $5 \times 10^{-9}$  per base pair per generation. If the length of one cycle is 10 million years, the final species could have accumulated about 50 million mutations in one cycle. This means that 5% of the bases have undergone mutations after 10 million generations. Assume each gene contained 600 bases to encode 200 amino acids on average and 200 bases for regulatory sequences to control gene expression, the gene had a chance of a single point mutation  $4 \times 10^{-6}$  per generation. Among 40,000 genes, 1,000 were immutable, and the chance for the remaining 39,000 genes to be hit with one single point mutation was only 15.6%, meaning even not a single gene would be subjected to one point mutation per generation. Over 10 million generations, the chance for each base pair to receive 1 point mutation is only 5%, which translated into one gene for 40 point mutations, and 39,000 genes for 1,560,000 point mutations. If mutation rate increased to 10 and the length of an evolution cycle increased to 20 million years, then average mutations for one gene would be 160 in one cycle. Be noted that the weight of 40 mutations on a protein variant is much heavier than on an all-new protein, indirectly indicating the importance of protein variants in the fast evolution stage.

The above estimate is very misleading. 40 point mutations per gene per 10 million years seemed too scanty to bring about a new species. However, the average number of offspring per generation per one pair of parent organisms would make huge difference. For easy estimate, if the average number of offspring was assumed to be 10, which was equivalent to 5 parents, and life span was 2 years, then the total number of offspring at 10 millionth years would be an astronomical number, about  $5^{5,000,000}$ . With such a large number, a single gene had undergone approximately  $5^{5,000,000}$  point mutations over 10 million years, making any possibilities possible even taking large mortality rates among intermediates into consideration. These mind-boggling numbers result in the enormous heterogeneity of the intermediate population in a cycle and are the basis of emergence of new species in explosive mode.

When we look at evolution from the whole genome, we could see a thread of events that go through the evolution cycle. Random mutations change the discrete bases in genes, and the discrete base changes result in discrete changes in amino acid sequences encoded by these mutated genes, which affect the biochemical properties of the proteins in a discrete manner. It can't be predicted that how the discrete changes in biochemical properties will change the proteins' functions in their normal environments. Regardless of the discrete changes from new proteins or existing proteins, the final effects are discrete, reflecting on the survivability of the organisms. If the mutations render protein products incompatible with the existing system, causing disruption of essential biochemical processes or collapse of cellular structures or death of organisms, they are lethal and would be eliminated from the populations immediately. Only neutral and beneficial mutations would be preserved and accumulate throughout the entire cycle to finally decide the forms of new species emerged at the end of a cycle.

It would be difficult at present to predict which protein factors played critical roles in determining the final fate of an intermediate. But it could be predicted with confidence that some protein factors play deterministic roles in the establishment of the morphology of an organism, for example, setting the general predisposition of organs in the body, deciding body shapes, brain size, development of feather not limb, etc. all of which are visible phenotype. However, coalescing and integrating all new and



changed properties into existing system is painfully long and purely of trial and error nature over tens of million years. More often intermediates would perish due to genetic changes that were incompatible with the existing activities before reaching the end of a cycle.

From evolution point of view, the genomes of ancestor organisms is more likely to be small and compact. A small genome allows recombinational events and point mutations to concentrate more on significant loci rather than on vast non-coding regions, which is more cost-effective and prolific. A small genome offers more margin for size increase due to DNA recombinational events like gene duplication. Gene duplication seemed to take place in every evolution cycle, and protein variants were more likely to be derived from active genes that were just duplicated in the cycle, as it was the most economical approach to obtain desired properties with minimum changes. Any “leftover” duplicated genes from earlier evolution cycles might have lost structures of an active gene, and couldn’t serve as suitable genetic sources for new properties. A direct measurable consequence is the genome size increase without an increase in protein coding gene counts after each evolution cycle.

It can be concluded that division of evolutionary timeline into slow and fast stages is scientific as two fundamentally distinct mechanisms are working behind each stage. A stage that took over 3.5 billion year to complete reflected the daunting difficulties in creation of thousands of all-new genes from random sequences and assimilation of those all-new properties into the existing system to generate a vibrant and robust system with great potential for further development. The net result of the slow stage is to build up protein coding gene counts that approach to the level of higher species, including mammals, while keep a morphology as simple as multicellular life could be. Apparent incompatibility between abundance of protein coding genes and simplicity of a morphology indirectly indicates the crudeness and coarseness of the large number of genes created and accumulated in this stage if they really could code for proteins. Regardless of the utility of those large number of genes for the pre-Cambrian organisms, they are the abundant ready-to-use materials to derive new properties to build more sophisticated forms and more variety features. Clearly slow evolution lays the solid foundations for rapid proliferation of new species, and is the preamble to the fast evolution stage.

The relative stability of protein coding gene counts across the entire post-Cambrian living kingdom argues well with the conclusion that no more than 20,000 protein coding genes are required to build an organism as sophisticated as humans, and the majority of protein coding genes in higher animals are descended from counterparts in lower species. The main task of the fast evolution is to derive new properties from existing properties and assimilate new properties into the existing system to achieve new species with distinct morphology and physiology, in stark contrast with the main task of slow evolution – creation and integration of new protein coding genes from random or semi-random DNA sequences to increase the gene counts. Comparing with slow evolution which is concentrated on creation, fast evolution emphasizes on reuse and recombine through mutation-based derivation. Because of this, post-Cambrian evolution occurs in cycle. In each cycle, random mutations on protein coding genes generate new properties with altered regulatory control, As the cycle proceeds, more new properties appear and change the organisms in an increasingly greater degree, finally resulting in new species as the cycle ends. New properties that appear in a cycle are laid on top of the properties from the previous cycle, and the new species are generally more sophisticated and more advanced than their ancestors virtually in all aspects of a living organism.

As species become more complex, the underlying biochemical and cellular machines become more delicate and intertwined, requiring great balance among biochemical processes, cellular structures, and all the way to cells, tissues, and organs. An implication is that a simple system is far facile to admit new components and as a result develop into many distinct new species. In contrast, a larger and more

intrinsically interlinked system is less tolerant for adding new things in and taking existing things out, so fewer new species. In addition, the size of offspring produced in one reproduction season is positively correlated to the size of new species from a single ancestor organism due to the enormous heterogeneity of intermediates in an evolution cycle. For example fish account for more than half of vertebrate species with over 33,000 species described.

Human evolution is an interesting thing to look at. Modern humans appear just 300,000 to 800,000 years ago, while earliest primates appeared at least 90 million years ago. Monkeys that are closer to humans more than many other primates appeared about 40 million years ago, and the ancestors of the gorillas split with the common ancestors of humans and chimpanzees about 10 million years. Chimpanzees, the closest relative of humans, split from early humans 8 million years ago. The exact time for these species to appear isn't important, but it's important to show clearly that the evolution cycle that leads to humans is well more than 50 million years long. DNA sequence comparisons show that genome sequences differ not as big as expected among these species. Single-nucleotide changes account for about 1% of differences, while gene duplications, deletions, and chromosomal rearrangement account for about 3% of the differences. This 1% differences seem to affect 70% of proteins although the differences in amino acid sequences can be as small as only a couple of amino acids.

Humans differ from chimpanzees, gorillas, and other primates so extensively in every aspect from morphology to physiology to the brain size. It seems unlikely that 1% of genome differences mainly from point mutations are able to account for all the differences between two species. There are no such genes that make humans humans or chimpanzees chimpanzees. Rather, this limited genetic changes have refined and honing every bit of genetic materials to form a system in which each gene expresses in such a precision manner in terms of cell type, timing, degree, coordination with others. Derivation and utilization of protein variants in humans are so fine tuned to meet the exact functions of the cell and tissue types. In all likelihood, since splitting with chimpanzees, the mutations on human genome have greatly optimized majority of the genes to achieve the best protein products, best expression in space and timing, best combinations of components to bring up the most beautiful living organism possible.

All multicellular organisms start from a fertilized egg, while the egg provides only components to walk the first step in the entire life process. From life standpoint, it's the genome that directs the organism to complete its life cycle without input of external guidance or instructions. From evolution standpoint, the next generation of species always arises from current generation of species. As a result, evolution always moves species forward. From civil engineering standpoint, the genome is the greatest blueprint ever for making things from simple (in the eye of evolution) to unthinkably complicated. It plans and then executes every facet of a building process from design, layout, materials, overall arrangement, maintenance, and all other aspects of engineering so precisely, flawlessly, and in greatest order, details, logic, and forms. A blueprint drawn from every genome can be put into a living marvel, logically arranged, aesthetically pleasing, and economically efficient. Truly the genome is indeed the finest thing ever in the universe.

## **7. Rethink of Natural Selection**

Natural selection is the process through which some organisms in a population adapt and change better to suit the environment than other organisms in the same population, and as a result survive better and reproduce more offspring. Differential survival and reproduction of individuals are due to genetic variations that produces some favorable traits to give them some surviving advantage in their natural habitat. Upon passage of those favorable traits onto their offspring over generations, individuals of later generations become a better fit for the environment and more common in the population. Through this

process of natural selection, favorable genetic variations, thus favorable traits, are transmitted through generations. After inheritable changes that underlie genetic variations in a population accumulated to a substantial amount over unknowable generations, the individuals that carried these changes became a distinctly different new species.

On the evolutionary timeline eukaryotic organisms appeared to reproduce sexually at the single celled stage about 2 billion years ago. Since then sexual reproduction seemed parallel with the evolution of eukaryotic organisms. Almost all modern eukaryotic organisms produce offspring through their sexual reproduction system. Sexual reproduction is costly and of low efficiency, but it is ubiquitous for all multicellular organisms, indicating that it has advantage over asexual reproduction. Main advantage seems to increase genetic diversity in the population and mitigate accumulation of harmful genetic mutations.

Adoption of sexual reproduction system confers eukaryotic organisms two sets of genomes, germline genome and somatic genome. Information flow between the two genomes is unidirectional from germline genome to somatic genome. As a result, mutations in germline genome will pass on to the somatic genome of next generation, while there exists no mechanism for the organism to transmit somatic mutations good or bad to the germline genome, making somatic mutations short lived to the life span of the mutation carrier. When we talk about mutations, it always refers to heritable germline mutations unless indicated otherwise.

Any mutations can exert one of the three consequences to the organism regardless of being germline mutations or somatic mutations – deleterious, neutral, or beneficial. The current understanding is that it's natural selection that determines if a mutation is deleterious, neutral, or beneficial. Beneficial mutations produce advantageous traits, which, under natural selection, allow the mutation carrier to survive better or reproduce more offspring, and eventually become more common in the population. Only beneficial and neutral mutations will be passed down from generation to generation.

Natural selection is a fundamental element of the evolution theory. It illustrates how extraordinary biodiversity on earth has been driven and shaped by natural selection in the entire timeline of evolution in a simple and elegant way. So far there are many examples to demonstrate natural selection at work, and the origin of giraffes' long necks is the classic one. Giraffe's ancestor inhabited in dry savannahs of Africa with open plains and woodlands. There trees were tall and hard to reach for animals of normal necks like deer or antelope. Some day certain genetic mutations occurred in the ancestor's genome, which made ancestor's necks grow longer. With a longer-neck, these individuals gained advantage to reach leaves on the high position, and as a result, were able to eat more and produce more offspring. As the genetic mutations passed down generation after generation over time, individuals' necks continued to grow longer until the neck reached the length as of today. Because the long-necks was a favorable trait which made individuals adapted to the dry savannahs better, these individuals finally became the most common in the population, and they were called giraffe. Giraffe is so different from its ancestors that it becomes a new species. It could have taken probably millions of years for giraffe's ancestor to slowly develop into giraffes as they are today. This explanation seems plausible and uncomplicated even to general public.

Nevertheless, the development of life through evolution over billions of years can't be as simple and straightforward as illustrated by natural selection at all. You will encounter unsurmountable blocks if you want to use natural selection to explain evolution a little deeper and in more details. To physically support a longer neck, giraffes will have to pump more blood to the upper body and change its body shape in order to run fast and keep body balanced. Therefore, giraffes must develop stronger skeletal,

cardiovascular, and nervous systems and more. To meet these broad requirements, many changes must be introduced and integrated into existing systems, including new or altered protein factors and a corresponding gene regulatory network to make sure that the new could be assimilated into the old and work together in a concerted fashion. The long neck could become a possibly viable direction of evolution only if all these conditions could be satisfied at the same time. However such large scale changes must be of far-reaching nature and well beyond what mutations and gene recombination that natural selection refers to were able to bring about.

If giraffes' ancestor had a compelling need for a long neck to survive better, and natural selection did select animals with long necks over their shorter-necked fellows, the reality would be that giraffes' ancestor wouldn't be able to evolve and become long-necked giraffes simply because of the compelling need and natural selection. If you sincerely would like to give natural selection some credit in the evolution of giraffe's long neck, what it did at its best was to check whether giraffes could survive in its natural habitat, not necessarily survive better when their necks were grew long. It's obvious that nature won't give you something just because you need it.

Attribution of evolution of giraffes' long necks to natural selection stood on flimsy ground. Giraffes' ancestor wouldn't be the only mammal living in such a habitat. Why did only giraffes develop such a long neck, while other mammals like deer or antelope remained normal necked and have survived ever since? Didn't those mammals need long necks in such a habitat to gain survival advantage? From survival point of view, an excessively large body size provides the animals with more survival disadvantages rather than advantages. A large physical body easily hinders its movement and reproduction and requires extra large food consumption to sustain normal life activity. All this seriously limits its population size and makes the animals more easily succumb to food shortage and natural disasters. Therefore, giraffes as a new species at the time of its appearance didn't gain any survival and reproduction advantage over normal-necked animals except the banal advantage to eat leaves on tall trees. Even this advantage might not be sure if the trees in the ancient habitats weren't really as tall as today. Natural selection can't be a fact for the neck to become this long.

Giraffes and its closet relative short-necked okapi diverged from their common ancestor about 11.5 million years ago, yet giraffes and okapi shared only about 20% identical proteins, attesting the great magnitude of genome changes during the evolution of giraffes. Giraffes appear in the fossil record around 4 million years ago. A time span of 7 million years seemed too short to endure the genetic changes of this grand scale that gave rise to giraffes that bear a long neck.

Assume that giraffes shared a common normal necked ancestor with deer or antelope. One individual ancestor suffered from some mutations in a gene coding for a protein factor that guided neck muscle development. The mutations somehow altered its biochemical properties and guided neck muscle to grow longer. In other words, the appearance of giraffes as a new species was likely to be triggered initially by some random mutations of similar kinds. The giraffes' neck development wasn't a single event isolated only to the necks, but affected giraffes in its entirety. The giraffes had to mass a large number of proteins with new or altered properties to build up a phenotype – a long neck and everything else that came with the long neck. All genes coding for these proteins must first be made available, and then be linked together through regulatory elements that control their differential expression in different tissues. This process was so complicated and interlinked, requiring the greatest coordination and integration in a relatively short time, which occurred far beyond the control of natural pressures for the animals to eat leaves on tall trees. If overall changes to the genotype couldn't support the long neck, which must be common during the process, giraffes could be deemed as biological defects and eliminated due to failure to grow, survive, or reproduce. A long neck would become inevitable if

cardiovascular system, nervous system and other impacted systems could make sweeping changes to suit the long neck as a whole in one evolution cycle. It's highly dubious that natural environments and habitats could play any roles in the evolution of giraffes.

Modern research hasn't shown something in nature that could direct an organism to change particular genes or mutate particular base pairs to gain desired beneficial traits. There is no mechanism through which an organism could be aware of which genes were involved in the evolution of a particular trait or engaged in a given biochemical process. All traits are developed through random changes in the genomes over long period of time, not due to any need for this or that kinds of traits. Organisms will accept any traits random mutations can bring about so long as the new traits are not lethal.

The best examples of evolution and natural selection came from Charles Darwin's observations of bird finches in Galápagos islands. Finches' bill sizes and shapes are attributed to each bird's adaptation to a specific type of food on the islands. For example, a thick beak adapted to feeding on crunchy seeds and arthropods, while a slender, pointy bill to catching tasty insects hiding between the leaves, and twigs or cactus spines to prying arthropods out of holes on tree trunks. There are more examples to the list. Curlew's long bill can probe deep into the mud and shallow water to catch aquatic invertebrates. Great egret's long legs allow the birds to walk in relatively deep water to search for fishes.

If we think a little more, it's not difficult to realize that appearances of birds' these highly specialized bills or legs in the course of evolution and their life styles are actually the chicken or the egg problem, a causality dilemma if you are stubborn enough to put them in order. A bird was considered to have developed long legs in order to adapt to deep water habitat, and developed a thick beak so it could feed on crunchy seeds and arthropods. There is an equally sound and even much more logic explanation. A bird had developed specialized bill or legs first. Because of the long legs, the birds gained the ability to enter the water and look for fish swimming in the water. This could be considered an active adaptation of the deep water habitat, in stead of nature forcing the birds to find food in the deep water for survival. Similarly because of the thick beak, the bird became able to feed on crunchy seeds and arthropods first, then they feed on crunchy seeds and arthropods as they are today.

Could natural force make a bird's legs long or beak thick so that those birds could survive better? In the later explanation, natural selection was no more than a type of adaptation to the natural environment. It's the birds that played active roles upon obtaining special traits that allowed them to adventure into a proper natural environment. A bird can't control what kind of favorable traits it can have, but it can put whatever favorable traits it already has to practical uses by actively finding a natural environment that best fits those favorable traits.

The active adaptation is more likely to be what has happened during billions of years of evolution. More examples could be found to support this view. Fish living in the dark cave is usually blind. It isn't the darkness that makes the fish blind, instead it's random mutations in the visual system that have caused fish blind, and blind fish could survive in the dark cave due to lack of predators there. Swimming is an ability beneficial to a mammal's survival, but not all mammals can swim. The enormous body sizes of many dinosaurs shouldn't be the result of natural selection to adapt to any natural environments. Excessively large antlers on some male deer could be detrimental to their survival when traced by their predators in dense woodland. Some traits are good, but not for every animal, while some traits are bad, but appear on some animals.

On the Galápagos islands again, a completely new finch species was created in the wild in just two generations by mating of two different species of finches. The importance of this observation was over

exaggerated. Mating between different species is not often. First different species don't attract each other for mating. Second fertilization couldn't occur due to recognition failure between an egg and a sperm. Third if fertilization succeeded, the hybrid offspring would carry two sets of proteins serving the same functions, but with different amino acid sequences, one from mother and one from father. These two sets of proteins were likely unable to interact with each other, thus disrupting normal biochemical processes and leading to the death of the hybrid organism. Fourth if hybrid offspring did develop normally, it was often sterile or reproduced with difficulty. If the hybrid offspring was able to reproduce, then it indicated that the two parent species were close enough for mating, and nothing more. It would not be scientific if the hybrid offspring was judged to be a new species just based on their appearance, food preferences, etc.

Almost all organisms enjoy a sort of ability to tolerate the environmental changes big or small, and this ability is built in the genome. A lot of organisms would have gone extinct without tolerance for changes. Animals like rats can tolerate a broad range of environments and live all over the earth because their tolerance is exceptionally strong and lax, while animals like some amphibian species can survive only in niche space with ecologically strict conditions because their tolerance is poor and rigid. It seems unlikely that it was natural selection that made rats able to tolerate a variety of natural environments and some amphibians tolerate no environment outside their current habitats. Favorable traits in a population are sufficient to enable the population to adapt to different kinds of environments without selective pressure. Natural selection doesn't seem to play much of a role in the evolution of species as it has been generally accepted in the scientific community.

Throughout the entire evolutionary course, natural environments had made profound impacts on the evolution of species through geologic and climate events. We learned natural selection from the basic biology classes, but what it really means could be misunderstood easily because of ambiguity caused by the word "natural". Natural selection will preserve any traits that make individuals survive and reproduce better in the population. New species can result if new traits made the organisms too different from the traits seen on their predecessors. Natural selection can be easily understood as a determining force that moves evolution forward by putting natural pressure on the organisms to develop and adapt. To many, natural selection is the engine that drives evolution. But they forgot about the fact that there exists no feedback mechanism between phenotype and genotype. A good phenotype or trait can't be made better simply by feeding back to the genome and asking for necessary changes to make me better.

The importance of natural selection as a theory in modern biology doesn't need to be emphasized more. However, can natural selection really explain evolution of species as it has been claimed for many many years? In free, natural habitats, individual organisms in the population carry more or less their own random heritable mutations, including mutations from DNA rearrangement. The phenotype of those mutations could be bad, neutral, or good, affecting more or less the survivability of the mutation carriers. Because mutations are of random nature, it is unknown what effect new mutations will have on the survivability of individuals, and the consequences are discrete beyond anyone's control. Random genomic changes that were detrimental to mutation carriers were hardly able to survive beyond birth.

Every object, living or non-living, has inherent properties that determine its behavior on the macro level. To a living organism, its behavior includes its special diet and unusual habitats. Why is it important to consider the behavior of organisms? The behavior determines what an organism will do to other individuals in the population and beyond and to natural environments as well. For organisms all this is encoded in their genomes. As a result, different organisms behave differently. Some organisms are quite hostile, even belligerent in behavior towards others, showing very competitive nature. Other

are weak and seemingly vulnerable to natural predators, but they have been in existence ever since they emerged from the evolution. Special behaviors shield weak organisms from the dangerous food chain and survive in their own safe niches.

In the free wildness, fierce competition, special behaviors, food preferences are the ultimate way to decide who will lead to differential survival and reproduction and dominate the population. Competition among individuals of same species or different species occurs spontaneously thanks to individuals' behavior. By controlling the behavior of a species, evolution has devised a mechanism to preserve and pass down mutations that could be translated into a behavior that gave rise to advantage in survival. Individuals with better survivability would finally become dominant in the population. Evolution has conferred all types of behaviors to millions of species, allowing them to survive and even thrive in a variety of environments that suit their behaviors. Because of this, there is tremendous biodiversity on the earth today. It must be emphasized that the appearance of any favorable mutations are not intentional, but out of random events on the genomes.

A trait in general is stable over time and has a quite complicated genotype behind. In an evolution cycle, a number of genes, newly created and variants of existing, must work together to engender new functionalities called traits. The new traits are always part of new species, and could impact the survival of new species in terms of food and natural predators. New species would perish from diet scarcity or increased risk of natural predators in the old habitats, unless they were able to migrate to places where food was ample and natural predators wouldn't endanger their existence. The appearance of a new trait wasn't necessarily to benefit the species to inhabit its native environment, rather it could be a rational way to disseminate the living organisms across lands and settle in different territories.

For example, the ancestor of giant pandas might be a mammal indigenous to an area where bamboo wasn't a widespread plant species. During the evolution, panda developed a unique digestive system that contained taste buds for bamboo and was equipped with teeth that could chew bamboo, and stomach that could digest bamboo. Such a peculiar digestive system prompted giant panda to migrate from its native habitat to places where bamboo was abundant. From genetic point of view, panda genome determines bamboo as its major diet, and bamboo diet in turn determines what kind of behavior panda will exhibit. It's panda's special behavior that prompted panda to select bamboo rich terrain to be its native habitat. A narrow appetite for bamboo put panda in a grave disadvantageous situation for its survival, but it can't be changed through natural selection. It isn't happening that a trait can be changed simply because it isn't a good one.

Natural selection is merely an empty shell without substance. It is too vague and superficial to explain evolution of species, especially too problematic and contrived to explain how a population can evolve to be sufficiently different and finally become different species. After we have a better understanding of how evolution works, the classic example of evolution – the origin of giraffes' long necks – can't be explained persuasively using natural selection. It would be too irrational to think that the dire need to eat leaves on tall trees would drive an animal's neck to grow long. On the other hand, if mutations had started to lengthen an animal's neck, nothing could stop it from growing long, unless the whole process went awry. If everything went smoothly, it would take at least tens of million years for giraffes' ancestor to develop into a phenotype that is sufficiently different to be a new species we call giraffe. This doesn't seem to be a process that natural selection theory would predict. Similarly, vultures' craving for dead animals couldn't be accounted for using natural selection theory either. The bird's whole biological system must be reshaped in order to feed on dead animals. First the bird must gain appetite for putrid carcasses to be its diet. Second the birds must develop a strong stomach to digest and kill infectious agents coming with the dead bodies. During the evolution, most bird intermediates didn't

have a strong stomach for dead animals, and they would perish from the infectious agents. Fortunately enough, one intermediate developed a strong stomach from random mutations which allowed this evolution cycle to continue to an end with the emergence of vultures. Third the birds must establish a nerve-muscle system that would allow the birds to look for targets from high positions. Forth, the birds must develop some peculiar behavior to support their strange diet. All this couldn't be achieved without genome wide changes that must have taken place in sync. In all likelihood, the appearances of giraffes, vultures, and numerous species were way beyond what natural selection could explain. Moreover, natural selection is unable to account for the explosive appearance of a large number of new species virtually in a simultaneous mode in a short period of time.

Everything on the planet earth is covered by sky, and similarly almost everything that is happening or existing on earth is covered by natural selection as why they look as we see today. River XX started from a place deep in the mountain range. In its early life, river XX was composed of many small branches, each of which started in different areas in the mountain and converged to become river XX when they flew out of the mountain. As time passed, one of the branch called branch X became wider, while many of other branches got blocked in their way out and emptied their water into branch X instead. After a long time evolution, branch X continued to widen and accepted a majority of water emptied from other branches. When main branch X flew out of mountain and merged with few remaining branches, they formed river XX, the largest river in the area known to people today. This is so similar to how an advantageous trait made the trait carrier the most common in the population. The formation of river XX seems to be the result of natural selection. Here is one more example to end this random thoughts. On a flat land stands a big stone, while its surrounding area is covered by small stones. It's evident that these small stones were left there after some big stones eroded by wind, rain or other natural elements. What can be drawn from these small stones is that the lone big stone has resisted the same natural elements that eroded its neighboring stones over years. Its survival seems fit well with natural selection theory. Is there anything that isn't the result of natural selection?

## **8. Summary**

Looking at the life forms that appear along the timeline, the evolutionary course could be divided into three stages as shown in Figure 2. Mystery of life began with the appearance of the primitive form of life in the first stage about 4 billion years ago, followed by the second stage, a 3.5 billion year long period in which life slowly but steadily evolves into multicellular forms with genomes reaching moderate sizes. The third stage is the stage characterized with explosion of all forms and complexity of life in a time span as short as 600 millions of years.

There is every reason to believe that nascent earth was a life-welcoming planet, in which a place called incubator of life was furnished with a mix of amino acids, nucleobases, sugars, lipids, and other inorganic and organic chemicals. Random chemical reactions occurred spontaneously and constantly in the incubator, generating all kinds of possible chemical products, including polypeptides, ribonucleic acid RNA. All these polymers were of random nature, various in length and sequence. As the amount of random polymers increased, some RNA happened to fold into structures similar to tRNA, rRNA and mRNA, and some polypeptides happened to fold into three dimensional structures with rudimentary biochemical properties, including preliminary enzymatic activities such as RNA polymerases, DNA polymerases, and ribonucleotide reductases, and structural components to form primitive ribosomes and primitive protein complex for DNA and RNA synthesis.

With the availability of the low grade protein complex for RNA and DNA synthesis, DNA grew longer in a random fashion and made copies through replication, while RNA was copied from DNA templates through transcription. Among RNA population, including those transcribed from DNA templates and



those randomly polymerized, were RNA molecules similar to modern day tRNA and rRNA. The rRNA like molecules could complex with some ribosomal-like proteins to form rudimentary ribosomes. The tRNA like molecules could contain a 3' end able to accept an amino acid and an anticodon in the opposite side. When tRNA molecules charged with amino acids aligned on the mRNA molecule attached to the ribosomal platform, adjacent amino acids reacted to form a peptide bond with an efficiency greater than random polymerization.

In the very early phase of the Stage 1, the life system was constantly changing in all its components. DNA sequences were quite random due to random elongation and error-prone replication, so the RNA and peptides derived from the DNA templates. As more peptides were produced from DNA templates to take over random polypeptides, some of them began to show nice properties to function as enzymes, ribosomal proteins, structural proteins, trans-membrane transporters, and so on. Gradual appearance of enzymes with an increased variety, better catalytic activities and higher specificities brought the early life into an enzyme era, making DNA replication, RNA transcription, and protein translation more reliable and consistent. Meantime, DNA sequences that served as templates for the peptide synthesis transformed slowly into gene-like structures, further increasing the reproducibility of protein molecules, indicating that the minimum genomes started to appear. Slow but steady improvement and maturation of the basic biochemical machineries marked the successful transition of early life away from randomness into consistent and disciplined operations. It's the infinite randomness at the beginning that generated infinite amount of random peptides, RNA and DNA. Only infinite amount of random stuff could serve as a cache of great treasure. Life was born out of sheer randomness.

Early life arising from randomness varied in all forms, dependent on the amino acids and bases available in the environments. Different forms of life were ultimately attributed to the use of different set of codons for protein translation. The earliest primitive cell – single celled life – formed when one particular minimal genome was enveloped in a lipid bilayer membrane. This single celled life relied on a single set of genetic codons corresponding to a single set of amino acids for protein synthesis, and is the common ancestor of all modern living organisms.

The early single celled life was too simple and too flimsy to withstand any adverse impacts from the environments. In stage 2, life had fully developed its biochemical processes, cellular structures, and genetic machine and greatly improved the overall efficiency, reliability, and survivability, successfully metamorphosing into full-fledged organisms,

The genetic system of early life was far from complete and robust, and its genome expansion was basically the continuation of Stage 1, largely random and of low efficiency. Increase in gene number allowed organisms to produce more enzymes of different kinds, which in turn allowed organisms to operate more metabolic pathways and perform genetic recombination with increased accuracy and efficiency. It was expected that numerous forms of life arose in the process due to randomness in the early phase of stage 2. Each form of life was likely to possess a unique set of proteins despite of using the same set of genetic codons. Existence of various life forms made it possible for multiple cells of different origins to merge into a single cell at the prokaryotic time. Merger accelerated the enlargement of genetic materials, widening the coverage of metabolic pathways, and finally compartmentalizing cellular structures into organelles, all of which were characteristic of an eukaryotic cell. Confinement of genome inside the nucleus ushered in the era of eukaryotic life.

Transition from prokaryotic life to eukaryotic life must be supported by additional set of new proteins, so the same for transition of single celled eukaryotic life to multicellular eukaryotic life. Multicellular organisms are not simple aggregates of cells of the same type, but the aggregates of cells differentiated

into different types packaged in a specific way. Data in Table 1 show dramatic increases not only in the genome size but also protein coding gene counts in selected eukaryotic organisms over selected prokaryotic organisms. Similar data can't be obtained for early life and full-fledged prokaryotic organisms because no baseline is available for comparison, but the data would be expected to be compatible with Table 1. Considering the ultra long period of stage 2 and buildup of multicellular life from the very bare necessities of early life, creation and assimilation of new genes, including regulatory elements and their protein products, seem to be the hardest bottleneck to break for evolution to move forward. Infinite randomness is again the sole means to achieve the data shown in Table 1, albeit at the cost of time. It can be concluded that infinite randomness is the sole driving force that has moved evolution forwards in stages 1 and 2.

The end of stage 2 is the beginning of stage 3, which started with Cambrian explosion. Stage 3 is different completely from stage 2. Organisms had so far amassed a large number of protein coding genes comparable even with mammals, and as a consequence, the mode of evolution changed from total randomness to reuse and assimilation, making evolution progress in the form of cycles. Organisms emerged from each cycle could be classified into the same class, a subdivision of a phylum in the biology classification system. For example, all amphibian animals evolved out of one cycle can be classified into Amphibia class, and all fishes that are classified into Actinopteria class belong to the same batch of fishes out of a single one cycle.

Not all early organisms were eligible for evolution. Only organisms with special genetic capacity would become ancestors of later organisms. Natural upheaval served as perturbation to push ancestor organisms from a disarmed state into an armed state, in which the genetic machine is more error prone than at normal time. Higher frequency of mutations caused genome-wide changes, bringing the organisms into an evolution cycle. All offspring that born in the cycle were intermediates of the cycle, and in an armed state. Upon entering the cycle, the genomes were constantly reshaped by random point mutations and gene duplication. New biochemical properties were derived from duplicated genes after action of point mutations. As perturbation disappeared, point mutations continued to slowly refine mutated genes, resulting in creation of new properties for the organisms. This process is a healing process that brings intermediates back to a disarmed state. During a cycle, majority of intermediates died from mutations of lethal nature, while lucky ones not only survived but emerged as new species. Evolution cycle brings about new species in explosive mode.

Evolution cycle is an important concept. It delimits the time period in which organisms begin to evolve from the time at which organisms are in a consistent, stable, calm, and disarmed state. As a result when we talk about evolution, we just focus on the evolution cycle. What happens in the cycle is what happens during evolution at a specific time on the history timeline. Another important thing to keep in mind is that evolution cycle can take up to tens of million years or generation to conclude, which, from a cumulative standpoint, implies that genome wide changes spread over millions or more generations and changes that happen to each generation must be limited in scope.

Reuse based evolution is the most distinguished characteristic of evolution in stage 3. Reuse greatly accelerated the emergence of new species. The result of reuse is the generation of protein variants. Duplicate genes encompass all genes of duplication origin, including active genes and pseudogenes. Gene duplication, as one type of DNA rearrangement, still occurs in modern day organisms. However, duplicate genes here refer only to those that led to functional protein variants after bombarded with random point mutations over tens of millions of years. Duplicate genes are free to diverge via gradual accumulation of random mutations along the evolution path so long as they are not expressed or their products are not lethal to the organism.

We can finally draw a picture to reveal the essence of evolution since Cambrian explosion from the reuse point of view. Gene duplication combined with random point mutations is the basis of reuse. In each evolution cycle, genes are duplicated from master genes from the previous evolution cycle. As a result, new species from this cycle are always more advanced and sophisticated than species which started the cycle. Simply put, reuse based evolution cycle is to make a new series of protein variants from proteins that are actively engaged in life processes in the ancestor species. When this new series of protein variants are integrated into the life system at different time and development stages, in different cells, tissues, organs according to their expression control, they reshape the organism in such a fundamental way that the organism has become a new species. The changes brought up by protein variants are broad and their impact on the existing biochemical processes and cellular organization is far reaching. For example, substitution of some subunits in a multisubunit protein with protein variants can bring subtle change to the biochemical properties of the protein, making it more desirable for the process or structures in the cell types. Variants of development inducing factors can change the final morphology of the organisms by tweaking the embryonic development processes. Protein variants can fill the functional void or improve old functions in the new species as in color vision of primates, stronger stomach of vultures, more sensitive olfactory buds in some organisms. Different protein variants will be expressed at different stage of the cycle, and accordingly change the look of organisms visibly. In a nutshell, in an evolution cycle, a large number of active gene products have been substituted with their variant counterparts at different time points and tissue locations over a long period of time, resulting in the re-establishment of a balanced system in which all components of new and old work together as a single unit just as in the ancestor organism, except the organism is no longer the same as the ancestor organisms both morphologically and physiologically. Any intermediates that can't re-establish such a balanced system perish in the cycle. When all intermediates complete metamorphoses into a number of new species, the evolution cycle completes as well.