

# Evolution with a Purpose



## 6.868 Final Paper - [Manolis Kamvysselis](#)

Some like Netscape [frames](#), and some [don't](#)

---

### Disclaimer

This article was written before i knew any biology. It has a lot of misconceptions about how things work, and uses all the wrong vocabulary. Please at least read "non-coding region" every time i wrote "introns". Despite the embarrassment, i leave this online for historical purposes only. Then again, some ideas are worth a thought, knowing that it was all theoretical and abstract, prior to any bio knowledge on my part.

---

## Motivation

On page 318 of *The Society of Mind*, Marvin Minsky suggests that "we can easily conceive how machinery could exist inside an animal, to purposefully direct some aspects of its evolution." Fascinated by the idea, but unhappy with my background in the matter (high-school biology), I ended up having to read quite a few pages to avoid making a report full of air. Starting with some basic assumptions and intuitions listed in Section IV, I searched a few interesting biology books I could find on the topic, and I propose here some mechanisms that would enable a new theory of evolution presented in section V.

### The metaphor - DNA is a computer program and exons are where the data is stored.

- The purpose of most of our DNA remains unclear to scientists today. Only one tenth of our genetic code is used to create all known proteins in our bodies.
- It is this paper's hypothesis that this tenth which has been mapped to protein creation is simply data stored in variables, and that the largest part describes how this data should be used.
- If we think of DNA as a huge computer program, most of it would describe operations to be performed and meta-information in some genetic programming language, and only a fraction of the program would contain genetic data that actually encodes a human's genotype.

### Purpose - to develop goals and constraints for such a language

- The purpose of this paper is to envision such a genetic language for encoding meta-information in a string of DNA.
- I tried to use biological evidence and knowledge about genetic mechanisms in basing my hypothesis, but the proposal is by no means based on biological observation, but merely a thought experiment
- The only contribution of this paper to genetics would be to pay more attention to introns as intelligently constructed rather than garbage accumulated over generations

### Vision - DNA encodes what species learned about survival and about Evolution

- Just like information about the individual is encoded in the DNA data (exons), in this meta-information (introns) would be accumulated knowledge about evolution itself.
  - This information would be used to favor mutations more likely to yield the fittest individuals. It would be thus guiding the evolutionary process, directing it, providing it with a purpose.
- 

## I. Darwin and Evolution.

The non-evolutionist humans are becoming extinct under the pressure of Darwin's theory of evolution, and this is a good thing. Under the same name of Darwin, however, seems to have disappeared the notion of purpose in the natural sciences. It is taboo to study why a certain mutation occurs, and how it is related to the environment, the previous mutations, the genetic history of the species.

### a. Random change and testing.

This fear of seeking purpose-directed mutations has come about to avoid mixing engineering and biology. Evolution doesn't have a purpose. Building humans or intelligent creatures is certainly not one of them. It is wrong to think that nature is aiming to populate all niches available, land, water, air. In fact, the beauty of Darwin's theory comes from this very fact, that there doesn't have to be a purpose for evolution to occur. All the magic comes from the fitness function that lets only the fittest survive.

## **b. Intermediate steps must be advantageous.**

However, as one tries to simulate evolutionary processes in purely random mutation and selection, the time it takes for a fit individual to appear is combinatorial in the number of elements that compose it. Simultaneous mutations yielding the ability to fly instantaneously could simply not have occurred. It is therefore thought that each small step towards wings was advantageous to the individuals that were thus able to survive enough generations until the final mutation which allowed them to fly. In fact those advantages need not be related to flying, but could be useful in providing heat dissipation, walking balance, or rain protection.

## **c. Co-evolution of environment and Individuals.**

Another notion that has to be incorporated in understanding species evolution is how they relate to a changing environment. Changes in individuals make changes in their environment to which individuals have to adapt. Individuals evolve and shape an environment which shapes them.

---

# **II. Strong evolutionary evidence unaccounted for by Darwin**

The Darwinian model of Evolution has successfully provided an explanation for many discoveries about our natural world, and revolutionized a field with the powerful idea of random mutation and natural selection. However, even though this theory is beautiful by its simplicity and powerful by its predictions, there are many important points that it misses about evolution in general.

## **a. Evolution speed is exponential in the number of possible permutations**

Even with the notion of advantageous intermediate steps, it is hard to believe that more complicated species evolved faster than less complicated species. Indeed, the more complex species become, the harder it should get for new interesting features to arise with simple random mutation. The number of possible mutations increases as the DNA becomes longer, and the likelihood of simultaneous mutations yielding interesting phenotypes is very slim. Evolution has demonstrated the opposite of what a random process would display. The more complicated species became, the less time it took to evolve to even more complicated species.

Time Frame:

Origin of the Earth	5 billion years
Bacteria and algae	3 billion years
Filamentous algae	2 billion years
Marine invertebrates	600 million
Fish, land plants	400 million
Amphibians	300 million
Dinosaurs	200 million
Mammals	150 million
Primates	20 million

## **b. Mutation rates do not correspond to evolutionary rates.**

The Homo species evolved more rapidly than frogs, chimps, etc. That is, humans developed more complex traits (eyes, brains) over a shorter period of time. More complex organisms exhibit a smaller number of generations, since each life cycle is longer. Instead of finding a higher mutation rate that would make up for this genetic slow-down, one observes a smaller number of mutations at each step of the reproduction of complex organisms. Mutation rates therefore do not reflex evolutionary rates.

This smaller number of mutations is due mostly to the precision of DNA replication in higher organisms, that allows for less errors during duplication. Complex error-checking mechanisms have been developed in eukaryotes and reduced the mutation rate by orders of magnitude. Somehow we have to believe that those mutations happening in more advanced species are somehow more effectively focused towards beneficial changes, or that they happen in genes likely to have an interesting change. More complex species would have accumulated knowledge also about the usefulness of mutations, and become better at evolution itself.

### c. Long periods of stability are interrupted by Evolutionary Bursts.

In a world of purely random mutations which occur uncontrolled anywhere in an organism's DNA, we would expect evolution to be more or less linear, an interesting trait appearing randomly every now and then, and followed by another interesting change only after a long time. However, it is a consensus nowadays that this is not how the evolutionary clock has been ticking. Instead, we observe long periods of stability between evolutionary bursts where new species are created and specialized in a relatively short period of time.

If every new species along the line to a bird was advantageous over the other species before it, we would expect to find every ancestor along the way outliving the species that preceded it, since every step would give individuals evolutionarily more fit than their ancestors. However, we find more species at the same evolutionary step, and not a progressive line of changed species. We have to assume that it took a shorter time for many changes to occur one after the other, once new features had started appearing. That is, in the formation of any single new trait, we have to assume that many steps occurred rapidly to achieve a finished and evolutionarily advantageous trait, before natural selection can work against an individual with wings too weak to fly and hands too long and inappropriate to any survival task.

Just like some animals seem to undergo multiple changes at roughly the same time, making their evolutionary clock tick faster, for others evolution seems to have frozen for millions of years. Nautilus for example has kept its present form for millions of years. If we assume the Darwinian model of random mutations in every DNA replication, it is highly unlikely that the same form persists unchanged for hundreds of thousands of generations.

Evolutionary rates thus seem to be changing from species to species, as if individuals could enter, voluntarily or not, an evolutionary mode where new traits appear, and then exit it for millions of years. Moreover, it seems that there exist different evolutionary modes, where specific characteristics are changed, for example, a posture-changing mode would act on the spinal cord structure, whereas an arm-changing mode would act on the length, size, orientation, of the arms. More than one such modes can be entered simultaneously, especially if the traits are related (an upright posture cannot be assumed unless the arms grow smaller). However, when the arms start changing for example, they will undergo more than one change before the animal leaves the arm-changing mode.

### d. Seconds of embryonic evolution reflect millennia of species evolution.

Before becoming a human baby, the human embryo seems to retrace the evolutionary ladder of humans, first looking like a fish, then reptiles, and finally earlier mammals. The same is true about embryonic stages of other species. Embryonic jellyfish look more like polyps than like adult jellyfish. Where does nature find all the information about the different species in an individual's genetic heritage? It seems that information is kept around in the DNA about the different species that an individual evolved from.

Why is this information re-expressed when a child undergoes its growth in its mother's womb. Many theories are possible. Maybe those species are better fit to the environment in which the embryo lives, and hence express their characteristics which would otherwise be hidden. This theory would assume that DNA is a huge rule-based system, where characters are expressed on demand. This is the case for some genes, such as the temperature-dependent black pigment of Himalayan mice, which gets expressed at the nose, tail, and toes of a mouse, where the temperature is low enough. However, although the environment can play a role in the expression of genetic information, assuming that the embryo can change its position on the evolutionary scale based on environmental factors seems a little far-fetched. This would moreover assume that this evolutionary process would change if the embryo were taken out of its liquid environment. Further we can ask, why aren't there organisms in some tropical zones which change from fish to mammal when the hot weather comes, and back to fish when it starts raining again.

Alternatively, we can imagine an encoding in human DNA of incremental differences which lead from bacteria to humans. As the beginning of DNA is read and expressed, the steps are followed all over again, taking the cell through the evolutionary process that led to a human. This would mean that characteristics build up on top of each other, and as the next one is read the previous one is either superseded or replaced. Mutations however are not encoded incrementally, since they change a gene when they occur. This would lead different embryos through different stages, depending on the mutations which have occurred in the parts of the DNA that encode the different evolutionary steps. That is, even though embryos of mice and men both go through a fish stage, the fishes themselves will be results of different mutations and hence different. When mutations occur very early in the string of DNA (where early is determined by the order in which it is read in a developing embryo) it is possible that the embryo does not develop at all. Hence, the most crucial parts of DNA should be encoded early in the string, so as to ensure a fail-fast system. We could even assume that the DNA would be selected against even at the spermatozoid stage, which could be a first expression of the initial DNA code. This can save a tremendous number of generations in evolutionary speed by selecting early against individuals which will not survive.

## III. Genetic evidence that enables higher-level language

We saw that the Darwinian model of random mutations does not successfully provide an explanation for many facts about the evolution of species and individuals. There are moreover genetic facts that suggest a more complex theory would be needed, and that provide hints on a more advanced model of evolution.

## **a. Non-translated parts of DNA exist in more complex organisms.**

For a long time, scientists thought that the entire DNA coded for proteins. In bacteria for example, there is an exact correspondence between genetic distance of codons (determined by mutation frequencies) and protein distance of amino-acids (measured in number of amino-acids). This is in fact the case for most early organisms. However, in eukaryotes, there exist intervening sequences, or introns, which are not translated.

It is still thought in the community, that these introns are the uninteresting regions of DNA, since they are not involved in determining the phenotype. The high mutation rate that they bear, as compared to the rest of the genetic code, has been interpreted as proof that they do not carry a genetic message, and thus can change without affecting a specie's evolution.

Instead, in this paper we postulate that like information about surviving is encoded in exons, information about evolving is accumulated in introns. For every piece of translated information in an exon, there would be some meta-information in an intron about how a gene came about, what it represents, what is its certainty, history, future, and importance. These different pieces of information will be analyzed in detail later in the paper when the functions of the language are described.

## **b. Universality of the Genetic Code broken for ending codons.**

The genetic code used to translate codons into amino acids is universal. That is, it was developed very early in evolution, at the stage of primitive cells from which all living species evolved. Because the code is so ancient, and in fact took such an incredibly long time to evolve, along with life, we will not interest ourselves to the how exactly a three-base code was developed. Let us simply hypothesize that maybe early in evolution only one of the three bases was used in each codon, determining the corresponding amino-acid. As the number of amino-acids raised, more bases of each three-base codon were used in the encoding (it is hard to believe that the number of codons itself evolved say from two to three, since no useful sequence encoded in two-digit code could have any particular property when interpreted on a three-digit code). Thus, very early in evolution, the universal genetic code was fixed to an encoding complex enough to encode for a rich variety of amino-acids, but compact enough to use a minimum number of codons. This balance between expressiveness and simplicity yielded a codeword with 3 bases, optimal in encoding 20 amino-acids.

The exceptions to the universal genetic code occur in the higher-level language structure of it. More specifically the termination codons, which are encoded by UGA, UAA, and UAG are changed in some simple organisms. For example, in a mycoplasma, UGA codes for tryptophan (just like UGG), and in certain species of the ciliates UAA and UAG code for glutamine (just like GAA and GAG). This suggests that termination came about later in the evolution scale, as higher-level constructs started appearing. Also, the fact that it appeared in different places for different organisms, along with the fact that it appears in every single living organism suggests that a higher-level construct was necessary in the development of organisms and their evolution. The ones which did not develop an ending codon were not fit for survival, and hence did not further evolve.

## **c. Existing Processes and Tools suggest a higher language of evolution.**

In microorganisms, the error rate of DNA replication is about one per billion or ten billion nucleotide replications. This cannot depend solely on the accuracy of hydrogen bonding of AT and GC pairs, in which case the error would be closer to 1 in 100. Indeed mechanisms exist that drive this error down drastically, such as the moving back and forth of the polymerase to get a second chance at the replication if a mismatch has occurred. There is evidence that a post-replication repair system in bacteria recognizes the new strand of DNA and changes incorrect bonds, such as AC, and replaces either A with G, or C with T, after correctly identifying which is the original strand of DNA and which is the duplicate on which the error has occurred). The complex behavior of such error-correction capabilities on DNA suggests the high-level language we envision in fact exists.

Along with error-prevention mechanisms, there exist site-specific mutagenesis. Specific mutagens which can induce specific mutations, such as changing AT to GC or a frame-shift, or any single-base change. In fact creation of a specific mutation is more the breaking of its prevention mechanism, but this breaking could be directed in a specific way.

Along with post-processing site-specific error-correction capabilities, there exist processes which duplicate genes, translocate them, cut them, or insert parts in them. From a computer scientist's point of view, a very complete data management system is hidden in the structure of DNA. There are many processes in nature for which we have yet not found an explanation, such as the use of Polymerase II to which no function has yet been assigned. Polymerase I is a helper to Polymerase III, which is the main DNA production enzyme. (The names correspond simply to their discovery date). Class 2 tRNAs contain an extra arm whose functionality is still unknown. Such crucial steps of genetics are unknown and could be dedicated to more complex mechanisms current theories could not account for. Instead of speculating on uses of those tools in interpreting higher genetic languages, we will spend the next section describing how basic functions needed in such a language of evolution appear to exist already in nature under one form or another, or how they could have evolved (or evolve some day) from existing mechanisms well understood in nature.

## IV. Functions of a Language for Evolution.

For a higher level genetic language to exist, basic functions must be present. The most basic form of controlling evolution is to specify mutation locations and rates from within the DNA string. The genetic code should also be able to keep backups of present states to fight against detrimental random changes. There must also be some way of understanding a purpose to the evolutionary steps followed along the way and establishing relations among genes. Once such relations are established, there must be a way to coordinate the evolution of such related genes. Finally, there must be some way of hiding partially completed evolution such that larger evolutionary steps can be made without being penalized for unneeded complexity in the phenotype.

### a. Specify mutation locations and rates.

There is proof that elements in DNA specify expression locations and rates (which genes should be expressed and how strongly). Min et al. [86] showed that deletion of an intron fragment showed a 2.5 fold reduction in the level of the gene's expression. In addition, the insertion of the same fragment upstream of the promoter showed transcriptional activity 3 times higher. They showed that not only the promoter region but also the first intron may be important for the regulation of mouse gene expression. Thus we can find encoded in the DNA a phenotype control mechanism to express interesting genes only, and to control the rate at which they translate.

Similarly we can envision gene-specific mutation control mechanisms that determine which genes of the DNA heritage should be preserved intact, and which should allow mutations to occur. Along with every gene could come an encoding of how much care should be taken in duplicating the gene. This encoding could be in the form of a sequence to which bind error-checking molecules, and that continue along the 5' - 3' path to check the work of replicons. Alternatively, we could picture that different types of replicons do a better or worse job at duplicating DNA, and that such a control structure would bind to particular types of replicons. Precise replicons could be requiring a more complex binding structure to start replication on a gene, while less precise replicons could replicate any gene. As organisms evolve, the more advanced genes could require more advanced replicons, which bind only to them, while less advanced replicons only bind to less important genes. Alternatively, this encoding could be addressed to the error checking and correction mechanisms, rather than replicons themselves, selecting if a gene replication should be left with errors or not.

A location-specific mutation inhibiting functionality seems to exist already. While in prokaryotes, cross-overs can occur anywhere within a gene or between genes, in higher organisms, most crossing-over occurs in the region between neighboring genes. Cross-over thus has no mutation effects, but instead is beneficial in enriching the genome by transmitting information on alleles coming from different parents onto the same one to transmit to the next generation. Again if some encoding exists on how beneficial is each trait, and how it combines with other traits, this selection can be done intelligently, transmitting together characters which are beneficial together (for example, large head from genes of father could be combined with large pelvis diameter from the mother genes, if a notion exists and can be extracted that the two are related).

It makes us think therefore, that recombination processes are non-random in higher organisms, and that there is some high-level understanding of DNA, as to which parts of the code should be preserved, and which are likely to be useful when changed. Genetic transformations can thus be directed to specific locations.

In 1927 was first published evidence that external factors could influence the mutation rate. X-rays were shown mutagenic in cereal plants. During WWII, Drosophila was also found Mutagenic, along with many other similar chemical compounds. important places to change - important places to remain the same

### b. Backup of information and Logging of changes and states.

Since errors can occur anywhere, and they often do, a mechanism for preserving the crucial parts of DNA could be encoded so as to ensure that those characters which are beneficial propagate unchanged. DNA could include backups of the genes that shouldn't be altered.

A simple way to do so is to duplicate the genes which are the most used, the most transcribed, since they are likely to be the most useful ones. The backup could be the result of a transposon's action, which would duplicate both the entire gene transcribed code, along with the intervening introns. Alternatively or simultaneously, only the transcribed code could also be backup up. The steps to achieve this process could be to randomly put back into the DNA code RNA sequences by reverse-transcription. This randomness will be more likely to keep backups of the most used gene sequences.

How these backups can be used is not an obvious matter. One can imagine that a partial match is made between strings of DNA and they are modified such that they all agree with what most of them agree on. Of course, this should not be done at every duplication, but instead at special points in evolution. Alternatively, checksums could be computed for different genes, so that smaller strings have to be compared every time instead of entire genes. This can be difficult if we use a hash function of the type we use in computer science, since any mutations that occurs would change the checksum code. Instead we need a function that varies little with every mutation such as for example a mod of the sum of the energy released during bindings of amino-acids.

Another possibility is that the exon-only parts of the gene are what the match is done on, the checksum itself. Since the introns are changing so frequently, matching on them would not yield correct results, and the checksum could be simply the exon part of the gene.



(Moreover, because of the tremendous parallel machinery that exists for DNA, having to match the entire length is not a problem, as long as the exons mutations are not that frequent.

Along with backing up data (either full gene or exons only), we can imagine a scheme where progressive changes are recorded in the DNA code, as a logging scheme. Since in the evolutionary progression not all steps and mutations are beneficial, a species can survive because of some beneficial changes, even though some errors have been introduced into its genome. A logging function could enable the DNA to go back to previous versions of a protein encoding when it cannot generate that protein anymore. This logging function can also be used to extract properties about the mutations that have occurred along the way to an evolutionary state, as it is described in the next section.

### **c. Extracting Gene Relations.**

We mentioned that an organism might like to back up interesting genes in its DNA. Determining what is interesting can be hard, since genetic cells don't necessarily know what is important to other parts of an organism's body. However, one could rely on the logging mechanism. When a mutation occurs, and when a gene is changed, or a new protein appears, some reinforcement can occur, where the organism learns that generating that type of protein can be beneficial. A mistake in what proteins are beneficial can be made, but would not persist by natural selection, so we don't have to worry too much about making mistakes at this point, since making correct assumptions is far more important. These reinforcement values therefore need not really exist, and the fact that something is written in the log could by itself be a reinforcement.

Now to determine what genes are important, one can just look back into the log file, and see what the last changes are. A goodness value could correspond to how recent a mutation is, or how important in the evolutionary process. If a mutation allowed for many other mutations to occur, then it has probably been beneficial (since it allows an organism to get out of a local maximum and explore more possibilities about the world around it, developing new characteristics. For example, the mutation that lead to the ability to breathe air opened up the ground to many other mutations that could occur once the organism had another world to explore.

A notion of time seems needed in determining which mutations have led to such new horizons to explore. A time stamp could be included along with each change in the log file (where time can be easily measured in number of mutations, by comparing two saved sequences where one is allowed to change in a constant fashion, while the other one is carefully checked for errors at every duplication).

Relations between genes can thus be determined by their proximity in the log. The fact that a certain factor was able to arise and perpetuate after a certain other had occurred can be interpreted as a codependency. This relation will be confirmed if a mutation reappears in the first factor after the second one has changed. What is learned by studying the log file is higher-level knowledge about evolving, what we can call heuristics of evolution. The more an organism has evolved, the more heuristics it will have developed, and the more apt it will be to further developing, without accidentally destroying useful characteristics. This explains the exponential nature of species evolution that was mentioned as one of the problems that Darwin failed to explain. Accumulating knowledge about evolving seems to be what can have led to such complex organisms and behaviors.

### **d. Referencing Related Genes without Pointers.**

Once such relations are established, genes could simply be transferred close to each other by a transposon, or they can be combined in subtle ways, so as to ensure that they evolve simultaneously. Examples for such genes would be upper arm structure and body weight, in the example of developing the ability to fly. If one mutates then the other will too, and if one doesn't change they other should not. If two genes which are related are too large to be moved, or if they are encoded in such a way that they cannot be moved, then they should reference each other in order to evolve simultaneously.

In the DNA, there is no need for pointers, nor tables of genes, which is the first thing that would spawn to one's mind for referencing genes. Because of the sporadic nature of chromosomes, pointing to them seems impossible. However, the advantage of DNA is the massively parallel processing that happens within a cell and within the nucleus, that makes such referencing unneeded. For searching a gene, an organism will simply look everywhere: millions of lookup molecules will try to match every gene of every chromosome. Bringing the matches together once they have been found is another story (irrelevant to the genetic code), but we can imagine some matcher molecule which binds to a lookup molecule when it has accomplished its goal, and then releases some substance which attracts other matchers.

To reference information from one part of DNA to the other, such a mechanism could be used. A part of DNA string could make reference to specific genes to guide their mutation rates and evolution in respect with the current gene. That is introns could be specifying not only the mutation rate of the current gene, but also favor or inhibit the mutation and expression rates of other related genes referenced therein.

### **e. Introns provide a working lab for protein invention.**

If a protein useful to the organism is fabricated from a particular gene, changing that gene will be deleterious to the individual whose genes will not propagate. It is unlikely that more than a few useful changes will occur simultaneously for the changed protein to be

suddenly functional, and more useful than the one it replaces. Therefore, for a new protein to develop, an organism needs a working space which does not directly affect its phenotype.

Such a working space is provided by the introns. Since they are not translated, drastic changes in them will not affect the survival of the individual, and of its descendants. In them, new codon combinations can be tried out, until a new working protein emerges from them, useful in the organism, and thus leading to the survival of the lineage which developed it without penalizing it along the way.

The mechanism for this process seems to already exist in the genetic system, specifically in the domain of introns. There are different mechanisms for duplicating entire genes and for moving them around, even across chromosomes. This paper does not claim that these mechanisms are directed into choosing a specific gene and copying it over (although we could envision a usage/usefulness factor associated with each gene and corresponding to what quantity of proteins is fabricated from it).

DNA sequences were found, which resembled the hemoglobin genes, but which were not transcribed or translated. These are called pseudogenes. They are similar but not identical to the standard globin genes, but carry within them a huge amount of divergence, which corresponds to millions of years.

## V. A Theory of Evolution

### a. DNA as a self-interpreted dynamic base of information.

The idea which has guided us throughout the paper is that within the DNA come the instructions on how to handle it. The interpreter for DNA code must be encoded within the DNA string itself, since there is no other way of propagating information from one generation to the next. The language specifications therefore evolve, just as the data carried evolves.

The beginning of a DNA string could code the formation of ribosomes, which then interpret the continuation of the string. This allows for a beautiful coevolution of language and interpreter. If we think of DNA in computer science terms, the data would correspond to exons, and the program to interpret it and use it would be encoded in the introns.

*But how is the program understood in the first place?* a critic will ask. In fact a paradox can arise: if introns determine the evolutionary rates of exons, what determines the rate at which introns change? That is, if evolutionary information encoded in introns determines how fast an organism will evolve, what determines how fast the organism will learn about evolving? We would need a higher level of abstraction to determine how introns would change and so on. This argument can lead to an infinite recursion of always higher levels of abstraction needed.

However, if we stick to the biological level of understanding, and we study how proteins are coded from DNA then RNA (or RNA alone in earlier organisms), we will realize that the first level of understanding is already available. It took billions of year, and trillions of generations, and countless millions of trials for the first amino-acid to be formed, even for the first carbonated molecule. How life first started is another question, but it seems that as soon as the process or a genetic language transmitted from one generation to the next was created, this language is able to evolve while improving itself. A simple process then becomes a more and more elaborate language of genetic transmission and evolution. Heuristics are added, higher levels of abstraction are reached, gene relations and mutation speeds are optimized. If one believes the arguments made earlier in the paper, it is clear that a higher-level language has in fact evolved and is used to direct evolution in meaningful ways.

### b. Tendencies instead of Characters.

We can suppose that when a complex organism mutates, instead of developing features, it develops tendencies. That is mutations occur in the intron level, instead of the exon level. That is changes in the program come first, and then guide changes in the data. Since the exon mutations are determined by the intron mutations, a change in the intron tendencies and goals will be reflected as many consecutive changes in the exon level.

An example of such a change could be in the intron determining the rate of mutations in the genes coding for the upper-arm structure of the animal. The introns would then encode a tendency towards changing the arm structure. This tendency will lead over a few generations to individuals with an arm structure increasingly different, but also prone to changing since the tendency is also transmitted from one generation to the next. If changing the arm structure is a desirable feature, the individuals will be more likely to survive, and the changed genes will persist, changing more and more every few generations. For example, such a tendency could have lead to increasingly long arms, which were useful in climbing trees and jumping, which would also favor light individuals who can climb more easily and higher. An ever changing structure could have lead to the actual flight capabilities. If however, the arm structure was optimal for the environment (where running for example is the main survival skill, and longer arms would mean heavier bodies and a less running-oriented equilibrium) the individuals carrying the change tendency will be likely to die by natural selection at the first generation, or at least very early, and the change in tendency will not be transmitted.

But when does the mutation rate slow down? It is clear that wings cannot grow indefinitely (in the case of the albatross for example, the wings are too large for the animal to land or walk gracefully, and albatross end up sleeping while flying). It is possible that an ancestor of albatross had this "change wings" tendency and kept growing larger wings. Every offspring with larger wings would die from inadequacy in hunting or landing or breeding. As generations went by, the intron specifying that wing mutations should be enhanced would eventually mutate again, lasting long enough for a stable formation to arise and establish itself.

This theory accounts for high bursts in evolution, where individuals do not change little by little, but instead when a new niche is exploited, species evolve more rapidly and drastically. This rapid changing cannot lead to inapt species, since they will be eliminated by natural selection. Moreover, when individuals change, they shape their changing environment to which they have to be ready to adapt. A high mutation rate during those changes can thus also account for the coevolution of changing individuals in a changing environment.

A lapse of natural selection can open the road to a wide variety of species, since those tendencies will not be eliminated early in their evolution, and are likely to stabilize in some advantageous future form. For example, after the death of dinosaurs, mammals were left with plenty of food and space, and were thus able to evolve rapidly and conquer many niches now left unfilled.

---

## Conclusion

It seems almost inevitable that more complex structures exist in the genome of species than the ones proposed by Darwin. A few operations were suggested for a possible language of evolution, which would enable a higher-level understanding of genetic data, which would not be simply transmitted blindly, but would evolve purposefully. A great deal has to be learned about existing mechanisms to lay the foundations on discovering new ones, and this paper is barely scratching the surface of ideas that could open up new horizons for a modern understanding of evolution.

---

## Acknowledgments

The ideas in this paper evolved from conversations with Jerome Lettvin, Patrick Winston and Randy Davis, from Marvin Minsky's Society of Mind, and from Douglas Lenat's Role of Heuristics in Learning by Discovery. Special thanks to Christina Carvey for lending me all of her biology books on genetics and molecular biology.

---

## References

- [Childs](#), Eric, et al., *Mutator Genes in E.Coli*, Carnegie Mellon University, 1996.
- Crow, James F., *Genetic Notes*, Eighth Edition, Macmillan Publishing Company, New York, 1983.
- Lenat, Douglas, *The Role of Heuristics in Learning by Discovery*, Machine Learning: An Artificial Intelligence Approach, Tioga Publishing Co., Palo Alto, 1983.
- Lewin, Benjamin, *Genes V*, Oxford University Press, 1994.
- Lewin, Roger, *Patterns in Evolution, The New Molecular View*, Scientific American Library, New York, 1996.
- MacIntyre, Ross J., *Molecular Evolutionary Genetics*, Plenum Press, New York, 1985.
- Maloy, Stanley R., *Microbial Genetics*, Second Edition, Boston, 1994.
- [Min](#), Bon-Hong, et al., *Identification of an Enhancer in the First Intron Involved in the Regulation of Mouse Vascular Smooth Muscle's Actin Gene*, Seoul 136-705, Korea, 1996.
- Minsky, Marvin, *The Society of Mind*, Touchstone Books, New York, 1985.
- Osawa, Syozo, *Evolution of the Genetic Code*, Oxford Science Publications, New
- Roff, Derek A., *Evolutionary Quantitative Genetics*, Boston, 1997
- Stryer, Lubert, *Biochemistry*, Fourth Edition, Stanford University, 1995. York, 1995