Non-adaptive evolution of genome complexity
Soojin V. Yi

Summary
Genome complexity is correlated with biological complexity. A recent paper by Michael Lynch proposes that evolution of complex genomic architecture was driven primarily by non-adaptive stochastic forces, rather than by adaptive evolution.\(^\text{1}\) A general negative relationship between selection efficiency and genome complexity provides a strong support for this hypothesis. The broad capacity of this theory is both its appeal and source for criticism. BioEssays 28:979–982, 2006. © 2006 Wiley Periodicals, Inc.

Introduction
Gene and genome structures vary dramatically across the tree of life. The difference between genomes of prokaryotes and eukaryotes is particularly striking. Prokaryotes generally boast compact genomes, consisting mostly of protein-coding nucleotides. In contrast, eukaryotes possess several classes of non-coding DNA such as intergenic regions, introns, untranslated regions (UTRs) and transposable elements. The number of genes and the amount of non-coding DNA vary tremendously within eukaryotes. As a result, genome size in eukaryotes varies by more than four orders of magnitude.\(^\text{2}\) As genome size increases, both the number of genes and the relative proportion of non-coding DNA in the genome increase. Generally speaking, genome size and biological complexity are correlated,\(^\text{3}\) both of them in an increasing order from prokaryotes, to unicellular eukaryotes, and to multicellular eukaryotes.\(^\text{1,4}\)

The distinction between the genomes of prokaryotes and eukaryotes has been considered such a fundamental divide that few studies have provided a general theory that can cross the barrier between them. Recently, Michael Lynch and his colleagues have published a series of papers that are firmly based on well-established population genetic theories to provide a new idea that can explain genome complexities of prokaryotes, unicellular and multicellular eukaryotes as a continuum that spans this divide.\(^\text{1,4,5}\) A central theme of this idea is that many characteristics of complex genomic structures have originated via non-adaptive, stochastic processes.

Stochastic accumulation of nearly neutral mutations
We will briefly review the population genetic theory underlying this idea. The fates of different alleles are determined by the interplay between two evolutionary forces, natural selection and genetic drift. Natural selection acts by changing allele frequencies according to their selective effects (designated as \(s\)) so that advantageous mutations can ultimately reach fixation, while deleterious mutations get eliminated from a population. Allele frequencies can also change due to random chance, which is referred to as genetic drift. The rate of genetic drift in a population is proportional to \(1/N_e\), where \(N_e\) is the effective population size.

To quantify the joint effect of natural selection and genetic drift on the fate of a new mutation, it is convenient to consider the fixation probability of a newly arising selected allele, relative to that of a neutral allele, which is given by \(\frac{4N_es}{1+C0}\) (Fig. 1).\(^\text{1,6,7}\) Therefore, the fate of a mutation is determined by its selective effect, relative to the effective population size to which it is introduced. For example, if a mutation with a slight selective disadvantage (\(s = -10^{-5}\)) arises in a population with \(N_e = 10^6\), it has practically no chance of reaching fixation. However, if a mutation with the same selective disadvantage occurs in a population with \(N_e = 30000\) (so that \(Nes\) is \(-0.3\)), the fixation probability is as much as 86% of that of a neutral allele. In other words, despite its selective disadvantage, this mutation behaves similarly to a neutral mutation. For mutations that satisfy \(|Nes| < 1\), genetic drift plays a substantial role in determining their fates; this class of mutations are referred to as `nearly neutral` mutations.\(^\text{6}\) Lynch proposes that mutations that satisfy \(|4Nes| < 1\), whose fixation probability is at least 88% of that of a neutral mutation, are effectively neutral.\(^\text{1}\) The key point here is that a mutation can be moderately deleterious in one population with large \(N_e\) and will be removed by natural selection, but could be nearly neutral in another population with small \(N_e\), thus reaching fixation via genetic drift (Fig. 2).
Reduced selection efficiency promotes evolution of genome complexity

Lynch\(^{(1,5)}\) and Lynch and Conery\(^{(4)}\) posits that mutations that have generated complex genomic architectures of multicellular eukaryotes fall into this 'effectively neutral' class of mutations. Genomes of species with small \(N_e\) provide population genetic environments in which nearly neutral mutations can passively accumulate, as a consequence of reduced selection efficiency. In a new paper, Lynch further details two crucial steps that advance the idea of stochastic accumulation of near neutral mutations to a general principle of genome evolution.\(^{(1)}\) First, he shows that there is a distinctive difference in the effective population sizes across the tree of life, so that genome complexity negatively correlates with the effective population size. Second, he discusses the molecular mechanisms that conferred slight selective disadvantages to the mutations that created introns, UTRs and modularized regulatory regions in early eukaryotes making these mutations susceptible to the effect of genetic drift.

Lynch first examines the distribution of the effective population sizes. \(N_e\) is commonly inferred from silent site diversity \((\pi_s)\), which is equivalent to \(4N_e\mu\) (where \(\mu\) is the neutral mutation rate).
Implications, criticisms and future prospects

On the one hand, the greatest appeal of this thesis is that it is a general theory, based upon solid population genetic principles. The idea that a general theory can explain the majority of genomic diversity without invoking cellular and physiological differences between different groups of organisms is greatly satisfying. On the other hand, the breadth itself raises a concern that such a broad idea may not be able to explain the astonishing amount of biological diversity.

An opposing view posits that the evolution of genome size is a highly complex process, and that determinants of genome size vary considerably among genomes, even between closely related organisms. Therefore, a general theory of genome size evolution is considered inadequate according to this view. Indeed, deviations from Lynch’s theory have been reported. For example, among mammals, the average genome size of herbivores (large $N_e$) is smaller than that of carnivores (small $N_e$), which is the opposite of what is predicted by Lynch’s theory. However, it should be noted that the above theory invokes a long-term stochastic process, and is best suited to explain a broad pattern. Significant deviations in specific examples are to be expected. In addition, it should be noted that, because $N_e$ of the contemporary populations, inferred from silent site diversity, measures genetic variation accumulated since the last coalescence of all individuals, it can be greatly affected by recent demographic histories. Comparisons between closely related species may not necessarily reflect long-term evolutionary patterns. It is notable that, in larger scales, the negative relationship between genome size and effective population size appears to be robust.

Yet another possibility is that the observed negative correlation between $N_e$ and genome size is a by-product of additional factors that control both of these variables. Genome size positively correlates with developmental rate. Developmental rate is in turn related to body size, which is negatively correlated with population size in a variety of eukaryotic species. Therefore, it is possible that a third factor is related to both genome size and population size, such as developmental rate or metabolic rate, is the cause for the observed negative correlation between genome size and the effective population size. Mutation rates are also known to be correlated with both body size and with silent site diversity (see above). Therefore, a careful evaluation of other factors is necessary to assess the relationship between genome size and effective population size.

A recent study re-evaluated the relationship between genome size and effective population size in ray-finned fishes, while controlling for body size and mutation rates. Genome size and the genetic variability were still strongly correlated after correcting for other factors, supporting the idea that the effective population size is the primary determinant of genome size, at least for this group of organisms.
Some theories propose that genome size itself is an adaptive trait. For example, increase in genome size was considered advantageous, because larger genomes may offer additional nucleotides that can serve as structural materials controlling cell volume.\(^{19}\) Even though this specific hypothesis has been challenged,\(^ {14}\) genome size itself can no doubt be a direct target of natural selection in different species: in particular, some parasitic eukaryotes have extremely reduced genomes, which is likely due to strong natural selection toward small genomes to suit their parasitic lifestyle.\(^ {20}\) A recent study proposes that the streamlined genomes of prokaryotes are the results of direct natural selection, resulting from the predator–prey relationship with earlier eukaryotes, who already possessed complex genomes.\(^ {21}\) Relative contributions of natural selection and genetic drift on genome evolution therefore remain to be resolved by future studies.

It should be clarified that Lynch’s theory does not deny the role of positive natural selection on genome evolution. On the contrary, adaptive evolution is an essential component of this theory. Once the slightly deleterious mutations (that increase genome complexity) have been permitted to proliferate via genetic drift, they provide a substrate for subsequent adaptive evolution. Returning to the example of introns, Lynch posits that once the introns and spliceosome were established, they can expand adaptively by providing substrates for the nonsense-mediated decay (NMD) pathway. The NMD pathway is an mRNA surveillance mechanism that searches for premature stop codons. For the NMD to work, genes would require the presence of at least one intron. The selective advantage associated with elimination of premature stop codons will then positively drive intron proliferation, until the number of introns and the efficiency of the NMD pathway reach an equilibrium. Similarly, once established, UTRs and regulatory modules can serve as substrates for secondary adaptive evolution, which will facilitate the proliferation of these non-coding elements, ultimately leading to the evolution of complex genomes found in multicellular eukaryotes.

In his concluding remarks, Lynch emphasizes that “the ideas presented [in this paper] are unlikely to be correct in every detail” yet the detailed hypotheses discussed in the paper are meant to provide a null model by which key elements of eukaryotic gene structure could evolve via non-adaptive processes. Lynch’s insight into the origins of complex genomic architecture in eukaryotes serves as both a provocative look into the evolutionary mechanisms that shape our genome and an appeal for evolutionary biologists to conduct further analyses into these fundamental questions.

**Acknowledgments**

SY acknowledges support from Georgia Tech and comments and discussions from Charles Warden, Todd Steelman, Eric Vigoda, James Thomas, and Michael Lynch.

**References**