Review of: "Non-Darwinian Molecular Biology"

Marco José[1]

1 Universidad Nacional Autónoma de México

Potential competing interests: The author(s) declared that no potential competing interests exist.

Comments on: Palazzo AF and Kejiou NS (2022) Non-Darwinian Molecular Biology.*Front. Genet.* 13:831068.

By Marco V. José

Theoretical Biology Laboratory, Instituto de Investigaciones Biomédicas, Universidad Nacional Autonóma de México, CP 04510, Ciudad de México

Completion of the modern evolutionary synthesis in the post-genomic era requires better understanding of the molecular processes of evolutionary change. Authors cite Sydney Brenner: “many in the field are drowning in a sea of data and starving for knowledge” (Brenner, 2003). About 3 decades ago, Sydney Brenner also said (Brenner 1991): “In one sense, everything in biology has already been published in the form of DNA sequences of genomes; but, of course, this is written in a language we do not yet understand. Indeed, I would assert that the prime task of biology is to learn and understand this language so that we could then compute organisms from their DNA sequences. ... We are at the dawn of proper theoretical biology.”

Authors present advances oriented to build a more modern conceptual framework for molecular biology that is non-Darwinian in outlook. They contend that cellular biochemistry is imperfect, by virtue of the chemical properties of macromolecules, and propose that genetic drift stymie their improvement and further exacerbates their messiness, especially in eukaryotes. Eukaryotes have evolved robust quality control systems (“global solutions”) which have a two-fold effect of ameliorating numerous entities that experience slightly deleterious mutations (“local problems”), while also permitting the proliferation of additional cellular sloppiness. In most organisms, transition mutations (e.g., G to A mutations) are more common than transversion mutations (e.g., G to C mutations). They propose a null hypothesis for adaptive evolution incorporating concepts of biochemical sloppiness, global solutions, and mutational bias into neutral evolution**.** An interesting proposal is that neutral evolutionary processes can increase the complexity of an organism which may in turn permit the evolution of novel adaptive traits. This is in part a consequence of the nearly neutral theory of molecular evolution, that the underlying mutation which allows enzyme promiscuity is not sufficiently deleterious to be selected against. As a result, enzymes accumulate slightly deleterious mutations and thereby protein activities and interactomes become even more messy and error-prone (Levy et al., 2009). Errors during transcription and translation, will produce proteins that differ from their canonical sequence that further contribute to biochemical sloppiness. Misincorporation of nucleotides is corrected by DNA repair enzymes and proof-reading machinery, both of which expend energy.

Given the many advantages of logic type control, one would expect that during ~ 3 billion years of evolution, where single cell life forms were likely the only forms of life, a biological logic type control system would have evolved to control at least

some biological processes. Sequential logic processes always require a clock to orchestrate the orderly processing of events (Penketh 2022).

The notion of global solutions is compatible with concepts as buffering and robustness. Global solutions are robust cellular mechanisms that maintain homeostasis, and they may have the potential to buffer not only environmental changes but also genetic changes. Authors use the example of chaperones, which not only prevent protein misfolding at high temperatures, but also promote the folding of proteins that have acquired destabilizing mutations. Other global solutions that increase the robustness of cells include RNA quality control mechanisms, such as non-sense mediated decay and the RNA export machinery, which retains mis-processed mRNAs and spurious transcripts in the nucleoplasm thus preventing their translation into potentially toxic proteins. If the deleteriousness of these suboptimal activities remains below the critical threshold required for purifying selection, they will not be eliminated and instead be subjected to evolution by drift (Lynch, 2007).

Another important global solution is the piwi-associated RNA (piRNA) system which represses transposon activity in the germline of most eukaryotes (Czech et al., 2018). Instead of eliminating every transposon, the piRNA system can be mobilized to suppress the deleteriousness of most transposable elements that have an RNA intermediate in their life cycle. Indeed, it is believed that the piRNA, small interference RNA (siRNA), and micro-RNA (miRNA) systems all evolved early in the evolution of eukaryotes to globally inhibit transposable elements and viruses (Shabalina and Koonin, 2008).

The ultimate effect of these global solutions/buffering systems is to simultaneously elevate the selection coefficient of numerous mutations that share a common problem. Global solutions reduce deleteriousness and promote evolvability.

Regarding the latter statement, a general statistical model has been recently developed which shows that macroevolutionary patterns and processes are consistent with Darwinian gradualism (Pagel et al. 2022). According to this model, directional changes do not alter evolvability. Evolvability changes occur at ancestral nodes and multiply the variance of a trait.

A key strength for the neutral theory of molecular evolution is that it provides molecular biologist with a null model for testing evolutionary hypotheses. Populations are perpetually evolving given the constant change in allele frequency and mutation rates. Therefore, it is important to understand whether these changes are due to stochastic versus adaptive processes. To invoke adaptive forces as the main driver of a molecular trait, one must demonstrate that the trait could not have arisen solely due to drift and biased mutations. In addition, organisms that have invested in global solutions that blunt the deleteriousness effects of messiness, will tolerate mutations that create a certain degree of messiness. This messiness will accumulate if it does not pose a burden above the drift barrier. To postulate adaptive forces as the main driver for the existence of some entity or some process, one must demonstrate that the trait is not simply a product of a messy organism.

Although the rate of evolutionary change is constant for a given protein-coding gene in different lineages, it has been documented that different protein-coding genes have different clock rates. Protein conservation in yeast species (i.e., how slow the clock runs for a given protein) correlates with its expression level (Pál et al., 2001). Highly essential enzymes evolve slowly (Aguilar-Rodríguez and Wagner, 2018). Adaptive models assume that most changes are due to positive selection. The null model is that most change is due to neutral, or slightly deleterious mutations. If we assume that most mutations are nearly neutral (the null hypothesis), why would these accumulate in lowly expressed genes? Global

solutions can blunt small changes but are not as effective against big changes. As such, highly expressed genes are under greater evolutionary constraint and are more optimized in terms of their structural stability and/or interactions than lowly expressed genes, which can more easily rely on global solutions to fold properly.

The neutral theory of molecular evolution has been used to develop a novel neutral amino acid substitution test (Zamudio et al, 2019). This neutrality test has been applied to unravel neutral mutations, positive and negative selection in the spike protein of SARS-CoV-2 (López-Cortés et al. 2021).

The complexity of a system is correlated with the number of parts and the number of*nonlinear* interactions between these parts. Authors point out that it has been naively assumed by some that biological complexity is a direct product of natural selection (Ekstig, 2015). Drift models do not explain one of the main dichotomies that we see in life on earth: that by most definitions, eukaryotes are more *complex* than prokaryotes.

Prokaryotes experience high levels of selection pressure, while eukaryotes do not (Lynch, 2007). In organisms where selection pressure is extreme, superfluous activity is wasteful and effectively eliminated by natural selection (Lynch, 2006). In organisms with a low effective population size, these extra features are not purged but instead are allowed to persist for extended evolutionary time and their deleterious effects are instead buffered by global solutions. Even within eukaryotes, lower effective population sizes correlate with longer and/or more numerous intergenic regions, introns, cryptic transcriptional start sites and other non-functional genomic entities (Lynch, 2007). Within this surplus of non-functional activity lies the raw substrates for the evolution of new components. A sloppy cellular environment full of junk contains many available substrates that can be tinkered with (Jacob, 1977; Jacob, 2001) and eventually exapted (Gould and Vrba, 1982) to form new functional parts. More recently, the detailed process of exaptation has been investigated, and surprisingly, the evolution of junk to functional entities often involves processes that do not rely on positive selection, but rather on neutral evolution.

Non-functional entities are inefficiently eliminated in organisms that evolve under weak selection regimes. In the case of junk RNA, their persistence allows them to explore sequence space over considerable evolutionary time, allowing them to potentially acquire additional activities, or what is known as “excess capacity” This is an example of the larger phenomenon, known as *constructive neutral evolution* (Stoltzfus, 1999; Stoltzfus, 2012). Note, that at no point was a new activity shaped by positive selection.

In contrast with this statement there are reports in which non-protein RNA-coding genes have undergone positive selection (Pollard et al. 2006). The hallmark of evolutionary shift of function is sudden change in a region of the genome that previously has been highly conserved owing to negative selection. It has been speculated that changes of this type in *FOXP21*, a gene involved in speech production, and *ASPM2*, which affects brain size, have had a significant role in the evolution of the human brain. The vast majority of the approximately 15 million changes in our genome since our common ancestor with the chimpanzee are likely to represent neutral drift 4,5, so systematic searches for potentially important evolutionary acceleration have focused exclusively on protein coding regions, where there is a more favorable signal-to-noise ratio.

Pollard et al. (2006) reported that the most dramatic of these ‘human accelerated regions’, HAR1, is part of a novel RNA gene (HAR1F) that is expressed specifically in Cajal–Retzius neurons in the developing human neocortex from 7 to 19 gestational weeks, a crucial period for cortical neuron specification and migration. HAR1F is co-expressed with reelin, a

product of Cajal–Retzius neurons that is of fundamental importance in specifying the six-layer structure of the human cortex. HAR1 and the other human accelerated regions provide new candidates in the search for uniquely human biology. The DNA should be recognized as a secure habitat for RNA molecules to preserve their genetic inheritance.

Alu elements are the most abundant repetitive elements in the human genome. Primate genomes have 10^6 copies. Tandemly arranged *Alu* RNA elements, are clearly correlated with the emergence of dimeric*Alu* elements during primate evolution. *Alu* elements play an important role in the regulation of gene expression at various levels, such as in alternative splicing when present in intronic regions of genes. *Alu* sequences in all *Homo sapiens´* chromosomes are closely associated to the structural condensing of DNA sequences in all nucleosomes (Sosa et al. 2013).

A quantitative unifying theory of modern evolutionary synthesis should consider several seminal works: 1. The finding of power-law correlations in DNA sequences which signals the presence of a high complexity, fractality, scale invariance in the heterogeneity of those sequences (Peng et al., 1992; Li and Kaneko, 1992) 2. The modern theory of dynamical systems explains how phenomena such as chaos, self-organization, synchronization, and complexity can emerge from nonlinear interactions within a system 3. The origin of universal allometric scaling laws in biology from genomes to ecosystems (West and Brown, 2005). There are universal scaling laws relating the metabolic rate, the rate of molecular evolution, and the molecular clocks that “tick” at a constant rate per unit of mass-specific metabolic energy flux rather than per unit of time (Gillooly et al. 2005). When combined with assumptions of the neutral theory (Kimura and Ohta, 1974), these allometric laws also can be used to characterize rates of molecular evolution. The first assumption is that molecular evolution is caused primarily by neutral mutations that randomly drift to fixation in a population, resulting in nucleotide substitutions (Kimura 1968). This assumption is consistent with theory and data demonstrating that deleterious mutations have only a negligible chance of becoming fixed in a population because of purifying selection, and that favorable mutation occur very rarely (Kimura 1968). Under this assumption, the rate of nucleotide substitution per generation is equal to the neutral mutation rate per generation and is independent of population size (Kimura and Ohta, 1974; Gillooly et al. 2005). On average, a certain quantity of metabolic energy transformation within a given mass of tissue causes a substitution in each gene regardless of body size, temperature, or taxon. Scaling theory predicts a 100,000-fold increase in substitution rates across the biological size range (g of whales to g of microbes) and a 34-fold increase in substitution rates across the biological temperature range () (Gillooly et al. 2005). 4. Critical scale invariance has been found from a primeval RNY genetic code to the modern Standard Genetic Code (SGC) in Eubacteria and Archaea (José et al. 2009). It was shown that the same scaling and statistical properties of protein coding genes have remain unaltered during the evolution of the SGC, using a renormalization group approach (José et al. 2009). The near universal SGC could be in part a result of the need species has to communicate using a common genetic language. The SGC must obey ecological needs for prey-predator interactions, food webs, host-parasite interactions, and so on.

If our cells are teeming with non-optimal reactions, mistake-riddled molecules, and non-adaptive processes, each being slightly deleterious, but not enough to be eliminated by natural selection, how do cells, and by extension multicellular organisms, manage to survive? This situation is exacerbated if the genome is constantly absorbing slightly deleterious mutations that increase messiness. If the deleteriousness of these suboptimal activities remains below the critical threshold required for purging selection, they will not be eliminated and instead be subjected to evolution by drift (Lynch, 2007; Sung et al., 2012). However, probably the biggest obstacle, is the selective maintenance of genes that encode the

additional machinery. Examples of this type of machinery are DNA repair enzymes and domains in polymerases responsible for proof-reading activity. The benefit of these genes, in terms of selection, must clear the drift barrier for natural selection to maintain them. It is for this very reason that the DNA mutation rate is strongly influenced by the effective population size (Lynch et al., 2016). Cryptic transcription factor binding events or exotic RNA species can exist while serving no benefit (Palazzo and Gregory, 2014; Palazzo and Lee, 2015). Even tissue-specific expression will be subjected to unique noise given that each cell type expresses a unique set of transcription factors that activate distinct cryptic transcription start sites scattered throughout the genome (Levy et al., 2009; Palazzo and Lee, 2015).
It has been shown that recombination is the major determinant of the evolution of (G+C)-content and therefore the selectionist models of isochore evolution have been dismissed. Different lines of evidence have suggested that (G+C)-content might be a consequence of the process of recombination (Duret and Arndt, 2008). Notably, analysis of the patterns of substitution in primate non-coding sequences have shown that recombination affects the relative rates of AT → GC and GC → AT substitutions (Galtier et. 2001; Galtier et al. 2007; Galtier et al. 2009). This effect might result from the *neutral* process of biased gene conversion (BGC) (Galtier et. 2001; Galtier et al. 2007; Galtier et al. 2009). According to this model, gene conversion (i.e., the copy/paste during meiotic recombination of one allele onto the other one at heterogozygous loci) is biased in favor of GC-alleles, which leads to an increase of probability of fixation of GC-alleles compared to AT-alleles). Thus, BGC should lead to an enrichment in (G+C)-content in genomic regions of high recombination compared to regions of low recombination. BGC may be one Achilles´ hill of the human genome (L. Duret personal communication).

In this article, concepts from molecular evolution, and the use of null models, have been used to understand the evolution of a wide array of entities and biochemical processes. In all these studies, many of the observed phenomena have features that are consistent with the null hypothesis—that they are largely shaped by neutral evolution. When these produce undesirable products that are nevertheless present in organisms, their effects are small enough that they do not clear the drift barrier and are largely blunted by global solutions. The proposal elaborated by the authors is based on observations and data. Terms like global, messiness, sloppiness, protein promiscuity, slightly deleterious, excess capacity, must be quantified possibly by a mathematical model.
Quoting Sean Eddy (Eddy, 2013): “The complexity of biology — its lack of unifying theory, its wealth of fascinating and crucial detail — is such that big experiments are nearly nonexistent in our field.”
“Everything is simple and neat— except, of course, the world” (Goldenfeld and Kadanoff, 1999)

References

Aguilar-Rodríguez J., Wagner, A. (2018). Metabolic determinants of enzyme evolution in a genome-scale bacterial metabolic network. *Genome Biol. Evol.* 10: 3076–3088. doi:10.1093/gbe/evy234.

Brenner, S. (2003). NOBEL LECTURE: Nature’s Gift to Science. *Biosci. Rep.* 23, 225–237. doi:10.1023/B:BIRE.0000019186.48208.f3

Brenner, S. (1991). Summary and concluding remarks. In Evolution of life (eds S. Osawa, T. Honjo). Berlin, Germany: Springer.

Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E. L., Fabry, M. H., Kneuss, E., et al. (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing.*Annu. Rev. Genet.* 52, 131–157. doi:10.1146/annurev-genet-120417-031441

Duret, L., and Arndt, P. F. (2008). The impact of recombination on nucleotide substitutions in the human genome, *PLoS Genetics*, **4**(5):1-19 (2008).

Eddy, S. R. (2013). The ENCODE Project: Missteps Overshadowing a success. Curr. Biol. 23, R259–R261. doi:10.1016/j.cub.2013.03.023

Ekstig, B. (2015). Complexity, Natural Selection and the Evolution of Life and Humans. Found. Sci. 20, 175–187. doi:10.1007/s10699-014-9358-y

Galtier, N., Piganeau, G., Mouchiroud, D., Duret L., (2001). GC-content evolution in mammalian genomes: The biased gene conversion hypothesis,*Genetics* 159, 907911 (2001).

Galtier, N., L. Duret L. (2007). Adaptation or biased gene conversion. Extending the null hypothesis of molecular evolution, *Trends Genet.* 23, 273-277.

Galtier, N., Duret, L., Glémin, S., Ranwez, V. (2009). GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates, *Trends Genet.* 25, 1-5 (2009).

Gillooly, J. F., Allen A.P., West G. B. Brown J. H. (2005). The rate of DNA evolution: Effects of body size and temperature on the molecular clock.*Proc. Natl. Acad. Sci. U.S.A.* 102, 140-145.

Goldenfeld, N., and Woese C. Biology's next revolution.*Nature* 2007, 445.

Goldenfeld, N., and Kadanoff, L.P. Simple lessons from complexity.*Science* 1999, 284: 87-89.

Gould, S. J., and Vrba, E. S. (1982). Exaptation-a Missing Term in the Science of Form.*Paleobiology* 8, 4–15. doi:10.1017/S0094837300004310

José, M. V., Govezensky T., García J. A., Bobadilla J. R. (2009). On the evolution of the standard genetic code: vestiges of scale invariance from the RNA World in current prokaryote genomes.*PLoS ONE* 4(2): e4340. doi:10.1371/journal.pone.0004340.

Jacob, F. (2001). Complexity and Tinkering. *Ann. N. Y. Acad. Sci.* 929, 71–73. doi:10.1111/j.1749-6632.2001.tb05708.x

Jacob, F. (1977). Evolution and Tinkering. *Science* 196, 1161–1166. doi:10.1126/science.860134

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. Nature 217, 624–626. doi:10.1038/217624a0

Kimura, M., and Ohta, T. (1974). On Some Principles Governing Molecular Evolution.*Proc. Natl. Acad. Sci.* 71, 2848–2852. doi:10.1073/pnas.71.7.2848

Levy, E. D., Landry, C. R., and Michnick, S. W. (2009). How Perfect Can Protein Interactomes Be.*Sci. Signal.* 2, pe11. doi:10.1126/scisignal.260pe11

Li, W., 1991. Expansion-modification systems. A model for spatial 1/f spectra.*Phys. Rev. A* 43, 5240–5260.

Li, W., Kaneko, K., 1992. Long-range correlations and partial 1/fa spectrum in a noncoding DNA sequence. *Europhys. Lett.* 17, 555–660.

López-Cortés G.I., Palacios-Pérez M., Zamudio G. S., Veledíaz H. F., Ortega E., José, M. V. (2021). Neutral evolution test of the spike protein of SARS-CoV-2 and its implications in the binding to

ACE2. *Sci. Rep.* 22;11(1):18847. doi: 10.1038/s41598-021-96950-z.

Pál, C., Papp, B., and Hurst, L. D. (2001). Highly Expressed Genes in Yeast Evolve Slowly. Genetics 158, 927–931.

Pagel M., O'Donovan C., Meade A. (2022). General statistical model shows that macroevolutionary patterns and processes are consistent with Darwinian gradualism. *Nature Commun.* 13, 1113. https://doi.org/10.1038/s41467-022-28595-z

Peng, C.-K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Sciortino, F. Simons, M., Stanley, H. E., (1992). Long-range correlations in nucleotide sequences. *Nature* 356, 168–170.

Peng, C.-K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L., 1994. Mosaic organization of DNA nucleotides. *Phys. Rev. E* 49, 1685–1689.

Penketh, P. G. (2022) Is DNA repair controlled by a biological logic circuit?*Theory in Biosciences* 141:41–47. https://doi.org/10.1007/s12064-021-00360-8

Pollard K. S. et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443, 167-172. doi:10.1038/nature05113.

Shabalina, S. A., and Koonin, E. V. (2008). Origins and Evolution of Eukaryotic RNA Interference. Trends Ecol. Evol. 23, 578–587. doi:10.1016/j.tree.2008.06.005

Sosa, D., Miramontes, P., Li, W., Mireles, V., Bobadilla, J. R., José, M. V. (2013). Periodic distribution of a putative nucleosome positioning motif in human, non-human primates, and Archaea: Mutual information analysis. *International Journal of Genomics.* http://dx.doi.org/10.1155/2013/963956

Stoltzfus, A. (1999). On the Possibility of Constructive Neutral Evolution.*J. Mol. Evol.* 49, 169–181. doi:10.1007/pl00006540

Stoltzfus, A. (2012). Constructive Neutral Evolution: Exploring Evolutionary Theory's Curious Disconnect. *Biol. Direct.* 7, 35. doi:10.1186/1745-6150-7-35

West, G. B., and Brown, J. H. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. The*Journal of Experimental Biology* 208, 1575-1592 Published by The Company of Biologists 2005. doi:10.1242/jeb.01589

Woese, C. R. A New Biology for a New Century. *Microbiol. Mol. Biol. Rev.* 2004, 68(2):173. DOI: 10.1128/MMBR.68.2.173-186.2004.

Zamudio G. S., Prosdocimi F., Farias S. T., José M. V. (2019). A neutral evolution test derived from a theoretical amino acid substitution model. *J. Theor. Biol.*. 467:31-38. https://doi.org/10.1016/j.jtbi.2019.01.027

Acknowledgments

MVJ was financially supported by Dirección General de Asuntos del Personal Académico (DGAPA), Universidad Nacional Autónoma de México, PAPIIT-IN201019.

Conflict of Interest: The author declares no conflict of interest.