

Does nothing in evolution make sense except in the light of population genetics?

Michael Lynch: Origins of Genome Architecture, Sinauer Associates, Sunderland Mass, 2007, 340 pp, hardback, ISBN-10: 0878934847

Lindell Bromham

Received: 13 October 2008 / Accepted: 21 October 2008 / Published online: 21 November 2008
© Springer Science+Business Media B.V. 2008

Abstract “*The Origins of Genome Architecture*” by Michael Lynch (2007) may not immediately sound like a book that someone interested in the philosophy of biology would grab off the shelf. But there are three important reasons why you should read this book. Firstly, if you want to understand biological evolution, you should have at least a passing familiarity with evolutionary change at the level of the genome. This is not to say that everyone interested in evolution should be a geneticist or a bioinformatician, but that a working knowledge of genetic change is an essential part of the intellectual toolkit of modern evolutionary biology, even if your primary focus is the evolution of behaviour or the diversity of communities. Secondly, this book provides excellent examples of another important tool in the biologist’s intellectual toolkit, but one that is rarely explained or illustrated to such an extent: null (or neutral) models. The role null models play in testing hypotheses in evolution is a central focus of this book. Thirdly, as an accomplished work of advocacy for a strictly microevolutionary view of evolution, this book provides grist for the mill for the important debate about whether population genetic processes are the *sine qua non* of evolutionary explanations.

Keywords Adaptation · Complexity · DNA · Gene duplication · Modularity · Neutral evolution · Null models · Population size

Introduction

By focussing on evolution of genomes, “*The Origins of Genome Architecture*” by Michael Lynch untangles the association between size, complexity and adaptation.

L. Bromham (✉)
Centre for Macroevolution and Macroecology, School of Biology, Australian National University,
Canberra, ACT 0200, Australia
e-mail: lindell.bromham@anu.edu.au

Despite the official line that we do not believe in evolutionary progress, biologists have always had a tendency to arrange organisms in an orthogenetic series: beginning with the oldest, smallest and simplest and ending with the newest, largest and most complex. To convince yourself of this, pick up an introductory biology textbook: there is a fair chance that the chapters will be arranged in a taxonomic series that starts with bacteria, progresses through plants, and ends with mammals. Even the evolution chapters will repeat this “Great Chain of Being”, relating the history of life as a series of stages leading to the arrival of humanity on the planet, as if this was the point that the rest of the biosphere had been leading up to. Consciously or unconsciously, we place humans at the top of the evolutionary tree and the apex of the ecological pyramid. And there has been an unspoken assumption that studies of the human genome would likewise put us at the top end of the scale. What a shock it has been to find that our genomes do not rank first in any measure of size or complexity.

The first fall from grace was genome size. It was discovered in the 1970s that genome size and organismal complexity did not seem to evolve hand in hand. Humans, for example, have smaller genomes than some cockroaches, ferns and amoebae. “Depending on one’s point of view, the puzzle was either solved or deepened as it became clear that a substantial fraction of many eukaryotic genomes consists of non-coding and putatively non-functional DNA” (Lynch 2007, p. 32). Thus humans rested their damaged egos on the assumption that, even if their genomes were not the biggest, they would undoubtedly contain more genes. However, upon gloriously entering the post-genomic era, humans found themselves faced with the embarrassing reality that it requires fewer genes to build their magnificent selves than it does to make a pufferfish or a rice plant. Even the humble nematode *Caenorhaditis elegans*, an animal studied because of the relative developmental simplicity required to build a body of only around a thousand cells, has nearly as many genes as the apparently more complex “higher” animals and plants.

If perishingly little of the human genome is taken up by genes, what is the rest? Some of the non-gene DNA is associated with genes. In complex genomes, protein-coding genes are divided into exons (sequences that code for an amino acid chain) and introns (non-amino-acid-coding sequences which must be excised from the gene transcript before it can be translated into a protein). In addition, some of the DNA on either side of gene sequences seems likely to be involved in gene regulation, since much of it is curiously well conserved between different species, indicating the action of selection maintaining important sequences. But even when all of the gene-associated DNA is added together—exons, introns and regulatory elements—it still forms a smaller proportion of the genome than the apparently functionless stretches of DNA between genes. This intergenic DNA contains a lot of repetitive sequences, which range from runs of the same few bases over and over (microsatellites) to multiple copies of virus-like genomes (mobile elements and endogenous retroviruses). Various adaptive explanations have been offered for excess DNA, but Lynch’s explanation for bulky genomes rests on a consideration of the role that population size plays in determining the relative power of natural selection and genetic drift.

Insertions that increase the amount of DNA in the genome apparently occur at a greater rate than deletions that reduce the amount of DNA. So what is stopping the

genome from expanding indefinitely? There must be a selective force stabilising genome size: in other words, there must be some kind of cost to having more DNA than you need. Lynch maintains that this cost is not the metabolic cost of making and keeping excess DNA, but the damage that excess DNA can do when it mutates. Any given nucleotide is vulnerable to mutation that will change its identity. While mutations provide essential raw material for evolution, very few mutations are actually of benefit to an organism. Most are harmless at best, or disastrous at worst. So the more DNA you have, the more chance you have for harmful mutations to occur.

The mutational burden of excess DNA is most obvious for functional DNA. A random mutation to a functional sequence, such as a gene or a regulatory element, could result in the catastrophic loss of an essential product. So, the more genetic elements it takes to build an organism, the more opportunities for damage to result in the loss of an essential component or service. “Each embellishment of the structure of a gene or its surrounding area increases the risk that the gene will be rendered defective by subsequent mutational processes” (p. 40). This is a bit like the Buddhist philosophy that the more possessions you have, the more potential for grief when you lose them. But Lynch also emphasises the cost of gain-of-function mutations: changes to apparently functionless DNA can make it code for something. This can accidentally introduce a new and destructive element to an otherwise well-balanced organism. In this way, the mutational burden extends to apparently functionless DNA, and therefore the risk of harm to the genome will increase with all forms of excess DNA. For example, a random change to a functionless DNA sequence could create a regulatory element that could switch on neighbouring genes inappropriately, potentially causing a gene expression mess. Since regulatory sequences are fairly short, and the amount of junk DNA is very large, this could be expected to happen by chance fairly often. Yet rogue regulatory elements are rarer than would be expected by chance in junk DNA, an observation that Lynch considers support for the existence of active selective against such mistakes.

If any excess DNA is costly, and provides no obvious benefits to the organism, why does it accumulate? Lynch’s thesis rests on three key points (neatly summarised on p. 40): (1), population size is the key determinant of the effectiveness of selection, because it determines the relative effect of genetic drift (random sampling) on the frequency of genetic variants in a population, and thus the level of disruption to the action of natural selection on allele frequencies; (2) genetic drift does not simply add noise to evolutionary change, it defines the kind of evolution that can proceed; (3) the accumulation of excess DNA depends on the population size (which determines the effectiveness of selection in removing excess DNA) and the mutation rate (which shapes the cost of excess DNA in terms of mutational burden). This argument rests on the wonderful simplicity of the underlying mechanics of molecular evolution.

Population size determines the kind of evolution that can occur

The triumvirate of biologists who orchestrated the merger of Darwinism and Mendelism—R. A. Fisher, J. B. S Haldane, and Sewall Wright—demonstrated the

effectiveness of selection for traits with even very small relative benefits or costs. The population genetic framework vindicated Darwin's faith in the power of natural selection to drive evolution, generation by generation, by the gradual accumulation of randomly generated heritable changes of small phenotypic effect. However, when the advent of new technology allowed scientists to actually estimate the amount of protein diversity in populations, they found that there was far more genetic variation in natural populations than expected if selection was the driving force of population genetics. This empirical challenge led Motoo Kimura and others to construct the neutral model for molecular evolution, showing that the excess variation could be explained if most mutations were selectively neutral, so had no effect on organismal fitness (see Kimura 1983). Neutral alleles would not be affected by selection, so their fate would depend on random processes, referred to as "genetic drift", because the frequency of a neutral allele would simply drift up and down until it was either fixed or eliminated by chance.

Substitution occurs when all members of the population carry a copy of the same mutation (allele) at a particular locus. The rate of substitution of neutral alleles is not affected by population size. But population size is an important factor in determining the substitution rate of non-neutral alleles. Large populations will have greater rates of substitution of advantageous alleles: with more genomes in the population, there is more chance a lucky advantageous mutation will arise, and selection is rigorously efficient at promoting alleles with even small selective advantages. Deleterious mutations will also be efficiently removed from large populations, as even a slight reproductive disadvantage will result in an inexorable decline in allele frequency. Allele frequencies are also subject to random effects; for example, a carrier of an advantageous allele that increases offspring survival may get buried in an avalanche, or a carrier of a deleterious allele that reduces the efficiency of foraging may have the good fortune to find a cache of high-energy food. In a large population, these chance events have little impact on the overall allele frequencies. But in a small population, allele frequencies are much more vulnerable to random fluctuations. The importance of population size in determining the tempo and mode of evolution (not just at the genetic but also at the phenotypic level) is underappreciated, and it is this deficit that "Origins of Genome Architecture" sets out to address.

The classic population genetic models generally assumed infinite (or at least very large) population sizes. But, as demonstrated by Tomoko Ohta (e.g., Ohta 1973), population size plays a critical role in determining the relative power of selection and drift in determining the fate of an important class of mutants. Strongly deleterious mutations, that dramatically reduce the chance of survival or reproduction, will disappear from the population because their carriers will reproduce less than others. But the fate of slightly deleterious mutations, for example, one that results in a minor change to an enzyme that makes it slightly less efficient, depends on the population size. In a large population, even a slight reduction in fitness will usually result in the mutation declining in frequency until it disappears from the population. But in a small population, the random sampling of alleles that occurs every generation has a proportionally greater effect on allele frequencies, such that it is possible for a slightly deleterious mutation to increase

in frequency by chance, and even go to fixation in the population. This means that in a small population, a larger fraction of the mutations will be governed by drift, not selection. In other words, slightly deleterious mutations will behave as if neutral in small populations, because their fate will be governed by drift and not selection, and thus a proportion of these undesirable mutations will be fixed by chance. So the proportion of alleles that are undergoing neutral evolution is strongly dependent on population size.

This is an extremely important result for evolutionary biology, because it tells us that the fate of a particular mutation (whether it will be lost from the population, fixed, or left to fluctuate as a polymorphism) depends not only on the properties of the mutation itself (whether advantageous, deleterious or neutral) but also on the properties of the population in which it arises. Deleterious mutations that would be removed by selection in a large population may go to fixation in a small population by chance. Advantageous mutations that would be fixed by selection in a large population have a higher risk of being lost by accident in a small population. This view of molecular evolution turns the focus on mutation as an important creative force in evolution, because recurrent mutations have a chance of going to fixation in small populations even if they are selectively disadvantageous.

Lynch places much emphasis on the relationship between organism size, population size, and the power of natural selection. Effective population size varies widely between organisms, influenced by many factors, including not only the census population, but also mating strategies, population structure and selection on linked alleles. But, in general, it is a fair prediction that larger-bodied organisms will have smaller populations than small-bodied organisms. At the extremes, most multicellular species have effective population sizes that are orders of magnitude smaller than most unicellular species. This means that a disadvantageous mutation that would be almost certainly eliminated by selection in a large population of unicellular organisms may be essentially immune to selection in a small population of large-bodied organisms. So rather than being eliminated from the population, this disadvantageous allele may find itself going to fixation by drift in a small population, despite the cost to fitness.

So the dynamics of molecular evolution depend very much on the type of organism you are looking at: "... Various forms of cellular life experience radically different population genetic environments—many forms of mutant alleles that are able to drift to fixation in multi-cellular eukaryotes are efficiently eliminated by selection in prokaryotic species. Without the theoretical and empirical results of population genetics, we would be unable to make such statements." (p. 96). Consideration of population size leads to a radical conclusion concerning the evolution of genomic architectural complexity. It is commonly felt that the evolution of complexity is something of an upward trajectory—complexity increases because complex organisms are, in some way, better than their simpler peers, so they are favoured by selection. "Where, then, is the direct supporting evidence for the assumption that complexity is entirely rooted in adaptive processes? No existing observations support such a claim, and, given the massive global dominance of unicellular species over multicellular eukaryotes, in terms of both species richness and numbers of individuals, if there is an advantage of

organismal complexity, one can only marvel at the inability of natural selection to promote it.” (p. 378). Lynch turns the selection-for-complexity argument on its head. Genome complexity is disadvantageous: fancier genomes have more targets for mutation, and mutation, on the whole, wrecks organisms. Genomic complexity evolves because organisms with small population sizes do not have the selective power to stop it. This nicely kicks the Great Chain of Being in the arse: complexity evolved not by selection for better organisms, but due to the failure of natural selection to maintain simplicity.

Lynch’s hard-nosed population genetic approach to genome evolution provides a counterfoil to the tendency in genomics for patterns of molecular evolution to be interpreted as the signs of adaptive processes. The great evolutionary biologist John Maynard Smith said that he had never met a birdwatcher who was not a naïve adaptationist (see Kohn 2004). Perhaps the same could be said for many bioinformaticians and genomicists, who tend to seek adaptive reasons for the complex patterns they observe in the genome. Is excess DNA a device for plumping out large cells (e.g., Gregory 2001)? Or is it a useless (and possibly harmful) burden that the cell carries to no benefit? Is alternative splicing a cunning mechanism for generating useful protein diversity (e.g., Harden 2008)? Or is it simply a sloppy mechanism that produces a lot of useless transcripts. Are mobile elements genomic viruses running amok, wrecking genes, disrupting expression and creating a metabolic and mutational burden, or are they the genome’s little helpers, soldiers of evolvability called upon in times of stress (e.g., Caporale 2003; McClintock 1983)? Adaptive explanations are fun to generate, but sometimes hard to test. One approach to testing adaptive explanation is to explore neutral models, in order to ask whether the trait you are interested in could have arisen by a non-adaptive process.

The importance of null models

The null model approach to evaluating selective explanations is playing an increasingly important role in many aspects of ecology and evolution (e.g., Hubbell 2001). A null model is used to ask whether the pattern you wish to explain could arise without the mechanism you wish to invoke operating. Lynch provides a nice example of applying null models to rates of morphological evolution. It is commonly assumed that rates of morphological evolution in mammals have been very fast and that this fast rate must be the signature of strong directional selection. Lynch tests this idea by using estimates of the rate of generation of morphological novelty in mice to derive a null model for the rate of neutral morphological change (that is, the possible rate of phenotypic change that could occur if all such changes made no difference to fitness). Since the observed rates of morphological change in skeletal characteristics are less than or equal to this calculated neutral rate, Lynch concludes that we do not need to invoke strong continuous selection to explain rates of morphological evolution in mammals. This is not to say that directional evolution has not operated on morphological evolution, merely that we are not necessarily required to invoke selection to explain our general observations on the rate of phenotypic evolution, because the same rate could be achieved without positive selection.

The null model approach is an essential tool in modern biology, yet it is rarely taught in a systematic way to biology students. Perhaps this is because null models are often formulated and explained in terms of statistics, and many biology students are resolutely statistics-averse. But null models do not have to be framed in terms of probabilities or algebra. Null models are essentially a rephrasing of Occam's Razor, which, whatever William of Occam's original statement or intention was, is commonly interpreted to mean that one should not leap to a fancy-pants explanation for a phenomenon when a simpler explanation will account for your observations, but that you should reject the simpler explanation as soon as it is shown to be unable to explain the data you have. Showing that a neutral model could produce the pattern of interest is not proof that it did so, but it shows that the neutral model must be held in mind as a possible explanation. A major advantage of neutral models is that we may find that we do not have to invoke an unknown selective force that we can only postulate and not observe.

To use a non-genetic example that is familiar to me, consider the hypothesis that middle-sized mammal species in Australia have a higher risk of extinction than smaller or larger mammal species (Burbidge and McKenzie 1989). Australia has suffered a depressingly high rate of terrestrial mammal extinctions in the past two centuries of European settlement: over half of the recent global mammal extinctions have been in Australia (Cardillo and Bromham 2001). It has been noted that the majority of these lost species are of intermediate size (defined as falling between 35 g and 5.5 kg in weight). This has been dubbed the "critical weight range", and many explanations have been offered for why middle-sized mammals should have been particularly badly hit: for example, they may be just the right size prey for introduced foxes and cats, and thus selected against in modern Australia. The observation that most of Australia's recently extinct terrestrial mammals fall within the critical weight range is undeniable. But a biological pattern does not always require a selective explanation. In this case, we must consider the fact that over half of Australian terrestrial mammal species, extinct or not, are middle-sized, so it is hardly surprising that the majority of extinctions have come from that group. So to test the hypothesis that species in the "critical weight range" are selected against, we first need a null model that tells us how many extinctions we would expect in the middle-size category if all species were equally likely to go extinct, regardless of size. When you simulate non-biased extinction by making a random draw of species from the Australian terrestrial mammal fauna, you find that most of them fall within the critical weight range. The null model does not disprove that middle-sized mammals have faced specific threats (c.f. Chisholm and Taylor 2007), but it does suggest that we should not automatically assume that this pattern requires a selective explanation, because it might not reflect selective extinction, but rather that middle-sized mammal species are more numerous than small or large ones (Cardillo and Bromham 2001).

Lynch rightly points out that while most biologists are comfortable with natural selection, many lack basic knowledge of neutral processes. We make sure our undergraduate students can apply selective thinking to a range of situations, but very few of them appreciate the role of random processes in evolution. Even many professional biologists rarely stop to consider neutral explanations for their data,

let alone test them. Neutral models are explained in academic books (e.g., Gotelli and Graves 1996; Hubbell 2001), but rarely if ever make it into the popular scientific press. Perhaps neutral models are considered less entrancing than selective stories. But neutral models have a beauty of their own, which arises from the ability to explain complex patterns with simple processes. More importantly, we have an incomplete understanding of evolution without considering neutral processes.

Of course, framing exactly the right null model is far from simple. Debate over whether a particular null model is asking the right question can result in many analyses and reanalyses. There are often multiple possible neutral explanations for a phenomenon. One neutral process that should be incorporated into any evolutionary model is phylogenetic inertia. When many organisms share a particular feature, we need to consider whether they do so simply because they all inherited it from a common ancestor rather than because they each have a “good reason” for having it. For example, there are many features of genome architecture that distinguish eukaryotes from prokaryotes—such as introns, bulky genomes, lack of operons—and Lynch explains each of these as arising from the reduction in population size. But an even “nuller” model requires no mechanism other than the weight of history, if the ancestral eukaryote had these features and its descendants, for whatever reason, did not lose them. I doubt that phylogenetic inertia over such long time periods is sufficient to explain the differences in genomic architecture between prokaryotes and eukaryotes, but ideally this null model should be formally considered and rejected before moving onto the fancier population genetic null models. So we have layers of null models: in order to call upon the selection-based model of genome architecture, you should first consider and reject the neutral explanation, but to call upon the neutral model, you should first consider and reject phylogenetic inertia.

Despite an inordinate fondness for neutral explanations, Lynch is not an inveterate nay-sayer. He accepts some patterns as the signature of an adaptive process, due to rejection of the neutral explanation, such as some of the functional variants produced by alternative splicing. He also accepts that the amount of non-coding DNA with some kind of conserved function (presumably in gene regulation) is twice that of the amount of coding DNA. But, on the whole, “*The Origins of Genome Architecture*” provides neutral models for the evolution of most major features of the eukaryotic genome, from centromeres to gene families to gene expression patterns. Producing a null model that fits observation does not prove the null model true. It does not even necessarily make the null model plausible. However, it does an important job in showing that the feature being explained could arise without selection operating, and therefore that selection might not necessarily have to be called upon to explain even quite complex features of the genome.

Neutral explanations for genome complexity

How can genomic complexity arise from a neutral process? In small populations, if mutation is biased to producing particular kinds of changes, then mutation becomes a “creative” force, giving rise to distinct patterns without the need to invoke

selection. Some of the mutational asymmetries arise as a by-product of the DNA replication processes, or as a result of variation in rates of transcription (i.e., how often a gene is expressed). There may even be metabolic costs involved in different kinds of mutation, as some nucleotides are “cheaper” to produce than others. Other patterns across chromosomes are more mysterious. For example, one of the most recognizable features of eukaryotic chromosomes is the centromere, the region at which two sister chromatids are joined, and from which point the two chromatids will be dragged apart at cell division. Clearly, centromeres have an important functional role: chromosomes without a functional centromere tend to get lost at cell division, producing unviable daughter cells. Yet it seems that very little of the centromere is actually involved in this function. In fact, despite a lot of effort, no one has succeeded in identifying specific sequences associated with centromere function in any species apart from yeast. Instead, most of the centromere is made up of large numbers of repetitive sequences, typically the corpses of transposable elements of one kind or another. So Lynch considers that one of the most notable features of eukaryotic genome organisation—the centromeres—may be a product of a mutational bias.

Mobile DNA provides a potent non-adaptive force for shaping the genome. Barbara McClintock discovered mobile elements in the 1950s, when her close scrutiny of maize chromosomes led to the heretical observation that some “jumping genes” could change their genomic location. There are several different classes of mobile elements in the genome, and they differ in their length, coding capacity and means of replication. Some mobile elements bear a close resemblance to the genomes of “free-living” viruses, with genes coding for proteins required for copying the viral genome, and even envelope genes that can make viral bodies. But the most common type of mobile elements in the human genome are LINE elements. “LINE” stands for Long Interspersed Elements (unlike fully-fledged organisms, which get proper names, viruses and mobile elements tend to be given pronounceable acronyms). LINES contain genes that encode proteins needed for their own replication: a RNA-binding protein that binds to viral mRNA transcripts, reverse transcriptase that is capable of making a DNA copy of the mRNA transcript, and an endonuclease that can cut the host DNA in the middle of a strand to allow the viral genome to be inserted. These genetic tools allow LINES to insert their DNA into the host genome, and persuade the host transcription machinery to make RNA copies of the LINE genome. These copies can then be reinserted elsewhere in the genome, thus increasing the number of LINE copies in the host genome. SINES (Short Interspersed Elements) replicate in the same way except that they do not contain genes for making their own proteins, so they rely on taking advantage of the reverse transcriptase and endonuclease produced by LINES (in this way SINES parasitize parasitic DNA, much like Swift’s smaller fleas that prey upon larger fleas: Bromham 2002). Transposons are simpler still, being short sequences bounded by repeated sequences that can bind together to loop the sequence and allow it to be cut out of the genome and then inserted elsewhere.

Mobile elements can create havoc in the genome. If your genome is full of mobile elements which hop about inserting randomly, then sooner or later one will land in a gene. In all probability, the insertion of a sodding great bit of DNA within a gene is

likely to destroy its ability to make a functional gene product. If the organism needs that gene product, then it's curtains for them. In addition, since mobile elements are built to ensure their own replication, many contain strong regulatory signals to attract the host transcription machinery. They may land upstream of a host gene and start screaming "Transcribe me! Transcribe me!", unwittingly turning on any adjacent genes. It is possible that this is how some retroviruses cause cancer (such as leukaemia viruses: see Bromham 2002). And, even at their most benign, mobile elements fill the genome with bits of DNA its owner does not need. Nonetheless, some researchers have suggested that mobile elements are good for you. McClintock noted that "jumping genes" were most active in a plant under stress, and concluded that they may be a deliberate mechanism for generating variation in "evolve or die" situations (see McClintock 1983). This is similar to the interpretation of the error-prone SOS mutation repair pathway in bacteria as an adaptive strategy for throwing up potentially useful mutations (e.g., Echols 1981; Hersh et al. 2004), rather than simply a disaster-prone last-ditch attempt to stay alive. Some scientists also consider that mobile elements have an important role in generating adaptations in the genome (e.g., Caporale 2003; Villareal 1997). Certainly the human genome has gained some useful bits of kit from the viral genomes that inhabit it. There are examples of viral genes or regulatory elements being co-opted into human development. The human placenta, for example, could not form without a gene derived from a retroviral envelope gene (Mi et al. 2000).

But the genome must pay a high cost for the presence of mobile elements. Can the occasional win really be maintained at the cost of frequent losses? If mobile element activity was an adaptation for generating variability in the host genome, then we would expect to see host-level adaptations that promote mobile element activity. No such adaptations have been identified. Instead, Lynch provides evidence that mobile element activity tends to occur in "boom and bust" cycles. Each element seems to go through an initial period of expansion, then reduction in replication rates, and eventually extinction from the genome. Most elements in the genome are therefore quite young, despite the fact that mobile elements have been entering the genome for tens of millions of years. It seems that relatively few mobile elements reach a stable equilibrium where they can maintain modest numbers in the genome without expanding to the detriment of the host's (and therefore the element's) chances of reproduction. This is not the pattern you would expect to see if mobile elements were retained by their hosts as an evolvability strategy. Far from being on the side of the host, Lynch maintains that it is theoretically possible for mobile element expansion to result in extinction of the host through "mutational meltdown", which he says could potentially result in the higher background extinction rates in larger animals. Experimental evidence suggests it is indeed possible for mobile elements to expand at substantial fitness cost to their host (Charlesworth and Langley 1989).

Is the genome designed for modularity or is it just a mess?

Even those who are comfortable with the construction of chromosome architecture by mutational bias, and consider that mobile elements are genomic parasites that do

not have our best interests at heart, may balk at considering gene duplication and regulation as a neutral process. Because duplicate genes can provide raw material for developing new traits, gene duplication is frequently regarded as an engine of evolvability. The implication is that one of the reasons we have so many duplicate genes is that selection has promoted the process of gene duplication itself. Our genomes are full of families of genes that have arisen by duplication then diverged in function. In some cases the duplication and diversification process has increased phenotypic complexity in a clearly advantageous manner. For example, the diversification of genes coding for G-protein-coupled receptors has led to a wide variety of proteins involved in taste, vision, smell and behaviour (e.g., Shi et al. 2003). In other cases, gene duplication has produced related genes that vary in expression pattern in time or space, though it may not be entirely clear whether this diversity is adaptive or not (e.g., the clusters of globin genes that make subunits of the blood protein haemoglobin: see Aguileta et al. 2006). But we may simply be witnessing the success stories of gene duplication, which, like advantageous mutations, may be a relatively rare phenomenon.

Lynch argues that duplicate genes are constantly being produced, but that most of them are quickly lost by mutational silencing, and thus the bulk of gene duplication is essentially a non-adaptive process: “a significant fraction of genes is nonessential and simply a reflection of stochastic expansion and contraction processes” (p. 45). Of course, some of these duplicate genes may well end up being extremely useful, providing raw material for adaptive evolution. But, like insertional mutation due to mobile elements, the occasional win from gene duplication might not negate an overall pattern of loss. This neutral explanation is counter to the tendency to attribute selective significance to any noticeable difference in gene number, function or expression patterns. Gene families with many members are often interpreted to represent a functional diversification associated with special adaptations of that lineage. But Lynch points out that the number of genes per gene family follows a pattern that could be explained by a stochastic process: most genes have no close relatives, many have few relatives, and few have many relatives. This null model for the distribution of gene family sizes suggest that we will always be able to find some gene-rich gene families in a genome, whether or not those families represent special adaptations.

Duplicate genes cannot simply be maintained as “spares”, back-up copies that can maintain function if the other copy is damaged by mutation. If a duplicate gene is truly excess to requirements, then there is no cost to losing it, so spare copies of genes are expected to accumulate mutations at the neutral rate, even if these mutations disable the gene. Instead, duplicates can generally only be maintained if they make unique and irreplaceable contributions to fitness, either through subfunctionalization (each copy loses a different part of the function of the original gene through deleterious mutation, making both copies essential for full function) or neofunctionalization (one or both copies develop new roles through advantageous mutation). Lynch’s argument for the primacy of subfunctionalisation is, again, typical of his approach. Unlike neofunctionalization, which is driven by positive selection, subfunctionalisation is initiated by the neutral processes of duplication and mutation. Under this scenario, duplication fills the genome with gene copies,

mutation knocks out their functions, but, by chance, duplicates may lose different sub-functions, making both copies necessary for continuing function. These sub-functionalised duplicates may then be subject to selection to further hone their specialised roles. The end result is that the genomes of large organisms get messier and more complex, requiring multiple genes to do the work previously accomplished by a single gene.

Lynch suggests that this process of duplication and subfunctionalisation may lead to speciation of isolated populations. If a single ancestral population is divided into two isolated populations, these populations will independently accumulate copies of different genes through random duplication of parts of the genome. These gene copies can diverge when mutations inactivate different functions in the duplicates, making it essential to keep both copies. By and by, the populations maintain original functions using different sets of genes. Mixing the two genomes together is unlikely to result in a full set of functions, some of which are now performed by different genes in different species. (Incidentally, a similar argument is put forward for the contribution of nuclearisation of organelle genes to speciation. The organelles that originated from endosymbiotic bacteria, such as mitochondria and chloroplasts, began with fully functional genomes, but over time many of their genes migrated to the host nuclear genome. Lynch proposes that the gradual movement of organelle genes to the nucleus could form an isolating mechanism between lineages, as the gene complements of the nuclear and organelle genomes must be compatible between potential parents in order to produce a whole functioning organism.)

Thus Lynch provides a neutral argument for gene duplication, a process that has been considered a hallmark of the selective advantage of modularity. He dishes out a similar treatment to another aspect of genome architecture commonly interpreted as an adaptation for modularity: the presence of non-coding sequences (introns) interspersed between the coding sequences (exons) in eukaryotic genes: “one of the most perplexing observations ever made in molecular genetics”. The average human gene contains over thirty times as much non-coding DNA as protein-coding sequences. These non-coding introns must be removed from the gene transcript, using complex splicing machinery, before it is translated into a protein. While it necessitates much more equipment and processing to produce a gene transcript, the piecemeal construction of genes has been interpreted as an adaptation for modular construction of proteins. By facilitating the variable construction of transcripts so to include or exclude particular exons, alternative splicing can allow different proteins to be produced from a single gene. Alternative splicing has been the cause of much excitement in the genomics literature in the past decade. But, while there are beautiful examples of mix-and-match proteins being generated through exon-shuffling, it is possible that in the majority of cases intron splicing result in non-adaptive variance between transcripts. Mistakes in splicing, for example, due to a mutation in a splice site that makes it unrecognizable to the splicesosomal machinery, can result in errant transcripts or non-functional proteins. Splicing errors are not trivial: “a third of human genetic disorders are attributable to mutations that cause defective splice-site recognition” (p. 238).

So even if there are benefits to having introns within genes (for example, as targets for recombination or mechanisms of alternative splicing), introns must place

a high burden on genome function. Not only do they inflate the genome, the presence of introns necessitates investment in machinery that removes them from gene transcripts before translation into gene products. Furthermore, these introns become mutational targets: random changes to the introns could ruin the gene product by altering splicing. Lynch's explanation for the vast numbers of introns in many eukaryote genes should, by now, be familiar: rather than having arisen for their contribution to generating novel proteins, introns colonised the genomes of those organisms lacking the population genetic clout to stop them. "For newly arisen introns having no functional significance for the products of their host genes, the primary force opposing their ability to spread throughout a population is their excess mutation rate to defective alleles, and because this force is expected to be quite weak, selection will be ineffective in preventing intron colonisation in populations experiencing substantial levels of genetic drift" (p. 254).

The primacy of population genetic thinking

The motto of Lynch's research group— "Nothing in evolutionary biology makes sense except in the light of population genetics"—is a paraphrase of the great Theodosius Dobzhansky's most quotable quote (Dobzhansky 1973). Few sensible biologists would deny that the population genetic framework is essential to evolutionary biology, and that it is too often ignored by people constructing evolutionary stories for the origin and maintenance of traits. But, in the provocative final chapter (which you should read even if you read none other in the book), Lynch takes his belief in the power of population genetic models a step further. He argues that the current population genetic framework that forms the basis of the Neodarwinian synthesis is all we need to understand evolution. Furthermore, Lynch considers that this framework is an essential tool for evolutionary biologists and that evolutionary explanations are unconvincing if they do not make use of this population genetic framework.

"As we move into the next phase of evolutionary biology, we can be confident of two things: the basic theoretical machinery for understanding the evolutionary process is well established, and we will soon be effectively unlimited by the availability of information at the DNA level" (p. 364). In other words, whatever exciting discoveries may arise from new data, they will be revealed using the same population genetic tools we have been using for nearly a century. Such statements can be read as a declaration of war against those who proffer evolutionary explanations without being aware of the underlying assumptions they are making about the way that new traits arise in individuals then become fixed in populations. It has always been frustrating to evolutionary biologists that workers from all fields of science (and beyond) feel qualified to put forward hypotheses in evolutionary biology, as if there were nothing more to being an expert on evolution than having learnt the phrase "survival of the fittest". "Evolutionary biologists have thought quite a lot about evolution, and individuals from outside the field who claim to have solved a major evolutionary enigma might want to consider why their ideas have not

previously come to the forefront. Have such ideas been ignored, or have they faded into the background because their feasibility is known to be marginal?" (p 372).

This tension is most evident in Lynch's annoyance with the vague but enthusiastic claims arising from the growing field of evolutionary development (evo-devo) that the Neodarwinian synthesis is inadequate (without demonstrating where it fails) and that they will furnish a new and better theory (without having done so yet). Evo-devo provides an example of the kind of evolutionary biology that Lynch does not like, because it is a field almost entirely devoid of population genetics. This does not simply mean that evo-devo papers do not contain equations. More deeply, evo-devo hypotheses generally do not even include a discussion of how variations that appear in individuals could rise in frequency in the population by drift or selection until they replaced the wild-type. Developmental biologists, on the whole, give no consideration to the theoretical framework that suggests mutations of large effect can rarely go to fixation, nor to the interaction between population size and selective co-efficient in determining the fate of mutations. Evo-devo jumps from the mutation in an individual to the substitution in a lineage with little formal consideration of the process in-between.

Does the failure to explore the population genetic mechanisms underlying the evolution of developmental or other traits invalidate hypotheses in evo-devo? No trait can skip the population genetic stage of evolution: any new mutation arising in a single individual must go through a stage of population polymorphism then subsequent fixation, either by selection or chance, before it can become a standard feature of a species. But it does not follow that no biologist can skip the population genetic stage if they wish to explain the origin of a trait. In an ideal world, we could furnish a complete explanation of the evolution of any trait, from its mutational origins to its developmental and phenotypic expression, its interaction with the environment and its rise or fall in frequency the population by selection or drift, followed by a complete history of the species bearing that trait as they compete with other species, respond to changing environment, speciate and go extinct. But most biologists operate at a specific part of this explanatory story (Calcott 2009). A behavioural biologist may be uninterested in the specific mutation underlying a behavioural change, as long as they can still measure the consequences of having that trait. An evolutionary biologist may ask whether a particular trait, such as sexual dimorphism, increases the speciation rate of a lineage without knowing what population genetic process brought the trait into being in the first place.

Clearly, investigating or modelling the substitution of developmental novelties within populations would put evo-devo on a much stronger footing, but the data required to test such hypotheses may be quite difficult to obtain. But this seems to be what Lynch requires: "[Mendelian genetics] has provided a solid foundation for the development of a mechanistic framework for understanding evolutionary processes with a level of mathematical rigor that has few rivals in the life sciences. Indeed, the general principles of populations genetics are now so well established that any proposed scenario for genomic evolution must remain in doubt until it has survived this theoretical gauntlet" (p. 67). It is not clear whether Lynch is simply asking us to bear microevolutionary processes in mind when considering the plausibility of evolutionary explanations, or whether he is demanding that each biologist provide a

formal population genetic model for the evolution of their trait of interest. At very least, he does require population genetic analyses before he will take explanations for the origins of genomic features seriously. For example, regarding the hypothesis that excess DNA in the genome performs a structural role in maintaining cell volume, he says that “the logic underlying the bulk-DNA hypothesis will remain unconvincing until it is demonstrated that: (1) heritable within-population variation in genome size significantly covaries with cellular features that are mechanistically associated with individual fitness and (2) mobile element proliferation is an easy means of achieving such variation with minimal negative side effects” (p. 34).

This first requirement—that evidence be furnished that heritable within-population variation is present and covaries with fitness—is setting the bar pretty high for making an adaptive argument convincing. Not only is this kind of data hard to obtain in most cases, this approach would be difficult to apply to features that originated under one selective regime are now maintained by another, for even if there is no such variation or no such microevolutionary advantage in current populations, it is possible that it operated in the past. Yet surely we can continue to conduct valuable evolutionary studies without measures of population variability, heritability and fitness. Although we only have a dozen *Archaeopteryx* specimens, and even fewer of most of the dino-birds, we can still use these to study the evolution of flight, even in the absence of information about the variation within these species in traits associated with flying ability. To argue that all biologists should be able to furnish a population genetic model for the origin of the trait they are interested promotes explanatory reductionism.

Lynch is right to say that those who prophesy the coming of new biological paradigms that will eclipse Neodarwinian explanations (e.g., Arthur 2000; Carroll 2000) must carry the burden of proof, and that it will take more than sweeping statements about the perceived inadequacy of microevolutionary explanations to do so. But to require a population genetic approach to all questions in evolutionary biology seems too stringent a criterion for determining what is a convincing hypothesis. Such an approach would deny the value of a macroevolutionary approach that compares patterns of evolution over time, between lineages, or across space. The majority of biologists that study macroevolution do so with the assumption that the microevolutionary processes that can be described by the population genetic framework underlie all macroevolutionary phenomena. But we often cannot get the data to test this assumption, and some questions we wish to ask may be best pursued by other means.

Conclusion

The Origins of Genome Architecture has several important messages, and it has the evidence and analysis to give these messages weight: a good evolutionary biologist should be able to construct and consider neutral explanations, appreciate the population genetic machinery that has served evolutionary biology so well, and be aware of the process of substitution that all new mutations must go through before becoming a fixed feature of a species. Whether or not you find yourself agreeing

with Lynch's neutral models for the origins of genome features, his arguments are put clearly and forcefully, so provide good fuel for further debate. This is a challenging book, not a "dummies' guide" to population genetics, but it should be accessible to anyone with a basic biology background. It provides a thorough introduction to population genetic thinking without getting bogged down in statistics (there are equations and graphs, but not at a level that will alienate the non-mathematical reader, who can elect to "hum through the equations" yet still get the main message). But this is not just a book of theory: it is illustrated throughout with observations from real data, which demonstrate that the beauty, diversity and complexity of the natural world extends to the genomic level, even if most nature fanciers are unaware of it.

An appreciation of the intricacies of genome structure, maintenance and expression is just as capable of filling the admiring biologist with "awe before the mystery of life" (Darwin 1876) as is the diversity of a rainforest. There is a lot of interesting molecular natural history in this book, such as the back-of-the-envelope calculation that the total amount of DNA in the world would, if unravelled and laid end to end, be around 10^{25} km, enough to wrap around the known universe many times over. Diversity at the genomic level is as rich as that at the phenotypic level. In this book I learned that while the mitochondrial genome of most organisms is a neat circle, in the *Amoebidium parasiticum* it is made up of hundreds of tiny fragments each containing only one or two genes, and in a particular kinetoplastid the mitochondrial genome consists of thousands of tiny circles, the DNA of which is gibberish without hefty post-translational editing (Burger et al. 2003). Why? who knows. It's wonderful all the same. Just as observations of animal behaviour or community structure or structural anatomy deepen our understanding of biology, so does an appreciation of the most basic level at which evolution occurs: the level of the genome. The message of this book, at its simplest, is that your genome is a glorious mess, an inefficient tangle of DNA. Yet, from this unholy muddle arises biological beauty and complexity.

Acknowledgements Many thanks to Brett Calcott, Kim Sterelny, Meg Woolfit and Rob Lanfear for helpful comments and discussions.

References

- Aguileta G, Bielawski JP, Yang Z (2006) Evolutionary rate variation among vertebrate β -globin genes: implications for dating gene family duplication events. *Gene* 380:21–29. doi:[10.1016/j.gene.2006.04.019](https://doi.org/10.1016/j.gene.2006.04.019)
- Arthur W (2000) The concept of developmental reprogramming and the quest for an inclusive theory of evolutionary mechanisms. *Evol Dev* 2:49–57. doi:[10.1046/j.1525-142x.2000.00028.x](https://doi.org/10.1046/j.1525-142x.2000.00028.x)
- Bromham L (2002) The human zoo: endogenous retroviruses in the human genome. *Trends Ecol Evol* 17:91–97. doi:[10.1016/S0169-5347\(01\)02394-1](https://doi.org/10.1016/S0169-5347(01)02394-1)
- Burbidge AA, McKenzie NL (1989) Patterns in the modern decline of Western Australia's vertebrate fauna: causes and conservation implications. *Biol Conserv* 50:143–198. doi:[10.1016/0006-3207\(89\)90009-8](https://doi.org/10.1016/0006-3207(89)90009-8)
- Burger G, Gray MW, Lang BF (2003) Mitochondrial genomes: anything goes. *Trends Genet* 19:709–716. doi:[10.1016/j.tig.2003.10.012](https://doi.org/10.1016/j.tig.2003.10.012)

- Calcott B (2009) Lineage explanations: explaining how biological mechanisms change. *Br J Philos Sci* (in press)
- Caporale LH (2003) Darwin in the genome: molecular strategies in biological evolution. McGraw-Hill, New York
- Cardillo M, Bromham L (2001) Body size and risk of extinction in Australian mammals. *Conserv Biol* 15:1435–1440. doi:[10.1046/j.1523-1739.2001.00286.x](https://doi.org/10.1046/j.1523-1739.2001.00286.x)
- Carroll RC (2000) Towards a new evolutionary synthesis. *Trends Ecol Evol* 15:27–32. doi:[10.1016/S0169-5347\(99\)01743-7](https://doi.org/10.1016/S0169-5347(99)01743-7)
- Charlesworth B, Langley CH (1989) The population-genetics of drosophila transposable elements. *Annu Rev Genet* 23:251–287. doi:[10.1146/annurev.ge.23.120189.001343](https://doi.org/10.1146/annurev.ge.23.120189.001343)
- Chisholm R, Taylor R (2007) Null-hypothesis significance testing and the critical weight range for Australian mammals. *Conserv Biol* 21:1641–1645
- Darwin CR (1876) The effects of cross and self fertilisation in the vegetable kingdom. John Murray, London
- Dobzhansky T (1973) Nothing in biology makes sense except in the light of evolution. *Am Biol Teach* 35:125–129
- Echols H (1981) SOS functions, cancer and inducible evolution. *Cell* 25:1–2. doi:[10.1016/0092-8674\(81\)90223-3](https://doi.org/10.1016/0092-8674(81)90223-3)
- Gotelli NJ, Graves GR (1996) Null models in ecology. Smithsonian Institution Press, Washington
- Gregory T (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc* 76:65–101. doi:[10.1017/S1464793100005595](https://doi.org/10.1017/S1464793100005595)
- Harden JH (2008) Quantitative and evolutionary biology of alternative splicing: how changing the mix of alternative transcripts affects phenotypic plasticity and reaction norms. *Heredity* 100:111–120. doi:[10.1038/sj.hdy.6800904](https://doi.org/10.1038/sj.hdy.6800904)
- Hersh MA, Ponder RG, Hastings PJ, Rosenberg SM (2004) Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. *Res Microbiol* 155:352–359. doi:[10.1016/j.resmic.2004.01.020](https://doi.org/10.1016/j.resmic.2004.01.020)
- Hubbell S (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, New Jersey
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
- Kohn M (2004) A reason for everything: natural selection and the English imagination. Faber & Faber, London
- Lynch M (2007) The origins of genome architecture. Sinauer Associates, Sunderland Mass
- McClintock B (1983) The significance of responses of the genome to challenge. In: Federoff N, Botstein D (eds) Nobel prize lecture: reprinted in The dynamic genome: Barbara McClintock's ideas in the century of genetics. 1992. Cold Spring Laboratory Press, New York
- Mi S, Lee X et al (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789. doi:[10.1038/35001608](https://doi.org/10.1038/35001608)
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98
- Shi P, Zhang J, Yang H, Zhang Y-P (2003) Adaptive diversification of bitter taste receptor genes in mammalian evolution. *Mol Biol Evol* 20:805–814. doi:[10.1093/molbev/msg083](https://doi.org/10.1093/molbev/msg083)
- Villareal LP (1997) On viruses, sex and motherhood. *J Virol* 71:859–865