Studies in History and Philosophy of Biological and Biomedical Sciences 58 (2016) 107–116

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

# Contingency, convergence and hyper-astronomical numbers in biological evolution

Ard A. Louis

*Rudolph Peierls Centre for Theoretical Physics, Univeristy of Oxford, 1 Keble Road, Ox1 3NP, United Kingdom*

ARTICLE INFO

*Article history:*
Available online 8 February 2016

*Keywords:*
Genotype–phenotype maps
Bias in evolution
Arrival of variation
Extended evolutionary synthesis

ABSTRACT

Counterfactual questions such as "what would happen if you re-run the tape of life?" turn on the nature of the landscape of biological possibilities. Since the number of potential sequences that store genetic information grows exponentially with length, genetic possibility spaces can be so unimaginably vast that commentators frequently reach of hyper-astronomical metaphors that compare their size to that of the universe. Re-run the tape of life and the likelihood of encountering the same sequences in such hyper-astronomically large spaces is infinitesimally small, suggesting that evolutionary outcomes are highly contingent. On the other hand, the wide-spread occurrence of evolutionary convergence implies that similar phenotypes can be found again with relative ease. How can this be? Part of the solution to this conundrum must lie in the manner that genotypes map to phenotypes. By studying simple genotype–phenotype maps, where the counterfactual space of all possible phenotypes can be enumerated, it is shown that strong bias in the arrival of variation may explain why certain phenotypes are (repeatedly) observed in nature, while others never appear. This biased variation provides a non-selective cause for certain types of convergence. It illustrates how the role of randomness and contingency may differ significantly between genetic and phenotype spaces.





## 1. Introduction

In Charles Darwin's theory of evolution, random variation provides the raw material for natural selection (Darwin (1989)). Variation proposes and natural selection disposes. But where does the variation come from? The modern evolutionary synthesis (Huxley, 1942) incorporated discrete Mendelian genetics into evolutionary theory, showing that the variation arises from mutations and recombinations that change and rearrange genes, leading to different phenotypes. Early reflections on the exponentially vast numbers of potential gene combinations suggested that variation is plentiful and probably isotropic, so that natural selection is the primary cause of evolutionary change. The size of these hyperspaces also naturally suggests that evolution is contingent because life can only explore a vanishingly small fraction of all genetic possibilities. On the other hand, the increasing evidence for widespread evolutionary convergence (Conway Morris, 2003, 2015; McGhee, 2011) suggests that nature can find the same solutions again and again. If the number of genetic possibilities is so vast, how can history appear to repeat itself?

It is important to remember that random mutations happen at the level of genotypes, while selection happens at the level of phenotypes which describe the different characteristics of organisms. Thus a full understanding of how evolution progresses needs to include a description of how genotypes map onto phenotypes. In this contribution I explore some recent theoretical studies of genotype to phenotype maps that may shed new light on the role of randomness and contingency in evolution. Even though mutations are fundamentally random with respect to outcomes, these studies suggest that there may nevertheless be strong bias in the kind of

*E-mail address:* ard.louis@physics.ox.ac.uk.

http://dx.doi.org/10.1016/j.shpsc.2015.12.014
1369-8486/

variation that arises. Certain types of variation are much more likely to arise than others, which may influence evolutionary outcomes.

But before proceeding, it may be helpful to reflect on the fundamental reason why the space of possible genotypes, or for that matter almost any combinatorial problem, grows so rapidly.

### 1.1. The power of exponentiating

Perhaps the earliest known reference to the power of exponential growth comes from an Old Babylonian tablet which dates from 1800–1750 BC. Soubeyran (1984). It describes the doubling of the number of barley corns for thirty consecutive days, ending up with no less than two "thousand," seven "hundred" and thirty-seven talents, half a mina, two and one-third shekels, and four barley-corns, which comes to about 47 metric tons of grain.[1] Such doubling problems were a popular subset of the so-called Silk Road problems (Friberg, 2005). They often took the form of 30 doublings, as in the tablet above, or else 64 doublings. A famous example of the latter comes from the epic poem *Shah-nama* (The Book of Kings) written by the Persian poet Firdausi around the turn of first millennium. It tells of the mythical Indian inventor of chess. Apparently the king was so pleased that he told the sage he could name his reward. The sage then asked to be given one barley corn for the first square of his board, two on the next, four on the next, and so on. While this at first seems like a very modest request, it would have totaled $2^{64}-1 = 18{,}446{,}744{,}073{,}709{,}551{,}615$ grains, weighing about 1,000 times the world's current annual wheat production. The shift from 30 to 64 in these silk-road doubling problems illustrates how quickly exponential growth leads to unearthly large numbers.

In the next section, I will explore how the power of exponential growth may have influenced some of the founding fathers of the neo-Darwinian Modern Synthesis (MS).

### 1.2. Hyper-astronomical numbers in the modern synthesis

Although Gregor Mendel and Charles Darwin were contemporaries, Darwin remained unaware of the far-reaching implications of Mendel's experiments in genetics for his theory of evolution by natural selection.[2] When Mendelian genetics was rediscovered in the late nineteenth and early twentieth centuries, it appeared to raise a significant objection to biometric formulations of Darwinian evolution: How can changes in discrete genes lead to the small continuous changes in phenotypic variation that Darwinian natural selection was said to act on?[3] This apparent conflict was solved between 1918 and the early 1930s by a triumvirate of great mathematical biologists—R. A. Fisher, J. B. S. Haldane, and Sewall Wright—who showed that if the traits of an organism are affected by many genetic loci, then, by the laws of statistics, many separate discrete changes will translate into effectively continuous variation, as required.

This early work on population genetics helped give birth to the fully fledged MS, also known as the neo-Darwinian synthesis. The term was coined by Julian Huxley with his book *Evolution, The Modern Synthesis* (1942) and other major figures in the movement included the aforementioned trio of Fisher, Haldane, and Wright, as well as the paleontologist George Gaylord Simpson, the ecologist E. B. Ford, the geneticist Theodosius Dobzhansky, and the evolutionary generalist Ernst Mayr who did much to cement the way the history of the MS is recounted.[4]

An important early set of arguments that fed into the MS can be derived from the geometry of discrete genetic spaces. Once you have many genes, it is natural to ask how many ways you can arrange them. Such thought experiments quickly lead to exponential growth and comparisons to the size of our universe, as can be seen in the following piece from Sewall Wright's hugely influential 1932 paper:

> "*Estimates of the total number of genes in the cells of higher organisms range from 1000 up…. With 10 allelomorphs in each of 1000 loci, the number of possible combinations is* $10^{1000}$ *which is a very large number. It has been estimated that the total number of electrons and protons in the whole visible universe is much less than* $10^{100}$." (Wright, 1932, pp. 356)

Stuart Kauffman (1995, pp. 167) has described such numbers as *hyper-astronomical* because they are beyond even the kinds of gigantic numbers that are used in astronomy. Of course such large numbers also easily obtain in all kinds of other combinatorial problems in physical sciences and engineering.[5]

Reflections on the hyper-astronomical size of these spaces may have influenced the further development of the MS. Here is Wright again in the same article: "*The population is thus confined to an infinitesimal portion of the field of possible gene combinations*" (Wright, 1932, pp. 356). These spaces are so (exponentially) vast that even over 3.8 billion years, life won't explore much more than a tiny fraction of all possible genetic combinations. Thus it naturally follows nature can only explore an unimaginably small fraction of all theoretically possible genomes.

What is not known, of course, is what fraction of that theoretical space of all genotypes can generate viable phenotypes. Given the fact that the majority of mutations appear to be deleterious, perhaps it is only an extremely small fraction. The rest of the space may be biologically sterile, "… *the howling wildernesses of the maladaptive, the 99.9% recurring of biological space where things don't work, the Empty Quarters of biological non-existence*." (Conway Morris (2003, pp. 309))

In Fig. 1 three schematics of the way the viable genotypes could be distributed are depicted. In each case only a small fraction of genotypes is viable. This immediately raises the question: how did life ever find the first viable genotypes? Given that we have evidence of life just a few hundred million years after the end of the violent "late heavy bombardment" of meteorites that characterized the birth of our planet, it would seem that life can find this viable region fairly quickly. With regard to Fig. 1, such an argument might favor the middle panel with a larger initial target, or the right panel with many small targets spread across the space.

Wright also pointed out that any genotype would have an exponentially large number of neighbors. Since he is also known for his advocacy of the role of genetic drift, random non-adaptive changes in genomes (Crow (2010)), he likely assumed that even if only a very small fraction of genetic possibilities are viable, the total space of potentially fruitful gene combinations remains unimaginably vast. Thus what we observe today in nature is only a small

[1] If one assumes that one barley corn, an ancient but tiny measure of weight, is about 0.05 g, then the total on day 30 is about 47 tons of barley. Interestingly, the scribe got his sums wrong, due perhaps to the complexity of Mesopotamian number systems (Soubeyran, 1984).

[2] As usual the history is more complex. There is evidence that Darwin had at least some indirect interaction with Mendel's work (see e.g. Sclater, 2006). Be that as it may, for all practical purposes, Mendel was forgotten.

[3] "*Natural selection can act only by the preservation and accumulation of infinitesimally small inherited modifications*" (Darwin, 1859, pp. 95).

[4] See Amundson (2007) for an opinionated counterpoint to Mayr's telling of the history of the MS.

[5] Another subfield of biology where the qualifier hyper-astronomical is frequently used is the number of possible connections in the brain (e.g. Edelman, 2001, pp. 38).

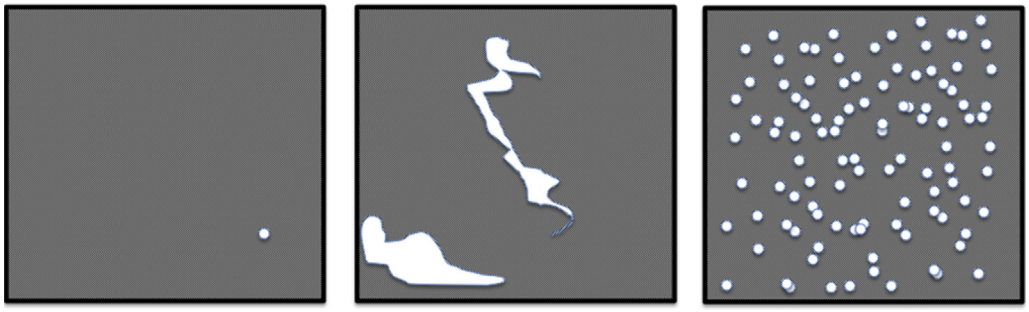

**Fig. 1.** Schematic of possible genotype space structures. White denotes biologically viable genetic combinations while dark grey unviable areas. **Left**: Only a very small fraction has any viable genotypes. **Center**: Most of genotype space is not viable, but there are a small number of fairly large areas where biologically possible genotypes do occur with higher probability. **Right**: Most of genotype space is not viable, but there are many small pockets where viable genotypes can occur. These may be interconnected. Other schematics can easily be imagined and it should be kept in mind that such two-dimensional representations are extremely limited.

fraction of what could be biologically possible. Thus Wright, and most of the other founders of the MS would have assumed that the current instantiation of genetic possibilities (life as we know it) is largely contingent, since it could just have well occupied a different part of genotype space.

Whether this genotypic contingency means that the phenotypes we observe are equally contingent depends very much on the structure of the mapping from genotypes to phenotypes. Unfortunately, these two types of contingency are sometimes conflated in the long literature on this topic, and one of the main points of this paper is to argue that they can be quite different. A similar conflation also followed interpretations of Wright's famous fitness landscape metaphor (Wright, 1932), where the axes were taken to refer to either genotypes or phenotypes by different subsequent authors (Pigliucci, 2008).

The hyper-astronomical size of these spaces appears again in Wright (1932, 365) "*under biparental reproduction a very low rate of mutation balanced by moderate selection is enough to maintain a practically infinite field of possible gene combinations within the species*." This is an early statement of another theme of the MS[6]: The variation for natural selection to act on is abundant.

Even if variation is abundant, one still needs to consider what its character will be. This question was famously formulated by one of the (re)discoverers of Mendelian genetics, the botanist Hugo de Vries, who wrote *"Natural selection may explain the survival of the fittest, but it cannot explain the arrival of the fittest"* (de Vries (1905), pp. 825–26).[7] In other words, the survival of the fittest may describe how selection prunes existing variation, but you may need another kind of theory to explain how novel phenotypic variation arrives in the first place.

By and large the MS answered this question by arguing that small gradual changes caused by mutations and recombination, when accumulated over time and acted on by natural selection, was sufficient to explain the arrival of the fittest. The same statistical arguments that explain how multiple changes at discrete loci can lead to small phenotypic changes, also quite naturally imply that the variation is effectively isotropic. For example, changes in genes could lead to an individual who is taller or shorter than the mean, but the probability of this occurring one way or the other is roughly equal. If variation is furthermore abundant, then it is not hard to see why another conclusion become part of the MS repertoire. Variation does not introduce a significant bias in evolutionary trajectories.

This last conclusion hardened as the MS matured. Consider for example by this influential statement from Ernest Mayr: "*Evolution is not primarily a genetic event. Mutation merely supplies the gene pool with genetic variation; it is selection that induces evolutionary change*" (Mayr (1963) pp 613). In other words, if we see an evolutionary outcome, then we should attribute the ultimate causes for this outcome primarily to selection, not to any possible biases in the kinds of variation that could have arisen.

There are many other ways to unpack the history of the MS, as well as the assumptions and implications of its theses. For example, there were a variety of complex historical reasons that the MS focused on the role of selection and underplayed the importance of variation. The relative importance of selection versus variation is one of the oldest controversies in evolutionary biology and predates the re-discovery of Mendel (see Amundson (2007) for a fuller discussion of these points). Moreover, in the last few decades the rise of evo-devo and other developments such as a better understanding of developmental plasticity, niche construction and even epigenetics have increased the calls for a new extended evolutionary synthesis (EES). see Pigliucci (2010), Laland et al. (2015), Noble (2015) for recent overviews of a complex discussion, and Laland et al. (2014) for an argument between authors calling for an extension of the MS, and authors claiming the MS is broad enough to incorporate these new developments. The subject is clearly complex. Here I argue am making a narrower argument, namely that the hyper-astronomically large space of possible gene combinations very likely influenced the founders of the MS.

## 2. The genetic code as a simple genotype–phenotype map

Wright's original formulation of the space of genotypes was limited by the fact that he didn't know exactly what genotypes were made of. That all changed when James Watson and Francis Crick published their classic 1953 paper on the structure of DNA (Watson 1953). Genetic information is encoded in the four nucleotides of DNA: ATCG. At this level, Wright's genotype space grows exponentially as $4^L$: it quadruples with $L$, the length of an

[6] See for example Dobzhansky (1949, 201): "It is the view of a majority of evolutionists that mutation and Mendelian recombination continually produce innumerable genetic materials, some of which are more and others less suitable for perpetuation in various environments."

[7] This phrase is originally due to Arthur Harris, whom de Vries quotes in the last line of his book (de Vries, 1905). Although de Vries is sometimes portrayed as a Mendelian anti-Darwinian (because of his saltationist viewpoints), he in fact held Darwin in very high esteem. As a young man he made a, for him, very inspirational visit to Darwin in 1878 (van der Pas 1970, p 187), engaged in extensive correspondence that included discussion of the origin of variation (van der Pas, 1970, pp. 200). In 1925 de Vries would say "*I was led to the study of heredity by … my love for Darwin*" (van der Pas, 1970, pp. 192). See also Stoltzfus and Cable (2014) for an extended argument that the early Mendelians, including de Vries, were less anti-Darwinian than is often claimed in the classic literature on this topic.

organism's DNA. As the ancient doubling problems showed, the sizes of such spaces rapidly become unearthly or hyper-astronomically large. For example, all combinations of DNA of length 79 nucleotides would weigh more than the earth (mass of $5.97 \times 10^{21}$ tons) and all of length 126 more than the estimated mass of the observable universe ($3 \times 10^{50}$ tons). Even the smallest viral genomes come out at several thousand nucleotides (HIV has about 9500), while humans have a genome made of about 3 billion nucleotides. The size of these genetic spaces is unfathomably large. Working out how all genotypes map to phenotypes is impossible even for something as small as a virus, let alone anything as complex as a mammal. Instead, to make progress, we need to study simpler systems. We therefore turn to what is perhaps the simplest non-trivial genotype-phenotype map: The mapping from the four letter alphabet used by DNA to the twenty amino acid alphabet that make up proteins.

This genetic code works by assigning three letter words of DNA to the amino acids, and can be viewed as a simple genotype–phenotype map. The genotypes are the $4^3 = 64$ ways of arranging four possible letters over words of three letters, while the phenotypes are the twenty amino acids plus three stop-codons (which mark the end of a gene).

The first thing to note is that there is *redundancy* in mapping, with on average about three codons per amino acid. There are more genotypes than phenotypes. Secondly, as can be seen in Fig 2A, there is also a *bias* in this genotype–phenotype map. For example, the amino acid tryptophan (W) only has just one codon mapping to it, whereas leucine (L) is encoded by six different codons and isoleucine (I) has three. So, if you chose a random set of three nucleotides of DNA, there is a six times larger chance of obtaining leucine than obtaining tryptophan.

Not long after the details of the mapping from codons to amino-acids were worked out, it was discovered that this redundancy correlated strongly with the relative frequency of amino acids found in biological protein sequences (Mackay (1967)). As shown in Fig. 2*B*, you are more likely to find leucine (L), with six codons mapping to it, than isoleucine (I), with three, and much less likely to find tryptophan (W), with only one codon.

The correlation is striking, but what causes it? In his original paper, Mackay (1967) followed the MS lead and provided an adaptive explanation: the code was optimized to correlate with the relative fitness of different amino acids. In other words, Leucine has more codons mapping to it than Isoleucine or Tryptophan do because this variation in redundancy (bias) has a positive fitness effect. But such arguments beg other questions, such as how did the code evolve this property? Once a code emerges, any further changes in the code will be hugely deleterious because a single change will affect many different proteins at once. A relatively small change would lead to huge negative consequences. It was for this reason that Francis Crick (1968) famously postulated that the genetic code was a frozen accident. It is not clear, at least with life as we know it, that the code could evolve in the way that Mackay suggested.

There is a completely different take on the correlation between codon frequency and amino acid abundance, based on the neutral theory of evolution, mainly associated with the with the great population geneticist Motoo Kimura. In a classic founding paper on the neutral theory, King and Jukes (1969) argued that this correlation between redundancy and amino acid coding frequency was strong evidence that many amino-acid substitutions had no discernable effect on fitness. Thus their coding frequency is simply caused by the rate at which variation arises, not by selection. This argument stands in contrast to the standard picture of the MS which would argue, which would argue that it is selection, not bias in the rate with which variation arrives that drives evolutionary change.

These unanswered questions about the genetic code are fascinating, in part because they may provide hints about the origin of life. There are a very large number of ways that the codons can be distributed among the amino acids. One might therefore think that, given Crick's arguments about the difficulty of the code evolving once it emerges, it would be impossible to search through this space. Thus the code is largely contingent, a frozen accident. Nevertheless, there is good evidence for a more interesting story. By comparing the properties of these many hypothetical codes, it has been shown fairly recently that the main code we find in nature is

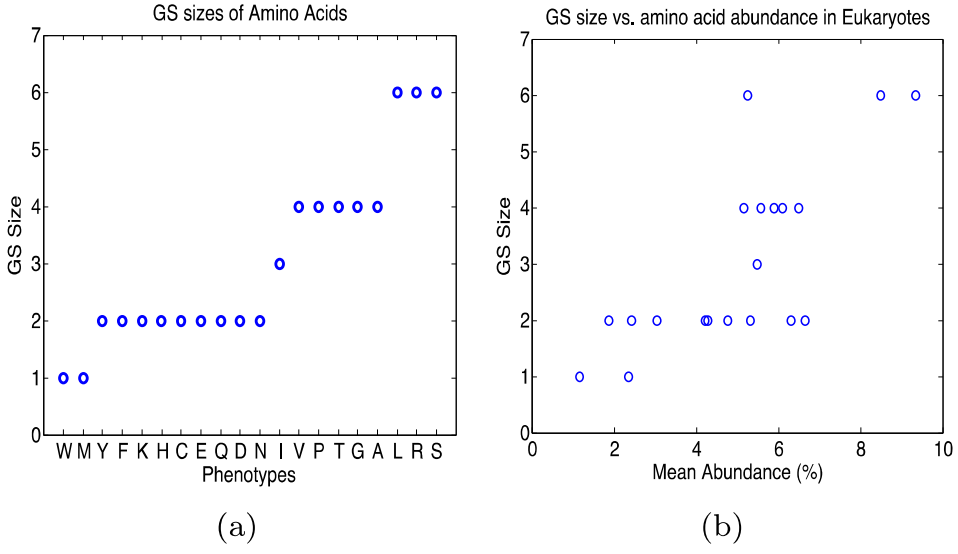


**Fig. 2.** Bias in mapping from codons to amino acids. In both graphs above the y-axis denotes the genotype set (GS) size, i.e. the number of codons per amino acid, which varies from 1 to 6. In *A*. the x-axis denotes the twenty different amino acids (each marked by a standard letter, e.g. methionine is M). In *B*. the x-axis denotes the mean abundance taken from the genomes of a set of five model eukaryotic organisms, namely *Arabidopsis thaliana* (a model plant), *Caenorhabditis elegans* (nematode worm), *Drosophila melanogaster* (fruit fly), *Homo sapiens* (humans), and *Saccharomyces cerevisiae* (yeast). There is a clear correlation between the GS and amino acid abundance in nature (Mackay, 1967). (Figure adapted from Dingle (2014).)

optimized in remarkable ways, for example for error reduction (see e.g. Freeland & Hurst, 1998 or Itzkovitz (2007)), so that it is "one in a million" in the words of Freeland and Hurst (1998).[8] Either Crick's frozen accident is an incredible stroke of luck, or there was some form of early (pre-Darwinian) optimization of the code. Perhaps this optimization also explains the codon bias, or else the pattern of redundancies could be a side effect of other processes. It should be kept in mind that even if the code was optimized in some kind of pre-Darwinian evolution, the use of amino acids would likely have been different than it is today and so it would be unlikely that this would optimize the code for current usage.

Our lack of understanding of the origin of life means that finding the ultimate cause of the genetic code's form is difficult. The jury is still out (For a good review of several competing theories of the origin of the genetic code see Koonin and Novozhilov (2009)). However, because this genotype–phenotype map is simple, so that properties of counterfactual alternative codes can be worked out, it sheds light on (or at least sharpens) some of the deep questions about the nature of our evolutionary history.

## 3. Hyper-astronomically large protein spaces and the Hoyle-salisbury paradox

The amino acids described above are strung together to form proteins, the molecular workhorses of the cell. If, for simplicity, we replace our genotype space by the space of amino acids (thus ignoring codon redundancy), then a mapping from all possible combinations of amino acids to proteins of length $L$ grows exponentially as $20^L$. Even for a relatively short protein made up of 58 amino acids, the set of $20^{58} \approx 10^{78}$ combinations would weigh more than the observable universe. The mean length of proteins in the human body is something like 476 amino acids. If one were to make all $20^{476}$ different combinations, then this would weigh more than about $10^{500}$ times the mass of the observable universe.[9]

Such hyper-astronomically large search spaces also naturally raise questions about how nature can ever find the sequence needed to perform a particular function. For example, the British astrophysicist and polymath Sir Fred Hoyle famously wrote "*I was constantly plagued by the thought that the number of ways in which even a single enzyme could be wrongly constructed was greater than the number of all the atoms in the universe*" (Hoyle, 1981). Even earlier, in the journal *Nature* (Salisbury, 1969), the American plant physiologist Frank Salisbury used several hyper-astronomical metaphors to emphasize the vast sizes of the protein search spaces, arguing that spaces were so large, and sparsely populated by viable folding proteins, that the standard picture of natural selection was not capable of explaining how viable proteins could evolve.[10] We might call this formulation of the search problem the *Hoyle-Salisbury paradox*.

In a famous response that pioneered a number of important concepts for the study of genotype–phenotype maps, John Maynard Smith (1970) took up Salisbury's challenge. Firstly, building on the neutral theory of Kimura and others, he pointed out that there was *redundancy* in the mapping from the "protein space" of all possible combinations of amino acids, to the phenotype space of viable proteins. Many different sequences map to the same folded and functional protein structure.[11] Secondly, he pointed out that these spaces are very likely to be *correlated*, that is if a sequence maps to a viable protein, then its neighbors are more likely than expected by random chance to contain viable proteins.[12] Finally, he pointed out that if the probability of having a viable neighbor is larger than the number of neighbors of a protein in protein space, then there can be a continuous mutational path connecting many different proteins that facilitates evolutionary exploration. He illustrated the third concept with a popular word game: Find a connection between two words by changing (mutating) just one letter at a time. The example he gave was:

WORD → WORE → GORE → GONE → GENE

There are $26^4 = 456{,}976$ possible 4 letter words in the English language. Only 4030 are valid words.[13] Each word has 100 words within one letter mutation (4 times 25 alternate letters) while the number of valid words is just below 1%, suggesting that in an uncorrelated space, the probability of having a valid word as a neighbor is slightly less than 1, which would mean that one would not expect large networks of words connected by single mutations. The game would be very difficult indeed. However, a direct calculation of the average number of neighbors of valid English words gives a much higher fraction of 13% which is due to correlations (Greenbury et al., 2015). Reasons for the correlations are not hard to see. Valid words need vowels, for example. So if a word has a vowel, then it is more likely to have valid neighbors. These kinds of correlations make the game much easier. If similar correlations hold in protein space (which recent calculations suggest is the case (Greenbury et al. (2015))), then once a biologically relevant protein is found, then it will be much easier to find other biologically relevant proteins in its direct mutational neighborhood. This could greatly enhance the probability of finding networks of linked phenotypes, and thus increase the rate at which evolution finds new selectable variation over the uncorrelated situation.

Another way to address the question of the probability of finding proteins from a random search has been to perform experiments using random protein libraries. For example, Keefe and Szostak (2001) used random length 80 sequences to deduce an estimated probability of $10^{-11}$ for generating ATP binding proteins from random sequences. This probability should be considered in the context that a mole of material is $6 \times 10^{24}$ particles, so that this probability is not that low. Nevertheless, estimates of the probability of finding viable solutions inside protein hyperspace continue to vary widely in the literature, a reflection of the fact that we don't

[8] There are approximately $1.5 \times 10^{84}$ different ways the 64 codons can be distributed over 20 amino acids and a stop codon (Schönauer & Clote, 1997).

[9] Some versions of multiverse theory based on string theory postulate $10^{500}$ universes (Douglas, 2003).

[10] The claim that protein spaces are too hyper-astronomically large to be searched by evolutionary mechanisms remains a common trope of the anti-evolutionary literature, see e.g. Meyer (2009) "*Another way to say that is the probability of finding a functional protein by chance alone is a trillion, trillion, trillion, trillion, trillion, trillion, trillion times smaller than the odds of finding a single specified particle among all the particles in the universe*".

[11] For more recent discussions, see e.g. Dill (1999) who argues that for many basic properties of a protein, the 20 amino acid alphabet can be simplified to just those that are hydrophobic or polar, reducing the scaling with length L from $20^L$ to $2^L$. He further argues that the identity of only about a third of the amino-acids really matters for folding, reducing the scaling down to $2^{L/3}$. Perhaps the most ambitious argument for the reduction of the effective size of this protein space can be found in Dryden, Thomson, & White (2008) who argue that most of the relevant phenotype space could have been explored during the history of life. The problem with such arguments based on redundancy is that they don't get rid of the exponential scaling. Make the length a bit longer and the system size becomes hyper-astronomical again. The concept of redundancy is also subtle. It depends, for example, on how tightly one defines the phenotype. Nevertheless, it is important to remember that redundancies can be very large so that the space of phenotypes is expected to be much smaller than the space of genotypes.

[12] For a more detailed investigation of the role of correlations in genotype-phenotype maps see Greenbury et al. (2015).

[13] At least according to the US Scrabble © dictionary http://www.wordfind.com/4-letter-words/.

really know the answer to this question, and that it can depend in detail on exactly what kind of question is being asked.

While the evidence for protein evolution is overwhelming, the question of exactly how the mapping from genotypes to phenotypes is navigated remains an open question. Maynard Smith's arguments, mainly 1) that redundancy greatly reduces the size of the phenotype space as compared to the size of the protein sequence space, 2) that correlations make it much more likely than one would expect from random chance to find biologically viable proteins connected by single point mutations in sequence space and 3) that this higher probability of connection leads to networks that facilitate the arrival of selectable variation, were important steps forward. Unfortunately it is still not possible to work out accurately what a protein's folded structure or function is just from its sequence, so that a theoretical search through all possible sequences is still well beyond current capacities. Perhaps the best way forward is by doing experiments, such as those of Keefe and Szostak (2001).

### 3.1. Aside: navigating hyper-astronomical conforational spaces in real time and Levinthal's paradox

Another famous invocation of hyper-astronomical metaphors comes from the problem of protein folding and was formulated by Cyrus Levinthal (1969). Proteins are long polymers, and many can fold reversibly from disordered configurations into well-defined three-dimensional structures that determine their function. Levinthal's argument goes like this: Consider a protein made up of 150 amino acids and assume for the sake of the argument that there are ten angles between each amino acid. To first order there are 10 angles between the first two amino acids, a further 10 between the next, and so on, leading to roughly $10^{150}$ different configurations. The maximum rate for sampling configurations is something like $10^{13}$ per second. Thus it would take $10^{133}$ times the age of the universe to sample all of them (yet another hyper-astronomical metaphor, but now a temporal one). At face value, this argument suggests that a protein could never find the very small subset of well-defined folded states in finite time. But we know from experiment that proteins can fold in microseconds. Witness *Levinthal's Paradox*.

Like many good paradoxes, Levinthal's paradox derives its rhetorical power from smuggling in a hidden, but seemingly uncontroversial, assumption. Here the culprit is taking for granted that searching through the space of protein configurations is like looking for a needle in a haystack or like finding a hole on a very large flat golf course: each configuration is equally likely to be scrutinized. Levinthal (1968) proposed that proteins sidestepped his paradox by evolving highly specified temporal pathways to the folded state so that only a small subset of configurations is ever explored. Modern statistical mechanical theories emphasize an ensemble picture where many different starting points all generate kinetic pathways down a "funneled landscape" to the correct folded state (Dill, 1999). Although the detailed pathway down the funnel may be different in each case, the outcome, arriving at the bottom of the funnel, is, for all practical purposes, always the same (golf would be a lot easier if courses were designed with the hole at the bottom of a much larger funnel shaped depression).

It is estimated that a small but significant fraction of all possible protein sequences fold well. Moreover, there is good evidence that evolution tinkers with the sequence not just to optimize the folded state, but also to enhance the ability of a protein to fold. Levinthal's paradox is therefore not solved by redundancy of the search space. The fraction of desired folded states remains a very small set of all possible configurations. Instead, it is overcome by a biased search that navigates through the hyperspace of possibilities in real time.

There is a lesson here for evolutionary search. If we could not observe proteins folding in real time, a theoretical argument such as that presented by Levinthal might seem quite compelling. By contrast, evolutionary time-scales are so long that we cannot directly observe the full evolutionary pathway over time. To an evolution sceptic, theoretical arguments such as those presented by Hoyle or Salisbury can appear to have considerable merit. The Levinthal paradox should give them pause. Perhaps it is best to simply point out the copious indirect evidence for evolutionary trajectories, for example from molecular phylogenetics.

Furthermore, studies of evolutionary convergence (Conway Morris, 2003; 2015; McGhee, 2011) demonstrate that evolution can indeed reach similar protein phenotypic endpoints multiple times, even though the search spaces are hyper-astronomically large. This suggests, by analogy to Levinthal's Paradox, that naïve assumptions about the search must break down.

So what assumptions need to be reconsidered? Firstly, the naïve, but incorrect, assumption that the size of the space is set by the number of possible sequences is clearly wrong. The widespread observation of neutral mutations and large sequence dissimilarity between similar proteins demonstrates that there is a lot of redundancy in the map. The size of the phenotype space is many orders of magnitude smaller than that of the genotype or sequence space. But whether this redundancy is enough to explain the success of evolutionary search is still an open question. The difficulties in artificial or directed evolution of proteins (Romero 2009) should flag up caution here.

Another option could be that something similar to the solution to the Levinthal paradox is at work in evolutionary design space. Fitness would rise as the protein gets closer to its final correct phenotype, thus introducing a fitness bias that extends well beyond the correct phenotype. For evolutionary search, such an adaptive solution to the "Blind Watchmaker Paradaox" (Dawkins, 1996) could, just as in protein folding, lead to an exponential speedup of the search time for a particular state. However, it depends on there being an overall funnel-like structure in design space, and in contrast to protein folding (Dill, 1999), it is not clear where this would come from.

Yet another possibility would be that the current instantiation of proteins is a fortunate accident. Even though the fraction of sequence space with biologically relevant solutions is small, correlations mean that as long as you start in a fruitful place, networks of connected viable proteins are relatively easy to find (Greenbury et al., 2015; Maynard Smith, 1970). Again, this could be (part of) the story, but it is hard to establish with certainty.

In the next section we will study RNA, which like proteins can fold into well defined structures and perform structural or catalytic functional roles. The advantage is that many properties of the genotype–phenotype map are easier to calculate, so that some of the questions raised above about redundancy, correlations or a funnel-like landscape in design space can be more directly addressed.

## 4. The RNA secondary structure genotype–phenotype map and the arrival of the frequent

In the cell, DNA copies its information to a sister molecule, RNA, which is also made up of a string of four different nucleotides. These instructions are then read by ribozymes, nanoscale factories that make proteins to order. However, RNA can do more than store information like DNA does. It can also catalyze reactions, or act as a structural element, just as proteins do. This ability to perform the dual tasks of replication and metabolism has made RNA a popular candidate for the origin of life.

Like proteins, RNA can fold into well-defined three-dimensional configurations. Because it only has a four letter alphabet, in principle this folding problem is considerably simpler than for proteins.

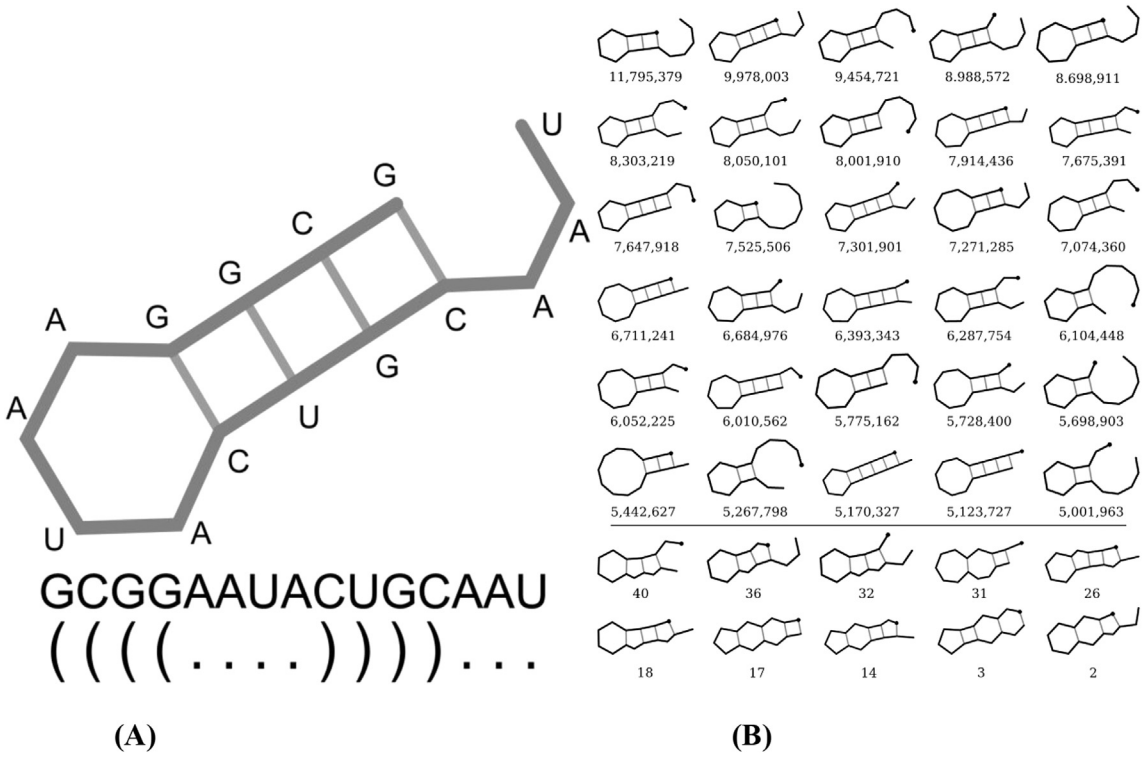


**Fig. 3.** *A*. The secondary structure (who sticks to whom) of a length 15 RNA sequence. The lines denote bonds. In this case, starting from the final G, one could describe the shape as: 1 binds to 12, 2 to 11, 3 to 10, 4 to 9, 5–8 don't bind, and 13–15 don't bind. Below the figure the sequence is given together with a more abstract notation of the shape, brackets denote bonds, and dots no bonds. Many different sequences can have the same secondary structure shape (the same dot-bracket sequence). *B*. For $L = 15$ RNA, there are 431 secondary structure shapes that sequences can map to. Above the line, the thirty most frequent secondary structure phenotypes (with the largest GS sizes) are shown, and below the line, the ten least frequent phenotypes (with the smallest GS sizes) are shown (adapted from Schaper (2012)).

Unfortunately, it is still not easy to solve the full three-dimensional shape of RNA just from the sequence. However, one can describe this folding problem in a more schematic way by looking only at the so-called secondary structure of RNA,[14] that is at a list of which nucleotide along the string binds to which other nucleotide. Fig. 3 illustrates some such secondary structures for sequences of length $L = 15$.

The great advantage of studying RNA secondary structures is that very fast algorithms exist to accurately calculate the free-energy minimum structure that corresponds to a particular sequence (Hofacker et al. 1994). Thus the genotype–phenotype map from sequence to phenotype (secondary structure) can be studied in much greater detail than typical protein genotype–phenotype maps can.

The number of sequences scales exponentially as $4^L$, which again rapidly becomes far too large to search comprehensively. But for smaller spaces one can exhaustively calculate all possible secondary structures and the results are instructive. Consider the example in Fig. 3. For $L = 15$ there are $4^{15} \approx 10^9$ or 1 billion different sequences, but only 431 distinct secondary structure phenotypes. So there is huge redundancy in this space, which greatly reduces the size of the phenotype search space.

There is another important feature to notice here. The number of genotypes mapping onto a phenotype (the Genotype Set (GS) size) is hugely biased. The most frequent secondary structure with bonds has about twelve million genotypes folding into it, while the least frequent only has two genotypes. In other words, a random mutation is six million times more likely to throw up the most frequent secondary structure than the least frequent secondary structure. That is a much bigger bias than the six-to-one maximum observed in the genetic code.

This biasing towards a small set of structures also means that for $L = 15$ RNA, just 26 (6%) of the 431 sequences gobble up 50% of the sequence space. We will call these the *frequent phenotypes*. In other words, if you select a sequence at random, you have a 50% chance of getting one of these 26 frequent phenotypes. That is a big drop from the one in a billion chance of finding a particular sequence (genotype).

Another related lesson from Fig. 3 concerns the difference between sampling over genotypes (the sequences) or over phenotypes (the secondary structures). There is a venerable tradition of theoretically studying the distribution of possible evolutionary outcomes (phenotypes). The set of possible forms, shapes or structures of organisms is sometimes called the morphospace[15] and Fig. 3*B* shows part of the RNA secondary structure morphospace of 431 possible structures. If one samples uniformly over phenotypes, then each one is more or less equally likely to appear. However, evolution proceeds by random mutations, so the search is in genotype space, and mapped through a genotype–phenotype map to the morphospace. What this example demonstrates is that the genotype–phenotype map heavily biases the search over morphospace to a small subset of "frequent" secondary structure phenotypes.

We have recently compared the results of randomly sampling over the space of genotypes to databases of known functional (so-

[14] For a good non-technical introduction, see Wagner (2005, chap. 4). See also the classic work by the Vienna group, reviewed for example in Schuster (2001).

[15] See McGhee (2007) for an overview of the concept of morphospaces, and McGhee's article in this issue for further discussion.

called non-coding) RNAs. Rather surprisingly, the distribution of phenotypic properties such as the number of stacks of contiguous base pairs per structure, or the mutational robustness of natural RNA almost perfectly track the random sampling over genotypes (Dingle et al., 2015). It is widely believed that structure is crucial to biological function. Yet these important features of RNA have the same distribution as if natural selection could be completely ignored. We argue elsewhere (Dingle et al., 2015) that an alternative argument is that natural selection uses structures that are presculpted by the genotype-phenotype map, and that it mainly acts on a small sub-set of the sequence.

Perhaps the best known example of convergence in RNA is the repeated emergence across the tree of life of the hammerhead ribozyme (Hammann, Luptak, Perreault, & de la Peña, 2012), an RNA motif that can cleave itself (it is catalytic, which is why it is called a ribozyme). In a famous set of experiments, Kourosh Salehi-Ashtiani and Jack W. Szostak (2001), recreated a very large number of independent evolutionary trajectories in their laboratory by making many random RNA sequences and selecting them on their self-cleaving activity. After several rounds of selection they found that the same hammerhead motif evolved again and again, from which they concluded:

> *Our results show that, despite the dominance of contingency (historical accident) in some recent discussions of evolutionary mechanisms [Gould], purely chemical constraints (that is, the ability of only certain sequences to carry out particular functions) can lead to the repeated evolution of the same macromolecular structures*. (Salehi-Ashtiani & Szostak, 2001, pp 84)

A typical length for a hammerhead ribozyme is about $L = 55$. The set of all possible $L = 55$ RNA sequences would weight 10 million tons, while the experiments are typically done on a few hundredths of a gram of material. So only the tiniest fraction of the space is searched in these experiments, and yet the same structure motifs are found, although of course the sequences are different. How can this be?

Our calculations for $L = 55$ show that the hammerhead ribozyme secondary structure has a very large number of sequences mapping to it (a large GS size) and so is particularly easy to find. Due to bias, experiments with randomised RNA strands, or for that matter evolution itself, can only access a tiny fraction of the full morphospace, and this may explain the convergence seen in the experiments by (Salehi-Ashtiani & Szostak, 2001) are many other self-cleaving enzymes out there in the full morphospace of possible phenotypes, but because these have smaller GS size, they are very unlikely to arise, and therefore even more unlikely to be selected and to fix (Schaper 2014). Or perhaps the rest of the phenotype space consists of "howling wildernesses of the maladaptive" as far as self-cleaving enzymes are concerned. Nature doesn't care. The only game in town is played by the frequent phenotypes.

We have recently completed the largest exhaustive calculation achieved to date of an RNA GP map, namely for $L = 20$ RNA (Schaper & Louis, 2014). There are $4^{20} \approx 1$ trillion different sequences, and just storing the information takes gigabytes of data. This is still very short for biologically relevant RNAs. Unfortunately, doing exhaustive calculations for larger systems is not feasible. So instead we have developed analytic methods that work for arbitrary lengths (Dingle, 2015). What this work shows is that the bias becomes much more pronounced for longer lengths. Whereas for $L = 15$ RNA 26/431 ~ 5% of structures take up the majority of genotypes, for $L = 55$ RNA this fraction drops to just 0.1% of the full morphospace (Dingle et al., 2015). So only a tiny fraction of the morphospace of shapes can effectively be searched by random undirected mutations.

The story told above about RNA is, admittedly for a fairly simple model system. But it has the big advantage is that we can quantitatively calculate properties of the whole genotype–phenotype map, something that is very rarely possible (yet) in biology. This allows us to calculate the whole set of possible phenotypes, and so get a sense of the counterfactual space: what could have happened, but did not.

As is often the case with detailed calculations, this study throws up unexpected results that simpler heuristic arguments may have missed. Perhaps the most striking finding is the extremely strong bias that dominates the evolutionary outcomes observed in nature.

It is also interesting to compare the properties of the RNA to secondary structure map to considerations from the MS that were influenced by less well-characterized genetic hyperspaces. The conclusion that nature can only search a very small fraction of all possible genotypes still holds, of course. RNA space grows exponentially as $4^L$, leading to the usual hyper-astronomical metaphors. For example, all possible RNAs of length $L = 126$ would weigh about the same as the mass of the visible universe, and much longer RNAs are found in nature. Moreover, we see explicitly for the relatively small space for $L = 15$ RNA in Fig. 3 that some of the very rare structures only have a handful of sequences mapping to them. Such structures will be very hard to discover by random mutations, and are therefore very unlikely to arise in nature.

While the conclusion that the genetic sequences found in nature for RNA are highly contingent also follows naturally from the size of the genetic spaces, the inference that this also holds for phenotypes is not so obvious. Firstly, the large redundancy reduces the size of the phenotype spaces (although the number of possible secondary structure phenotypes also grows exponentially with size, albeit with a lower exponent than the number of genotypes, so can also becomes unsearchably large). Secondly, the strong bias means that the fraction of the phenotype space that is accessible to evolution is greatly reduced further. Thirdly, the nature of these high-dimensional spaces means that in just a few steps you can reach almost any frequent phenotype from any other frequent phenotype (Schuster 2011; Wagner, 2005). More generally, although the volume of the genotype space grows exponentially as $4^L$, the maximum distance between genomes only grows linearly as $L$. Consider $L = 15$ RNA. There are 1 billion sequences, but you can transform any sequence to any other in at most 15 steps. And the frequent phenotypes are typically connected on average by a much smaller number of steps than the maximum (Schuster 2011; Wagner, 2005). Clearly these search spaces are very different from the low-dimensional spaces we are used to thinking of. Wright's influential two-dimensional landscape with fitness in the third dimension (Wright, 1932) is such a caricature of these high-dimensional spaces that I am unsure if this influential metaphor restricts our imagination more than it illuminates it. Nevertheless, if the space of all genotypes were represented in two dimensions, as in Fig. 1, then it probably is best illustrated by the rightmost panel because many different initial conditions (genotypes) can easily lead to the same frequent phenotypes. The geometry of the mapping from genotypes to phenotypes means that outcomes may be much more predictable and likely to repeat than one might think from initial reflections on the space of genotypes.

Another conclusion from the MS, that there is ample variation for natural selection to act on, probably remains true for RNA in part due to the fact that any two sequences can be fairly easily connected by mutations, suggesting that a lot of variation is locally available near a given sequence (Schuster 2011; Wagner, 2005). Nevertheless, that variation is mainly limited to the frequent phenotypes.

Finally, the more controversial conclusion of the MS, that variation is isotropic, is dramatically contradicted for this RNA system.

The anisotropic bias in the arrival of variation is so strong that it provides the dominant predictive power as to which secondary structure phenotypes could potentially appear in nature (Dingle et al., 2015). We have called this effect where only a subset of all structures are available due to such coding constraints the "arrival of the frequent" (Schaper & Louis, 2014).

One must be careful not to construe this dominance of variation for determining distributions too quickly as a victory for the neutralist camp over the selectionist camp. Consider the simpler example of the genetic code. Yes, it is true that certain amino acids are much more likely to arise by random variation than others, but that doesn't mean that what fixes doesn't do so because of natural selection. For example, about half the amino acids are hydrophobic, which means that it is not favorable to be in water and so they are preferentially found on the inside of a protein fold. If a protein has evolved a fold, and needs to select for more thermodynamic stability, it can sometimes achieve this by filling an internal cavity with a hydrophobic amino acid. Depending on the details, some hydrophobic amino acids will be better than others, but perhaps several will do the job. Once one appears (variation) and is selected for, then it prevents others from fixing in its place as any further evolutionary advantage will be relatively small. So hydrophobic amino acids like leucine, with six codons, are more likely to appear and thus to fix than closely related hydrophobic acids like isoleucine, with just three codons. Similar effects may be in play for RNA, such that multiple secondary structures could solve a problem, but the most frequent structures are proposed the most often as potential variation, and hence are most often chosen by natural selection to "survive." What is interesting is not whether selection is playing a role. It almost certainly does in the convergence of the hammerhead ribozyme for example. The interesting question is where the causal force comes from. If the arrival of variation was not strongly biased, one might expect nature to search fairly uniformly among different possible phenotypes. Instead, bias means that variation strongly restricts the palette of possibilities that natural selection can work with. If that bias is strong enough, it may be said to be the cause for a particular evolutionary outcome. For example, why does the hammerhead ribozyme repeatedly emerge in nature? The ultimate cause for a self-cleaving ribozyme is most likely that it provides a selective advantage. But the ultimate cause that the hammerhead ribozyme is the particular self-cleaving enzyme found, and the ultimate cause for its convergence is probably not selection but rather the non-adaptive process we have called the arrival of the frequent (Schaper et al., 2014, Dingle et al., 2015).

## 5. Conclusions

In summary then, we have shown that genotype spaces grow exponentially and rapidly reach sizes that can only be described as hyper-astronomical. Clearly nature can't search through spaces that are so large, and it seems like a small step to go from the size of these spaces to conclusions about the contingency of evolutionary outcomes. Reflections on the size of these spaces are likely to have influenced the founders of the MS. However, the ubiquity of convergence suggests that there may be deeper patterns (Conway Morris, 2008) at play that bias evolutionary search to similar outcomes, possibly through different pathways. Furthermore, the Levinthal paradox warns us that impossibility arguments invoking hyper-astronomically large numbers can be vulnerable to hidden assumptions.

For a few model systems, a full genotype–phenotype map can be calculated, which can shed light on the overall space over which evolution can search. The relatively simple example of the genetic code already shows modest bias towards certain outcomes, and that bias appears to be recapitulated in the natural frequencies of amino acids. Much stronger bias is found in the RNA to secondary structure mappings, where it suggests that only an exponentially small fraction of all possible phenotypes, the frequent ones, dominate what natural selection can work with. Since these frequent phenotypes constitute a tiny subset of the space of all phenotypic possibilities they are more likely to be found more than once, which may help explain convergence.[16]

It is too early to tell whether these forays into RNA secondary structure genotype–phenotype maps can be extended to more complex biological systems. Preliminary work on simplified models of protein folding, protein quaternary structure self-assembly, gene networks, and development all suggest that bias in the introduction of variation plays a much more important role than has been acknowledged in the MS (Dingle et al., 2015; Schaper 2014). Whether the adaptationists are right that convergent evolution mainly signals the power of unfettered natural selection also remains an open question. While selection surely plays an important role in convergence, it may be that the ultimate causes for some important examples of convergence are more closely entwined in the deep structure of a full genotype–phenotype map that heavily biases nature's possibilities in certain preferred directions. The jury is out.

## References


Amundson, R. (2007). *The changing role of the embryo in evolutionary thought: Roots of evo-devo*. Cambridge: Cambridge University Press.

Conway Morris, S. (2003). *Life's solution: Inevitable humans in a lonely universe*. Cambridge: Cambridge University Press.

Conway Morris, S. (Ed.). (2008). *The deep structure of biology: Is convergence sufficiently ubiquitous to give a directional signal?*. West Conshohocken, PA: Templeton Press.

Conway Morris, S. (2015). *The runes of evolution: How the universe became self-aware* West Conshohocken. PA: Templeton Press.

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology, 38*, 367–379.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Dawkins, R. (1996). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. New York: Norton.

Dill, K. A. (1999). Polymer principles and protein folding. *Protein Science, 8*, 1166–1180.

Dingle, K. (2014). *Probabilistic bias in genotype-phenotype maps*. DPhil. Thesis. University of Oxford.

Dingle, K., Schaper, S., & Louis, A. A. (2015). Phenotype bias strongly constrains the evolution of non-coding RNA. *Journal of the Royal Society Interface Focus, 5*, 20150053.

Dobzhansky, T. (1949). Genetic structure of natural populations. *Carnegie Institution of Washington Year Book, 48*, 201–213.

Douglas, M. (2003). The statistics of string/M theory vacua. *Journal of High Energy Physics, 0305*, 46.

Dryden, D. T. F., Thomson, A. R., & White, J. H. (2008). How much of protein sequence space has been explored by life on earth? *Journal of the Royal Society Interface, 5*, 953–956.

Edelman, G. M. (2001). Building a picture of the brain. In G. M. Edelman, & J.-P. Changeux (Eds.), *The brain* (pp. 37–70). Somerset, NJ: Transaction Publishers.

Freeland, S. J., & Hurst, L. D. (1998). The genetic code is one in a million. *Journal of Molecular Evolution, 47*(3), 238–248.

Friberg, J. (2005). *Unexpected links between Egyptian and Babylonian Mathematics*. Singapore: World Scientific.

Greenbury, S. F., Schaper, S., Ahnert, S. E., & Louis, A. A. (2015). Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. http://arxiv.org/abs/1505.07821.


[16] In the literature convergence proper is often distinguished from parallel evolution, where the same character emerges more than once because two different organisms share a common ancestor that primes them to find the same solution. The biasing in the genotype–phenotype map could be construed as causing parallel evolution because it is in a sense an "internal constraint." But that would be a mistake, as this effect does not depend on sharing common ancestry. In other words, a frequent phenotype may re-appear several times, but never share the same evolutionary pathway.

Hammann, C., Luptak, A., Perreault, J., & de la Peña, M. (2012). The ubiquitous Hammerhead ribozyme. *RNA, 18*(5), 871–885.
Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie – Chemical Monthly, 125*, 167–188.
Hoyle, F. (1981). The universe: Past and present – reflections. *Engineering & Science, Nov 1981*, 8–12.
Huxley, J. (1942). *Evolution, the modern synthesis*. New York: Harper & Brothers.
Itzkovitz, S., & Alon, U. (2007). The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Research, 17*(4), 405.
Kauffman, S. A. (1995). *At home in the universe: The search for laws of self-organization and complexity*. New York: Oxford University Press.
Keefe, A. D., & Szostak, J. W. (2001). Functional proteins from a random-sequence library. *Nature, 410*, 715–717.
King, J. L., & Jukes, T. H. (1969). Non-Darwinian evolution. *Science, 164*, 788.
Koonin, E. V., & Novozhilov, A. S. (2009). Origin and evolution of the genetic code: The Universal Enigma. *IUBMB Life, 61*(2), 99–111.
Laland, K. N., Uller, T., Feldman, M. W., Wray, G. A., Hoekstra, H. E., Futuyma, D. J., et al. (2014). Does evolutionary theory need a rethink? *Nature, 514*, 161–164.
Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., et al. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. *The Royal Society Publishing Proceedings B, 282*(1813), 20151019.
Levinthal, C. (1968). Are there pathways for protein folding? *Journal de chimie physique, 65*, 44–45.
Levinthal, C. (1969). How to fold graciously. In J. T. P. DeBrunner, & E. Munck (Eds.), *Mossbauer spectroscopy in biological systems: Proceedings of a meeting held at Allerton House, Monticello, IL* (pp. 22–24). Champaign: University of Illinois Press.
Mackay, A. L. (1967). Optimization of the genetic code. *Nature, 216*, 159–160.
Maynard Smith, J. (1970). Natural selection and the concept of a protein space. *Nature, 225*, 563.
Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Belknap Press of Harvard University Press.
McGhee, G. R., Jr. (2007). *The geometry of evolution: Adaptive landscapes and theoretical morphospaces*. Cambridge: Cambridge University Press.
McGhee, G. R., Jr. (2011). *Convergent evolution: Limited forms most beautiful*. Cambridge: Massachusetts Institute of Technology Press.
Meyer, S. M. (2009). *Signature in the cell*. New York: Harper Collins.
Noble, D. (2015). Evolution beyond neo-Darwinism: A new conceptual framework. *Journal of Experimental Biology, 218*, 7–13.
Pigliucci, M. (2008). Sewall Wright's adaptive landscapes: 1932 vs. 1988. *Biology & Philosophy, 23*, 591–603.
Pigliucci, M., & Muller, G. (Eds.). (2010). *Evolution – The extended synthesis*. Cambridge: The MIT Press.
Romero, P. A., & Arnold, F. H. (2009). Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology, 10*, 866–876.
Salehi-Ashtiani, K., & Szostak, J. W. (2001). In vitro evolution suggests multiple origins for the Hammerhead ribozyme. *Nature, 414*, 82–84.
Salisbury, F. B. (1969). Natural selection and the complexity of the gene. *Nature, 224*, 343.
Schaper, S. (2012). *On the significance of neutral spaces in adaptive evolution*. DPhil. Thesis. University of Oxford.
Schaper, S., & Louis, A. A. (2014). The arrival of the frequent: How bias in genotype-phenotype maps can steer populations to local optima. *PLoS One, 9*(2), e86635.
Schönauer, S., & Clote, P. (1997). How optimal is the genetic code? In D. Frishman, & H. Mewes (Eds.), *Computer Science and Biology, Proceedings of the German Conference on Bioinformatics (GCB'97)* (pp. 65–67) Oxford: Oxford University Press.
Schuster, P. (2001). Evolution in Silico and in Vitro: The RNA model. *Biological Chemistry, 382*, 1301–1314.
Sclater, A. (2006). The extent of Charles Darwin's knowledge of mendel. *Journal of Biosciences, 31*(2), 191–193.
Soubeyran, D. (1984). Textes mathématiques de Mari. *Revue d'Assyriologie, 78*, 19–48.
Stoltzfus, A., & Cable, K. (2014). Mendelian-Mutationism: The forgotten evolutionary synthesis. *Journal of the History of Biology, 47*, 501–546.
van der Pas, P. W. (1970). The correspondence of Hugo de Vries and Charles Darwin. *Janus, 57*, 173–213.
de Vries, H. (1905). In D. T. MacDougal (Ed.), *Species and varieties: Their origin by mutation*. Chicago: The Open Court Publishing Company.
Wagner, A. (2005). *Robustness and evolvability in living systems*. Princeton: Princeton University Press.
Watson, J. D., & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature, 171*, 737–738.
Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones (Ed.), *Transactions and General Addresses: Vol. 1. Proceedings of the Sixth International Congress of Genetics, Ithaca, New York, 1932* (pp. 356–366). Menasha, WI: Brooklyn Botanic Garden.