Communications in Information and Systems
Volume 23, Number 4, 359–392, 2023

The maximum genetic diversity theory of molecular evolution

Shi Huang*

The study of evolution has a long history and early theories of evolution included one by Lamarck and the other by Darwin. Advances in Mendelian genetics resulted in the modern synthesis in the 1930s merging Darwin's theory with genetics. However, the reality of genetic diversity remains poorly understood. The unlocking of protein sequences in the early 1960s revealed a shocking phenomenon, genetic equidistance, which then led to an ad hoc hypothesis known as the molecular clock. This in turn inspired the neutral theory by Kimura, which negates the role of natural selection in molecular evolution. The neutral theory has served as a useful null model but remains an unsatisfactory account for genetic diversity. In the first decade of the 21st century, we fortuitously rediscovered the long-overlooked genetic equidistance phenomenon, which inspired us to propose the maximum genetic diversity hypothesis as a comprehensive evolutionary theory. Analytical tests have shown that genetic distances observed today are mostly at maximum saturation rather than still increasing with time as misread by the molecular clock and the neutral theory. The maximum genetic diversity theory posits that macroevolution from simple to complex taxa involves a punctuational increase in epigenetic complexity and a corresponding loss in the maximum genetic diversity that a taxon can tolerate. It rekindles some of Lamarck's ideas and fully grants the proven virtues of Darwin's and Kimura's theories. The theory will rewrite molecular phylogeny and help solve difficult biomedical problems including the mystery of the purpose of sexual reproduction.

Keywords and phrases: Genetic diversity, genetic equidistance, molecular clock, neutral theory, maximum genetic diversity theory.

1. Introduction

The theoretical investigation of evolutionary phenomena has a long history. It first started by observing the phenotypes of various biological organisms.

*Corresponding author.

Early evolutionary thought came from the French naturalist Jean-Baptiste Lamarck. In 1809, he published "Philosophie Zoologique", proposing an evolutionary theory consisting of two forces. The first is a complexifying force accounting for the evolutionary progress from simple to complex. The second is an adaptive force. He believed that the effect of environments on traits can be passed on through generations.

Charles Darwin and Alfred Wallace independently came up with the idea of natural selection and jointly presented their papers on natural selection to the Linnean Society of London in 1858. Three other English naturalists also independently had similar ideas from 1813 through 1859, including Edward Blyth, Patrick Matthew, and William Wells. The racism and expansion of the British Colonial Empire at the time may have played a role in the popularity of the law of the jungle among the culture of the English men [1, 2]. The key concepts are common descent and natural selection to filter variations.

In the early 20th century, the rediscovery of Mendelian genetics led to the modern synthesis of Darwin's theory with genetics. The modern synthesis claims that adaptive evolution is a result of natural selection acting on genetic mutations or variations. While the modern synthesis is widely acknowledged to be a correct mechanism for the microevolution of traits, its claim that macroevolution involves the same mechanism as microevolution is more controversial and has had little experimental support. Regardless, however, this theory remains at best incomplete as it cannot account for a shocking finding in molecular evolution that was first found in the early 1960s, the genetic equidistance phenomenon. Two different theories of molecular evolution have since been proposed that were both inspired by this finding. Here we review and compare these two competing theories, i.e., the neutral theory and the maximum genetic diversity theory. There is also a recently published textbook by Prof. David Bickel that compared these two theories [3]. We have also introduced these theories in a recent Chinese textbook [4].

2. Genetic equidistance, molecular clock, and the neutral theory of molecular evolution

2.1. Genetic equidistance or molecular equidistance and the molecular clock

In the early 1960s, molecular or genetic differences between species were measured for hemoglobin, cytochrome C, and fibrinopeptides [5, 6, 7]. Genetic distance is a measure of molecular differences between species and

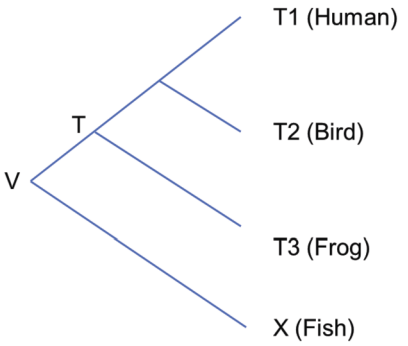

Figure 1: Illustration of the genetic equidistance result and its interpretation known as the molecular clock.
Comparison for tetrapod species (T1–T3; human, bird, frog), which are known to have a most recent common ancestor (T), and another species (X; fish). Species X is the outgroup species and is distantly related to species T1–T3, the ingroup species. Evolutionary lineages leading to species T1–T3 separated from the lineage leading to X at the same point, V. Furthermore, species T1–T3 are products of an evolutionary process that has run for exactly the same amount of time since V, because they share a common ancestor V. Therefore, if a given protein is equally different when we compare the same fish protein with proteins from different tetrapods, then the rate at which differences accumulate is similar among tetrapods (T1–T3). The tacit assumption here is that the molecular distance among the species is not at maximum level and hence can be used to infer the rate at which sequence differences accumulate.

is represented by the percentage difference in orthologous protein or DNA sequences. There were two kinds of results depending on whether the reference species to which one was comparing other species is the most complex or the least complex (complexity is defined as the number of cell types). If the reference species is human, one would observe 'the gradually increased amount of difference found when human hemoglobin is compared with hemoglobin from progressively more distant species' [5, 6, 7]. This inspired Zuckerkandl and Pauling to informally propose the universal molecular clock that hemoglobins from different species are changing at a steady and similar rate of 1.4×10^{-7} amino acid substitutions per year [5, 6, 7, 8].

If on the other hand the reference species is the least complex among the species being compared, e.g., when fish is the reference in comparison to human, bird, and frog (Figure 1), one would observe a highly unusual result, the genetic equidistance result that fish is approximately equidistant

Table 1: Percentage differences among species in Dot1

YEASA:	Saccharomyces cerevisiae (Baker's yeast)
CAEEL:	Caenorhabditis elegans (worm)
STRPU:	Strongylocentrotus purpuratus (Purple sea urchin)
DANRE:	Danio rerio (Zebrafish)
XENTR:	Xenopus tropicalis (Western clawed frog)
TAEGU:	Taeniopygia guttata (Zebra finch)
MYOLU:	Myotis lucifugus (Little brown bat)
HUMAN:	Homo sapiens (Human)

	YEASA	CAEEL	STRPU	DANTE	XENTR	TAEGU	MYOLU
CAEEL	71						
STRPU	73	69					
DANTE	70	66	29				
XENTR	71	68	31	9			
TAEGU	72	67	32	8	5		
MYOLU	72	67	31	6	3	2	
HUMAN	72	67	31	7	3	2	1

to human, bird, and frog [6]. Similarly, yeast cytochrome C is equidistant to all multicellular organisms such as fish, frog, birds, horse, and human [6]. An example of the genetic equidistance phenomenon using the Dot1 protein is shown in Table 1 [9]. The genetic equidistance result has later been confirmed by us for nearly all proteins involving 15 species, e.g., bacteria is equidistant to yeast and human in all 196 protein orthologs that one could find for the three species [9, 10]. Equidistance here does not mean numerically equivalent but does mean that the distance is similar in numerical values. To statistically show the genetic equidistance phenomenon, we have used multiple genes [9, 10]. For example, for a simple species C, its equidistance to two other more complex species B and A can be shown by examining a set of randomly picked genes. The idea is that if approximately half of all proteins show more identity between C and B and the other half show more identity between C and A, this would mean equidistance of C to A and B. The statistical significance of this can be examined by chi-squared test. On the other hand, if significantly more proteins (let's say 90% of proteins examined) show higher identity between C and B than between C and A, this would mean non-equidistance or C is closer to B than to A. No one could have anticipated the genetic equidistance result, and nearly no one today is aware of this result.

Perhaps because of the influence of the modern synthesis that flatly ignores the issue of complexity or an evolutionary progress towards higher complexity as embedded in Lamarck's theory, Margoliash, in first reporting

the equidistance result of cytochrome C, only noted time to be the determining factor of genetic distance: "It appears that the number of residue differences between cytochrome c of any two species is mostly conditioned by the time elapsed since the lines of evolution leading to these two species originally diverged. . . . If elapsed time is the main variable determining the number of accumulated substitutions, it should be possible to estimate roughly the period at which two lines of evolution leading to any two species diverged." [6]. With that, he formally proposed the universal molecular clock hypothesis that different species have approximately the same substitution rate, which implies that genetic distance between species is strictly determined by time alone.

The molecular clock idea as inspired by molecular distance in protein sequence alignment has two claims:

1. Different species have similar substitution rates.
2. The substitution rate of any given gene is constant over time.

Although the two distance results from the above two kinds of alignment can both give rise to the idea of a universal molecular clock and are just two sides of the same coin, the equidistance result however was unusual for two reasons. First, unlike the distance result with humans as the reference, deducing a molecular clock from the equidistance to a least complex reference species was more straightforward involving no calculations and fossil dating. As all more complex species are equidistant in evolutionary time to the least complex reference species, equidistance in molecular divergence must mean equal rates of substitution for all species (Figure 1, under the tacit assumption that the molecular distance is still increasing with time rather than at maximum level). Second, the molecular distance results with humans as the reference has in fact been presented in textbooks as evidence for Darwin's theory or the modern synthesis. The message is that the more similar in phenotypes, the closer the molecular distance (this pattern has many exceptions). However, the equidistance result is in fact completely unexpected from the modern synthesis: how can similar amounts of genetic changes result in vastly different amounts of phenotypic changes? Relative to the phenotypic changes from fish to frog, the changes from fish to human are much greater. Despite being considered by some scientists as 'one of the most astonishing findings of modern science' [11], the genetic equidistance result has largely disappeared from our collective consciousness. It has only been occasionally acknowledged as the key finding in inspiring the molecular clock hypothesis [8].

2.2. The neutral theory of molecular evolution

The molecular clock hypothesis was initially met with great resistance among classical evolutionary biologists as Darwin's theory clearly cannot explain why there should exist similar substitution rates among different species. Nonetheless, some researchers have treated the molecular clock as a genuine reality and have in turn proposed a number of theories to explain it [12, 13, 14, 15, 16, 17]. The neutral theory is known to have accounted for the molecular clock [15, 17], even though it is widely acknowledged to be an incomplete explanation [18, 19]. However, Kimura believes that the best evidence for his neutral theory is the molecular clock [20].

The abstract of the Kimura paper on the neutral theory has only one sentence: "Calculating the rate of evolution in terms of nucleotide substitutions seems to give a value so high that many of the mutations involved must be neutral ones." [15]. But this calculation was the logic of the molecular clock hypothesis, which has two implicit assumptions that were taken for granted, without deliberation. One is that the observed genetic distance always increases with time. The other is that every nucleotide in a genome is freely changeable or mostly neutral (there are no nucleotide positions that would cause lethality when changed). As a matter of fact, the mere use of the equation $r = d/2t$ for deriving the mutation rate is already guilty of begging the question: it has already assumed the conclusion, the same mutation rate, for the two species concerned when it assumes the distance to have been contributed by both equally.

Research concerning neutral processes in genetics started in the 1920s and culminated in Kimura's neutral theory [15, 21]. From the beginning researchers did not seek to deny the importance of natural selection but instead were interested in how neutral processes affected adaptive evolution [22]. Fisher and Wright asked questions and developed techniques of population genetics that are also relevant to the neutral theory [23, 24]. Haldane's genetic load argument was instrumental in inspiring Kimura to develop his neutral theory [25]. Kimura was inspired by the notion that if most variants are not neutral, the genetic load would be too high to be tolerable [15]. In addition, some researchers have suggested that much of molecular evolution is neutral [26, 27]. However, Kimura first combined population genetics theory with molecular evolution data to arrive at a theory of neutral evolution [15]. While the neutral idea has been used unsuccessfully to explain electrophoretic protein polymorphisms or diversities within species [28, 29], the data that directly inspired Kimura was from sequence variation among species.

The neutral theory makes these claims:

1. The observed sequence variations between different species are mostly neutral with no effects on fitness or not under natural selection. DNA segments that are assumed to play no functional roles or not under natural selection are viewed as junk DNA [30]. Some estimation has concluded that the human genome is mostly, 90%, junk DNA [31, 32].
2. Substitution rate equals mutation rate. Evolutionary rate as measured in generation time rather than years is constant and similar among taxa [15, 21].
3. The infinite-site model is part of the neutral framework [33]. This model has been widely used for interpreting observed polymorphisms and for constructing phylogenetic trees. A corollary of the infinite site assumption includes many unrealistic notions such as infinite genetic distance/diversity, no recurrent or back mutations, and no stage of evolution could be at equilibrium. However, this assumption has been invalidated by experimental data [34, 35].
4. The neutral theory classifies mutations into deleterious, neutral, and advantageous. This assumption overlooks the fact that all mutations including advantageous ones have a deleterious aspect as random noises to an ordered system. It does not recognize the fact that most variations appear neutral as a result of balancing selections [36, 37, 38, 39, 40].
5. The neutral theory views non-conservation as non-function [41]. This assumption treats many poorly conserved genomic elements as neutral junk, such as repetitive and viral elements, and is key to the conclusion of only 8% functional genome in humans [31]. However, functional genomics research in recent years has revealed that there are essentially no junk DNAs [42]. All DNA should function in a mechanical code, i.e., the mapping between the local sequence and the local deformability of DNA [43].
6. The neutral theory treats synonymous mutations as neutral [17, 44]. This assumption has now been invalidated by a systematic study comparing the fitness effects of synonymous versus non-synonymous mutations [45].

The neutral theory is also used as an explanatory theory to explain genetic diversity within species. Effective population size is deduced to be related to genetic diversity and often tautologically derived by using the very genetic variables that it is meant to predict or explain [28]. However, it is widely acknowledged that the neutral theory is not a satisfying account

for genetic diversity [46, 47, 48]. Most in the field today merely view the neutral theory as a null model useful for testing if a site is under natural selection. However, researchers interested in building phylogenetic trees still generally assume the neutral theory to be a true account of nature and often unfoundedly present the trees based on the neutral assumption as uncertainty-free, e.g., the out of Africa model of modern human origins [49].

3. The maximum genetic diversity theory of molecular evolution

3.1. The maximum genetic diversity theory

Both the molecular clock and the neutral theory are useful ideas in many respects but are nonetheless flawed or mistaken in their core claims. As shown by studies in the last two decades, the genetic equidistance result in fact remains unexplained by them. The molecular clock interpretation of the genetic equidistance results is in fact a classic tautology, a mere ad hoc restatement of a distance phenomenon [50]. It has not been verified by any independent observation and has on the contrary been contradicted by a large number of observations [19, 50, 51, 52]. Direct measurements of germline mutation rate across vertebrate species have found as large as 40 fold difference in mutation rates among different species [53]. Most in the field today do not believe that there is a universal molecular clock and are comfortable working with a relaxed clock that treats different taxa to have different mutation rates. However, few have realized the great cost of the negation of the universal molecular clock: the original genetic equidistance result would remain unexplained.

If the molecular clock is not a real phenomenon, it is only to be expected that no explanatory theory of nature could account for it. Indeed, the neutral theory has not fully explained the molecular clock even though it was inspired by it [18, 19]. The observed rate is measured in years but the neutral theory predicts a constant rate per generation. Also, the theory predicts that the clock will be a Poisson process, with equal mean and variance of mutation rate. Experimental data have shown that the variance is typically larger than the mean. Ohta's "nearly neutral theory" explained to some extent the generation time issue by observing that large populations have faster generation times and faster mutation rates, but remains unable to account for the great variance issue [54].

In 2005, we fortuitously re-discovered the genetic equidistance result by performing protein alignments using our favorite gene PRDM2 and soon

afterwards realized the existence of the early literature on this result [6]. We further realized that the molecular clock interpretation is merely a tautology [50]. In 2008, we published the maximum genetic diversity (MGD) hypothesis and reinterpreted the genetic equidistance result [55, 56]. We also discovered a new characteristic of the equidistance result, the overlap feature, that has not even been appreciated let alone explained by any theory [57]. Here we summarize the terminology and concepts of the theory.

Microevolution: The term microevolution in the popular theory refers to evolutionary change within a species or small group of organisms, especially over a short period. In our new theory, we use the term to describe evolutionary changes both within a species and between species over both short and long periods so long these changes do not involve major epigenetic changes.

Macroevolution: The term macroevolution in the popular theory applies mainly to the evolution of whole taxonomic groups over long periods of time. In our new theory, we use the term to describe new species formation that involves both genetic and epigenetic changes, particularly involving increases in epigenetic complexity. We use the number of cell types to define and quantify complexity. As the number of cell types is directly related to the number of epigenetic programs (each cell type is a unique epigenetic program) or epigenetic complexity, advances in complexity during macroevolution is thus major epigenetic changes.

Maximum genetic diversity: The central problem of the field has always been the old riddle of what determines genetic diversity [28, 46, 58, 59]. Is it mostly determined by natural selection or neutral drift? The study of variation/diversity has a long history [60]. Understanding the amount and nature of genetic variation has been fundamental to evolutionary research ever since the synthesis of natural selection with genetics in the 1930s and remains so today. For over two decades before 1960s, there were two schools of thought in hot debate, with one believing in very little genetic variation within a population of a species, and the other hypothesizing extensive genetic variation maintained by natural selection [28, 59]. The debate was mostly fruitless because there was little experimental data on genetic variation. In the mid-1960s, the application of protein electrophoresis method showed for Drosophila, humans and other organisms that there were extensive variations in many proteins [28, 59]. The debate shifted to how to best explain such high diversity and the mysteriously narrow range of genetic diversity levels seen across taxa that vary markedly in their census population

size [46]. There were again two main schools of thought that were naturally evolved from the earlier two schools, with the school originally believing in little genetic variation now favoring neutral drift, and the other sticking with natural selection. The debate remains unsettled and has been neglected in the last three decades with the focus of the field shifted to generating more and more diversity data with newer techniques.

Genetic diversity within a species is commonly measured as the nucleotide diversity, defined by Nei and Li in 1979 as the average number of nucleotide differences per site between two DNA sequences in all possible pairs in the sample population [61]. For a kind of organisms with multiple taxa of similar phenotypes, the genetic diversity of the kind with respect to a protein or DNA sequence is the percentage of positions in the sequence that differ among taxa. Genetic distances between species in specific genes were first reported in the early 1960s [5, 6, 7]. The genetic diversity of an individual is commonly measured by the amount of heterozygosity, which is equivalent to the sequence difference or genetic distance between the father and the mother of the individual. It can also be measured by the minor allele content (MAC) or the total amount of minor alleles in an individual genome. Minor alleles are in general less fit and represent deviations from the fit.

Many authors use the term genetic diversity to refer only to genetic variations within a species or population [46, 62]. But many also use the term to refer to genetic variations both within and between populations or species [63]. Genetic diversity among species is commonly measured as genetic distances. As the popular theory of molecular evolution, the neutral theory, was directly inspired by a genetic distance phenomenon between species (genetic equidistance) and its interpretation the molecular clock and is commonly used to interpret the genetic diversity patterns both within and between species, it would be inconsistent if the term genetic diversity covers only genetic variations within a species or population.

The existing concepts of genetic diversity remain unchanged in our new theory. However, we have added to them a new notion, the upper limit concept. Genetic diversity or distance as measured in most genes would increase with time but cannot do so up to 100% nonidentity. The increase would stop at an upper limit level with many sequence positions remain unchanged. The upper limit of the percentage of positions in the sequence that differ among taxa within a kind of organism is called the maximum genetic diversity of the organism. The upper limit of the percentage of positions in the sequence that differ among individuals within a taxon is called the maximum genetic diversity of the taxon.

The maximum genetic diversity measures the fraction of positions in the sequence that are free to change without negatively impacting the core physiology of the members of a taxon or the taxa of a kind. These changes may be beneficial, deleterious, or neutral with regard to fitness depending on circumstances. The positions not free to change are considered conserved and are required for the core physiology of a taxon. Positions that are identical among the compared species are considered as conserved. These unchanged positions consist of two types. The first type includes the positions essential for the barebone or minimal function of a gene. A change in such a position would alter the biochemical activity of the gene in a test tube. The second type includes the positions essential for more complex species but not for simple species. As species become more complex, more positions in a gene will become unchanged or involved in more complex traits. Changes in these positions would not affect the barebone function of a gene or the activity in a test tube or even short-term phenotypes in living organisms, but may affect long term survival of the more complex species but not the simple ones.

The major building blocks for biological organisms are DNA and the architectural plans of how to use the DNA parts are the epigenetic programs. The more the number of cell types, the more the number of ways of using the same set of DNAs, and the more complex the organism [64, 65, 66, 67, 68, 69, 70, 71]. Epigenetic programs are not only inherited during mitotic cell division but are also often transmitted through the germline to the next generation [72, 73, 74, 75, 76, 77]. The epigenetic complexity of a taxon is the average number of cell types of its individual members. Taxa of higher epigenetic complexity are considered more complex. Humans can in most cases intuitively and correctly judge the complexity differences. For unicellular organisms or for species with similar number of cell types, the number of epigenetic genes may be used to infer higher complexity. So yeasts are more complex than bacteria because yeasts have more epigenetic proteins. Yeasts have several histone acetylases and SET domain histone methyltransferases while bacteria have none.

The maximum genetic diversity (MGD) hypothesis contains a pair of self-evident intuitions or axioms [55, 56, 78, 79]. Axiom 1 posits that the more complex the phenotype, the greater the restriction on the choice of molecular building blocks. Complex/ordered system needs higher precision building parts. In biology, this means that there is an inverse relationship between genetic diversity and epigenetic complexity. Complexity is defined as the number of cell types. Axiom 2 says that for any system that can allow a limited level of random errors or noises in molecular building parts, such

errors may be beneficial, deleterious, or neutral depending on circumstances. Limited errors at optimum level are more likely to be beneficial than deleterious because they are after all within tolerable levels and confer economy in construction and strongest possible adaptive capacity or robustness to environmental challenges. Obviously, one only needs to substitute "errors in building blocks" with "genetic diversity" to get the equivalent concept in biology. Axiom 2 in fact highlights the valid parts of Kimura's and Darwin's theories.

Since the original reasoning, we have now come up with two additional self-evident reasons to support the maximum genetic diversity theory that there exists an inverse relationship between genetic diversity and species complexity or higher brain function.

1. Matter or randomness and consciousness or cognition are opposite to each other. High randomness inside the body of an individual hence must result in poor mental function, and the measure of randomness is genetic diversity as diversity originates from random mutational events. Thus, complex species with higher cognitive capacity must have lower randomness or genetic diversity. Here, the meaning of randomness is based on common sense routine experiences and may not be relevant to the quantum world. It is well established that reaction time is correlated with cognition and reflects neurological efficiency [80]. Such efficiency could be harmed by random noises or genetic diversities. In fact, slower reaction time has been found to be associated with autozygosity or the enrichment of deleterious variants [81].

2. Deleterious variants or mutations harmful to a trait could be rescued by other mutations [82]. Populations with greater genetic diversity can more easily rescue or tolerate harmful mutations, which would make it hard for natural selection to maintain the quality of a trait and to eliminate harmful variants. So, suppressing genetic diversity is necessary for maintaining traits at a high quality level and for removing harmful variants.

The maximum genetic diversity hypothesis makes these claims:

1. Maximum genetic diversity tends to be higher for simpler taxa and lower for more complex taxa. The reason is that the members of more complex taxa rely on more sequence positions, which are for that reason conserved, leaving fewer positions free to change. Complex taxa also require higher precision DNA, leaving fewer positions free to change. Furthermore, complex taxa have higher cognition capacity and must suppress randomness or genetic diversity as consciousness and randomness are opposite to each other.

2. Macroevolution is a change in organismal complexity, and most of the time results in an increase in complexity that is mirrored by an increase in the precision of the building parts or a decrease in the allowed range of the standard deviations (stdev) for the parts. Microevolution is an increase in genetic diversity within the allowed stdev ranges without much change in complexity. It covers both evolution within species and evolution from one species to others. At saturation phase, microevolution means genetic turnovers at the equilibrium level of genetic diversity. The gradual evolution of sequences takes place at the microevolution level but cannot be extrapolated to the scale of macroevolution, as Gould and Eldredge had concluded largely on the basis of the fossil record [83]. Note that the definition of macroevolution under the MGD theory is different from the standard definition under the popular theory.
3. The positions that are conserved in simpler taxa tend to also be conserved in more complex taxa. In other words, the positions that are free to change in more complex taxa tend to also be free to change in simpler taxa.
4. Genetic distance among taxa and genetic diversity within a taxon is mostly at the optimum level today after a very long evolutionary time, especially so for fast-evolving sequences [84]. Any level higher or lower than the optimum would be negatively selected. The optimum level is equal to the maximum level and can be defined by comparing the genetic diversity of a control population with those of patient populations. If the genetic diversity of the patient population of a particular disease is greater than that of the control population, while the genetic diversity of the patient population of another disease is smaller than that of the control population, one can infer that the genetic diversity level is at an optimum.
5. Genetic variants are mostly functional or under balancing selection rather than neutral (under both positive and negative selection).
6. Genetic distance or molecular distance between two taxa of different complexity is not contributed equally by mutations in the two lineages but rather is mostly contributed by mutations in the simpler lineage.
7. Non-conservation is not non-function. Fast-changing non-conserved sequences play more important roles in adaptation to the environment than the slowly changing conserved housekeeping genes.
8. Lower MGD means higher homozygosity, which is however very different from the higher homozygosity due to inbreeding. Lower MGD results in higher fitness traits because there are more common alleles

or good alleles becoming homozygous. In contrast, inbreeding leads to lower fitness traits (inbreeding depression) due to homozygosity in minor alleles or deleterious alleles. Inbreeding shows long ROH (runs of homozygosity) but lower MGD does not.

9. The origin of the first life involves a reduction in the randomness of the life-building molecules, which is in principle similar to the reduction in genetic diversity (or randomness) during the step-wise increase in complexity in macroevolution. They all involve the same complexifying force or anti-randomness force.

To use conservation as an index of function is only measuring one of two kinds of sequences, the essential ones for internal system physiology that have little to do with adaptation to the outside environments. To maintain the long-term integrity of the system, such sequences cannot change. For living fossils to be possible, these sequences should be highly stable. On the other hand, sequences involved in adaptation to environments must be fast-changing because environmental changes are usually fast. Flu viruses escape neutralizing antibodies every few years, and the fast-changing non-conserved sites in these viruses are absolutely critical for their survival but not essential for their physiology.

Comparing the claims of the two competing theories, the maximum genetic diversity theory makes the following claims that are the direct opposites of the claims by the neutral theory. 1. Most variants are not neutral. 2. Genetic distances or diversities are mostly at upper limit levels. 3. The infinite site model is unrealistic. 4. Most DNA sequences are under selection and hence not informative to phylogenetic inferences. 5. All mutations have a deleterious aspect. 6. Non-conservation means adaptive function rather than no function. 7. Synonymous substitutions are also functional. 8. Genetic distance or molecular distance between two taxa of different complexity is not contributed equally by mutations in the two lineages. 9. Increase in complexity requires suppression of genetic diversity. 10. Evolution or turn-over of alleles are very fast rather than slow [84].

3.2. Genetic equidistance phenomenon reinterpreted

The maximum genetic diversity hypothesis explains the genetic equidistance phenomenon as a result of maximum genetic distance [55, 56, 78]. Over a long evolutionary time and for fast-evolving DNAs, the genetic distance between species has reached the maximum level. The distance between the ingroup species and a simpler outgroup taxon is mainly determined by the maximum

genetic diversity of the simpler outgroup (Figure 1). This distance is equal to the maximum distance allowed within members of the simpler outgroup, e.g., the distance between humans and fishes equals the maximum distance between different taxa of fishes. Changes in the lineage leading to the simpler outgroup mask any changes in the lineages leading to the ingroup taxa.

This notion that the maximum genetic diversity of a simple kind of outgroup organism determines the distance between the outgroup and the more complex clade can be illustrated by the example of cytochrome c. The maximum diversity in this protein sequence is about 70% difference within bacteria, for example, between Bordetella parapertussis and Paracoccus Versutus. The maximum distance between bacteria and mammals is about 65% difference, such as between Bordetella parapertussis and Pan troglodytes. Within fungi, the maximum diversity is about 40% difference, for example, between Aspergillus oryzae and Yarrowia lipolytica. The maximum distance between fungi and mammals is about 43% difference, such as between Aspergillus oryzae and Pan troglodytes.

There are in fact two kinds of genetic equidistance results. For long evolutionary timescale or for fast-evolving sequences, one would observe "maximum genetic equidistance": different species are equidistant to a species of lower or equal complexity. The original result of Margoliash is maximum genetic equidistance. For short evolutionary timescale or for slow evolving sequences, one observes "linear genetic equidistance" where the molecular clock holds and the distance is still linearly related to time: when ingroup species have similar mutation rates, they would be equidistant to a lower or equal complexity outgroup.

This explanation of the genetic equidistance result by the MGD theory can also be easily illustrated by a simple thought experiment. If we can create a yeast, a fish, and a human being by using identical genes for their shared homologs and let the three organisms diverge for an infinite amount of time or about 500 million years with each organism remains phenotypical largely the same as today, a gene in yeast would have changed a lot to a maximum of, say 50%, while its homolog in fish would have changed to a maximum of, say, 30%, and its homolog in human would have changed very little, say less than 1%. Any more changes than 50% would be lethal to yeast; any more changes than 30% would be lethal to fishes; and any more changes than 1% would be lethal to humans. The reason that a gene in yeast can change much more than in fish, which is still more than in human, is because a gene in human encounters far more functional constraints than its homolog in fish or in yeast. Thus the genetic distance between yeast and human or fish is mainly determined by the mutations in yeast. In this

case, the 50% change in yeast would account for the genetic distance of 50% identity between yeast and human or between yeast and fish, as well as 50% identity in within species distance in yeast. The 30% change in fish would account for the genetic distance of 30% identity between fish and human. In contrast, the neutral theory would predict that both human and fish can also, like yeast, change up to 50% or more and would have a genetic distance of 50% identity.

3.3. Testing the two different interpretations of genetic equidistance

There are two ways to test which of the two different interpretations of genetic equidistance is true, the molecular clock versus the maximum genetic diversity theory. The first is the overlap feature [57, 85]. When aligning a protein sequence from three taxa, the conserved positions would show identical amino acids for all taxa. The non-conserved positions include two different types. One type shows mutation in only one of the three taxa while the other two taxa have identical amino acids or are non-mutated. The other type shows that each taxon has its own unique amino acid at the same position. This indicates that at least two taxa have each had a unique mutation at that same position, and hence their mutations overlapped at that position. Such position is termed an overlap position. As one mutation is enough to lead to a difference of one between two taxa, overlapped mutations at the same position do not increase molecular distances and are hence hallmarks of mutation saturation. At maximum saturation, each free-to-change position in a taxon would all have mutated, and thus there would be a higher fraction of overlap positions among the non-conserved positions than expected by chance. In fact, the original equidistance result of Margoliash shows a high fraction of overlap positions that are consistent with the predictions of the maximum genetic diversity theory but far more than that by the molecular clock or neutral theory [9, 57].

The other way to test the molecular clock versus the maximum genetic diversity theory is the genetic non-equidistance result despite equidistance in time. The maximum genetic diversity theory predicts that maximum equidistance would only result when the outgroup is less complex than the sister species. If the outgroup is more complex, then its maximum distance with the ingroup sister species would be determined by the MGD of each of the ingroup species, which may not be the same for all the ingroup species. However, the molecular clock would predict genetic equidistance to the outgroup

regardless if the outgroup taxon is more or less complex. For example, human is the outgroup to the Sauropsida clade containing snake and bird. The molecular clock predicts that humans should be equidistant to snakes and birds in protein sequence. However, the maximum genetic diversity theory predicts that birds should be closer to humans than snakes should because birds should be more complex than snakes. The actual data in fact validated the maximum genetic diversity theory [3, 86, 87]. Genetic non-equidistance to humans despite equidistance in time has also been found for sister species within the teleost fish clade, the arthropod phylum, the Porifera phylum, and the fungi kingdom. In all five cases where the difference in complexity of the ingroup sister species can be inferred (octopus vs. cockle, terebratulina vs. lingula, bird vs. snake, dragonfly vs. louse, and smut vs. yeast), the more complex species always shows greater sequence similarity to humans in fast-evolving genes, fully conforming to the predictions of the maximum genetic diversity theory but not that of the molecular clock [86]. Also, by whole genome sequencing analysis, two new world monkeys are found to be non-equidistant in nucleotide sequence to humans with the most primitive monkey marmoset to be more distant to humans than the owl monkey [88].

For species within a clade that are of similar degree of complexity, it is also possible to observe genetic non-equidistance to a more complex outgroup. The sequence of a more complex taxon and any of its closely related sequences should be well tolerated by the lower clade (but not vice versa) and so may exist as a part of the normal variation of the lower clade. Therefore, within a lower clade, certain taxon may by chance happen to have a genome more similar to the higher outgroup than another taxon. We tested and confirmed this expectation by using the cytochrome c protein. For the mollusc clade, the Asian clam showed lower protein nonidentity to human than the Yesso scallop (Table 2, 17% vs 33% nonidentity). However, this is a chance phenomenon that cannot be consistently observed for most genes. Thus, for the mitochondrial ND1 gene, the Asian clam showed slightly higher nonidentity to human than the Yesso scallop (data not shown). Likewise, by chance, it is even possible for a unicellular species like choanoflagellate to show slightly higher sequence similarity to a more complex species such as scallops and insects than a simple metazoan like amphioxus (Table 2). As expected, the genetic equidistance phenomenon still holds in these cases, e.g., choanoflagellate is approximately equidistant (24–33% nonidentity) to various metazoan taxa (sponge, cnidaria, insect, mollusc, lamprey, zebrafish, and mammal) and so is amphioxus (18–22% nonidentity) to different vertebrates (lamprey, zebrafish, and human, Table 2). Lamprey is approximately equidistant (12–19% nonidentity) to different jawed-vertebrates (mammal and zebrafish, Table 2). These results further confirmed the genetic equidistance

Table 2: Percentage differences among species in cytochrome c

Asian clam	Corbicula fluminea	$APY24038.1$
Yesso scallop	Mizuhopecten yessoensis	$XP\_021373283.1$
Choanoflagellate	Monosiga brevicollis MX1	$XP\_001748907.1$
Amphioxus	Branchiostoma floridae	$XP\_035688830.1$
Sponge	Oopsacas minuta	$KAI6652077.1$
Cnidaria	Actinia tenebrosa	$XP\_031569134.1$
Insect	Holotrichia oblita	$KAI4469925.1$
Mammal	Homo sapiens	$NP\_061820.1$
Zebrafish	Danio rerio	$NP\_001002068.1$
Lamprey	Petromyzon marinus	$XP\_032831412.1$

	Mammal	E. Scallop	A. clam	Sponge	Cnidaria	Insect	Lamprey	Zebrafish
Asian clam	17	31		20	19	31	18	18
Yesso scallop	33		31	35	34	29	28	28
Choanoflagellate	25	33	24	26	26	27	29	26
Amphioxus	22	35	22	17	17	28	22	18
Lamprey	19	28	18	24	23	29		12

phenomenon by using a larger variety of traits or complexities. A caveat here is that our analysis involved just one protein (cytochrome c) and was intended to give a flavor of what the data may look like for the original protein used to discover the genetic equidistance phenomenon among some of the species that have yet to be examined. More analyses involving more genes and more species, similar to what Luo and Huang have done [9], need to be done in the future.

3.4. What does the universal molecular clock really mean?

The universal molecular clock is really about the constant rate of complexity increases. The first molecular evidence for the constant albeit discontinuous advances in complexity is in fact the maximum genetic equidistance phenomenon. Defying the Darwinian gradualistic worldview, there have always been researchers since the time of Darwin who have appreciated the discontinuous nature of macro-evolutionary changes [11, 60, 83, 89, 90, 91, 92]. Underlying the direction towards higher complexity and the constant rate of increase in complexity may be the mysterious 'complexifying force' first proposed by Lamarck.

3.5. Is the maximum genetic diversity theory useful?

The maximum genetic diversity theory has been instrumental in directing productive research on both evolutionary problems and important biomedical problems. The theory does not mean discarding the old assumptions but merely making them more limited in their scopes. One must carefully select those DNAs that may follow those assumptions.

Phylogenetics The maximum genetic diversity theory should help resolve difficult historical problems such as the phylogenetic tree of life. Past methods have no concept of maximum distance and use mostly non-informative distance data for inferring phylogeny. They often produce self-conflicting results and results inconsistent with the fossil records and morphology [93]. We developed the slow clock method based on the maximum genetic diversity theory [86]. The method makes use of only slow evolving sequences that have fewer overlap positions, and thus ensures the linear relationship between distance and time. Its results therefore will be more objective and independent of the variations in sequence selections and investigators. In keeping with the intuitive logic, important genes mutate less often [94]. Slowly evolving sequences are more likely to meet the neutral criteria and in fact are indeed more neutral [85]. They are unlikely to be under positive selection since their low speed of change makes them too slow to meet adaptive needs. Relative to fast-evolving sequences, they are also less likely to be under negative selection by way of collective effects over an MGD threshold since their low speed means that they are not a major contributor to the collective effects of variants. Also, such sequences are unlikely to be under pressure to reduce their tolerable number of changeable positions as a result of complexity increases, because their slow speed of mutations means that they are less likely to be disruptive to increased complexity. Thus, their MGD levels are more likely to be similar in different species. We have used the slow clock method to re-establish a correct primate phylogeny that re-establishes the intuitive common sense that humans and pongids are two separate groups, which has long been the consensus view of paleoanthropologists [86, 95]. Orangutans have less reasoning ability than chimpanzees and humans [96]. Supporting our results, a Miocene ape fossil of 11.6 million years ago in Europe, Danuvius guggenmosi, showed bipedal features, consistent with a separation of the homo lineage from the apes earlier than 12 million years ago [97]. Also, new findings push back the oldest evidence of C4 grass–dominated habitats in Africa—and globally—by more than 10 million years to 17–21 million years ago [98, 99]. Such habitats are believed by the popular Savanna theory to have favored an upright posture and selected for bipedalism.

To truly neutral sequences still at the linear phase of divergence, many of the assumptions of the neutral theory such as the infinite sites model would be valid. Thus phylogenetics research can largely proceed as before except that one has now a standard to separate the neutral from the noninformative DNAs. One must now distinguish two different kinds of high sequence similarity, one due to less time of separation and the other because of common construction resulting in using similar parts (convergent evolution).

The out of Africa model of modern human origins is based on the molecular clock and the neutral theory [49, 100]. The high genetic diversity of Africans is interpreted to mean a deeper evolutionary time for Africans if one assumes the molecular clock [49, 100]. Also, the infinite site model is assumed in order to infer the derived allele status, which is critical for rooting the phylogenetic tree in Africans by using the outgroup rooting method [101, 103]. However, both of these assumptions are invalid according to the maximum genetic diversity theory and experimental data [34, 35]. By using informative variants and allowing recurrent and back mutations, we have built a new model of modern human origins, the out of East Asia model [104, 105]. The out of East Asia model is consistent with the multiregional model in terms of autosomal evidence, which indicates that the major races have separated for 2 million years as originally claimed by the multiregional model [106]. However, uniparental DNA data indicates a single origin in East Asia at a more recent time. Others have also found that all Eurasian Y chromosome and X chromosome may have originated in East Asia [107, 108]. The likely scenario is that modern humans first evolved in East Asia as marked by a new modern version of uniparental DNAs and then migrated to Europe and Africa and admixed with local less modern people. Admixture led to replacement of uniparental DNAs and autosomal DNAs so that Europeans or Africans would have modern uniparental DNAs but largely local autosomal DNAs. A real example of this scenario is the Saami people in Finland who have East Asian Y chromosome haplotype N but European autosomes [109]. Ancient human DNA should be very informative in falsifying the incorrect models. Our analysis of ancient DNA samples have confirmed the out of East Asia model and invalidated the out of Africa model [105, 110]. In contrast, researchers who believe in the out of Africa model have yet to report any ancient DNA evidence for their model but have instead found support for the out of East Asia model, i.e., ancient DNA samples of 40000–45000 year old found in Europe and East Asia are East Asian like rather than African like [111, 112].

Biomedical problems Most complex traits and diseases are partly inheritable and presumably caused by polymorphic genetic variations such as SNPs. The neutral theory views most such variations to be nonfunctional and neutral and hence the study of complex traits and diseases has in the past focused on searching for a few functional variants. Although such GWAS studies have met some successes in identifying a number of variants, these variants account for only a small fraction of the total trait variation and

their functional roles typically remain unclear. The maximum genetic diversity theory predicts that complex diseases may be caused by excess genetic noise over a threshold and may serve to prevent an infinite increase in genetic diversity. Complex traits evolved as a result of suppressing genetic noises and hence should be susceptible to damage by excess noises. Also, insufficient amount of genetic diversity may hurt adaptive capacities such as immunity. The quantitative variations in a complex trait may correlate with the number of genetic variations.

Results from our efforts in testing the maximum genetic diversity theory have shown the expected that higher minor allele contents (MAC) or noises correlate with many complex diseases. These include association of MAC with higher lung cancer incidence in mice and humans [36, 39, 113]. Also, Parkinson's disease patients have higher MAC than controls and a selected set of 37000 minor alleles can predict 2% of Parkinson's patients [37]. Other diseases that show higher MAC include Schizophrenia, Type 1 diabetes, Type 2 diabetes, lung cancer, and Alzheimer's [113, 114, 115, 116, 117]. An efficient method of identifying target genes of complex traits has been established using the MAC concept [40]. MAC may control traits by regulating a set of target genes whose expressions are associated with both MAC and traits [40].

To directly examine the self-evident antagonistic relationship between cognition or consciousness and randomness or genetic diversity, we have performed a study analyzing the genotype and phenotype data from more than 400,000 people in the UK [118]. We calculated multiple measures of genetic diversity for each individual, and examined which traits these measures were associated with using linear regression analysis that has controlled for confounding factors. Among the 17 traits examined, only educational attainment, which is highly correlated with cognition or IQ, has the most robust relationship with genetic diversity, and it is an inverse association. This association is likely to be causal, since only the brain-expressed genes, but not the brain-non-expressed genes, showed an association. This result is likely to be free from the interference of confounding factors, because the correlation of non-synonymous variants is significantly higher than that of synonymous variants or intronic variants. Consistently, animal studies have also revealed an inverse relationship between learning and memory and genetic diversity [36, 39]. Therefore, the highest level of genetic diversity of the San people in Southern Africa may actually be the reason for their lowest cognition or civilization level and strong immunity [80, 119, 120, 121], and may have little to do with long evolutionary time or being human ancestors. Low cognition abilities are subject to natural selection pressure, and so genetic diversity

must be also under natural selection rather than being mainly time-related and not subject to natural selection as assumed by the molecular clock and neutral theory.

Several articles have studied the relationship between autozygosity or runs of homozygosity (ROH) caused by inbreeding and various traits and diseases, and found that ROH is negatively correlated with educational attainment [81, 122, 123]. The finding is consistent with the MGD theory because autozygosity caused by inbreeding means that the overall level of harmful variants or minor alleles is high. Because there is no correlation between the heterozygosity in the normal range and ROH (even if the heterozygosity is very different between two individuals, the ROH can be similar), our finding of the inverse correlation between heterozygosity and cognition involves mostly the variations in heterozygosity in the normal range and so is not related to ROH and not inconsistent with the finding of an inverse correlation between ROH and cognition [118]. Indeed, when we removed samples with very low heterozygosity (probably from inbreeding), we saw a stronger inverse correlation between heterozygosity and educational attainment, indicating that also in our analysis, there was an inverse correlation between autozygosity due to inbreeding and educational attainment (Wang and Huang, unpublished). There is little difference in ROH between the normal populations of different racial groups (the population without inbreeding). So, autozygosity or ROH cannot explain the difference in cognition of different racial groups. However, there are significant differences in heterozygosity among the normal populations of different racial groups, which can explain the observed differences in cognition among these racial groups.

It is known that males are more variable than females in intelligence test scores or overrepresented at both high and low levels of performance when the average scores of males and females are similar [124]. This remains poorly understood. The maximum genetic diversity provides a straightforward explanation. Males have just one X chromosome and so fewer heterozygotes than females when autosome heterozygosities of both sexes are the same. The males' more homozygous genomes can be either the best or the worst, depending on whether it's the good or bad alleles that are homozygous.

With regard to very small timescales and small progresses in complexity, can the MGD theory and the genetic equidistance phenomenon still hold? Within humans, there are many different population or ethnic groups. It has been reported in the literature that Africans have lower brain volumes and IQ scores than Europeans who in turn have lower values than East Asians [80]. This correlates well with the genetic diversity levels of these ethnic

groups. As we have shown that genetic diversity inversely correlates with educational attainment, genetic diversity in humans is in fact at the maximum level [118]. Therefore, one should expect to see the genetic equidistance phenomenon among human ethnic groups, i.e., Africans should be equidistant to Europeans and East Asians and the distance between Africans and non-Africans should be equal to the maximum distance between Africans. This is indeed the case for the fast-evolving DNAs or a randomly selected set of genomic SNPs [104]. In contrast, for slowly evolving DNAs yet to reach maximum genetic diversity, the distance between Africans is smaller than that between Africans and non-Africans, which is to be intuitively expected. The dramatic difference here between fast and slowly evolving DNAs is strong evidence validating the maximum genetic diversity theory.

Why sex? Almost all eukaryotes reproduce sexually, through a meiosis which generates haploid gametes from a diploid cell. The purpose of sex has long remained a mystery. The common explanation is that sexual reproduction increases genetic diversity [125, 126]. However, asexual organisms such as bacteria generally have much higher genetic diversity than eukaryotes. There is also the suggestion that sexual reproduction can remove chromosomal and epigenetic abnormalities or other deleterious mutations [126]. However, such abnormalities could also be removed by natural selection of abnormal phenotypes.

The maximum genetic diversity theory offers a straight-forward solution to the mystery of sex. According to the theory, macroevolution from a simple taxon to a higher complexity taxon requires a reduction in genetic diversity (at the nucleotide level). The reduction in genetic diversity in an individual of the simple taxon is necessary for the individual to become the incipient individual of the more complex new taxon. As the overall level of genetic variation in an offspring is mostly determined by the inheritance of the combination of single nucleotide variants carried by the parents, sexual reproduction can either increase or decrease the genetic variation in an offspring relative to the parents but asexual reproduction can only increase the genetic variation in an offspring. Thus, sexual reproduction is essential for reducing genetic diversity necessary for the macroevolution of higher complexity.

3.6. Future directions

To further establish the maximum genetic diversity theory, future studies may focus on several major areas. First, large scale functional genomics studies may show that nearly all bases are functional, which would provide direct

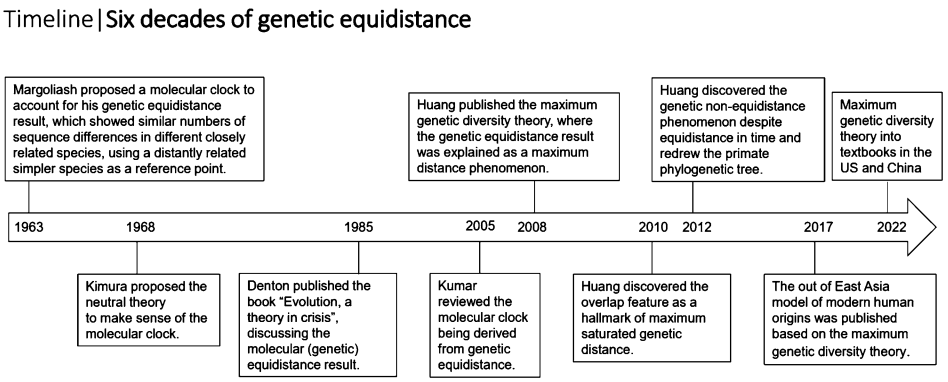

Figure 2: Time line regarding the genetic equidistance phenomenon.

evidence invalidating the neutral theory and confirming the MGD theory. Second, more studies on more populations or species need to be performed to show that genetic diversities within and among species are at the upper limit level. Third, the genetic equidistance phenomenon needs to be further established by studying more species and more genes. As this phenomenon directly inspired both the neutral theory and the MGD theory, it would nullify both these theories if it turns out to be unreal. Fourth, more ancient DNA studies may further invalidate the out of Africa model and confirm the out of East Asia model. By confirming the phylogenetic tree based on the MGD theory, the theory itself gets validated as well. Fifth, along the same line, more findings regarding Miocene hominid fossils may help confirm the human-pongid separation and falsify the human-chimpanzee clade, thereby validating phylogenetic methods and results based on the MGD theory. Sixth, better understanding of cognition may further establish the MGD concept that random noise is detrimental to order or cognition. Finally, the collective effects of a large number of variants on traits need to be better understood at the molecular level.

4. Conclusion

The genetic equidistance phenomenon is the most astonishing finding in molecular evolution (Figure 2). One cannot understand evolution without understanding molecular evolution. One cannot understand molecular evolution without understanding the genetic equidistance phenomenon. This phenomenon was first interpreted by the molecular clock hypothesis and the neutral theory, and later re-interpreted by the maximum genetic diversity

theory. Direct tests of the predictions of these two competing theoretical frameworks have invalidated the molecular clock and the neutral theory and confirmed the maximum genetic diversity theory. Genetic distances or diversities today are mostly at optimum equilibrium. This new understanding may shed new light on many evolutionary questions.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 81171880) and the National Basic Research Program of China (Grant No. 2011CB51001). This paper is dedicated to the 60th year anniversary of the discovery of the genetic equidistance phenomenon. We thank the reviewers for the very insightful comments that have helped to improve the manuscript.

Funding statements

The author was supported by grants from the National Natural Science Foundation of China (Grant No. 81171880) and the National Basic Research Program of China (Grant No. 2011CB51001).

Competing interests

The author has no competing interests to declare.

References

[1] Lewontin, R., What is Evolutionary Theory, 2003. https://www.youtube.com/watch?v=n6W_FzjaKlw.

[2] Newman, S., Remembering Richard Lewontin. Biopolitical Times, Center for Genetics and Society, 2021. https://www.geneticsandsociety.org/biopolitical-times/remembering-richard-lewontin-1929-2021.

[3] Bickel, D., Testing hypotheses of molecular evolution, in: Phylogenetic Trees and Molecular Evolution. Springer Briefs in Systems Biology, 2022: Springer, Cham.

[4] Huang, S. and Z. Zhu, Systems Biology and Evolutionary Theory, Systems Biology, ed. Zhu, Z., R. Zhang, and L. Wang: 2022, Southeast University Press, Nanjing.

[5] Zuckerkandl, E. and L. Pauling, Molecular disease, evolution, and genetic heterogeneity, in: Horizons in Biochemistry, ed. M. Kasha and B. Pullman, 1962: New York: Academic Press.

[6] Margoliash, E., Primary structure and evolution of cytochrome c. Proc. Natl. Acad. Sci., 1963. 50: 672–679.

[7] Doolittle, R.F. and B. Blombaeck, Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. Nature, 1964. 202: 147–152.

[8] Kumar, S., Molecular clocks: four decades of evolution. Nat Rev Genet, 2005. 6(8): 654–662.

[9] Luo, D. and S. Huang, The genetic equidistance phenomenon at the proteomic level. Genomics, 2016. 108(1): 25–30.

[10] Yuan, D. and S. Huang, Genetic equidistance at nucleotide level. Genomics, 2017, 109: 192–195.

[11] Denton, M., Evolution: a theory in crisis. 1985, Chevy Chase, MD: Adler & Adler.

[12] Van Valen, L., Molecular evolution as predicted by natural selection. J. Mol. Evol., 1974. 3: 89–101.

[13] Clarke, B., Darwinian evolution of proteins. Science, 1970. 168(934): 1009–1011.

[14] Richmond, R.C., Non-Darwinian evolution: a critique. Nature, 1970. 225(5237): 1025–1028.

[15] Kimura, M., Evolutionary rate at the molecular level. Nature, 1968. 217(5129): 624–626.

[16] Kimura, M. and T. Ohta, On the rate of molecular evolution. J. Mol. Evol., 1971. 1: 1–17.

[17] King, J.L. and T.H. Jukes, Non-Darwinian evolution. Science, 1969. 164: 788–798.

[18] Ayala, F.J., Molecular clock mirages. BioEssays, 1999. 21: 71–75.

[19] Pulquerio, M.J. and R.A. Nichols, Dates from the molecular clock: how wrong can we be? Trends Ecol Evol, 2007. 22(4): 180–184.

[20] Kimura, M. and T. Ohta, Protein polymorphism as a phase of molecular evolution. Nature 1971. 229: 467–479. MR0862479

[21] Kimura, M., The Neutral Theory of Molecular Evolution. 1983, Cambridge: Cambridge University Press.

[22] Leigh, E.G., Jr., Neutral theory: a historical perspective. J Evol Biol, 2007. 20(6): 2075–2091.

[23] Fisher, R.A., The Genetical Theory of Natural Selection. 1930, Oxford, U.K.: Oxford University Press. MR1785121

[24] Wright, S., The roles of mutation, inbreeding, crossbreeding and selection in evolution, in: Proc. Sixth Int. Congr. Genet., 1932. 1: 356–366.

[25] Haldane, J.B.S., The cost of natural selection. J. Genet., 1957. 55: 511–524.

[26] Sueoka, N., On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA, 1962. 48: 582–592.

[27] Freese, E., On the evolution of base composition of DNA. J. Theor. Biol., 1962(3): 82–101.

[28] Lewontin, R.C., The Genetic Basis of Evolutionary Change. 1974, New York and London: Columbia University Press.

[29] Lewontin, R.C. and J.L. Hubby, A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. Genetics, 1966. 54(2): 595–609.

[30] Palazzo, A.F. and T.R. Gregory, The case for junk DNA. PLoS Genet, 2014. 10(5): e1004351.

[31] Rands, C.M., et al., 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. PLoS Genet, 2014. 10(7): e1004525.

[32] Doolittle, W.F. and T.D.P. Brunet, On causal roles and selected effects: our genome is mostly junk. BMC Biol, 2017. 15(1): 116.

[33] Kimura, M. and J.F. Crow, The number of alleles that can be maintained in a finite population. Genetics, 1964. 49: 725–738. MR0862480

[34] Demeulemeester, J., et al., Biallelic mutations in cancer genomes reveal local mutational determinants. Nat Genet, 2022. 54(2): 128–133.

[35] Harpak, A., A. Bhaskar, and J.K. Pritchard, Mutation rate variation is a primary determinant of the distribution of allele frequencies in humans. PLoS Genet, 2016. 12(12): e1006489.

[36] Yuan, D., et al., Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. Sci China Life Sci, 2014. 57(9): 876–888.

[37] Zhu, Z., et al., Enrichment of minor alleles of common SNPs and improved risk prediction for Parkinson's disease. PLoS ONE, 2015. 10(7): e0133421.

[38] Zhu, Z., et al., Collective effects of SNPs on transgenerational inheritance in Caenorhabditis elegans and budding yeast. Genomics, 2015. 106(1): 23–29.

[39] Yuan, D., et al., Minor alleles of common SNPs quantitatively affect traits/diseases and are under both positive and negative selection. arXiv:1209.2911, 2012.

[40] Zhu, Z., et al., Collective effects of common SNPs in foraging decisions in Caenorhabditis elegans and an integrative method of identification of candidate genes. Sci. Rep., 2015. doi: 10.1038/srep16904.

[41] Kimura, M. and T. Ohta, On some principles governing molecular evolution. Proc Natl Acad Sci USA, 1974. 71(7): 2848–2852.

[42] Quinodoz, S.A., et al., RNA promotes the formation of spatial compartments in the nucleus. Cell, 2021. 184(23): 5775–5790, e30.

[43] Basu, A., et al., Deciphering the mechanical code of the genome and epigenome. Nat Struct Mol Biol, 2022. 29(12): 1178–1187.

[44] Kimura, M., Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet Res, 1968. 11(3): 247–269.

[45] Shen, X., et al., Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature, 2022. 606(7915): 725–731.

[46] Leffler, E.M., et al., Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol, 2012. 10(9): e1001388.

[47] Buffalo, V., Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. Elife, 2021. 10.

[48] Kern, A.D. and M.W. Hahn, The neutral theory in light of natural selection. Mol Biol Evol, 2018. 35(6): 1366–1371.

[49] Cann, R.L., A.C. Stoneking, and A.C. Wilson, Mitochondrial DNA and human evolution. Nature, 1987. 325: 31–36.

[50] Huang, S., The genetic equidistance result of molecular evolution is independent of mutation rates. J. Comp. Sci. Syst. Biol., 2008. 1: 092–102.

[51] Avise, J.C., Molecular Markers, Natural History and Evolution. 1994, New York, NY: Springer.

[52] Nei, M. and S. Kumar, Molecular Evolution and Phylogenetics. 2000, New York: Oxford University Press.

[53] Bergeron, L.A., et al., Evolution of the germline mutation rate across vertebrates. Nature, 2023. 615(7951): 285–291.

[54] Ohta, T., Slightly deleterious mutant substitutions in evolution. Nature, 1973. 246(5428): 96–98.

[55] Huang, S., Inverse relationship between genetic diversity and epigenetic complexity. Nature Precedings 2009. doi: 10.1038/npre.2009.1751.2.

[56] Huang, S., Histone methylation and the initiation of cancer, in: Cancer Epigenetics, ed. T. Tollefsbol. 2008, New York: CRC Press.

[57] Huang, S., The overlap feature of the genetic equidistance result, a fundamental biological phenomenon overlooked for nearly half of a century. Biological Theory, 2010. 5: 40–52.

[58] Aquadro, C.F., Why is the genome variable? Insights from Drosophila. Trends Genet, 1992. 8(10): 355–362.

[59] Lewontin, R.C., Twenty-five years ago in Genetics: electrophoresis in the development of evolutionary genetics: milestone or millstone? Genetics, 1991. 128(4): 657–662.

[60] Bateson, W., Materials for the Study of Variation Treated with Especial Regard to Discontinuity in the Origin of Species. 1894, London: Macmillan.

[61] Nei, M. and W.H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA, 1979. 76: 5269–5273.

[62] Ellegren, H. and N. Galtier, Determinants of genetic diversity. Nat Rev Genet, 2016. 17(7): 422–433.

[63] Nevo, E., Genetic diversity, in: Encyclopedia of Biodiversity, ed. S.A. Levin. 2001: Elsevier Inc.

[64] Yang, J., R. Lusk, and W.H. Li, Organismal complexity, protein complexity, and gene duplicability. Proc Natl Acad Sci USA, 2003. 100(26): 15661–15665.

[65] Levine, M. and R. Tjian, Transcription regulation and animal diversity. Nature, 2003. 424(6945): 147–151.

[66] Vinogradov, A.E. and O.V. Anatskaya, Organismal complexity, cell differentiation and gene expression: human over mouse. Nucleic Acids Res, 2007. 35(19): 6350–6356.

[67] Bonner, J.T., Perspective: the size-complexity rule. Evolution, 2004. 58: 1883–1890.

[68] Carroll, S.B., Chance and necessity: the evolution of morphological complexity and diversity. Nature, 2001. 409(6823): 1102–1109.

[69] McShea, D.W., Metazoan complexity and evolution: is there a trend? Evolution, 1996. 50: 477–492.

[70] Bonner, J.T., The Evolution of Complexity. 1988, Princeton, NJ: Princeton University Press.

[71] Vogel, C. and C. Chothia, Protein family expansions and biological complexity. PLoS Comput Biol, 2006. 2(5): e48.

[72] Remy, J.J., Stable inheritance of an acquired behavior in Caenorhabditis elegans. Curr Biol, 2010. 20(20): R877–R878.

[73] Anway, M.D., et al., Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science, 2005. 308(5727): 1466–1469.

[74] Hitchins, M.P., et al., Inheritance of a cancer-associated MLH1 germline epimutation. N Engl J Med, 2007. 356(7): 697–705.

[75] Cropley, J.E., et al., Germ-line epigenetic modification of the murine A vy allele by nutritional supplementation. Proc Natl Acad Sci USA, 2006. 103(46): 17308–17312.

[76] Huypens, P., et al., Epigenetic germline inheritance of diet-induced obesity and insulin resistance. Nat Genet, 2016. 48(5): 497–499.

[77] Fitz-James, M.H. and G. Cavalli, Molecular mechanisms of transgenerational epigenetic inheritance. Nat Rev Genet, 2022. 23(6): 325–341.

[78] Hu, T., et al., The genetic equidistance result, misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. Sci China Life Sci, 2013. 56: 254–261.

[79] Huang, S., New thoughts on an old riddle: What determines genetic diversity within and between species? Genomics, 2016. 108(1): 3–10.

[80] Lynn, R., Race Differences in Intelligence. 2006, Augusta, GA: Washington Summit Publishers.

[81] Clark, D.W., et al., Associations of autozygosity with a broad range of human phenotypes. Nat Commun, 2019. 10(1): 4957.

[82] Parts, L., et al., Natural variants suppress mutations in hundreds of essential genes. Mol Syst Biol, 2021. 17(5): e10138.

[83] Gould, S.J. and N. Eldredge, Punctuated equilibrium comes of age. Nature, 1993. 366(6452): 223–227.

[84] Rudman, S.M., et al., Direct observation of adaptive tracking on ecological time scales in Drosophila. Science, 2022. 375(6586): eabj7484.

[85] Wang, M., et al., Enrichment in conservative amino acid changes among fixed and standing missense variations in slowly evolving proteins. PeerJ, 2020. 8: e9983. doi: 10.7717/peerj.9983.

[86] Huang, S., Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers. Sci China Life Sci, 2012. 55: 709–725.

[87] Bickel, D., A generalization of null hypothesis significance testing with applications to replication failures, molecular evolution models, and bounded parameter spaces. Zenodo, 2021. doi: 10.5281/zenodo.5123388.

[88] Mao, Y., et al., Structurally divergent and recurrently mutated regions of primate genomes. bioRxiv, 2023. doi: 10.1101/2023.03.07.531415.

[89] Jenkin, F., The origin of species. The North British Review, 1867. 46: 277–318.

[90] Goldschmidt, R., The Material Basis of Evolution. 1940, New Haven, Connecticut: Yale University Press.

[91] Forsdyke, D.R., Evolutionary Bioinformatics. 2011, New York: Springer. MR3015081

[92] Heng, H.H., Genome Chaos: Rethinking Genetics, Evolution, and Molecular Medicine. 2019, Academic Press Elsevier: Cambridge, MA, USA.

[93] Oyston, J.W., et al., Molecular phylogenies map to biogeography better than morphological ones. Commun Biol, 2022. 5(1): 521.

[94] Monroe, J.G., et al., Mutation bias reflects natural selection in Arabidopsis thaliana. Nature, 2022. 602(7895): 101–105.

[95] Martin, M.K., The Wrong Ape for Early Human Origins: The Chimpanzee as a Skewed Ancestral Model. 2023, Maryland, USA: Lexington Books.

[96] Herrmann, E., V. Wobber, and J. Call, Great apes' (Pan troglodytes, Pan paniscus, Gorilla gorilla, Pongo pygmaeus) understanding of tool functional properties after limited experience. J Comp Psychol, 2008. 122(2): 220–230.

[97] Bohme, M., et al., A new Miocene ape and locomotion in the ancestor of great apes and humans. Nature, 2019. 575(7783): 489–493.

[98] Peppe, D.J., et al., Oldest evidence of abundant C(4) grasses and habitat heterogeneity in Eastern Africa. Science, 2023. 380(6641): 173–177.

[99] MacLatchy, L.M., et al., The evolution of hominoid locomotor versatility: Evidence from Moroto, a 21 Ma site in Uganda. Science, 2023. 380(6641): eabq2835.

[100] Johnson, M.J., et al., Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol, 1983. 19(3-4): 255–271.

[101] Ingman, M., et al., Mitochondrial genome variation and the origin of modern humans. Nature, 2000. 408: 708–713.

[102] Li, J.Z., et al., Worldwide human relationships inferred from genome-wide patterns of variation. Science, 2008. 319: 1100–1104.

[103] Underhill, P.A., et al., Y chromosome sequence variation and the history of human populations. Nat Genet, 2000. 26(3): 358–361.

[104] Yuan, D., et al., Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. bioRxiv, 2019. doi: 10.1101/101410.

[105] Xia, Z., et al., Ancient uniparental DNAs in distinguishing the competing theories of molecular evolution and modern human origins. Research Square, 2021. doi: 10.21203/rs.3.rs-1087730/v1.

[106] Wolpoff, M.H., X.Z. Wu, and A.G. Thorne, Modern homo sapiens origins: a general theory of hominid evolution involving the fossil evidence from east Asia, in: The Origins of Modern Humans: A World Survey of

the Fossil Evidence, ed. F.H. Smith and F. Spencer. 1984, New York: Alan R. Liss.

[107] Hallast, P., et al., A Southeast Asian origin for present-day non-African human Y chromosomes. Hum Genet, 2021. 140(2): 299–307.

[108] Skov, L., et al., Extraordinary selection on the human X chromosome associated with archaic admixture. Cell Genom, 2023. 3(3): 100274.

[109] Tambets, K., et al., The western and eastern roots of the Saami – the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am J Hum Genet, 2004. 74(4): 661–682.

[110] Zhang, Y. and S. Huang, The Out of East Asia theory of modern human origins supported by recent ancient mtDNA findings. Acta Anthropologica Sinica, 2019. 38(4): 491–498.

[111] Hajdinjak, M., et al., Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. Nature, 2021. 592(7853): 253–257.

[112] Yang, M.A., et al., 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. Curr Biol, 2017. 27(20): 3202–3208, e9.

[113] Lei, X., et al., Collective effects of common SNPs and risk prediction in lung cancer. Heredity, 2018. doi: 10.1038/s41437-018-0063-4.

[114] He, P., et al., Accumulation of minor alleles and risk prediction in schizophrenia. Sci Rep, 2017. 7(1): 11661.

[115] Gui, Y., X. Lei, and S. Huang, Collective effects of common SNPs and genetic risk prediction in type 1 diabetes. Clin Genet, 2017. 93: 1069–1074.

[116] Lei, X. and S. Huang, Enrichment of minor allele of SNPs and genetic prediction of type 2 diabetes risk in British population. PLoS ONE, 2017. 12(11): e0187644.

[117] Chen, H., et al., The relationship between the minor allele content and Alzheimer's disease. Genomics, 2020. 112(3): 2426–2432.

[118] Wang, M. and S. Huang, The collective effects of genetic variants and complex traits. J. Hum. Genet., 2023, 68: 255–262.

[119] Tishkoff, S.A., et al., The genetic structure and history of Africans and African Americans. Science, 2009. 324(5930): 1035–1044.

[120] Nedelec, Y., et al., Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell, 2016. 167(3): 657–669, e21.

[121] Quach, H., et al., Genetic adaptation and neandertal admixture shaped the immune system of human populations. Cell, 2016. 167(3): 643–656, e17.

[122] Howrigan, D.P., et al., Genome-wide autozygosity is associated with lower general cognitive ability. Mol Psychiatry, 2016. 21(6): 837–843.

[123] Joshi, P.K., et al., Directional dominance on stature and cognition in diverse human populations. Nature, 2015. 523(7561): 459–462.

[124] Feingold, A., Sex differences in variability in intellectual abilities: a new look at an old controversy. Review in Educational Research, 1992. 62: 61–84.

[125] Barton, N.H., The "new synthesis". Proc Natl Acad Sci USA, 2022. 119(30): e2122147119.

[126] Gorelick, R. and H.H. Heng, Sex reduces genetic variation: a multidisciplinary review. Evolution, 2011. 65(4): 1088–1098.

Shi Huang
Center for Medical Genetics
Hunan Key Laboratory of Medical Genetics
School of Life Sciences
Central South University
110 Xiangya Road
Changsha, Hunan 410078
P.R. China
E-mail address: shihuang1@gmail.com

Received November 16, 2022