

THE MAXIMUM GENETIC DIVERSITY THEORY: A COMPREHENSIVE FRAMEWORK FOR UNDERSTANDING EVOLUTIONARY PROCESSES

Shi Huang*

Center for Medical Genetics, Department of Genetics, Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 110 Xiangya Road, Changsha, Hunan 410078, P.R. China

Abstract

A good understanding of evolutionary processes is essential for addressing a variety of important scientific questions. In the early 1960s, the discovery of protein sequences revealed an unexpected phenomenon known as genetic equidistance, which led to the development of the molecular clock hypothesis. This hypothesis later inspired Kimura's neutral theory (NT). While the NT has served as a useful null model, it has struggled to fully explain genetic diversity. Systematic neutrality tests of the human mitochondrial genome have raised questions about the validity of the NT. As the molecular clock has fallen out of favor, this has also affected the standing of the NT and reintroduced the genetic equidistance phenomenon (GEP) as an unresolved issue. In the early 2000s, renewed interest in this phenomenon led to the formulation of the maximum genetic diversity (MGD) hypothesis, which offers a more comprehensive approach to understanding evolution. Analytical tests suggest that current genetic distances are primarily at maximum saturation rather than steadily increasing, as previously proposed by the molecular clock and NT. The MGD theory posits that macroevolution – from simple to complex organisms – involves a punctuated rise in epigenetic complexity accompanied by a decrease in the maximum genetic diversity a taxon can sustain. This theory has undergone extensive testing, contributing to our evolving understanding of evolutionary processes.

Keywords

genetic diversity • genetic equidistance phenomenon (GEP) • molecular clock • neutral theory • maximum genetic diversity (MGD) theory

Glossary

Genetic equidistance phenomenon (GEP): A simpler species is equally distantly related to two or more complex species in a sequence alignment when the genetic distance is measured by the identity matrix or percentage non-identity.

Maximum genetic diversity (MGD): The upper limit in genetic distance or diversity, as measured by the identity matrix or percentage differences, that can be reached during evolution over long periods.

Microevolution: In the MGD theory, the term encompasses evolutionary changes both within a species and between species over both short and long periods, as long as these changes do not involve major epigenetic modifications.

Macroevolution: In the MGD theory, the term describes new species formation that involves both genetic and epigenetic

changes, particularly involving increases in epigenetic complexity.

Introduction

A robust framework of evolutionary theories underpins all biological phenomena and is essential for understanding the mysteries that captivate us, including our own origins. The theoretical exploration of the evolutionary process has a rich history. Early evolutionary thought was pioneered by the French naturalist Jean-Baptiste Lamarck, who, in 1809, published 'Philosophie Zoologique' (1). Lamarck proposed an evolutionary theory based on two forces: a complexifying force that drives evolutionary progress from simple to complex forms, and an adaptive force that suggests environmental effects on traits can be inherited across generations.

* Corresponding author e-mail: shihuang1@gmail.com

Charles Darwin and Alfred Wallace independently conceived the idea of natural selection, presenting their joint papers to the Linnean Society of London in 1858. Similar concepts were also independently developed by three other English naturalists – Edward Blyth, Patrick Matthew, and William Wells – between 1813 and 1859. The era's racism and the expansion of the British Colonial Empire may have influenced the popularity of the 'survival of the fittest' ideology within English society (2, 3). Central to this theory are the concepts of common descent and natural selection as a mechanism to filter variations.

Among these five individuals, Blyth believed that the role of natural selection was very limited, merely maintaining the basic *status quo* of species, eliminating individuals that were not adapted, and ensuring the long-term stable existence of species. Wallace believed that natural selection did not apply to the emergence of human consciousness. Only Darwin considered natural selection to be nearly omnipotent, and only Darwin provided the most systematic and comprehensive description of examples supporting the concept of natural selection. His monumental work, 'On the Origin of Species', was published in 1859.

Based on theoretical advancements in the past two decades, Blyth's understanding of natural selection appears to be more in line with reality. The primary role of natural selection may be to maintain species stability. However, Blyth failed to note that natural selection may also play a role in the microevolution of species or species formation involving no major epigenetic or phenotypic changes (further discussion to follow).

In the early 20th century, the rediscovery of Mendelian genetics led to the modern synthesis, which combined Darwin's theory with genetic principles. The modern synthesis posits that adaptive evolution results from natural selection acting on genetic mutations. While this synthesis is widely accepted as a valid explanation for microevolutionary changes, its assertion that macroevolution operates through the same mechanisms is more controversial and has few if any experimental support. As one of the founders of the modern synthesis Simpson puts it: 'If the two proved to be basically different, the innumerable studies on microevolution would become relatively unimportant and would have minor value in the study of evolution as a whole' (4). Additionally, this theory remains incomplete as it fails to account for a surprising discovery in molecular evolution made in the early 1960s: the genetic equidistance phenomenon (GEP).

Since the discovery of the GEP, two theories of molecular evolution have emerged: first the neutral theory (NT) (plus the molecular clock hypothesis) and then the maximum genetic

diversity (MGD) theory. Both theories acknowledge the proven virtues of natural selection. The MGD theory also incorporates the valid concepts of the NT but considers its core claim to be flawed. This review examines these competing theories. Notably, there is a recent textbook by Bickel (5) that provides a comparative analysis of these theories. Additionally, these theories are introduced in a recently published Chinese textbook (6).

The GEP

In the early 1960s, molecular or genetic differences between species were measured for hemoglobin, cytochrome C, and fibrinopeptides (7–9). In these studies, genetic distance, a measure of molecular differences between species, was represented by the percentage non-identity in orthologous protein sequences, which is termed the identity matrix. The results varied depending on whether the reference species used for comparison was the most complex or the least complex, with complexity defined by the number of cell types. When humans were used as the reference species, it was observed that there was a gradually increasing amount of difference in hemoglobin when compared with progressively more distant species (7–9). This observation led Zuckerkandl and Pauling (7) to informally propose the universal molecular clock that hemoglobins from different species are changing at a steady and similar rate (10).

If, on the other hand, the reference species is the least complex among those being compared – for instance, when fish is the reference in comparison to humans, birds, and frogs (Figure 1) – a highly unusual result was observed: the GEP (8). This shows that fish is approximately equidistant to humans, birds, and frogs. Similarly, yeast cytochrome C is equidistant to all multicellular organisms. An example of the GEP using the Dot1 protein is shown in Table 1 (11). The GEP has been confirmed for nearly all proteins involving 15 species; for instance, bacteria are equidistant to yeast and human in all 196 protein orthologs available for the three species (11, 12). Data demonstrating the GEP have occasionally been presented in some papers without formal acknowledgment (13). Here, equidistance means that the distances are similar, though not necessarily identical, in numerical values.

To statistically demonstrate the GEP, multiple genes have been analyzed (11, 12). For example, for a simple species C, its equidistance to two more complex species B and A can be demonstrated by examining a set of randomly selected genes. 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 suggests that C is equally distant from

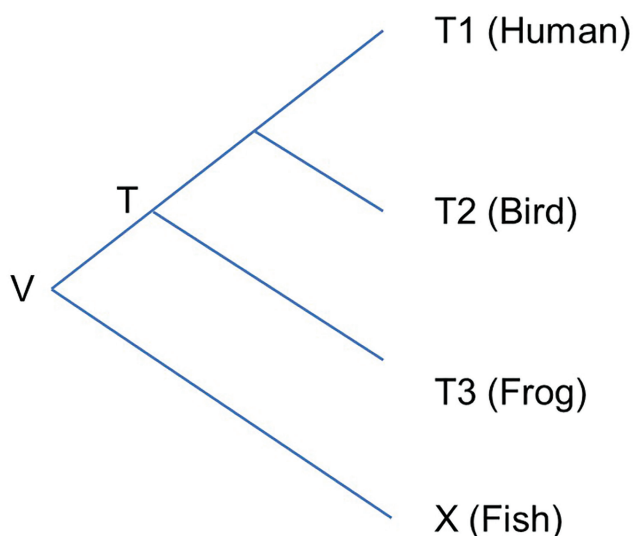


Figure 1. Illustration of the genetic equidistance result and its interpretation known as the molecular clock. Comparison for tetrapod species (T1–T3; human, bird, frog), which are known to have a most recent common ancestor (T), and another species (X; fish). Time flows from the past (left) to the present (right). Species X is the outgroup species and is equally distant to species T1–T3, the ingroup species, both in terms of time of separation and sequence difference, as measured by the identity matrix. This is illustrated by the lines linking X to T1, T2, and T3 being of equal length, indicating the same level of divergence in both time and sequence. The GEP refers to the fact that X is equally different in protein sequence to T1, T2, and T3. 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 been ongoing for the same duration since their common ancestor, V. Therefore, if a given protein exhibits equal divergence when comparing the same fish protein with proteins from different tetrapods, it suggests that the rate at which differences accumulate is similar among tetrapods (T1–T3). The implicit assumption here is that the molecular distance among the species has not yet reached its maximum level, allowing us to infer the rate at which sequence differences accumulate. If, however, the distance has reached an upper limit, it would no longer be related to time or mutation rate, rendering the inference of the rate invalid. GEP, genetic equidistance phenomenon.

A and B. The statistical significance of this can be assessed using a chi-squared test. Conversely, if a significantly higher number of proteins (e.g., 90%) show greater identity between C and B than between C and A, this indicates non-equidistance, meaning C is closer to B than to A. The GEP was unanticipated and remains largely unrecognized today.

It should be noted that while the two types of alignments yield different results – one consistent with intuitive expectations and the other not – the GEP specifically refers to the alignment with the simpler species as the reference species, as measured by the identity matrix. The alignment demonstrating the GEP was performed by Margoliash (8), whereas Zuckerkandl and Pauling (7) did not use this approach.

The molecular clock hypothesis

Perhaps due to the influence of the modern synthesis, which disregards the issue of complexity or evolutionary progress toward higher complexity as proposed in Lamarck's theory, Margoliash (8), in his initial report of the GEP, emphasized time as the determining factor of genetic distance. With this observation, he formally proposed the universal molecular clock hypothesis, suggesting that different species have approximately the same substitution rate, implying that genetic distance between species is determined solely by time (10).

The molecular clock hypothesis 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 types of distance results from the above alignments can both lead to the idea of a universal molecular clock and are merely two sides of the same coin, the GEP 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 and did not require calculations or fossil dating. Since all more complex species are equidistant in evolutionary time to the least complex reference species, equidistance in molecular divergence implies 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 a maximum level).

Second, the molecular distance results with humans as the reference have been presented in textbooks as evidence for Darwin's theory or the modern synthesis, suggesting that the more similar the phenotypes, the closer the molecular distance (although this pattern has many exceptions). However, the equidistance result is unexpected from the perspective of the modern synthesis: how can similar amounts of genetic change or a similar number of sequence differences between a simpler species and two more complex species result in vastly different degrees of phenotypic changes? Relative to the phenotypic changes from fish to frog, the changes from fish to human are far more pronounced. Despite being regarded by some scientists as 'one of the most astonishing findings of modern science' (14), the GEP has largely faded from collective consciousness. It has only been occasionally acknowledged by experts as the key finding that inspired the molecular clock hypothesis (10).

Table 1. Percentage non-identity among species in Dot1.

Species	Percentage non-identity						
	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

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

The table presents alignment results as measured by the identity matrix (percentage non-identity) for the Dot1 protein among a group of species, in the order from simple to complex. Equidistance to the simpler species can be seen as shown in each column. For example, yeast is equidistant to all the more complex species listed, with approximately 72% genetic difference.

Numerous studies indicate that the molecular clock has only limited applicability, suitable only for similar or closely related species. Mutation rates among different vertebrate species can vary by up to 40 times (15). Living fossil species have lower mutation rates (16). Today, almost no experts agree that there is a universal molecular clock applicable to all species (17). The relaxed clock is commonly used today that acknowledges different mutation rates among species (18). However, researchers are often unaware of the significant implications of rejecting the 'universal molecular clock'. They seldom realize that there is the GEP that has been overshadowed, misinterpreted, and replaced by the 'universal molecular clock'. The collapse of the 'universal molecular clock' implies that the GEP once again becomes an unsolved mystery.

The NT of molecular evolution

The molecular clock hypothesis was initially met with significant resistance from classical evolutionary biologists, as Darwin's theory could not explain the existence of similar substitution rates among different species. Nonetheless, some researchers have treated the molecular clock as a genuine reality and subsequently proposed various theories to explain it (19–24). The NT of Kimura (22) is one such theory known to account for the molecular clock (24), even though it is widely acknowledged to be an incomplete explanation (25, 26). However, Kimura believes that the best evidence for his NT is the molecular clock (27).

The abstract of the Kimura (22) paper on the NT consists of a single 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'. However, this calculation followed the logic of the molecular clock hypothesis, which includes two implicit assumptions that were taken for granted without deliberation. The first is that observed genetic distance always increases with time. The second is that every nucleotide in a genome is freely changeable or mostly neutral, meaning that alterations to any nucleotide position would not result in lethality. In fact, the mere use of the equation $r = d/2t$ to derive the mutation rate is an example of begging the question: it assumes the conclusion – that the mutation rate is the same for the two species – by presuming that the genetic distance was contributed equally by both species.

Research concerning neutral processes in genetics began in the 1920s and culminated in Kimura's NT (22, 28). From the outset, researchers did not seek to deny the importance of natural selection but were instead interested in how neutral processes influenced adaptive evolution (29). Fisher (30) and Wright (31) posed questions and developed techniques in population genetics that are also relevant to the NT. Haldane's (32) genetic load argument played a crucial role in inspiring Kimura to develop his NT. Kimura (22) was influenced by the idea that if most variants were not neutral, the genetic load would be too high to be sustainable. Additionally, some researchers have suggested that much of molecular evolution is neutral (33, 34). However, Kimura (22) was the first to combine population genetics theory with molecular evolution data to formulate a theory of neutral evolution. While the NT has been used (unsuccessfully) to explain electrophoretic protein polymorphisms within species (35, 36), the data that directly inspired Kimura to propose the NT came from sequence variation among species (37, 38).

The NT 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 (39). Some estimation has concluded that the human genome is mostly, 90%, junk DNA (40, 41).
2. Substitution rate equals mutation rate. Evolutionary rate as measured in generation time rather than years is constant and similar among taxa (22, 28).
3. The infinite-site model is part of the neutral framework (42). 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 (43–45).
4. The NT 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 (only quasi-neutral) as a result of selections (46–51).
5. The NT views non-conservation as non-function (52). This assumption treats many poorly conserved genomic elements as neutral junk, such as repetitive and viral elements, and is the basis for deducing the conclusion of only 8% functional genome in humans (40). However, functional genomics research in recent years has revealed that there are essentially no junk DNAs (53). All DNA should function in a mechanical code, i.e., the mapping between the local sequence and the local deformability of DNA (54), or the spatial grammar, i.e., the location-dependent function of transcription factors (55), or the optimal positioning of the enhancers (56). Also, at least 25% of the 1 million predicted regulatory DNA elements in the human genome are derived from transposable elements (57).
6. The NT treats synonymous mutations as neutral (24, 58). This assumption has now been invalidated by systematic studies (59–62).
7. Mutations in different sites are independent.
8. Effective population size (N_e) is deduced to be related to genetic diversity (π) as in the formula $\pi = 4N_e\mu$ (mutation rate μ) and often tautologically derived by using the very genetic variables that it is meant to predict or explain (35).
9. The Hardy–Weinberg law: allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences. This law underlies nearly all methods in phylogenetic studies of human populations. Direct tests of this law is rare but a recent test has disproven it (46).

The NT is also used as an explanatory theory to explain genetic diversity within species. However, it is widely acknowledged that the NT is not a satisfying account for genetic diversity (46, 63–65). Some have even declared it dead as an explanatory theory of nature, while acknowledging its value as a useful null model (66). Direct experimental tests of the NT have disproven it (46). A multiyear, genome-wide analysis of a population of the microcrustacean *Daphnia pulex* demonstrates that temporal variation in selection intensity has a significant influence on nucleotide polymorphism and divergence (46). Most nucleotide sites experience fluctuating selection with mean selection coefficients near zero, with little covariance in selection strength across time intervals, and with selection distributed across large numbers of genomic islands of linked sites. Therefore, it is natural selection, rather than the presumed neutral drift, that explains the patterns of genetic diversity in real life.

Most in the field today merely view the NT 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 NT to be a true account of nature and often unfoundedly present the trees based on the neutral assumption as uncertainty-free, e.g., the recent out of Africa model of modern human origins (67). Most phylogenetic methods require the neutral assumption, including methods such as the bootstrap intended to test the robustness of a phylogenetic tree (68).

The MGD theory of molecular evolution

Both the molecular clock and the NT provide valuable insights in many areas. However, they have limitations and may not fully account for the complexities of evolutionary processes. As shown by studies in the last two decades, the GEP remains unexplained by these theories. The molecular clock interpretation of the GEP is, in fact, a classic tautology – a mere ad hoc restatement of a distance phenomenon (69). It has not been verified by any independent observations, and on the contrary, has been contradicted by a significant number of observations (15, 16, 26, 69–71). Some have even declared it to be merely a mirage (25).

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 NT has not fully explained the molecular clock even though it was inspired by it (25, 26). The observed rate is measured in years but the NT 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 (72) '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.

Around 2005, open access to genome sequencing results of various species allowed us to compare the sequences of fish, frog, and human for the retinoblastoma interacting zinc-finger-protein 1/PR domain containing 2 (RIZ1/PRDM2) tumor suppressor gene that we had been studying (73, 74). We first used the human RIZ1 protein sequence as a query to search for homologous proteins in the GenBank database, which is a routine practice, and obtained the expected alignment identity results showing that humans are closer to frogs than to fish. Next, we conducted an unconventional experiment by using the fish RIZ1 protein as the query, fully expecting – like most researchers would – that fish would be closer to frogs than to humans. To our surprise, we observed the GEP, with fish being approximately equally different from both frogs and humans. This led us to recognize that the molecular clock interpretation of the GEP might be more of a tautology than a definitive explanation (69). After three years of in-depth research, we published the MGD hypothesis in 2008, which offers a new interpretation of most major evolutionary phenomena (74, 75). We also later identified a new characteristic of the GEP, known as the overlap feature, which has not yet been recognized or explained by existing theories (76).

Here we summarize the terminology and concepts of the MGD theory.

Microevolution

In popular evolutionary theory, the term 'microevolution' refers to evolutionary changes within a species or a small group of organisms, especially over short periods. In the MGD theory, however, the term encompasses evolutionary changes both within a species and between species over both short and long periods, as long as these changes do not involve major epigenetic modifications.

Macroevolution

The term macroevolution in the popular theory applies mainly to the evolution of whole taxonomic groups over long periods

of time. In the MGD theory, the term describes new species formation that involves both genetic and epigenetic changes, particularly involving increases in epigenetic complexity. The number of cell types is used 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 reflected by major epigenetic changes. Each neuron is likely a unique cell type with a unique synaptic connectome. A typical human neuron has an average of about 1000 to 10,000 synapse connections with other neurons. Even within the same type of seemingly homogeneous neurons, epigenetic heterogeneity or diversity exists and plays a role in memory formation, rather than being neutral or simply noise (77). Therefore, complexity is largely defined by the number of types of neurons. Given that phenotypes may change throughout an organism's lifetime, the complexity of its phenotype refers to the peak level of complexity during its healthy lifespan.

MGD

The central problem of the field has always been the old riddle of what determines genetic diversity (35, 63, 78, 79). The study of variation and diversity has a long history (80). 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 (35, 79). 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 (35, 79). 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 (63). Two main schools of thought evolved from the earlier two: one originally believed in little genetic variation and now favors neutral drift, while the other continues to support natural selection. The debate remains unsettled and has been largely neglected over the past three decades as the field has shifted its focus to generating increasing amounts of diversity data using newer techniques.

Genetic diversity within a species is commonly measured as the nucleotide diversity, defined by Nei and Li (81) in 1979 as the average number of nucleotide differences per site between two DNA sequences in all possible pairs in the sample population. In this sense, genetic diversity is equivalent to genetic distance or sequence mismatches. The new concept of MGD is mainly in regard to the form of genetic diversity as measured by genetic distance. For a group of organisms with multiple taxa exhibiting similar phenotypes,

the genetic diversity of the group with respect to a gene 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 (7–9). The genetic diversity of an individual is commonly measured by the amount of heterozygosity, which is equivalent to the sequence difference 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 optimal genetic state.

It is important to distinguish between low genetic diversity and low heterozygosity. While low genetic diversity is manifested as both low heterozygosity and low MAC or a low number of deleterious alleles, low heterozygosity by itself, associated with higher amount of deleterious alleles without low MAC, could result from inbreeding (82).

Many authors use the term genetic diversity to refer only to genetic variations within a species or population (63, 83). But many also use the term to refer to genetic variations both within and between populations or species (84). The popular theory of molecular evolution, the NT, was directly inspired by a genetic distance phenomenon between species (the GEP) and its interpretation, the molecular clock (37, 38). This theory is commonly used to interpret genetic diversity patterns both within and between species. Therefore, it would be inconsistent if the term 'genetic diversity' only covered genetic variations within a species or population.

The existing concepts of genetic diversity remain unchanged in the MGD theory. However, the new theory introduces an additional concept: the upper limit concept. Genetic diversity or distance, as measured in most genes by identity matrix, would increase with time but cannot reach 100% non-identity. The increase would stop at an upper limit level, leaving many sequence positions unchanged. The upper limit of the percentage of positions in the sequence that differ among taxa within a group of similar organisms is called the MGD of the organism. The upper limit of the percentage of positions in the sequence that differ among individuals within a taxon is called the MGD of the taxon.

While the concept of an upper limit does imply that genetic distance/diversity will remain unchanged over time, it does not signify the cessation of evolution. Evolution is generally understood as a process involving mutations followed by natural selection (Darwin's theory) or neutral drift (Kimura's theory), and these processes are expected to continue indefinitely. However, while these processes would lead to increases in genetic distance/diversity during the early stages

of evolution, they would not alter the optimum level of genetic distance/diversity once the upper limit is reached. At the upper limit stage, what continues to change is the turnover of alleles. In other words, while the MGD level is maintained by one set of variants/alleles at time point A, the same level of MGD would be maintained at a later time point B by a new set of variants/alleles.

The level of MGD measures the fraction of positions in a sequence that can change without adversely affecting the core physiology of the members of a taxon or the taxon itself within a group of similar organisms containing multiple taxa. These changes may be beneficial, deleterious, or neutral with regard to fitness, depending on circumstances. Positions not free to change are considered conserved and are essential for the core physiology of a taxon. Positions that are identical among the compared species are considered conserved. These unchanged positions consist of the following three types:

1. Positions essential for the barebone or minimal function of a gene, where any change would alter the biochemical activity of the gene in a test tube.
2. Positions essential for more complex species but not for simpler species. As species become more complex, an increasing number of positions in a gene become unchangeable or are involved in more complex traits. Changes in these positions might not affect the basic function of a gene or its activity in a test tube, but they could influence the long-term survival of more complex species, while having little to no impact on simpler species.
3. Positions essential for adaption to a specific environment. Such positions may be fixed. Environmental selection may lead to homozygosity of variants and reduce the MGD level of a taxa.

Genetics, together with epigenetics, play important roles in determining phenotypes, and if genetic diversity has no limit, it implies that phenotypic diversity also has no limit. To understand the limits of genetic diversity, one can consider whether phenotypic diversity has limits. For example, humans exhibit extensive phenotypic diversity, such as variations in skin color, hair color, and eye size, but these are all minor quantitative variations within certain boundaries. Fundamental aspects of nearly all human traits, like having two eyes or one mouth, do not change, suggesting that there are immutable sequences in the genome. The existence of unchanging phenotypes indicates that phenotypic diversity is not infinite, implying that genetic diversity must also have an upper limit. The NT assumes no upper limit on genetic diversity, suggesting that genetic differences between human individuals could

reach 100%, which implies unlimited phenotypic differences – a notion that can be easily disproven.

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 (85–92). Epigenetic programs are not only inherited during mitotic cell division but are also often transmitted through the germline to the next generation (93–99). 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 Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain histone methyltransferases while bacteria have none.

The MGD theory contains three self-evident intuitions or axioms (74, 75, 100–103).

Axiom 1 posits that the more complex the phenotype, the greater the restriction on the choice of molecular building blocks. A complex or ordered system requires higher precision in its building parts. In biology, this means there is an inverse relationship between genetic diversity and epigenetic complexity (75).

The idea of *Axiom 1* was first noted nearly 100 years ago by Fisher (30, 104), as well put by Orr (104): ‘*Fisher further showed that mutations of a given size are less likely to be favorable in complex than simple organisms. The reason is intuitively clear: A random mutation of some size is more likely to disrupt (not improve) a complex than simple organism for the same reason a random change of some size is more likely to disrupt (not improve) a complex than simple machine. Changing the length of an arbitrary mechanical part by one inch, for instance, is more likely to derail the function of a microscope than a hammer.*’ However, the prevailing evolutionary theories have long overlooked, and in fact, contradict this notion.

Axiom 2 states that in any system allowing a limited level of random errors or noises in molecular building parts, such errors may be beneficial, deleterious, or neutral depending on the circumstances. Limited errors at an optimum level are more likely to be beneficial than deleterious because they are within tolerable limits, conferring economy in construction

and the strongest possible adaptive capacity or robustness to environmental challenges. In biology, substituting ‘errors in building blocks’ with ‘genetic diversity’ yields an equivalent concept. *Axiom 2* highlights the valid aspects of both Kimura’s NT and Darwin’s theory of natural selection.

Axiom 3. Deleterious variants or mutations harmful to a trait could be rescued by other mutations (105). 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. Therefore, suppressing genetic diversity is necessary for maintaining traits at a high-quality level and for removing harmful variants.

For neutral variants, genetic diversity π has the formula $\pi = 4N_e\mu$, and is primarily determined by the mutation rate (μ) and the effective population size (N_e). However, MGD is driven by species complexity C and environmental factors E , as reflected in the formula: $MGD = E/C$. Environmental factors may include geographical and ecological pressures, competition for resources, social or behavior changes, and many other factors that influence the survival and adaptation of species. A narrow environmental selection can reduce MGD, as limited environmental factors would restrict the ability of species to diversify genetically, thus lowering the overall diversity.

The MGD theory makes the following claims:

1. MGD tends to be higher for simpler taxa and lower for more complex taxa (Figure 2). The reason may be that the members of more complex taxa rely on more sequence positions for their more complex physiology, which are for that reason conserved, leaving fewer positions free to change. Complex taxa also require higher precision genomes, leaving fewer positions free to change.
2. Macroevolution involves changes in organismal complexity and often results in an increase in complexity, which is mirrored by an increase in the precision of the building parts or a decrease in the allowed range of the standard deviations for the parts (Figure 2A). Although there are some rare instances of apparent loss of complexity, such as the loss of eyesight in nearly blind naked mole rats, these changes are relatively minor and do not detract from the overarching trend of progression toward higher complexity. Moreover, it remains unclear whether there may be compensatory increases in complexity in other aspects, leading to overall minimal net changes in complexity. Microevolution, on the other hand, is an increase in genetic diversity within the

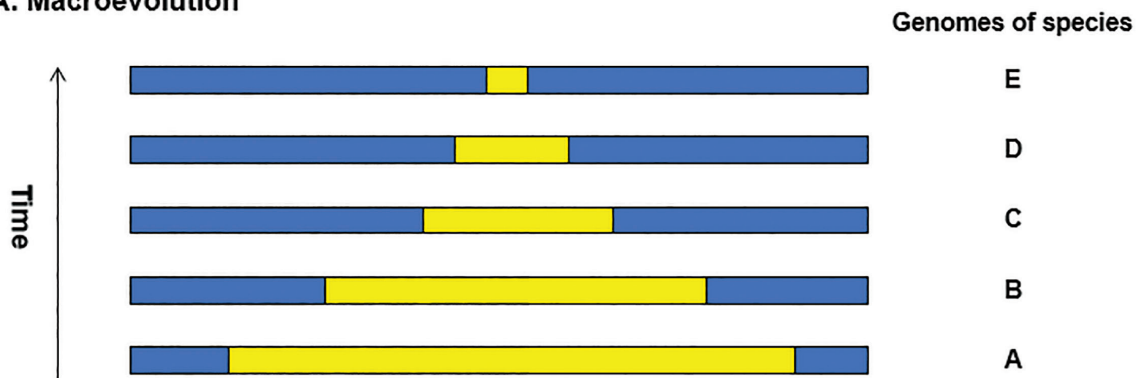
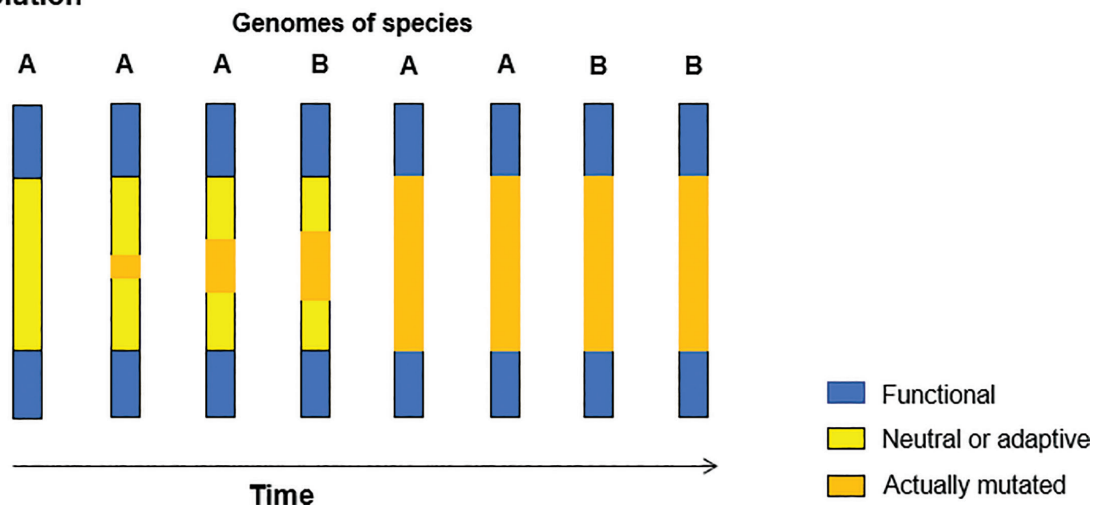
A. Macroevolution**B. Microevolution**

Figure 2. Schematic representation of the MGD theory of evolution. (A) Macroevolution. As species evolves from simple to complex (taxon A to taxon E), the maximum level of genetic diversity that a taxon can tolerate gets reduced. (B) Microevolution. Accumulation of random mutations within the tolerated range of genetic diversity leads to speciation (from taxon A to taxon B) without much changes in epigenetic complexity or the level of MGD that a taxon can tolerate. MGD, maximum genetic diversity.

allowed standard deviation ranges without significant change in complexity (Figure 2B). It encompasses both evolution within species and evolution from one species to another. At the saturation phase, microevolution involves turnover of alleles at the equilibrium level of genetic diversity. The gradual evolution of sequences occurs at the microevolution level but cannot be extrapolated to the scale of macroevolution, as Gould and Eldredge (106) had concluded largely on the basis of the fossil record.

3. The positions that are conserved in simpler taxa tend to also be conserved in more complex taxa. 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 (107). Any level higher or lower than the optimum would be negatively selected. As genetic diversity, so long it is within the maximum level, facilitates adaptation, it would be positively selected to quickly reach the optimum level. The optimum level is equal to the maximum level and can be defined in humans 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. The optimum concept here means a Pareto optimum or simply the best that can be achieved

due to a balance between positive and negative selection at a particular time point under a specific level of epigenetic or organismal complexity (108). As time, environments, and complexity changes, the optimum level of nucleotide diversity will also change.

5. Genetic variants are mostly functional or under balancing selection and quasi-neutral (under both positive and negative selection) rather than neutral.
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 runs of homozygosity (ROH) 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 fundamentally 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.

Using conservation as an index of function measures only one of two kinds of sequences: those essential for internal system physiology that have little to do with adaptation to external environments. To preserve the long-term integrity of the system or essential traits, such as the presence of only two eyes in humans, such sequences must remain unchanged. For living fossils to be possible, these sequences should be highly stable. On the other hand, sequences involved in adaptation to changing environments must be able to evolve quickly, as environmental changes tend to occur rapidly. For instance, flu viruses escape neutralizing antibodies every few years, and the fast-changing non-conserved sites in these viruses are crucial for their survival but not essential for their basic physiological functions.

Comparing the claims of the two competing theories, the MGD theory makes many assertions that directly oppose those of the NT (Table 2). (a) Most variants are not neutral. (b) Genetic distances or diversities are mostly at upper limit levels. (c) The infinite site model is unrealistic. (d)

Table 2. The MGD theory versus the NT.

Features	MGD theory	NT
Micro-evo linear stage	NT and natural selection	NT and natural selection
Micro- and Macro-evo	Different mechanisms	Same
Role of epigenetics	Yes	No
Noises suppressed	Yes	No
Neutral regions	Complexity dependent	Complexity independent
Time frames	Any	Relevant to short times
Mutation rates	Any	True for slow rates
Genetic equidistance	Maximum and linear	Linear only
MGD	Yes	No
Genetic variants	Mostly non-neutral	Mostly neutral
Recurrent mutations	Common	Rare
Quasi-neutrality	Common	Rare
Turnover of alleles	Fast and common	Slow and rare
Non-conservation	Adaptive function	No function
Physiological selection	Yes	No
Independent tests	Confirmed	Rejected
Axiom based	Yes	No

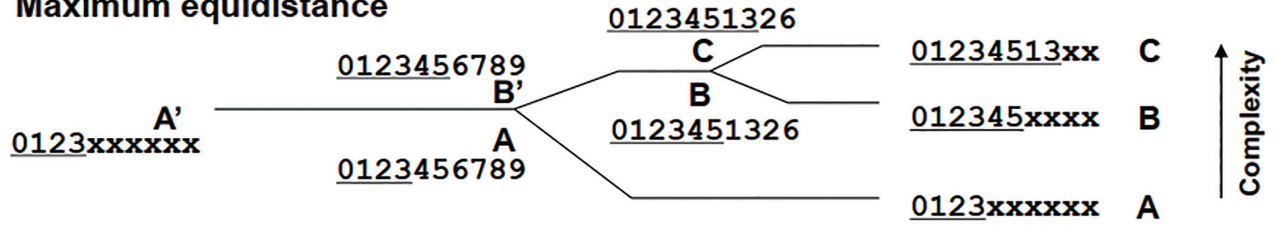
MGD, maximum genetic diversity; NT, neutral theory.

The table compares the major features of the two different theories of molecular evolution.

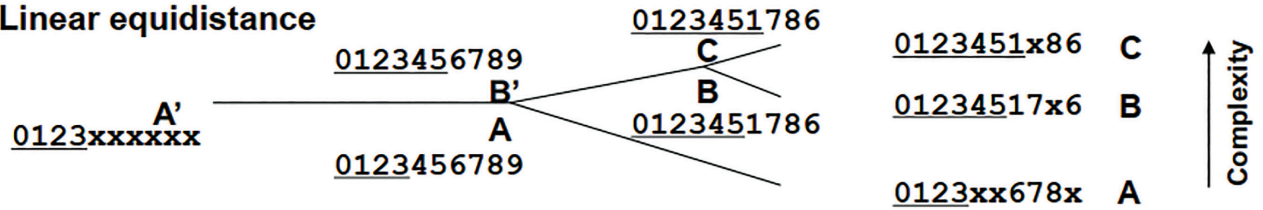
Most DNA sequences are not informative to phylogenetic inferences. (e) All mutations have a deleterious aspect. (f) Non-conservation means adaptive function. (g) Synonymous substitutions are also functional. (h) Genetic or molecular distance between two taxa of different complexities is not equally influenced by mutations in each lineage. (i) Increase in complexity requires suppression of genetic diversity. (j) Evolution or turnover of alleles are very fast rather than slow (46, 107). (k) Physiological selection determines the fate of many variants.

The GEP reinterpreted

The MGD hypothesis explains the GEP as a result of maximum genetic distance (74, 75, 100). Over a long evolutionary period, for fast-evolving DNAs, the genetic distance between species has reached the maximum level. The distance between ingroup species and a simpler outgroup taxon is primarily determined by the MGD of the simpler outgroup (Figures 2 and 3). 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

A. Maximum equidistance**B. Cyt C**

1	GDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKN	Hs
1	GDVEKGKKLFFVQCAQCHTVEAGGKHVGPNNLHGLIGRKTGQAAGFAYTDANKA	Dm
1	GSAKKGATLTKTRCLQCHTVEKGGPHKVGPNLHGIFGRHSGQAEGYSYTDANIK	Sc
		Hs Dm Sc
55	KGIIWGEDTLMLEYLENPKKIYIPGTMIFVGIKKKEERADLIAYLKAT	Hs 22 36
55	KGITWNEDTLFEYLENPKKIYIPGTMIFAGLKKPNERGDLIAYLKAT	Dm 36
55	KNVLWDENNMSSEYLTNPKKIYIPGTMIFAGLKKEDRNDLITYLKAT	Sc

C. Linear equidistance**D. Txnd9**

61	QKQEWLSKGHGGEYREIPSERDFFQEVKESENVVCHFYRDSTFRCKILDRHLAILSKKHLE	Hs
61	QKQEWLSKGHGGEYREIPSEREFFQEVKESENVVCHFYRDSTFRCKILDRHLAILSKKHLE	Pa
61	QKQEWLSKGHGGEYREIGSERDFFQEVKESEKVVCHFYRDTTFRCKILDRHLAILAKKHLE	Mm
		Hs Pa Mm
121	TKFLKLNVEKAPFLCERLRIKVIPTLALLKDGKTQDYVVGFTDLGNTDDFTTETLEWRLG	Hs 2 6
121	TKFLKLNVEKAPFLCERLRIKVIPTLALLKDGKTQDYVVGFTDLGNTDDFTTETLEWRLG	Pa 6
121	TKFLKLNVEKAPFLCERLRIKVIPTLALLRDGKTQDYVVGFTDLGNTDDFTTETLEWRLG	Mm

Figure 3. The genetic equidistance result and the MGD theory. (A) Maximum genetic equidistance. A 10 amino acid peptide is used to illustrate the evolution process. When the protein is fast evolving, the observed equidistance today would be maximum distance with a HOR. The figure shows 4 overlap positions with an overlap ratio 1. The distance of C–A is 60%, the same as that of that of B–A. This is a schematic representation of the original Margoliash genetic equidistance result. (B) An example of maximum genetic equidistance. Alignment of human, drosophila, and yeast cytochrome C proteins. Human differs from drosophila in 22 amino acid positions. Human and drosophila are equidistant to yeast with 36 amino acid differences. There are 12 overlap positions (in red color) and the overlap ratio is $12/22 = 55\%$. Other mutant positions are colored in green, blue and orange. (C) Linear genetic equidistance. When the protein is slowly evolving, assuming molecular clock holds, the observed equidistance today would be linear distance with a low overlap ratio. Here every substitution in any species would mean an increase in distance. The figure shows 0 overlap position with an overlap ratio 0. The distance of C–A is 50% and equals that of B–A. (D) An example of linear equidistance. Human, orangutan, and mouse *TXND9* gene alignment. There are two amino acid difference between human and orangutan, which are equidistant to mouse with six amino acid differences. The overlap ratio is $0/2 = 0$. HOR, high overlap ratio; MGD, maximum genetic diversity.

of fishes. Changes in the lineage leading to the simpler outgroup mask any changes in the lineages leading to the ingroup taxa (Figure 3).

This notion that the MGD level 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,

such as 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 GEP (Figure 3). For long evolutionary timescales or 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 (8) is an example of maximum genetic equidistance. For short evolutionary timescales or slowly-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 GEP by the MGD theory can be easily illustrated by a simple thought experiment. Imagine we create a yeast, a fish, and a human 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 remaining phenotypically largely the same as today. Over this time, a gene in yeast would change significantly, up to a maximum of, say, 50%, while its homolog in fish would change up to a maximum of 30%, and its homolog in humans would change very little, say <1%. Any further changes beyond 50% would be lethal to yeast, beyond 30% would be lethal to fish, and beyond 1% would be lethal to humans.

The reason that a gene in yeast can change much more than in fish, which in turn can change more than in humans, is that a gene in humans encounters far more functional constraints than its homolog in fish or yeast. Thus, the genetic distance between yeast and humans 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 humans or between yeast and fish, as well as 50% identity within species distance in yeast. The 30% change in fish would account for the genetic distance of 30% identity between fish and humans. In contrast, the NT would predict that both humans and fish can also, like yeast, change up to 50% or more, resulting in a genetic distance of 50% identity.

Testing the different evolutionary theories

There are two ways to test which of the two different interpretations of genetic equidistance is true, the molecular clock versus the MGD theory. The first is the overlap feature of mutations (76, 109). When aligning a protein sequence from three taxa, the conserved positions would show identical amino acids for all three taxa. The non-conserved positions can be categorized into two types. One type involves mutation in only one of the three taxa, while the other two taxa retain either identical amino acids or remain unchanged. The second type consists of positions where each taxon has a unique

amino acid at the same location (red colored positions, Figure 3). This indicates that at least two taxa had gone through a unique mutation at that same position, and hence their mutations overlapped at that position. Such positions are referred to as overlap positions.

Since one mutation is enough to lead to a difference of one between two taxa as measured by identity matrix, 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 have mutated, resulting in a higher fraction of overlap positions among the non-conserved positions than expected by chance. This is termed the high overlap ratio (HOR) phenomenon.

In fact, Margoliash's GEP shows a high fraction of overlap positions, which aligns with the prediction of the MGD theory and significantly exceeds what is predicted by the molecular clock and NT (Figure 3) (11, 76).

It is important to note that the HOR type of saturation is different from that seen in the 'long branch attraction (LBA)' in phylogenetic trees (110). In LBA, saturation means convergent mutations leading to the same amino acid residue or nucleotide among (across) multiple taxa, leading to a shortening of branch length. Although they were derived independently, these shared alleles can be misinterpreted in phylogenetic analyses as being shared due to common ancestry. However, for the HOR phenomenon, independent mutations at the same site among different taxa would generally lead to different taxa having different amino acids rather than the same, since the probability of an independent mutation changing to the same amino acid is about 20 times lower than that of mutating to a different amino acid. Thus, HOR is expected to be more common in nature compared to the saturation in the case of LBA. These two types of saturation are essentially just different aspects of the same phenomenon: one more commonplace and manifesting as HOR, while the other is less common and manifests as LBA.

The other way to test the molecular clock versus the MGD theory is the genetic non-equidistance result despite equidistance in time. The MGD 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 MGD 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 MGD theory (5, 111, 112). 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 intuitively 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 MGD theory but not that of the molecular clock (111). It has recently been shown that octopus indeed has the lowest heterozygosity level among mollusks, and that the highest heterozygosity level among mollusks is found for the least complex solenogastres (113). 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 (114).

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.

There are also many other independent tests that have confirmed the MGD theory:

1. Latest evidence for Axiom 1 includes the following. Lower species have more common variants that are however harmful or non-tolerated in humans (115, 116). The highest MGD level is found for the Human Immunodeficiency Virus (HIV) at 40% bp/bp (117), while the lowest is found in humans at 0.1% bp/bp, which is half that of chimpanzees (118), one-third of that in macaques (129), and one-eighth of that in lemurs (120). The malaria parasite endemic to Asian macaque monkeys, *Plasmodium knowlesi*, exhibits higher genetic diversity and, as demonstrated by supersaturation

mutagenesis experiments, is more tolerant of mutations than *Plasmodium falciparum*, which is endemic in humans (121). The so-called 'Random Genome Project' verifies this axiom: mammalian cells repress random DNA that yeast transcribes, confirming the expectation that complex organisms are less tolerant of random errors or noises than simple ones do (122).

2. The overall level of individual genetic diversity is inversely related to educational attainment or cognitive function (123–127), with most genetic variations being harmful to the human brain (128). Consistently, the more socially tolerant crab eating macaque has lower genetic diversity than the less tolerant rhesus macaque (129, 130), and the more tolerant bonobo has lower genetic diversity than the less tolerant chimpanzee (131, 132). It is possible that the high genetic diversity observed in the San people of Southern Africa could be a factor contributing to their relatively under developed culture, while also providing them with stronger immunity (133–136), and may have little to do with longer evolutionary time or being human ancestors. Also, animal model studies have found associations between genetic diversity and various complex traits, including cognitive function and immune function (47).
3. The genetic diversity level of patient groups with many complex diseases is higher than that of normal control groups, indicating that the genetic diversity level of normal individuals is at an upper limit state (48, 137, 138). Many human traits are subject to stabilizing selection, implying there is an optimal degree for a quantitative trait (139).
4. Around 40% of human single nucleotide polymorphisms (SNPs) in database of SNPs (dbSNP) carry recurrent mutations or more than two alleles, directly invalidating the infinite site assumption of the NT framework (140).
5. Functional genomics studies have found that most of the genome is functional (141–143). For example, short tandem repeats (STRs), which account for 5% of the human genome, almost all have regulatory functions in gene expression (141). Also, large fragments of random foreign DNA, differing in Guanine (G) and Cytosine (C) (GC) content, introduced into yeast either mingle with other active yeast chromosomes or segregate into an inactive compartment, a process predicted by the underlying DNA sequence. If most of the DNA were neutral, this would be unlikely to occur (144). During the aging process, mutations correlate with alterations in DNA methylation, directly supporting the inverse relationship between genetic diversity/mutation and epigenetic complexity/order as posited by the MGD theory (145).

6. Quantifying constraints in the human mitochondrial genome has shown that most variants in mitochondrial DNA (mtDNA), including those in the non-coding region, are subject to selection (146). This challenges numerous phylogenetic trees constructed using mtDNA genes that assume the molecular clock and the NT.
7. Bacterial robustness to mutations is not affected by evolutionary generations, which does not align with popular theoretical expectations. The robustness of a species is primarily determined by its blueprint: the simpler the construction of a species, the stronger its robustness and the greater its tolerance to mutations (147).
8. The existence of conserved sequences that are related to complexity but not to enzyme function per se. The length of such sequences increases with complexity. This type of sequences, termed complexity associated protein sectors, has been directly shown to exist (148).
9. Genome as a whole is a building part, and this part can vary in size is just like any part should have an allowable standard deviation. Thus, complex species tend to have less variation in genome size relative to lower species (39). Protozoa, a group of unicellular organisms, exhibit a genome size variation where one species' genome is approximately 20,000 times larger than that of another, whereas in mammals, the variation is less than 10-fold. The truly significant question regarding genome size should be why protozoa have such a vast variation in genome size, rather than why the onion has such a large genome. Notably, Africans, well-known to have greater genetic diversity, also consistently have higher copy number of duplicated gene families than non-Africans (149), implying slightly greater genome size variation. A bigger genome has been found to provide environment-dependent growth benefits in grasses (150).
10. A number of mechanisms exist that suppress the transmission of new mutations to the next generation prior to birth, including the abortion of mutant embryos, the DNA repair system in germline reproductive cells being more efficient than somatic cells (151), and older mothers having higher capacities to prevent mutant mtDNA from getting transmitted to their daughters (152). This is physiological selection as opposed to natural selection (of adult individuals by environments). The important question implied here is, how can the Darwinian mechanism, which needs as many as possible mutant adult individuals in order for natural selection to work, lead to mechanisms that would reduce the number of such individuals or undermine itself? A related question is, if genetic diversity provides the raw material for Darwinian natural selection and lower genetic diversity is associated with each advance in species complexity, how can natural selection undermine itself during evolution?
11. Trait diversity is commonly believed to be shaped by environmental diversity, according to the prevailing view. However, this perspective overlooks the idea proposed by the MGD theory, which posits that genetic diversity – and, by extension, trait diversity – is primarily influenced by species complexity. Recent findings have shown that the habitats within Africa seem less diverse or complex than outside (153). Eurasia, where humans have lived for ~2 million years, has a much more diverse environments than Africa. So, humans living in Eurasia have experienced a broader, and hence more complex, range of environments than those who have stayed in Africa as well as non-human primates such as African apes that have lived in Africa. So, Eurasians should be expected to have higher genetic diversity than bonobos or even Africans according popular views or the finding of reduction in environmental diversity leading to a decrease in genetic diversity (154). But the reality is the exact opposite. So, the correlation between environmental complexity and genetic diversity is contradicted by the low genetic diversity of Eurasians compared to bonobos or Africans. Also, recent study shows that drylands act as a global reservoir of plant phenotypic diversity and challenge the prevailing view that harsh environmental conditions reduce plant trait diversity (155). These studies are in line with the MGD theory that trait diversity or genetic diversity is mostly determined by species complexity.

What does the universal molecular clock really mean?

The molecular clock hypothesis contains both valid and flawed elements. During the linear phase, where genetic distances are driven by truly neutral mutations, the molecular clock holds and can explain the phenomenon of linear genetic equidistance. However, this concept clearly cannot be applied to the phenomenon of maximum genetic distance.

The universal molecular clock was (mis)inspired by the phenomenon of maximum genetic distance and reflects the constant rate of complexity increases. The first molecular evidence for the constant, albeit discontinuous, advances in complexity is, in fact, the phenomenon of maximum genetic equidistance. Although this progression towards higher complexity is non-linear, it generally appears roughly linear over long periods, leading to the superficial appearance of a universal molecular clock. 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 (14, 80, 106, 156–159). Underlying the direction towards higher complexity and the constant rate of increase in complexity may be the mysterious ‘complexifying force’ first proposed by Lamarck (1).

Dissenting perspectives on the MGD theory

Since the publication of the MGD theory in 2008, there has been a lack of formal dissenting viewpoints from experts in the field, with the exception of a recent preprint by Zhang (160), which is notably dismissive. We have written a detailed rebuttal of Zhang’s paper (161). In brief, Zhang attempts to deny the validity of the GEP while defending the molecular clock and NT, a stance that is contradictory, as discrediting the GEP would undermine the molecular clock itself. Furthermore, he does not challenge the upper limit concept of the MGD theory. As a result, his paper is self-defeating and lacks the rigorous scholarship necessary for a meaningful contribution to the scientific discourse.

Zhang also raised a common criticism of the MGD theory, which argues that the saturation issue it raises is not new. However, these critics fail to recognize the distinction between partial saturation and the upper limit concept of the MGD theory. The field has primarily focused on partial saturation, where genetic distance between species still increases with time, albeit at a non-linear rate compared to the early stages of species divergence. The methods used in the field to address partial saturation cannot be applied to full saturation at the upper limit level. This upper limit could have been reached a long time ago, making it

impossible to model the number of mutations that have occurred in the past.

Conclusion

The GEP is the most astonishing finding in molecular evolution, so remarkable that it was misunderstood for many decades and directly inspired two distinct theoretical frameworks of molecular evolution (Figure 4). To fully grasp evolution, one must first understand molecular evolution, and to understand molecular evolution, the GEP is essential. This phenomenon was initially interpreted by proposing the molecular clock hypothesis and in turn the NT, but has since been re-interpreted by the MGD theory (Table 2). Direct tests of the predictions from these two competing theoretical frameworks have raised doubts about the validity of the molecular clock and NT, while providing support for the MGD theory. Currently, genetic distances or diversities appear to be largely at an optimal equilibrium. Although turnover of alleles due to mutations, natural selection, and neutral drift will continue indefinitely, such standard evolutionary processes will not affect the MGD level unless there is an increase in species complexity or a dramatic change in environmental selection. This evolving understanding may offer fresh insights into various evolutionary questions such as the origin of modern humans (162, 163).

Future research may further support the MGD theory in several ways:

1. Large-scale functional genomics studies may show that almost all bases have functions.

Timeline | Six decades of genetic equidistance

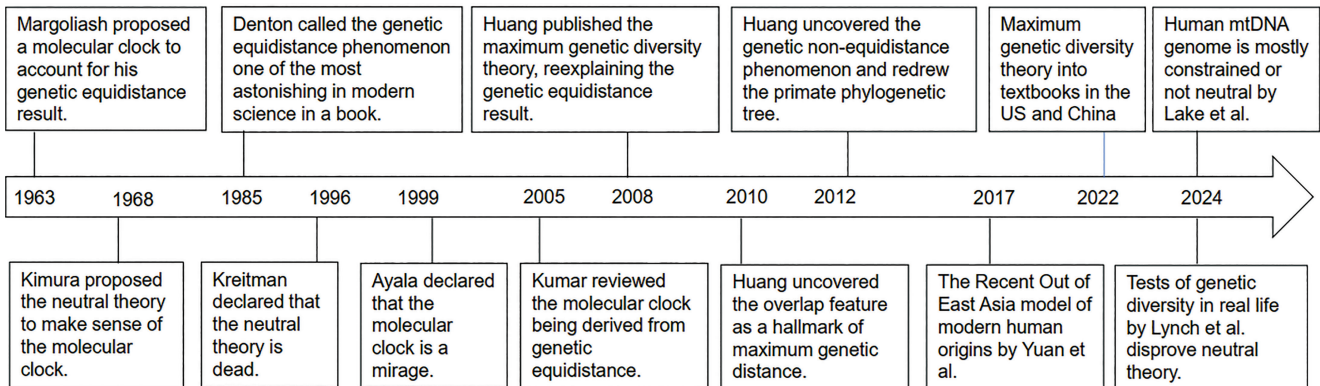


Figure 4. Time line regarding the GEP. GEP, genetic equidistance phenomenon.

2. Studies on more populations or species may further demonstrate that genetic diversity is at its maximum limit. For individuals within the same ethnicity, compare the genetic diversity of the population of children's versus that of their parents, do the parents with higher than average genetic diversity tend to produce children with lower genetic diversity as expected by the MGD theory? Do the parents with lower than average genetic diversity tend to produce children with higher genetic diversity?
3. Research on more species and more genes may further establish the GEP.
4. By confirming phylogenetic trees based on the MGD theory the theory itself will also be validated.
5. Better understanding of cognitive abilities may further establish the inverse relationship between MGD and cognitive functions.
6. More studies may lead to a better understanding of the collective effect of numerous genetic variants in determining complex traits.

The understanding of fundamental evolutionary theories will inevitably evolve into a more refined and comprehensive framework. This will greatly enhance humanity's ability to understand our shared destiny, both in the present and future, and contribute to the creation of a global community with a shared future.

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). We extend our gratitude to Shaoyuan Wu and the reviewers for the critical review of 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. Lamarck JB. Zoological philosophy: an exposition with regard to the natural history of animals. Chicago: University of Chicago Press; 1984.
2. Lewontin R. *What is evolutionary theory*. 2003. Santa Fe Institute. Available from: https://www.youtube.com/watch?v=n6W_FzjaKlw. (Accessed date: November 11, 2003)
3. Newman S. *Remembering Richard Lewontin*. Biopolitical Times, Center for Genetics and Society; 2021. Available from: <https://www.geneticsandsociety.org/biopolitical-times/remembering-richard-lewontin-1929-2021>. (Accessed date: July 6, 2001)
4. Simpson GG. *Tempo and mode in evolution*. New York: Columbia University Press; 1944.
5. Bickel D. Testing hypotheses of molecular evolution. *Phylogenetic trees and molecular evolution. Springer briefs in systems biology*. Cham: Springer; 2022. p. 71–77.
6. 黄石 和 朱作斌, 系统生物学与进化理论. 系统生物学, ed. 朱作斌, 张潇, and 王亮. 2022, 南京: 东南大学出版社.
7. Zuckerkandl E, Pauling L. Molecular disease, evolution, and genetic heterogeneity. In: Kasha M, Pullman B. (eds.) *Horizons in biochemistry*. New York: Academic Press; 1962. p.189–225.
8. Margoliash E. Primary structure and evolution of cytochrome C. *Proceedings of the National academy of Sciences of the United States of America*. 1963;50: 672–679. doi: 10.1073/pnas.50.4.672
9. Doolittle RF, Blombaeck B. Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature*. 1964;202: 147–152. doi: 10.1038/202147a0
10. Kumar S. Molecular clocks: four decades of evolution. *Nature Reviews Genetics*. 2005;6(8): 654–662. doi: 10.1038/nrg1659
11. Luo D, Huang S. The genetic equidistance phenomenon at the proteomic level. *Genomics*. 2016;108(1): 25–30. doi: 10.1016/j.ygeno.2016.03.002
12. Yuan D, Huang S. Genetic equidistance at nucleotide level. *Genomics*. 2017;109: 192–195. doi: 10.1016/j.ygeno.2017.03.002
13. Copley RR, Schultz J, Ponting CP, Bork P. Protein families in multicellular organisms. *Current Opinion in Structural Biology*. 1999;9: 408–415. doi: 10.1016/S0959-440X(99)80055-4
14. Denton M. *Evolution: a theory in crisis*. Chevy Chase, MD: Adler & Adler; 1985.
15. Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, et al. Evolution of the germline mutation rate across vertebrates. *Nature*. 2023;615(7951): 285–291. doi: 10.1038/s41586-023-05752-y
16. Brownstein CD, MacGuigan DJ, Kim D, Orr O, Yang L, David SR, et al. The genomic signatures of evolutionary stasis. *Evolution; International Journal of Organic Evolution*. 2024;78(5): 821–834. doi: 10.1093/evolut/qpae028
17. Mello B, Schrago CG. Modeling substitution rate evolution across lineages and relaxing the molecular clock. *Genome Biology and Evolution*. 2024;16(9): evae199. doi: 10.1093/gbe/evae199
18. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biology*. 2006;4(5): e88. doi: 10.1371/journal.pbio.0040088
19. Van Valen L. Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution*. 1974;3: 89–101. doi: 10.1007/BF01796554

20. Clarke B. Darwinian evolution of proteins. *Science*. 1970;168(934): 1009–1011. doi: 10.1126/science.168.3934.1009
21. Richmond RC. Non-Darwinian evolution: a critique. *Nature*. 1970;225(5237): 1025–1028. doi: 10.1038/2251025a0
22. Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217(5129): 624–626. doi: 10.1038/217624a0
23. Kimura M, Ohta T. On the rate of molecular evolution. *Journal of Molecular Evolution*. 1971;1: 1–17. doi: 10.1007/BF01659390
24. King JL, Jukes TH. Non-Darwinian evolution. *Science*. 1969;164: 788–798. doi: 10.1126/science.164.3881.788
25. Ayala FJ. Molecular clock mirages. *Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology*. 1999;21(1): 71–75. doi: 10.1002/(SICI)1521-1878(199901)21:1<71::AID-BIES9>3.0.CO;2-B
26. Pulquerio MJ, Nichols RA. Dates from the molecular clock: how wrong can we be? *Trends in Ecology & Evolution*. 2007;22(4): 180–184. doi: 10.1016/j.tree.2006.11.013
27. Kimura M, Ohta T. Protein polymorphism as a phase of molecular evolution. *Nature*. 1971;229: 467–479. doi: 10.1038/229467a0
28. Kimura M. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press; 1983.
29. Leigh EG Jr. Neutral theory: a historical perspective. *Journal of Evolutionary Biology*. 2007;20(6): 2075–2091. doi: 10.1111/j.1420-9101.2007.01410.x
30. Fisher RA. *The genetical theory of natural selection*. Oxford, UK: Oxford University Press; 1930.
31. Wright S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*. 1932;1: 356–366.
32. Haldane JBS. The cost of natural selection. *Journal of Genetics*. 1957;55: 511–524. doi: 10.1007/BF02984069
33. Sueoka N. On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the National academy of Sciences of the United States of America*. 1962;48: 582–592. doi: 10.1073/pnas.48.4.582
34. Freese E. On the evolution of the base composition of DNA. *Journal of Theoretical Biology*. 1962;3(1): 82–101. doi: 10.1016/S0022-5193(62)80005-8
35. Lewontin RC. *The genetic basis of evolutionary change*. New York and London: Columbia University Press; 1974.
36. Lewontin RC, Hubby JL. 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. doi: 10.1093/genetics/54.2.595
37. Dietrich MR. The origins of the neutral theory of molecular evolution. *Journal of the History of Biology*. 1994;27(1): 21–59. doi: 10.1007/BF01058626
38. Suarez E, Barahona A. The experimental roots of the neutral theory of molecular evolution. *History and Philosophy of the Life Sciences*. 1996;18: 55–81.
39. Palazzo AF, Gregory TR. The case for junk DNA. *Plos Genetics*. 2014;10(5): e1004351. doi: 10.1371/journal.pgen.1004351
40. Rands CM, Meader S, Ponting CP, Lunter G. 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *Plos Genetics*. 2014;10(7): e1004525. doi: 10.1371/journal.pgen.1004525
41. Doolittle WF, Brunet TDP. On causal roles and selected effects: our genome is mostly junk. *BMC Biology*. 2017;15(1): 116. doi: 10.1186/s12915-017-0460-9
42. Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. *Genetics*. 1964;49: 725–738. doi: 10.1093/genetics/49.4.725
43. Demeulemeester J, Dentre SC, Gerstung M, Van Loo P. Biallelic mutations in cancer genomes reveal local mutational determinants. *Nature Genetics*. 2022;54(2): 128–133. doi: 10.1038/s41588-021-01005-8
44. Harpak A, Bhaskar A, Pritchard JK. Mutation rate variation is a primary determinant of the distribution of allele frequencies in humans. *Plos Genetics*. 2016;12(12): e1006489. doi: 10.1371/journal.pgen.1006489
45. Porubsky D, Dashnow H, Sasani TA, Logsdon GA, Hallast P, Noyes MD, et al. A familial, telomere-to-telomere reference for human de novo mutation and recombination from a four-generation pedigree. *bioRxiv*. 2024: 2024.08.05.606142. doi: 10.1101/2024.08.05.606142
46. Lynch M, Wei W, Ye Z, Pfrender M. The genome-wide signature of short-term temporal selection. *Proceedings of the National academy of Sciences of the United States of America*. 2024;121(28): e2307107121. doi: 10.1073/pnas.2307107121
47. Yuan D, Zhu Z, Tan X, Liang J, Zeng C, Zhang J, et al. Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. *Science China Life Sciences*. 2014;57(9): 876–888. doi: 10.1007/s11427-014-4704-4
48. Zhu Z, Yuan D, Luo D, Lu X, Huang S. Enrichment of minor alleles of common SNPs and improved risk prediction for Parkinson's disease. *Plos One*. 2015;10(7): e0133421. doi: 10.1371/journal.pone.0133421
49. Zhu Z, Man X, Xia M, Huang Y, Yuan D, Huang S. Collective effects of SNPs on transgenerational inheritance in *Caenorhabditis elegans* and budding yeast. *Genomics*. 2015;106(1): 23–29. doi: 10.1016/j.ygeno.2015.04.002
50. Yuan D, Zhu Z, Tan X, Liang J, Zeng C, Zhang J, et al. Minor alleles of common SNPs quantitatively affect traits/diseases and are under both positive and negative selection. *arXiv*. 2012;1209.2911. doi: 10.1007/s11427-014-4704-4
51. Zhu Z, Lu Q, Wang J, Huang S. Collective effects of common SNPs in foraging decisions in *Caenorhabditis elegans* and an integrative method of identification of candidate genes. *Scientific Reports*. 2015;5: 16904. doi: 10.1038/srep16904
52. Kimura M, Ohta T. On some principles governing molecular evolution. *Proceedings of the National academy of Sciences of the*

- United States of America*. 1974;71(7): 2848–2852. doi: 10.1073/pnas.71.7.2848
53. Quinodoz SA, Jachowicz JW, Bhat P, Ollikainen N, Banerjee AK, Goronzy IN, et al. RNA promotes the formation of spatial compartments in the nucleus. *Cell*. 2021;184(23): 5775–5790.e30. doi: 10.1016/j.cell.2021.10.014
54. Basu A, Bobrovnikov DG, Cieza B, Arcon JP, Qureshi Z, Orozco M, et al. Deciphering the mechanical code of the genome and epigenome. *Nature Structural & Molecular Biology*. 2022;29(12): 1178–1187. doi: 10.1038/s41594-022-00877-6
55. Duttke SH, Guzman C, Chang M, Delos Santos NP, McDonald BR, Xie J, et al. Position-dependent function of human sequence-specific transcription factors. *Nature*. 2024;631(8022): 891–898. doi: 10.1038/s41586-024-07662-z
56. Eder M, Moene CJI, Dauban L, Leemans C, Steensel BV. Functional maps of a genomic locus reveal confinement of an Enhancer by its target gene. *bioRxiv*. 2024. doi: 10.1101/2024.08.26.609360
57. Du AY, Chobirko JD, Zhuo X, Feschotte C, Wang T. Regulatory transposable elements in the encyclopedia of DNA elements. *Nature Communications*. 2024;15(1): 7594. doi: 10.1038/s41467-024-51921-6
58. Kimura M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research*. 1968;11(3): 247–269. doi: 10.1017/S0016672300011459
59. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature*. 2022;606(7915): 725–731. doi: 10.1038/s41586-022-04823-w
60. Lu J, Wu CI. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proceedings of the National academy of Sciences of the United States of America*. 2005;102(11): 4063–4067. doi: 10.1073/pnas.0500436102
61. Nyerges A, Chiappino-Pepe A, Budnik B, Baas-Thomas M, Flynn R, Yan S, et al. Synthetic genomes unveil the effects of synonymous recoding. *bioRxiv*. 2024: 2024.06.16.599206. doi: 10.1101/2024.06.16.599206
62. Rodriguez A, Diehl JD, Wright GS, Bonar CD, Lundgren TJ, Moss MJ, et al. Synonymous codon substitutions modulate transcription and translation of a divergent upstream gene by modulating antisense RNA production. *Proceedings of the National academy of Sciences of the United States of America*. 2024;121(36): e2405510121. doi: 10.1073/pnas.2405510121
63. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, et al. Revisiting an old riddle: what determines genetic diversity levels within species? *Plos Biology*. 2012;10(9): e1001388. doi: 10.1371/journal.pbio.1001388
64. Buffalo V. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's paradox. *eLife*. 2021;10: e67509. doi: 10.7554/eLife.67509
65. Kern AD, Hahn MW. The neutral theory in light of natural selection. *Molecular Biology and Evolution*. 2018;35(6): 1366–1371. doi: 10.1093/molbev/msy092
66. Kreitman M. The neutral theory is dead. Long live the neutral theory. *Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology*. 1996;18(8): 678–683. discussion 683. doi: 10.1002/bies.950180812
67. Cann RL, Stoneking AC, Wilson AC. Mitochondrial DNA and human evolution. *Nature*. 1987;325: 31–36. doi: 10.1038/325031a0
68. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution; International Journal of Organic Evolution*. 1985;39(4): 783–791. doi: 10.2307/2408678
69. Huang S. The genetic equidistance result of molecular evolution is independent of mutation rates. *Journal of Computer Science and Systems Biology*. 2008;1: 092–102. doi: 10.4172/jcsb.1000009
70. Avise JC. Molecular markers, natural history and evolution. New York, NY: Springer; 1994.
71. Nei M, Kumar S. *Molecular evolution and phylogenetics*. New York: Oxford University Press; 2000.
72. Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature*. 1973;246(5428): 96–98. doi: 10.1038/246096a0
73. Huang S. Histone methyltransferases, diet nutrients, and tumor suppressors. *Nature Reviews Cancer*. 2002;2: 469–476. doi: 10.1038/nrc819
74. Huang S. Histone methylation and the initiation of cancer. In: Tollefsbol T. (ed.) *Cancer epigenetics*. New York: CRC Press; 2008. p. 109–158.
75. Huang S. Inverse relationship between genetic diversity and epigenetic complexity. *Nature Precedings*. 2009. doi: 10.1038/npre.2009.1751.2
76. 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. doi: 10.1162/BIOT_a_00021
77. Santoni G, Astori S, Leleu M, Glauser L, Zamora SA, Schioppa M, et al. Chromatin plasticity predetermines neuronal eligibility for memory trace formation. *Science*. 2024;385(6707): eadg9982. doi: 10.1126/science.adg9982
78. Aquadro CF. Why is the genome variable? Insights from *Drosophila*. *Trends in Genetics*. 1992;8(10): 355–362. doi: 10.1016/0168-9525(92)90281-8
79. Lewontin RC. Twenty-five years ago in genetics: electrophoresis in the development of evolutionary genetics: milestone or millstone? *Genetics*. 1991;128(4): 657–662. doi: 10.1093/genetics/128.4.657
80. Bateson W. Materials for the study of variation treated with especial regard to discontinuity in the origin of species. London: Macmillan; 1894.

81. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*. 1979;76: 5269–5273. doi: 10.1073/pnas.76.10.5269
82. Steux C, Szpiech ZA. The maintenance of deleterious variation in wild Chinese rhesus macaques. *Genome Biology and Evolution*. 2024;16(6): evae115. doi: 10.1093/gbe/evae115
83. Ellegren H, Galtier N. Determinants of genetic diversity. *Nature Reviews Genetics*. 2016;17(7): 422–433. doi: 10.1038/nrg.2016.58
84. Nevo E. Genetic diversity. In: Levin SA. (ed.) *Encyclopedia of biodiversity*. Amsterdam: Elsevier Inc; 2001. p. 662–677. doi: 10.1016/B978-0-12-384719-5.00065-4
85. Yang J, Lusk R, Li WH. Organismal complexity, protein complexity, and gene duplicability. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(26): 15661–15665. doi: 10.1073/pnas.2536672100
86. Levine M, Tjian R. Transcription regulation and animal diversity. *Nature*. 2003;424(6945): 147–151. doi: 10.1038/nature01763
87. Vinogradov AE, Anatskaya OV. Organismal complexity, cell differentiation and gene expression: human over mouse. *Nucleic Acids Research*. 2007;35(19): 6350–6356. doi: 10.1093/nar/gkm723
88. Bonner JT. Perspective: the size-complexity rule. *Evolution; International Journal of Organic Evolution*. 2004;58: 1883–1890. doi: 10.1111/j.0014-3820.2004.tb00476.x
89. Carroll SB. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*. 2001;409(6823): 1102–1109. doi: 10.1038/35059227
90. McShea DW. Metazoan complexity and evolution: is there a trend? *Evolution; International Journal of Organic Evolution*. 1996;50: 477–492. doi: 10.2307/2410824
91. Bonner JT. *The evolution of complexity*. Princeton, NJ: Princeton University Press; 1988.
92. Vogel C, Chothia C. Protein family expansions and biological complexity. *PLoS Computational Biology*. 2006;2(5): e48. doi: 10.1371/journal.pcbi.0020048
93. Remy JJ. Stable inheritance of an acquired behavior in *Caenorhabditis elegans*. *Current Biology: CB*. 2010;20(20): R877–R878. doi: 10.1016/j.cub.2010.08.013
94. Anway MD, Cupp AS, Uzumcu M, Skinner MK. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*. 2005;308(5727): 1466–1469. doi: 10.1126/science.1108190
95. Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. *New England Journal of Medicine*. 2007;356(7): 697–705. doi: 10.1056/NEJMoa064522
96. Cropley JE, Suter CM, Beckman KB, Martin DI. Germ-line epigenetic modification of the murine A_{vy} allele by nutritional supplementation. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(46): 17308–17312. doi: 10.1073/pnas.0607090103
97. Huypens P, Sass S, Wu M, Dyckhoff D, Tschöp M, Theis F, et al. Epigenetic germline inheritance of diet-induced obesity and insulin resistance. *Nature Genetics*. 2016;48(5): 497–499. doi: 10.1038/ng.3527
98. Fitz-James MH, Cavalli G. Molecular mechanisms of transgenerational epigenetic inheritance. *Nature Reviews Genetics*. 2022;23(6): 325–341. doi: 10.1038/s41576-021-00438-5
99. Yu J, Zhang Y, Fang Y, Paulo JA, Yaghoubi D, Hua X, et al. A replisome-associated histone H3-H4 chaperone required for epigenetic inheritance. *Cell*. 2024;187: 5010.e–5028.e. doi: 10.1016/j.cell.2024.07.006
100. Hu T, Long M, Yuan D, Zhu Z, Huang Y, Huang S. The genetic equidistance result, misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. *Science China Life Sciences*. 2013;56: 254–261. doi: 10.1007/s11427-013-4452-x
101. Huang S. New thoughts on an old riddle: what determines genetic diversity within and between species? *Genomics*. 2016;108(1): 3–10. doi: 10.1016/j.ygeno.2016.01.008
102. Huang S. The maximum genetic diversity theory of molecular evolution. *Communications in Information and Systems*. 2023;23: 359–392. doi: 10.4310/CIS.2023.v23.n4.a1
103. Zhu Z, Han C, Huang S. New insights shed light on the enigma of genetic diversity and species complexity. *Science China Life Sciences*. 2024;67: 2774–2776. doi: 10.1007/s11427-023-2610-2
104. Orr HA. Adaptation and the cost of complexity. *Evolution; International Journal of Organic Evolution*. 2000;54(1): 13–20. doi: 10.1111/j.0014-3820.2000.tb00002.x
105. Parts L, Batté A, Lopes M, Yuen MW, Laver M, San Luis BJ, et al. Natural variants suppress mutations in hundreds of essential genes. *Molecular Systems Biology*. 2021;17(5): e10138. doi: 10.15252/msb.202010138
106. Gould SJ, Eldredge N. Punctuated equilibrium comes of age. *Nature*. 1993;366(6452): 223–227. doi: 10.1038/366223a0
107. Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, et al. Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science*. 2022;375(6586): eabj7484. doi: 10.1126/science.abj7484
108. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, et al. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science*. 2012;336(6085): 1157–1160. doi: 10.1126/science.1217405
109. Wang M, Wang D, Yu J, Huang S. 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
110. Bergsten J. A review of long-branch attraction. *Cladistics: the International Journal of the Willi Hennig Society*. 2005;21: 163–193. doi: 10.1111/j.1096-0031.2005.00059.x
111. Huang S. Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing

- tarsiers. *Science China Life Sciences*. 2012;55: 709–725. doi: 10.1007/s11427-012-4350-7
112. 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
 113. Chen Z, Baeza JA, Chen C, Gonzalez MT, González VL, Greve C, et al. A genome-based phylogeny for Mollusca is concordant with fossils and morphology. *Science*. 2025;387(6737): 1001–1007. doi: 10.1126/science.ads0215
 114. Mao Y, Harvey WT, Porubsky D, Munson KM, Hoekzema K, Lewis AP, et al. Structurally divergent and recurrently mutated regions of primate genomes. *Cell*. 2024;187(6): 1547–1562.e13. doi: 10.1016/j.cell.2024.01.052
 115. Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, et al. The landscape of tolerated genetic variation in humans and primates. *Science (New York, N.Y.)*. 2023;380(6648): eabn8153. doi: 10.1126/science.abn8197
 116. Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science (New York, N.Y.)*. 2007;316(5822): 222–234. doi: 10.1126/science.1139247
 117. Kuiken C, Korber B, Shafer RW. HIV sequence databases. *AIDS Reviews*. 2003;5(1): 52–61.
 118. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 2005;437(7055): 69–87. doi: 10.1038/nature04072
 119. Zhang S, Xu N, Fu L, Yang X, Li Y, Yang Z, et al. Comparative genomics of macaques and integrated insights into genetic variation and population history. *bioRxiv*. 2024: 2024.04.07.588379. doi: 10.1101/2024.04.07.588379
 120. Orkin JD, Kuderna LFK, Hermosilla-Albala N, Fontseré C, Aylward ML, Janiak MC, et al. Ecological and anthropogenic effects on the genomic diversity of lemurs in Madagascar. *Nature Ecology and Evolution*. 2025;9(1): 42–56. doi: 10.1038/s41559-024-02596-1
 121. Oberstaller J, Xu S, Naskar D, Zhang M, Wang C, Gibbons J, et al. Supersaturation mutagenesis reveals adaptive rewiring of essential genes among malaria parasites. *Science*. 2025;387(6734): eadq7347. doi: 10.1126/science.adq7347
 122. Camellato BR, Brosh R, Ashe HJ, Maurano MT, Boeke JD. Synthetic reversed sequences reveal default genomic states. *Nature*. 2024;628(8007): 373–380. doi: 10.1038/s41586-024-07128-2
 123. Wang M, Huang S. The collective effects of genetic variants and complex traits. *Journal of Human Genetics*. 2023;68: 255–262. doi: 10.1038/s10038-022-01105-1
 124. Chen CY, Tian R, Ge T, Lam M, Sanchez-Andrade G, Singh T, et al. The impact of rare protein coding genetic variation on adult cognitive function. *Nature Genetics*. 2023;55(6): 927–938. doi: 10.1038/s41588-023-01398-8
 125. Ganna A, Genovese G, Howrigan DP, Byrnes A, Kurki M, Zekavat SM, et al. Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nature Neuroscience*. 2016;19(12): 1563–1565. doi: 10.1038/nn.4404
 126. Ganna A, Satterstrom FK, Zekavat SM, Das I, Kurki MI, Churchhouse C, et al. Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. *American Journal of Human Genetics*. 2018;102(6): 1204–1211. doi: 10.1016/j.ajhg.2018.05.002
 127. Sha Z, Sun KY, Jung B, Barzilay R, Moore TM, Almay L, et al. The copy number variant architecture of psychopathology and cognitive development in the ABCD® study. *medRxiv*. 2024: 2024.05.14.24307376. doi: 10.1101/2024.05.14.24307376
 128. Wainberg M, Forde NJ, Mansour S, Kerrebijn I, Medland SE, Hawco C, et al. Genetic architecture of the structural connectome. *Nature Communications*. 2024;15(1): 1962. doi: 10.1038/s41467-024-46023-2
 129. Zhang S, Xu N, Fu L, Yang X, Ma K, Li Y, et al. Integrated analysis of the complete sequence of a macaque genome. *Nature*. 2025. doi: 10.1038/s41586-025-08596-w
 130. Joly M, Micheletta J, De Marco A, Langermans JA, Sterck EHM, Waller BM. Comparing physical and social cognitive skills in macaque species with different degrees of social tolerance. *Proceedings. Biological Sciences/the Royal Society*. 2017;284(1862): 20162738. doi: 10.1098/rspb.2016.2738
 131. Kuderna LFK, Gao H, Janiak MC, Kuhlwillm M, Orkin JD, Bataillon T, et al. A global catalog of whole-genome diversity from 233 primate species. *Science*. 2023;380(6648): 906–913. doi: 10.1126/science.abn7829
 132. de Manuel M, Kuhlwillm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*. 2016;354(6311): 477–481. doi: 10.1126/science.aag2602
 133. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009;324(5930): 1035–1044. doi: 10.1126/science.1172257
 134. Lynn R. *Race differences in intelligence*. Augusta, GA: Washington Summit Publishers; 2006.
 135. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell*. 2016;167(3): 657–669.e21. doi: 10.1016/j.cell.2016.09.025
 136. Quach H, Rotival M, Pothlichet J, Loh YE, Dannemann M, Zidane N, et al. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell*. 2016;167(3): 643–656.e17. doi: 10.1016/j.cell.2016.09.024

137. Lei X, Yuan D, Zhu Z, Huang S. Collective effects of common SNPs and risk prediction in lung cancer. *Heredity*. 2018;121: 537–547. doi: 10.1038/s41437-018-0063-4
138. He P, Lei X, Yuan D, Zhu Z, Huang S. Accumulation of minor alleles and risk prediction in schizophrenia. *Scientific Reports*. 2017;7(1): 11661. doi: 10.1038/s41598-017-12104-0
139. Sanjak JS, Sidorenko J, Robinson MR, Thornton KR, Visscher PM. Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National academy of Sciences of the United States of America*. 2018;115(1): 151–156. doi: 10.1073/pnas.1707227114
140. Amos W, Elhaik E. Unexpected D-tour ahead: why the D-statistic, applied to humans, measures mutation rate variation not Neanderthal introgression. *bioRxiv*. 2025. Available from: <https://www.biorxiv.org/content/10.1101/2024.12.31.630954v2>. doi: 10.1101/2024.12.31.630954 (Accessed date: December 31, 2024)
141. Horton CA, Alexandari AM, Hayes MGB, Marklund E, Schaepe JM, Aditham AK, et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*. 2023;381(6664): eadd1250. doi: 10.1126/science.add1250
142. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414): 57–74. doi: 10.1038/nature11247
143. Mattick JS. A Kuhnian revolution in molecular biology: most genes in complex organisms express regulatory RNAs. *Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology*. 2023;45(9): e2300080. doi: 10.1002/bies.202300080
144. Meneu L, Chopard C, Serizay J, Westbrook A, Routhier E, Ruault M, et al. Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus. *Science*. 2025;387(6734): eadm9466. doi: 10.1126/science.adm9466
145. Li A, Evans DS, Cummings S, Ideker T. Somatic mutation as an explanation for epigenetic aging. *Nature Aging*. 2025. doi: 10.1038/s43587-024-00794-x
146. Lake NJ, Ma K, Liu W, Battle SL, Laricchia KM, Tiao G, et al. Quantifying constraint in the human mitochondrial genome. *Nature*. 2024;635: 390–397. doi: 10.1038/s41586-024-08048-x
147. Couce A, Limdi A, Magnan M, Owen SV, Herren CM, Lenski RE, et al. Changing fitness effects of mutations through long-term bacterial evolution. *Science*. 2024;383(6681): eadd1417. doi: 10.1126/science.add1417
148. Halabi N, Rivoire O, Leibler S, Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. *Cell*. 2009;138(4): 774–786. doi: 10.1016/j.cell.2009.07.038
149. Jeong H, Dishuck PC, Yoo D, Harvey WT, Munson KM, Lewis AP, et al. Structural polymorphism and diversity of human segmental duplications. *Nature Genetics*. 2025;57(2): 390–401. doi: 10.1038/s41588-024-02051-8
150. Simpson KJ, Mian S, Forrester EJ, Hackel J, Morton JA, Leitch AR, et al. Bigger genomes provide environment-dependent growth benefits in grasses. *The New Phytologist*. 2024;244(5): 2049–2061. doi: 10.1111/nph.20150
151. Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, et al. The mutational landscape of human somatic and germline cells. *Nature*. 2021;597(7876): 381–386. doi: 10.1038/s41586-021-03822-7
152. Ru Y, Deng X, Chen J, Zhang L, Xu Z, Lv Q, et al. Maternal age enhances purifying selection on pathogenic mutations in complex I genes of mammalian mtDNA. *Nature Aging*. 2024;4(9): 1211–1230. doi: 10.1038/s43587-024-00672-6
153. Zeller E, Timmermann A, Yun KS, Raia P, Stein K, Ruan J. Human adaptation to diverse biomes over the past 3 million years. *Science*. 2023;380(6645): 604–608. doi: 10.1126/science.abq1288
154. Exposito-Alonso M, Booker TR, Czech L, Gillespie L, Hateley S, Kyriazis CC, et al. Genetic diversity loss in the Anthropocene. *Science*. 2022;377(6613): 1431–1435. doi: 10.1126/science.abn5642
155. Gross N, Maestre FT, Liancourt P, Berdugo M, Martin R, Gozalo B, et al. Unforeseen plant phenotypic diversity in a dry and grazed world. *Nature*. 2024;632(8026): 808–814. doi: 10.1038/s41586-024-07731-3
156. Jenkin F. The origin of species. *The North British Review*. 1867;46: 277–318.
157. Goldschmidt R. *The material basis of evolution*. New Haven, CT: Yale University Press; 1940.
158. Forsdyke DR. *Evolutionary bioinformatics*. New York: Springer; 2011.
159. Heng HH. *Genome chaos: rethinking genetics, evolution, and molecular medicine*. Cambridge, MA, USA: Academic Press Elsevier; 2019.
160. Zhang Y. The genetic equidistance and maximum genetic diversity hypothesis: smoke and mirrors? *bioRxiv*. 2023. doi: 10.1101/2023.02.14.528494
161. Huang S. A rebuttal to Zhang's critique of the genetic equidistance phenomenon and maximum genetic diversity hypothesis. *Zenodo*. 2025. doi: 10.5281/zenodo.14927073
162. Yuan D, Lei X, Gui Y, Zhu Z, Wang M, Zhang Y, et al. Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. *bioRxiv*. 2017. doi: 10.1101/101410
163. Huang S. Examining models of modern human origins through the analysis of 43 fully sequenced human Y chromosomes, *Communications in Information and Systems*, in press. *bioRxiv*. 2023. doi: 10.1101/2023.11.09.566475