The genetic equidistance result: misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later

HU TaoBo[1,2], LONG MengPing[1,3], YUAN DeJian[1], ZHU ZhuBing[1], HUANG YiMin[1] & HUANG Shi[1*]

[1]*State Key Laboratory of Medical Genetics, Xiangya Medical School, Central South University, Changsha 410078, China;*
[2]*The Second Xiangya Hospital, Xiangya Medical School, Central South University, Changsha 410011, China;*
[3]*Xiangya Hospital, Xiangya Medical School, Central South University, Changsha 410008, China*

In 1963, Margoliash discovered the unexpected genetic equidistance result after comparing cytochrome c sequences from different species. This finding, together with the hemoglobin analyses of Zuckerkandl and Pauling in 1962, directly inspired the *ad hoc* molecular clock hypothesis. Unfortunately, however, many biologists have since mistakenly viewed the molecular clock as a genuine reality, which in turn inspired Kimura, King, and Jukes to propose the neutral theory of molecular evolution. Many years of studies have found numerous contradictions to the theory, and few today believe in a universal constant clock. What is being neglected, however, is that the failure of the molecular clock hypothesis has left the original equidistance result an unsolved mystery. In recent years, we fortuitously rediscovered the equidistance result, which remains unknown to nearly all researchers. Incorporating the proven virtues of existing evolutionary theories and introducing the novel concept of maximum genetic diversity, we proposed a more complete hypothesis of evolutionary genetics and reinterpreted the equidistance result and other major evolutionary phenomena. The hypothesis may rewrite molecular phylogeny and population genetics and solve major biomedical problems that challenge the existing framework of evolutionary biology.

genetic equidistance, overlap feature, evolution, molecular clock, neutral theory, maximum genetic diversity hypothesis, microevolution, macroevolution

Among the most interesting phenomena in biology are reproduction, heredity, and evolution. Biological evolution includes both phenotypes and molecules, which are modeled by the theory of natural selection and the neutral theory, respectively. Together, the two make up the modern evolutionary theory (MET). The neutral theory was proposed by Kimura, Jukes, and King to explain the observation that molecules evolved at a steady, clock-like rate [1,2]. But is the molecular clock a genuine universal phenomenon?

1 A brief history of the molecular clock hypothesis and the neutral theory: how the molecular clock replaced the genetic equidistance phenomenon

Genetic distance is a measure of molecular differences between species and is represented by the percentage differ-

*Corresponding author (email: huangshi@sklmg.edu.cn)

ence in orthologous protein or DNA sequences. For a gene shared among any three species of different complexity (as inferred by intuition), one can ask two related questions, which are really different aspects of the same question. One is easily answered by anyone with a middle-school education: which, frog or fish, is closer to human in sequence identity of a given shared gene? In contrast, most people, including most professional biologists, would fail to answer the other: which, frog or human, is closer to fish? Here, the same intuition that leads most to answer the first question correctly would inevitably led them to the wrong answer that frog is closer to fish. So, why can most people not answer this question when it has been extensively studied for nearly half a century? The straightforward conclusion is that something fundamental, at the level of middle-school textbooks, may be missing in our present knowledge of biology.

For any three or more species of different epigenetic or organismal complexity (as judged by intuition), one can perform two kinds of sequence alignment. The first aligns data from a complex organism, such as human, against those from simpler or less complex species that evolved earlier, such as frog and fish, and demonstrates that humans are more closely related to frogs than to fishes (Figure 1(A)). The second method aligns data from a simpler organism, such as fish, against those from more complex ones, such as frog and human, and shows that fishes are approximately equidistant to frogs and humans (Figure 1(B)). This genetic equidistance result was not predicted by anyone. Based on
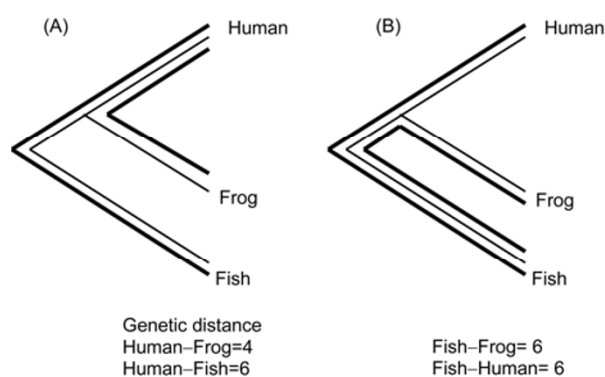
Figure 1 The molecular clock interpretation of genetic distance. The cladogram assumes branch length (time) equals genetic distance (sequence dissimilarity), because the other variable in determining genetic distance, mutation rate, is assumed by the molecular clock hypothesis to be similar or constant across all species. Thin black lines indicate branch length. Thick black lines indicate measured genetic distance between human and frog, human and fish, and fish and frog (in arbitrary schematic units). (A), Genetic distance measured using human as baseline, resulting in a linear progression of genetic distance. (B), Genetic distance using fish as baseline, resulting in all members of Tetrapoda being equally distant from fish. Both A and B can be interpreted equally well by the molecular clock assumption, but only A can also be interpreted by the theory of natural selection. As a result of teaching A rather than B as evidence of evolution, nearly all people with an evolutionary mindset interpret A via natural selection and readily predict A correctly but B incorrectly.

either of these two alignment results, simple mathematical calculations would imply a molecular clock, i.e., different species have very similar substitution rates.

Based on this introduction, we can now briefly review the history of the molecular clock hypothesis [3]. A breakthrough in protein sequencing methods in the late 1950s made possible the comparison of homologous protein sequences from different species. In 1962, Zuckerkandl and Pauling performed the first type of alignment described above using hemoglobin sequences and informally discussed the idea of a molecular clock [4]. Margoliash in 1963 aligned cytochrome c sequences from seven different species using both methods and made a formal statement of the molecular clock after noticing the genetic equidistance result [5]. The equidistance phenomenon has since been shown to hold for nearly all proteins and species [6,7].

The molecular clock interpretation of the equidistance result was *ad hoc* and not a deduction of any natural law or theory. The premise for using the formula $r=d/2t$ to deduce mutation rate (r) is that genetic distance (d) always positively correlates with time (t). But this unspoken assumption was never seriously debated and not based on any evidence. Numerous follow up studies, including ours, have falsified the concept of a universal molecular clock [6,8–18]. Work on fossil sequences also invalidated the molecular clock [19,20]. The equidistance result was found to be independent of mutation rate variations [6]. The molecular clock is now considered to only exist in limited situations, and almost no experts today acknowledge a universally constant clock. However, most researchers are unaware of the long overlooked equidistance result and do not realize the huge cost of a failure of the molecular clock hypothesis. The equidistance result has now become an unsolved puzzle.

The equidistance result has not appeared in textbooks or primary research literature since its initial discovery. It only appears in books by scholars who openly doubt evolution, and has been used to challenge evolutionary theory [21]. The interpretation of the equidistance result via the molecular clock misled many to treat the clock as a genuine reality. A number of hypotheses have been proposed to explain it [1,2,22–25]. The 'Neutral Theory', put forward in a 1968 paper by Kimura, and independently by King and Jukes, has become the favorite [1,2]. The abstract of the Kimura paper has only one sentence: "Calculating the rate of evolution in terms of nucleotide substitutions seems to give a value so high that many of the mutations involved must be neutral ones." But this calculation has two implicit assumptions that were taken for granted, without deliberation. One, as mentioned above, is that observed genetic distance always increases with time. The other is that every nucleotide in a genome is freely changeable (there are no nucleotide positions that would cause lethality when changed). Obviously, the apparent high rate of mutations of the so-called molecular clock directly inspired Kimura to propose the neutral theory. The theory posits that sequence differences between

species are due to random drift of neutral mutations that are neither beneficial nor deleterious. Neutral mutations and random drift, rather than natural selection, drive molecular evolution. The theory seems to predict a molecular clock measured in generations and Kimura in turn viewed the molecular clock as the best evidence for his theory [25], but this is obviously a tautology.

The neutral theory has remained highly controversial. Many results were found to invalidate it [10,26]. The observed substitution rate or molecular clock is measured in years, but the substitution rate predicted by the neutral theory is measured in generations. 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. Of course, one should acknowledge that the neutral theory is not without value and has proven valuable in limited situations [3,27].

2 The maximum genetic diversity hypothesis and the genetic equidistance phenomenon

We have been studying the genetics, epigenetics, and evolutionary theory of cancer. Since 2003, we have also had an interest in the general theory of evolution and the problem of complexity. Around 2005, open access to genome sequencing results of various species allowed us to compare sequences of fish, frog, and human for the RIZ1 tumor suppressor gene that we had been studying. We were naturally very surprised to find the genetic equidistance phenomenon and quickly realized the absurdity of the molecular clock interpretation. After three years of extensive studies, we published the maximum genetic diversity (MGD) hypothesis in 2008 and reinterpreted all major evolutionary phenomena, including the genetic equidistance result [8,28]. We also discovered a new aspect of the equidistance result, the overlap or coincident substitution feature, which has never been explained by any theory [27]. The feature means that a maximum distance between species will ultimately be reached. Recurrent mutations at the same sites do not increase genetic distance. Sequences with such mutations do not satisfy the infinite sites assumption of the neutral theory, but constitute the majority of any genome.

It seems regrettable that we failed to note the inspiration for the MGD idea, but the subconscious influence of the traditional Chinese culture may have played a role. Our past research discovered the RIZ1- or PRDM-family histone methyltransferases and the phenomenon of two differently-sized protein products of the RIZ1 gene, one that suppresses tumors and another that does the opposite. Our review article in 2000 used the phrase "yin-yang" in the title [29]. In 2002, to highlight another review paper of ours, the editors of "Nature Review Cancer" printed a Tai-Ji yin-yang symbol on the table-of-contents page of the issue containing

our paper [30]. From these personal experiences, we gained great admiration for the remarkable intuitive and perceptive power of the ancient Chinese saints. We naturally think that any fundamental unipolar phenomenon in nature must have an opposing phenomenon to keep it from going to extremes. It is impossible for genetic distance to increase indefinitely. So, what limits it?

Obviously, differences in biological complexity exist. The study of evolution must account for biological complexity, but its definition is controversial, and no one definition is universally accepted or experimentally proven. Past evolutionary research has overlooked biological complexity and has not considered it an important issue. Of course, all theories have unproven assumptions that are essentially subjective intuitions or axioms that have yet to be disproven. The practical way to judge a theory is not whether an assumption is proven but whether it can best explain nature without contradiction. An incorrect assumption may explain some phenomena by chance but would not be able to account for all relevant observations. Thus, the standard for judging a theory must and can only be whether it explains all relevant phenomena without contradiction. The failure of the neutral theory to account for the genetic equidistance result is sufficient to deem some of its assumptions incorrect.

Heritable phenotypes are determined by the primary sequences of DNA, or genotypes, as well as how the genotypes are used or expressed, often termed epigenotypes or epigenetic programs. Genes may be viewed as the major building blocks of biological organisms and epigenetic programs as the architectural plans. Epigenetic programs are not only inherited during mitotic cell division but are also transmitted through the germline to the next generation, and sometimes over many generations [31–34]. We therefore define complex organisms as those that have complex epigenetic programs in terms of the number of cell types and epigenetic molecules.

This definition of biological complexity is consistent with the general definition of complex systems proposed by a pioneer in complexity research, QIAN Xuesen, that a complex system is a large one comprising many subsystems with complex relationships among them [35]. Furthermore, we assume that humans can in most cases intuitively and correctly judge complexity differences among species. That humans are more complex than monkeys is an intuitive assumption that may not be verifiable today but has not been and is unlikely to be contradicted in the future. The ancient Chinese sages viewed humans as the soul of all species and equal to the creative natural forces of Heaven and Earth. This belief may be subjective, but all axioms of science are subjective and only relevant to humans. To assume the non-existence of complexity differences or a lack of evolutionary progress towards complexity is even more subjective and flawed a notion because it is counter-intuitive.

Based on this definition of complexity, we proposed that

simpler systems or machines can tolerate more variation in their building blocks above the atomic level. The more complex the system, the more restriction would be placed on the choice of building blocks. An equivalent concept in constructing biological organisms is easy to deduce. Thus, populations of simple organisms with low epigenetic complexity can tolerate more DNA variation and have higher genetic diversity. There exists an inverse relationship between genetic diversity and epigenetic complexity. Genetic diversity is defined here as genetic distance or dissimilarity in DNA or protein sequences between individuals or species.

A related assumption is that any system can allow a limited level of random errors or noise in constructing its parts. Such errors may be beneficial, deleterious, or neutral, depending on circumstances. Limited errors at an optimum level are more likely to be beneficial than deleterious because they are, after all, within tolerable levels and confer economy in construction and the strongest possible adaptive capacity or robustness to environmental challenges. Obviously, one only needs to substitute "errors in building blocks" to "genetic diversity" to infer the equivalent concept in biology. This intuition underlies the MET.

Based on these intuitive concepts, we deduced the MGD hypothesis to unify genetics and evolution. First, any species has a certain level of epigenetic complexity, which limits the amount of genetic diversity. Genetic diversity also puts a limit on the level of epigenetic complexity. Second, the maximum genetic diversity allowed for a complex organism is smaller than that allowed for a simple organism. Finally, macroevolution from simple to complex organisms is associated with a punctuated increase in epigenetic complexity and, consequently, a punctuated loss in genetic diversity. However, microevolution involves no major changes in epigenetic complexity, the linear stage of which can be basically described by the MET.

The MGD hypothesis explained nearly all major evolutionary phenomena, and here we focus on the genetic equidistance result [6,8,28]. Over long evolutionary time scales or for fast-evolving sequences, we define "maximum genetic equidistance" in terms of maximum distance: different species are equidistant to a species of lower or equal complexity, and such distances do not change with time (Figure 2(A) and (B)). The original result of Margoliash is maximum genetic equidistance. For short evolutionary time scales or for slow-evolving sequences, we define "linear genetic equidistance" in which the molecular clock holds and distance is still linearly related to time: when sister species have similar mutation rates, they would be equidistant to a less or equally complex outgroup, and such distances still increase with time (Figure 2(C) and (D)).

Both definitions emphasize that equidistance is in reference to a less complex species. If, however, a more complex species is compared, then genetic non-equidistance may be observed: for long timescale evolution or fast evolving sequences, when two or more species of different complexity are compared to a more complex species, the genetic distance between the least and most complex species would be greater than that between other species and the most complex species [8,9]. For example, when octopus and cockle are compared to human, the distance between cockle and human is greater than that between octopus and human. Octopi are the most intelligent invertebrates. This genetic non-equidistance phenomenon was discovered when we tested the MGD hypothesis and is strong evidence for the MGD and against the molecular clock hypothesis.

When two sister species are compared with an outgroup, there are often certain nucleotide positions where each of the three species is different, indicating that at least two species have independently undergone substitutions at these positions. This is termed the overlap feature of the equidistance phenomenon, or coincident substitutions (Figure 2(A) and (B)) [27]. If, after speciation, two species randomly accumulate substitutions with similar rates, as assumed by the clock/neutral theory, then the chance for a substitution in one species to occur coincidentally at the same overlap position where the other species also has a substitution should largely follow probability theory. This probability should be inversely related to the total number of changeable or tolerable positions and positively related to mutation rate.

The observed number of overlap positions in cases of maximum genetic equidistance are consistent with the expectations of the MGD hypothesis but far more than predicted by the neutral theory [27]. For linear genetic equidistance in microevolution, both theories give similar predictions that match well with observations. This congruence occurs because the short timescale of microevolution does not permit major changes in complexity and hence in the number of tolerable positions. Although the MGD hypothesis tends to give lower estimates of the number of tolerable positions, the difference is not enough to make the calculation of overlap positions much different from the predictions by the neutral theory. The overlap feature is the best evidence for distinguishing both micro- and macroevolution and maximum and linear genetic equidistance and strongly invalidates the MET's denial of these distinctions (Figure 2).

According to the MGD hypothesis, the maximum distance between sister species and an outgroup is mainly determined by the MGD of the lower-complexity outgroup, resulting in the maximum genetic equidistance phenomenon (Figure 2(A) and (B)). During evolution, more genetic diversity should be better, because it increases the adaptive capacity of the species. Thus, it would be positively selected and quickly reach MGD. The vast difference in mutation rates in different genes or parts of the human genome may exceed 10^4-fold. Most of our genome has relatively fast mutation rates, and most genetic distances observed today
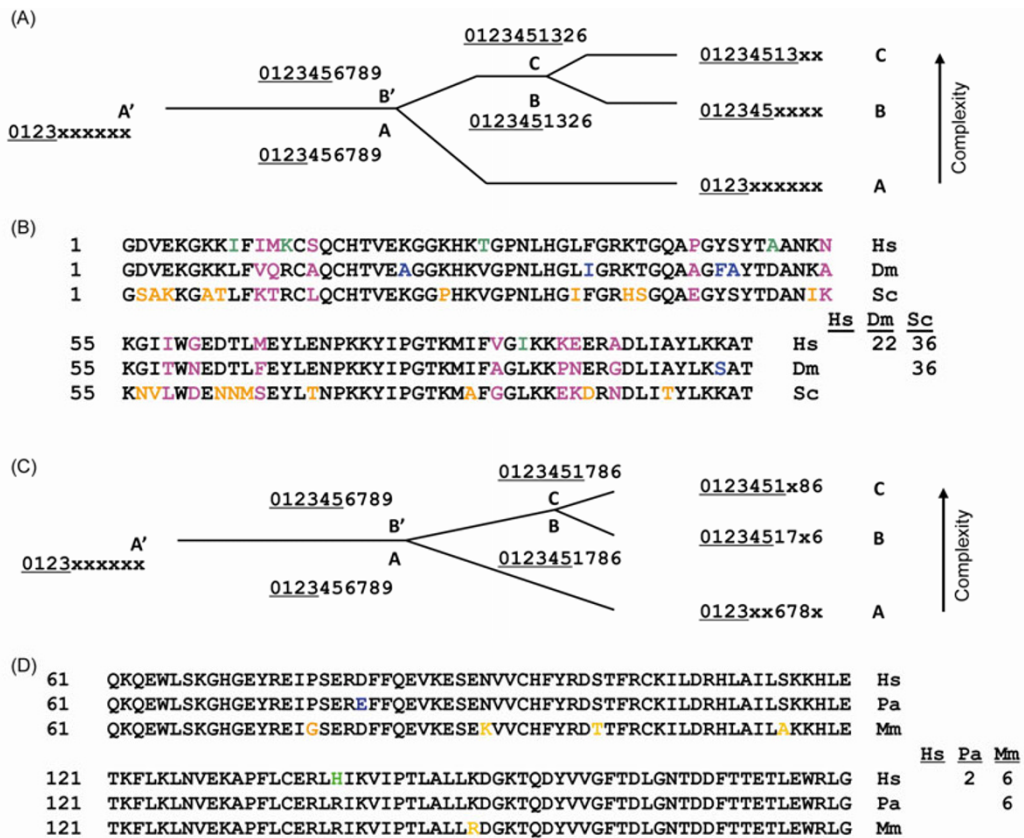
Figure 2   The genetic equidistance result and the maximum genetic diversity (MGD) hypothesis. (A), Maximum genetic equidistance. Similar to Figure 1, a 10 amino acid peptide is used to illustrate the evolution process. When the protein evolves quickly, the observed equidistance today would be maximum distance with a large overlap ratio. The figure shows four overlap positions for an overlap ratio of 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* at 22 amino acid positions. Human and *Drosophila* are equidistant to yeast, with 36 amino acid differences. There are 12 overlap positions (in pink) and the overlap ratio is 12/22=55%. Other types of mutant positions are shown in green, blue, and orange. (C), Linear genetic equidistance. When the protein evolves slowly, assuming a molecular clock holds, the observed equidistance today would be linear distance with a small overlap ratio. Here every substitution in a species would increase distance. The figure shows no overlap position, an overlap ratio 0. The distances of both C–A and B–A are 50%. (D), An example of linear equidistance. In this alignment of human, orangutan, and mouse TXND9 genes, there are two amino acid differences between human and orangutan, which are equidistant to mouse with six amino acid differences. The overlap ratio is 0/2=0.

are maxima. For most species today, the variations in their genomes are mostly at an optimum equilibrium. The only exceptions are in the slowest evolving parts of the genome [9]. The molecular clock interpretation of the maximum genetic equidistance result is really about the constant rate of complexity increases.

Also, the MGD hypothesis 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 sister species would be determined by the MGD of each species, which may not be the same for the sister species.

Since the publication of the MGD hypothesis, new evidence supporting it has constantly emerged. For example, the MGD predicts 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 has recently been shown to exist [36]. We have termed such sequences Complexity-Associated Protein Sectors or CAPS (Figure 3). In addition, the ENCODE project recently found that at least 80% of the human genome is functional [37]. The MGD is the only known scientific hypothesis that predicts a nearly 100% functional human genome.

3   Practical value of the MGD to major biomedical problems

A common saying is that to know the present or even the future is to know the past. A theory that correctly explains the past would necessarily have irreplaceable value for solving real-world problems today. For example, the correct interpretation of microevolution by the MET has helped solving the problem of bacteria drug resistance. As a more complete theory, the MGD is expected to have even more
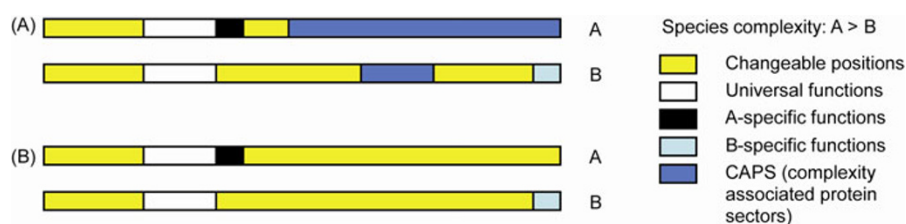
Figure 3 A schematic representation of different types of variant positions in a protein sequence alignment. (A), Schematics of a protein shared by a complex species A and a simple species B based on the MGD hypothesis. Protein sector blocks schematically represent clusters of residues that may actually be scattered in different places. (B), Schematics of the same protein according to the neutral theory.

practical value.

First, the MGD hypothesis should help resolve difficult historical problems such as the phylogenetic tree of life. Past methods have no concept of maximum distance and used mostly non-informative distance data for inferring phylogenies. Such analyses often produced self-conflicting results and conclusions inconsistent with the fossil record. We have developed the slow clock method based on the MGD [9]. The method uses only slowly-evolving sequences that have no overlap positions, and thus ensures the linear relationship between distance and time. Its results therefore will be more objective and independent of the variations in sequence selection and among investigator. Slowly-evolving sequences are more likely to meet the neutral criteria. They are unlikely to be under positive selection since their low rates of change are unlikely to meet adaptive needs. They are also unlikely to be under negative selection since their low speed means that they are often yet to reach MGD. Also, such sequences are unlikely to be under pressure to reduce their tolerable number of changeable positions as a result of complexity increases, because their slow rate of mutation means that they are less likely to be disruptive to increased complexity. Thus, their MGD levels are more likely to be similar in different species. We have used the slow clock method to re-establish that human and pongids are two separate groups, which has long been the consensus view of the paleoanthropologists [9].

Second, the MGD may rewrite population genetics. The present theoretical foundation for population genetics is the neutral theory. The value of the theory is its description of the linear phase, which has been retained by the MGD. What the neutral theory lacks and the MGD provides is a consideration of the plateau phase of the evolutionary process when genetic distance no longer changes with time. We predict that genetic diversity of a typical population is mostly at an optimum level. Most of what is considered to be neutral variation by the neutral theory will turn out to be only seemingly neutral or neutral in the sense of the traditional Chinese 'The Way of the Neutral (middle/mean)', i.e., a net result of yin and yang selection.

Finally, the MGD may help solve major biomedical problems. Most complex traits and diseases are partly heritable and presumably caused by polymorphic genetic variations, such as single nucleotide polymorphisms. The neutral theory views most variation as nonfunctional and neutral and hence the study of complex traits and diseases has in the past focused on searching for a few functional variants. Although such GWAS studies have successfully identified a number of variants, they account for only a small fraction of the total trait variation, and their functional roles typically remain unclear. The MGD predicts that complex diseases may be caused by excess genetic noise over a threshold and may serve to prevent infinite increase in genetic diversity. Complex traits evolved as a result of suppressing genetic noise and hence should be susceptible to damage by excess noise. Also, insufficient genetic diversity may hurt adaptive capacities, such as immunity. Quantitative variation in a complex trait may correlate with the amount of genetic variation. Our recent results confirm these predictions [38].

4 Reactions to the MGD hypothesis

We have since 2008 published a series of papers on the MGD hypothesis. Numerous anonymous reviewers and readers have commented on these papers. We discuss a few common reactions here. First, from the outset, we note that no one has provided a contradiction to the MGD, even though we have stated that one contradiction is sufficient for us to admit defeat. Also, no one has denied that the MET has numerous contradictions or has defended the interpretation of the maximum genetic equidistance result by the MET. Second, there are published papers that agree with our ideas and cite our papers [39]. Third, questions about the definitions of complexity are common; our views on the subject are detailed above. Fourth, some critics dispute the stability of epigenetic inheritance. However, there is already evidence of stable transgenerational epigenetic inheritance [31]. Also, epigenetic inheritance observed today in laboratory conditions is not about complexity changes and may be very different from historical events of epigenetic complexity increases. Fifth, why should independently-evolved species of similar complexity share complexity-related sequences by convergent evolution? In other words, if there are three independently evolved and phenotypically different species A, B, and C having 100, 1000, or 10000 cell types, respectively, why is distance B–C smaller than A–C? This is because A has lower complexity than B, which has

lower complexity than C, and maximum distance is only related to species complexity. Finally, some felt that maximum distance is not common sense. Here, one only needs to imagine the obvious absurdity of no MGD. Can one imagine that the phenotype of a human being can be encoded by an infinite number of non-identical sequences in a genome of the same size? Can one imagine that the function of cytochrome C can be encoded by an infinite number of non-identical sequences of ~100 amino acids in length?

5   Comparison of existing theories of evolution

Since the publication of Darwin's theory more than 150 years ago, numerous new evolutionary hypotheses have been published that purported to be better than Darwin's. A personal website gave a summary of an incomplete list of these various theories (http://home.wxs.nl/~gkorthof/), but it does not include publications written in languages other than English. We have noticed at least a few dozen new evolutionary theory books or papers written in Chinese by Chinese authors. However, no hypothesis has so far succeeded in posing even a remote threat to the MET. The reason is very simple. None of these new hypotheses can be used by professional biologists to direct their daily research, can provide a better alternative account of the questions that interest professionals, or can solve problems that the MET cannot solve. Especially for molecular evolution and population genetics, none of these hypotheses are relevant; the genetic equidistance phenomenon alone is sufficient to deem all of them inadequate. In contrast, relative to all other hypotheses, the MGD has explained more evolutionary

phenomena, is more practically useful, and has yet to meet any contradiction within its domain of relevance (Table 1).

6   Conclusion

The MGD hypothesis is a new genetics and evolutionary hypothesis that incorporates the proven virtues of existing theories and contains novel concepts to overcome their shortcomings. It has both abstract universal intuition and the capacity to account for a vast amount of data. We hope that this hypothesis can help solve real-world problems and contribute to the development of a healthy and harmonious human society.

Table 1   Comparison of maximum genetic diversity (MGD) and modern evolutionary theory (MET)

	MGD	MET
Microevo linear stage	MET	MET
Micro and macro evo	Different	Same
Role of epigenetics	Yes	No
Noise suppressed	Yes	No
Neutral regions	Neutral region sizes inversely related to organism complexity	Majority of the genome is neutral
Time frames	Any	Short
Mutation rates	Any	Slow
Genetic equidistance	Maximum/linear	Linear
Maximum distance	Yes	No
Overlap ratio	Any	Small
Contradictions	None	Numerous
Practical values	Large	Small

1   Kimura M. Evolutionary rate at the molecular level. Nature, 1968, 217: 624–626

2   King J L, Jukes T H. Non-Darwinian evolution. Science, 1969, 164: 788–798

3   Kumar S. Molecular clocks: four decades of evolution. Nat Rev Genet, 2005, 6: 654–662

4   Zuckerkandl E, Pauling L. Molecular disease, evolution, and genetic heterogeneity. In: Kasha M, Pullman B, eds. Horizons in Biochemistry. New York: Academic Press, 1962

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

6   Huang S. The genetic equidistance result of molecular evolution is independent of mutation rates. J Comp Sci Syst Biol, 2008, 1: 92–102

7   Copley R R, Schultz J, Ponting C P, et al. Protein families in multicellular organisms. Curr Opin Struct Biol, 1999, 9: 408–415

8   Huang S. Inverse relationship between genetic diversity and epigenetic complexity. Preprint available at Nature Proceedings 2009, http://dx.doi.org/10.1038/npre.2009.1751.2

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

10 Pulquerio M J, Nichols R A. Dates from the molecular clock: how wrong can we be? Trends Ecol Evol, 2007, 22: 180–184

11 Laird C D, McConaughy B L, McCarthy B J. Rate of fixation of nucleotide substitutions in evolution. Nature, 1969, 224: 149–154

12 Jukes T H, Holmquist R. Evolutionary clock: nonconstancy of rate in different species. Science, 1972, 177: 530–532

13 Goodman M, Moore G W, Barnabas J, et al. The phylogeny of human globin genes investigated by the maximum parsimony method. J Mol Evol, 1974, 3: 1–48

14 Langley C H, Fitch W M. An examination of the constancy of the rate of molecular evolution. J Mol Evol, 1974, 3: 161–177

15 Li W H. Molecular Evolution. Sunderland, MA: Sinauer Associates, 1997

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

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

18 Gago S, Elena S F, Flores R, et al. Extremely high mutation rate of a hammerhead viroid. Science, 2009, 323: 1308

19 Huang S. Molecular evidence for the hadrosaur *B. canadensis* as an outgroup to a clade containing the dinosaur *T. rex* and birds. Riv Biol,

2009, 102: 20–22

20 Huang S. Ancient fossil specimens are genetically more distant to an outgroup than extant sister species are. Riv Biol, 2008, 101: 93–108

21 Denton M. Evolution: a Theory in Crisis. Chevy Chase, MD: Adler & Adler, 1986

22 van Valen L. Molecular evolution as predicted by natural selection. J Mol Evol, 1974, 3: 89–101

23 Clarke B. Darwinian evolution of proteins. Science, 1970, 168: 1009–1011

24 Richmond R C. Non-Darwinian evolution: a critique. Nature, 1970, 225: 1025–1028

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

26 Ayala F J. Molecular clock mirages. BioEssays, 1999, 21: 71–75

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

28 Huang S. Histone methylation and the initiation of cancer. In: Tollefsbol T, ed. Cancer Epigenetics. New York: CRC Press, 2008

29 Jiang G L, Huang S. The yin-yang of PR-domain family genes in tumorigenesis. Histol Histopathol, 2000, 15: 109–117

30 Huang S. Histone methyltransferases, diet nutrients, and tumor suppressors. Nat Rev Cancer, 2002, 2: 469–476

31 Remy J J. Stable inheritance of an acquired behavior in *Caenorhabditis elegans.* Curr Biol, 2010, 20: R877–878

32 Anway M D, Cupp A S, Uzumcu M, et al. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science, 2005, 308: 1466–1469

33 Hitchins M P, Wong J J, Suthers G, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med, 2007, 356: 697–705

34 Cropley J E, Suter C M, Beckman K B, et al. Germ-line epigenetic modification of the murine Avy allele by nutritional supplementation. Proc Natl Acad Sci USA, 2006, 103: 17308–17312

35 Qian X S, Yu J Y, Dai R W. A new discipline of science: the study of open complex giant system and its methodology (in Chinese). Chinese Journal of Nature, 1990, 13: 3–10

36 Halabi N, Rivoire O, Leibler S, et al. Protein sectors: evolutionary units of three-dimensional structure. Cell, 2009, 138: 774–786

37 The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature, 2012, 489: 57–74

38 Yuan D, Zhu Z, Tan X, et al. Minor alleles of common SNPs quantitatively affect traits/diseases and are under both positive and negative selection. Presented at the 62nd Annual Meeting of the American Society of Human Genetics, San Francisco, CA, USA, 2012

39 Ho M W. Development and evolution revisited. In: Hood K E, Halpern C T, Greenberg G, et al., eds. Handbook of Developmental Science, Behavior and Genetics. New York: Blackwell Publishing, 2010