

1 **Testing the basic tenet of the molecular clock and neutral theory by using**
2 **ancient proteomes**

3

4 Tiantian Liu and Shi Huang*

5

6

7 Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School
8 of Life Sciences, Xiangya Medical School, Central South University, 110 Xiangya
9 Road, Changsha, Hunan 410078, China

10

11 *Corresponding author: huangshi@sklmg.edu.cn

12

13 **Key words:** molecular clock, neutral theory, MGD theory, evolution, ancient proteins

14

15 **Abstract**

16 Early research on orthologous protein sequence comparisons by Margoliash in
17 1963 discovered the astonishing phenomenon of genetic equidistance, which has
18 inspired the *ad hoc* interpretation known as the molecular clock. Kimura then
19 developed the neutral theory and claimed the molecular clock as its best evidence.
20 However, subsequent studies over the years have largely invalidated the universal
21 molecular clock. Yet, a watered down version of the molecular clock and the neutral
22 theory still reigns as the default model for phylogenetic inferences. The seemingly
23 obvious tenet of the molecular clock on evolutionary time scales remains to be
24 established by using ancient sequences: the longer the time of evolutionary
25 divergence, the larger the genetic distance. We here analyzed the recently published
26 Early Pleistocene enamel proteome from Dmanisi and found that ancient proteins
27 were not closer to an outgroup than their orthologs from the extant sister species were.
28 Together with a previous study, the combined results showed that most ancient
29 proteins were in fact more distant to the outgroup. The results are unexpected from
30 the molecular clock but fully predicted by the notion that genetic distances or
31 diversities are largely at optimum saturation levels as described by the maximum
32 genetic diversity (MGD) theory.
33

34 **Introduction:**

35 Since the early 1960s, protein and later DNA sequence comparisons have
36 become widely used in building evolutionary trees [1-4]. Margoliash in 1963
37 discovered an astonishing finding, genetic equidistance where sister species are
38 about equidistant to a simpler outgroup, and made an *ad hoc* interpretation of it by
39 assuming a molecular clock [2, 5]. The molecular clock hypothesis assumes a
40 constant and similar evolutionary rate among different species [1, 2, 5]. Thus, gene
41 non-identity between species is thought to be largely a function of time. The molecular
42 clock has been widely used in phylogenetic inferences and produced many
43 controversial conclusions contradictory to phylogenies built by other independent
44 methods, including the human relationship with the great apes and the origin of
45 modern humans in Africa rather than Asia.[6-8].

46 The constant and similar mutation rate (i.e., molecular clock) interpretation of the
47 equidistance result has not been verified by any independent observation and has on
48 the contrary been contradicted by a large number of facts [9-15]. Nonetheless,
49 researchers have treated the molecular clock as a genuine reality and have in turn
50 proposed a number of theories to explain it [16-21]. The 'Neutral Theory' has found
51 wide acceptance [19-21], even though it is widely acknowledged to be an incomplete
52 explanation for the clock [13, 22], and an incomplete explanatory theory of nature in
53 general [23-26]. The observed rate is measured in years but the Neutral theory
54 predicts a constant rate per generation. Also, the theory predicts that the clock will be

55 a Poisson process, with equal mean and variance of mutation rate. Experimental data
56 have shown that the variance is typically larger than the mean.

57 Ohta's "nearly neutral theory" explained to some extent the generation time issue
58 by observing that large populations have faster generation times and faster mutation
59 rates but remains unable to account for the great variance issue [27]. With the neutral
60 and nearly neutral theory, molecular evolution has been treated as the same as
61 population genetics or microevolution. However, the field still lacks a complete theory
62 as Ohta and Gillespie had acknowledged [28]. The field has unfortunately yet to pay
63 attention to the equidistance result, which has been considered by some as "one of
64 the most astonishing findings of modern science"[29, 30].

65 While it is widely acknowledged that there is no universal molecular clock (vastly
66 different species diverged for very long time do not have similar mutation rates), the
67 seemingly obvious tenet of the molecular clock notion and the neutral theory, i.e., the
68 longer the evolutionary divergence, the larger the genetic distance or sequence
69 divergence, remains widely popular in phylogenetic inferences and has yet to be
70 formally tested or established for species divergence over evolutionary time scales. A
71 most direct, simple, and strait forward test of the molecular clock would be to use
72 ancient proteins or DNAs from fossil species. Ancient fossil species are expected to
73 show less sequence divergence from an outgroup than its extant sister species
74 (Figure 1). Recent advances in protein sequencing methodology have made such test
75 feasible. We here analyzed the recently published Early Pleistocene enamel
76 proteome from Dmanisi [31] and found that ancient proteins are not closer to an

77 outgroup than their orthologs from their extant sister species are. The results are
78 unexpected from the molecular clock but are fully expected from a recently developed
79 alternative framework.

80

81 **Materials and Methods:**

82 Proteome sequences from Dmanisi were from the supplementary materials of the
83 article by Cappellini [31]. Identification of closest extant proteins and comparisons
84 with outgroups were performed using BLASTP against the protein database in
85 Genbank. Extant proteins with the highest identity to the ancient samples were
86 considered the closest extant proteins. Both the ancient and the extant orthologs were
87 then aligned with the outgroup protein to determine which was closer to the outgroup.

88

89 **Results:**

90 The recently published Early Pleistocene enamel proteome from Dmanisi had a
91 total of 10 proteins from 6 different fossil samples [31]. The ancient species
92 represented include one Equidae, one Rhinocerotidae, and four Bovidae. Some
93 species such as Bovidae were represented by more than one sample and some
94 proteins were sequenced from more than one sample. However, different peptide
95 fragments were sequenced from different samples. Thus, each of the total 22 protein
96 sequences was unique and different from others.

97 Using the ancient sequences, we searched the Genbank to identify the orthologs
98 from the closest extant sister species. We then compared both the ancient and the

99 extant proteins to an outgroup. We first tested human as the outgroup to determine
100 which shared more identity with human. The peptide fragments of each protein were
101 each individually examined for identities between the outgroup and the sister species
102 consisting of the ancient species and its closest extant species. Gaps were not counted
103 and the same length and region of alignment were maintained for the ancient and its
104 extant sister species. The results from all peptide fragments of a protein were then
105 combined to obtain the total number of identical residues and the total alignment
106 length (Table 1, and Supplementary Table 1). For a total of 22 proteins, five among
107 them were too short to be informative in terms of revealing any differences between
108 the two sisters. Of the remaining 17 proteins, fifteen showed lower number of identical
109 residues between the ancient samples and the outgroup relative to between the
110 extant proteins and the outgroup while two showed the opposite, indicating that the
111 ancient proteins were significantly more distant to the outgroup human than the extant
112 proteins were ($P < 0.01$, Chi squared test).

113 To verify the above results, we next tested a different outgroup *Sus scrofa*. Of the
114 24 proteins, four showed no difference between the sisters and hence were non
115 informative due to probably the short length covered; 15 showed lower number of
116 identical residues between the ancient samples and the outgroup relative to between
117 the extant proteins and the outgroup and 3 showed the opposite (Table 2), indicating
118 that the ancient proteins were significantly more distant to the outgroup *Sus scrofa*
119 than the extant proteins were ($P < 0.01$, Chi squared test).

120 The ancient sequences analyzed above had all isoleucines converted into
121 leucines, as standard tandem mass spectrometry (MS/MS) cannot differentiate
122 between these two isobaric amino acids. Leucines are about 2 fold more frequent
123 than isoleucines in mammalian proteins. Thus, some leucines in the above analyzed
124 ancient samples may in fact be isoleucines and would cause artificial non-identities
125 with the outgroup. We next focused only on peptides that showed non-identities
126 between the ancient samples and the sister taxon in amino acid positions not
127 involving leucine/isoleucines. Three of five informative proteins showed lower identity
128 between the ancient and the outgroup human relative to that between its extant sister
129 taxon and human while two showed the opposite. With *Sus scrofa* as the outgroup,
130 two of four informative proteins showed lower identity between the ancient and the
131 outgroup relative to that between its extant sister taxon and the outgroup while two
132 showed the opposite (Table 3). As the number of informative proteins were limited,
133 the results were inconclusive with regard to ancient proteins being more distant to the
134 outgroup but did show such a trend. At least, there was no indication that the ancient
135 proteins were closer to the outgroup, as predicted by the molecular clock. Together
136 with a previous study of ours that showed four informative ancient proteins to be all
137 more distant to the outgroup [10], the combined studies showed that 7 or 6 ancient
138 proteins were more distant to the outgroup while 2 were closer. Overall, these results
139 were not expected by the molecular clock hypothesis.

140

141 **Discussion**

142 Our results here showed that ancient proteins were not closer to an outgroup
143 than their orthologs from the closest extant species of the ancient taxon, confirming a
144 previous study of ours using ancient proteins [10]. These results remained even when
145 uncertain amino acid positions due to technical limitations such as
146 leucines/isoleucines were excluded from the analyses. The ancient proteins used
147 here were from ~2 million years ago and those from the previous study included a
148 dinosaur protein from 68 million years ago. The timeframe concerned is thus long
149 enough for ancient proteins to have lower number of substitutions than their orthologs
150 from their extant sister taxons. And yet 7 or 6 ancient proteins were more distant to
151 the outgroup while only 2 were closer. The results therefore were unexpected if the
152 basic tenet of the molecular clock and neutral theory is true that non-identities in
153 sequences is largely a function of time. That most of the ancient proteins were more
154 distant to the outgroup was even more unexpected from the molecular clock and
155 neutral theory. What then may explain the seemingly unexpected observations here?

156 In recent years, a more complete molecular evolutionary theory, the maximum
157 genetic distance or diversity (MGD) hypothesis, has been making steady progress in
158 solving both evolutionary and contemporary biomedical problems [6, 8, 25, 32-41].
159 That reality is largely at maximum genetic diversity or distance no longer changing
160 with time is *a priori* expected and supported by numerous facts [12, 25, 42, 43].
161 Genetic distance can increase with time up to a point when maximum saturation is
162 reached. At such maximum/optimum saturation, genetic distance would no longer be
163 related to time but are determined by physiological selection and environmental

164 selection. The MGD theory has solved the two major puzzles of genetic diversity, the
165 genetic equidistance phenomenon and the much narrower range of genetic diversity
166 relative to the large variation in population size [25, 26]. The primary determinant of
167 genetic diversity (or more precisely MGD) is species physiology with complex
168 physiology being compatible only with lower levels of random genetic
169 noises/diversities [25, 44]. The genetic equidistance result of Margoliash in 1963 is in
170 fact the first and best evidence for maximum distances rather than linear unsaturated
171 distances as mis-interpreted by the molecular clock and in turn the neutral theory [2, 6,
172 11, 23, 25, 43]. Two contrasting patterns of the equidistance result have now been
173 recognized, the maximum and the linear [6, 23]. The neutral theory explains only the
174 linear pattern, which however represents only a minority of any genome today. The
175 link between traits/diseases and the amount of SNPs shows an optimum/maximum
176 (Pareto optimum) genetic diversity level maintained by selection, thereby providing
177 direct experimental disproof for the neutral assumption for common SNPs [32-38].
178 More direct functional data invalidating the neutral assumption have also been found
179 [45].

180 The results here are fully predicted by the MGD theory (Figure 1). If the observed
181 genetic distance or non-identities of today and of the relatively recent past such as 2
182 million years ago as in the case of the Dmanisi samples here were at maximum
183 saturation, it would not be determined by time but by physiology and environmental
184 adaptations. As the ancient taxon and its closest present day sister species would be
185 very close in physiology, their genomes would be highly similar if not identical in those

186 parts involved in physiology. However, their genomes would be different in the parts
187 involved in adaptation to environments since environmental conditions from ~2 million
188 years ago are expected to be different from today. As the outgroup species are from
189 the present day time, they are expected to share some adaptive variants with all
190 present day species as a common adaptive strategy to today's environment. Hence,
191 the sister species of the ancient taxon would be expected to be closer to the outgroup
192 than the ancient is. The degree of this closeness would be related to the similarities
193 between the ancient environments and today's.

194 Using ancient protein or DNA sequences can be very effective in testing
195 evolutionary theories. Our results here show that the basic tenet of the molecular
196 clock and neutral theory does not hold for evolutionary time scales when maximum
197 mutation saturation has been reached. This conclusion is expected to be further
198 confirmed when more ancient sequences become available in the future.

199

200 **Acknowledgments:**

201 Supported by the National Natural Science Foundation of China grant 81171880 and
202 the National Basic Research Program of China grant 2011CB51001.

203

204 **References:**

- 205 [1] Zuckerkandl E, Pauling L. Molecular disease, evolution, and genetic
206 heterogeneity, Horizons in Biochemistry. New York: Academic Press, 1962.
207 [2] Margoliash E. Primary structure and evolution of cytochrome c. Proc. Natl.
208 Acad. Sci. 1963;50:672-9.

-
- 209 [3] Doolittle RF, Blombaeck B. Amino-Acid Sequence Investigations Of
210 Fibrinopeptides From Various Mammals: Evolutionary Implications. *Nature*
211 1964;202:147-52.
- 212 [4] Fitch WM, Margoliash E. Construction of phylogenetic trees. *Science*
213 1967;155:279-84.
- 214 [5] Kumar S. Molecular clocks: four decades of evolution. *Nat Rev Genet*
215 2005;6:654-62.
- 216 [6] Huang S. Primate phylogeny: molecular evidence for a pongid clade excluding
217 humans and a prosimian clade containing tarsiers. *Sci China Life Sci* 2012;55:709-25.
- 218 [7] Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL.
219 Radiation of human mitochondria DNA types analyzed by restriction endonuclease
220 cleavage patterns. *J Mol Evol* 1983;19:255-71.
- 221 [8] Yuan D, Lei X, Gui Y, Zhu Z, Wang D, Yu J, et al. Modern human origins:
222 multiregional evolution of autosomes and East Asia origin of Y and mtDNA. *bioRxiv*
223 2017;doi: <https://doi.org/10.1101/106864>.
- 224 [9] Huang S. Molecular evidence for the hadrosaur *B. canadensis* as an outgroup
225 to a clade containing the dinosaur *T. rex* and birds. *Riv. Biol.* 2009;102:20-2.
- 226 [10] Huang S. Ancient fossil specimens are genetically more distant to an
227 outgroup than extant sister species are. *Riv. Biol.* 2008;101:93-108.
- 228 [11] Huang S. The genetic equidistance result of molecular evolution is
229 independent of mutation rates. *J. Comp. Sci. Syst. Biol.* 2008;1:092-102.
- 230 [12] Huang S. Inverse relationship between genetic diversity and epigenetic
231 complexity. *Nature Precedings* 2008;doi: doi.org/10.1038/npre.2009.1751.2.
- 232 [13] Pulquerio MJ, Nichols RA. Dates from the molecular clock: how wrong can
233 we be? *Trends Ecol Evol* 2007;22:180-4.
- 234 [14] Nei M, Kumar S. *Molecular evolution and phylogenetics*. New York: Oxford
235 University Press, 2000.
- 236 [15] Ayala FJ. Molecular clock mirages. *BioEssays* 1999;21:71-5.
- 237 [16] Van Valen L. Molecular evolution as predicted by natural selection. *J. Mol.*
238 *Evol.* 1974;3:89-101.
- 239 [17] Clarke B. Darwinian evolution of proteins. *Science* 1970;168:1009-11.
- 240 [18] Richmond RC. Non-Darwinian evolution: a critique. *Nature* 1970;225:1025-8.
- 241 [19] Kimura M. Evolutionary rate at the molecular level. *Nature* 1968;217:624-6.
- 242 [20] Kimura M, Ohta T. On the rate of molecular evolution. *J. Mol. Evol.*
243 1971;1:1-17.
- 244 [21] King JL, Jukes TH. Non-Darwinian evolution. *Science* 1969;164:788-98.
- 245 [22] Ayala FJ. Molecular clock mirages. *BioEssays* 1999;21:71-5.
- 246 [23] Hu T, Long M, Yuan D, Zhu Z, Huang Y, Huang S. The genetic equidistance
247 result, misreading by the molecular clock and neutral theory and reinterpretation
248 nearly half of a century later. *Sci China Life Sci* 2013;56:254-61.
- 249 [24] Kern AD, Hahn MW. The Neutral Theory in Light of Natural Selection. *Mol*
250 *Biol Evol* 2018;35:1366-71.
- 251 [25] Huang S. New thoughts on an old riddle: What determines genetic diversity
252 within and between species? *Genomics* 2016;108:3-10.

-
- 253 [26] Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al.
254 Revisiting an old riddle: what determines genetic diversity levels within species? PLoS
255 Biol 2012;10:e1001388.
- 256 [27] Ohta T. Slightly deleterious mutant substitutions in evolution. Nature
257 1973;246:96-8.
- 258 [28] Ohta T, Gillespie JH. Development of Neutral and Nearly Neutral Theories.
259 Theor Popul Biol 1996;49:128-42.
- 260 [29] Denton M. Evolution: a theory in crisis. Chevy Chase, MD: Adler & Adler,
261 1986.
- 262 [30] Denton M. Evolution, still a theory in crisis. Seattle, WA: Discovery Institute
263 Press, 2016.
- 264 [31] Cappellini E, Welker F, Pandolfi L, Ramos-Madriral J, Samodova D, Ruther
265 PL, et al. Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus
266 phylogeny. Nature 2019;574:103-7.
- 267 [32] Zhu Z, Lu Q, Wang J, Huang S. Collective effects of common SNPs in
268 foraging decisions in *Caenorhabditis elegans* and an integrative method of
269 identification of candidate genes. Sci. Rep. 2015;doi:10.1038/srep16904.
- 270 [33] Zhu Z, Yuan D, Luo D, Lu X, Huang S. Enrichment of Minor Alleles of
271 Common SNPs and Improved Risk Prediction for Parkinson's Disease. PLoS ONE
272 2015;10:e0133421.
- 273 [34] Lei X, Yuan J, Zhu Z, Huang S. Collective effects of common SNPs and risk
274 prediction in lung cancer. Heredity 2018;doi:10.1038/s41437-018-0063-4.
- 275 [35] He P, Lei X, Yuan D, Zhu Z, Huang S. Accumulation of minor alleles and risk
276 prediction in schizophrenia. Sci Rep 2017;7:11661.
- 277 [36] Lei X, Huang S. Enrichment of minor allele of SNPs and genetic prediction of
278 type 2 diabetes risk in British population. PLoS ONE 2017;12:e0187644.
- 279 [37] Gui Y, Lei X, Huang S. Collective effects of common SNPs and genetic risk
280 prediction in type 1 diabetes. Clin Genet 2017;93:1069-74.
- 281 [38] Yuan D, Zhu Z, Tan X, Liang J, Zeng C, Zhang J, et al. Scoring the collective
282 effects of SNPs: association of minor alleles with complex traits in model organisms.
283 Sci China Life Sci 2014;57:876-88.
- 284 [39] Yuan D, Huang S. Genetic equidistance at nucleotide level. Genomics
285 2017;10.1016/j.ygeno.2017.03.002.
- 286 [40] Luo D, Huang S. The genetic equidistance phenomenon at the proteomic
287 level. Genomics 2016;108:25-30.
- 288 [41] Biswas K, Chakraborty S, Podder S, Ghosh TC. Insights into the dN/dS ratio
289 heterogeneity between brain specific genes and widely expressed genes in species of
290 different complexity. Genomics 2016;108:11-7.
- 291 [42] Huang S. Histone methylation and the initiation of cancer, Cancer
292 Epigenetics. New York: CRC Press, 2008.
- 293 [43] Huang S. The overlap feature of the genetic equidistance result, a
294 fundamental biological phenomenon overlooked for nearly half of a century. Biological
295 Theory 2010;5:40-52.

296 [44] Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al.
297 Comparative population genomics in animals uncovers the determinants of genetic
298 diversity. Nature 2014;515:261-3.

299 [45] Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, et al.
300 Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic
301 Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell
302 2019;24:724-35 e5.

303

304

305

306

307 **Table 1. Identities between human as the outgroup and the ancient or present**

308 **day species**

Sample	Protein	Length	Identities to <i>Homo sapiens</i>		
			Ancient	Present	Difference
16857	COL1A2	613	546	554	-8
16639	AMELX	89	70	77	-7
16635	ENAM	249	184	189	-5
16635	AMELX	173	146	150	-4
16857	COL3A1	507	385	389	-4
16632	AMELX	139	117	120	-3
16632	MMP20	130	109	112	-3
16635	AMBN	118	102	104	-2
16635	AMTN	19	14	16	-2
16638	AMELX	50	41	43	-2
16635	MMP20	57	48	49	-1
16638	AMBN	139	119	120	-1
16639	AMBN	115	103	104	-1
16641	AMBN	92	82	83	-1
16641	AMELX	59	54	55	-1
16632	ALB	27	20	20	0
16632	ENAM	182	147	147	0
16638	MMP20	44	42	42	0
16639	ENAM	123	98	98	0
16641	ENAM	67	57	57	0
16635	ALB	106	77	76	1
16638	ENAM	126	106	105	1

309

310

311 **Table 2. Identities between *Sus scrofa* as the outgroup and ancient or present**

312 **day species**

Sample	Protein	Length	Identities to <i>Sus scrofa</i>		
			Ancient	Present	Difference
16857	COL1A2	621	588	596	-8
16639	AMELX	87	74	79	-5
16632	AMELX	136	97	101	-4
16635	AMELX	172	140	144	-4
16635	AMBN	121	107	109	-2
16638	AMBN	139	124	126	-2
16638	ENAM	127	111	113	-2
16857	COL3A1	367	307	309	-2
16632	MMP20	126	113	114	-1
16635	AMTN	25	13	14	-1
16639	AMBN	118	108	109	-1
16639	ENAM	122	103	104	-1
16641	ENAM	75	66	67	-1
16641	AMBN	99	91	92	-1
16641	AMELX	58	54	55	-1
16632	ALB	26	18	18	0
16632	ENAM	183	143	143	0
16638	AMELX	49	44	44	0
16638	MMP20	44	42	42	0
16635	ALB	110	78	77	1
16635	ENAM	248	196	195	1
16635	MMP20	44	37	34	3

313

314 **Table 3. Identities between the outgroup and ancient or present day species.**

315 Amino acids differing between the ancient and its extant sister taxon are in bold.

Sample	Protein	Peptide	Length	Identities to human			Identities to <i>Sus scrofa</i>		
				Ancient	Present	Difference	Ancient	Present	Difference
16635	ALB	VFDELKPLVDEPVNLVKEN	19	13	11	2	15	13	2
16635	AMBN	ANQLNAPGRLGLMSSEEM PGRRGGPMAY	28	23	24	-1	24	25	-1
16635	ENAM	KQQSKTDPAPETQKPDQP QPEESPPKQHLKQPAATKH EEEARLPPAFPSFGNGLFP YHQP	70	35	37	-2	41	41	0
16635	MMP20	GPRKTFPGKXMPHAPPHN PS	21	18	19	-1	18	16	2
16638	ENAM	FFGYFGYHGFGRRPPYYSE EMFEDFEKPKKEE	31	30	29	1	28	29	-1

316

317

318

319 **Figure Legends:**

320 **Figure 1. Schemes for testing the molecular clock hypothesis by using ancient**

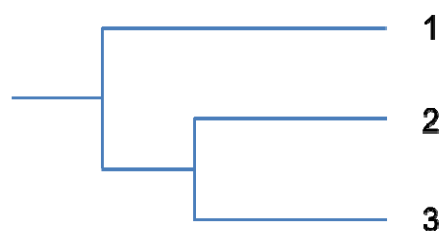
321 **proteins.** Extant species are represented by numbers 1, 2, and 3. Ancient fossil

322 species is represented by the number 4 with 2 representing the closest extant sister

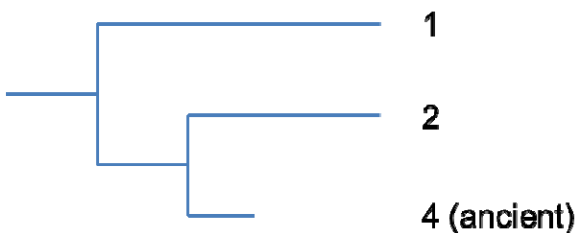
323 species of 4 and 1 representing the outgroup.

324

325



Molecular clock prediction:
1-2 identity = 1-3 identity
1-2 identity < 1-4 identity



MGD theory prediction:
1-2 identity = 1-3 identity
1-2 identity > 1-4 identity

326

327

328