1 **The Out of East Asia model versus the African Eve model of modern human**

2 **origins in light of ancient mtDNA findings**

3

4 **Ye Zhang and Shi Huang**

5 Key Laboratory of Medical Genetics, School of Life Sciences, Xiangya Medical School,

6 Central South University, 110 Xiangya Road, Changsha, Hunan 410078, P.R. China

7

10

11 **Abstract**

12   The first molecular model of modern human origins published in 1983 had the

13   mtDNA phylogenetic tree rooted in Asia. This model was subsequently overlooked

14   and superseded by the African Eve model in 1987 that was premised on the

15   unrealistic infinite site assumption and the now failed molecular clock hypothesis. We

16   have recently developed a new framework of molecular evolution, the maximum

17   genetic diversity (MGD) hypothesis, which has in turn led us to discover a new model

18   of modern human origins with the roots of uniparental DNAs placed in East Asia.

19   While the African mtDNA Eve model has haplotype N as ancestral to R, our Asia

20   model places R as the ancestor of all. We here examined ancient mtDNAs from the

21   literature focusing on the relationship between N and R. The data showed that all

22   three oldest mtDNAs were R with the 45000 year old Ust-Ishim a basal type and the

23   two ~40000 year old samples sub-branch of R. Among the numerous mtDNAs of

24   39500-30000 year old, most were R subtype U and only two were N samples, the

25   39500 year old Oase1 and the 34425 year old Salkhit. These N types are basal and

26   hence likely close to the root of N. These ancient DNA findings suggest that basal R is

27   ~5000 years older than basal N, thereby confirming the East Asia model and

28   invalidating the African Eve model.

29

30

31 **Introduction**

32   Investigation into the question of human origins has a long history and two

33 competing models termed "Multiregional" and the "Recent Out-of-Africa" hypothesis

34 have been proposed [1]. In the Multiregional model [2-4], recent human evolution is

35 seen as the product of the early and middle Pleistocene radiation of *Homo erectus*

36 from Africa. Thereafter, local differentiation led to the establishment of regional

37 populations which evolved to produce anatomically modern humans (AMH) in

38 different regions of the world. *Homo* has been a single species since the genus first

39 appeared in the fossil record ~2.3-2.8 million years ago. The model has ample

40 evidence from fossils and Paleolithic cultural remains but consistent molecular

41 evidence has been lacking until very recently [5].

42   The single origin Recent Out of Africa model assumes that there was a relatively

43 recent common ancestral population for *Homo sapiens* which already showed most if

44 not all of the anatomical features shared by present day people. This population

45 originated in Africa ~200 ky ago, followed by an initiation of African regional

46 differentiation, subsequent radiation from Africa, and final establishment of modern

47 regional characteristics outside Africa [1, 6]. These modern Africans replaced the

48 archaic *Homo* in Eurasia with limited genetic mixing [7-11]. Support for this model

49 comes from the African location of the earliest fossils of some AMH features [12, 13]

50 and the molecular clock and neutral theory interpretation of the greater genetic

51 diversity in Africans [6]. However, completely modern humans older than 80000 years

52 have yet to be found in Africa but did exist in Daoxian South China, although the

53   Daoxian site has no skeletons other than teeth, leaving it uncertain whether features

54   other than teeth are also completely modern [14, 15].

55   While long overlooked, there is in fact a third model based on mtDNA analyses. In

56   1983, researchers derived the first mtDNA phylogenetic tree and rooted the tree in

57   Asia [16]. Unfortunately, this Asia model was superseded 4 years later by the African

58   Eve model without any valid reasons [6]. The African Eve model assumes the

59   molecular clock while the Asia model not. Given that the universal molecular clock is

60   widely acknowledged to be unreal [17-22], a reexamination of these models is clearly

61   warranted. Furthermore, the Africa model requires the neutral theory and the infinite

62   site assumption, which, while largely sound as a null model and a framework for

63   pre-saturation evolutionary processes, are not *a priori* valid and have met with great

64   difficulties as an explanatory framework for most molecular evolutionary phenomena

65   [23-27]. Obviously, inferring human origins by using genetic diversity data must wait

66   until one has a complete understanding of what genetic diversity means.

67   We have in recent years examined the longstanding genetic diversity riddle and

68   developed a more complete evolutionary account, the maximum genetic diversity

69   (MGD) hypothesis, that has been productive in addressing both evolutionary and

70   biomedical problems [21, 28-36]. The MGD theory has solved the two major puzzles of

71   genetic diversity, the genetic equidistance phenomenon and the much narrower range

72   of genetic diversity relative to the large variation in population size [27, 28]. The

73   genetic equidistance result of Margoliash in 1963 is in fact the first and best evidence

74   for MGD rather than linear distance as mis-interpreted by the molecular clock and in

75   turn the neutral theory [19-21, 28, 34, 37, 38]. The MGD regards the genome to be

76   mostly functional and under stabilizing selection.

77   Based on the MGD theory, we have developed the "slow clock" method that only

78   uses slow evolving DNAs for demographic inferences because they are more likely to

79   be neutral and away from the saturation phase of evolution that is associated with fast

80   evolving sites [5, 34, 39, 40]. Our analyses have led to a new scheme of modern

81   human origins largely consistent with the multiregional model in terms of autosomes

82   where the first split among major human groups is placed at ~1.9 million years ago [5].

83   Different from the multiregional model, however, our new scheme has solved the

84   origin of the modern uniparental DNAs that obviously can only have originated from a

85   single region as the coalescence time of the modern mtDNA lineages is much shorter

86   than 1.9 million years and as archaic humans, such as Neanderthals, Denisovans,

87   and Heidelbergensis, clearly carried distinct lineages [41]. Our model placed the roots

88   of both mtDNA and Y chromosome in East Asia [5], which independently confirms the

89   1983 finding on the Asia origin of mtDNA [16]. Our model furthermore specifically

90   places the least differentiated haplotype R0 or R* as the ancestor of all mtDNA

91   haplotypes, which is in contrast to the African Eve model that puts R downstream of

92   haplotype N (Figure 1). The R0 haplotype is most common today in the Southern

93   Chinese group in the 1000 genomes project, implicating the origin of the modern

94   mtDNA lineage in Southern China [5].

95   The Africa model assumes the infinite site assumption to classify haplotype

96   relationships based on inference of ancestral and derived alleles with the allele
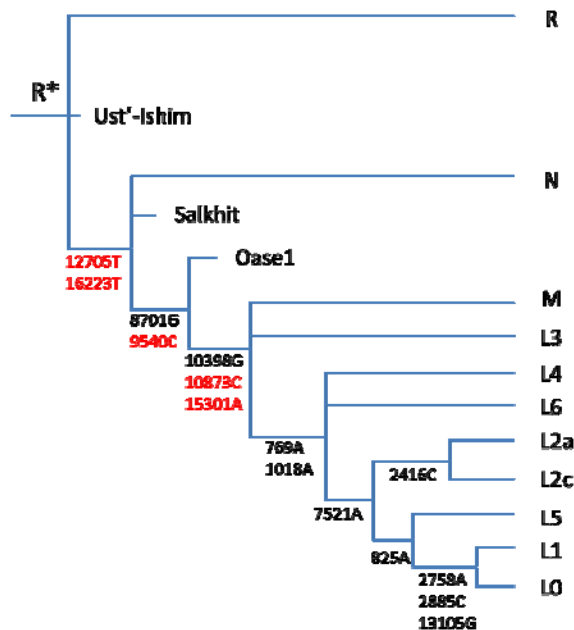
Figure 1. mtDNA phylogenetic tree of the Out of East Asia model. Only major

braches and representative mutations are shown. Slow evolving sites (altering protein

and RNA sequences) in black and fast sites in red. Also shown are several

representative ancient DNAs.

identical to that of chimpanzees as the ancestral. However, the infinite site model is in

fact unrealistic and there is overwhelming evidence of mutation saturation as

explained by the MGD theory, which makes the inference of ancestral alleles

meaningless. Actually, the Africa Eve phylogenetic tree cannot even meet the

minimum standard of a scientific model, i.e., self-consistency, as it contains huge

number of back mutations (reversions to an ancestral allele) that violate the infinite

side assumption required to build the Africa model in the first place. These mutations

can be found in the official mtDNA tree, phylotree (http://www.phylotree.org/), and are

designated by an exclamation mark (!) following the position numbers in phylotree.

113   The total number of these back mutations in phylotree is 1180 per our manual

114   counting. For example, of the 5 defining mutations for haplotype N, one at site 15301

115   is a back mutation.

116   　　Another premise (inferred from the molecular clock) for the Africa model is that

117   the original ancestor group should today show highest genetic diversity. However, as

118   the MGD theory shows, different groups may show different maximum genetic

119   diversity and human groups today are at saturation level of genetic diversity for the

120   fast evolving DNAs [5].

121   　　Unlike the Africa model, our Asia model, as well as that of Johnsen et al [16],

122   however, is based on the common sense that the original type should be the least

123   differentiated. The original ancestor should mostly carry major alleles at SNP sites

124   because as the ancestor population grew in size only a small fraction of the group

125   should carry alternative mutant alleles (mutations are rare events). Indeed, the 45000

126   year old Ust'-Ishim's mtDNA had fewer SNPs to begin with and the one he did have

127   are all common alleles in today's population (>67%) except one site 16150 [10]. The

128   Asia model also regards mutations in mtDNA to be functional, which has ample

129   evidential support [26, 33, 42, 43], and classifies haplotypes based on sharing of

130   alleles with more weight on the slow evolving sites (altering protein or RNA sequences)

131   [5]. Hence, the Asia model is inherently more sound and self-consistent than the

132   Africa model due to stronger theoretical foundations and far more realistic

133   assumptions. Nonetheless, direct experimental test of these models is required to be

134   absolutely certain. Recent discoveries in ancient DNAs have made such test possible

8

135   and we here found the Asia model strongly supported by the ancient mtDNAs findings

136   from the recent literature.

137

Results and discussion

139   The relationship between N and R is the most obvious difference between the

140   Africa and the Asia model. Of the three oldest mtDNAs, all turned out to be R, the

141   45000 year old Ust-Ishim from Siberia [10], the 40328 year old Tianyuan from

142   Northern China [44], and the 39805 year old Fumane2 from Riparo Bombrini Italy [45].

143   Tianyuan was haplotype B as reported. Fumane2 was closest to extant HV or H

144   haplotype (KP34013 and KF523402) per our analyses of the NCBI database. Only the

145   oldest R sample Ust-Ishim was found to be basal, consistent with expectations.

146   Among relatively younger samples of 39500-30000 years old, two N samples

147   were found, the 39500 year old Oase1 from Romania [9] and the 34425 year old

148   Salkhit from Mongolia [46]. However, in contrast to the rarity of N, numerous samples

149   of R subtype U were found in Europe as recently summarized by Deviese et al [46],

150   indicating that R* basal samples may only be expected near the time when the root of

151   R lived, which should be close to 45000 years ago as marked by the R* sample

152   Ust-Ishim. If R is indeed the root of modern mtDNA, then the modern lineage should

153   not be much older than 45000 years old. In fact, while partially modern skeletons can

154   be found in many sites in Africa and Asia at 310000-80000 years ago [14, 15, 47-53],

155   no human skeletons that were fully modern for all key skeletal features have been

156   found prior to 50000 years ago. As the demise of Neanderthals is thought to be

9

157   associated with the arrival of modern humans and occurred at ~40000 years ago [45],

158   one can infer that the appearance of fully modern humans may not be much earlier

159   than 40000 years ago. If it is much earlier, one would have the difficulty to explain the

160   long lag before the demise of Neanderthals.

161   Both the N samples were found to be basal N* and not directly ancestral to any

162   present day haplogroups [46]. As these two N* samples had ages of 39500-34425

163   years and no N subtypes were found in the numerous mtDNA samples from

164   40000-30000 years ago, which is in contrast to the abundantly found R subtype U,

165   one can infer that the basal N* ancestor probably lived at 39500-34425 years ago and

166   cannot be much earlier than 39500 years ago or older than R*. It is therefore highly

167   unlikely to detect the presence of N in future samples older than 45000 years or the

168   oldest R.

169   There were two M samples aged 34000-35000 years, Goyet Q116-1 and Goyet

170   Q376-3 from Belgium, found among samples older than 30000 years, and one M

171   sample of 28000 years old from La Rochette France [54, 55]. As M is downstream of R

172   and N in the Asia model but is parallel to N in the Africa model, the rarity of M relative

173   to R and N in samples older than 35000 years is consistent with the Asia model but

174   not the Africa model.

175   That R* haplotype was found to be 5000 year older than N* haplotype as

176   discussed above is consistent with the 2 SNP difference between R and N (Figure 1)

177   and the mtDNA mutation rate of 2.67×10^{-8} substitution per site per year (5000 x 16500

178   x 2.67×10^{-8} = 2.2 substitution) as recently used by others [46]. This suggests that the

179   ages of the oldest R* and N* samples found so far were likely to be close to the real

180   ages for the R and N ancestors. If the Africa model is true, the age of the original

181   African L0 ancestor of all modern humans would be 120000 years old (35 SNPs

182   separate L0 and N). However, as there are no completely modern humans found in

183   Africa that are more than 50000 years old, the L0 lineage may be unrealistic.

184   The original mtDNA type should leave more descendants today than the subtypes

185   because mutations are rare events and occur only to rare individuals in a large

186   population. We next asked whether the R haplotype is more popular today than other

187   types by examining the data collected by the mitomap database

188   (https://www.mitomap.org/MITOMAP). There are 64531 R samples in the database,

189   representing 53.6% of the whole collection. While the database may not be free of

190   sampling bias, one can nonetheless infer that it is highly likely for the R group to be

191   larger than any other haplogroup as it is found larger than all the other haplogroups

192   combined.

193   In contrast to the Asia model, the Africa model requires rare individuals with new

194   mutations to have a reproductive advantage over ancestral populations, hence

195   accounting for why the R group is much larger than any of its ancestor group, or why

196   L0 is smaller than the mutant group L1'2'3'4'5'6, why L1 is smaller than the mutant

197   group L2'3'4'5'6, why L5 is smaller than the mutant group L2'3'4'6, why L2 is smaller

198   than the mutant group L3'4'6, why L6 is smaller than the mutant group L3'4, and why

199   L4 is smaller than the mutant group L3. Such repeated reproductive advantages that

200   must be granted to each round of new mutations leading to a new haplotype are in

201   plain contradiction to the neutral mutation premise required to build the Africa model in

202   the first place.

203   In summary, the Out of East Asia mtDNA model is inherently more sound and

204   self-consistent than the African Eve model due to stronger theoretical foundations and

205   far more realistic assumptions. The original R and N are likely to have ages close to

206   those of the so far found oldest R* and N* samples, which showed that R* is older

207   than N* by 5000 years. These results plainly invalidate the African Eve model and

208   confirm the Out of Asia model. We expect our conclusion here to be further

209   strengthened by future ancient DNA studies.

210

211   **Methods:**

212   Sequence alignment and mismatches were performed using blastn at the NCBI

213   database. The number of back mutations in the mtDNA tree was inferred by counting

214   the exclamation marks (!) that designate the back mutations in the official mtDNA tree

215   phylotree (http://www.phylotree.org/). The number of R samples in the mitomap

216   database (https://www.mitomap.org/MITOMAP) were calculated by counting the

217   number of haplotypes having the R defining SNP 16223C.

218

222

References:

[1] Stringer CB, Andrews P. Genetic and fossil evidence for the origin of modern humans. Science 1988;239:1263-8.

[2] Thorne AG, Wolpoff MH. Regional continuity in Australasian Pleistocene hominid evolution. Am J Phys Anthropol 1981;55:337-49.

[3] Wolpoff MH, Wu XZ, Thorne AG. Modern homo sapiens origins: a general theory of hominid evolution involving the fossil evidence from east Asia. New York: Alan R. Liss, 1984.

[4] Wu X. On the origin of modern humans in China. Quaternary International 2004;117:131-40.

[5] Yuan D, Lei X, Gui Y, Zhu Z, Wang D, Yu J, et al. Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. bioRxiv 2017:doi: https://doi.org/10.1101/106864.

[6] Cann RL, Stoneking AC, Wilson AC. Mitochondrial DNA and human evolution. Nature 1987;325:31-6.

[7] Green RE, Krause J, et., al. A draft sequence of the Neandertal Genome. Science 2010;328:710-22.

[8] Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science 2012.

[9] Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature 2015;524:216-9.

[10] Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 2014;514:445-9.

[11] Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science 2014;343:1017-21.

[12] Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 2017;546:289-92.

[13] White TD, Asfaw B, DeGusta D, Gilbert H, Richards G, Suwa G, et al. Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 2003;423:742-7.

[14] Scerri EML, Thomas MG, Manica A, Gunz P, Stock JT, Stringer C, et al. Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? Trends Ecol Evol 2018;33:582-94.

[15] Liu W, Martinon-Torres M, Cai YJ, Xing S, Tong HW, Pei SW, et al. The earliest unequivocally modern humans in southern China. Nature 2015;526:696-9.

[16] Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 1983;19:255-71.

[17] Ayala FJ. Molecular clock mirages. BioEssays 1999;21:71-5.

[18] Ayala FJ. On the virtues and pitfalls of the molecular evolutionary clock. J Hered 1986;77:226-35.

[19] 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. Sci China Life Sci 2013;56:254-61.

266 [20] Huang S. The genetic equidistance result of molecular evolution is independent of
267 mutation rates. J. Comp. Sci. Syst. Biol. 2008;1:092-102.
268 [21] Huang S. The overlap feature of the genetic equidistance result, a fundamental biological
269 phenomenon overlooked for nearly half of a century. Biological Theory 2010;5:40-52.
270 [22] Kumar S. Molecular clocks: four decades of evolution. Nat Rev Genet 2005;6:654-62.
271 [23] Hahn MW. Toward a selection theory of molecular evolution. Evolution 2008;62:255-65.
272 [24] Kreitman M. The neutral theory is dead. Long live the neutral theory. Bioessays
273 1996;18:678-83; discussion 83.
274 [25] Ohta T, Gillespie JH. Development of Neutral and Nearly Neutral Theories. Theor Popul
275 Biol 1996;49:128-42.
276 [26] Ballard JW, Kreitman M. Is mitochondrial DNA a strictly neutral marker? Trends Ecol Evol
277 1995;10:485-8.
278 [27] Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al. Revisiting an
279 old riddle: what determines genetic diversity levels within species? PLoS Biol
280 2012;10:e1001388.
281 [28] Huang S. New thoughts on an old riddle: What determines genetic diversity within and
282 between species? Genomics 2016;108:3-10.
283 [29] Zhu Z, Lu Q, Wang J, Huang S. Collective effects of common SNPs in foraging decisions
284 in Caenorhabditis elegans and an integrative method of identification of candidate genes. Sci.
285 Rep. 2015:doi:10.1038/srep16904.
286 [30] Zhu Z, Yuan D, Luo D, Lu X, Huang S. Enrichment of Minor Alleles of Common SNPs and
287 Improved Risk Prediction for Parkinson's Disease. PLoS ONE 2015;10:e0133421.
288 [31] Zhu Z, Man X, Xia M, Huang Y, Yuan D, Huang S. Collective effects of SNPs on
289 transgenerational inheritance in Caenorhabditis elegans and budding yeast. Genomics
290 2015;106:23-9.
291 [32] Biswas K, Chakraborty S, Podder S, Ghosh TC. Insights into the dN/dS ratio
292 heterogeneity between brain specific genes and widely expressed genes in species of different
293 complexity. Genomics 2016;108:11-7.
294 [33] Zhu Z, Lu Q, Zeng F, Wang J, Huang S. Compatibility between mitochondrial and nuclear
295 genomes correlates wtih quantitative trait of lifespan in Caenorhabditis elegans. Sci. Rep.
296 2015:doi:10.1038/srep17303.
297 [34] Huang S. Primate phylogeny: molecular evidence for a pongid clade excluding humans
298 and a prosimian clade containing tarsiers. Sci China Life Sci 2012;55:709-25.
299 [35] Huang S. Inverse relationship between genetic diversity and epigenetic complexity.
300 Preprint available at Nature Precedings 2009:doi.org/10.1038/npre.2009.1751.2.
301 [36] Huang S. Histone methylation and the initiation of cancer, Cancer Epigenetics. New York:
302 CRC Press, 2008.
303 [37] Margoliash E. Primary structure and evolution of cytochrome c. Proc. Natl. Acad. Sci.
304 1963;50:672-9.
305 [38] Luo D, Huang S. The genetic equidistance phenomenon at the proteomic level. Genomics
306 2016;108:25-30.
307 [39] Yuan D, Huang S. On the peopling of the Americas: molecular evidence for the
308 Paleoamerican and the Solutrean models. bioRxiv 2017;bioRxiv 130989; doi:
309 https://doi.org/10.1101/130989

310 [40] Zhang Y, Lei X, Chen H, Zhou H, Huang S. Ancient DNAs and the Neolithic Chinese
311 super-grandfather Y haplotypes. bioRxiv 2018;doi: https://doi.org/10.1101/487918.
312 [41] Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga JL, et al. A mitochondrial
313 genome sequence of a hominin from Sima de los Huesos. Nature 2014;505:403-6.
314 [42] Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, van der Lingen CD, von der
315 Heyden S, et al. Mitochondrial DNA is unsuitable to test for isolation by distance. Sci Rep
316 2018;8:8448.
317 [43] Towarnicki SG, Ballard JWO. Mitotype Interacts With Diet to Influence Longevity, Fitness,
318 and Mitochondrial Functions in Adult Female Drosophila. Front Genet 2018;9:593.
319 [44] Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early
320 modern human from Tianyuan Cave, China. Proc Natl Acad Sci U S A 2013;110:2223-7.
321 [45] Benazzi S, Slon V, Talamo S, Negrino F, Peresani M, Bailey SE, et al. Archaeology. The
322 makers of the Protoaurignacian and implications for Neandertal extinction. Science
323 2015;348:793-6.
324 [46] Deviese T, Massilani D, Yi S, Comeskey D, Nagel S, Nickel B, et al. Compound-specific
325 radiocarbon dating and mitochondrial DNA analysis of the Pleistocene hominin from Salkhit
326 Mongolia. Nat Commun 2019;10:274.
327 [47] Hershkovitz I, Weber GW, Quam R, Duval M, Grun R, Kinsley L, et al. The earliest modern
328 humans outside Africa. Science 2018;359:456-9.
329 [48] Groucutt HS, Grun R, Zalmout IAS, Drake NA, Armitage SJ, Candy I, et al. Homo sapiens
330 in Arabia by 85,000 years ago. Nat Ecol Evol 2018;2:800-9.
331 [49] Athreya S, Wu X. A multivariate assessment of the Dali hominin cranium from China:
332 Morphological affinities and implications for Pleistocene evolution in East Asia. Am J Phys
333 Anthropol 2017;164:679-701.
334 [50] Zhao L, Zhang L, Du B, Nian XM, Zheng Y, Zhang Z, et al. New discovery of human fossils
335 and associated mammal faunas in Bijie Guizhou. Acta Anthropologica Sinica 2016;35:24-35.
336 [51] Curnoe D, Ji X, Shajin H, Tacon PSC, Li Y. Dental remains from Longtanshan cave 1
337 (Yunnan, China), and the initial presence of anatomically modern humans in East Asia.
338 Quaternary International 2016;400:180-6.
339 [52] Bae CJ, Wang W, Zhao J, Huang S, Tian F, Shen G. Modern human teeth from Late
340 Pleistocene Luna Cave (Guangxi, China). Quaternary International 2014;354:169-83.
341 [53] Li ZY, Wu XJ, Zhou LP, Liu W, Gao X, Nian XM, et al. Late Pleistocene archaic human
342 crania from Xuchang, China. Science 2017;355:969-72.
343 [54] Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of
344 Ice Age Europe. Nature 2016;534:200-5.
345 [55] Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, et al. Pleistocene
346 Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial
347 Population Turnover in Europe. Curr Biol 2016;26:827-33.
348
349