28. J. Savarino, M. Legrand, *J. Geophys. Res.* **103**, 8267 (1998).
29. P. D. Jones, M. E. Mann, *Rev. Geophys.* **42**, 2003RG000143 (2004).
30. A. Moberg, D. M. Sonechkin, K. Holmgren, N. M. Datsenko, W. Karlen, *Nature* **433**, 613 (2005).
31. M. O. Andreae, P. Merlet, *Global Biogeochem. Cycles* **15**, 955 (2001).
32. C. McEvedy, R. Jones, *Atlas of World Population History* (Penguin, London, 1978), pp. 368.
33. W. Denevan, *The Native Population of the Americas in 1492* (Univ. of Wisconsin Press, Madison, WI, 1992), pp. 386.
34. B. Glaser, L. Haumaier, G. Guggenberger, W. Zech, *Naturwissenchaften* **88**, 37 (2001).
35. C. MacFarling Meure, thesis, University of Melbourne (2004).
36. We thank the staff of the Australian Antarctic Program, especially Casey Station, for field support; A. Smith for firn-air sampling assistance; the Bureau of Meteorology (Australia) for Cape Grim archive-air collection assistance; R. Francey, P. Steele, C. Allison, and S. Coram at CSIRO for logistical and technical help; and especially B. Ruddiman and B. Allan for valuable discussions. Supported by NSF (grant no. OPP0087357); NOAA/Climate Modeling and Diagnostics Laboratory; NIWA, New Zealand (Foundation for Research Science and Technology grant no. C01X0204); and the Australian Government's Antarctic Climate and Ecosystems Cooperative Research Centre and CSIRO Atmospheric Research.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5741/1714/DC1
Materials and Methods
References and Notes

23 May 2005; accepted 28 July 2005
10.1126/science.1115193

Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans

Patrick D. Evans,[1,2] Sandra L. Gilbert,[1] Nitzan Mekel-Bobrov,[1,2] Eric J. Vallender,[1,2] Jeffrey R. Anderson,[1] Leila M. Vaez-Azizi,[1] Sarah A. Tishkoff,[4] Richard R. Hudson,[3] Bruce T. Lahn[1]*

The gene *Microcephalin* (*MCPH1*) regulates brain size and has evolved under strong positive selection in the human evolutionary lineage. We show that one genetic variant of *Microcephalin* in modern humans, which arose ~37,000 years ago, increased in frequency too rapidly to be compatible with neutral drift. This indicates that it has spread under strong positive selection, although the exact nature of the selection is unknown. The finding that an important brain gene has continued to evolve adaptively in anatomically modern humans suggests the ongoing evolutionary plasticity of the human brain. It also makes *Microcephalin* an attractive candidate locus for studying the genetics of human variation in brain-related phenotypes.

The most distinct trait of *Homo sapiens* is the exceptional size and complexity of the brain (*1*, *2*). Several recent studies have linked specific genes to the evolution of the human brain (*3*–*12*). One of these is *Microcephalin* (*7*, *8*); mutations in this gene cause primary microcephaly [MCPH; Online Mendelian Inheritance in Man (OMIM) accession 251200] (*13*, *14*). MCPH is defined clinically as severe reductions in brain size coupled with mental retardation, but remarkably, an overall retention of normal brain structure and a lack of overt abnormalities outside of the nervous system (*15*–*17*). This led to the notion that the brains of MCPH patients function normally for their size and that genes underlying MCPH are specific developmental regulators of brain size (*15*–*17*).

Microcephalin is one of six known loci, named *MCPH1* through *MCPH6*, for which recessive mutations lead to MCPH (*14*, *18*–*23*). For four of these, the underlying genes have been identified as *Microcephalin* (*MCPH1*), *CDK5RAP2* (*MCPH3*), *ASPM* (*MCPH5*), and *CENPJ* (*MCPH6*) (*14*, *21*, *23*). Patients with loss-of-function mutations in *Microcephalin* have cranial capacities about 4 SD below the mean at birth. As adults, their typical brain size is around 400 cm^3 (whereas the normal range is 1200 to 1600 cm^3), and the cerebral cortex is especially small (*13*, *14*). *Microcephalin* is suggested to control the proliferation and/or differentiation of neuroblasts during neurogenesis. This postulate was consistent with several observations. First, mouse *Microcephalin* is expressed prominently in the proliferative zones of the embryonic brain (*14*). Second, the Microcephalin protein contains several copies of the BRCT domain that is found in cell cycle regulators, such as *BRCA1* (*14*, *24*). Finally, cell culture studies indeed suggested a role of *Microcephalin* in regulating cell cycle (*25*–*27*).

The finding that *Microcephalin* is a critical regulator of brain size spurred the hypothesis that it might have played a role in brain evolution (*16*, *28*). Consistent with this hypothesis, phylogenetic analysis of *Microcephalin* revealed signatures of strong positive selection in the lineage leading to humans (*7*, *8*). Here, we examine the possibility that positive selection has continued to operate on this gene after the emergence of anatomically modern humans.

The human *Microcephalin* locus has 14 exons spanning about 236 kb on chromosome 8p23 (*14*) (Fig. 1). We previously sequenced all the exons in 27 humans (*8*). When reanalyzing the data, we noticed that one haplotype had a much higher frequency than the other haplotypes. Additionally, this haplotype differed consistently from the others at position 37995 of the genomic sequence (counting from the start codon) or position 940 of the open reading frame. This polymorphism falls in exon 8 and changes amino acid residue 314 from an ancestral aspartate to a histidine. (This polymorphism is described as G37995C with G denoting the ancestral allele.)

To investigate whether positive selection has acted on the high-frequency haplotype, we resequenced 23.4 kb of a 29-kb region centered around the G37995C polymorphism (Fig. 1). Sequencing was performed on a panel of 89 individuals from the Coriell Institute, which broadly represents human diversity (see SOM). To assign the ancestral state of polymorphisms, we also sequenced the common chimpanzee. Several GC-rich segments were not sequenced because of technical difficulties. The resulting sequence data contained 220 polymorphic sites, including 213 single-nucleotide polymorphisms (SNPs) and 7 insertion/deletion polymorphisms (indels) (table S1).

Haplotypes were inferred using the PHASE 2.1 program (*29*, *30*). A total of 86 haplotypes

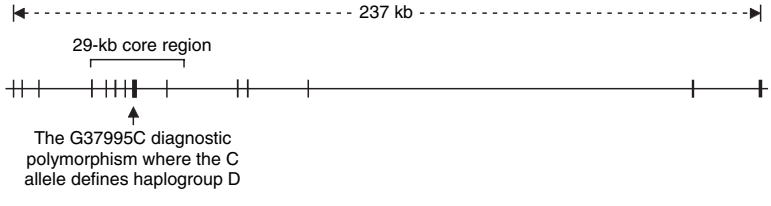


Fig. 1. Genomic structure of the human *Microcephalin* gene. The region sequenced in the 89-individual Coriell panel is bracketed.

[1]Howard Hughes Medical Institute, Department of Human Genetics, [2]Committee on Genetics, [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. [4]Department of Biology, University of Maryland, College Park, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu

were identified along with their frequencies (Fig. 2 and table S2). One haplotype, denoted 49, had a much higher frequency than the other haplotypes. It had the derived C allele at the G37995C SNP site and corresponded to the high-frequency haplotype in the aforementioned exon-only polymorphism survey (*8*). In the Coriell panel, haplotype 49 had a frequency of 33% (59 out of 178 chromosomes) and is found in all the populations sampled in the panel. The remaining 85 haplotypes varied in frequency from 0.6 to 6.2% (1 to 11 chromosomes).

Positive selection on an allele can increase the frequency of the haplotype bearing the allele while maintaining extended linkage disequilibrium (LD) around that allele (*31*–*36*). Our data on haplotype 49 are consistent with these signatures of selection. We formally tested the statistical significance of positive selection using the previously established coalescent model (*37*, *38*). Given the slight uncertainty in haplotype inference, we considered only the 18 individuals in the Coriell panel who are homozygous for haplotype 49 (table S1).

By simulation, we calculated the probability of obtaining 18 or more individuals (out of 89) who are homozygous for a single haplotype across a region of 220 segregating sites under neutral evolution. Here, recombination and gene conversion rates were set to values previously established for the *Microcephalin* locus (*39*), and a demographic model with a severe bottleneck followed by exponential growth was assumed (see SOM). Prior studies have shown that the bottleneck specified here is likely to be much more stringent than that associated with the real demographic history of human populations (*40*, *41*); thus, the test is conservative (*38*). Under these parameters, the probability of obtaining 18 homozygotes out of 89 is highly significant ($P = 0$ based on 5,000,000 replicates).

We then tested several additional demographic models, including (i) constant size, (ii) very ancient expansion, (iii) very recent expansion, (iv) repeated severe bottlenecks with subsequent expansion, and (v) population structure with between two and five subpopulations (see SOM). All produced exceedingly significant results. Even though the exact demographic history of humans is yet to be defined, our tests are highly significant under a broad range of demographic scenarios, which furthers the argument that the statistical significance is unlikely to be altered by reasonable variations in the supposed human demography. We also tested the significance of the inferred haplotype data (i.e., the significance of having 59 copies of haplotype 49 among 178 chromosomes), which similarly produced highly significant results. These data strongly suggest that haplotype 49 was driven to high frequency by positive selection. However, our data do not address whether the positive selection is frequency-dependent selection, heterozygote advantage, or simple additive positive selection.

Using the G37995C polymorphism as a diagnostic site, we divided all the haplotypes into two groups: those that carry the derived C allele and those that carry the ancestral G allele. We designated the former group as haplogroup D (where D stands for "derived"). It includes 43 haplotypes that together have a 70% frequency in the Coriell panel, and haplotype 49 is the predominant member (table S2). Although the derived C allele at the G37995C site only provides an operational definition for haplogroup D, several observations make evident that haplogroup D is systematically different from the non-D haplotypes. First, this haplogroup consists exclusively of haplotype 49 or its minor variants, whereas non-D haplotypes show much greater sequence divergence from haplogroup D chromosomes. This greater divergence is because haplogroup D and non-D haplotypes have multiple fixed differences relative to each other in addition to G37995C (table S2). The only exceptions are a few recombinant haplotypes between D and non-D chromosomes (discussed below). Second, for sites that are polymorphic within haplogroup D chromosomes (excluding recombinants between D and non-D chromosomes), the non-D chromosomes are invariably monomorphic for the ancestral alleles. These data indicate that haplogroup D constitutes a genealogical clade of closely related haplotypes that is altogether separate from the more distantly related non-D haplotypes (again, excluding recombinants between D and non-D chromosomes, which represent mixed genealogies).

Collectively, the above observations support an evolutionary scenario with two aspects. First, haplotype 49 swept from a single copy to high frequency in a short period of time. Second, during the sweep, minor variants of haplotype 49 emerged through rare mutations and recombinations. These variants, together with haplotype 49, make up haplogroup D. Haplotype 49 evidently represents the most recent common ancestor (MRCA) of haplogroup D, because it consistently has the ancestral allele for the sites polymorphic within haplogroup D.

We next estimated the coalescence age (i.e., time to MRCA) of haplogroup D chromosomes in the Coriell panel. We used the average number of mutations from the MRCA of a haplogroup clade to its descendant lineages as a molecular clock for estimating the age of the clade (*42*, *43*). This approach is known to be unbiased by demographic history (*42*). The age of haplogroup D was found to be ~37,000 years, with a 95% confidence interval of 14,000 to 60,000 years. In comparison, the coalescence age of all the chromosomes in the Coriell panel is about 1,700,000 years. The emergence of anatomically modern humans has been estimated to be 200,000 years before present (*44*). Haplogroup D is obviously much younger, which indicates that positive selection was at work in a period considerably postdating the emergence of anatomically modern humans in Africa. We note that the age of haplogroup D coincides with the introduction of anatomically modern humans into Europe about 40,000 years ago, as well as the dramatic shift in the archeological record indicative of modern human behavior, such as art and the use of symbolism (i.e., the "Upper Paleolithic revolution") (*45*).

If haplogroup D indeed experienced a recent selective sweep, it should show low poly-

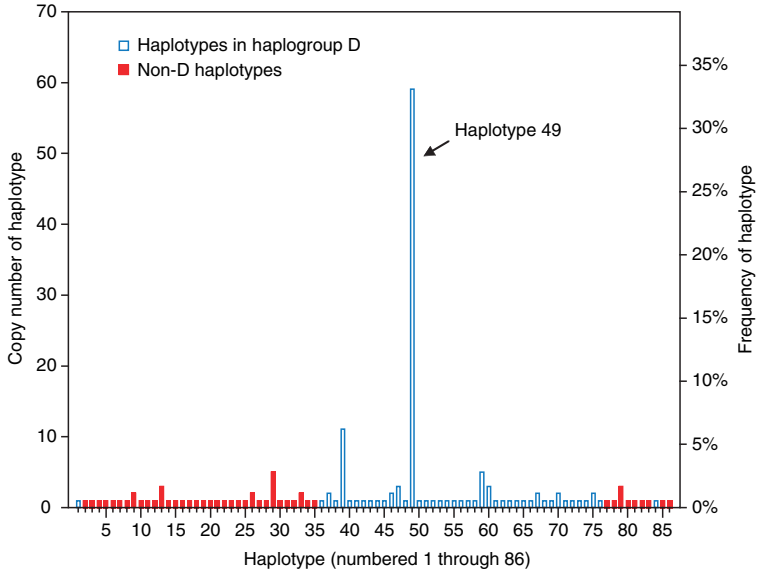


Fig. 2. Frequencies of 86 inferred *Microcephalin* haplotypes in the 89-individual Coriell panel. Haplotypes in haplogroup D are indicated by blue-edged bars; non-D haplotypes are indicated by solid red bars.

morphism and an excess of rare alleles (*46*). To confirm this, we calculated nucleotide diversity (π) and Tajima's *D* for the 47 individuals who are homozygous for haplogroup D chromosomes, and we compared these values to those of the non-D chromosomes. The π value of the D chromosomes is lower, by a factor of 12, than that of the non-D chromosomes (0.000077 and 0.00092, respectively), even though the D chromosomes represent about 70% of the chromosomes in the panel. Tajima's *D*, which is a summary statistic for the frequency spectrum of alleles, is –2.3 for haplogroup D (whereas it is –1.2 for the non-D chromosomes). This strongly negative Tajima's *D* indicates a starlike genealogy for haplogroup D chromosomes (*47*). Thus, both summary statistics contrast sharply between D and non-D chromosomes and are consistent with the recent age and rapid expansion of haplogroup D. We note that these calculations do not provide a statistically stringent test of positive selection, because they are done on subsets of the genealogy. Nevertheless, they do reveal qualitative signatures of positive selection that further corroborate the more stringent statistical tests described earlier.

Another sign of a positive selective sweep is extended LD around the selected allele. This is apparent in the region of *Microcephalin* investigated here, where haplogroup D chromosomes show near-complete LD across the entire region. The only exceptions are haplotypes 1, 68, and 84 (each found in a single copy in the Coriell panel), which are recombinants between D and non-D chromosomes as evidenced by recombination tracts (table S2). The remaining 121 copies of haplogroup D chromosomes show no evidence of recombination. By comparison, the non-D chromosomes do not display any significant LD across the region.

To probe the extent of LD beyond the 29-kb core region, we sequenced the Coriell panel for two segments of about 3 kb each, situated at the beginning and end of the gene separated from each other by about 235 kb. In these flanking regions, there is clear evidence of LD decay from the core region, which supports the idea that selection has most likely operated on a site (or sites) around the core region. Our present data cannot resolve the exact site(s) of selection, and the G37995C nonsynonymous SNP used to define haplogroup D is just a candidate.

To obtain a more detailed frequency distribution of haplogroup D across the globe, we analyzed a much larger human population panel containing 1184 globally diverse individuals. We genotyped the diagnostic G37995C SNP in this panel to infer the frequency of haplogroup D chromosomes (Fig. 3). Geographic variation was observed, with sub-Saharan populations generally having lower frequencies than others. The statistic for genetic differentiation, F_{ST}, is 0.48 between sub-Saharans and others, which indicates strong differentiation (*48*) and is significantly higher than the genome average of 0.12 ($P < 0.03$ based on previously established genomewide F_{ST} distribution) (*49*). Such population differentiation may reflect a Eurasian origin of haplogroup D, local adaptation, and/or

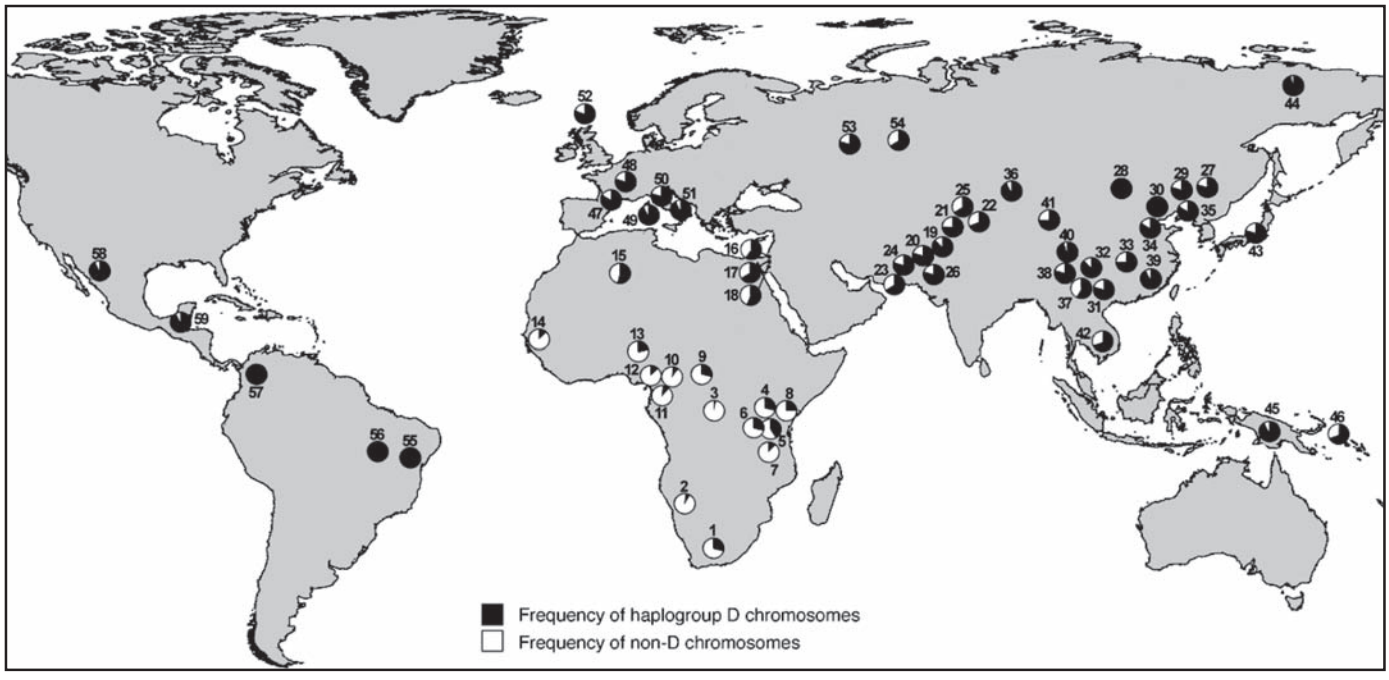


Fig. 3. Global frequencies of *Microcephalin* haplogroup D chromosomes (defined as having the derived C allele at the G37995C diagnostic SNP) in a panel of 1184 individuals. For each population, the country of origin, number of individuals sampled, and frequency of haplogroup D chromosomes are given (in parentheses) as follows: 1, Southeastern and Southwestern Bantu (South Africa, 8, 31.3%); 2, San (Namibia, 7, 7.1%); 3, Mbuti Pygmy (Democratic Republic of Congo, 15, 3.3%); 4, Masai (Tanzania, 27, 29.6%); 5, Sandawe (Tanzania, 32, 39.1%); 6, Burunge (Tanzania, 28, 30.4%); 7, Turu (Tanzania, 23, 15.2%); 8, Northeastern Bantu (Kenya, 12, 25%); 9, Biaka Pygmy (Central African Republic, 32, 26.6%); 10, Zime (Cameroon, 23, 8.7%); 11, Bakola Pygmy (Cameroon, 24, 10.4%); 12, Bamoun (Cameroon, 28, 17.9%); 13, Yoruba (Nigeria, 25, 24%); 14, Mandenka (Senegal, 24, 16.7%); 15, Mozabite [Algeria (Mzab region), 29, 53.5%]; 16, Druze [Israel (Carmel region), 44, 60.2%]; 17, Palestinian [Israel (Central), 40, 63.8%]; 18, Bedouin [Israel (Negev region), 44, 54.6%]; 19, Hazara (Pakistan, 20, 85%); 20, Balochi (Pakistan, 23, 78.3%); 21, Pathan (Pakistan, 23, 76.1%); 22, Burusho (Pakistan, 25, 66%); 23, Makrani (Pakistan, 24, 62.5%); 24, Brahui (Pakistan, 25, 78%); 25, Kalash (Pakistan, 24, 62.5%); 26, Sindhi (Pakistan, 25, 78%); 27, Hezhen (China, 9, 77.8%); 28, Mongola (China, 10, 100%); 29, Daur (China, 10, 85%); 30, Orogen (China, 10, 100%); 31, Miaozu (China, 9, 77.8%); 32, Yizu (China, 10, 85%); 33, Tujia (China, 10, 75%); 34, Han (China, 41, 82.9%); 35, Xibo (China, 9, 83.3%); 36, Uygur (China, 10, 90%); 37, Dai (China, 9, 55.6%); 38, Lahu (China, 10, 85%); 39, She (China, 9, 88.9%); 40, Naxi (China, 10, 95%); 41, Tu (China, 10, 75%); 42, Cambodian (Cambodia, 11, 72.7%); 43, Japanese (Japan, 27, 77.8%); 44, Yakut [Russia (Siberia region), 25, 98%]; 45, Papuan (New Guinea, 17, 91.2%); 46, NAN Melanesian (Bougainville, 18, 72.2%); 47, French Basque (France, 24, 83.3%); 48, French (France, 28, 78.6%); 49, Sardinian (Italy, 26, 90.4%); 50, North Italian [Italy (Bergamo region), 13, 76.9%]; 51, Tuscan (Italy, 8, 87.5%); 52, Orcadian (Orkney Islands, 16, 81.3%); 53, Russian (Russia, 24, 79.2%); 54, Adygei [Russia (Caucasus region), 15, 63.3%]; 55, Karitiana (Brazil, 21, 100%); 56, Surui (Brazil, 20, 100%); 57, Colombian (Colombia, 11, 100%); 58, Pima (Mexico, 25, 92%); 59, Maya (Mexico, 25, 92%).

demographic factors such a bottleneck associated with human migration out of Africa 50,000 to 100,000 years ago.

Previous studies have shown that *Microcephalin* is a specific regulator of brain size (*13*, *14*) and that this gene has evolved under strong positive selection in the primate lineage leading to *Homo sapiens* (*7*, *8*). Here, we present compelling evidence that *Microcephalin* has continued its trend of adaptive evolution beyond the emergence of anatomically modern humans. The specific function of *Microcephalin* in brain development makes it likely that selection has operated on the brain. Yet, it remains formally possible that an unrecognized function of *Microcephalin* outside of the brain is actually the substrate of selection. If selection indeed acted on a brain-related phenotype, there could be several possibilities, including brain size, cognition, personality, motor control, or susceptibility to neurological and/or psychiatric diseases. We hypothesize that D and non-D haplotypes have different effects on the proliferation of neural progenitor cells, which in turn leads to different phenotypic outcomes of the brain visible to selection.

References and Notes

1. J. N. Spuhler, *The Evolution of Man's Capacity for Culture* (Wayne State Univ. Press, Detroit, MI, 1959).
2. J. H. Jerison, *Evolution of the Brain and Intelligence* (Academic Press, New York, 1973).
3. W. Enard *et al.*, *Nature* **418**, 869 (2002).
4. J. Zhang, *Genetics* **165**, 2063 (2003).
5. P. D. Evans *et al.*, *Hum. Mol. Genet.* **13**, 489 (2004).
6. N. Kouprina *et al.*, *PLoS Biol.* **2**, E126 (2004).
7. Y. Q. Wang, B. Su, *Hum. Mol. Genet.* **13**, 1131 (2004).
8. P. D. Evans, J. R. Anderson, E. J. Vallender, S. S. Choi, B. T. Lahn, *Hum. Mol. Genet.* **13**, 1139 (2004).
9. R. J. Ferland *et al.*, *Nat. Genet.* **36**, 1008 (2004).
10. H. H. Stedman *et al.*, *Nature* **428**, 415 (2004).
11. F. Burki, H. Kaessmann, *Nat. Genet.* **36**, 1061 (2004).
12. S. Dorus *et al.*, *Cell* **119**, 1027 (2004).
13. A. P. Jackson *et al.*, *Am. J. Hum. Genet.* **63**, 541 (1998).
14. A. P. Jackson *et al.*, *Am. J. Hum. Genet.* **71**, 136 (2002).
15. W. B. Dobyns, *Am. J. Hum. Genet.* **112**, 315 (2002).
16. G. H. Mochida, C. A. Walsh, *Curr. Opin. Neurol.* **14**, 151 (2001).
17. C. G. Woods, J. Bond, W. Enard, *Am. J. Hum. Genet.* **76**, 717 (2005).
18. E. Roberts *et al.*, *Eur. J. Hum. Genet.* **7**, 815 (1999).
19. L. Moynihan *et al.*, *Am. J. Hum. Genet.* **66**, 724 (2000).
20. C. R. Jamieson, C. Govaerts, M. J. Abramowicz, *Am. J. Hum. Genet.* **65**, 1465 (1999).
21. J. Bond *et al.*, *Nat. Genet.* **32**, 316 (2002).
22. G. F. Leal *et al.*, *J. Med. Genet.* **40**, 540 (2003).
23. J. Bond *et al.*, *Nat. Genet.* **37**, 353 (2005).
24. T. Huyton, P. A. Bates, X. Zhang, M. J. Sternberg, P. S. Freemont, *Mutat. Res.* **460**, 319 (2000).
25. S. Y. Lin, S. J. Elledge, *Cell* **113**, 881 (2003).
26. X. Xu, J. Lee, D. F. Stern, *J. Biol. Chem.* **279**, 34091 (2004).
27. M. Trimborn *et al.*, *Am. J. Hum. Genet.* **75**, 261 (2004).
28. S. L. Gilbert, W. B. Dobyns, B. T. Lahn, *Nat. Rev. Genet.* **6**, 581 (2005).
29. M. Stephens, N. J. Smith, P. Donnelly, *Am. J. Hum. Genet.* **68**, 978 (2001).
30. M. Stephens, P. Donnelly, *Am. J. Hum. Genet.* **73**, 1162 (2003).
31. S. A. Tishkoff *et al.*, *Science* **293**, 455 (2001).
32. P. C. Sabeti *et al.*, *Nature* **419**, 832 (2002).
33. E. Wang *et al.*, *Am. J. Hum. Genet.* **74**, 931 (2004).
34. T. Bersaglieri *et al.*, *Am. J. Hum. Genet.* **74**, 1111 (2004).
35. E. E. Thompson *et al.*, *Am. J. Hum. Genet.* **75**, 1059 (2004).
36. H. Stefansson *et al.*, *Nat. Genet.* **37**, 129 (2005).
37. R. R. Hudson, *Oxf. Surv. Evol. Biol.* **7**, 1 (1990).
38. R. R. Hudson, *Bioinformatics* **18**, 337 (2002).
39. A. Kong *et al.*, *Nat. Genet.* **31**, 241 (2002).
40. E. Zietkiewicz *et al.*, *J. Mol. Evol.* **47**, 146 (1998).
41. H. Harpending, A. Rogers, *Annu. Rev. Genomics Hum. Genet.* **1**, 361 (2000).
42. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, *Genetics* **161**, 447 (2002).
43. R. Thomson, J. K. Pritchard, P. Shen, P. J. Oefner, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360 (2000).
44. I. McDougall, F. H. Brown, J. G. Fleagle, *Nature* **433**, 733 (2005).
45. R. G. Klein, *The Human Career: Human Biological and Cultural Origins* (Univ. of Chicago Press, Chicago, 1999).
46. M. Bamshad, S. P. Wooding, *Nat. Rev. Genet.* **4**, 99 (2003).
47. F. Tajima, *Genetics* **123**, 585 (1989).
48. S. Wright, *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1978).
49. J. M. Akey, G. Zhang, K. Zhang, L. Jin, M. D. Shriver, *Genome Res.* **12**, 1805 (2002).

50. We thank the Coriell Institute for Medical Research, the Centre d'Etude du Polymorphisme Humain (CEPH), and A. Froment for human DNA samples. We thank H. M. Cann, S. Dorus, E. E. Eichler, N. M. Pearson, A. Di Rienzo, M. Kreitman, and J. K. Pritchard for technical support and/or helpful discussions. Supported in part by the Searle Scholarship and the Burroughs Wellcome Career Award (to B.T.L.), and David and Lucile Packard Career Award, the Burroughs Wellcome Career Award, and NSF grant BCS-0196183 (to S.A.T).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5741/1717/DC1
Materials and Methods
Tables S1 and S2
References and Notes

18 April 2005; accepted 14 June 2005
10.1126/science.1113722

Ongoing Adaptive Evolution of *ASPM*, a Brain Size Determinant in *Homo sapiens*

Nitzan Mekel-Bobrov,[1,2] Sandra L. Gilbert,[1] Patrick D. Evans,[1,2] Eric J. Vallender,[1,2] Jeffrey R. Anderson,[1] Richard R. Hudson,[3] Sarah A. Tishkoff,[4] Bruce T. Lahn[1]*

The gene *ASPM* (*abnormal spindle-like microcephaly associated*) is a specific regulator of brain size, and its evolution in the lineage leading to *Homo sapiens* was driven by strong positive selection. Here, we show that one genetic variant of *ASPM* in humans arose merely about 5800 years ago and has since swept to high frequency under strong positive selection. These findings, especially the remarkably young age of the positively selected variant, suggest that the human brain is still undergoing rapid adaptive evolution.

Homozygous null mutations of *ASPM* cause primary microcephaly, a condition characterized by severely reduced brain size with otherwise normal neuroarchitecture (*1*). Studies have suggested that *ASPM* may regulate neural stem cell proliferation and/or differentiation during brain development, possibly by mediating spindle assembly during cell division (*1*, *2*). Phylogenetic analysis of *ASPM* has revealed strong positive selection in the primate lineage leading to *Homo sapiens* (*3*–*5*), especially in the past 6 million years of hominid evolution in which ASPM acquired about one advantageous amino acid change every 350,000 years (*4*). These data argue that *ASPM* may have contributed to human brain evolution (*3*–*6*). Here, we investigate whether positive selection has continued to operate on *ASPM* since the emergence of anatomically modern humans.

Human *ASPM* has 28 exons with a 10,434–base pair open reading frame (*1*) (fig. S1). We resequenced the entire 62.1-kb genomic region of *ASPM* in samples from 90 ethnically diverse individuals obtained through the Coriell Institute and from a common chimpanzee (*7*). This revealed 166 polymorphic sites (table S1). Using established methodology (*7*), we identified 106 haplotypes. One haplotype, numbered 63, had an unusually high frequency of 21%, whereas the other haplotypes ranged from 0.56% to 3.3% (fig. S2). Moreover, this haplotype differed consistently from the others at multiple polymorphic sites (save for a few rare haplotypes that are minor mutational or recombinational variants of haplotype 63, as discussed later) (table S2). Two of these polymorphic sites are nonsynonymous, both in exon 18, and are denoted A44871G and C45126A (numbers indicate genomic positions from the start codon, and letters at the beginning and end indicate ancestral and derived alleles, respectively). These two sites reside in a region of the open reading frame that was shown previously to have experienced par-

[1]Howard Hughes Medical Institute, Department of Human Genetics, [2]Committee on Genetics, [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. [4]Department of Biology, University of Maryland, College Park, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu

Science Supporting Online Material

Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans

Patrick D. Evans, Sandra L. Gilbert, Nitzan Mekel-Bobrov, Eric J. Vallender, Jeffrey R. Anderson, Leila M. Vaez-Azizi, Sarah A. Tishkoff, Richard R. Hudson, Bruce T. Lahn

DOI: 10.1126/science.1113722

Materials and Methods

Sequence acquisition and preliminary analysis

A panel of 89 human samples from the Coriell Institute that broadly represent worldwide populations was used for resequencing. It includes 9 sub-Saharan Africans (Coriell numbers: 17341–17349), 7 North Africans (17378–17384), 9 Iberians (17091–17097, 17099, 17100), 7 Basques (15883–15887, 16185, 16188), 9 Russians (13820a, 13838, 13852, 13876, 13877, 13911–13914), 9 Middle Easterners (17331–17340), 9 South Asians (17021–17024, 17026–17030), 8 Chinese (16654, 16688, 16689, 17014, 17015, 17017–17019), 1 Japanese (11587), 8 Southeast Asians (17081, 17083, 17085–17090), 6 Pacific Islander (17385–17388, 17390, 17391), and 7 Andeans (17301, 17302, 17306–17310). A common chimpanzee (*Pan troglodytes*) was also included in the sequencing. Double-stranded sequences in regions of interest were obtained by PCR amplification followed by sequencing of PCR products. Sequenced regions include 24750–26292, 26988–29992, 30561–32132, 32841–42938, 43006–44351, 45808–49406, 50123–50908, and 52305–53776 (the first base of the initiation codon of *Microcephalin* is defined as position 1). The core region used for haplotype analysis spans 29027 bases (24750–53776), of which 23416 bases were sequenced. Sequence chromatograms were aligned by the Sequencher software (Gene Codes Corporation, Ann Arbor, MI). Polymorphisms were detected by direct visual inspection of sequence chromatograms. The ancestral alleles of polymorphisms were called using the chimpanzee sequence as outgroup. Inference of haplotypes from the diploid sequence data was performed using the PHASE 2.1 software as described (*S1*, *S2*), which is available online at http://www.stats.ox.ac.uk/mathgen/home.html. Nucleotide diversity (π) and Tajima's *D* were calculated using the program DnaSP 3.51, as described previously (*S3*). To avoid uncertainties of haplotype inference, the 47 individuals who are homozygous for haplogroup D chromosomes were used for the calculation of π and Tajima's *D* of this haplogroup. Inferred haplotypes were used to calculate π and Tajima's *D* for the non-D chromosomes. Recombinants between D and non-D chromosomes were excluded from the calculation.

Genotyping

Genotyping of the G37995C nonsynonymous polymorphism in *Microcephalin* was performed on a panel of 1184 human samples. This panel does not overlap with the Coriell panel described above. It consists of the HGDP CEPH diversity panel as described previously (*S4*), minus the following two sets of samples. One is a set of duplicated samples that needed to be removed, including HGDP00472, HGDP00452, HGDP00457, HGDP00980, HGDP00650, HGDP00583, HGDP00111, HGDP00220, HGDP00813, HGDP01233, HGDP00762, HGDP00770, HGDP00657, HGDP00658, HGDP00660, and HGDP01149. The other is a set of samples that failed to be

genotyped due to technical reasons (e.g., poor DNA quality), including HGDP01263, HGDP00633, HGDP00635, HGDP00636, HGDP00644, HGDP00579, HGDP00581, HGDP00584, HGDP00698, HGDP00700, HGDP00722, HGDP00723, HGDP00724, HGDP00725, HGDP00730, HGDP00731, HGDP00732, HGDP00734, HGDP00746, HGDP00076, HGDP00090, HGDP00109, HGDP00115, HGDP00122, HGDP00125, HGDP00141, HGDP00254, HGDP00281, HGDP00782, HGDP00783, HGDP01023, HGDP01193, HGDP01311, HGDP01334, HGDP00766, HGDP00768, HGDP00662, HGDP00520, HGDP00666, HGDP01077, HGDP01386, HGDP01402, HGDP00890, HGDP00707, HGDP00708, HGDP00995, HGDP00998, HGDP01010, and HGDP00841. The CEPH panel originally contained 1064 individuals, and had 999 individuals remaining after removing the above two sets of samples. Demographic information for the HGDP CEPH diversity panel is available online at http://www.cephb.fr. In addition, the panel contained 185 sub-Saharan African samples collected by S. A. Tishkoff and A. Froment (sample collection was approved by the Institutional Review Board at the University of Maryland). The samples included 23 Turu, 32 Sandawe, 28 Burunge, and 27 Masai individuals from Tanzania; they also included 24 Bakola Pygmy, 28 Bamoun, and 23 Zime individuals from Cameroon. To perform genotyping, a small region encompassing the G37995C polymorphism was amplified by PCR, followed by sequencing of the PCR product. Genotype was scored by visual inspection of the sequence chromatograms. F_{ST} was calculated as described previously (*S5*). The exact formulas are available on pages 143–155 of (*S6)*.

Statistical analysis

To test the statistical significance that the frequency of haplotype 49 departs from neutral expectation, we used a previously described simulation method based on the coalescent process as implemented in the ms software (*S7*, S*8*). First, the following parameters were specified: the number of chromosomes, the number of segregating sites, recombination rate, gene conversion rate, and demographic model. Recombination rate of the *Microcephalin* region was set at the locus-specific value of 1.9 cM/Mb as obtained in a previous genomewide survey (*S9*), and gene conversion rate was set to be the same as recombination rate with an average tract length of 100 bp. The gene conversion model was as previously described (*S10*), which assumes that the tract length is geometrically distributed. Nine demographic models were tested:

1) constant population with an effective size of 10^4,

2) an ancient population expansion from 10^4 at 5,000 generations ago exponentially to 10^7 today,

3) a recent population expansion from 10^4 at 1,000 generations ago exponentially to 10^7 today,

4) a severe bottle neck starting 5,000 generations ago that reduced the population from 10^4 instantly to 10^3 and lasted until 2,500 generations ago at which point the population started to expand exponentially to 10^7 today,

5) repeated bottlenecks for five successive rounds starting 7000 generations ago, each from 10^4 instantly to 10^3 for 500 generations followed by exponential recovery back to 10^4 over another 500 generations, except at the end of the fifth bottleneck 2500 generations ago which was followed by exponential growth to 10^7 today,

6) population structure where the initial 178 chromosomes were split equally into 2 different subpopulations under constant population size with 1 migration per generation, and

7 to 9) population structure where the initial 178 chromosomes were split equally into 3 to 5 different subpopulations with 1 migration per generation. Command lines in the ms program to input the above demographic models were as follows:

<u>1) Constant population size:</u>

./ms 178 100000 –s 220 –r 11.6104 29027 –c 1 100 |./samh 18| wc

2) Ancient population expansion:
./ms 178 100000 –s 220 –r 11610.4 29027 –c 1 100 –G 55262.04223 –eG 0.000125 0 |./samh 18| wc

3) Recent population expansion:
./ms 178 100000 –s 220 –r 11610.4 29027 –c 1 100 –G 276310.2112 –eG 0.000025 0 |./samh 18| wc

4) Several bottleneck:

./ms 178 100000 –s 220 –r 11610.4 29027 –c 1 100 –G 147365.446 –eG 0.0000625 0 –eN 0.000125 0.001 |./samh 18| wc

5) Repeated bottlenecks with subsequent expansion:

./ms 178 100000 –s 220 –r 11610.4 29027 –c 1 100 –G 147365.446 –eG 0.0000625 0 –eN 0.000075 0.001 –eG 0.000075 184206.8074 –eG 0.0000875 0 –eN 0.0001 0.001 –eG 0.0001 184206.8074 –eG 0.0001125 0 –eN 0.000125 0.001 –eG 0.000125 184206.8074 –eG 0.0001375 0 –eN 0.00015 0.001 –eG 0.00015 184206.8074 –eG 0.0001625 0 –eN 0.000175 0.001 |./samh 18| wc

6) Population structure with 2 subpopulations:

./ms 178 100000 –s 220 –r 11.6104 29027 –c 1 100 –es 0.0 1 .5 –eM 0.0 1.0 |./samh 18| wc

7) Population structure with 3 subpopulations:

./ms 178 100000 –s 220 –r 11.6104 29027 –c 1 100 –es 0.0 1 0.3333 –es 0.0 1 0.5 –eM 0.0 1.0 |./samh 18| wc

8) Population structure with 4 subpopulations:

./ms 178 100000 –s 220 –r 11.6104 29027 –c 1 100 –es 0.0 1 .25 –es 0.0 1 .333 –es 0.0 1 .5 –eM 0.0 1.0 |./samh 18| wc

9) Population structure with 5 subpopulations:

./ms 178 100000 –s 220 –r 11.6104 29027 –c 1 100 –es 0.0 1 0.2 –es 0.0 1 0.25 –es 0.0 1 0.333 –es 0.0 1 0.5 –eM 0.0 1.0 |./samh 18| wc

Age estimation

We estimated the age of haplogroup D using a mutation-based method as previously described (*S11*). This method simply relies on averaging the number of mutations along each lineage from the most recent common ancestor (MRCA) to the sampled chromosome. This averaging produces an estimate of the time to MRCA that is unbiased by demographic history (*S11*). Let t denote the time to MRCA for haplogroup D in units of mutations. The value of t could be estimated as follows: To start, we decided to focus only on the 47 individuals who are homozygous for haplogroup D chromosomes (rather than using all the inferred copies of haplogroup D). This avoided uncertainties in haplotype inference. We also note that there are no evident recombinants between D and non-D types among these 47 individuals, which is important because the absence of such recombinants is a necessary condition for our methodology (*S11*, *S12*). Using chimpanzee sequence as an outgroup, we deduced the MRCA sequence of haplogroup D, which happens to be the same as the sequence of haplotype 49. We next added up the total number of mutations separating the MRCA and the 94 chromosomes sampled in the 47 individuals. This number was 93, which was divided by 94 to yield $\hat{t}$, the estimate of t, at 0.989. This value was then divided by 23416 (the total length of DNA sequenced) to yield an estimate for the number of mutations per base ($\hat{T}$) of 4.2×10^{-5}. By comparing human and chimpanzee sequences in this region, the rate of human-chimpanzee nucleotide divergence (D) in this region was estimated at 0.0136 mutations per base. Finally, human-chimpanzee divergence time (L) was set at 6×10^6 years. Most estimates of this time is between 5×10^6

and 6×10^6 years. We chose the upper one to be conservative. The estimated time to MRCA in years was then obtained, using the simple formula $(2\hat{T}/D)^* L$ as described previously (*S11*), at 37,281 years before present. The coalescence age of the entire Coriell panel was calculated in a similar manner. There are a total of 8136 mutations between the 178 chromosomes in the Coriell panel and the deduced MRCA sequence, which leads to an age estimate of 1,722,347 years. We note that owing to recombination, this estimated age is actually the average of multiple coalescence ages corresponding to multiple recombination blocks that coalesce independently.

The 95% confidence interval (CI) for the age of haplogroup D was estimated by an analytical approach that is an extension of a previously described method (*S12*). Let y_i denote the number of differences between the MRCA and the i^{th} chromosome. The value of $\hat{t}$ would be $(\sum_{i=1}^{n} y_i)/n$, where n is the number of chromosomes sampled. The variance of $\hat{t}$ is $[\sum_{i=1}^{n} \text{var}(y_i) + 2\sum_{i<j} \text{cov}(y_i, y_j)]/n^2$. If we assume an infinite-sites model, each y_i is Poisson distributed with mean t. The $\text{var}(y_i)$ is simply t, and the $\text{cov}(y_i, y_j)$ is simply $t - t_{ij}$, where t_{ij} is the time of the most recent common ancestor of chromosome i and chromosome j (*S11*, *S12*). Therefore the variance of our estimate is $t/n + 2[\sum_{i<j}(t - t_{ij})]/n^2$. There are $n(n-1)/2$ terms in this sum, so this can be written as $t - 2[\sum_{i<j}(t_{ij})]/n^2$ or $t - [(n-1)/n]\bar{t}_{ij}$, where $\bar{t}_{ij}$ is the average time to the most recent common ancestor of a pair of chromosomes. $\bar{t}_{ij}$ can be estimated as one-half the average pairwise differences between the 94 chromosomes, calculated as $(1/2)\sum_{k=1}^{m}\{2f_k(m - f_k)/[m(m-1)]\}$ or $\sum_{k=1}^{m}\{f_k(m - f_k)/[m(m-1)]\}$, where f_k is the count of the derived allele at the k^{th} polymorphic site and m is the total number of polymorphic sites. So we can estimate the variance of $\hat{t}$ by $\hat{t} - [(n-1)/n]\sum_{k=1}^{m}\{f_k(m - f_k)/[m(m-1)]\}$. For the 94 haplogroup D chromosomes sampled in the 47 individuals, there are 34 SNP sites. Let N_x designate the number of sites where the count of the derived allele is x. For our data, N_1 = 23, N_2 = 2, N_3 = 5, N_4 = 1, N_{15} = 2, N_{17} = 1, and all others N_x values are zero. Thus, based on our data, the estimate for the variance of $\hat{t}$ is 0.094, and the estimate for the standard error of $\hat{t}$ is $\sqrt{0.094}$ = 0.307. Assuming that the $\hat{t}$ estimator is roughly normally distributed, the 95% CI of $\hat{t}$ would be approximately 0.376 to 1.60. This corresponds, in units of years, a CI of 14175 to 60387 years before present. We note that this CI does not consider uncertainties in mutation rate. It also does not consider uncertainties in the estimated human-chimpanzee divergence time, which can only be inferred from fossil records and molecular data, and cannot be directly observed.

References and Notes

S1. M. Stephens, N. J. Smith, P. Donnelly, *Am. J. Hum. Genet.* **68**, 978 (2001).

S2. M. Stephens, P. Donnelly, *Am. J. Hum. Genet.* **73**, 1162 (2003).

S3. J. Rozas, R. Rozas, *Bioinformatics* **15**, 174 (1999).

S4. H. M. Cann *et al.*, *Science* **296**, 261 (2002).

S5. B. S. Weir, C. C. Cockerham, *Evolution* **38**, 1358 (1984).

S6. B. S. Weir, *Genetic Data* Analysis (Sinauer Associates, Sunderland, 1990).

S7. R. R. Hudson, *Oxford Surv.Evol. Biol.* **7**, 1 (1990).

S8. R. R. Hudson, *Bioinformatics* **18**, 337 (2002).

S9. A. Kong *et al.*, *Nat. Genet.* **31**, 241 (2002).

S10. C. Wiuf, J. Hein, *Genetics* **155**, 451 (2000).

S11. R. Thomson, J. K. Pritchard, P. Shen, P. J. Oefner, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360 (2000).

S12. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, *Genetics* **161**, 447 (2002).

Supplementary Table S1. Genotype data of *Microcephalin* in the 89 individuals of the Coriell panel.

Genomic position from start codon	25082	25134	25233	25250	25457	25480	25567	25604	25749	26018	26040	26162	26212	26215	26252	26256	26281	27149	27297	27307	27336	27370	27401	27629	27636	27701	27819	27961	28012	28089	28146	28182	28190	28285	28565	28986	29290	29678	29703	29742	29821	29848	30857	30884	30943	30979	31062	31126	31232	31369	31449	31482	31563	31641	31708-31712	
Population	**ID #**																																																							
Subsaharan Africans	17341	R	C	Y	T	Y	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	R	G	C	C	C	G	G	Y	C	W	A	C	T	T	5G's
	17342	R	C	Y	T	Y	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	G	R	S	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	R	G	C	C	C	G	G	T	C	W	A	C	T	T	5G's
	17343	A	C	C	T	T	G	G	C	Y	Y	A	C	C	A	A	C	G	C	C	G	A	C	G	R	G	A	C	A	C	G	S	G	A	A	C	G	T	C	A	Y	A	G	C	C	C	G	G	T	C	A	A	S	T	T	5G's
	17344	A	S	Y	T	T	G	G	C	Y	Y	A	M	M	A	A	C	G	C	C	G	A	C	G	G	R	G	A	Y	A	M	G	C	G	G	A	R	Y	G	T	C	W	T	A	G	C	C	C	G	G	T	C	A	A	G	T	W	5G's
	17345	A	C	C	T	T	G	G	C	Y	C	M	C	C	A	A	C	G	C	C	G	A	C	G	R	G	A	C	A	C	G	C	G	G	A	R	C	G	T	Y	A	Y	A	G	C	Y	C	G	G	T	C	A	A	C	T	T	5G's
	17346	A	C	T	T	T	K	G	C	C	C	M	C	C	M	R	W	C	G	C	C	Y	G	R	Y	G	G	G	A	Y	A	C	G	S	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	W	A	S	T	W	5G's
	17347	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	C	T	T	5G's
	17348	A	C	C	T	T	G	G	C	Y	Y	A	C	C	A	A	C	G	C	C	G	A	C	G	R	G	A	C	R	C	G	C	G	G	A	R	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	W	A	C	T	T	5G's
	17349	R	C	Y	T	Y	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	G	R	S	A	C	A	C	G	C	G	G	A	R	C	G	T	Y	A	T	R	G	C	C	Y	G	G	T	Y	W	A	C	T	T	5G's
North Africans	17378	R	C	C	T	Y	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	R	A	S	A	C	A	C	G	C	R	A	G	C	G	T	T	A	T	R	G	C	Y	C	G	G	T	C	A	A	C	T	T	4G's
	17379	A	C	T	T	T	K	G	C	C	C	M	C	C	M	R	W	C	G	C	Y	G	R	Y	G	G	G	A	Y	A	C	G	S	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	W	A	S	T	W	5G's
	17380	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	C	G	G	A	G	C	A	T	T	A	T	G	G	C	C	C	G	G	G	C	C	A	A	C	T	T	5G's
	17381	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17382	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	A	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17383	R	C	Y	T	Y	G	G	C	Y	Y	A	C	M	A	A	C	G	C	C	G	A	C	G	R	S	A	Y	A	C	G	S	G	A	R	C	G	T	Y	A	T	R	G	C	C	C	G	G	T	C	A	A	G	T	T	5G's
	17384	R	C	Y	T	Y	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	A	Y	G	R	G	A	C	A	C	G	C	G	G	A	R	C	R	T	Y	A	T	R	G	C	C	C	G	G	Y	C	W	A	C	T	T	5G's
Iberians	17091	A	C	T	T	T	K	G	C	C	C	M	C	C	M	R	W	C	G	C	Y	G	A	Y	G	G	G	A	Y	A	C	G	S	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	W	A	S	T	T	5G's
	17092	A	C	Y	T	T	K	G	C	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	R	R	G	A	C	A	C	G	C	R	R	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	17093	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	G	A	A	G	C	G	T	T	A	T	A	G	C	C	T	C	G	G	T	C	A	A	C	T	T	4G's
	17094	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	C	G	A	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17095	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	C	G	A	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17096	A	C	Y	T	T	K	G	C	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	A	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	17097	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	Y	C	C	G	G	T	C	T	A	C	T	T	5G's
	17099	A	C	Y	T	T	K	G	C	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	17100	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	A	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
Basques	15883	A	C	Y	T	T	K	S	C	Y	Y	M	C	C	R	W	Y	G	C	Y	G	A	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	15884	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	15885	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	15886	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	R	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	15887	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	16185	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	A	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	S	G	T	C	W	A	C	T	T	Het
	16188	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
Russians	13820a	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	13838	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	13852	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	13876	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	G	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T	4G's
	13877	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	Y	T	5G's
	13911	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	13912	A	C	C	T	T	G	G	C	T	T	C	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T	4G's
	13913	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	13914	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	R	T	C	T	A	C	T	T	5G's
Middle Easterners	17331	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17332	A	C	T	T	T	K	G	C	C	C	M	C	C	M	R	W	C	G	C	Y	G	R	Y	G	G	G	A	Y	A	C	G	S	G	A	A	C	G	T	C	A	T	A	G	C	C	Y	C	G	G	T	C	W	A	S	T	T	5G's
	17333	R	C	C	T	Y	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	R	A	G	A	C	A	C	G	C	R	A	G	C	R	T	T	A	T	R	G	C	Y	C	G	G	Y	C	A	A	C	T	T	4G's
	17334	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	Y	C	C	G	G	T	C	T	A	C	T	T	5G's
	17335	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	A	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	17336	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	Y	C	C	G	G	T	C	T	A	C	T	T	5G's
	17337	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17339	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17340	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	Y	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	Het
South Asians	17021	A	C	Y	T	T	K	G	Y	Y	Y	M	C	C	R	W	C	G	Y	Y	G	R	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
	17022	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	R	Y	G	R	G	A	C	A	C	G	S	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	W	M	S	T	T	5G's
	17023	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17024	A	C	Y	T	T	K	G	Y	Y	Y	M	C	C	R	W	C	G	Y	Y	G	R	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
	17026	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17027	A	C	C	T	T	G	G	Y	T	T	A	C	C	A	A	C	G	Y	C	R	A	C	R	A	G	A	C	A	C	G	C	R	A	G	C	G	T	T	A	T	A	G	C	Y	C	G	G	T	C	A	A	C	T	T	Het
	17028	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	R	R	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	C	C	G	G	T	C	W	A	C	T	T	Het
	17029	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17030	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	R	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T	4G's
Chinese	16654	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	R	T	G	G	G	A	C	A	C	R	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	16688	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	16689	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17014	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17015	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17017	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	S	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17018	A	C	T	W	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17019	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
Japanese	11587	A	C	Y	T	T	K	G	Y	Y	Y	M	C	C	R	W	C	G	Y	Y	G	R	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
Southeast Asians	17081	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17083	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17085	A	C	Y	T	T	K	G	Y	Y	Y	M	C	C	R	W	C	G	Y	Y	G	R	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
	17086	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17087	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17088	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17089	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17090	A	C	Y	T	T	K	G	Y	Y	Y	M	C	C	R	W	C	G	Y	Y	G	A	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
Pacific Islanders	17385	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17386	A	C	Y	T	T	K	G	C	Y	Y	M	C	C	R	W	C	G	C	Y	G	A	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
	17387	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17388	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17390	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17391	A	C	Y	T	T	K	G	Y	Y	Y	M	C	C	R	W	C	G	Y	Y	G	R	Y	G	R	G	A	C	A	C	G	C	G	A	R	C	G	T	Y	A	T	A	G	C	Y	C	G	G	T	C	W	A	C	T	T	5G's
Andeans	17301	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	Y	T	A	C	T	T	5G's
	17302	A	C	Y	T	T	K	G	C	Y	Y	C	C	C	R	W	C	G	C	Y	G	A	Y	R	R	G	A	C	A	C	G	C	R	A	R	C	G	T	Y	A	T	A	G	C	C	Y	C	G	G	T	C	W	A	C	T	T	Het
	17306	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	Y	T	5G's
	17307	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	S	C	T	G	G	R	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17308	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17309	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	S	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T	5G's
	17310	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	C	T	G	G	T	G	G	G	G	A	C	A	C	G	C	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	Y	T	5G's
Chimpanzee		A	C	C	T	T	G	G	C	C	C	A	C	C	A	A	C	G	C	C	G	G	C	G	G	G	A	C	A	C	G	C	G	A	A	C	A	T	C	A	T	A	G	T	T	C	G	G	T	C	A	A	C	T	T	5G's

Note: Individuals are highlighted in light blue background across all segregating sites if they are homozygous for haplogroup D chromosomes, not highlighted if homozygous for non-D haplotypes, and highlighted in yellow if heterozygous between D and non-D haplotypes. Individuals homozygous for haplotype 49 (which is the most abundant member of haplogroup D) have their identification numbers colored in red. Haplogroup D is defined as having the derived C allele at the diagnostic SNP site G37995C. The genomic position of this diagnostic site is colored in red. For three segregating sites, the ancestral or derived status of alleles could not be assigned because neither allele is the same as the chimpanzee nucleotide. The genomic positions of these three sites are colored in blue.

31879	31898	31954	32048	32862	32937	32940	32977	33165	33258	33308	33706	33768	33951	34135	34475	34732-34738	34843	35017	35324	35344	35371	35685-35692	35857	36044	36080	36582	36654	36658	36673	36711	36759	36887	36972	36986	36998	37000	37006	37013	37047	37089	37191	37215	37284	37331	37358	37369	37415	37651	37824	37845	37918	37966	37995	38173	38774	38783	38793	38836	38837	39221
G	A	C	A	C	Y	G	S	A	C	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	M	A	T	R	T	Y	G	C	G	S	A	G	T	G	G	G	A	A	S	Y	R	A	G	A	S	A	M	G	S	G	C	C	A	C	G	K
G	A	C	A	C	Y	G	S	A	Y	C	A	R	A	M	A	Het	T	A	G	G	C	Het	T	G	M	A	T	R	T	Y	G	C	G	S	A	G	T	G	G	G	A	A	S	Y	R	A	G	A	G	R	M	G	S	G	C	C	A	C	G	K
G	A	C	A	Y	C	G	G	A	T	C	A	R	A	M	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	T	C	G	S	G	G	A	G	T	G	G	G	A	A	C	T	A	A	G	A	G	A	C	G	G	G	C	C	R	C	G	G
G	M	M	A	C	C	K	G	R	T	Y	A	G	A	A	A	7T's	T	R	G	A	C	8T's	T	G	A	A	T	G	T	C	G	C	G	G	A	G	T	G	G	G	A	A	C	T	R	A	G	A	G	A	C	G	G	G	C	C	A	C	G	K
G	A	C	A	Y	C	G	G	A	T	C	A	R	A	M	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	T	C	G	C	G	G	A	G	T	G	G	R	A	A	C	T	A	A	G	A	G	A	C	G	G	G	C	Y	A	C	G	G
G	A	C	A	C	Y	G	G	R	Y	Y	A	G	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	R	C	G	G	A	G	T	G	G	G	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	C	A	C	G	K
C	A	C	A	C	C	G	C	A	T	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	M	A	T	R	T	Y	G	C	G	S	A	G	T	G	G	G	A	A	S	Y	R	A	G	A	G	R	M	G	S	S	C	C	A	C	G	K
G	A	C	A	C	Y	G	G	A	Y	C	A	G	A	M	A	Het	T	A	G	R	C	9T's	T	G	A	A	K	R	T	Y	G	C	G	G	A	G	Y	G	G	G	A	R	S	T	R	W	G	R	G	A	C	G	S	G	C	C	A	C	G	K
G	A	C	A	C	Y	G	S	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	M	A	T	R	T	Y	G	C	G	S	A	G	T	G	G	G	A	A	S	Y	R	A	G	A	G	R	M	G	S	G	C	C	A	C	G	K
G	A	C	A	C	C	G	S	A	T	C	A	R	A	A	A	7T's	T	A	G	A	C	8T's	T	G	M	A	T	G	Y	C	G	C	G	S	A	G	T	G	R	G	A	A	C	T	R	A	G	A	G	R	M	K	G	G	C	C	A	C	G	G
G	A	C	A	C	Y	G	G	R	Y	Y	A	G	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	G	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	C	A	C	G	K
G	A	C	A	C	C	G	C	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	C	A	T	G	T	C	G	C	G	C	A	G	T	G	G	G	A	A	C	C	A	A	G	A	G	A	A	G	G	G	C	C	A	C	G	G
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	R	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	R	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	C	G	S	R	T	Y	A	R	A	A	A	7T's	T	R	G	A	C	8T's	T	G	M	A	T	G	T	C	G	C	G	S	A	G	T	G	G	G	A	A	C	Y	A	A	G	A	G	R	M	G	G	G	C	C	A	C	G	G
G	A	C	A	C	Y	G	S	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	M	A	T	R	T	Y	G	C	G	S	A	G	T	G	G	G	A	A	S	Y	R	A	G	A	G	A	M	G	S	G	C	C	A	C	G	K
G	A	C	A	C	Y	G	G	A	Y	C	A	G	A	C	A	Het	T	A	G	G	C	Het	T	G	A	A	T	A	T	Y	G	C	G	G	R	R	T	G	G	G	R	A	S	T	G	A	G	A	G	A	C	G	S	G	Y	C	A	C	R	K
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	G	G	K
G	A	C	A	C	C	G	G	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	C	C	G	C	G	G	A	G	T	G	A	G	A	A	C	T	G	A	G	A	G	A	C	K	G	G	C	C	A	C	G	G
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	M	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	R	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	G	K
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	R	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	R	K
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	R	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	A	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	R	G	R	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	Y	C	A	R	A	M	A	Het	W	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	G	K
G	A	C	A	C	T	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	S	A	G	A	C	K	S	G	C	C	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	C	G	G	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	C	C	G	C	G	G	A	G	T	G	A	G	A	A	C	T	G	A	G	A	G	A	C	G	G	G	C	C	A	C	G	G
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	G	K
G	A	C	A	C	C	G	G	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	C	C	G	C	G	G	A	G	T	G	A	G	A	A	C	T	G	A	G	A	G	A	C	T	G	G	C	C	A	C	G	G
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	R	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	K	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	R	Y	Y	A	G	A	M	A	Het	T	R	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	G	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	C	A	C	G	K
G	A	C	A	C	C	G	S	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	M	A	T	G	Y	C	G	C	G	S	A	G	T	G	R	G	A	A	C	Y	R	A	G	A	G	A	C	K	G	G	C	C	A	C	G	G
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	R	K
G	A	C	A	C	T	G	G	A	C	C	A	G	R	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	G	K
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	Y	A	C	G	K
G	M	C	R	C	T	G	G	A	C	C	A	G	A	C	A	Het	T	A	G	G	C	Het	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	S	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	T	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	C	G	G	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	Y	C	G	C	G	G	A	G	T	G	R	R	A	A	C	T	R	A	G	A	G	A	C	K	G	G	C	Y	A	C	G	G
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	Y	Y	G	C	R	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	R	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	C	G	G	A	T	C	A	A	A	A	A	7T's	T	A	G	A	C	8T's	T	G	A	A	T	G	C	C	G	C	G	G	A	G	T	G	A	G	A	A	C	T	G	A	G	A	G	A	C	T	G	G	C	C	A	C	G	G
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	K	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	S	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	Y	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	Y	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	T	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	C	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	T	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	T	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	R	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	T	A	C	R	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	Het	T	G	A	A	T	R	T	Y	G	C	G	G	A	G	T	G	G	R	A	A	S	T	R	A	G	A	G	A	C	G	S	G	C	T	A	C	G	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	Y	G	G	A	Y	C	A	R	A	M	A	Het	T	A	G	R	C	8T's	T	G	A	A	T	R	Y	Y	G	C	G	G	A	G	T	G	R	G	A	A	S	T	G	A	G	A	G	A	C	K	S	G	C	C	A	C	R	K
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	S	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	R	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	C	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	G	T
G	A	C	A	C	T	G	G	A	C	C	A	G	A	C	A	8T's	T	A	G	G	C	9T's	T	G	A	A	T	A	T	T	G	C	G	G	A	G	T	G	G	G	A	A	G	T	G	A	G	A	G	A	C	G	C	G	C	C	A	C	A	T
G	A	C	A	C	C	G	G	A	T	C	A	C	A	C	A	7T'S	T	A	G	G	C	8T's	T	G	A	A	T	A	T	C	G	C	A	G	A	G	T	G	G	G	A	A	C	T	G	A	G	A	G	A	C	G	G	G	C	C	A	C	G	T

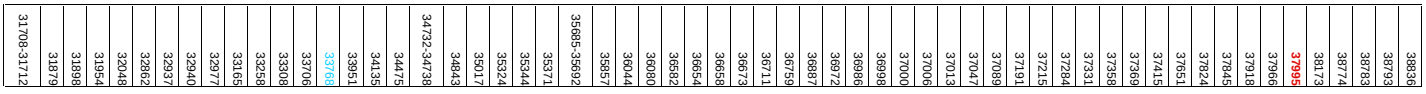
39235	39246	39323	39401	39426	39561	39622	39658	39671	39802	39997	40071	40160	40254	40262	40287	40343-40356	40505	40506	40542	40624	40626	40660	40828	40866	40907	40916	41000	41038	41111	41112	41234	41274	41320	41567	41741	41745	41766	41800	41807	41836	41874	41893	41922	41973	42431	42481	42601	42649	43102	43134	43145	43264	43338-43341	43607	43868	43993
C	W	W	A	M	A	G	R	Y	A	T	R	Y	G	G	A	Het	T	R	M	C	R	T	A	Y	T	W	M	A	C	G	T	T	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	R	Y	A	T	R	Y	G	G	A	Het	T	R	M	C	R	T	A	Y	T	W	M	A	C	G	T	T	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	R	Y	C	K	Y	G	4 bp insertion	R	M	R
Y	T	T	A	A	A	G	A	C	A	A	G	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	A	A	A	C	G	T	W	C	G	G	C	C	T	A	G	T	G	C	T	G	G	A	T	C	G	T	G	4 bp insertion	A	A	G
C	T	T	A	A	A	G	A	C	A	A	K	G	C	C	G	R	K	No deletion	T	A	A	C	A	T	A	T	T	A	A	A	Y	G	T	K	W	C	C	G	G	C	C	S	C	T	A	G	G	T	G	C	C	T	Y	G	G	G	A	T	C	G	T	S	Het	A	A	G
C	T	T	A	A	A	G	A	C	A	A	G	G	C	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	C	G	T	W	C	C	G	G	C	C	C	T	A	G	G	T	G	C	C	C	G	G	G	A	T	S	K	T	G	4 bp insertion	A	A	G
C	W	T	A	M	A	G	A	Y	A	A	K	R	C	G	G	A	Het	T	R	M	C	R	T	A	T	T	W	M	A	Y	G	T	W	C	C	G	G	C	C	S	C	T	A	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	R	Y	A	T	R	Y	G	G	A	Het	T	R	M	C	R	T	A	Y	T	W	M	A	C	G	T	T	S	G	G	C	C	T	A	R	W	G	C	Y	G	S	R	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	A	Y	A	T	R	Y	G	G	A	Het	W	R	M	C	R	T	A	T	T	W	M	A	C	G	T	T	C	G	G	C	C	T	A	G	W	G	Y	T	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	R	Y	A	T	R	Y	G	G	A	Het	T	R	M	C	R	T	A	Y	T	W	M	A	C	G	T	T	C	G	G	C	C	T	R	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	T	T	A	A	A	G	R	C	A	K	G	C	G	G	A	No deletion	W	A	A	C	A	Y	A	Y	T	A	A	R	C	G	T	W	C	G	G	C	C	T	A	G	T	G	C	C	G	S	R	T	C	G	T	G	4 bp insertion	A	A	G
C	W	W	A	M	A	G	R	Y	A	K	G	R	Y	G	G	R	Het	T	R	M	C	R	T	A	Y	T	W	M	A	C	G	T	T	C	G	G	C	C	T	R	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	T	T	A	A	A	G	G	C	A	A	T	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	C	T	A	A	A	C	G	T	T	C	G	G	C	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	4 bp insertion	A	A	G
C	A	A	A	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A
C	A	A	R	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A
C	T	T	A	A	A	G	R	C	A	K	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	Y	T	A	A	A	Y	G	T	W	C	G	G	C	C	T	A	G	T	G	C	C	G	S	R	T	C	G	T	G	4 bp insertion	A	A	G
C	W	W	A	M	A	G	R	Y	A	T	R	Y	G	G	A	Het	T	R	M	C	R	T	A	Y	T	W	M	A	C	G	T	T	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	A	Y	A	K	R	Y	G	G	A	Het	T	R	M	C	R	T	A	T	T	W	M	R	C	G	T	W	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	A	A	A	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A
C	A	W	A	C	A	G	A	Y	C	K	R	Y	G	G	A	Het	T	R	C	C	R	T	A	T	T	W	C	R	C	G	T	W	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	C	R
C	W	W	A	M	A	G	A	Y	A	K	R	Y	G	G	A	Het	T	R	M	C	R	T	A	T	T	W	M	R	C	G	T	W	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	A	Y	A	K	R	Y	G	G	A	Het	T	R	M	C	R	T	A	T	T	W	M	R	C	G	T	W	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	A	A	A	C	M	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A
C	A	A	A	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A
C	A	A	A	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A
C	W	W	A	M	A	G	A	Y	A	K	R	Y	G	G	A	Het	T	R	M	C	R	T	A	T	T	W	M	R	C	G	T	W	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	W	W	A	M	A	G	A	Y	A	K	R	Y	G	G	A	Het	W	R	M	C	R	T	A	T	T	W	M	R	C	G	T	W	C	G	G	C	C	T	A	R	W	G	C	Y	G	S	A	Y	C	K	Y	G	4 bp insertion	R	M	R
C	A	A	A	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	C	C	T	A	A	A	G	C	T	G	C	A	C	C	T	C	G	4 bp insertion	G	C	A

44055	44101	44158	44288	45926	45938	45991	46114	46230	46246	46550	46635	46660-46662	46772	47074	47169	47271	47741	47756	47792	47833	48094	48200	48519	48695	49021	49124	49184	49253	49320	50148	50301	50451	50681	50718	50719	50737	52631-52638	52654	52729	52930	52988	53006	53090	53223	53243	53248
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	R	G	Het	G	S	A	K	G	S	G	C	A
T	G	C	G	G	T	A	R	A	A	A	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	C	A	G	G	G	G	C	A
T	G	C	G	G	T	A	G	A	A	A	G	Het	G	A	G	C	T	C	A	C	S	G	C	Y	C	A	K	T	C	C	C	C	Y	C	G	R	Het	G	S	A	K	S	G	G	C	G
T	G	C	G	G	T	A	G	A	A	A	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	C	A	G	G	G	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	C	A	K	T	C	Y	Y	S	Y	C	G	G	Het	G	S	A	K	G	S	G	C	R
G	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	M	G	G	Het	G	S	A	K	G	S	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	Y	C	A	G	T	S	C	C	C	Y	C	G	R	Het	G	S	R	K	G	C	G	Y	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	R	G	Het	G	S	A	K	G	S	G	C	A
T	G	C	G	G	T	A	G	A	A	A	S	No 3 bp deletion	G	A	G	C	C	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	R	G	8 bp insertion	G	S	A	G	G	G	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	R
T	G	C	G	G	T	A	G	A	A	A	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	A	G	8 bp insertion	G	C	A	G	G	G	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
T	G	C	G	G	T	A	G	A	A	A	G	No 3 bp deletion	G	A	G	C	T	Y	M	C	C	G	C	Y	C	A	G	T	C	C	C	C	C	C	R	G	8 bp insertion	G	C	A	G	G	G	G	C	R
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	R	G	Het	G	S	A	K	G	S	G	C	A
G	S	Y	G	G	T	R	G	G	A	G	G	No 3 bp deletion	G	A	S	Y	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	R
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
T	G	C	G	G	T	A	G	A	A	A	G	No 3 bp deletion	G	A	G	C	C	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	G	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	Y	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	K	G	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	T	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	R	G	C	Y	C	A	C	C	G	Y	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	S	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
T	G	C	G	G	T	A	G	A	A	A	S	No 3 bp deletion	G	A	G	C	C	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	G	A	G	G	G	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
T	G	C	G	G	T	A	G	A	A	A	G	No 3 bp deletion	G	A	G	C	C	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	G	A	G	G	G	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	S	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	G	G	C	R
T	G	C	G	G	T	A	G	A	A	A	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	R	G	8 bp insertion	G	S	A	G	G	G	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	R	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	S	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	T	G	S	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	M	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	Y	C	G	C	T	M	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	R	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
T	G	C	G	G	T	A	G	A	W	A	G	No 3 bp deletion	G	A	G	C	C	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	S	A	G	G	G	G	C	A
K	S	C	G	G	T	A	G	R	W	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
T	G	C	G	G	T	A	G	A	W	A	G	No 3 bp deletion	G	A	G	C	C	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	G	G	8 bp insertion	G	G	A	G	G	G	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	G	G	K	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	S	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	K	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	C	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	S	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
K	S	C	G	G	T	A	G	R	A	R	G	No 3 bp deletion	G	A	G	Y	Y	C	A	C	C	G	C	T	M	A	K	T	C	Y	Y	C	Y	C	G	G	Het	G	G	A	K	G	S	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	R	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	C	T	C	A	C	C	A	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	C	C	G	G	T	A	G	G	A	G	G	No 3 bp deletion	G	A	G	Y	T	C	A	C	C	G	C	T	A	A	T	T	C	T	T	C	T	C	G	G	No insertion	G	G	A	T	G	C	G	C	A
G	G	C	G	G	T	G	G	G	A	T	G	No 3 bp deletion + 4 bp	G	A	G	C	T	C	A	C	C	G	C	T	C	A	T	T	C	C	C	C	C	C	G	A	No insertion	G	C	A	G	G	C	G	C	A

Supplementary Table S2. Haplotype data of *Microcephalin* in the 89 individuals of the Coriell panel.

Haplotype	Occurence (Genomic position from start codon →)	25082	25134	25233	25250	25457	25480	25567	25604	25749	26018	26040	26162	26212	26215	26252	26256	26281	27149	27297	27307	27336	27370	27401	27629	27636	27701	27819	27861	28012	28089	28146	28182	28190	28285	28565	28986	29290	29678	29703	29742	29821	29848	30857	30884	30943	30979	31062	31126	31232	31369	31449	31482	31563	31641
1	1	A	C	C	T	T	G	G	C	C	C	A	C	C	A	A	C	G	C	C	G	A	C	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
2	1	A	C	C	T	T	G	G	C	C	C	A	C	C	A	A	C	G	C	C	G	A	C	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	C	A	G	C	C	C	G	G	T	C	A	A	C	T	T
3	1	A	C	C	T	T	G	G	C	C	C	A	C	C	A	A	C	G	C	C	G	A	C	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	C	A	G	C	C	C	G	G	T	C	A	A	C	T	T
4	1	A	C	C	T	T	G	G	C	T	C	C	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
5	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	T
6	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	T
7	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	T
8	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	G	C	G	C	G	A	G	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	C	T	T
9	2	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
10	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
11	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
12	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
13	3	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
14	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
15	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
16	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
17	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
18	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
19	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	C	G	T	C	A	A	C	T	T
20	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	C	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
21	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	G	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
22	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	A	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	C	C	G	G	T	C	A	A	C	T	T
23	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	A	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	C	C	G	G	T	C	A	A	C	T	T
24	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	C	C	A	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
25	1	A	C	C	T	T	G	G	C	T	T	A	C	C	A	A	C	G	T	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
26	2	A	C	C	T	T	G	G	C	T	T	C	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
27	1	A	C	C	T	T	G	G	C	T	T	C	C	C	A	A	C	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
28	1	A	C	C	T	T	G	G	T	T	T	A	C	C	A	A	C	G	T	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
29	5	A	C	C	T	T	G	G	T	T	T	A	C	C	A	A	C	G	T	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
30	1	A	C	C	T	T	G	G	T	T	T	A	C	C	A	A	C	G	T	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
31	1	A	C	C	T	T	G	C	C	T	T	A	C	C	A	A	T	G	C	C	G	A	C	A	A	G	A	C	A	C	G	C	A	A	G	C	G	T	T	A	T	A	G	C	T	C	G	G	T	C	A	A	C	T	T
32	1	A	C	T	T	T	G	G	C	C	C	A	C	A	A	A	C	G	C	C	G	A	C	G	G	G	A	T	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	T
33	2	A	C	T	T	T	G	G	C	C	C	A	C	A	A	A	C	G	C	C	G	A	C	G	G	G	A	T	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	T
34	1	A	C	T	T	T	G	G	C	C	C	A	C	A	A	A	C	G	C	C	G	A	C	G	G	G	A	T	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	A
35	1	A	C	T	T	T	G	G	C	C	C	A	C	A	A	A	C	G	C	C	G	A	C	G	G	G	A	T	A	C	G	G	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	A	A	G	T	A
36	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
37	2	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
38	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
39	11	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
40	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
41	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
42	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
43	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
44	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
45	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	C	T
46	2	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	C	T
47	3	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	T	C	C	G	G	T	C	T	A	C	T	T
48	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	A	T	G	G	G	G	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
49	59	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
50	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
51	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
52	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
53	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
54	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
55	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
56	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
57	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
58	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
59	5	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
60	3	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
61	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
62	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
63	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
64	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
65	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
66	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
67	2	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
68	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	C	C	T	T
69	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	A	T	C	T	A	C	T	T
70	2	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	T	G	G	T	T	T	A	C	T	T
71	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	T	C	G	G	T	C	T	A	C	T	T
72	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	A	C	C	C	G	G	T	C	T	A	C	T	T
73	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	A	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
74	1	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	C	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
75	2	A	C	T	T	T	T	G	C	C	C	C	C	C	G	T	C	C	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
76	1	A	C	T	A	T	T	G	C	C	C	C	C	C	G	T	C	G	C	T	G	G	T	G	G	G	A	C	A	C	G	C	G	A	A	C	G	T	C	A	T	A	G	C	C	C	G	G	T	C	T	A	C	T	T
77	1	A	G	T	T	T	G	G	C	C	C	A	A	A	A	A	C	G	C	C	G	A	C	G	G	G	A	T	A	A	G	G	G	A	G	T	G	T	C	T	T	A	G	C	C	C	G	G	T	C	A	A	G	T	A
78	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	C	C	A	A	G	T	T
79	3	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	A	T	T	A	T	G	G	C	C	C	G	G	C	C	A	A	C	T	T
80	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	G	A	C	A	C	G	C	G	A	G	C	A	T	T	A	T	G	G	C	C	C	G	G	C	C	A	A	C	T	T
81	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	C	T	T
82	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	C	T	T
83	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	C	T	T
84	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	C	T	T
85	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	C	T	T
86	1	G	C	C	T	C	G	G	C	T	T	A	C	C	A	A	C	G	C	C	G	A	C	G	A	C	A	C	A	C	G	C	G	A	G	C	G	T	T	A	T	G	G	C	C	C	G	G	T	C	A	A	G	T	T
Chimpanzee		A	C	C	T	T	G	G	C	C	C	A	C	C	A	A	C	G	C	C	G	G	C	G	G	G	A	C	A	C	G	C	G	A	A	C	A	T	C	A	T	A	G	T	T	C	G	G	T	C	A	A	C	T	T

Note: Haplotypes are highlighted across all segregating sites in light blue background if they are members of haplogroup D, and not highlighted if they are non-D haplotypes. Haplotype 49 (which is the most abundant member of haplogroup D) has its identification number colored in red. Haplogroup D is defined as having the derived C allele at the diagnostic SNP site G37995C. The genomic position of this diagnostic site is colored in red. For three segregating sites, the ancestral or derived status of alleles could not be assigned because neither allele is the same as the chimpanzee nucleotide. The genomic positions of these three sites are colored in blue.



38837	39221	39235	39246	39323	39401	39426	39561	39622	39658	39671	39802	39997	40071	40160	40254	40262	40287	40343-40356	40505	40506	40542	40624	40626	40660	40828	40866	40907	40916	41000	41038	41111	41112	41234	41274	41320	41567	41741	41745	41766	41800	41807	41836	41874	41893	41922	41973	42431	42481	42601	42649	43102	43134	43145	43264	43338-43341	43607
G	T	C	A	A	A	C	A	G	A	T	A	T	A	T	G	G	A	14 bp deletion	A	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	G	C	C	T	A	G	T	A	G	A	G	T	T	G	C	A	C	C	T	C	G	No insertion	G
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	T	G	G	A	T	C	G	T	G	No insertion	A
G	G	T	C	T	T	A	A	A	G	A	C	A	G	G	C	G	G	A	No deletion	T	A	A	C	C	A	T	A	T	T	T	A	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	T	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	T	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	A	T	A	C	A	G	G	A	C	C	A	G	G	C	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	T	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	14 bp deletion	T	G	C	C	G	T	A	T	T	T	C	A	C	G	T	T	C	G	G	G	C	C	T	A	G	T	A	G	A	G	T	T	G	C	A	C	C	T	C	G	No insertion	G
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	C	G	T	T	C	G	G	G	C	C	T	A	G	T	G	C	C	T	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	T	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	A	A	A	C	A	T	A	T	T	T	A	A	G	C	G	T	T	A	C	G	G	G	T	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T	T	T	A	A	A	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	T	G	C	G	G	A	No deletion	A	A	A	C	A	T	A	T	T	T	A	A	G	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	G	G	C	G	G	A	No deletion	A	A	A	C	A	T	A	T	T	T	A	A	G	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	G	G	C	G	G	A	No deletion	A	A	A	C	A	C	A	T	T	A	A	G	C	G	T	T	A	C	G	G	G	C	C	T	A	G	T	G	C	C	G	G	A	T	C	G	T	G	No insertion	A
G	G	C	T	T	A	A	A	G	A	C	A	G	G	C	G	G	A	No deletion	T	A	A	C	A	T	A	T

43868	43993	44055	44101	44158	44288	45926	45938	45991	46114	46230	46246	46550	46635	46660-46662	46772	47074	47169	47271	47741	47756	47792	47833	48094	48200	48519	48695	49021	49124	49184	49253	49320	50148	50301	50451	50681	50718	50719	50737	52631-52638	52654	52729	52930	52988	53006	53090	53223	53243	53248
C	A	G	C	C	G	G	T	A	G	G	A	G	G	No deletion	G	A	G	C	T	C	A	C	C	G	C	T	C	C	A	G	T	C	C	C	C	C	T	C	G	A	No insertion	G	G	A	T	G	C	G	C	A
A	G	T	G	C	G	G	T	A	G	G	A	A	A	No deletion	G	A	G	C	T	C	A	C	C	G	C	T	C	A	G	T	C	C	C	C	C	C	C	G	G	8 bp insertion	G	C	A	G	G	G	G	C	A
A	G	T	G	C	C	G	G	T	A	G	G