GENETICS
Current Issue Information for Authors Editorial Board Subscribe Submit a Manuscript

Genetics. 2004 Jan; 166(1): 351–372. PMCID: PMC1470693

The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations.

Gabor T Marth, Eva Czabarka, Janos Murvai, and Stephen T Sherry

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA.
marth@bc.edu

Abstract

We have studied a genome-wide set of single-nucleotide polymorphism (SNP) allele frequency measures for African-American, East Asian, and European-American samples. For this analysis we derived a simple, closed mathematical formulation for the spectrum of expected allele frequencies when the sampled populations have experienced nonstationary demographic histories. The direct calculation generates the spectrum orders of magnitude faster than coalescent simulations do and allows us to generate spectra for a large number of alternative histories on a multidimensional parameter grid. Model-fitting experiments using this grid reveal significant population-specific differences among the demographic histories that best describe the observed allele frequency spectra. European and Asian spectra show a bottleneck-shaped history: a reduction of effective population size in the past followed by a recent phase of size recovery. In contrast, the African-American spectrum shows a history of moderate but uninterrupted population expansion. These differences are expected to have profound consequences for the design of medical association studies. The analytical methods developed for this study, i.e., a closed mathematical formulation for the allele frequency spectrum, correcting the ascertainment bias introduced by shallow SNP sampling, and dealing with variable sample sizes provide a general framework for the analysis of public variation data.

Full Text

The Full Text of this article is available as a PDF (690K).

Selected References

These references are in PubMed. This may not be the complete list of references from this article.

- Harpending H, Rogers A. Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet. 2000;1:361–385. [PubMed]
- Hey J. Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol Biol Evol. 1997 Feb;14(2):166–172. [PubMed]
- Hey J, Harris E. Population bottlenecks and patterns of human polymorphism. Mol Biol Evol. 1999 Oct;16(10):1423–1426. [PubMed]
- Ingman M, Kaessmann H, Päbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature. 2000 Dec 7;408(6813):708–713. [PubMed]

- Jorde LB, Watkins WS, Bamshad MJ. Population genomics: a bridge from evolutionary history to genetic medicine. Hum Mol Genet. 2001 Oct 1;10(20):2199–2207. [PubMed]
- Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989 Dec;123(4):887–899. [PMC free article] [PubMed]
- Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB. Signatures of population expansion in microsatellite repeat data. Genetics. 1998 Apr;148(4):1921–1930. [PMC free article] [PubMed]
- Akey Joshua M, Zhang Ge, Zhang Kun, Jin Li, Shriver Mark D. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 2002 Dec;12(12):1805–1814. [PMC free article] [PubMed]
- Kondrashov Alexey S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat. 2003 Jan;21(1):12–27. [PubMed]
- Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000 Sep 28;407(6803):513–516. [PubMed]
- Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet. 1999 Jun;22(2):139–144. [PubMed]
- Kuhner MK, Yamato J, Felsenstein J. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics. 1995 Aug;140(4):1421–1430. [PMC free article] [PubMed]
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics. 1995 Jun;140(2):783–796. [PMC free article] [PubMed]
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860–921. [PubMed]
- Cardon Lon R, Abecasis Gonçalo R. Using haplotype blocks to map human complex trait loci. Trends Genet. 2003 Mar;19(3):135–140. [PubMed]
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet. 1999 Jul;22(3):231–238. [PubMed]
- Marth Gabor, Schuler Greg, Yeh Raymond, Davenport Ruth, Agarwala Richa, Church Deanna, Wheelan Sarah, Baker Jonathan, Ward Ming, Kholodov Michael, et al. Sequence variations in the public human genome data reflect a bottlenecked population history. Proc Natl Acad Sci U S A. 2003 Jan 7;100(1):376–381. [PMC free article] [PubMed]
- Clark Andrew G. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. Curr Opin Genet Dev. 2003 Jun;13(3):296–302. [PubMed]
- Mullikin JC, Hunt SE, Cole CG, Mortimore BJ, Rice CM, Burton J, Matthews LH, Pavitt R, Plumb RW, Sims SK, et al. An SNP map of human chromosome 22. Nature. 2000 Sep 28;407(6803):516–520. [PubMed]
- Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics. 2000 Sep;156(1):297–304. [PMC free article] [PubMed]
- Di Rienzo A, Wilson AC. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci U S A. 1991 Mar 1;88(5):1597–1601. [PMC free article] [PubMed]
- Payseur Bret A, Cutter Asher D, Nachman Michael W. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Mol Biol Evol. 2002 Jul;19(7):1143–

1153. [PubMed]
- Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, Barch DH. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics. 1998 Mar;148(3):1269–1284. [PMC free article] [PubMed]
- Pluzhnikov Anna, Di Rienzo Anna, Hudson Richard R. Inferences about human demography based on multilocus analyses of noncoding sequences. Genetics. 2002 Jul;161(3):1209–1218. [PMC free article] [PubMed]
- Ewens WJ. The sampling theory of selectively neutral alleles. Theor Popul Biol. 1972 Mar;3(1):87–112. [PubMed]
- Przeworski Molly. The signature of positive selection at randomly chosen loci. Genetics. 2002 Mar;160(3):1179–1189. [PMC free article] [PubMed]
- Fay JC, Wu CI. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol. 1999 Jul;16(7):1003–1005. [PubMed]
- Przeworski M, Hudson RR, Di Rienzo A. Adjusting the focus on human variation. Trends Genet. 2000 Jul;16(7):296–302. [PubMed]
- Ptak Susan E, Przeworski Molly. Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. 2002 Nov;18(11):559–563. [PubMed]
- Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993 Mar;133(3):693–709. [PMC free article] [PubMed]
- Reich DE, Goldstein DB. Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci U S A. 1998 Jul 7;95(14):8119–8123. [PMC free article] [PubMed]
- Gabriel Stacey B, Schaffner Stephen F, Nguyen Huy, Moore Jamie M, Roy Jessica, Blumenstiel Brendan, Higgins John, DeFelice Matthew, Lochner Amy, Faggart Maura, et al. The structure of haplotype blocks in the human genome. Science. 2002 Jun 21;296(5576):2225–2229. [PubMed]
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, et al. Linkage disequilibrium in the human genome. Nature. 2001 May 10;411(6834):199–204. [PubMed]
- Gonser R, Donnelly P, Nicholson G, Di Rienzo A. Microsatellite mutations and inferences about human demography. Genetics. 2000 Apr;154(4):1793–1807. [PMC free article] [PubMed]
- Reich David E, Schaffner Stephen F, Daly Mark J, McVean Gil, Mullikin James C, Higgins John M, Richter Daniel J, Lander Eric S, Altshuler David. Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet. 2002 Sep;32(1):135–142. [PubMed]
- Griffiths RC, Tavaré S. Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci. 1994 Jun 29;344(1310):403–410. [PubMed]
- Relethford JH, Jorde LB. Genetic evidence for larger African population size during recent human evolution. Am J Phys Anthropol. 1999 Mar;108(3):251–260. [PubMed]
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science. 2001 Feb 16;291(5507):1304–1351. [PubMed]
- Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol. 1992 May;9(3):552–569. [PubMed]
- Wall Jeffrey D, Pritchard Jonathan K. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 2003 Aug;4(8):587–597. [PubMed]
- Rybicki Benjamin A, Iyengar Sudha K, Harris Trent, Liptak Rachel, Elston Robert C, Sheffer Roberta, Chen Kang Mei, Major Marcie, Maliarik Mary J, Iannuzzi Michael C. The distribution of long range admixture linkage disequilibrium in an African-American population. Hum Hered. 2002;53(4):187–196.

[PubMed]
- Wall JD, Przeworski M. When did the human population size start increasing? Genetics. 2000 Aug;155(4):1865–1874. [PMC free article] [PubMed]
- Weber James L, David Donna, Heil Jeremy, Fan Ying, Zhao Chengfeng, Marth Gabor. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet. 2002 Oct;71(4):854–862. [PMC free article] [PubMed]
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001 Feb 15;409(6822):928–933. [PubMed]
- Wiehe T. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theor Popul Biol. 1998 Jun;53(3):272–283. [PubMed]
- Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M. Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol. 1994 Oct;66(5):761–775. [PubMed]
- Wooding Stephen, Rogers Alan. The matrix coalescent and an application to human single-nucleotide polymorphisms. Genetics. 2002 Aug;161(4):1641–1650. [PMC free article] [PubMed]
- Sherry ST, Harpending HC, Batzer MA, Stoneking M. Alu evolution in human populations: using the coalescent to estimate effective population size. Genetics. 1997 Dec;147(4):1977–1982. [PMC free article] [PubMed]
- Yu N, Zhao Z, Fu YX, Sambuughin N, Ramsay M, Jenkins T, Leskinen E, Patthy L, Jorde LB, Kuromori T, et al. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol Biol Evol. 2001 Feb;18(2):214–222. [PubMed]
- Sunyaev SR, Lathe WC, 3rd, Ramensky VE, Bork P. SNP frequencies in human genes an excess of rare alleles and differing modes of selection. Trends Genet. 2000 Aug;16(8):335–337. [PubMed]
- Zhao Z, Jin L, Fu YX, Ramsay M, Jenkins T, Leskinen E, Pamilo P, Trexler M, Patthy L, Jorde LB, et al. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci U S A. 2000 Oct 10;97(21):11354–11358. [PMC free article] [PubMed]
- Tavaré S, Balding DJ, Griffiths RC, Donnelly P. Inferring coalescence times from DNA sequence data. Genetics. 1997 Feb;145(2):505–518. [PMC free article] [PubMed]
- Zhivotovsky LA, Bennett L, Bowcock AM, Feldman MW. Human population expansion and microsatellite variation. Mol Biol Evol. 2000 May;17(5):757–767. [PubMed]
- Tishkoff Sarah A, Williams Scott M. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet. 2002 Aug;3(8):611–621. [PubMed]