Semi-qualitive evaluation of variations in mutation rates for human evolution by phylogeny-based methods.

Abstract

The relative variations of mutation rates were evaluated by fitting small perturbations introduced to a non-parametric distance-based phylogenetic tree of human evolution. The proposed over-simplified approach allows qualitative comparison of mutation rates between populations and qualitative reconstruction of changes in mutation rate spectrum over time. The method was applied to eight populations, independently to autosomal chromosomes, to mtDNA and to Y chromosomes of the same individuals. The results are consistent with several independent measurments of mutation rate spectrum and provide a broader view to a dynamics of variable-rate human evolution.

Introduction

With the advances of molecular biology and sequencing technologies, the subjects of genealogy and the origins of modern humans attract extensive and deep attention; new experimental data and data processing methods could provide specific results about both genealogical relations and mankind origins. However, the volume of collected sequenced data is extremely large, and applying different bioinformatics methods to the same data sometimes could resulted in somewhat contradictorily statements concerning the question of human evolution. But anyway, several conclusions represent the consensus between various kinds of data interpretation, among which are the following:
(a) Evidences from both Y-chromosome data and Mt DNA data resulted in the statement that the most recent common ancestors of modern humans from both paternal and maternal line did exist in the per-historical time, not earlier than ca. 200 thousand years ago.
(b) Evidences from both paleontology and sequenced data resulted in the conception about extensive expansion of modern humans to Eurasia started ca. 50 thousands years ago.
(c) A wide diversity of human genotypes in modern African population, relatively to all other populations, and several fossil findings, suggests that the modern humans emerged in African continent, before ca. 200 thousand years ago.

The reconstruction of phylogenetic trees is, generaly speaking, the main approach used in the interpretation of sequenced data of this kind; it was first applied to Mt DNA data (Vigilant 1991) and to Y-chromosome data (Thomson 2000), to drawn the conclusions about female and male MRCA of humans. The Bayesian inference in phylogenetic analysis is, at present, the most powerful and precise tool to reconstruct phylogenetic trees attributed with the estimates of

divergence times for each node of tree. This methodology was applied, at least, to the interpretation of Y-chromosome data about male MRCA of humans (Thomson 2000), and to a lot of more specific studies about human evolution [Hey 2005, Elhassan 2014,]. The application of Bayesian inference to the interpretation of autosomal sequencing data require the extension of methodology, and the view to human origins using the analysis of autosomal sequences appeared more recently [Schiffels 2014]; these results allow to confirm and extend the consensus view to human origins mentioned above.

The most of the methods for computational phylogeny, including Bayesian inference, are strongly based on the conception of molecular clocks, and the estimates of divergence dates, for a case of human evolution, require the knowledge of mutation rate. The exact values of mutation rate could be the source of some controversy in the interpretation of the experimental data; as the example, the male and female MRCA of humans may live in different times, as it could be suggested by independent interpretations of Mt DNA and Y-chromosome data. Applying the same or similar methodology to the interpretation of several types of sequencing data could, however, resolve the logical inconsistence arose from the issue above [Poznik 2007]. Another example of somewhat controversal conclusions is a discussion about the precise time of human male MRCA [Mendez 2013, Elhaik 2014, Mendez 1014, Elhaik 2015] initiated by the evidence about the African Y-chromosome haplotype which is too distant from the most of human male haplotypes observed earlier. The related open questions in the recovering the story of human origins include the evidences about the presence of modern humans outside Africa before the well-documented period of human expansion [Hershkovitz 2018, Posth 2017 Pagani 2017], the speciation of early modern humans in African continent [Hublin 2017], certain controversy between the observed "founder effect" and the estimates of minimal population size for human groups [Hawks 2000].

For the other biological objects, the applications of phylogenetic tree reconstruction may require some extension of molecular clock concept, in order to adequately describe the biological data. Several methods which allow to expand the molecular clock models were proposed since the early development of phylogenetic analysis [Langley 1974, Sanderson 1997, Sanderson 2002]. For conventional implementations of Bayesian inference in phylogeny, the concept of relaxed molecular clocks is widely used [Drummond 2006], where the mutation rates assumed to obey some statistical distribution in the branches of phylogenetic tree. However the applicability of the approach is somewhat limited and may require external calibration, partly due to a computational complexity of the underlying probabilistic model [Wertheim 2010, Drummond 2010, Dornburg 2011].

The variations in mutation rate are also observed for human populations, using several methods independent from phylogenetic tree reconstruction, like the processing of pedigree relations [Henn 2008, Ge 2009, Narsimhan 2017, Haris 2017]. Also, the averaged estimates of mutation rates using these methods are normally differ from rates derived from calibrated phylogentic methods [Scally 2016, Balanovsky 2017]. The possible sources of the observed variations include

historical variations in the time interval between generations, presence of some specific mutations in enzymes responsible for DNA reparation pathways; but more important are the evidences about the presence of a complex machinery which connects phenotype differences (in a wide sense), epigenetic variations and mutations in the genome [Bell 2011, Jónsson 2017].

The effects caused by variations in mutation rate should be markedly observed on autosomal data, but the expansion of relaxed clock model to the interpretation of genomic data for diploid genomes within Bayesian inference is a non-trivial and still unresolved problem. And, to our best knowledge, the conventional models of relaxed clock models find no application in the analysis of paternal and maternal genetic data for humans, possibly due to the limitations mentioned above. So, due to the complexity of the exact reconstruction of the dated phylogenetic tree, the attempt to semi-qualitatively estimate variations in mutation rates for human evolutionary tree by tools of phylogenetic analysis would be meaningful.

Conventional phylogenetic analysis normally includes the choice of evolutionary model from several alternatives; some calibration of divergence dates is often introduced. Despite the vast number of successful studies where Bayesian inference was used to a reconstruction of phylogeny, some controversy in the interpretations may present depending on the choice of the methodology and evolutionary model. So, for a semi-qualitative estimates, the most simple phylogenetic tree reconstructions using distance matrix methods was intentionally choosen, avoiding most of the assumptions about specific evolutionary models. The implementations of Bayesian inference in phylogeny include sophisticated machinery of inter-dependent statistical models, and in this situation one could expect some hidden systematic biases introduced on any stage of data processing. So, the another point of the study is to provide interpretations of the data which could possibly detect the presence of systematic biases in a conventional results of human origins, normally obtained using complex evolution models and Bayesian inference. And, at last, the requirements of precision and rigor of the statistical approaches may restrict the flexibility and extent of the models, and it is possible to consider semi-qualitative estimates, to reduce the number of implicit and explicit assumptions accepted in precise approaches.

New Approaches

For a distance-based phylogenetic trees constructed with conventional methods, the tree distances were evaluated used linear regression, with Tikhonov regularization ('ridge regression') applied to avoid negative values for branch lengths. As a result, the terminal nodes of the rooted tree would have unequal height relatively to root. If one assumes, that leaf nodes of a tree in the reconstruction of human evolution should correspond to the same historical time, the distances of the tree calculated using linear regression introduce the uncertainty to the reconstruction of evolution. To reduce partially this uncertainty, on a second approximation the branch lengths were updated using recursive root-to-leafs algorithm. The update value for each branch was proportional to the length of the branch and to average discrepancy for two adjacent clades of the tree. The assumption about unequal mutation rates was implicitly introduced on this stage. The topology of the tree was reconstructed using conventional distance-based methods.

The root of the reconstructed tree was chosen by enumeration of tree branches, using the average

discrepancy of terminal node heights as a criteria. The comparison of branch lengths for terminal nodes before and after the recursive update of node heights is applied, this allows to calculate relative change of mutation rate for each branch of the tree.

The method was applied to the subset of individuals from several ethnic groups available from '1000 genomes' project, almost repeating the subset of individuals used in [Schiffels 2014] where the whole autosomal sequenced data on these genomes was used to drawn the conclusions of human origins. For these individuals, Mt DNA and Y-chromosome data (when appropriate) is also available from 1000 genomes project, thus allowing to reconstruct the three independent evolutionary trees. X-chromosome data on these subjects should obviously be treated using an enhanced evolutionary model within Bayesian inference; this data was also processed in the study, in order to expand the scope of expected qualitative results.

The approach could not provide reliable estimates of divergence times or exact topology of relations between selected individuals. However, the distribution of relative changes in mutation rates across the constructed trees allows to present the dependencies of rate variations from populations and across the time scale, comparable with published data from other sources. Expanded this way view to a human evolution could suggest several opinions concerned the unresolved questions in this important subject.

Results

1. The evaluation of tree topologies and distributions of distances in a results of a pipeline.

The comparive presentations of estimated times of human MRCA are shown on fig 1, for all four datasets. Three alternative tree topologies were compared (UPGMA, neighbour joining (NJ) and iterative taxon addition). Error bars indicate relative variations of mutation rates obtained by several methods; for MtDNA, the interval was based on the estimates from 1.26e-8 mutations per year [Herrnstadt 2002] to 2.38e-8 mutations per year [Poznik 2007]; similarly, for Y chromosome values are ranged from 6.12e-10 per year [Mendez 2013] to 1.24 e-9 per year [Thomson 2000], and, for autosomal data, from 1.25e-8 per generation [Schiffels 2014] to 2.5e-8 per generation [Nachman 2000]. A geometric mean of these values was used to draw tree heights on fig. 1.

The most conventional and straightforward measures were used to calculate distance matrices for the analysis.The relative number of substitutions was used as a measure for non-recombinant types of data (X and Mt chromosomes). For diploid autosomal chromosomes the simple measure was used, similar to Genpofad measure [], so that the distance from heterozygotic allele to homozygotic was taken as half of the distance between two homozygotic alleles. After that, the distances were calibrated using the direct processing of sequence data available from 1000 genomes site, for several selected individuals from the whole set. In this way, the values of distances become in an adjustment with mutation rates specified above.

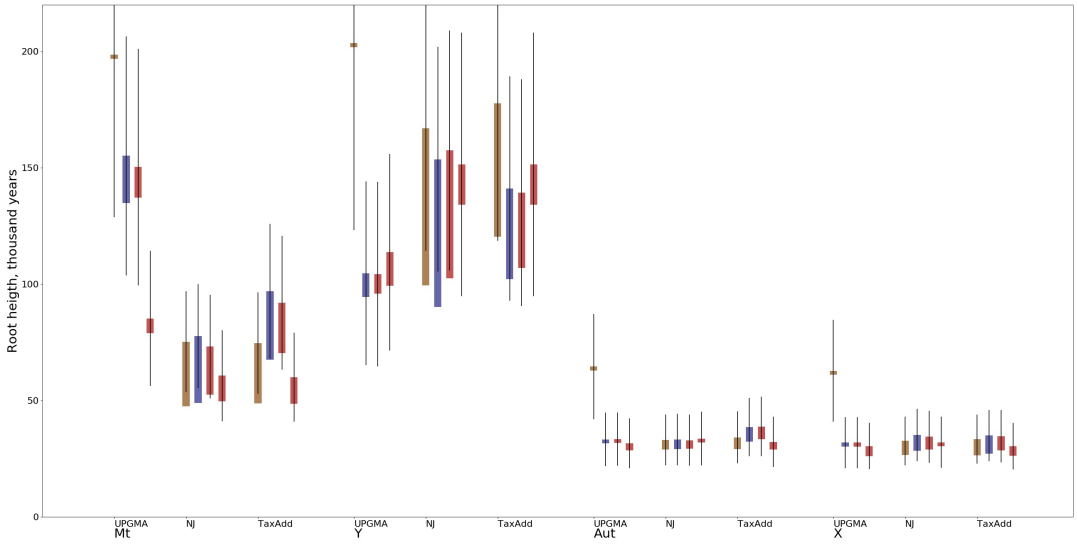

Fig. 1
The comparison of heights of the trees obtained by several distance-based methods for four kinds of data (Mitochondrial genome, Y chromosome, autosomal SNP and X-chromosome SNP). Y axis - time of estimated common origin (tree root). For each type of data, the results for three topologies are show: UPGMA, Neighbour-joining and TaxAdd (iterative taxon addition). For each topology, results for four steps of pipeline are shown: initial tree, tree after distance updates using linear regression, tree after appying the iterative correction of branch distances; tree with optimal position of root (after applying 3 previous steps). The colored bars indicate the uncertainty arose from the unequal positions of leaf nodes. The error bars indicate the interval of published values for mutation rates.

Four bars on fig. 1 for each kind of topology represent, correspondingly, the original tree distances, the distances after applying a linear regression rescaling, the distances after applying the iterative update of branch lengths, and the final result obtained by selection of best branch to be used as a tree root. As it could be seen from fig. 1, the choice of the tree root could be crucial for the estimates of the root height. Also, difference for the heights of the trees shown on fig 1 is clearly seen between trees for paternal and maternal lines and trees for diploid chromosomes. The systematic biases of the approach and possible hypotheses about real world trends in evolution which could partially explain observed differences are discussed in the next sections.

The phylogenetic trees itself, for the UPGMA topology after selection of root branch, are presented on fig. 2, for all four datasets. The ethnic groups which are compared in the study are labeled with color notations. Only approx. half of individuals are presented on the Y-chromosome tree, because the initial subset included samples for both men and women.

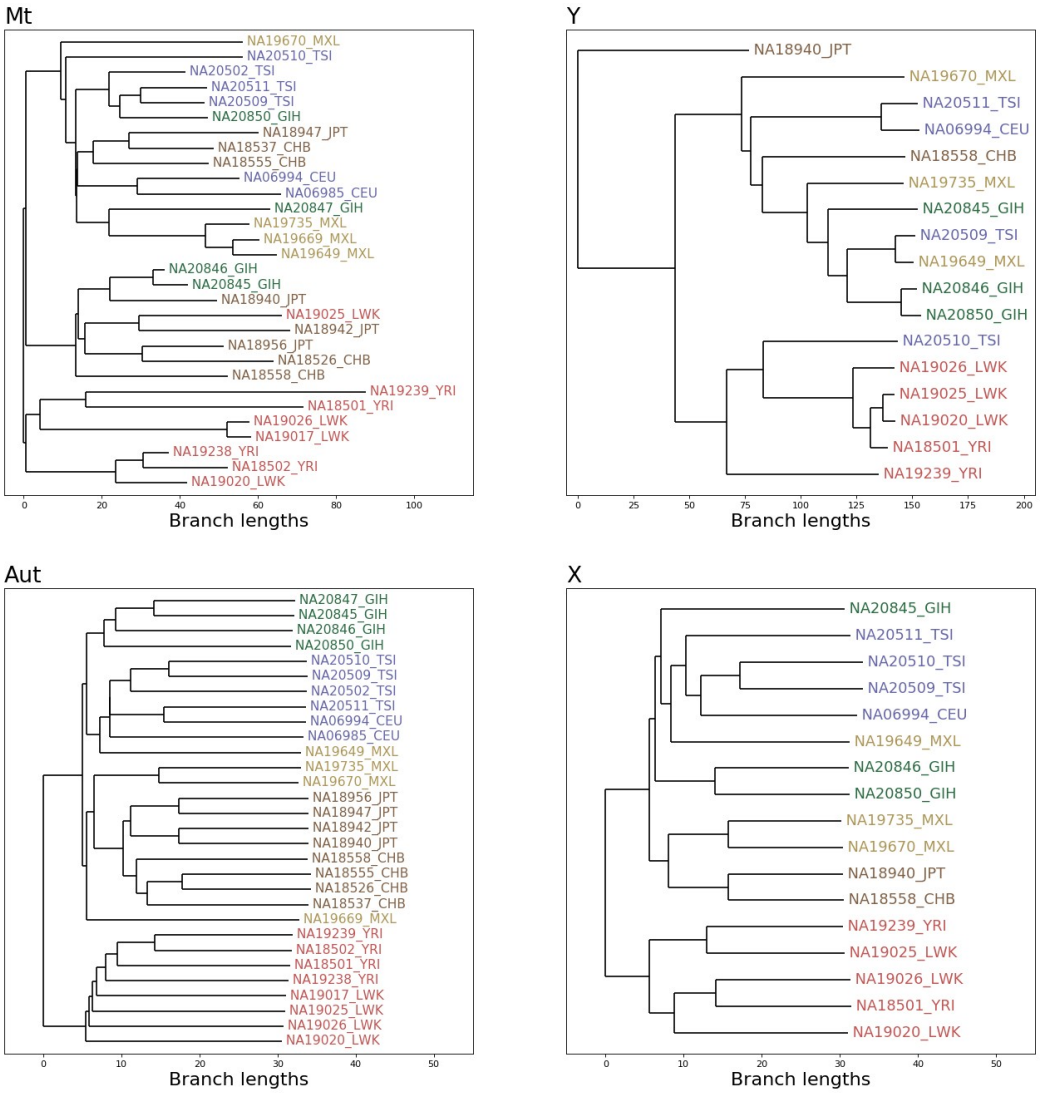

Fig. 2
The phylogenetic trees constructed for selected individuals from "1000 genomes" project using four kinds of genomic data. The identifiers of leafs are composed from the identifier of sample and identifier of ethnic group. The color notations indicate population group for the individual: Green - South Asia, Blue - Europe, Beige - Native Americans, Brown - East Asia, Red - Africa.

The relative lengths of the branches for autosomal data on fig. 2 are obviously inadequate; the distances between individuals from the same population do not relect the expected proximity of relations. This observation demonstrates that the distance-based methods couldn't correctly describe the relations in diploid genome, and advanced models like the approach proposed in [Schiffels 2014] are much more

adequate in the reconstruction of the precise dated phylogenetic tree. However the individuals are grouped almost correctly by their ethnicity for autosomal data on fig. 2, as all the samples in the original study [Schiffels 2014] wee carefully selected to be the most representative for the corresponding populations. However, comparable tree topologies for Y and Mt chromosomes, and even for X chromosome, for the same individuals and with the use of similar methodology, demonstrate the traces of complexity in a genealogy of each individual, and a degree of uncertainty in the precise definition of ethnic groups. The unexpectedly short branches around the tree root and origins of all ethnic groups for MtDNA and Y chromosome data could be partly explained by presence of the same systematic bias in the methodology, as it was observed for autosomal data. Another aspects of the obtained distribution of ethnic groups are considered in the Discussion section.

2. Estimated variations of mutation rates between populations are comparable by value and by direction of change with results obtained by other approaches.

The distribution of mutation rates for all nucleotide triplets was systematically studied in [Harris 2017] for individual genomes from “1000 genomes” project. The data provided in this study allows a comparison with estimates of a relative variations in mutation rate between populations, calculated from distance-based phylogenetic trees obtained in this study as described above. The resulted variations are presented on fig. 3, for both approaches.

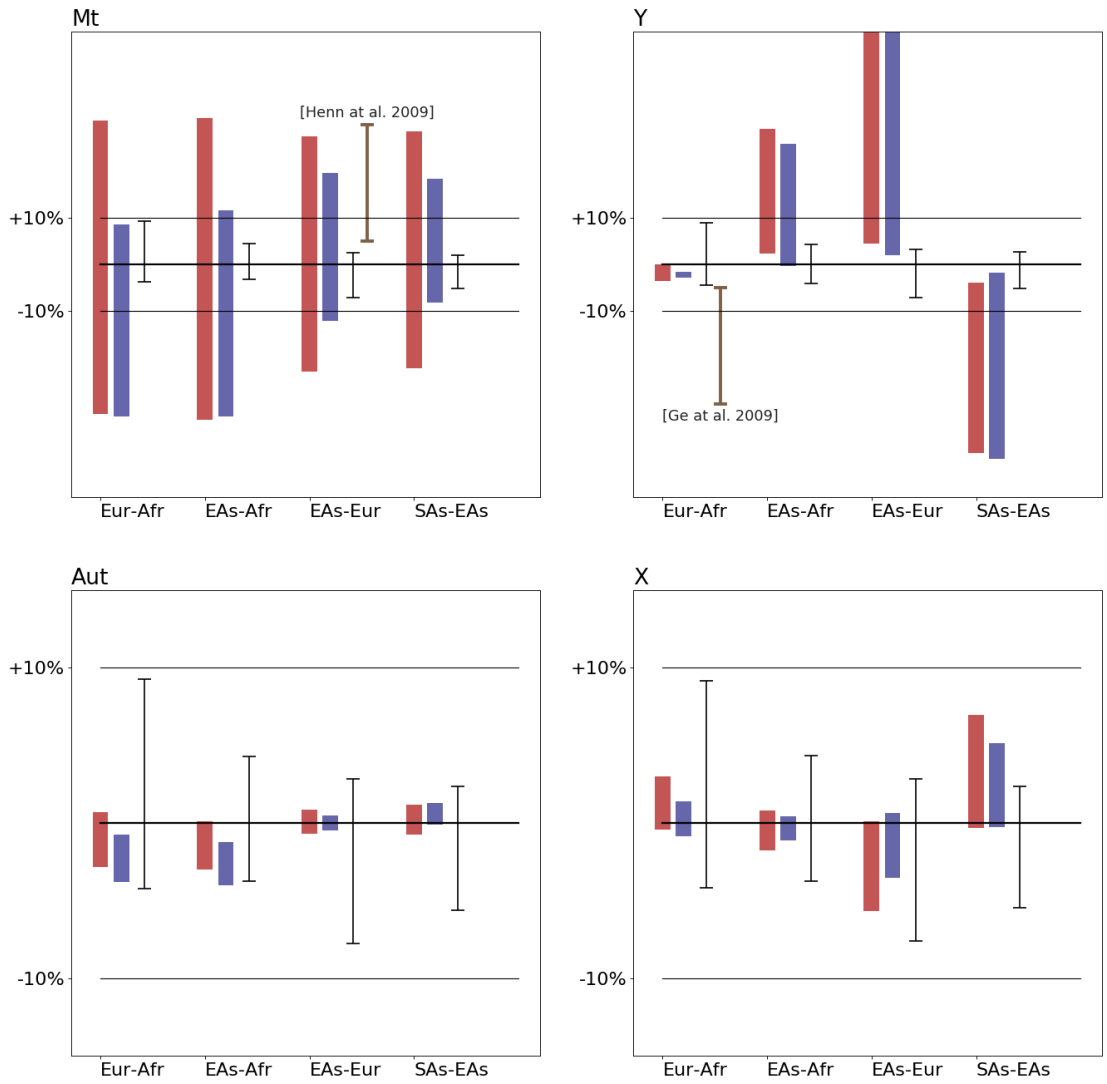

Fig. 3
Variations of mutation rates between populations, estimated as described in the main text, in a comparison with published values of variations for several groups.

Red bars - variations estimated using a relation between branches adjacent to terminal nodes. Blue bars - variations estimated using a relation between distances from root to terminal nodes. Thin error bars - variations derived from the results presented at [Harris]. Thick brown error bars - variations evaluated in the two another studies.Four pairs of population groups are presented, in the order: Europe versus Africa, East Asia versus Africa, East Asia versus Europe, South Asia versus East Asia.

In addition, genealogy-based variations of mutation rate between Japanese and European MtDNA haplogroups [Henn 2009] and between African Americans and Caucasians in Texas population for Y

chromosome [Ge 2009] are presented on fig. 3 for a qualitative comparison.

As the relative changes of mutation rates are estimated for each branch of phylogenetic tree, there are two possible straightforward ways to reduce these values to the average variations between populations. In a first method, only the terminal branches were considered to calculate average. In a second method, distances to tree root from each terminal node were used for averaging. The values calculated by the second method represent relative variations in mutation rate, more distant in time. Both estimates are shown on fig. 3. In a presentation of results for study [Harris 2017], the bars on fig. 3 show the significance intervals for rate variation; the median values for variation are often have opposite directions in a comparison of two approaches, but the limits of significance interval in the latter approach allows to conclude about the consistency of the proposed approach relatively to results of a study used for comparison. And the directions of variations from two other independent studies are also in agreement with the directions of results obtained by the proposed method.

The results for four types of data are often in disagreement with each other; the directions of change for autosomal data and MtDNA are consistent with each other, but are sometimes opposite to the directions for Y-dna data. But, as it was observed for historical variations of mutation rate for these three types of data, the autosomal and mithochondrial mutation rates have a trend to decelerate ([Henn 2009, Scally 2017]), and Y-chromosome mutation rate has a trend to accelerate [Balanovsky 2017], in a qualitative agreement with the discrepancy described above.

Discussion

The results of the presented study demonstrate the ability of the proposed simplified approach of distance based phylogenetic analysis to qualitatively describe the variations of mutation rates for a real-word genomic data, in the application to the human evolution. One of the particular consequences of the well-proven observation about the presence of variation in mutation rates for different human populations, is that a degree of heterogeneity within some population could point not only to a distance in time to the origins of the population, but also to a relative increase of mutation rate for this population. The phylogenetic relations between human races, reconstructed using distance-based methods and presented on fig. 2, look much more different than trees derived from likelihood-based methods and normally used as an additional proof of the out-of-Arica model. Part of this difference could arise from the biases introduced by simplified distance-based approach. But the assumption, that a mutation rate for African population in the pre-historical times was significantly higher than for the human populations expanded to Eurasia, could partly explain the high variability of genotypes observed in the present-days African population, which is conventionally used as an argument to support a long-time development of humans within Africa before the expansion. Of course, the proposed simplified approach could not be used to exactly reconstruct the evolution of humans in the pre-historic times and confirm or disprove the out-of-Africa model and other possible hypotheses about human origins. But the presented results could point to certain logical incompleteness of the conventional hypothesis about origins of humans, in the light of the expanded view to the process of evolution.

Also the question of presence or absence of bottlenecks in the population size should be discussed in this expanded view. The presence of founder effect is observed in several studies for human ethnic groups, being in some controversy with the estimates of minimal population size. But, as the presence of bottlenecks in population size is normally detected with implicit or explicit assumption about neutrality of mutation process in the genomes or genes used for the analysis, this question could be revisited in the view of possible high variations in mutation rate. A number of founders for the

presented evolutionary trees could be semi-qualitatively estimated, as it is described in methods section. The idea of the estimator is to use the distribution of heterozygous alleles evaluated from the reconstructed tree, and to compare it with distributions obtained from coalescent simulations with several different sizes of founding population. The detected increase in relative heterozygocity, in both real-world distribution and in simulations where a number of founders was less than 500, give a clue that the absence of the bottleneck in the evolution of early humans also could be doubted.

The assumption of the alternative scenario about presence of the bottleneck in the times comparable with the beginning of human expansion to Eurasia have long-coming consequences. As one of these consequences, the marriages between close relatives should assumed to be a normal practice, and this imply that the number of deleterious genes in the genomes of early humans should be much lower than in present-day human genomes. Another consequence is that this alternative scenario is much closer to the proverbs from Bible than most of the scenarios in the recent studies of human evolution, and the names 'Adam' and "Eva' for genetic ancestors of modern humans could mean more than just an curious allusions with ancient scriptues. However the qualitative nature of the results allows to keep the acceptable degree of uncertainy in the consequences mentioned above, and the proposed suggestions should not be compared with the theories which in past opposed Darwinian evolution theory. The degeneration theory, which was resulted in many negative and even tragic consequences, could be recalled as the example of such a theories, as it is described in historical review [Pick 1993]. However, many of the inconsistencies and controversies accumulated in the present-day studies of human evolution, in particular by the tools of bioinformatics, arose due to implicit introduction of several oversimplified models beyond the complex software and statistical machinery. This kind of problems could be classified as a case of reductionist approach, as well as the approaches which in past were opposed by Darwinian evolution theory, and are anyway in explicit or hidden disagreement with somewhat paradoxal meaning of Holy Scriptures. Without any intention to contribute to a philosophical or sociological context of the subject under study, some of the qualitative results obtained by means of bioinformatics and presented above allows to expand the view to the human evolution and possibly point to the directions which could resolve some of the questions arose in this area of science.

Methods

The SNP data on 1000 genomes project available on EBI ftp site was used to construct distance matrices. For Mt DNA and Y chromosomes, all SNP presented in the repository were used. For autosomal chromosomes and X chromosomes, SNPs for the analysis were filtered based on a subset of SNP used in FamilyTreeDna genotyping data, in order to limit the analysis to most important and statistically valuable polymorphisms.

Distance matrices for Mt DNA and Y chromosome were constructed using counts of mismatches in SNP for pair of individuals, similarly to Jukes-Cantor model of site substitutions. For autosomal data, distances were calculated based on the GENPOFAD distance for polymorphic sites [Joly 2015]; for a case of diploid alleles as it is presented on available 1000 genomes SNP data, this measure is reduced to a simple 4x4 matrix. Only samples included to Y-chromosome subset were used to construct trees from X chromosome data, so a binary measure was applied to these SNPs as well as to Mt and Y SNPs.

The topology of phylogenetic trees presented in the results section was generated using UPGMA method. software []. As the alternative, neighbour-joining method and

TaxAdd_Balme (iterative taxon addition), as it is implemented in FASTME software [], were tested for a comparison of tree topologies.

Samples NA06994 (Europe), NA18558 (China), NA19020 (Africa) and NA20846 (India) were used for the calibration of distances. To calculate the reference genetic distances between samples, preprocessed sequencing reads available from 1000 genomes project site were aligned to human genome (version 37) using bowtie2; the alignments were processed and compared using a conventional samtools/vcftools pipeline, for all chromosomes and mitochondrial genome.

To characterize the distribution of polymorphic sites in the reconstucted tree with a single parameter, the idea was used as described below. Let the population is founded by a limited number of ancestors, and each of them had a certain proportion of polymorphic sites in the genome. In this case, in a certain number of next generations, some of the sites which were polymorpic in the genomes of founders, will be polymorhic in the new generations. But the frequences of polymorphic alleles in these sites will be not random, but will correlate with each other and reflect the history of a kind. Accortding to this idea, an average similarity between distribution of allele frequences in polymorphic sites was evaluated from a phylogenetc tree and a composition of polymorpic sites on terminal nodes on the tree. For non-terminal nodes, the frequences of alleles were calculated from the composition of alleles on descendant nodes. To be more precise, the frequences of alleles for each site could be presented as a vector of dimension 3, and a euclidean distance between these vectors was used as a measure, for any pair of sites. 1000 randomly selected pairs were used to calculated the average degree of independency. The distribution of the average similarity parameter on the nodes of the tree, as a function of distance to the root, is shown of fig. 4, for human autosomal sequences processed as described above, and for simulated data. By the construction, the lower values of averaged distance coorespond to lesser number of anccestral genotypes, so the presence of slope in the regression lines on fig. 4 both for simulations with small number of founders and for human genomes could be explained by the presence of the botleneck in the evolution of early humans.

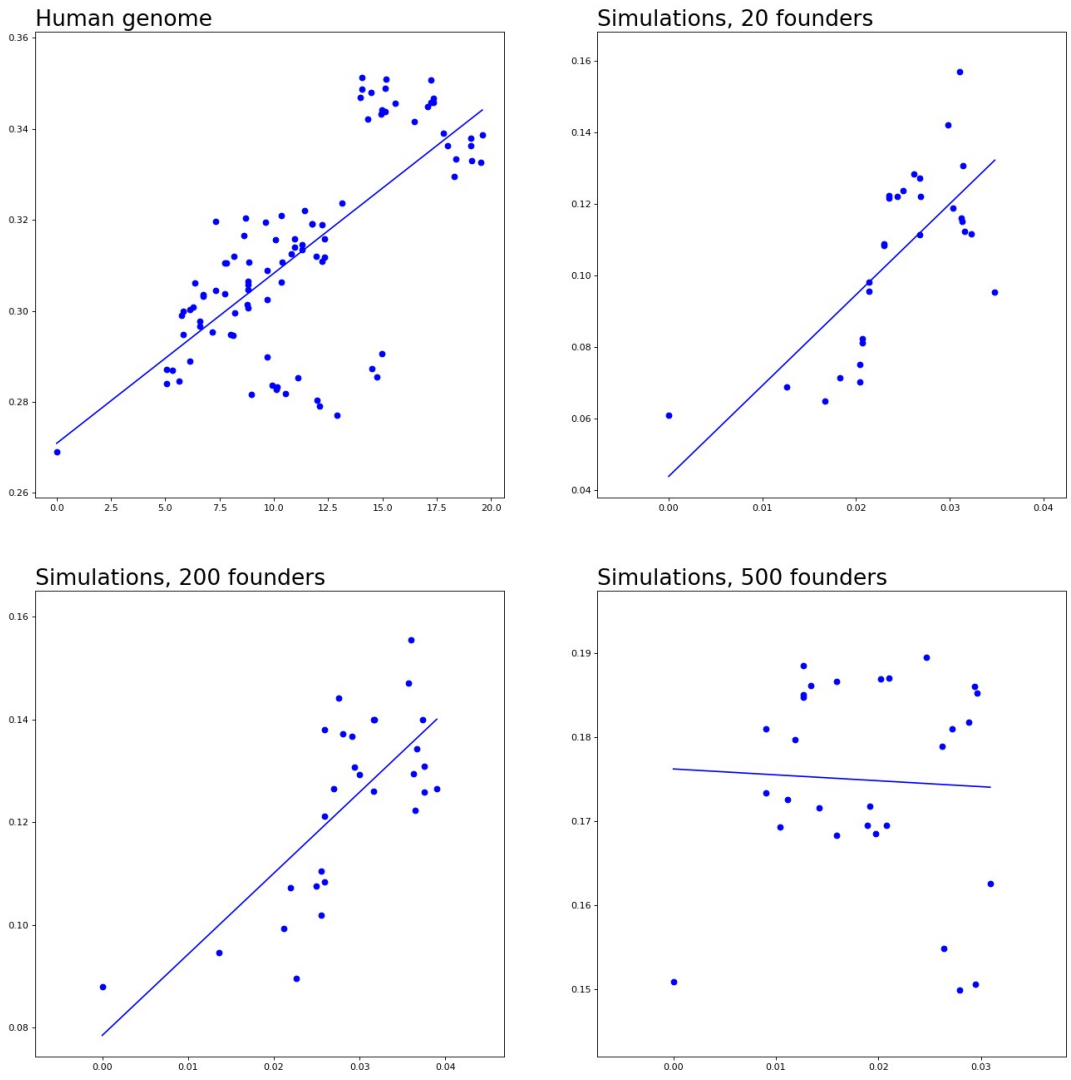

Fig. 4 The degree of independency in the distribution of heterozygotic alleles, for human autosomal sequences and simulated data. Three kinds of simulated data present different sizes of initial population size, as it described in ttiles of the sections. On all four sections, points correspond to nodes of the tree; the time from tree root is shown on X axis and estimated degree of heterozygocity on the node is shown on Y axis. The lines on the charts present a slope of trivial linear regression.

Cosi2 coalescent simulator [Shlyakhter 2014] was used to generate the sequences used on fig. 3 for a comparison with real-world data. The sequences were generated assuming the model of exponential growth for 500 generations and 50000 as a size of presnt-day population; 1 mln as a length of the genome and 60 as sample size were assumed as the parameters.

Software on C++ and python used to run the analysis is deposited in github.com (sferanchuk/h_evol),

together with data tables and shell scripts sufficient to reproduce the results.
Lapack library for linear algebra was used to implement regression algorithm;
Phylo and Matplotlib python packages were used to run the analysis and prepare the figures.

References

Cosi2: an efficient simulator of exact and approximate coalescent with selection", by Ilya Shlyakhter Pardis C. Sabeti and Stephen F. Schaffner, Bioinformatics (2014) 30 (23): 3427-3429. doi: 10.1093/bioinformatics/btu562

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson A C. Science. 1991;253:1503–1507.

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(13):7360-7365.

Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R. A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. Am J Hum Genet. 2011;88:814–818

Inferring human population size and separation history from multiple genome sequences.
Schiffels S, Durbin R.
Nat Genet. 2014 Aug;46(8):919-25. doi: 10.1038/ng.3015

Mendez FL, Krahn T, Schrack B, et al. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. Am J Hum Genet. 2013;92:454–459.

Elhaik E, Tatarinova TV, Klyosov AA, Graur D. The ‘extremely ancient' chromosome that isn't: a forensic bioinformatic investigation of Albert Perry's X-degenerate portion of the Y chromosome. Eur J Hum Genet. 2014;22:1111–1116

Mendez FL, Veeramah KR, Thomas MG, Karafet TM, Hammer MF. Reply to “The ‘extremely ancient' chromosome that isn't” by Elhaik et al. Am J Hum Genet. 2014

Elhaik E, Tatarinova TV, Klyosov AA, Graur D. Reply to Mendez et al: the “extremely ancient” chromosome that still isn’t. European Journal of Human Genetics. 2015;23(5):567-568. doi:10.1038/ejhg.2014.227.

Science. 2013 Aug 2;341(6145):562-5. doi: 10.1126/science.1237619.
Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females.
Poznik GD1, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, Bustamante CD.

Shang H, Tong H, Zhang S, Chen F, Trinkaus E. An early modern human from Tianyuan Cave, Zhoukoudian, China. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(16):6573-6578. doi:10.1073/pnas.0702169104.

Nature. 2011 Nov 2;479(7374):521-4. doi: 10.1038/nature10484.
The earliest evidence for anatomically modern humans in northwestern Europe.

Higham T1, Compton T, Stringer C, Jacobi R, Shapiro B, Trinkaus E, Chandler B, Gröning F, Collins C, Hillson S, O'Higgins P, FitzGerald C, Fagan M.

J Hum Evol. 2007 Mar;52(3):243-61. Epub 2006 Oct 1.
The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo).

Barker G1, Barton H, Bird M, Daly P, Datan I, Dykes A, Farr L, Gilbertson D, Harrisson B, Hunt C, Higham T, Kealhofer L, Krigbaum J, Lewis H, McLaren S, Paz V, Pike A, Piper P, Pyatt B, Rabett R, Reynolds T, Rose J, Rushworth G, Stephens M, Stringer C, Thompson J, Turney C.

Rapid evolution of the human mutation spectrum.
Harris K, Pritchard JK.
Elife. 2017 Apr 25;6 0.84-1.15

Nat Commun. 2017 Aug 21;8(1):303. doi: 10.1038/s41467-017-00323-y.
Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes.

Narasimhan VM1, Rahbari R2, Scally A3, Wuster A2,4, Mason D5, Xue Y2, Wright J5, Trembath RC6, Maher ER7,8, van Heel DA9, Auton A10, Hurles ME2, Tyler-Smith C2, Durbin R2.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, et al. (2008)
Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82: 583–592.

Mol Biol Evol. 2009 Jan;26(1):217-30. doi: 10.1093/molbev/msn244. Epub 2008 Nov 4.
Characterizing the time dependency of human mitochondrial DNA mutation rate estimates.

Henn BM1, Gignoux CR, Feldman MW, Mountain JL.

Henn, B.M., L.L. Cavalli-Sforza, and M.W. Feldman, The great human expansion.
Proceedings of the National Academy of Sciences of the United States of America, 2012.
109(44): p. 17758-17764.

Methods in Ecology and Evolution 2015, 6, 938–948
doi: 10.1111/2041-210X.12343
Flexible methods for estimating genetic distances from single nucleotide polymorphisms
Simon Joly1*, David Bryant 2 and Peter J. Lockhart 3

Pick, Daniel (1993). Faces of degeneration : a Europeam disorder, c. 1848-c. 1918 (1st pbk. ed.). Cambridge: Cambridge University Press. ISBN 052145753X
Pick, Daniel (1993) Faces of degeneration: a European disorder, c.1848–1918. Cambridge, UK: Cambridge University Press. ISBN 9780521457538

DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines

Jordana T BellEmail author, Athma A Pai, Joseph K Pickrell, Daniel J Gaffney, Roger Pique-Regi, Jacob F Degner, Yoav GiladEmail author and Jonathan K PritchardEmail author
Genome Biology201112:R10
https://doi.org/10.1186/gb-2011-12-1-r10© Bell et al; licensee BioMed Central Ltd. 2011
Received: 3 October 2010Accepted: 20 January 2011Published: 20 January 2011

Shendure, J. & Akey, J. M. The origins, determinants, and consequences of human mutations. Science 349, 1478–1483 (2015).

Science. 2015 Sep 25;349(6255):1478-83. doi: 10.1126/science.aaa9119. Epub 2015 Sep 24.

Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.

Journal of Molecular Evolution

September 1974, Volume 3, Issue 3, pp 161–177 | Cite as

An examination of the constancy of the rate of molecular evolution
Authors
Authors and affiliations
Charles H. LangleyWalter M. Fitch

PLoS Biol. 2006 May;4(5):e88. Epub 2006 Mar 14.
Relaxed phylogenetics and dating with confidence.

Drummond AJ1, Ho SY, Phillips MJ, Rambaut A.

Wertheim JO, Sanderson MJ, Worobey M, Bjork A. Relaxed Molecular Clocks, the Bias–Variance Trade-off, and the Quality of Phylogenetic Inference. Systematic Biology. 2010;59(1):1-8. doi:10.1093/sysbio/syp072.

Bayesian random local clocks, or one rate to rule them all
Alexei J DrummondEmail author and Marc A SuchardEmail author
BMC Biology20108:114
https://doi.org/10.1186/1741-7007-8-114© Drummond and Suchard; licensee BioMed Central Ltd. 2010

Relaxed Clocks and Inferences of Heterogeneous Patterns of Nucleotide Substitution and Divergence Time Estimates across Whales and Dolphins (Mammalia: Cetacea)
Alex Dornburg,*,1 Matthew C. Brandley,#,1 Michael R. McGowen,,2 and Thomas J. Near1,3

Nachman, M.W. and S.L. Crowell, Estimate of the mutation rate per nucleotide in humans. Genetics, 2000. 156(1): p. 297-304.

Sanderson M.J. 2002. Estimating absolute rates of molecular evolution
and divergence times: a penalized likelihood approach. Mol. Biol.
Evol. 19:101–109.

Current Biology 23, 553–559, April 8, 2013 ª2013 Elsevier Ltd All rights reserved

A Revised Timescale for Human Evolution
Based on Ancient Mitochondrial Genomes
Qiaomei Fu,1,3,19 Alissa Mittnik,2,19 Philip L.F. Johnson,4
Kirsten Bos,2,5 Martina Lari,6 Ruth Bollongino,7
Chengkai Sun,8 Liane Giemsch,9,10 Ralf Schmitz,9
Joachim Burger,7 Anna Maria Ronchitelli,11 Fabio Martini,12
Renata G. Cremonesi,13 Ji#rı´ Svoboda,14,15 Peter Bauer,16
David Caramelli,6 Sergi Castellano,1 David Reich,17,18
Svante Pa¨a¨bo,1 and Johannes Krause2,

Nature. 2003 Jun 12;423(6941):742-7.
Pleistocene Homo sapiens from Middle Awash, Ethiopia.
White TD1, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC.

Nature. 2017 Jun 7;546(7657):289-292. doi: 10.1038/nature22336.
New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens.
Hublin JJ1,2, Ben-Ncer A3, Bailey SE4, Freidline SE1, Neubauer S1, Skinner MM5, Bergmann I1, Le Cabec A1, Benazzi S6, Harvati K7, Gunz P1.

Nature. 2017 Jun 7;546(7657):293-296. doi: 10.1038/nature22335.
The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age.
Richter D1,2,3, Grün R4,5, Joannes-Boyau R4,6, Steele TE1,7, Amani F8, Rué M9,10, Fernandes P9,11, Raynal JP1,11, Geraads D1,12, Ben-Ncer A8, Hublin JJ1,13, McPherron SP1.

Science. 2018 Jan 26;359(6374):456-459. doi: 10.1126/science.aap8369.
The earliest modern humans outside Africa.
Hershkovitz I1,2, Weber GW3, Quam R4,5,6, Duval M7,8, Grün R7,9, Kinsley L9, Ayalon A10, Bar-Matthews M10, Valladas H11, Mercier N12, Arsuaga JL5,13, Martinón-Torres M8,14, Bermúdez de Castro JM8,14, Fornai C3,15, Martín-Francés L8,16, Sarig R2,17, May H18,2, Krenn VA3,15, Slon V18, Rodríguez L5,19,20, García R5,19, Lorenzo C21,22, Carretero JM5,19, Frumkin A23, Shahack-Gross R24, Bar-Yosef Mayer DE25,26, Cui Y27, Wu X27, Peled N28, Groman-Yaroslavski I29, Weissbrod L29, Yeshurun R29, Tsatskin A29, Zaidner Y29,30, Weinstein-Evron M29.

Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals
Cosimo Posth,a,1,2 Christoph Wißing,3 Keiko Kitagawa,1,4 Luca Pagani,5,6 Laura van Holstein,7 Fernando Racimo,8 Kurt Wehrberger,9 Nicholas J. Conard,1,10 Claus Joachim Kind,11 Hervé Bocherens,3,12 and Johannes Krauseb,1,2

Philos Trans R Soc Lond B Biol Sci. 2016 Jul 19; 371(1699): 20150137.
doi: 10.1098/rstb.2015.0137
PMCID: PMC4920338

PMID: 27325834
Mutation rates and the evolution of germline structure
Aylwyn Scally

PLoS Biol. 2005 Jun; 3(6): e193.
Published online 2005 May 24. doi: 10.1371/journal.pbio.0030193
PMCID: PMC1131883
PMID: 15898833
On the Number of New World Founders: A Population Genetic Portrait of the Peopling of the Americas
Jody Hey 1
Andy G. Clark, Academic Editor

The Episode of Genetic Drift Defining the Migration of Humans out of Africa Is Derived from a Large East African Population Size
Nuha Elhassan,# 1 Eyoab Iyasu Gebremeskel,# 1 , 2 Mohamed Ali Elnour, 1 Dan Isabirye, 3 John Okello, 3 Ayman Hussien, 1 Dominic Kwiatksowski, 4 Jibril Hirbo, 5 Sara Tishkoff, 5 and Muntaser E. Ibrahim 1 , *

Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach
Michael J. Sanderson
Molecular Biology and Evolution, Volume 19, Issue 1, 1 January 2002, Pages 101–109,
https://doi.org/10.1093/oxfordjournals.molbev.a003974
Published: 01 January 2002 Article history

Journal of Anthropological Sciences
the JASs is published by the Istituto Italiano di Antropologia www.isita-org.com
Vol. 95 (2017), pp. 1-5
JASs Research news JASs Research news
A Cover Story for a Nature cover: genetic signature
of human expansions into Eurasia revealed
by a panel of worldwide high coverage genomes
Luca Pagani

Nature. 2017 Sep 28;549(7673):519-522. doi: 10.1038/nature24018. Epub 2017 Sep 20.
Parental influence on human germline de novo mutations in 1,548 trios from Iceland.
Jónsson H1, Sulem P1, Kehr B1, Kristmundsdottir S1, Zink F1, Hjartarson E1, Hardarson MT1, Hjorleifsson KE1, Eggertsson HP1, Gudjonsson SA1, Ward LD1, Arnadottir GA1, Helgason EA1, Helgason H1, Gylfason A1, Jonasdottir A1, Jonasdottir A1, Rafnar T1, Frigge M1, Stacey SN1, Th Magnusson O1, Thorsteinsdottir U1,2, Masson G1, Kong A1,3, Halldorsson BV1,4, Helgason A1,5, Gudbjartsson DF1,3, Stefansson K1,2.

Hum Genet. 2017 May;136(5):575-590. doi: 10.1007/s00439-017-1805-8. Epub 2017 Apr 28.
Toward a consensus on SNP and STR mutation rates on the human Y-chromosome.
Balanovsky O1,2.