What do our genes tell us about our past?

Rob DeSalle

The Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024
e-mail: desalle@amnh.org

Summary - *The use of DNA sequences to elucidate the history of relationships of organisms is widespread, and focus on our species has been intense. This paper examines some simple aspects of using genetic information to analyze relationships within and amongst humans. Clonal markers (mtDNA and Y chromosomal DNA) have always shown a high degree of structure and robustness when analyzed for hierarchical structure. Results from genome wide phylogenetic structure in many organismal systems suggests instead that recombining genetic elements like the X chromosome and the autosomes will give conflicting information from genome region to genome region. In addition, the evolutionary signal from the different chromosomal regions will show a high degree of incongruence with each other, as do adjacent regions of chromosomes. This incongruence and lack of hierarchical structure is discussed in the context of what we know about human populations and the theoretical underpinnings of tree building based analysis of human populations.*

Keywords - *Genomics, Phylogeny, Incongruence, Neanderthal, Coalescence.*

A brief history of treelike human history

The response to the question that titles this paper has many answers, all of them dependent on the definition of history. Human history can be defined at many different levels and in many different contexts. For instance, certainly, genes can tell us a lot about the place of humans in the animal tree of life and genes have also gone a long way to resolving the Human-Chimp-Gorilla trichotomy problem (Ruvolo *et al.*, 1991; Mailund, 2014). However, answering the question posed in the title of this paper over the broad range of definitions of history would be a huge task, hence I focus in this paper on one aspect of human history that has pervaded the study of human genetics and human origins since Darwin. This aspect concerns a discussion of how we study and perceive relationships of individuals or populations of *Homo sapiens*. I also confine the discussion in this paper to the area of understanding the relationships of the people living on this planet today – a single miniscule slice through time, but nonetheless an important one. Since Darwin's wonderfully poetic description of the "great tree of life" in On the Origin, naturalists and biologists have focused deeply on tree thinking as a way to describe relationships of entities and indeed as a way to depict history. Hierarchy then becomes an important aspect of any attempt to understand our history, but is it really relevant to our understanding of the historical divergence within our own species.

The well known attempt by Aristotle to make sense of hierarchy in the context of humans led to his "scala naturae". One only need to look at the sly dismemberment of *scala naturae* on the cover of Olivier Rieppel's book Fundamentals of Comparative Biology to see why Aristotle's way of organizing *Homo sapiens* relationship with nature was at best shallow and at worst misleading. The book cover shows the scala naturae arrangement of plants, animals and man in a linear fashion (plants→animals→man). Hierarchy is evident in the scheme because of the arrows,

but bifurcation is not a part of Aristotle's way of organizing things. That is, until the characters that can be used to describe these three groups of organisms are considered. Here Rieppel turns Aristotle upside down and uses the "vegetative soul", the "sensitive soul" and "the rational soul" as characters to change how the history of these organisms can be viewed in a bifurcating diagram. This way of constructing and interpreting trees has become a standard way of thinking about the natural world because of the richness with which such thinking allows hypothesis testing and because of their clear explanatory power.

The scholarly website "The genealogical world of phylogenetic networks" (http://phylonetworks.blogspot.com/2013/09/public-availability-of-phylogenetic-data.html), chronicles the development of tree thinking and tree construction. The authors of this website point out that trees have been used since the mid 1700's to illustrate relationships of humans. A good number of these have used characters from the study of language to accomplish the tree building. Morris (2012) suggests that the earliest real evolutionary branching diagram of humans occurred in a publication by Kieth (1915), that shows clear branching of four major "lineages" of humans – African, Australian, Mongoloid and Caucasian. Curiously, the figure shows no resolution with respect to the ordering of these so-called lineages by leaving them in an unresolved polychotomy. One branching diagram of particular interest not mentioned by the authors of this website is John Sparks' Histogram of Evolution (www.flickr.com/photos/13964815@N00/3307665413/sizes/o/). Sparks attempted to nudge the entirety of evolutionary history of the planet Earth into a single histomap. The bottom one third of this now rather popular poster depicts human evolution as anthropology saw it in the 1930's. While the diagram is clearly Euro-centric, Sparks gives it a bifurcating pattern and indeed the topology of his hierarchical diagram reflects fairly closely the understanding of human relationships at the time when a bifurcating pattern is assumed as the best way to present these relationships.

Morris (2013) points out that some of the depictions of human history at this level are clearly reticulate. Specifically Hooton (1946) at several points in his career depicted human racial relationships using the metaphor of the vertebrate circulatory system. This metaphor led to highly reticulate diagrams of human relationships. So even before genetic methods were applied to understanding divergence in our species, the stage was set for a dichotomous view of our history.

More recently and based on genetic information, human relationships are represented with trees (Nei & Roychoudhry, 1993; Nievergelt *et al.*, 2007; Krause *et al.*, 2010) and reticulating diagrams (Zerjal *et al.*, 2003; Campbell & Tischkoff, 2010), or hybrids of both (Pickerel & Pritchard, 2012; Pickerel *et al.*, 2012). The use of these broad approaches proves that Rieppel's (2010, p. 475) statement "*The history of biological systematics documents a continuing tension between classifications in terms of nested hierarchies congruent with branching diagrams (the 'Tree of Life') versus reticulated relations*" is equally descriptive of the situation in human relationship studies.

Incongruence and phylogenetic analysis

Part of the tension between tree thinking and the more reticulate way of thinking is that genes in the genome of organisms have different evolutionary histories. This phenomenon can clearly be shown by surveying studies at the species boundary or with closely related species for the level of congruence of trees from single genes for a circumscribed group. The overwhelming theme of such a survey is massive incongruence between genetic elements with respect to phylogenetic signal. The incongruence of phylogenetic inference of the many genes in a genome is so extreme that in one group of bacteria, no single gene recovers the same branching pattern as the fully accepted taxonomic topology for the organisms (Bonaventura *et al.*, 2010). Moreover, under some criteria, no single gene recovers the tree

generated by analyzing all genes concatenated together into a single matrix (which is very close in topology to the taxonomic one). A review of the degree of incongruence of trees for diploid eukaryotic organisms (Rosenfeld *et al.*, 2012) reveals that the range of single genes in a genome that are incongruent with the accepted taxonomically derived topology ranges from 22% (higher primate comparisons (Hobolth *et al.*, 2011) to 50% (yeast and frutiflies, Rokas *et al.*, 2003 and Pollard *et al.*, 2008 respectively). In one study of subspecies in the genus *Mus* (White *et al.*, 2009), only 33% of the genes analyzed agreed in topology with the accepted genealogy for the subspecies (Rosenfeld *et al.*, 2012). This percentage is what one would expect if all of the genes in the genome were sorting randomly.

It is not outlandish to suggest that trees constructed from individual genes in the human genome will be as incongruent with each other phylogenetically at the level of relationships of populations of *Homo sapiens*. Mailund *et al.* (2014) suggest that about 25% of the genes in human chimp gorilla comparisons are incongruent with the accepted topology. While this question has not been directly examined in detail for within *Homo sapiens*, one bit of information that can be used to infer the frequency with which individual SNPs differ in inference from accepted topology comes from Reich *et al.* (2010). This paper describes the dynamics of sequence change in Denisova (D), Neanderthal (N) and *sapiens* Yoruba (S) genomes. Only SNPs that are identified as informative and that are transversions are used in the following description. In this study, 46,362 SNPs support the accepted topology of these three terminals ((D,N),S) and 22,012 SNPs support the other two rejected topologies [((S,D),N) and ((S,N),D)]. If we take the Yoruba genome as a representative of the human genome and there is no reason not to, this suggests that around 30% of the informative SNPs based on transversions give conflicting information with respect to the accepted topology of these genus *Homo* entities.

In addition, one can use the data from this paper (Reich *et al.*, 2010) to do the so-called ABBA/BABA tests that are used to quantitate introgression from Neanderthal and Denisova genomes into *sapiens* genomes. In this test two modern human genomes such as San (H1) and Han (H2) in conjunction with either a Neanderthal or Denisova genome (Ac) are used to determine which of the two topologies ((H1, Ac), H2) = ABBA or ((H2,Ac),H1) = BABA is supported by the SNPs. In nearly every case of comparison, on a gross level the ABBA:BABA ratio is 1:1 (from Table 1 in Reich *et al.*, 2010; it should also be noted that the slight skews from 1:1 are used to demarcate introgression). This suggests that the 30% or so incongruent topologies are represented roughly in a 1:1 ratio between one kind of living human genome and the other being more closely related to the archaic genome. It is more than likely that these numbers would be more extreme if all of the informative SNPs were used (ie including the transitions) and might be closer to 50% supporting the accepted topology and 50% supporting the two rejected ones as in other similar studies done at the same taxonomic level. The point here is not to challenge the ABBA-BABA results but to show that phylogenetic patterns from these data contain rampant phylogenetic incongruence. What this means is that current methods we use to construct trees when using large numbers of individual *Homo sapiens* as terminals, will contain a great deal of conflicting information.

Is coalescence the phylogenetic cure to incongruence and hence the road to a resolved *Homo sapiens* tree?

If we can expect this much incongruence in data sets using individual *Homo sapiens* as terminals for phylogenetic analysis, is tree building even worth the effort, to establish relationships of living humans? There are methods that can be used to generate resolved phylogenetic hypotheses using incongruent information. Currently in systematics there are two major ways of utilizing genome level data in a tree building context. The first simply concatenates the data into a single

matrix and treats the data in this concatenated fashion. Partitioned models can be applied to different regions of the concatenated matrix, but the overall hypothesis derived from the concatenated analysis is based on the interplay of all of the data simultaneously. The other approach attempts to use the coalescent in a gene by gene fashion to incorporate coalescent theory into the tree building process. The coalescent is an important evolutionary concept that can be used for phylogenetics at the species boundary and leads to individual trees under the coalescent for each genetic element (i.e.: gene, linkage group, etc.) in a data set. Methods have been developed that then summarize the gene trees (called summary species tree approaches) into a single overall species tree (Liu *et al.*, 2009a,b; Edwards *et al.*, 2007; Liu *et al.*, 2015). In many cases the results from both approaches converge, but when the results don't there are some extreme differences between the concatenation approaches and summary species tree approaches. Several researchers (Gatesy & Springer, 2013, 2014; Springer & Gatesy, 2014; Simmons & Gatesy, 2015), have suggested that coalescent theory and summary species tree approaches have limitations based on the problems they are applied to, especially in deep phylogenetic scenarios.

Clonal and recombining markers

Another connected problem that confuses the issue of tree generation using individual *Homo sapiens* as terminals is that different genes used to do this are inherited in different ways. The most obvious example of this concerns the clonally inherited genes of the mitochondrial DNA (mtDNA) and the Y chromosome. Due to the cell biology of both of these markers the inferences made from them shown in the literature give well resolved trees that are interpretable as histories of the mtDNA and the Y chromosome, and they should. For these markers inferences about human history are usually correctly stated as being relevant to the history of the markers. In other words, the mtDNA analyses indicate maternal history and the Y chromosomal analyses indicate paternal history. By clearly claiming that these clonal markers do not make inferences about individual *Homo sapiens*, but rather about a demographic slice of our species, the studies deliver a valid and useful inference about our species.

On the other hand, autosomal markers are usually used to say something about the individual *Homo sapiens* they come from or at best the population they reside in. Each of the individual 20,000 or so genes on our autosomes in our genomes will also give trees revealing clonal evolutionary history of that element up to a point. But because of different coalescent processes such as lineage sorting, recombination and the lack of clonality of the individual genes, such trees will start to diverge from a clonal picture and the demographic focus of the history from such a gene becomes less and less useful and less and less accurate because of recombination. If one wants to tell 20,000 different "stories" about human history based on the 20,000 or so genes in our genomes then I suppose this might be a useful approach. However if one wants to tell a single unified story about individual *Homo sapiens* by combining the data, then as Tattersall & DeSalle (2011) point out "*The bottom line here, then, is that hierarchical structuring of humans using phylogenetic trees based on the entire genome gives an unrecognizable and unresolved bush.*"

One eye trees, the other clusters

The above statement (with apologies to Paul Klee) refers to the idea that the alternative to constructing a tree to analyze data is to use some variant of clustering. Weiss & Long (2009) point out that these two approaches to analyzing human genomic population level data are pervasive in modern human genomics. They also warn against the use of trees as accurate descriptors of hierarchical structure. Hence many human genomics papers turn to representing their results based on Bayesian K-means clustering, best implemented by the program called Structure (Pritchard *et al.*, 2000). Another approach is to

use principal components analysis to try to discover underlying structure in data sets. While these approaches do not indicate hierarchy of *Homo sapiens* populations, they are capable of indicating divergence of human populations. Results from these approaches are controversial with respect to what they mean about human population divergence (Fujimara *et al.*, 2015).

Smouse (1998) also turned a famous quote into a statement about tree building in a paper entitled "To tree or not to tree" and rejected tree building as a valid approach below the species level. He realized that at certain levels of biological organization, using trees was not only inappropriate, but actually detrimental. Others (Nixon & Wheeler, 1992; Goldstein & DeSalle, 2000) have also recognized the problem and have gone so far as to refer to the point at which tree building no longer is useful as "*the line of death*".

There is a relatively extensive literature on the line of death in the systematics literature starting with Hennig's (Hennig *et al.*, 1966) original description of the problem in his book Phylogenetic Systematics. The extensive nature of this literature exists because systematists realized early on that trees below the species level (systems experiencing tokogenesis to use Hennig's terminology) were meaningless with respect to hierarchy of the entities in the population being studied (Goldstein & DeSalle, 2000). To many, systematists' attempts to tree build at this level are considered inscrutable and unnecessary. One should not expect a hierarchical set of relationships when recombination and admixture abound and have destroyed it. But one can rely on clonal markers like the Y chromosomal and mtDNA data though. Trees using these markers give a nice hierarchical view of the evolution of mtDNA and Y chromosomal lineages.

Why then do we sometimes get dragged into thinking recombining genomic data can give us hierarchical inferences? Nei & Roychoudhry (1993), Nievergelt *et al.* (2007), Krause *et al.* (2010), Zerjal *et al.* (2003), Pickerel & Pritchard (2012), Pickerel *et al.* (2012) and more recently Banda *et al.* (2015), Hoffmann *et al.* (2015), Lazaridis *et al.* (2014) and Prüfer *et al.* (2014)

all use bifurcating diagrams in one way or another to represent human populations. More than likely it is because there are some small regions of the genome that support preconceived notions of hierarchy of human populations. Kittles & Weiss (2003) used an extreme ascertainment strategy to demonstrate that there are indeed loci in the human genome that allow for very distinct clustering of culturally preconceived geographic groupings. The ascertainment strategy used, was to search for those genes that produced patterns that clustered geographical groups of people together in their preconceived groups. In this way they were able to rank all of the genes in their study with respect to how the genes agreed with the preconceived geographic clustering. When they used the top 10% ranked genes, they not surprisingly obtained a tree with individual *Homo sapiens* from their predetermined geographic groups branching together. While demonstrating that there is a subset of the genome that recovers a pattern based on geography is interesting, if one takes the bottom 10% of the genes and constructs a tree, then a completely unresolved tree is obtained (Kittles & Weiss, 2003).

I suggest here that this conundrum is very similar and related to the so-called Lewontin's fallacy (Edwards, 2005). Edwards pointed out that Lewontin's claim of lack of hierarchy in human populations (Lewontin, 1972) is based on ignoring the correlation structure of the data. Edwards claims that when this correlation structure is taken into account, populations can be delineated and distinguished from one another. This view is not in line with modern systematics where the weight of the evidence is used to determine hierarchy. Some systematists' solution to the problem would be to focus on the overall picture of variation and not be tempted to "chuck out" genetic information. If the data are concatenated in this way, for recombining genetic elements (like the X chromosome and autosomes) one can predict that highly unresolved trees will be generated. Some systematists will also attempt to use coalescent approaches to deal with the different signal in the genes due to different coalescent times. This approach is not

defensible at the level of populations within a species, due to the reasons I discussed above. It follows that using just the correlation structure of the data is not in line with a modern systematics view.

Conclusions

I further suggest that the underlying correlation structure of the data at the population level using human genetics is not useful in systematics and hence not useful in producing a hierarchical picture of individual *Homo sapiens*. Instead such data are simply polymorphisms that are either non-informative in a systematic context, or homoplasious. I suggest that analyses using the underlying correlation structure of the data with methods like K-means clustering and multivariate statistical analysis are telling us something different about our history than what systematics would want to say. What the correlation structure is informing us about is our ancestry and while this topic is indeed very interesting, it has nothing to do with taxonomy or hierarchical arrangement of individual *Homo sapiens.* Finally, I suggest that we would make better progress in understanding the events and processes that led to our current highly variable and wonderfully diverse species if we abandon the tree metaphor for understanding the relationships of individual *Homo sapiens*. In reality no such tree exists and theoretically it is useless and inscrutable to strive for one.

Acknowledgements

References

Banda Y., Kvale M.N., Hoffmann T.J., Hesselson S.E., Ranatunga D., Tang H. & Risch N. 2015. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics*, 200: 1285-1295.

Bonaventura M.D., Lee E., DeSalle R. & Planet P. 2010. A whole-genome phylogeny of the family Pasteurellaceae. *Mol. Phylogenet. Evol.,* 54: 950-956.

Campbell M.C. & Tishkoff S.A. 2010. The evolution of human genetic and phenotypic variation in Africa. *Curr. Biol.,* 20: R166-R173.

Edwards S.V., Liu L. & Pearl D.K. 2007. High-resolution species trees without concatenation. *Proc. Natl. Acad. Sci. USA*, 104: 5936-5941.

Edwards A.W.F. 2003. Human genetic diversity: Lewontin's fallacy. *BioEssays*, 25: 798-801.

Fujimura J.H., Bolnick D.A., Rajagopalan R., Kaufman J.S., Lewontin R.C., Duster T., Ossorio P. & Marks J. 2014. Clines Without Classes How to Make Sense of Human Variation. *Sociological Theory,* 32: 208-227.

Gatesy J. & Springer M.S. 2013. Concatenation versus coalescence versus "concatalescence". *Proc. Natl. Acad. Sci. USA,* 110: E1179-E1179.

Gatesy J. & Springer M.S. 2014 Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Mol. Phylogenet. Evol.*, 80: 231-266.

Goldstein P.Z. & DeSalle R. 2000. Phylogenetic species, nested hierarchies, and character fixation. *Cladistics,* 16: 364-384.

Hennig W., Davis D.D. & Zangerl R. 1999. *Phylogenetic systematics.* University of Illinois Press.

Hobolth A., Dutheil J.Y., Hawks J., Schierup M.H. & Mailund T. 2011.Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.*, 21: 349-356.

Hoffmann T. J., Van Den Eeden S. K., Sakoda L. C., Jorgenson E., Habel L. A., Graff R. E. &

Witte J. S. 2015. A Large Multiethnic Genome-Wide Association Study of Prostate Cancer Identifies Novel Risk Variants and Substantial Ethnic Differences. *Canc. Discov.*, 5: 878-891.

Hooton E.A. 1946. *Up From the Ape*, 2nd edition. Macmillan, New York.

Keith A. 1915. *The Antiquity of Man*. Williams & Norgate, London.

Kittles R.A. & Weiss K.M. 2003. Race, ancestry, and genes: implications for defining disease risk. *Annu. Rev. Genomics Hum. Genet.,* 4: 33-67.

Krause J., Fu Q., Good J., Viola B., Shunkov M., Derevianko A. & Pääbo S. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, 464: 894-897

Lazaridis I., Patterson N., Mittnik A., Renaud G., Mallick S., Kirsanow K. & Dugoujon J.M. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature, 513*: 409-413

Lewontin R.C. 1972. The Apportionment of Human Diversity. *Evol. Biol.*, 6: 381-398.

Liu L., Yu L., Kubatko L., Pearl D.K. & Edwards S.V. 2009a. Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet.* Evol., 53: 320-328.

Liu L., Yu L., Pearl D.K. & Edwards S.V. 2009b. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.*, 58: 468-477.

Liu L., Xi Z., Wu S., Davis C.C. & Edwards S.V. 2015 Estimating phylogenetic trees from genome-scale data. *Ann. N.Y. Acad. Sci.*, 1360: 36-53.

Mailund T., Munch K. & Schierup M. H. 2014. Lineage sorting in apes. *Ann. Rev. Genet.*, 48: 519-535.

Morris D. 2012. Human races, networks and fuzzy clusters. *(http://phylonetworks.blogspot. com/2012/08/human-races-networks-and-fuzzy-clusters.html)*

Morris D. 2013 The genealogical world of phylogenetic networks. *(http://phylonetworks.blogspot.com/2013/09/ public-availability-of-phylogenetic-data.html*

Nei M. & Roychoudhury A.K. 1993. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.*, 10: 927-943.

Nievergelt C.M., Libiger O. & Schork N.J. 2007. Generalized analysis of molecular variance. *PLoS Genetics,* 3: e51.

Nixon K. C. & Wheeler Q. D. 1990. An amplification of the phylogenetic species concept. *Cladistic,* 6: 211–223.

Pickrell J.K. & Pritchard J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics,* 8: e1002967

Pickrell J.K., Patterson N., Barbieri C., Berthold F., Gerlach L., Güldemann T., Kure B., Mpoloka S.W., Nakagawa H., Naumann C. *et al*. 2012. The genetic prehistory of southern Africa. *Nature Comm.,* 3: 1143-1147.

Pollard D.A., Iyer V.N., Moses A.M. & Eisen M.B. 2006. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genet.*, 2: e173.

Pritchard J.K., Stephens M. & Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155: 945-959.

Prüfer K., Racimo F., Patterson N., Jay F., Sankararaman S., Sawyer S., Heinze A., Renaud G., Sudmant P.H., de Filippo C. *et al.* 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505: 43-49.

Reich D., Green R.E., Kircher M., Krause J., Patterson N., Durand E.Y., Viola B., Briggs A.W., Stenzel U., Johnson P.L. *et al.* 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature,* 468: 1053-1060.

Rieppel O. 2010. The series, the network, and the tree: changing metaphors of order in nature. *Biol. Philos.,* 25: 475-496.

Rieppel O. 1988. *Fundamentals of comparative biology*. Birkhauser Verlag, Berlin

Rokas A., Williams B.L., King N. & Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature,* 425: 798–804.

Rosenfeld J. A., Payne A. & DeSalle R. 2012. Random roots and lineage sorting. *Mol. Phylogenet. Evol.*, 64: 12-20.

Ruvolo M., Disotell T., Allard M.W., Brown W.M & Honeycutt R.L. 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proc. Natl. Acad. Sci. USA*, 88: 1570-1574.

Simmons M.P. & Gatesy J. 2015. Coalescence vs. concatenation: sophisticated analyses vs. first principles applied to rooting the angiosperms. *Mol. Phylogenet. Evol.*, *91*: 98-122.

Smouse P.E. 1998. To tree or not to tree. *Mol. Ecol.*, 7: 399-412.

Springer M.S. & Gatesy J. 2014. Land plant origins and coalescence confusion. *Trend Plants Sci.*, 19: 267-269.

Tattersall I. & DeSalle R. 2011. *Race?: Debunking a scientific myth*. Texas A&M University Press, College Station.

Weiss K. M. & Long J.C. 2009. Non-Darwinian estimation: My ancestors, my genes' ancestors. *Genome Res.*, 19: 703-710.

White M.A., Ané C., Dewey C.N., Larget B.R. & Payseur B.A. 2009. Fine scale phylogenetic discordance across the house mouse genome. *PLoS Genet.*, 5: e1000729.

Zerjal T., Xue Y., Bertorelle G., Wells R.S., Bao W., Zhu S., Qamar R., Ayub Q., Mohyuddin A., Fu S. *et al*. 2003. The genetic legacy of the Mongols. *Am. J. Hum. Genet.*, 72: 717-721.