The Proceedings of the International Conference on Creationism

Volume 6
Print Reference: Pages 111-116

Article 12

2008

The "Eve" Mitochondrial Consensus Sequence

Robert W. Carter
FMS Foundation

Dan Criswell
Institute for Creation Research

John Stanford
Cornell University

In A. A. Snelling (Ed.) (2008). Proceedings of the Sixth International Conference on Creationism (pp. 111–116). Pittsburgh, PA: Creation Science Fellowship and Dallas, TX: Institute for Creation Research.

The "Eve" Mitochondrial Consensus Sequence

Robert W. Carter, FMS Foundation, 7160 Stone Hill Rd., Livonia, NY 14487
Dan Criswell, Institute for Creation Research, 1806 Royal Lane, Dallas, TX 75229
John Sanford, Department Horticultural Science, NYSAES, Cornell University, Geneva, NY 14456

Abstract

We have calculated the consensus sequence for human mitochondrial DNA using over 800 available sequences. Analysis of this consensus reveals an unexpected lack of diversity within human mtDNA worldwide. Not only is more than 83% of the mitochondrial genome invariant, but in over 99% of the variable positions, the majority allele was found in at least 90% of the individuals. In the remaining 0.22% of the 16,569 positions, which we conservatively refer to as "ambiguous," every one could be reliably assigned to either a purine or pyrimidine ancestral state. There was only one position where the most common allele had an allele frequency of less than 50%, but this has been shown to be a mutational hot spot. On average, the individuals in our dataset differed from the Eve consensus by 21.6 nucleotides. Sequences derived from sub-Saharan Africa were considerably more divergent than average. Given the high mutation rate within mitochondria and the large geographic separation among the individuals within our dataset, we did not expect to find the original human mitochondrial sequence to be so well preserved within modern populations. With the exception of a very few ambiguous nucleotides, the consensus sequence clearly represents Eves mitochondrial DNA sequence.

Keywords

Mitochondrial Eve, Human evolution

Introduction

In order to develop a biblical model of human genetic history, we have calculated a consensus sequence of full-length mitochondrial DNA (mtDNA) from a worldwide sampling of nationalities and ethnic groups. The developed model stands in direct opposition to the Recent African Origins Hypothesis (RAO) (Cann, Stoneking, & Wilson, 1987; Stoneking & Soodyall, 1996), currently the most popular model of human evolution.

There are several critical assumptions on which RAO relies—all part of the "Standard Neutral Model" of Kimura (1968)—and most of the assumptions have been openly questioned in the evolutionary literature (Carter, 2007). These assumptions include the need for mutations to accumulate in all lineages at an equal rate (a molecular clock), that mtDNA undergoes no recombination, and that all new mutations are free from natural selection.

If the molecular clock is violated, a reliable phylogenetic tree for worldwide mtDNA haplotypes cannot be built. Tests for a molecular clock have failed in African L2 clades of mtDNA (Howell, Elson, Turnbull, & Herrnstadt, 2004; Torroni et al., 2001;). This is a grave difficulty for RAO because haplogroup L2a is the most common haplogroup specific to Africa (Salas et al., 2004). The median-joining algorithm (Bandelt, Forster, & Rohl, 1999) is commonly used to create mtDNA phylogenetic trees, but due to the underlying mechanics of the algorithm, clades with a faster-than-average clock will have a disproportionate influence on tree structure (Carter, 2007). This brings into sharp focus the problems associated with lineages that violate the molecular clock assumption. Several recent studies have raised the specter of non-clock like evolution of mtDNA and have openly, but politely, questioned RAO theory (for example, Howell et al., 2004; Zsurka et al., 2007).

A second major assumption behind RAO is that mtDNA undergoes no recombination. This has been debated often in the evolutionary literature but at least one of the newer studies seems to have found conclusive evidence for mitochondrial recombination (Zsurka et al., 2007). If true, many phylogenetic studies will need reassessment for it is the pure maternal inheritance of mtDNA (that is, no input from the paternal side and no recombination of mixed maternal lineages) that allows for a clear-cut phylogeny to be constructed. This might be an explanation for at least some of the homoplasy (identical mutations occurring in parallel lineages) found in the mitochondrial family tree (Zsurka et al., 2007), especially for the African sequences.

Several studies indicate that selection may also

operate on mtDNA (Kivisild et al., 2006; Mishmar et al., 2003; Ruiz-Pesini, Mishmar, Brandon, Procaccio, & Wallace, 2004). Indeed, it is hard to imagine that all mitochondrial mutations (essentially all negative) are free from selection, for all cellular functions rely on the efficient working of the mitochondrion and mitochondrial efficiency is especially important at the extremes of human habitation at both low and high latitudes.

Most mtDNA phylogenies use chimpanzee mitochondrial sequences as an out group. Not only is this a product of circular reasoning, but it also skews the resulting trees towards the consensus of human and chimpanzee mtDNA sequences. The human and chimp mtDNA sequences are substantially different, we do not know the ancestral chimp sequence, and we do not know the degree of degeneration that has occurred in chimp lineages. Each of these factors will affect placement of the root. Without chimp, one is free to explore alternative root placement options. This is essentially what we have done in our consensus calculation. Even though a consensus is not necessarily the same as the ancestral sequence, if the major clades are adequately sampled it should be close (Gao et al., 2003; Gaschen et al., 2002; Nickle et al., 2003).

The genetic facts, apart from the formulation of historical scenarios, are clear: (a) There was a single dispersal of mankind with three main mitochondrial lineages interspersed within the clans. (b) This dispersal either passed through, or originated within, the Middle East. (c) These things happened in the recent past. (d) The dispersion was essentially tribal in nature, with small groups pushing into previously-uninhabited territory. In addition, genetic evidence indicates that male lineages are much more geographically specific than female lineages, with female "migration rates" up to eight-fold higher than males (Seielstad, Minch, & Cavalli-Sforza, 1998; Stoneking, 1998)—a direct confirmation of the Babel account where the initial, well-mixed population split up and migrated according to paternal lineage. These facts are very consistent with a biblical scenario.

While the biblical model fits very well with the data collected by many evolutionary studies, the main difficulty comes from the mtDNA clades from sub-Saharan Africa. These become much less problematic when given proper consideration. Simple visual analysis (see Nordborg, 2001) of any published mtDNA phylogenetic tree (for example, Torroni et al., 2006) indicates that the African clades have had different historical population histories, with the African clades forming a cascading pattern with deep branches and the non-African lineages forming a star-like pattern with short branches. The evolutionary explanation is that these groups have been in Africa for tens of thousands of years longer than the lineages that left Africa. However, there are a number of alternative explanations, all of which support the biblical model. For instance, if the groups that eventually made up the African populations were restricted to smaller tribe sizes until recently, drift would have occurred more quickly and they would have diverged from the rest of the world, and from each other, at a higher rate. Likewise, if the African groups have a different DNA repair system than the others (either defective or differential), this would also explain their more rapid divergence. While these are only a theoretical considerations, they serve to illustrate the large number of assumptions implicit in RAO theory. Generation time is another consideration. Evolutionary models assume equal generation times among all subpopulations, but cultural and genetic factors could easily influence generation time. For example, if the average age of marriage in one population was 20 years old and the average age of marriage in another was 18 years old, a 10% difference in generation time results. Analysis of the life spans of the patriarchs shows how average age of marriage changed dramatically downward in the first generations after the Flood, but there is no indication that this change occurred at the same rate in all populations. Average lifespan differences among populations might also skew generation time differences.

Methods

Publicly available sequences were collected and carefully culled to remove the many sequences with documented errors. The resulting set of 827 sequences is a best effort at generating a set of reliable mtDNA sequences and they should be free from the most common errors described in the literature. A sequence alignment was created in BioEdit (Hall 1999) manually. Using the Revised Cambridge Reference Sequence (rCRS) (Andrews, Kubacka, Chinnery, Lightowlers, Turnbull, & Howell, 1999) as a template for nucleotide numbering and BioPerl (Stajich et al., 2002) for all calculations, a hash table was constructed that included all variant positions with sequence names and nucleotide positions as keys. Output data included a list of variable positions, alleles, and allele frequencies. The consensus (Eve1.0) was constructed by taking the majority allele at each variable position and adding in the invariant sites from the rCRS. Areas with common length polymorphisms were identified and treated separately. These "poly-x" sites were composed of at least three length variations and were treated as single nucleotide positions for all analyses. More detailed methods can be found in Carter (2007).

Human mitochondrial genetic history was

modeled using Mendel's Accountant (Sanford, Baumgardner, Brewer, Gibson, & ReMine, 2007a, b), with parameters designed to mimic the mitochondrial genome (for example, genome size = 16,500 nucleotides, one linkage block). A population of 1,000 individuals (a biblically-reasonable size) was allowed to freely interbreed under realistic constraints for 150 and 10,000 generations. Our population size of 1,000 is considerably smaller than most evolutionary estimates of historic human population size, but recent data suggest an effective human population size of just a few thousand individuals (Tenesa et al., 2007). With an average generation time of 30 years (Tremblay & Vézina, 2000), there have only been c.a. 150 generations in the 4,500 years since the Flood. The 10,000-generation model run is more consistent with evolutionary models.

Results

The Eve mitochondrial consensus sequence is unambiguous. Invariant positions made up 83.9% of the mitochondrial genome, and nearly half of the variable positions (43.8%) were due to the presence of private mutations (that is, at each of these positions, a single sequence in the database carried an alternate allele). Each variable site had more than one allele, but 99% of these sites had a primary allele frequency of 0.90 or greater. That is, for nearly all variant sites, there was a strongly dominant allele. There were only 36 positions (0.22% of the 16,569 nucleotides in human mtDNA) that had a primary allele frequency of less than 0.90. Many of these were not simple polymorphic nucleotides, but were "poly-x" sites (see Methods). For the non-poly-x sites, the ancestral site was consistently and clearly either a purine (nearly all alleles were either "A" or "G") or a pyrimidine (nearly all alleles were either "C" or "T"). The most variable position, 309, is a poly-C tract which is clearly a mutational hotspot and shows a high rate of heteroplasmy within individuals (Carter, 2007). Thus, even the most variable position does not challenge the model of a single invariant human mitochondria in the recent past. This is strong evidence of a young mitochondrial genome.

Pair-wise differences from Eve1.0 and among all sequences in the dataset are given in Figure 1. On average, individuals differed from the consensus at only 22.6 positions, with sequences from sub-Saharan Africa varying at up to 89 positions. Carter (2007) extensively tested inclusion of alternate sequence datasets to see how much the consensus varied according to sample selection and found very little effect of sample selection, even with quite disparate regional distributions of included sequences. Thus, even though the African sequences are quite divergent from the main, inclusion of many more such

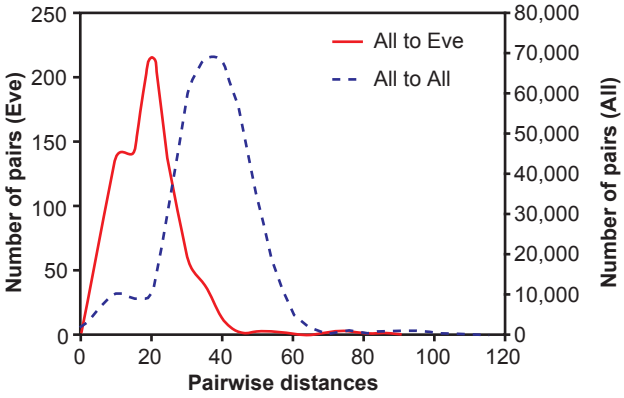

Figure 1. Pair-wise nucleotide differences between all sequences in our dataset and Eve1.0 and among all sequences. The average number of differences between Eve and the worldwide distribution of sequences was 21.6.

sequences will not change the consensus appreciably. At most, more comprehensive sampling would change a handful of nucleotides in the consensus sequence. This is mainly due to the many private mutations that characterize the African sequences. Private mutations are best explained as very recent mutations that have not yet spread into the population. This is also further evidence of a young mitochondrial genome.

Due to genetic drift, it is expected that an old population will harbor many highly-variable positions, with those positions that are unconstrained containing any one of the four possible nucleotides. This is very seldom seen in our dataset. Only 112 of the variable positions carried three alleles and only 6 carried four, and most of these multi-allelic positions still had a strongly dominant consensus allele. This suggests that there has not been enough time to accumulate poly-allelic sites in the human mitochondrial gene pool.

By conventional thinking, young populations should have almost exclusively low allele frequencies, while old populations should have allele frequencies ranging from nearly zero to nearly 1. The program Mendel's Accountant (Sanford, Baumgardner, Brewer, Gibson, & ReMine, 2007a, b), a numerical simulation program, was employed to predict allele frequency distributions in young versus old human populations. Figures 2 and 3 show the expected allele frequency distribution within mitochondria in a model human population after 150 and 10,000 generations, respectively, using realistic population parameters and a realistic mutation rate. There is a very clear difference in the distributions of alleles that have accumulated within the populations. In the younger population, most mutations appear in 0–1% of the population, and there are no new mutations that have accumulated in substantial numbers (for example, allele frequencies of more than 10%). In the older population, there are many more mutations

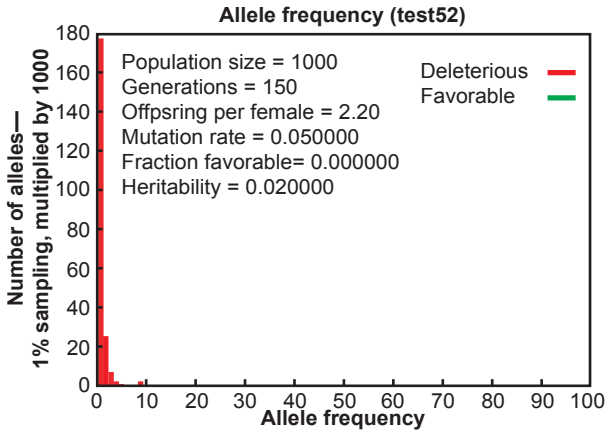

Figure 2. Mitochondrial allele frequencies in a population of 1,000 individuals after 150 generations. These results were generated with Mendel's Accountant under realistic population parameters (as shown).

appearing at higher frequencies, with some drifting all the way to fixation (that is, an allele frequency of 1.0). When population sub-structure (mating primarily within tribes) is modeled, the young and older populations show even more striking differences in the distribution of allele frequencies (not shown).

Allele frequencies were calculated using the current mitochondrial dataset of 827 individuals (Figure 4). The results are strikingly similar to Figure 2, with many mutations in the 0–1% range (indicating a high mutation rate) and very few greater than 10% (an indication of a young genome). Due to the consensus calculation, however, there will be no mutations detected in more than 50% of the population and there is no way to calculate "fixed" mutations. Any fixed mutation would have been called an invariant allele and any "mutation" at greater than 50% would have automatically been assigned to the ancestral state. Due to the nature of the evidence presented in this

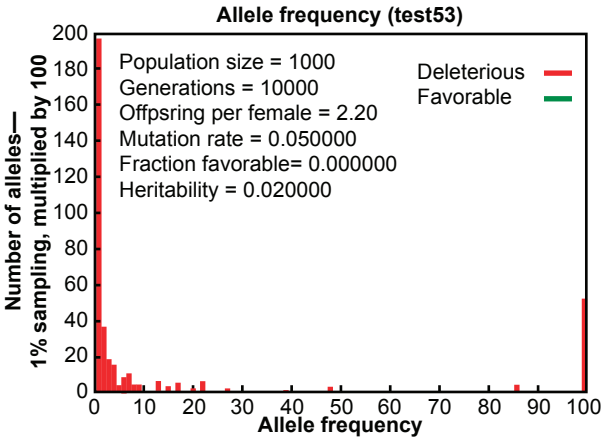

Figure 3. Mitochondrial allele frequencies in a population of 1,000 individuals after 10,000 generations. This is the same population under the same parametric constraints as in Figure 2.

paper (for example, lack of high-frequency alternate alleles), it seems unlikely that any mutations have gone to fixation in human mtDNA.

Discussion

There are several reasons why we claim that Eve1.0 is nearly identical to the real Eve mitochondrial sequence. We feel that we have a fair representation of most of the world's population, and including more sequences from the under-sampled populations (for example, sub-Saharan Africa) should not change the consensus significantly. The only possible positions that might change with the addition of more samples are the small number of sites that are currently ambiguous (0.22% of sites have a primary allele frequency less than 0.90), but even those will stay within the purine/pyrimidine states that we have identified as ancestral. If there were many positions with no dominant allele, we could not reasonably infer the original Eve sequence. Essentially, all mutant alleles are rare. The mitochondrial genome is subject to high mutation rates (as evidenced from the high degree of private mutations), but the lack of significant worldwide variation indicates a young mitochondrial genome.

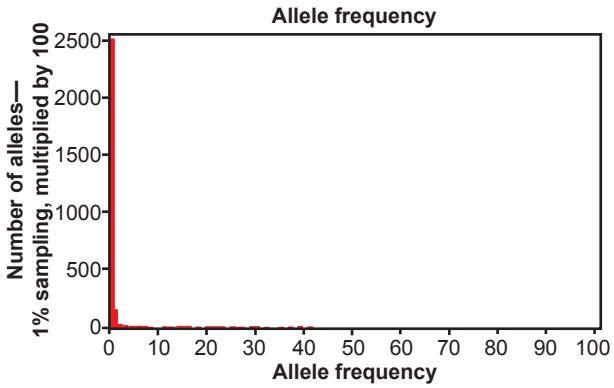

Figure 4. Mitochondrial allele frequencies in a worldwide sampling of mitochondrial DNA. The 827 sequences in the dataset were used to generate these data and they reflect a young mitochondrial genome when compared to Figures 2 and 3.

The African mitochondrial lineages are still problematic, but mainly because of lack of data. Africa may have been subject to different population structure than rest of world. Harsh conditions (the only significant world population to span the equator), small tribal units, shorter generation times (controlled by either genetics or culture), shorter average life spans, etc., plus perhaps a defective or differential DNA repair system could all contribute to higher African mitochondrial diversity. African mitochondria appear to have more mutations, relative to the consensus, but why is there is no evidence of impaired metabolism? Most mutations are nearly

neutral so the correlation between mutation count and fitness should be very weak.

The Eve1.0 consensus is identical to the root node for macrohaplopgroup R in the evolutionary nomenclature of mtDNA clades (Figure 5a). We were surprised to find that our results confirmed the evolutionary studies in this way, but after examining our methods in light of theirs, we understood that the two approaches mirrored one another mathematically. Our placement of Eve in clade R is much different from their conclusion of an ancestor much "earlier" in superhaplogroup L. Their conclusion is biased by: (a) assuming that humans and chimpanzees had a common ancestor; (b) using chimpanzee sequences as an out group; and (c) giving equal weight to sequences that may be accumulating mutations more rapidly (Carter, 2007).

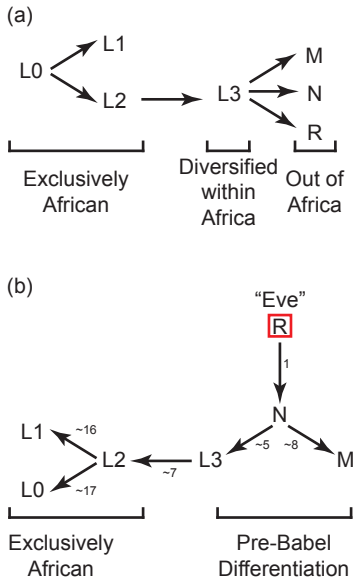

Figure 5. Alternate Mitochondrial Tree Possibilities. (a) The "Out of Africa" model of human mitochondrial history. According to Out of Africa proponents, the most ancient extant human mitochondrial lineage is L0, which originated in Africa many thousands of years ago. L1 and L2 diverged from L0 within Africa. L3 also originated within Africa, but three of its major branches, M, N, and R, were involved in the "Out of Africa" event that led to the colonization of the rest of the world by modern humans. N and R are only separated by a single nucleotide difference, and evolutionary models assign R to the most derived state, but they are shown here equidistant from L3 for simplicity. (b) One of several possible models of biblical mitochondrial history. Eve1.0 is identical to the root node of macrohaplogroup R. From R, the closely related M, N, and L3 lineages diverged (small numbers indicate the approximate number of mutations that separate the nodes for each lineage, not the number of mutations that separate individuals among the various clades). One of the L3 lineages entered Africa and gave rise to the African-specific lineages, L2, L1, and L0.

As more sequences become available for inclusion in a world-wide consensus, there is a possibility that a few of the consensus nucleotides will change, especially those that separate the majority of the world population from the clades specific to sub-Saharan Africa. However, most of the differences we see in these lineages are due to rare, homoplasic, or private alleles so generally these should have no affect on the consensus calculation. Only a few positions will be subject to change due to sample composition (that is, those with the least dominant consensus alleles). Several of these positions drive the major breaks in tree topology, so we expect future revisions in Eve to be minimal and to follow the general outline of the model given in Figure 5b.

One confounding factor in our work is the mitochondrial bottleneck that occurred during the Genesis Flood. It is not possible to tell how much mitochondrial diversity existed prior to the Flood, how closely related the three available mitochondrial lineages were (one each from the three daughter's-in-law of Noah), or if one of the three mitochondrial lineages went extinct in the first few generations (not likely with an exponentially growing population). There is also a slight possibility that Noah and his wife had children after the flood (this is not excluded by the text). Since our calculations are based on a consensus of available sequences, it is possible that Eve1.0 could be off by a few nucleotides. This could position Eve within macrohaplogroup N (only one mutational difference separates the root nodes of R and N), or, possibly, beyond N and in the direction of L3.

From an evolutionary perspective, the apparent youth of the human mitochondrial genome might be explained by an extreme (effectively homogenizing) population bottleneck, about 100,000 to 200,000 years ago and lasting for tens of thousands of years. Our allele frequency analysis indicates this relatively "recent" date is still too long ago to explain existing human mitochondrial variation. Furthermore, numerical simulations (Sanford, Baumgardner, Brewer, Gibson, & ReMine, 2007a) of human population bottlenecks demonstrate that a bottleneck that is severe enough to homogenize a population will either cause severe genetic degeneration or extinction (purifying selection breaks down in small populations). An integral aspect of homogenization via a population bottleneck is the systematic fixation of nearly neutral deleterious mutations. Therefore, the pre-bottleneck population should always be superior to the post-bottleneck population. This fact is not compatible with the evolutionary scenario where sub-humans go into a disastrous bottleneck, and then modern humans come out the other end.

References

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics, 23*, 147.

Bandelt, H.-J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16*, 37–48.

Cann, R.L., Stoneking, M., & Wilson, A.C. (1987). Mitochondrial DNA and human evolution. *Nature, 325*, 31–36.

Carter, R.W. (2007) Mitochondrial diversity within modern human populations. *Nucleic Acids Research, 35*(9), 3039–3045.

Gao, F., Bhattacharya, T., Gaschen, B., Taylor, J., Moore, J.P., Novitsky, V., et al. (2003). Response to Nickle et al. 2003. *Science, 299*, 1517–1518.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B.H., Bhattacharya, T., & Korber, B. (2002). Diversity considerations in HIV-1 vaccine selection. *Science, 296*, 2354–2360.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symposium Series, 41*, 95–98.

Howell, N., Elson, J.L., Turnbull, D.M., & Herrnstadt, C. (2004) African haplogroup L mtDNA sequences show violations of clock-like evolution. *Molecular Biololgy and Evolution, 21*, 1843–1854.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature, 217*, 624–626.

Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., et al. (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics, 172*, 373–387.

Mishmar, D. Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G, Hosseini, S., et al. (2003). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences, 100*, 171–176.

Nickle, D.C., Jensen, M.A., Gottlieb, G.S., Shriner, D., Learn, G.H., Rodrigo, A.G., & Mullins, J.I. (2003). Consensus and ancestral state HIV vaccines. *Science, 299*, 1515–1517.

Nordborg, M. (2001). Coalescent theory. In D. Balding, M. Bishop, & C. Cannings (Eds.), *Handbook of statistical genetics* (pp.179–212). Wiley: Chichester, UK.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., & Wallace, D.C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science, 303*, 223–226.

Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., & Carracedo, A. (2004). The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *American Journal of Human Genetics, 74*, 454–465.

Sanford, J.S., Baumgardner, J., Brewer, W., Gibson, P., & ReMine, W. (2007a) Mendel's accountant: A biologically realistic forward-time population genetics program. *Scalable Computing: Practice and Experience 8*(2), 147–165.

Sanford, J.S., Baumgardner, J., Brewer, W., Gibson, P., & ReMine, W. (2007b) Using computer simulation to understand mutation accumulation dynamics and genetic load. In Y. Shi, C.D. van Albada, J. Dongara, & P.M.A. Sloot (Eds.), *International Conference on Computational Science 2007, Part II, Lecture Notes in Computational Science 4488* (pp.386–392).

Seielstad, M.T., Minch, E., & Cavalli-Sforza, L.L. (1998). Genetic evidence for a higher female migration rate in humans. *Nature Genetics, 20*, 278–280.

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Research, 12*, 1611–1618.

Stoneking, M. (1998). Women on the move. *Nature Genetics, 20*, 219–220.

Stoneking,M., & Soodyall,H. (1996). Human evolution and the mitochondrial genome. *Current Opinion in Genetics & Development, 6*, 731–736.

Tenesa A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard M.E., & Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Research, 17*, 520–526.

Torroni, T., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., et al. (2001). Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *American Journal of Human Genetics, 69*, 1348–1356.

Torroni, A., Achilli, A., Macaulay, V., Richards, M. and Bandelt, H.-J. (2006). Harvesting the fruit of the human mtDNA tree. *Trends in Genetics, 22*, 339–345.

Tremblay, M., & Vézina, H. (2000). New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *American Journal of Human Genetics, 66*, 651–658.

Zsurka, G., Hampel, K.G., Kudina, T., Kornblum, C., Kraytsberg, Y., Elger, C.E., Khrapko, K., & Kunz, W.S. (2007). Inheritance of mitochondrial DNA recombinants in double-heteroplasmic families: potential implications for phylogenetic analysis. *American Journal of Human Genetics, 80*, 298–305.