

THE DENISOVA CAVE mtDNA AND NUCLEAR SEQUENCES: ITS RELATION TO HUMAN EVOLUTION

Niccolo Caldararo, Ph.D.

Dept of Anthropology

San Francisco State University

1600 Holloway Ave.

San Francisco, Ca. 94132

415-453-9064 cald@sfsu.edu

Michael Guthrie

Dept of Biological Sciences

City College of San Francisco

54 Phelan Ave.

San Francisco, Ca. 94112

Key Words: Denisova Cave sequence, Neandertal inserts, Genbank aDNA

Abstract: The Altai (Denisova Cave) fossil mtDNA sequence was purported to be so different from anatomically modern humans that it was suggested to be a different species, yet the bone material it was drawn from seems to have the physiological landmarks of that species designation. This is difficult to determine given the small fragment of a juvenile manual phalanx. nDNA was drawn from a molar tooth and DNA analysis indicated both bone samples were from the same population but that the relationship of this population to modern humans is separated by over 1my as well as from Neandertals. Further, it is claimed that present day Melansians share genetic material with the Denisova population. What significance this has is unclear especially regarding theories of the evolution of modern humans and warrants further study of the sample sequence. Examination of the published sequence found that the alignment of segments in the mtDNA hypervariable regions could be aligned with that of anatomically modern humans if one introduced an insert at a position found in Neandertals. Some other points of interest arise from a reconsideration of the sequences for other published samples and Neandertals from the same perspective. The implications for systematics and human phylogenetic analysis are considerable.

The recent publication of a mtDNA sequence by Krause, et al.(2010) produced a proposal by the authors that the differences between this sequence, that of modern humans and Neandertal sequences indicated that the Denisova individual was probably derived from an unknown hominid population that shared its last common ancestor with AMH and Neandertals before 1.0 mya. The reason for this speculation was the great number of base pair differences between the Denisova sample and AMH and Neandertal samples. While there is significant evidence of degradation present in the reported sequence which parallels degraded mtDNA as Caldararo and Gabow (2000) argued in a paper in Ancient Biomolecules, and Caldararo (2004) extended in a later analysis, some sequences do align with published samples. Our analysis leads us to conclude, however, that the Denisova sample was significantly degraded and the resulting sequence up to 16193 contains corrupted mtDNA. After 16193 if one reads into the sequence a break as appears in the Krings, et al. 1997 paper, as an insert of a cytosine between 16193 and 16194 the sequence aligns as human given human reference samples in GenBank and presented by Caldararo & Gabow and Krings, et al., 1997. We note this insert in the Caldararo & Gabow 2000 paper where we align several modern reference sequences with several ancient sequences both identified from Neandertal material (Feldhofer consensus) and early moderns. There is another insert between 16262 and 16263 in the Krings et al. (1997) sequence for Feldhofer but this does not appear in the Denisova sequence. If one adds the Denisova Cave sequence to our alignment (see figure 1) and begin the reading from the Neandertal insert we find that the Denisova sequence then fits the reference sequence of Anderson and that used by Krings (1997) with only a few bp variations (18). The same is not true when one reaches 16264. There is also an insert between 16263 and 16264 in the Neandertal sequence but if one reads ignoring the Neandertal insert using the Denisova data the reading is nearly identical to the contemporary

modern human sequence. In this Nature Precedings article the mtDNA sequences are published as a pdf file separately.

This changes the appearance of the Denisova sequence and makes it appear to reflect a combination of Neandertal sequences and AMH sequences consistent with the recent analysis of the Neandertal genome by Green, et al, (2010). In the entire sequence from 16020 to 16409 the adjusted Denisova sequence agrees with the Neandertal sequence of Krings, et al. 1997, 17 times where both do not agree with the Anderson sequence. But the Neandertal sequence differs from the Denisova sequence but agrees with the Anderson sequence in other 124 locations but of these 113 are found up stream of the C insert between location 16193 and 16194. Of the other ancient mtDNA sequences we provide in our 2000 chart, Betty (1996) agrees 19 times with Denisova where variations occur from the Anderson sequence, Horai and Hayasaka (1990) 17 times, Handt (1996) 8 times and Handt (1994) 4 times. The Betty sample is from Australian Aborigines and the Horai and Hayasaka from Japan.

The recognition of the Neandertal inserts in the Denisova sequence changes the reading considerably and indicates that the sequence before the insert at 16194, perhaps ending at 16181 is corrupt either from degradation or during preparation for sequencing. There may be another explanation for the lack of sequence alignment before that location and the substantial agreement after it. Nevertheless, this finding argues for modern human status of the Denisova sample and against a new species designation as suggested from the original analysis (Krause, et al., 2010). Variations in mtDNA in populations and their significance given the natural history of mitochondria have been noted by Ballard and Whitlock (2004) with a caution of their use to build phylogenetic relationships. The production of algorithms to assess the relevance of gaps in alignment, deletions, insertions, etc. has resulted from a number of philosophical assumptions

about their placement and frequency and length (Needleman & Wunch, 1970; Sellers, 1974). Historically these included a variety of indexes, including similarity indexes, distance indexes and others (Li & Graur, 1991). Understanding the meaning in evolution of such alignment differences depends on the assumptions behind the indexes rather than how the gaps have affected selection or how they reflect mechanisms of the chemistry of molecular evolution.

The agreement of algorithms has become more important than the biology of the organisms. In the same way the mathematics and philosophy (assumptions) of the Ptolemaic Geocentric system retarded the development of astronomy, we are seeing a focus on algorithms supersede analysis of the molecular biology of the gene in evolutionary context. Leslie White (1949) perhaps produced one of the most concise commentaries showing how mathematics is dependent on cultural assumptions, but John Maynard Keynes (1926) reminded economists in his work on statistics that probability is often mistaken for reality. Caldararo discusses this in the issue of various ideas of the species concept and ancient mtDNA in an article in *The Linnean* (2003). There are a number of definitions of inserts, mtDNA sequences transferred to the nDNA and mutations in sequences of both nDNA and mtDNA, the latter move the frame and can be either silent or damaging to replication of proteins. This latter case is the situation with the Neandertal insert. One has to keep in mind that the published sequences are of varied quality (Carter, 2007 see <http://nar.oxfordjournals.org/content/35/9/3039.full> for details) However, inserts are not unusual in the human mitochondrial genome, reference to the MITOMAP database produces many locations for inserts and deletions (<http://www.mitomap.org/bin/view.pl/MITOMAP/PolymorphismsControl>). Mutations in the D-loop area are also associated with some diseases as in cancers (Lee, et al., 2005) and in aging, though recent surveys have shown that cumulative levels of base substitutions in mtDNA can be

very low (Shokolenko, et al., 2009). Some of the philosophical aspects of phylogenetic tree building were discussed by Caldararo in an article in 2002 and Caldararo & Guthrie regarding constructing programs especially using Neutral Theory, but also concerning the nature of the mitochondrial genome and its inheritance and relating variations in the Y chromosome to nDNA as well as interpretations of the Mungo sequence published in 2001. The points raised in these papers demonstrate that the philosophical basis of gene sequence variation is neither clearly integrated between these three sources of information and variation, nor is it defined concerning what variations mean to the species status of individuals possessing these variations. More recent critics of these assumptions have questioned the veracity of molecular clocks in general (Schwartz & Maresca, 2006). As Caldararo and Gabow (2000) demonstrated in an analysis of chimpanzee DNA and the reported Feldhofer mtDNA, the variations between interfertile chimpanzees were greater than those reported between modern humans and Neandertals. So the idea that Neandertals should be considered a different species seems illogical. Rather we look at the variations as similar to those reported for wolves, coyotes and dogs, all interfertile (Rutledge, Patterson & White, 2010). One wonders how different Neandertal DNA would be compared with a sequence of a surviving sample from a population dated to that of the Feldhofer sample but from an anatomically similar skeleton to that of Cro Magnon. Krause, et al. (2009) criticize work on early modern aDNA as liable to contamination and instead use Neandertal mtDNA as a reference for work on early modern mtDNA as in the Kostenki remains. This begs the question of the authenticity of the Neandertal mtDNA and potential contamination. Authenticity here depends on an arbitrary size of Neandertal mtDNA molecules retrieved by PEC. However, they argue that fragment size alone cannot be used to gauge authenticity citing Green, et al. 2009. Further they find in the Kostenki samples parallels to the degradation process seen in Neandertal

samples in length, nucleotide misincorporations and fragmentation patterns (Krause, et al. (2009) The discovery of a segment sequencing haplogroup U in the Kostenki sample was claimed to be authentic while it can equally indicate contamination.

In another paper (Reich, et al., 2010) report on the sequencing of nDNA from samples of human bone from the same site. This study combined findings from an earlier sequencing of the Neandertal nuclear DNA which argued that present-day humans share common ancestors with Neandertals about 800,000 years ago and that a population split occurred between premodern populations leading to modern humans and Neandertals at about 270,000 to 440,000 years ago. It also asserted that Neandertals shared more genetic variants with present-day humans in Eurasia than with present-day humans in Africa. Applying this interpretation to their sequence from Denisova Reich, et al. (2010) argue that they found the Denisova Cave population contributed between 4 and 6% of its genetic material to the genomes of present-day Melanesians.

Degradation of Neandertal mtDNA samples demonstrated “drastically “ different levels of contamination as reported by Green, et al. (2006). In some widely separated sites where samples were found (France, Russia and Uzbekistan) only “around 1% of the mtDNA displayed Neandertal-like sequences. Of course, if one has interpreted a degraded sample sequence as “Neandertal-like” then one is simply searching for degradation parallels and not authentic species sequences. One sample from Croatia and one from Spain contained around 5% to 75% Neandertal-like sequences according to Green, et al., (2006).

In another paper by Green, et al. 2010, the authors report on a draft sequencing of the Neandertal genome. The results show “more shared genetic variants with present-day humans in Eurasia than with present-day humans in sub-Saharan Africa.” These two conclusions confound

current theories of modern human evolution. The Out of Africa/Replacement theory espoused by Stringer & Andrews, (1988) applies the perspective that modern humans evolved in South Africa and migrated north eventually replacing all other premodern hominids, with no gene flow or contribution to present-day human populations by the premoderns. The Regional Continuity theory of Wolpoff and Caspari (1997) presents the idea of geographic populations in Europe, Central Asia, North and South Asia and Australia and Africa all evolving from premodern status to anatomically modern human status as a widespread single species maintained by gene flow. A third alternative that has become popular in recent years is a combination of and out of Africa waves of migration or gene flow over the past 2 million years with comprehensive interbreeding patterned by alternating periods of isolation (Brauer, 1989; Hawks and Wolpoff, 2001; Templeton, 2005). A more complex situation is emerging from an analysis of a recent skull from Central Africa dated about 13,000 (Harvati, et al., 2011). Regarding the genetic information from Denisova, Henry Harpending was quoted as saying, ““We had a number of hints that there was something else in Melanesian genome, an admixture from some other group,” he said. “To discover that it was from this particular group suggests that it was pretty widespread in Asia.” (see <http://articles.latimes.com/2010/dec/24/science/la-sci-new-hominid-20101224>).

At first blush it would appear that the Reich, et al. 2010 conclusion would support the third theory of African waves of migration and interbreeding. However, the analysis that Templeton (2005) produced does not provide for the complete isolation of Melanesian populations as proposed by Reich, et al. 2010. Rather the Reich, et al. 2010 proposal could inadvertently give life to an earlier theory of human evolution of the polygenesis proponent, Carlton Coon (1962). Coon argued that hominids had evolved to the *Homo erectus* grade all over the old world and then evolved in situ into perceived geographically distinct races today. This did not preclude

gene flow at times, but was not necessarily considered by Coon to be significant. The issue of the nature of population variants and the contributions of haplotypes needs to be clarified to not only determine the species status of Neandertals and the Denisova remains, but how we consider population diversity in general. An example of this appears in the Green, et al. (2010) study where a Yoruba individual has a divergence estimate to the human genome sequence about 14% greater than previous estimates for an African American individual and the heterozygosity measured in another Yoruba individual. Such individual variation needs to be considered in understanding the comparability of ancient DNA samples from populations dating tens of thousands of years ago. We are ignorant of the population diversity of the past and the role sedentary behavior had played in forming present human diversity, especially under the pressure of epidemic diseases (Caldararo, 1996;Caldararo & Gabow, 2000). In this regard, by analyzing HLA haplotypes Abi-Rached, et al. (2011) have identified adaptive introgression from archaic to modern humans using data from the Denisova aDNA.

It seems to us, from our analysis of the Denisova mtDNA sequence that the Reich, et al. 2010 conclusion is unnecessary and in error. A simpler explanation is arrived at by the Neandertal insert concept we have proposed which leaves us with an early modern population in the process of the out of Africa wave theory of Hawkes, Wolpoff and Templeton. We are suspect of the nDNA sequence Reich, et al. have produced as nuclear DNA is notoriously more liable to degradation than mtDNA (Caldararo & Gabow, 2000) . While we commend the authors of these different studies of Neandertal DNA and the Denisova samples efforts to eliminate contamination, the sequences do not appear without considerable difficulty to assure authenticity.

What is also interesting is that the Genbank Neandertal Feldhofer sequence (FM865407.1) differs substantially from the clones published in the original Cell article by Krings, et al. (1997). In the original paper the various clones are shown aligned together with the Anderson reference sample in Figures 4 and 5 of their paper beginning in Figure 4 at 16,022 in Figure 4 and running to 16,401 in Figure 5 with the Neandertal consensus of the clones from the different laboratories at the bottom. For example an “A” appears at position 16,037 in the Anderson reference sample but a consensus of the clones for the Neandertal finds a “G” at this site. No other differences are seen in the sequence for the Feldhofer as consensus before this site except in two clones at position 16,036. In the Genbank sequence for Feldhofer we find 10 differences listed. These identical differences are found in the Genbank sequences for Neandertal isolates from Mezmaiskaya 1, Vindija 33.25, and Sidron 1351e. It is difficult to explain this situation and we will not venture an explanation but would hope that one is available from the laboratories involved. It may be that the human sequences are contaminated as has been found to be true of many of the non-Primate archived sequences discovered by Mark Longo and his colleagues (Longo, 2011; Phillips, 2011). Forster (2003) in an earlier paper detailed problems with human mtDNA published data, reinforced by the work of Yao, Bravi & Bandelt, 2004.

It often seems that the state of molecular phylogenetics is like morphological genetics before Huxley's "new systematics" in the 1930s. It seems in the condition molecular studies of viruses was before the concept of “quasi-species was introduced. Caldararo discussed this issue in a review of a book on the evolution of HIV in the American Journal of Human Biology in 2001 [http://onlinelibrary.wiley.com/doi/10.1002/1520-6300\(200102/03\)13:2%3C289::AID-AJHB1047%3E3.0.CO;2-W/abstract](http://onlinelibrary.wiley.com/doi/10.1002/1520-6300(200102/03)13:2%3C289::AID-AJHB1047%3E3.0.CO;2-W/abstract)). We see the same passion to name each difference in sequence as a species as in Darwin's day a new species was named with every difference in

morphology. Stephen J. Gould's discussion of this regarding Darwin's redefinition of von Baer's work on embryology and recapitulation is worth revisiting (Gould, 1977). It does, however, seem clear that assumptions of human diversity in the past as in the present are in need of revision. Work by Fraumene, Belle, Castri, et al., 2006, show much more diversity in a population in Sardinia than would be expected given present assumptions of human diversity. Since present phylogenetic analysis is based on these assumptions, new paradigms of the role of diversity should be explored.

Bibliography

Abi-Rached, Laurent; Jobin, Matthew J.; Kulkarni, Subhash, et al., (2011) "The shaping of Modern Human immune systems by multiregional admixture with Archaic Humans," Science, 7 October, Vol. 334 no. 6052 pp. 89-94.

Betty, D.J.; Chin-Atkins, A.N.; Croft, L.; Sraml, M. and Easteal, S., (1996) "Multiple independent origins of the COII / rRNA interspersed 9-bp mtDNA deletion in Aboriginal Australians," Am. J. Hum. Genet., v. 58:428-433.

Ballard, J.W. & Whitlock, M.C., (2004) "The incomplete natural history of mitochondria," Mol.Evol., Apr. v. 13, n. 4, pp 729-744.

Brauer, Gunter, (1989) "The Evolution of Modern Humans: a Comparison of the African and non-African Evidence," in Paul C. Mellars and Chris B. Stringer ed. The Human Revolution. Princeton: Princeton University Press, pp. 123-153.

Caldararo, Niccolo (1996) "The HIV/AIDS epidemic: its evolutionary implications for human ecology with special reference to the immune system," The Science of the Total Environment, v. 191, pp. 245-269.

Caldararo, Niccolo, (2002) "Ancient DNA and human origins: the role of gene sequence variation and the species concept," Human Nature Review, v. 2, 317-321. <http://www.human-nature.com/nibbs/02/caldararo.html>.

Caldararo, Niccolo (2003) "Human phylogenetics and ancient DNA: the role of gene sequence variation in the species concept," The Linnean, v. 19: 27-29.

Caldararo, Niccolo & Guthrie, Michael, (1998) "Mitochondrial DNA, the Y chromosome and the origins of modern humans," Homo, v. 49/2: 225-240.

Caldararo, Niccolo,(2004) "Some effects of the use of ultrasonic devices in conservation and the question of standards for cleaning objects," Objects Specialty Group Postprints, American Institute for Conservation of Historic and Artistic Works, v.11, pp 126-153.

Caldararo, Niccolo & Gabow, Steve, (2000) "Mitochondrial DNA analysis and the place of Neandertals in Homo," Ancient Biomolecules, v. 3: 135-158.

Carter, Robert W., (2007) "Mitochondrial diversity within modern human populations," Nucleic Acids Research, v. 35, n. 9, pp. 3039-3045.

Coon, Carleton, (1962) The Origin of Races, N.Y., Knopf.

Forster, P., (2003) "To err is human," Annals of Human Genetics, v. 67, pp. 2-4.

Fraumene, Cristina, Belle, Elise M.S., Castri, Loredana, et al., (2006), "High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates," Mol. Biol. Evol., v. 23, n. 11, pp. 2101-2111.

Gould, S.J., (1977) Ontogeny and Phylogeny, Cambridge Mass., Belknap Press.

Green, Richard E., Krause, J. & Ptak, Susan, et al., "Analysis of one million base pairs of Neandertal DNA," Nature, v. 444, 16 November 2006:330-336.

Green, Richard E.; Krause, Johannes; Briggs, Adrian W., et al., (2010) "A draft sequence of the Neandertal genome," Science, v. 328:710-722.

Handt, O., Richards, M., Tromsdorf, M., Kilger, C., et al., (1994) "Molecular genetic analyses of the Tyrolean Ice Man," Science, v. 264, pp. 1775-1778.

Handt, O.; Krings, M.; Ward, R.H. & Paabo, S., (1996) "The retrieval of ancient human DNA sequences," Am.J.Hum.Genet.,v. 59, pp. 368-376.

Harvati, Katerina, Stringer, Chris, Grun, Rainer, et al., (2011) "The Later Stone Age calvaria from Iwo Eleru, Nigeria: Morphology and Chronology," PLoS ONE, 6(9):e24024. Doi: 10.1371/journal.pone.0024024.

Hawks, John D. & Wolpoff, Wilford H., (2001) "The accretion model of Neandertal evolution,"Evolution, v. 55, (7), pp. 1474-1485.

Horai, S. & Hayasaka, K., (1990) "Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA," Am. J. Hum. Genet., v. 46, pp. 828-842.

Keynes, John Maynard, (1921) A Treatise on Probability, Cambridge, Cambridge U. Press.

Krause, Johannes; Fu, Qiaomei; Good, Jeffrey M.; Viola, Bence; Shunkov, Michael; Derevianko, Anatoli and Paabo, Svante, (2010) "The complete mitochondrial DNA genome if an unknown hominin from southern Siberia," Nature, v. 464:894-897.

Krings, M., Stone, A., Schmitz, R.W., et al., (1997), "Neandertal DNA sequences and the origin of modern humans," Cell, v. 90, pp. 19-30.

Lee, H.C., Yin, P.H., Wu., CC, et al., (2005), "Mitochondrial genome instability and mtDNA depletion in human cancers," Ann. NY Acad. Sci., May, 1042:109-122.

Li, Wen-Hsiung & Graur, Dan, (1991) Fundamentals of Molecular Evolution, Sunderland, Sinauer Associates.

Longo, Mark, J. O'Neill, Michael; O'Neill Rachel J., (2011) "Abundant Human DNA Contamination Identified in Non-Primate Genome Databases," PLoS ONE 6(2): e16410. doi:10.1371/journal.pone.0016410.

Needleman, S.B. & Wunsch, C.D., (1970) "A general method applicable to the search of similarities in the amino acid sequence of two proteins," J.Mol.Biol., v. 48, pp. 443-453.

Phillips, Melissa Lee, (2011) "Genome databases suffer from the human touch," Scientific American, February 2011:2.

Reich, D., Green, Richard E., Kircher, Martin, et al., (2010) "Genetic history of an archaic hominin group from Denisova Cave in Siberia," Nature, v. 468, Dec. 23/30, pp1053-1060.

Rutledge, Linda, Patterson, Brent R. & White, Bradley, (2010) "Analysis of *Canis* mitochondrial DNA demonstrates high concordance between the control region and ATPase genes," BMC Evolutionary Biology, v. 10, pp. 215-228.

Sellers, P.H., (1974) "On the theory and computation of evolutionary distances," SIAM J. Appl. Math., v. 26, pp. 787-793.

Shokolenko, Inna, Venediktova, Natalia, Bochkareva, Alexandra, et al., (2009) "Oxidative stress induces degradation of mitochondrial DNA," Nucleic Acids Research, v. 37, n. 8, pp. 2539-2548.

Stringer, C.B. & Andrews, P., (1988) "Genetic and fossil evidence for the origin of modern humans," Science, v. 239, 11 March, pp. 1263-1269.

Swartz, Jeffrey H. & Maresca, Bruno, (2006) " Do molecular clocks run at all? A critique of molecular systematic," Biological Theory, v. 1, n. 4, pp.357-371.

Templeton, Alan R., (2005) "Haplotype trees and modern human origins," Yearbook of Physical Anthropology, v. 48:33-59.

Wolpoff, M.H. & Caspari, R., (1997) Race and Human Evolution, Simon & Schuster, New York.

Yao, Yong-Gang, Bravi, Claudio, Bandelt, Hans-Jurgen, (2004) "A call for mtDNA data quality control in forensic science," Forensic Science International, v. 141, pp. 1-6.

