The Time of Origin and the Rate of Formation of Haplogroup I and its Subclades I1 and I2

-- William E. Howard III --

Abstract:

Using a correlation approach to analyze Y-DNA haplotypes in Haplogroup I, we investigate how the numbers of pairs within the same haplogroup vary as a function of time, and we can estimate the rate of change of its formation. We have estimated when the progenitors of Haplogroups I1 and I2 lived (~27000 BC for Hg I2 and ~9000 BC for Hg I1), dates that are consistent with, but slightly higher than past estimates. This analysis is arguably the first time that changes in pairs of haplotypes have been followed in time from the progenitor of the subclades to the present. Events in the evolutionary history of these subclades began during and at the end of the last glacial maximum, suggesting that the rate of formation of the Hg I2 subclade was influenced by where and when early carriers of this subclade lived. This rate accelerated as early hunter-gatherers transitioned to farming, reaching a maximum at about 6000 BC for Hg I2 and 1000 BC for Hg I1. The rate of change in the number of pairs of Haplogroup I and the growth rate of the world's population between 27,000 BC and 2000 BC have the same slope.

Introduction:

Using the RCC correlation approach to analyze Y-DNA haplotypes in Haplogroup I, we have investigated 281 testees whose Y-DNA can be divided into the two main subclades of Haplogroup I (Hg 1), Hg I1 and Hg I2[1]. The analysis was done separately on the I1 and I2 groups and on their sum. The goal of this analysis is to investigate (1) to the time of origin of Hg I and its two main components, Hg I1 and I2; (2) the rate of formation of some typical lines from their progenitors down to the present; (3) the extent to which this evolution may have been influenced by events in the history of the earth and of human accommodation to the environment; and (4) the uncertainties that mutations introduce in estimates of the times and rates of formation.[2]

The Growth Curves of Haplotype I – The Rates of Formation of Its Subclades

We first derive the RCC matrix and determine the percentage that is filled at selected time intervals from the present back to many tens of thousands of years ago. At dates long ago, the RCC matrix is completely filled, indicating that all testees were related to every other testee in the sample. The entire matrix is filled at an RCC of 673, a date that corresponds to 29,000 years ago (~27,000 BC). It is the most probable date when the

---

[1] The correlation approach is introduced and discussed in Howard 2009. The standard deviation of the date estimates in this analysis is covered in the Appendix. They are of the order of 40% when one pair's RCC leads to a date. If n pairs are used, the uncertainly decreases by the square root of (n-1).

[2] As long as the sample size is sufficiently large, the rate of change should be independent of sample size. Our sample size of 281 haplotypes should be sufficient for the formation rates we find.

progenitor of Hg I2 lived. This date also designates the origin of Haplogroup I.[3] Then, using a time slice matrix[4], we can study how the numbers of the most recent common ancestors (MRCAs) of pairs of testees diminish in the matrix as these relationships change from the past to the present.

The RCC matrix that results from the group of 281 Hg I shows the time distance between each pair of testees. The RCC value of each pair of testees is proportional to the time to their most recent ancestor (TMRCA). A time slice matrix allows the high RCC value to be slowly decreased from its highest observed value to zero. The percentage of the RCC matrix that is filled at various values of RCC in increments of 20 is shown in Figure 1. We note that the last value plotted at the lower right in Figure 1 is the percentage filled between RCC 0 and 20. This range is typical of pairs of testees whose MRCA can be most easily identified by the pedigrees of the members who appear between RCC 0 to 20 in the matrix.

Figure 1: The percentage of the RCC matrix filled as a function of RCC. Time progresses from left to right.
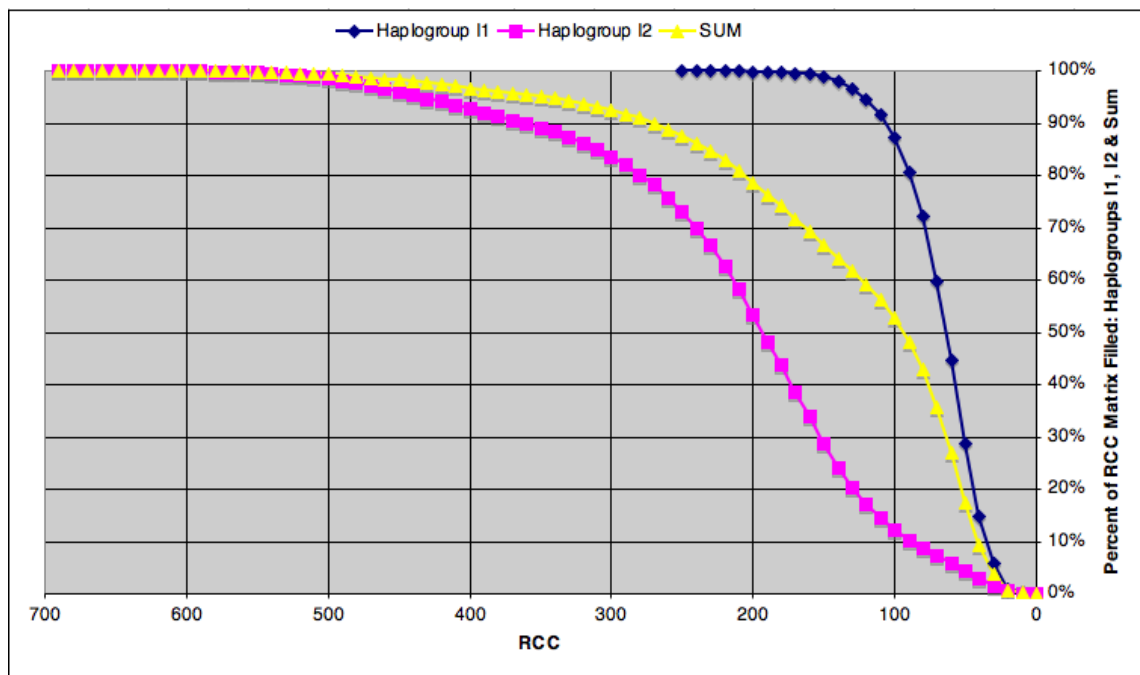

Figure 1 shows RCC as the abscissa, which has been calculated directly from the correlation of the haplotype pairs, prior to the conversion of RCC to time. The RCC time scale has been calibrated using over 100 RCC values derived from pedigrees where the corresponding time difference between a testee pair and their respective MRCAs is well-

---

[3] It can be argued that if there are more haplotypes in the sample, a higher value of RCC will be found, but when the sample is statistically as large as it is here, the RCC(max) value should be close to the date of origin as shown by a comparison of this date with others in Table 1.

[4] Howard 2009 op cit.

known (Howard 2009). Since a change of 10 in RCC equates to a time difference of about 433 years, we can derive the relationship in Figure 2.

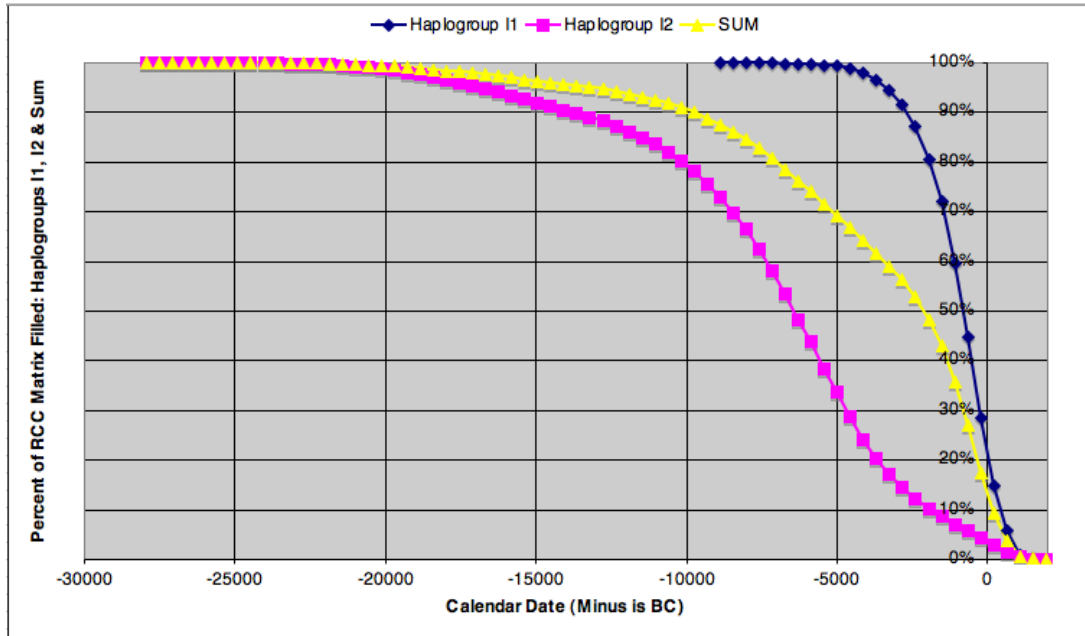Figure 2: The percentage of the RCC matrix filled as a function of time.

Figure 2 shows the <u>percentage</u> of pairs of haplotypes indicative of the I1 and I2 components of the Hg I population at the dates on the graph.

Figures 1 and 2 appear to have some fine structure, which can be further analyzed by deriving the <u>rate of change</u> of the rate of change of pair formation. Thus, Figures 3 and 4 show the rate of change that took place in Hgs I1 and I2, as functions of RCC and date, respectively.

Figure 3: The Rate of Change that took place in the formation of Haplogroups I1 and I2, as a function of RCC.
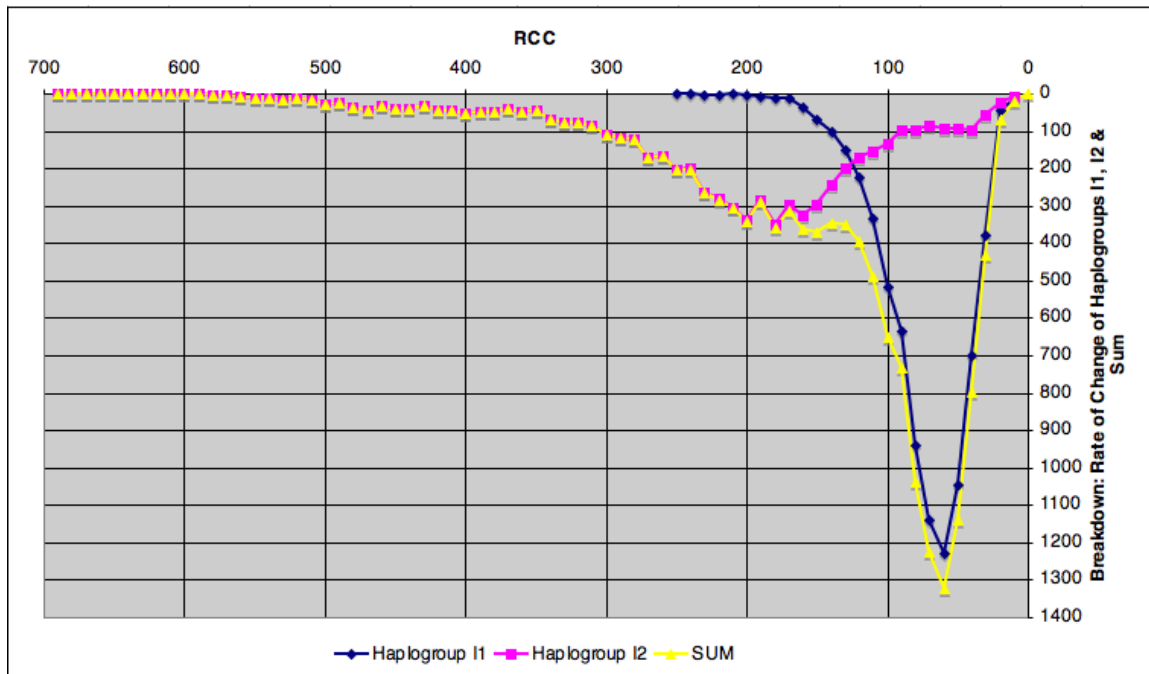
For example, the total number of pairs of I1 and I2 haplotypes was found to be 7619 and 6440, respectively. In Figures 3 and 4 we plot the difference in the numbers filled at each increment of 10 in RCC. For example, in Hg I1 we found that 4551 out of 7619 (59.7%) were filled at RCC 70, while at RCC 80, 5493 (72.1%) were filled. In Figure 4, the difference 942 is plotted against RCC 80.

Figure 4: The Rate of Change that took place in the formation of Haplogroups I1 and I2, as a function of date.
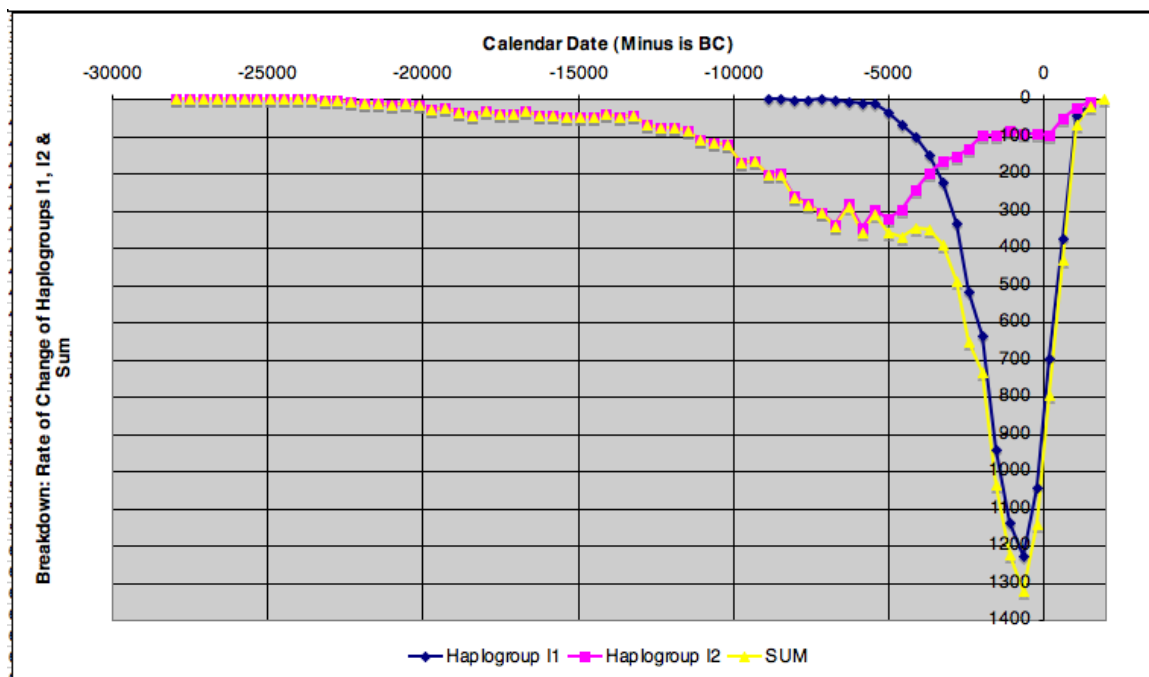
The previous figures show that Haplogroup I2 is the older of the two subclades. The oldest pairs in HgI2 can be used to point to the time when the progenitor of Haplogroup I probably lived. Figure 5 shows a histogram of the RCC matrix of HgI2, presented on a log plot. An extrapolation of the points at the high RCC end of that plot to the bottom of the chart shows that the progenitor lived somewhere between RCC 650 and 700. This value should be compared to the largest value of testee pairs found in the entire RCC matrix of Haplogroup I. That value, RCC (max) is at 673, the point found earlier when the RCC matrix is completely filled. The corresponding time is 29,000 years ago, or about 27,000 BC. The standard deviation of this time is estimated to be about 15 percent.

Figure 5: A Log Histogram Plot of the Frequency of Occurrence of RCC Values of Pairs of Testees in the RCC Matrix of HgI2 Testees
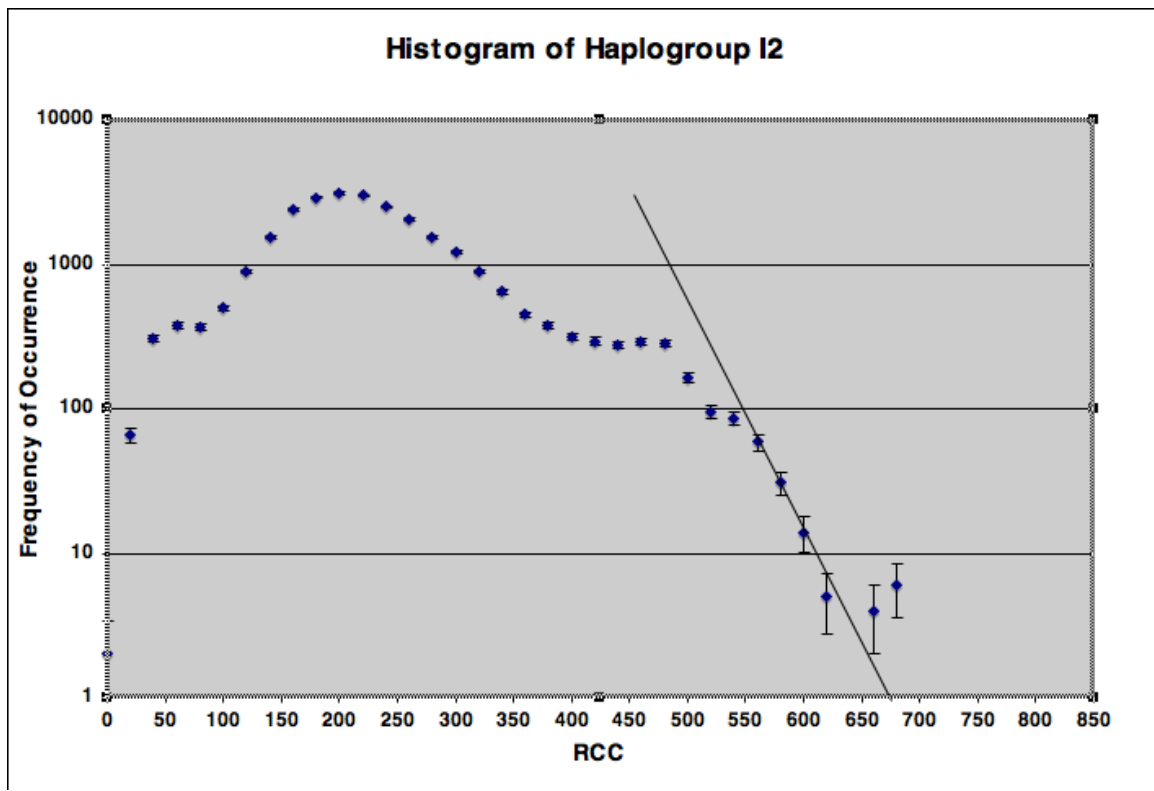
We can now summarize these findings in Table 1 where we compare them with those of other researchers.

Comparison of Dates of Origin for Haplogroup I and its Components I1 and I2

Table 1: Comparison of the dates of origin of Hg I and its components[5].

Reference[6]	Estimated Date of Origin: I	Estimated Date of Origin: I1	Estimated Date of Origin: I2
This paper	Probably coincident with origin of I2, 29,000 BC	8450 BC, with major velocity increase from 5000 BC to peak velocity at ~ 1000 BC	27,000 BC. Slow velocity to 20,000 BC, then major velocity increase from 13,000 BC to peak velocity at 6000 BC
ISOGG	26,000 BC	< 3000 BC – 1000 AD (I1b)	Close to time of origin of Hg I ~ 26,000 BC
Nordtvedt	~ 20,000 BC	~ 2500 BC	Close to time of origin of Hg I ~ 19,000 BC
Wikipedia	28,000 – 22,000 BC	6000 – 2000 BC	16,000 -- 13,000 BC
Europedia	~23,000 BC	8000 – 6000 BC	15,000 BC
Place of Origin (the bunching of locations is typical of recent origin of the haplogroup	Inferred Hg IJ clade in West Asia or Middle East ~ 38,000 – 33,000 BC	Scandinavia, Iceland, NW Europe	Caucasus, Central or SE Europe

One or more of the following situations may have caused our estimates of the time a progenitor lived to be somewhat older than the estimates derived by others:
- Sensitivity of the analysis. Variance analysis used by other researchers may have identified the TMRCA at a more recent time when many haplotype pairs were grouped whereas the RCC intercluster matrix can identify individual MRCAs.
- Unless an error has occurred in testing or unless many unexpected mutations have occurred, the pair of testees who have the largest value of RCC should be the best date indicator as long as numbers in our analysis are reasonably large. Even then, the progenitor may have lived even earlier.
- Since the Mathematica application optimizes the entire marker string of each testee's haplotype when it positions a testee on the phylogenetic tree, the earliest junction point on the tree will tend to be an underestimate of the time when the progenitor lived. We have to select the largest RCC found for the group as its time of origin. This high value of RCC (max) is located in the intercluster RCC matrix, not on the tree.[7]

---

[5] The dates estimated or derived in this paper have standard deviations of the order of 10% in the absence of unknown systematic errors. Thus, it would be rare to see errors over 30% unless they are systematic.
[6] See web page references at the end of this paper.
[7] The dated phylogenetic tree only shows the time of the earliest pair (N=2, not 1) since Mathematica uses

- The earliest RCC in the sample may point to too recent a date if lines of descent from earlier male ancestors have died out.

Comparison of the Rate of Change in the Formation of Haplogroup I and the Growth Rate of the World's Population

No definitive data are available for the world's population during the early time when the numbers of pairs of haplotypes within Haplogroup I were evolving[8]. Using the available data it is possible to derive estimates, and from those estimates derive a rough rate of change in the numbers of humans between 30,000 and 2000 BC. We then can express the result in mathematical form and derive the rates of change in population numbers at the times when pairs of Hg I were evolving. A comparison of those rates is shown in Figure 6.

Figure 6: A Comparison of the Rate of Change in the Formation of the Sum of the Two Haplotype I Subclades and the Growth of the World's Population.
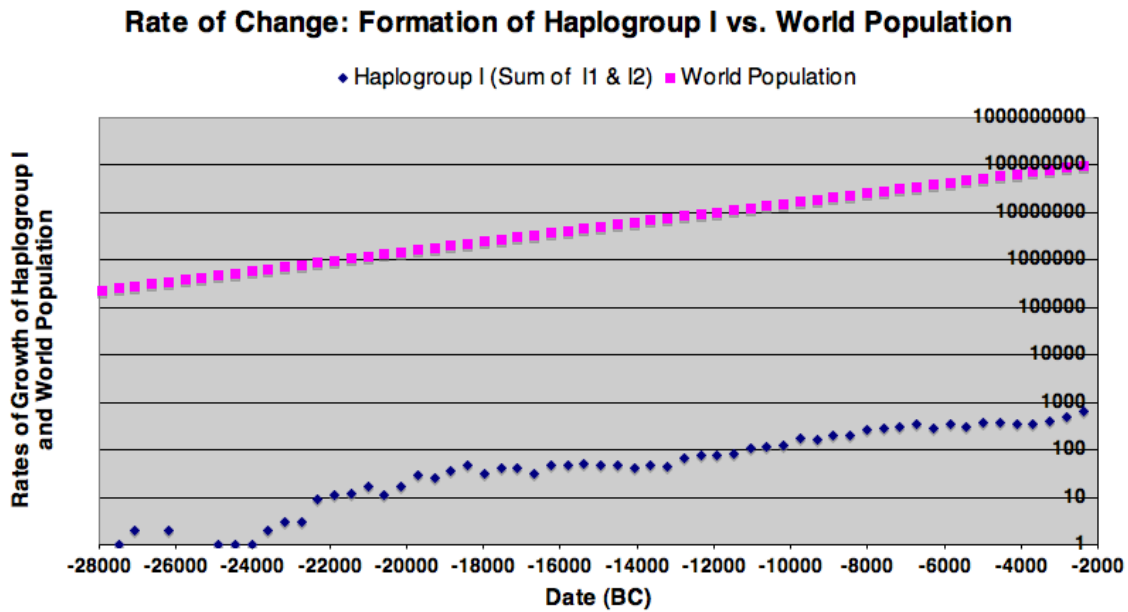
The parallel nature of the two relationships shown in Figure 6 indicates that pairs of Hg I and the world's population were growing at comparable but slow, exponential rates. In the absence of reasons why they should be different, this result is not surprising, but it

---

the 'average' option. This option was selected because it shows values that are consistent with the averages of the RCC intercluster matrices. Because the tree presents averages, it is not expected to show the times of the oldest pairs in the selection since those values are extremes and not averages. Thus we choose a combination of RCC (max) and the extrapolation of the histogram to estimate the age of the haplogroup.

[8] Four population estimates for this time period are given in the summary paper by Scott Manning (see web pages) who, in turn, refers to a more definitive paper by McEvedy and Jones (1978).

does show that the growth of at least this haplogroup tracked the growth of the general population.

Possible Relationships between the Dates of Haplogroup I and Events that Affected the Environment and Human Development:

Haplogroup I is distinctly European. It originated during or just after the last glacial maximum when other new haplogroups were also forming.

What do we know about history of mankind in this distant past and how might changes in man's environment and his accommodation to change correlate with the evolution of the two components of Hg I?[9]

Haplogroup I2 – The older component of Hg I.

The early dates on the graphs correspond to the late Stone Age, which lasted over two million years until 6000 BC -- 2500 BC in Europe, North Africa and Asia when the age of metals began. Our analysis indicates that the progenitor of Hg I2 lived at the end of the Stone Age, about 27,000 BC, at a time when ice sheets were covering the earth from latitudes 40-50 degrees to the pole. During that time humans were forced to migrate to lower, more hospitable latitudes where game was more plentiful but where human habitation barely survived at near the replacement level. Hg I2 carriers were nomadic hunter-gatherers in that era.

Figure 4 shows from the Y-DNA evidence that the numbers of Hg I2 males increased imperceptibly from that time until the end of the last glacial maximum near 19,000 -- 20,000 BC when their rate of change very slowly increased, but it remained relatively constant until about 13,000 BC. The graph clearly shows two periods in the evolution of Hg I2 when its population grew very slowly, viz., 27,000 BC to 20,000 BC, and 19,000 BC to 13,000 BC. The more recent of these periods coincided with the recession of the glaciers when the hunter-gatherers took many generations to move slowly to the north, following game. Over fewer generations the small numbers of people in tribes tended to remain in narrowly defined territories. Living conditions were still harsh and survival was difficult, permitting procreation at only the replacement level. The small numbers of males in these tribes probably contributed to genetic drift, a situation that probably was more important to genetic diversity than natural selection. By 10,000 BC, humans had spread to all habitable sections of the globe.

Between 13,000 BC and about 6000 BC our analysis shows that the rate of change in the formation of the Hg I2 population accelerated, reaching a peak about 6000 BC, just as the formational rate of change in the population of Hg I1 was beginning to increase. Earlier in this period, 11,000-7000 BC (the early Neolithic), farming began in warmer climates like parts of Turkey, Syria, Greece and India and progressed to northern climates around 5000-2000 BC (the later Neolithic) in Italy, Scotland and Ireland. During this time, when

---

[9] The chronological histories by Grun and Trager and the New Vistas reference have been valuable in linking dates derived in this Y-DNA analysis to events in history (see References and Web papers).

the number of Hg I2 carriers accelerated and when the progenitor of Hg I1 lived, hunter-gatherers were making the transition to farming. The oscillations in Figure 4 arguably show fluctuations in the evolution of Hg I2 carriers between 7000 BC and 5000 BC, but its causal tie to human history is not apparent. This is the time of the invention of the wheel and the domestication of cattle.

The density of present day carriers of Hg I2 peaks up in the central and Eastern European population, including the Balkans, Sardinia and the region between the Ukraine and Romania. Since Hg I2 has this earlier history, it is nearer the Y-DNA migration out of Africa and appears en route to Scandinavia and northern Europe where carriers of Hg I1 peaks up, reflecting their later migratory history.

By this time, the planting of crops had supplemented hunting. Animal domestication, better living conditions and more dependable and nutritious supplies of food permitted a growth in population that was not available to their ancestors. On the other hand, farmers were shorter than the hunter-gatherers; they suffered more from malnutrition and disease, resulting in high infant mortality.

While it is not the intent of this paper to analyze deeper subclades of Hg I1 and I2, some researchers point out that new lineages in Hg I2 occurred around 6000 BC. It is tempting to interpret the abrupt changes of evolutionary velocity in Figure 4 between 5000 and 7000 BC as a manifestation of the bursts of these new lineages.

Haplogroup I1 – The younger component of Hg I

The earliest pair of Hg I1 haplotypes is located in the RCC matrix at RCC 240, corresponding to a date of about 8450 BC. Figures 3 and 4 show the earliest signs of Hg I1 activity at 6000 BC, a time when humans lived in the broadly defined Mesolithic era. This era varies in definition from region to region from the Holocene period around 9500 BC, to the introduction of farming.

Farming began in warmer climates like parts of Turkey, Syria, Greece and India around 11,000-7000 BC in the early Neolithic and progressed to northern climates around 5000-2000 BC in the later Neolithic in Italy, Scotland and Ireland. During this time the number of I2 carriers accelerated as hunter-gatherers made the transition to farming. The progenitor of I1 lived toward the end of this period when farming was becoming well-developed. Once farming began, farmers began to crowd out the hunter-gatherers and the evolutionary velocity of the carriers of Hg I1 began to dominate the speed of formation of Hg I2 as shown by the cross-over point on Figure 4 between 5000 – 3000 BC. During the period in history when Hg I1 and Hg I2 carriers first co-existed, after about 5000 BC, migrations from southeastern Europe were taking place during the early introduction of farming.

Thus our analysis from the Y-DNA evidence indicates that the progenitor of Hg I1 lived toward the beginning of the period when farming was becoming more prevalent. The

major growth of Hg I1, shown by this analysis, took place more recently. The rate of formation increased about 5000 BC and peaked at about 1000 BC.

Conclusions:

Our dates for the origin of Haplogroup I and its subclades I1 and I2 agree with those of other researchers within their respective date uncertainties. This study has shown:
- First, that RCC(max) is a valid indicator of age,
- The date of origin determined by RCC(max) and the date derived by the extrapolation of the histogram are mutually supportive. We have seen from the plot of the log of the histogram vs. RCC that it indicates a date near RCC(max) when it is extrapolated to Log N=0. So the histogram appears to be a good check on the age of the sample we are using.
- Third, the spread of points where the extrapolation MIGHT hit on the Log N= 0 axis can be used give a reasonable idea of what the error bars should be on RCC(max). This point may be used unless there is a good reason to think that RCC(max) is a testing error.
- The slope in the growth of the number of pairs of Haplogroup I is the same as the growth of the world's population.

We believe that the principal contributions of this paper are:
1. the development of a method that uses correlation techniques and a time slice RCC matrix to derive:
   a. the date of origin for the sample of haplotypes we select and,
   b. the rate of formation of components of a haplogroup; and,
2. the analysis and investigation of relationships between the rate of haplotype growth and historical events that involve man's surroundings and his human development.
3. the uncertainties that mutations introduce in estimates of the times and rates of formation of haplogroup components.

Acknowledgements:

References:

Chandler, John F., Estimating Per-Locus Mutation Rates, Journal of Genetic Genealogy 2:27-33, 2006

Grun, Bernard, The Timetables of History: A Horizontal Linkage of People and Events, Simon and Schuster, New York, 1975 (Based on Werner Stein's Kulturfahrplan, 1946)

Howard, William E. III, The Use of Correlation Techniques for the Analysis of Pairs of Y-STR Haplotypes, Part 1: Rationale, Methodology and Genealogy Time Scale, Journal of Genetic Genealogy, 5,(2): 256-270, 2009.

McEvedy, Colin and Jones, Richard, Atlas of World Population History, Middlesex, England: Penguin, 1978.

Trager, James, The People's Chronology: A Year-by-Year Record of Human Events from Prehistory to the Present, Holt, Rinehart and Winston, New York, 1979.

Wolfram Research, Inc., Mathematica, Version 8.0, Champaign, IL., 2010.

Web Pages:

International Society of Genetic Genealogy (for Haplogroup I):
http://www.isogg.org/tree/ISOGG_HapgrpI11.html

Wikipedia Summary of Haplogroup I:
http://en.wikipedia.org/wiki/Haplogroup_I-M253_(Y-DNA)
http://en.wikipedia.org/wiki/Haplogroup_I2_(Y-DNA)

Eupedia Summary of Haplogroups I1 and I2:
http://www.eupedia.com/europe/Haplogroup_I1_Y-DNA.shtml
http://www.eupedia.com/europe/origins_haplogroups_europe.shtml#I1

Manning, Scott (2008):
http://www.scottmanning.com/content/year-by-year-world-population-estimates/
See also: Durand, J. "A Long-Range View of World Population Growth." The Annals of The American Academy of Political and Social Science: World Population 369 (1967).
https://spreadsheets.google.com/pub?key=pbb0aoD3hdM-HgD-knTGXIA

New Vistas on the Distant Past:
http://www.buildinghistory.org/distantpast/peoplingeurope.shtml
http://www.buildinghistory.org/distantpast/haplogroupi.shtml

Ken Nordtvedt's Web site (Includes his tree for Haplogroup I):
http://knordtvedt.home.bresnan.net/

Wolfram Research:
http://en.wikipedia.org/wiki/Mathematica
http://www.wolfram.com/mathematica

Submitted to the Journal of Genetic Genealogy on 27 July 2012

Dr. William E. Howard III
1653 Quail Hollow Court
McLean, VA 22101-3234
703 532-8975; Email: wehoward@post.harvard.edu; Skype: wehowardiii