Human yDNA Ethnographic and Genographic Atlas: Open-Source Data Modeling and Compilation

Adam Cherson
Greencore Environmental Information Services
New York, NY

ABSTRACT

This paper presents an open source web portal of human genography: the Atlas of Ancestral Genography (http://atlas.xyvy.info). A dataset compilation has been prepared from yDNA haplogrouping research published in the human genetics literature. The compiled dataset has been modeled through ethnographic and demographic statistics to create a unique global atlas of human yDNA genography. The portal enables exploration of the prepared data by means of various interactive charts, maps, graphs, and visualizations. Onsite modeling of many well known sociopolitical, economic, and ecological indicators, as functions of yDNA haplogroupings, is also available. A new statistic, the Haplogroup Diversity Index, is modeled and presented. Researchers may download the raw data in CSV format for offsite modeling.

INTRODUCTION

The Atlas of Ancestral Genography is founded upon the work of human geneticists developing the field of human haplogrouping since the early 1990s (Table S1; http://www.xyvy.info/TableS1.xls). These trailblazers have been criss-crossing the globe collecting gender-linked DNA samples, analyzing single-nucleotide polymorphisms

(SNPs) and single tandem repeats (STRs), and gradually perfecting a categorization system for the results. Together with this labor of data gathering, human geneticists have constructed a robust theory of human ancestry based on gender-linked chromosome mutations. According to this theory, any human being alive today, or at any other time, can be categorized according to mutations on the two gender-linked chromosomes [1]. These chromosomes, the mtDNA and the yDNA chromosomes, pass largely intact from parents to offspring going back to the very first human mother and father that ever lived. In the course of time, however, mutations have arisen on these gender-linked chromosomes; the presence of this or that mutation defines the ancestral branch of the family tree from which any given individual stems [2]. Standardized data libraries of these mutations together with a letter based classification system, known as the human haplogroup system, have emerged, one for mtDNA [3] and the other for yDNA [4], allowing geneticists across the globe to identify any person's gender-linked genes according to the haplogroup system.

The following example serves to illustrate the theory and practice of haplogroup classification. Supposing about 54,000 years ago there was a male human living his life somewhere along the Rift Valley of Eastern Africa and one of this individual's sons is born having the M60 mutation on the y-chromosome. The label M60 is one of many generally accepted ISOGG [4] markers used to define membership in one of these haplogroups. The M60 marker in particular is used to define the B haplogroup. At the location defined as M60 this individual and all of his offspring share the same mutated genetic coding and will continue to do so until another mutation happens to shake things up at that precise spot. So any person determined to have the M60 y-chromosome

mutation is in the B haplogroup and is ipso facto a direct male descendent of the first M60 mutant. To continue with the illustration, supposing that 5,000 years ago, an M60 individual produces a son having the M236 mutation in addition to the M60, and so now all the male descendants from that 5,000 year old son have both the M60 and the M236 mutations. A new branch, or sub-clade, is formed and any person with these two mutations is considered to be a member of the B1 haplogroup. This process of mutation and sub-cladding continues over time producing extensive haplogroup trees. The B line for instance goes as far as B2b1a1b1 as of this writing [4]. A similar historical narrative accompanies the history of mtDNA mutations.[1]

Human populations at any scale, from local to global, can be sampled for their haplogroup proportions (e.g., A-65%, G-30%. R-5%); the structure of these proportions, together with any deductive interpretations which logically follow, provide an objective account of the genetic, demographic, migrational, political, and other forms of human history, from the dawn of the species until the present. There are many such studies in the published literature, only a few of which are cited here for illustration [5,6,7,8]. The Atlas of Ancestral Genography has been developed to allow researchers from all fields to access and model this body of knowledge easily and quickly.

The presentation of haplogroup data in the Atlas is organized in several ways. First, the data may be viewed as national data, indexed by country. Second, they may be viewed as ethnic data, indexed by ethnic clusters such as Aborigine or Zhuang. Third, the data are viewable directly by haplogroup. Fourth, the data are presented in relation to various well known environmental and socioeconomic indicators, indexed by indicator.

[1] The Atlas currently offers yDNA data. Future plans include the development of a comparable mtDNA derived Atlas along with tools allowing for comparative explorations of each gender-linked dataset.

Fifth, the raw data set is available for download, allowing for scholarly use with statistical software. The entire list of visualization models is indexed for easy reference (Table S2; http://www.xyvy.info/TableS2.xls). Instructional videos explaining the use of each visualization are available by hyperlink from the portal's start page.[2] With these haplogrouping tools at their disposal, researchers may explore and model populations according to their needs.

METHODS

Demographic information for each of the world's nations was compiled from various public databases [9,10,11]. The demographic compilation identifies and quantifies distinct cultural, linguistic, and religious groups of people within each country and further agglomerates these people into broader 'people clusters'. Using these data, people cluster inventories for each nation were formed, starting with the most numerous people clusters within each country and continuing until at least 97% of that nation's total population was included in the cluster inventory. I will use the Afghani case to illustrate this process. The Pashtun from the largest people cluster in the country, 43.33% of the population. Following this group comes the Persian cluster at 26.98% and then the Uzbeks at 9.58%. I continue with this process until covering at least 97% of the total population. It should be emphasized that the broader people cluster level brings together numerous heterogeneous 'people'. The Persian people cluster in Afghanistan, for instance, consists of various people: the Qizilbash, Afghani and Tajiki Persians, non-Afghani and non-Tajiki Persians, Pahlavani, and Warduji. However, in this case as in

[2] Videos: https://www.youtube.com/user/Pecologix/videos

others, an extensive literature search revealed published data for only one of the specific people within the Persian cluster, the non-Afghani and non-Tajiki Persians. In such cases the best possible extrapolation from available data was used to stand for the entire cluster. In other cases, data for each of the people within the cluster were available. This process was repeated with each people cluster in each country, creating a unique haplogroup profile for each people cluster in each country. As discussed further below, the Atlas is actively engaged in the continuous improvement of the database by obtaining newly published studies for periodic update. The complete list of people clusters and population percentages used is available in the raw database file (Data S1; http://atlas.xyvy.info/data-files, $8). The complete list of individual peoples comprising each broader people cluster in each country is available in Supplementary Table S3; http://www.xyvy.info/TableS3.xls.

Database construction in this fashion revealed that some 1.5 billion people around the world live in 'non-homeland' countries (the term homeland is defined here simply as the country housing the largest population for that given cluster and makes no assertion regarding native, indigenous, or ancestral claims). For instance, 55.1 million Anglo-Celts live in the United Kingdom, while 44.9 Anglo-Celts live outside of the United Kingdom (Table S4; http://www.xyvy.info/TableS4.xls). In order to account for these diaspora populations, each national inventory was also made to include people clusters whose numbers within that country may not have arisen to a major proportion, but whose global diaspora populations are substantial. So returning to the Afghani case, the data shows that only about 76,000 people belonging to the Tajik people cluster live in Afghanistan, while the global Tajik diaspora numbers some 1.7 million Tajiks living outside the home

country of Tajikistan. Therefore, the small Tajik cluster is included in the Afghan inventory to assure coverage of the large Tajik diaspora in the results. Using this method of inclusion, the Atlas accounts for 99.6% of the world's population and 98.2% of the world's diasporatic populations. Table S5 (http://www.xyvy.info/TableS5.xls) summarizes the treatment of diaspora populations in the database.

After the selection of people clusters within each national inventory came the task of finding yDNA data in the published literature for each people cluster, on the basis of as many heterogeneous peoples within the cluster as could be found. It was often possible to locate specific reports for the heterogeneous peoples or people clusters. Occasionally more than one report was found for the same peoples or cluster. In some cases two or more reports where combined to form a larger, more accurate sample. In other cases the most recent and comprehensive among them was selected. In this way, individual statistics for each people cluster in each nation were created from the best, most granular data available, and according to generally accepted statistical principles (Table S6; http://www.xyvy.info/TableS6.123).

As any researcher in this young field is aware, the past ten years has seen a variety of haplogroup naming systems. This nomenclature confusion has been largely resolved by the ISOGG standardization (the current Atlas adopts the 2011 ISOGG standards). However, when the lack of contemporary data (meaning older than about 10 years) required the use of earlier publications, the older data were converted from one of several prior naming systems into the current ISOGG standard. Following the creation and standardization of the statistics, the demographic and yDNA data were combined into

country by country summaries, providing a haplogroup census for each nation and people cluster as seen in the raw data file (Data S1; http://atlas.xyvy.info/data-files, $8).

Using the prepared data, various analytical models were created and are displayed graphically within the portal. One such model is the Haplogroup Diversity Index (Figure 1). This index applies the well known Shannon diversity formula to the prepared haplogroup-demographic data, replacing the traditional species population statistics used by ecologists with the haplogroup population statistics prepared for this study. Using an interactive map, users may observe the haplogroup diversity statistic for each of the nations covered by the Atlas. Other available models include the well known Human Development Index [12], the Environmental Protection Index [13], and the Peace Index [14], each presented according to haplogroup results.

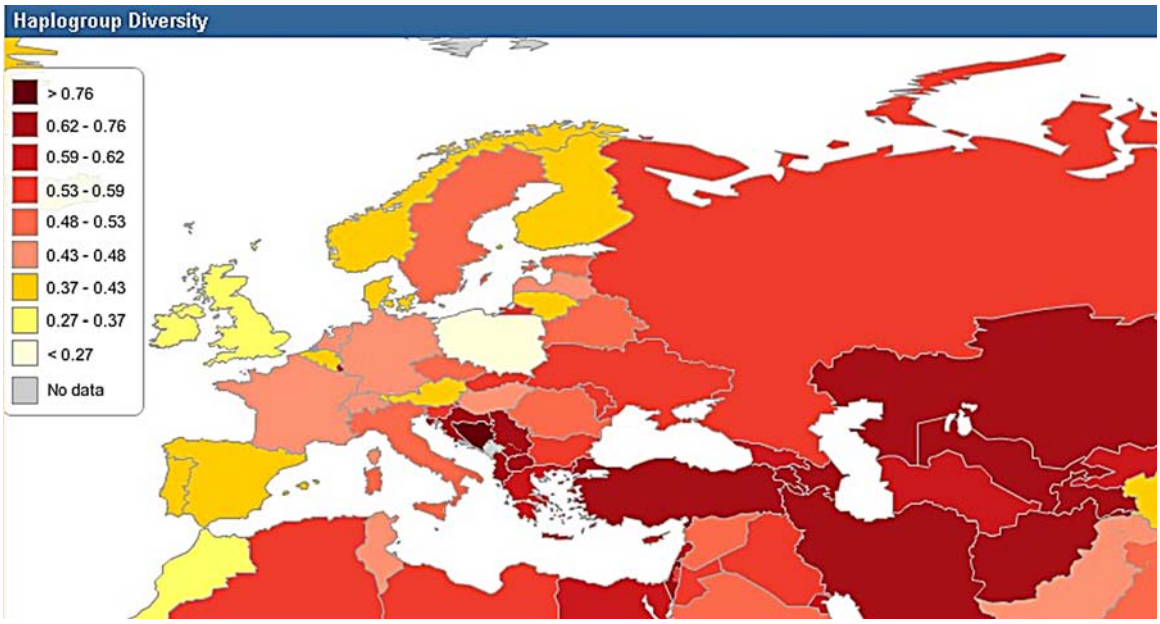

Figure 1: Interactive Map Displaying European Haplogroup Diversity (charts are active only online)

RESULTS

Ethnic Views

The Atlas includes haplogroup data for 249 ethnic groups, accounting for 99.6% of world population (Table S7; http://www.xyvy.info/TableS7.xls). Users may view haplogroup charts for each of these individually as shown in Figure 2.

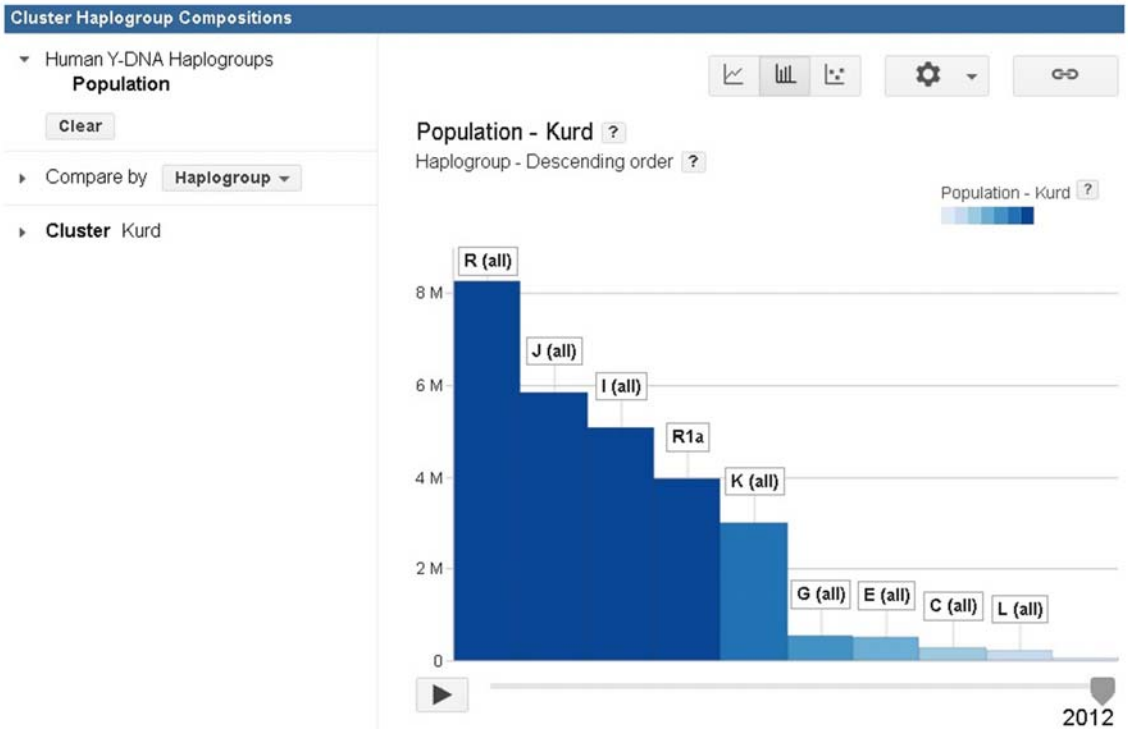

Figure 2: Visualization Showing Kurdish Haplogroup Composition (charts are active only online)

National Views

The Atlas includes haplogroup data for 123 world nations, representing all the nations of the world for which this type of data are available (Table S8;

http://www.xyvy.info/TableS8.xls). Users may view haplogroup charts for each of these individually as shown in Figure 3.

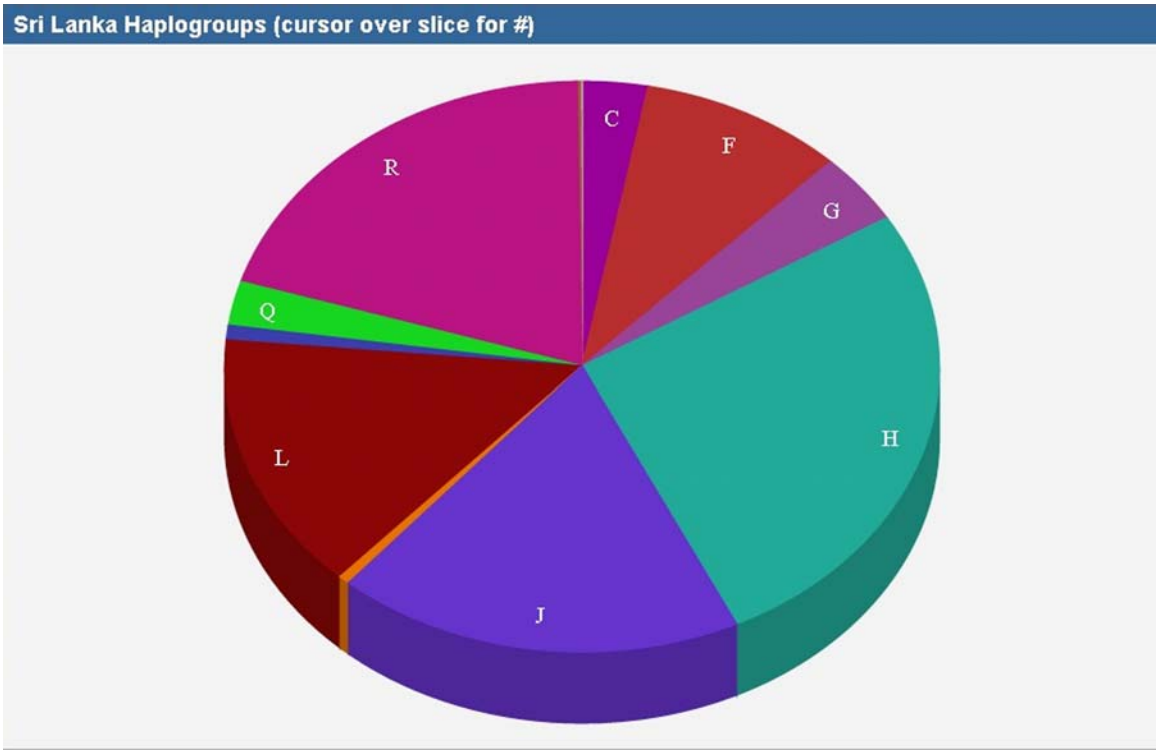

Figure 3: Pie Chart Showing Sri Lankan Haplogroup Distribution (charts are active only online)

Haplogroup Views

The Atlas includes various maps allowing views of the world according to haplogroups (Table S9; http://www.xyvy.info/TableS9.xls) . One such view allows users to view global distributions of each haplogroup individually as shown in Figure 4.

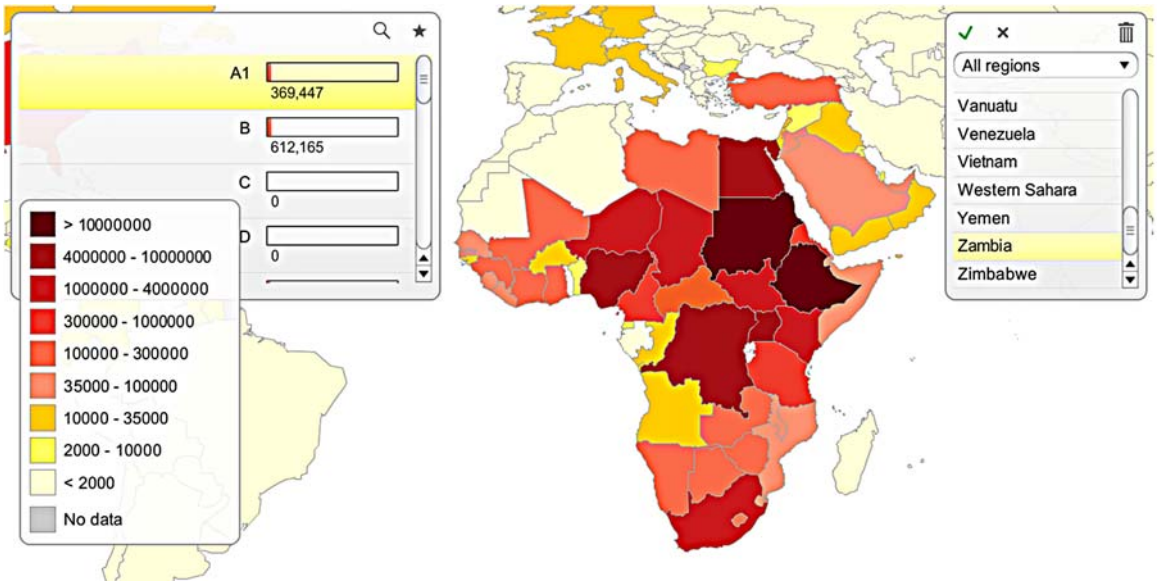

Figure 4: Map Showing Distribution of Haplogroup A1 in Africa (charts are active only online)

Environmental Modeling Views

As an example of modeling uses of the haplogroup data, the Atlas includes a series of visualizations depicting models of environmental genography according to haplogroups (Table S10; http://www.xyvy.info/TableS10.xls). One such visualization shows each haplogroups percentage share of global population and percentage share of land area (Figure 5).

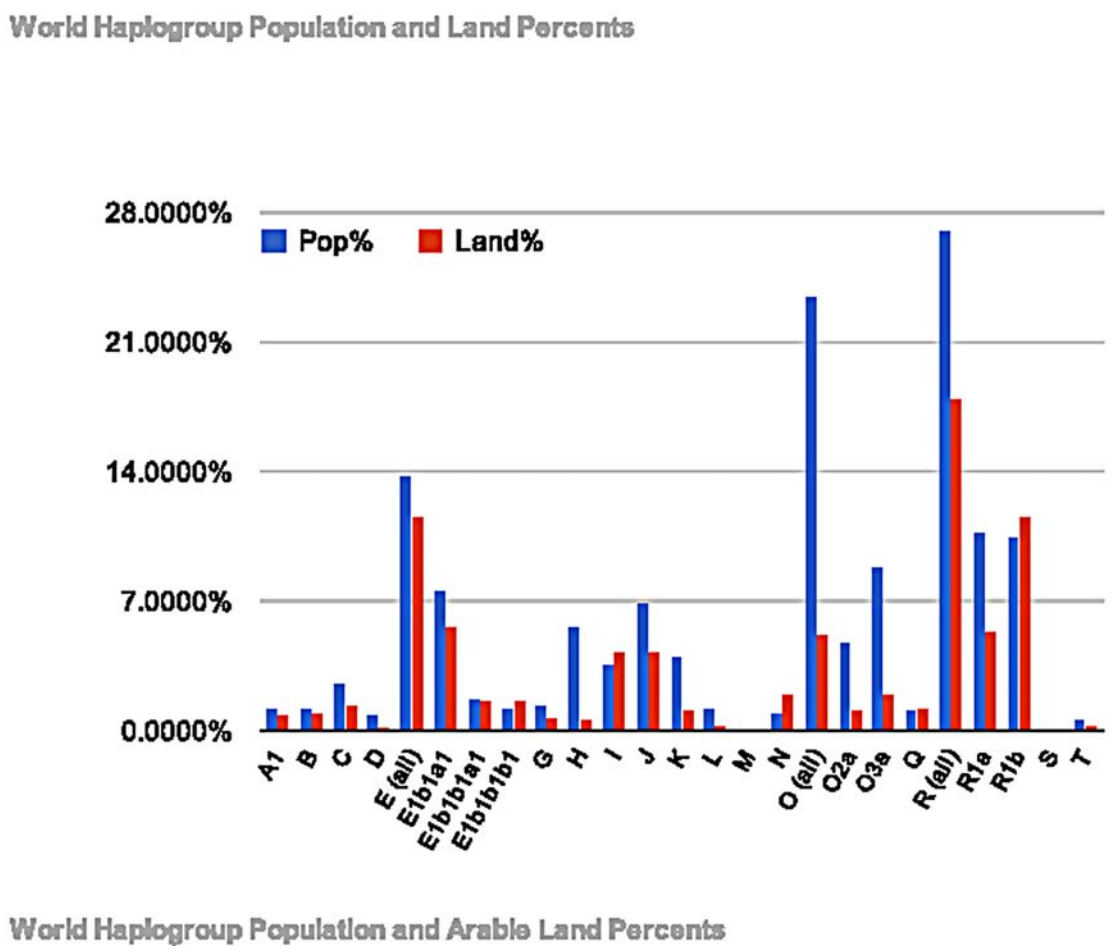

Figure 5: Visualization Showing Global Population % and Land % For Each Haplogroup (charts are active online)

Interactive Socioeconomic Maps

The Atlas offers a highly interactive map application allowing the user to view any one of a series of important indicators and indices according to haplogroup (Table S11; http://www.xyvy.info/TableS11.xls). So for instance, the user may see a map of the world showing the comparative ecological footprints of nations where the O2a haplogroup is dominant (Figure 6).

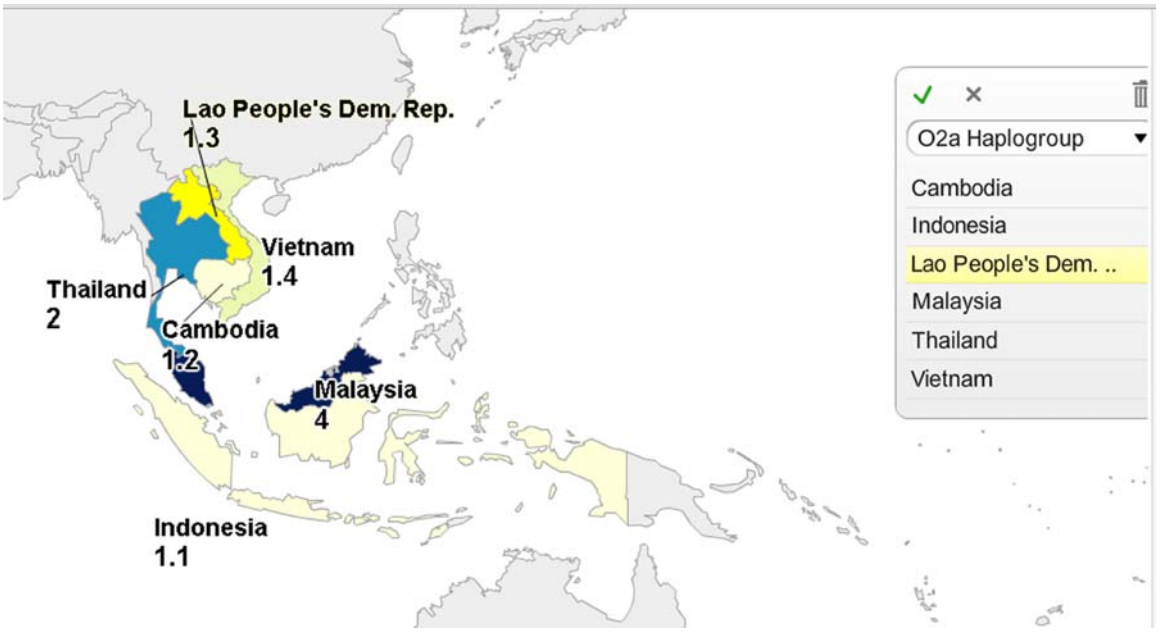

Figure 6: Ecological Footprints of O2a Haplogroup Nations (charts are active only online)

DISCUSSION

An initial review of the maps presented in the Atlas leads to the inescapable conclusion that human history thus far has fostered the creation of homogenous haplogroup populations rather than populations of randomly mixed up haplogroups. To contextualize this with our above example, B-M60 individuals tend to live in populations with other B-M60 individuals. Although each of these B-M60 individuals is unique with respect to their full set of 23 chromosomes, as far as their gender-linked chromosomes are concerned, they are exactly the same. So for instance if one takes a sampling of yDNA from Spain, one would soon discover that about 60% of all the males there share the R-M207 mutation. One would also find that about another 13% share the E-M96

mutation and about another 7.25% share the J-M304. Nearly 80% of males in Spain carry one of these three mutations.

A similar pattern is manifested in each nation of the world, such that in 66% of countries, a single haplogroup constitutes the majority of males living in that nation, while in 99% of countries, two haplogroups constitute a majority of the males living in that nation. This may seem obvious in places like Finland, where most of the people are what one would call Finns. But it also holds true for large, multi-ethnic, and diverse countries like India, China, the United Kingdom, Congo (Kinshasa), Egypt, and the USA. In fact, only in Sri Lanka and Uzbekistan, among the world's nations, do the two most common haplogroups within the country fall short of constituting a majority. Results such as these indicate the importance of gender-linked ancestry to an understanding not only of ancient human history, but also of recent and current sociopolitical circumstances.

The Atlas has conceivable applications in many fields of study: Anthropology, Archaeology, Political Science, Medicine, Criminology, Psychology, Sociology, Epidemiology, Environmental History, to name a few. A good example of haplogroup modeling comes from the latter. Using the data offered by the Atlas it is possible to model, for example, the amount of arable land forest land, and fresh water controlled by each haplogroup. The results of this modeling show that certain genetic branches of the human family tree have essentially been expanding their use and possession of the Earth's natural resources while other branches have seen their portions stagnating or dwindling. At this early stage of haplogroup research, the results of this study give rise to numerous questions for further inquiry: Could this mechanism, whereby one or more genetic branches enjoys access to natural resources while excluding genetic cousins from

these same resources, be one of the main forces driving all of human history? Could a vestigial attachment to this driving force, which may have been adaptive under earlier historical circumstances, now be eclipsing modern day impulses to cooperate for the good of the common environment or other common goods? Could this driving force be manifested superficially in other ways: as political ideology, religious fervor, real-economik or real-politik, or any number of other socio-politically recognizable constructs. These lines of inquiry suggest that a deeper understanding of the relationship between the forces underlying human haplogroup genography and current sociopolitical behaviors could be an essential component of sustainable human survival. Despite the vast potentials of haplogroup data modeling, researchers must remain cognizant of the risks, assumptions, and potential errors of interpretation that accompany such an endeavor, especially one depending on compiled data (Appendix S1). Furthermore, all researchers and observers should become acquainted with the special risks inherent in the study of human genography (Appendix S2).

The Atlas data portal includes several ancillary applications of high value. First, it serves as an aggregator of continuing haplogroup research by means of an RSS stream[3] delivering the latest published research in the field; the research aggregator is accessible from the portal as well as by email alerts[4], facebook[5], and twitter feed[6]. Second, the Atlas serves as an active clearinghouse of data and data gaps that researchers may refer to when designing their own research agendas in coming years. The comprehensive data-gap list includes the names of populations that have yet to be sampled or have been only

[3] RSS Feed: http://nature.xyvy.info/category/genomics/feed/
[4] Email Alerts: http://feedburner.google.com/fb/a/mailverify?uri=PclgGenomics&loc=en_US
[5] Facebook: https://www.facebook.com/geneticancestrygenography
[6] Twitter: @GeneticAncestry

minimally sampled as well as suggestions for deeper haplogroup structure research. As such it becomes a compendium of what is already known, how well it is known, and of what still needs to be known (Table S12; Table S13; http://www.xyvy.info/TableS12.xls; http://www.xyvy.info/TableS13.xls). The published literature in this field, while not yet complete, is sufficiently mature to support meaningful and robust data modeling projects. As the research data matures further, the Atlas can continue to improve as well.

CONCLUSION

The Atlas of Ancestral Genography represents a unique compilation of human yDNA haplogroup genography and ethnography, available for online visualization and modeling to scholars, researchers, genealogists, and the general public from anywhere in the world. The data offered by this portal open up an unexplored resource to scholars in a variety of fields, allowing for a panoply of new insights into human behavior and social history.

REFERENCES

1. Jobling, M, et al. (2003) The human Y chromosome: an evolutionary marker comes of age. Nature Reviews Genetics 4: 598-612; Berniell-Lee, G, et al. (2007) SNPlexing the human Y-chromosome: A single-assay system for major haplogroup screening. Electrophoresis 28: 3201-3206.

2. Sykes, B (2002) The Seven Daughters of Eve. New York: W.W. Norton & Company. 320p.; Olson, S (2003) Mapping Human History: Genes, Race, and Our Common Origins. New York: Mariner Books. 304p.

3. van Oven M, et al. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2): E386-E394. Available: http://www.phylotree.org. Accessed: 2 April 2014.

4. International Society of Genetic Genealogy (2014) YDNA Haplogroup Tree 2011, Version: 6.105. Available: http://www.isogg.org. Accessed: 2 April 2014.

5. Battaglia V, et al. (2008) Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. European Journal of Human Genetics 17: 820-830.

6. Abu-Amero KK, et al. (2009) Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC Genetics 10: 59.

7. de Filippo C, et al. (2010) Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. Am J Phys Anthropol. 141(3): 382-94.

8. Karafet TM, et al. (2010) Major EastWest Division Underlies Y Chromosome Stratification across Indonesia. Mol. Biol. Evol. 27(8):1833-1844.

9. Lewis, MP, Simons, GF, and Fennig, CD (eds.) (2013) Ethnologue: Languages of the World, Seventeenth edition. Dallas, Texas: SIL International. Available: http://www.ethnologue.com. Accessed: 2 April 2014.

10. Central Intelligence Agency (2013-14) The World Factbook, 2011 Version. Washington, DC: Central Intelligence Agency. Available: https://www.cia.gov/library/publications/the-world-factbook/index.html. Accessed: 2 April 2014.

11. Joshua Project (2014) Datasets, 2011 Version. Colorado Springs, CO: Joshua Project. Available: http://joshuaproject.net/resources/datasets. Accessed: 2 April 2014.

12. United Nations Development Programme (2014) Human Development Index, 2011 Version. New York: United Nations Development Programme. Available: https://data.undp.org/dataset/Human-Development-Index-HDI-value/8ruz-shxu. Accessed: 2 April 2014.

13. Hsu, A, et al. (2014) The Environmental Performance Index, 2011 Version. New Haven, CT: Yale Center for Environmental Law and Policy. Available: http://www.epi.yale.edu. Accessed: 2 April 2014.

14. Institute for Economics and Peace (2010) Global Peace Index. Available: http://economicsandpeace.org/research/iep-indices-data/global-peace-index. Accessed: 2 April 2014.

APPENDIX S1: ASSUMPTIONS, RISKS, AND POTENTIAL ERRORS

Haplogroup Data

Data Gaps: In some cases, specific sampling studies for a given people cluster were not found. In these cases a variety of methods were used based on the best available data. Sometimes this involved using a related cluster from a neighboring country whose population had been studied and reported. For instance, if data exists for Pygmys in the Central African Republic but none for Burundi, these data were transposed from one country to the other. In other cases, for lack of any known data, a haplogroup model covering a regional territory was used. For instance certain Bantu people clusters in central African countries are not reported and so a generic, central African, Bantu haplogroup identity was used. The specific sources and methodologies used for each cluster are described in Table S1. Lists of currently unknown haplogroup data are available in Tables S12 and S13.

Sample Sizes: Sample sizes vary from study to study. Some haplogroup information is reported on the basis of only a handful of individuals from the population and thus may present an inaccurate picture of the true identity of the population. The sample size, 'n', for people cluster calculations may be found in Table S6 or in the source literature listed in Table S1.

Founder Effects/Distortions: There is a presumption made in this study that haplogroup identity for a given people cluster remains intact through migration events. For example, once the haplogroup profile for, say, the Greek cluster in Greece has been determined, that same profile is used in any other country where a Greek cluster arises, unless specific Greek haplogroup data for that other country are available in the reported literature. The presumption therefore is that the Greek cluster population of Australia shares the same haplogroup profile with the Greek cluster populations in Greece or Mexico or anywhere else. This presumption may prove to be incorrect since it could well be the case that emigrants from, say, Greece did not migrate according to the same haplogroup proportions as are found in Greece. Cultural or economic factors in the land of origin may have lead to a higher proportion of certain sub-haplogroups migrating to another place and therefore founding a diaspora cluster exhibiting an entirely different proportion of haplotypes from that at home. In some cases, where available data allow, this type of 'founder effect' has been included in the Atlas. For instance, since most of the East Indian laborers brought by the British to the New World came from around Uttar Pradesh, India, data from that part of India were used when calculating for East Indian haplotypes in the former British colonies of the New World.

Lab Errors and Study Methodologies: Since no primary field research sampling of population haplogroups was done for this study, the results herein are entirely dependant on the accuracy of the reported scientific literature. Since a broad array of studies from many countries was used, some coming from as early as 1999, there exists the possibility of laboratory error in the reported results. Likewise, the Atlas is dependent upon the methodologies used in these studies to select sample subjects and to interpret the sampled

DNA data into haplogroup categories. When given the option to choose between studies covering the same people clusters, selection criteria are for the largest and most robust sampling methodologies. As for haplogroup classifications, wherever available haplotype allowed, the reported haplogroup designations were confirmed from the haplotypes and the ISOGG marker system (e.g., M60, P30, etc.). In studies using older classification systems, original classification results have been updated according to the ISOGG tree of 2011, attention being paid to go to no further granularity than the original data allows. So for instance, the Rosser (2000) classification scheme allows classification only to a certain level of granularity. When Rosser assigns a people cluster into his HG category 1, we can only interpret this as an R1b (and can reach no conclusion whether the type is R1b1 or R1b1b1, etc.

Varying Levels of Haplogroup Granularity: As laboratory methods and classification trees have evolved over the past ten years, a more nuanced picture of sub-haplogroups has emerged. Today, we have complex trees involving numerous nested sub-haplogroups. This means that not all studies reach the same level of classification granularity. For instance, certain studies may have determined that an individual haplotype belongs to the E1 group while another may take its genetic testing all the way to the E1a1a1a1b level. The latter result means that the individual exhibits the E1 mutation plus other mutations down to that sub-level, but the former result could also be an E1a1a1a1b—it was simply not tested down to that level. What this means is that comparison of sub-haplogrouping between studies is not possible using current data. In spite of this reality, it is still possible to analyze the data at the major haplogroup level: in the above example, both studies concur in the E and E1 mutations for these individuals. Therefore, there is enough cladistic consistency across studies to produce a robust, global result at the major haplogroup or major sub-haplogroup level. As more testing is reported, greater data minor sub-haplogroup resolution will become possible.

Ecosocionomic Analyses

Erroneous Association of yDNA with Race or Ethnicity: yDNA identity is not to be misconstrued as a surrogate for 'race' or ethnicity. Gender-linked markers are a completely distinct form of classification. One need only remember that the sex chromosomes form only a small fraction of the complete human genome and, as far as is currently known, have no impact directly on physiognomy. To understand this, consider a person whose ancestral grandfather was a central African man of the B haplogroup. This person mates with an Ethiopian woman to produce a son. Now imagine that this son and all future grandsons follow the same pattern (B haplotype son mates with Ethiopian woman). After countless generations a male descendant of this line looks like an Ethiopian (rather than a central African), but his yDNA haplogroup remains the B type prevalent in central Africa. This same scenario could be repeated between members of any racial or ethnic group such that any haplogroup could be theoretically exhibited by members of any race or ethnicity. For the same reason, members of the same 'race' or ethnicity could belong to different ancestral haplogroups and often do. One need only observe haplogroup diversity in any of the people clusters portrayed in this study to see the mechanism in action. People whom we could consider Scandinavian-looking, for

instance, are often members of the I or the R or the N haplogroups. As mentioned earlier, the primary objective of this study is the exploration of associations between gender-linked ancestral markers and complex ecosocionomic indicators.

Indicators and Values Equally Distributed Within Nations: When the haplogroup is used as the organizational principle for socioeconomic analysis, the presumption is being made that all haplogroups within a national group are proportionally equal participants in the underlying socioeconomic processes. Thus, if we say that the US, as a whole, has an ecological deficit of X and we know that 15% of the population in the US consists of haplogroup K, then we assume that 15% of X can be credited to haplogroup K. This type of presumption is commonplace in the use of 'per capita' socioeconomic analysis and this study is no different in this regard.

Non-Penetrance of Gender Linked DNA: One of the hypotheses explored by this work is the relationship between ancestral haplogrouping statistics and other ecosocionomic statistics. The presumption is that gender-linked DNA is somehow able to 'penetrate' through to the level of phenotype, social structure, or behavior, either directly or indirectly. The null hypothesis is that gender-linked genes simply determine gender and then have no other impact on human attributes. There are studies beginning to appear that explore the validity of this hypothesis as well as the relationships between genes and human behaviors generally:

Association of gender-linked genes and immunity:

> Yousefi S, et al. (2008) Catapult-like release of mitochondrial DNA by eosinophils contributes to antibacterial defense. Nature Medicine 14: 949-953. doi:10.1038/nm.1855

Association of gender-linked genes and specific neurons:

> Kimura K, et al. (2008) Fruitless and Doublesex Coordinate to Generate Male-Specific Neurons that Can Initiate Courtship. Neuron 59(5): 759-769. doi:10.1016/j.neuron.2008.06.007

Association of gender-linked genes and HLA typing:

> Thorsby E (2012) The Polynesian gene pool: an early contribution by Amerindians to Easter Island. Phil. Trans. R. Soc. B 367(1590): 812-819. doi: 10.1098/rstb.2011.0319

Association of gender-linked genes and coronary artery disease:

> Mearns BM (2012) Coronary artery disease: Y chromosome link to CAD risk. Nature Reviews Cardiology 9: 187. doi:10.1038/nrcardio.2012.24

Association of genes with anthropology and genography:

Novembre J, et al. (2008) Genes mirror geography within Europe. Nature 456: 98-101. doi:10.1038/nature07331

Association of cultural traits and socioeconomic indicators:

Gorodnichenko Y, Roland G (2011) Individualism, innovation, and long-run growth. PNAS 108(Supp. 4): 21316-21319. doi: 10.1073/pnas.1101933108.

Association of genes with behavior and political orientation:

Alford, J, et al. (2005) Are Political Orientations Genetically Transmitted? American Political Science Review 99(2):153-167.

Arceneaux, K, et al. (2012) The genetic basis of political sophistication. Twin Res Hum Genet 15(1):34-41.

Bouchard, T, McGue, M (2003) Genetic and Environmental Influences on Human Psychological Differences. Journal of Neurobiology 54(1):4-45.

Fowler, J, Schreiber, D (2008) Biology, Politics, and the Emerging Science of Human Nature. Science 322(5903): 912-914.

Fowler, JH, et al. (2008) Genetic Variation in Political Behavior. American Political Science Review 102(2): 233-248.

Hatemi, P, et al. (2009) Genetic and Environmental Transmission of Political Attitudes Over the Life Time." Journal of Politics 71(3): 1141-1156.

Hatemi, P, et al. (2010) Not by Twins Alone: Using Extended Family Design to Investigate Genetic Influence on Political Beliefs." American Journal of Political Science 54(3):798-814.

Hatemi, PK, et al. (2011) Integrating social science and genetics: News from the political front. Biodemography and social biology 57(1): 67-87.

Hatemi, PK, et al. (2009) Is There a Party in Your Genes. Political Research Quarterly 62(3): 584-600.

Hatemi, PK, McDermott, R (2012) The genetics of politics: discovery, challenges, and progress. Trends Genet .

Hatemi, PK, et al. (2012) Introduction: What is a `gene' and why does it matter for political science? Journal of Theoretical Politics 24(3): 305-327.

Klemmensen, R, et al. (2012) The genetics of political participation, civic duty, and political efficacy across cultures: Denmark and the United States. Journal of Theoretical Politics 24(3): 409-427.

Martin, N, et al. (1986) Transmission of Social Attitudes. Proceedings of the National Academy of Sciences 83: 4364-4368.

Oskarsson, S, et al. (2012) The genetic origins of the relationship between psychological traits and social trust. Twin Research and Human Genetics 15(1): 21-33.

Smith, K, et al. (2012) Biology, Ideology, and Epistemology: How Do We Know Political Attitudes Are Inherited and Why Should We Care?" American Journal of Political Science 56(1): 17-33.

Weber, C, et al. (2011) Genetics, Personality, and Group Identity." Social Science Quarterly 92(5): 1314-1337.

APPENDIX S2: GENOGRAPHY AND BIOLOGICAL DETERMINISM

The subject of genography and genetic ancestry often triggers uneasiness due to its association with racist genetics, eugenics, and biological determinism. As written by Gould in his 'Mismeasure of Man', the notion that "the social and economic differences between human groups—primarily races, classes, and sexes—arise from inherited, inborn distinctions and that society, in this sense, is an accurate reflection of biology…" is not scientifically sustainable. The study of environmental genography is in complete accord with Gould's verdict.

Genography does not purport to explain social differences on the basis of genetics. However, the fact that these associations are not biologically determined does not negate the fact that associations do exist between ancestral genetic haplogroups and certain contemporary aspects of society and culture. For instance, it is possible to examine access to, or use of, certain natural resources according to ancestral haplogroup. Similarly, it is possible to estimate the current population densities of the ancestral haplogroups. These are but two example applications of genographic data for social science and policy purposes. I think Gould's point is that these differences are not pre-determined; they are rather management choices.

Therefore, while genetic differences are not determinative of social and economic differences, looking at social data through the lens of ancestral genetics can provide important information as to the actual state of the distribution of many social goods across the globe as well as the morality of these distributions. Furthermore, there is no reason to reject the hypothesis that social and economic differences between human groups are the result of cultural differences and that these cultural differences are in turn associated with (albeit not caused by) ancestral genealogies. That these associations may be arbitrary and not biologically determined does not negate their existence nor the utility of studying the associations, nor does it deny the possibility that ancestral genetic haplogroup formations precede and are likely to be a causative factor in the divergence and geography of human cultures from prehistoric times until the present.