Paper 5.5: Haplogroup E Report.

Abstract: Haplogroup E-M96 represents a significant evolutionary marker for understanding the evolutionary history of populations in Mediterranean Europe, southeastern Europe, the Middle East, North Africa, and East Africa. Additionally, Haplogroup E-M96 represents almost all the Y-chromsome genetic diversity in Sub-Saharan Africa.

Within the E-M96 main haplogroup several variants stand as especially informative mutations for deciphering the correlation between linguistic and genetic diversity. E-U174 and E-U175 mutations carry the Bantu expansion southwards from West Central Africa. Proto-Berber and E-M81 co-expanded across North-Africa. E-V22 and E-M34 represent Afro-Asiatic agriculturalists that entered North and East Africa during the Neolithic. E-V12, E-M33, and E-M41 are genetic relics of pre-agricultural Nilo-Saharan populations. E-V32 represents the demic expansion of Nilo-Saharan and/or Afro-Asiatic pastoralists into East Africa. Pastoralism later expanded from this region with E-M293. E-V13 raises the possibility that some prehistoric Europeans may have spoken a proto-Afro-Asiatic language.

Posted September 2, 2017
Genetic-Linguistic Interface
Dr. Michael St. Clair
Stuttgart, Germany
mstclair@genlinginterface.com

Please cite as: St. Clair, Michael R. 2017d. “Paper 5.5: haplogroup E report.” *The Genetic-Linguistic Interface.* Series 5 Papers: language prehistory from a Y-Chromosome perspective. Posted November 16, 2017. https://genlinginterface.com

Table of Contents for Paper 5.5 (Haplogroup E).

SECTION	PAGE
Section 1. Overview.	1
Section 2. Evolution of Agriculture in Southwest Asia.	2
Section 3. Origins of Afro-Asiatic.	3
Section 4. Green Cluster Mutations.	4
Section 5. Red Cluster Mutations.	6
Section 6. The Purple Cluster E-V6 Mutation.	8
Section 7. The Blue Cluster.	8
Section 8. The Orange Cluster E-M33 Mutation.	9
Section 9. The Yellow Cluster E-M41 Mutation.	10
Section 10. Conclusions.	11
Bibliography for Paper and Tables.	11
Figure 5.5.1. Geographic Distribution of Afro-Asiatic Languages.	15
Figure 5.5.2. Overview of Afro-Asiatic and it Main Branches.	16
Figure 5.5.3. Overview of Linguistically Significant Haplogroup E Variants.	17
Figure 5.5.4. E-Z827 and the Green Cluster Mutations.	18
Figure 5.5.5. E-M78 and the Red Cluster Mutations.	19
Figure 5.5.6. E-V38 and the Blue Cluster Mutations.	20
Table 5.5.1. Survey of E-M81 Populations.	21
Table 5.5.2. Survey of E-M34 Populations.	23
Table 5.5.3. Survey of E-M293 Populations.	25
Table 5.5.4. Survey of E-V13 Populations.	26
Table 5.5.5. Survey of E-V22 Populations.	28
Table 5.5.6. Survey of E-V12 Populations.	30
Table 5.5.7. Survey of V65 Populations.	32
Table 5.5.8. Survey of E-V32 Populations.	33
Table 5.5.9. Survey of E-V6 Populations.	35
Table 5.5.10. Survey of E-M33 Populations.	36

Paper 5.5. Haplogroup E-M96.

The Genetic-Linguistic Interface.

Section 1. Overview.

The reader is invited to review Figure 5.1.1 from Paper 5.1. Both haplogroup E-M96 and haplogroup D-M174 diverged from DE-M145. According to Poznik et al. (2016: Supplementary Table 10), this occurred roughly 62 thousand years ago. Furthermore, as explained previously in Paper 5.4, Section 3, the DE-M145 mutation and its downstream variants, D-M174 and E-M96, have a unique Alu insertion (YAP) polymorphism. However, despite the phylogenetic closeness, the phylogeographic distribution of E-M96 and D-M174 are very much different. As explained in the previous Paper (5.4, Section 3), haplogroup D-M174 plays a rather modest role in representing the genetic diversity of East Asia. Moreover, within this region haplogroup D-M174 only represents a significant evolutionary marker for three populations: the Japanese, Tibetans, and Andaman Islanders. Haplogroup E-M96, on the other hand, represents a significant evolutionary marker for understanding the evolutionary history of populations in Mediterranean Europe, southeastern Europe, the Middle East, North Africa, and East Africa. Additionally, haplogroup E-M96 represents almost all the Y-chromosome genetic diversity in Sub-Saharan Africa, where over ninety-two percent of men have a variant of this haplogroup (Luis et al. 2004).

Among the geneticists (e.g. Abu-Amero et al. 2009) most support the position that haplogroups E-M96 and D-M174 evolved outside of Africa in the Middle East. Then, haplogroup D-M174 migrated to East Asia about 50 thousand years ago (see Paper 5.4 for more details). Haplogroup E-M96, on the other hand, "back-migrated" to Africa by around 56 thousand years ago (Poznik et al. 2016). Interestingly, some argue (e.g. ISOGG 2017) that haplogroup E-M96 evolved in Africa because almost all of the sub-haplogroups of E-M96 evolved on the African continent. However, as suggested by Poznik et al. (2016), a more "parsimonious interpretation" of the data places the origins of E-M96 in the Middle East because otherwise three haplogroups (D-M174, C-M130 and FR-M89) would have been part of the out of African migration, which seems inconsistent with the genetic evidence and archaeological record.

On the African continent, about 50 thousand years ago, diversification of haplogroup E-M96 began with the evolution of the E1-P147 and E2-M75 mutations. Since then haplogroup E-M96 has undergone extensive diversification producing what seems to be an extremely complex arrangement of phylogenetic relationships (see, for example, the ISOGG 2017 website). Indeed, among the eighteen main haplogroups listed at the bottom of Figure 5.1.1 from Paper 5.1, Haplogroup E-M96 has arguably the most complex internal phylogenetic structure of mutational variants. Thus in order to facilitate a discussion of linguistically significant E-M96 variants, our presentation of data for this haplogroup has been divided in six different "clusters" each with a color designation: orange, yellow, blue, red, green and purple (Sections 5.5.4 through 5.5.10).

The origins and expansion of languages in Africa seem to correlate well with the expansion of agriculture on this continent. The reader may recall that the herding of cattle in East Africa correlates well with Nilo-Saharan languages (see Paper 5.2, Hg. A, Section 3). Similarly, the cultivation of sorghum and millet in West Central Africa carries the history of Niger-Congo languages (see Paper 5.3, Hg. B, Section 4). In this present discussion of haplogroup E-M96 the reader now encounters another important language family on the African continent, languages classified as Afro-Asiatic. Arguably, Afro-Asiatic languages also co-expanded with agriculture like Nilo-Saharan and Niger-Congo. In the case of Afro-

Asiatic, the origin of the agriculture expansion is centered in Southwest Asia (or the Middle East). This agricultural expansion involved the cultivation of crops such wheat and barley, as well as the herding of goats and sheep. Since several haplogroup E mutations record the co-expansion of Afro-Asiatic languages and agriculture from Southwest Asia into Africa, Sections 2 and 3 (below) provides necessary background information that facilitates a discussion of the haplogroup E-M96 data.

Section 2. Evolution of Agriculture in Southwest Asia.

The 2005 book *First Farmers: the Origins of Agricultural Societies* by Peter Bellwood provides an excellent resource for linguists who wish to explore the worldwide correlation between the origins of agriculture and the expansion of languages. In chapter three of the book (pp. 44-66) he explores the origins of agriculture in Southwest Asia, focusing on a region often identified in the literature as the "Fertile Crescent." This region encompasses parts of contemporary Egypt, Israel, Jordan, Lebanon, Turkey, Syria, Iraq and Iran. The transition to agriculture in the Fertile Crescent was facilitated by the domestication of cereals such as wheat and barley, and legumes such as chickpeas and lentils, from wild sources. Additionally, the agricultural transition in the Fertile Crescent involved the domestication of goats and sheep. The success of agriculture in Southwest Asia partly stems from improved climatic conditions following the Last Ice Age. Another factor that ensured the success of this transformation was the development of pottery.

Prior to the adoption of agriculture in Southwest Asia, and elsewhere in the world for that matter, the human tribe practiced hunter-gather techniques in order to survive. The evolution of agricultural in Southwest Asia generally follows a series of cultural transitions that begin with the Natufians, followed by the Pre-Pottery Neolithic A and Pre-Pottery Neolithic B, and then finally the development of pottery itself. The Natufians stand as an important cultural transition because they were the last hunter-gatherers in the Middle East. According to Bellwood (2005) about 14.5 thousand years ago the Natufians appeared near the Sea of Galilee in what is now present-day Israel. Bar-Josef (1998) paints a picture of everyday Natufian life which centered on the hunting of gazelles and other animals. Moreover, and more significantly, he reports that they "practiced intensive and extensive harvesting of wild cereals" that grew abundantly in the region at the time. According to the description provided by Bellwood (2005) this abundant supply of food allowed the Natufians to construct semi-permanent settlements, something that is unusual for hunter-gatherers. These cultures are generally nomadic.

The Natufian thrived until about 13 thousand years ago when the Younger Dryas cold snap suddenly appeared. For a period of about seven hundred years, global temperatures sank considerably. Weather conditions in Southwest Asia became cold and arid, and with that the abundant supply of wild cereals disappeared. Once again the Natufians became nomads and ultimately disappeared from the archaeological record (see Blockley and Pinhasi 2011).

Then almost as suddenly as it began, the Younger Dryas ended and warmer weather returned. This created ideal climatic conditions that produced, once again, what must have been a seemingly inexhaustible abundance of wild cereals (Bar-Yosef 1998; Bellwood 2005). Amid this abundance, for reasons not entirely clear, a significant human innovation occurred. People began to domesticate the wild cereals and legumes that their Natufian ancestors had previously gathered. The Pre-Pottery Neolithic A culture stands as the initial Southwest

Asian culture that embraced this new development. They and their descendants thrived and by around 10.5 thousand years ago large farming settlements appeared such as the one at Abu Hureyra in northern Syria. This development signaled the evolution of another cultural transition in the region, the Pre-Pottery-Neolithic B culture. One of the significant innovations that occurred during this period was the development of pastoralism, the herding of goats and sheep, which were once wild animals that people had managed to domesticate.

About nine thousand years ago the development of pottery ushered in a new cultural transition in Southwest Asia. This development allowed people to cook their food more efficiently and facilitated the storage of grain after harvesting. Around this time the climate in Southwest Asia also became more arid. According to Bellwood (2005) this change in climate was accompanied by deforestation that human settlements had brought as well as less productive soil due to over-farming. These conditions caused many people in Southwest Asia to abandon sedentary crop agriculture. Instead of cultivating crops, some turned to sheep and goat herding as a food source. By around 6.4 thousand years ago some of these Southwest Asian pastoralists herded their goats and sheep out of the region into Egypt (Kuper and Kroepelin 2006).

Section 3. Origins of Afro-Asiatic.

The Afro-Asiatic language family contains 376 languages (Ethnologue 2017). These languages are distributed throughout the Middle East, as well as in North Africa, East Africa, and West Central Africa. Figure 5.5.1 provides an informative map prepared by Roger Blench. This map clearly shows that most of the diversification within the Afro-Asiatic language family has occurred within Africa. Figure 5.5.2 further illustrates this diversification according to the Ethnologue (2017) classification standard. As shown by the figure, Afro-Asiatic is subdivided into six main branches: Egyptian, Semitic, Chadic, Cushitic, Omotic, and Berber. As inferred by the present-day distribution of these six main branches, Semitic evolved in Southwest Asia, while Chadic, Cushitic, Omotic, and Berber evolved in Africa. This scenario, of course, assumes that the current distribution of Arabic follows the historical spread of Islam.

Long-standing opinion among linguists (e.g. Ehret 2004) places the prehistoric origins of Afro-Asiatic languages somewhere in East Africa. This opinion follows the idea that most of the diversification within Afro-Asiatic occurred in Africa (e.g. Hetzron 2009). However, Bellwood (2005: 207-210), based on his interpretation of the archaeological data, suggests that Afro-Asiatic languages initially evolved in Southwest Asia and co-expanded out of this region with the spread of agriculture. Interestingly, linguistic data may also support this model of Afro-Asiatic origins. Using linguistic reconstruction, Militarev (2002) presents a proto-Afro-Asiatic lexicon of farming terminology. Based on the reconstructions, he suggests that the Natufians, agriculture and Afro-Asiatic co-evolved in Southwest Asia. Finally, another reason for identifying Southwest Asia as the putative homeland of Afro-Asiatic languages is the Y-chromosome data as presented below in Sections 4 and 5 (below).

Section 4. Green Cluster Mutations.

This paper employs color clustering as a tool for explaining a Y-chromosome haplogroup with a rather complicated internal phylogeny. The reader is now invited to review Figure 5.5.3 which provides an overview of the six E-M96 color clusters that will be used. As shown by the figure, the E1b1-P2 mutation unites the blue, red, green and purple clusters. The blue cluster represents downstream variants of the E1b1a-V38 mutation that evolved in West Central Africa. The green, red and purple clusters, on the other hand, first evolved in East Africa. Trombetta (2015) suggest that the diversification of E-P2 into these west and east variants occurred around 48 thousand years ago.

Once again the reader's attention is directed to Figure 5.5.3. Note that the green, red and purple clusters evolved from E-M35, which evolved from E-M215, which evolved from E-P2. Trombetta et al. (2015) suggest that E-M35 arose in East Africa about 25 thousand years ago. This date is important as it provides time depth for the expansion of red and green cluster mutations out of East Africa into Egypt, and eventually into the Levant region of the Middle East. This second "out-of-Africa" migration probably followed the Nile River as it would have been an ideal corridor for human expansions (see Cruciani et al 2004; Luis et al. 2004; Cruciani et al. 2007; Cadenas et al. 2008). Note: the DR-M168 mutation represents the first out-of-Africa migration about 100 thousand years ago. See Section 1 (above).

Focusing now on the green cluster, Figure 5.5.4 reflects that mutations within this cluster are variants of the E-Z827 haplogroup. One of the downstream variants is the E-PF1961 mutation. An interesting study from 2016 (Lazaridis et al.) was able to extract three ancient DNA samples from a Natufian archaeological site in Israel. As the reader may recall from Section 2 (above), the Natufians were the last hunter-gathers of Southwest Asia. Two of the samples belong to E-PF1961, which is an ancestral marker for E-M34.

The E-M34 mutation has a wide distribution, currently found in populations of Mediterranean Europe, southeastern Europe, the Middle East, North Africa, and East Africa (see Table 5.5.2). For linguists, the E-M34 mutation is significant as it represents a potential back-to-Africa marker that co-expanded from Southwest Asia into East Africa with goat and sheep herders about six thousand years ago. This Neolithic migration, as the reader may recall from Section 3 (above), appears to have carried Afro-Asiatic languages from Southwest Asia to East and North Africa. The ancient Natufian Y-chromosome data, as just presented, supports this position because the E-PF1961mutation is an upstream marker from E-M34. Thus, E-M34 probably evolved in the Middle East. Additionally, analysis of contemporary genetic data also supports Middle Eastern and agricultural origins of Afro-Asiatic languages in East Africa. Cruciani et al. (2004) suggest that E-M34 arose in the Levant based on their interpretation of the data and wider distribution of this marker outside of Africa. Additionally, Cadenas et al. (2008) date the E-M123 mutation in the Middle East to about 11 thousand years ago, which potentially places the evolution of the downstream E-M34 mutation during the Southwest Asian Neolithic.

Another significant green cluster mutation for linguists is E-M81. The position of the mutation within the green cluster phylogeny (see Figure 5.5.4), and the contemporary distribution of the mutation (see Table 5.5.1), suggest that E-M81 arose somewhere in Northwest Africa. This stems from the observations that the mutation exhibits a clinal

frequency pattern across North Africa, with very low frequency among Egyptians, whereas the frequency climbs to around eighty percent among the Berbers of Morocco. In their 2004 analysis of the contemporary data Arredi et al. identify the E-M81mutation as a Neolithic marker. They suggest that goat and sheep pastoralism from Southwest Asia produced a "demic diffusion" of this mutation across North Africa and this expansion spread proto-Berber languages across the region. The term "demic diffusion" describes a scenario where a group adopts agriculture. This produces a sudden and rapid clinal population explosion across a region because agriculture supports far more people per square kilometer than hunter-gathering food economies. Y-chromosome mutations occasionally ride the coattails of such expansions. This explains why the E-M81 mutation has a low frequency in Egypt and a high frequency in Morocco.

The E-M81mutation consistently attains a high frequency among Berber populations (e.g. Bosch et al. 2001; Ennafaa et al. 2011; Fadhlaoui-Zid, et al. 2011; Trombetta et al. 2015). Accordingly this marker has become not only the genetic signature of the North African Neolithic but also the genetic signature of Berber languages. Since the Tuareg people of the Sahara desert speak languages classified within Berber branch of the Afro-Asiatic language family, researchers have regarded these nomads as descendants of the North African Berbers. The genetic evidence, and more specifically, the elevated frequency of E-M81 among the Tuareg, further suggests Berber origins for this population (see Pereira et al. 2010; Ottoni et al. 2011).

A report detailing ancient DNA having the E-M81 mutation was recently posted by Fregel et al. (2017) on the bioRxiv website. The samples come from two remains found at the Ifri n'Amr o'Moussa archaeological site in Morocco. The remains are about five thousand year old and as such, they relics of the Neolithic in North Africa. These data provide additionaly support for contemporary DNA studies that equate the E-M81 marker as the genetic signature of the North African Neolithic.

The researcher Roger Blench (2014) posted a rather interesting paper on his website that presents an anthropological and linguistic perspective of the Berber people and language. According to the paper, based on a comparison of grammar, Semitic is the closest Afro-Asiatic branch to Berber. Blench suggests that the Berber branch split from Afro-Asiatic language family around 6.5 thousand years ago. However, as he suggests, such a great time depth seem inconsistent with close linguistic similarities as found among the twenty-six contemporary Berber languages. Bench argues that a leveling of linguistic differences among the Berber languages occurred about two thousand years ago. This date is based on Neo-Punic and Latin lexical borrowings found in contemporary Berber languages. Blench suggests that the expansion of the Roman Empire into North Africa created a need for a *lingua franca* among the Berber. By this time the Berbers used camels and this brought an opportunity to trade with the Romans, especially along their southern frontier in North Africa, the so-called "limes." Thus, a *lingua franca* among the Berbers facilitated trade with the Romans. According to Bench, the adoption of a common trade language among the Berbers ultimately leveled linguistic diversity among this people. Blench further writes that the influence of the Berbers in North Africa later diminished after the spread of Islam throughout the region.

It should be noted that a discussion of the proto-Berber expansion across North Africa will continue in Paper 5.10, Section 3 and the discussion of haplogroup J-M304. A variant of

this main haplogroup, the J1-M267 mutation, which has origins in Southwest Asia, co-expanded with E-M81.

A final linguistically significant green cluster mutation is E-M293. Trombetta et al. (2015) estimate that this mutation evolved about 3.5 thousand years ago. In their 2008 study, Henn et al. suggest that the mutation evolved in Tanzania among the Datooga people. This population speaks a Nilo-Saharan language. However, as demonstrated by the data in Table 5.5.3, the E-M293 mutation attains a significant frequency among several different populations, not only the Datooga. Furthermore, the languages spoken by these populations not only belong to the Nilo-Saharan language family, but also to Afro-Asiatic, Niger-Congo, and Khoe-Kwadi. Additionally, E-M293 is found among the Sandawe and Hadza, two populations that speak a language classified as an isolate.

Henn et al. (2008) suggest that the E-M293 mutation represents the genetic signature of an expansion of East African pastoralism and the herding of cattle, goats and sheep. According to the study, the migration began about two thousand years ago and covered territory that the Bantus migrated through about fifteen hundred years later. The migration of East African pastoralists apparently followed a corridor to South Africa that was free of the tsetse fly, a blood sucking insect capable of transmitting diseases which devastate livestock. This facilitated a demic diffusion scenario and the frequency of E-M293 expanded rapidly. Many of those with the mutation then joined hunter-gatherer societies. For linguists, what is particularly interesting about the E-M293 mutation is that agricultural expansions have the potential of producing a series of language shifts among the populations that is recorded by a genetic mutation. In the case of the southern expansion of East African pastoralism and E-M293, the initial population may well have spoken a Nilo-Saharan language. As the migration expanded southwards, people shifted languages and adopted those that fall within the Afro-Asiatic, Niger-Congo, and Khoe-Kwadi language families, or the two isolates, Hadza and Sandawe.

Section 5. Red Cluster Mutations.

As noted previously in the discussion of green cluster mutations, haplogroup E-M35 evolved about 25 thousand years ago in East Africa. This date provides time depth for the co-migration of red and green cluster mutations out of East Africa into Egypt and later, the Middle East and Europe. Green cluster mutations are variants of the E-Z827 marker and red cluster mutation are variants of the E-M78 marker (see Figure 5.5.3). Within the red cluster, five mutations represent potentially significant markers for linguists: E-V12, E-V32, E-V65, E-V13, and E-V22. Based on phylogenetic relationships as shown in Figure 5.5.5, and frequency data shown in Tables 5.5.4 through 5.5.8, the E-V12, E-V32, and E-V65 mutations probably evolved in Northeastern Africa, whereas E-V22 and E-V13 probably evolved in Southwest Asia.

As the reader may recall from Paper 5.2, Section 3, and the discussion of haplogroup A-M13, about ten thousand years ago Holocene climate change transformed the Sahara desert into a savannah type ecosystem complete with rivers and lakes. Then about seven thousand years ago the rain stopped and the Sahara became once again a desert. As result of the so-called “desertification” of the Sahara, people either congregated along the Nile River in Egypt, or alternatively, moved with their herds of cattle, goats and sheep into the Sudan and East Africa (e.g. Kuper and Kröpelin 2006). Those that settled along the Nile eventually

adopted sedentary agriculture and cultivated crops that came from Southwest Asia. The pastoralists, on the other hand, herded sheep and goats that came from Southwest Asia, and cattle that probably have an African origin (see Bellwood 2005: 97-103).

Hassan et al. (2008) suggest that E-V12 and V22 represent the genetic relics of the desertification of the Sahara. According to Cruciani et al. (2007), the E-V22 mutation evolved about ten years ago. Additionally, as noted earlier, the E-V22 mutation appears to have evolved in Southwest Asia. Based on this data, it appears that E-V22 may well represent a Neolithic back-to-Africa migration of farmers or pastoralists that spoke a proto-Afro-Asiatic language. Moreover, E-V22 may well have co-migrated into North Africa with the "green cluster" E-M34 mutation that has been described previously in Section 4.

E-V12, on the other hand, appears to have expanded after the arrival of agriculture in North Eastern Africa within a population that may well have spoken a Nilo-Saharan language. This scenario is supported by the earlier discussion that places the origins of the E-V12 mutation in North Africa. Additionally, in-situ origins and expansion of E-V12 is supported by dating estimates from Cruciani et al. (2007) who suggest that E-V12 evolved about fourteen thousand years ago. Since the E-V22 and E-V12 mutations currently attain a significant frequency in Nilo-Saharan and Afro-Asiatic speaking populations, it appears that language shift has occurred quite frequently in North and East Africa. In other words, the data paint a scenario suggesting that since prehistoric times where Nilo-Saharan speaking populations have shifted to Afro-Asiatic, and Afro-Asiatic populations have shifted to Nilo-Saharan.

As shown by Figure 5.5.5, the E-V32 marker is a downstream variant of the E-V12 mutation that was discussed in the previous paragraph. As noted earlier, E-V32 appears to have evolved in Northeast Africa. Dating estimates from Cruciani et al. (2007) suggest that this occurred about eight thousand years ago. According to the same study, E-V32 currently represents eighty-two percent of E-M78 (or red cluster) variation in East Africa. Frequency data from Table 5.5.8, along with its estimated evolution date, suggest that the expansion of E-V32 in East Africa follows a demic diffusion model. In other words, the marker potentially represents an expansion of Nilo-Saharan cattle herders, or alternatively, an expansion of Afro-Asiatic speaking pastoralists, or alternatively both, from Egypt to East Africa.

Focusing now on the E-V65 mutation, very little information for this marker is currently available. As noted earlier, E-V65 probably evolved in North Africa. Dating estimates from Cruciani et al. (2007) suggest that this occurred about four thousand years ago. Data from Table 5.5.7 indicate that E-V65 attains a significant frequency among Arab populations in North Africa, whereas the frequency among Berber populations is low. This mysterious variation in frequency numbers seems to be a topic worthy of additional research.

The E-V13 mutation is the only haplogroup E-M96 variant that attains a significant frequency in Europe. As shown by Table 5.5.4, E-V13 attains a significant frequency among the populations of the Balkans and in Greece. More moderate frequencies are observed elsewhere in Europe, such as among the Italians and the Hungarians. Most studies suggest that E-V13 entered Europe during the Mesolithic (Battaglia et al. 2009; Regueiro et al. 2012; Karachanak et al. 2013). Regueiro et al. (2012) in their study of Serbs estimate the presence of E-V13 in the Balkans by 12 thousand years ago. Thus, the arrival of E-V13 may follow the disintegration of the Natufian culture during the Younger Dryas (see discussion in Section 2).

It should be noted that a recent study (Trombetta et al. 2015) suggests that E-V13 evolved around eight thousand years ago. Here, researchers favor a Neolithic or latter arrival of the mutation in Europe. Regardless, the presence of E-V13 in the Balkans, either during the Mesolithic or Neolithic, raises an interesting question: Were proto-Afro-Asiatic languages part of the linguistic inventory of prehistoric Europe?

From the frequency data tables provided for this present paper (5.5) the reader may notice that in addition to E-V13, other haplogroup E mutations also appear among the populations of Mediterranean Europe and Iberia. It should be emphasized, once again, that only E-V13 attains a significant frequency in Europe. Moreover, E-V13 has a clear prehistoric presence on this continent. Turning now to appearance of the green cluster E-M81 mutation in Iberia, the literature almost always treats this as a genetic relic of the Islamic (or Umayyad) conquest of the peninsula in the year 711. While this mutation may attain a significant frequency among a few isolated populations in Spain and Portugal, it should be noted that the overall frequency of E-M81 in Iberia is otherwise low (e.g. Regueiro et al. 2015). As such the Islamic conquest of Iberia added little to the gene pool in contemporary Spain and Portugal. For haplogroup E-M96 variants other than E-V13 and E-M81, a pattern surfaces whereby these mutations are generally found at a low frequency among populations residing on the Mediterranean islands of Europe or along the European Mediterranean coast. Here, historical trade and cultural exchange between North Africa and Europe may well provide an explanation (e.g. Cruciani et al. 2007).

Section 6. The Purple Cluster E-V6 Mutation.

The only mutation within the purple cluster is E-V6 (see Figure 5.5.3). Very little information is available about this marker and future research in this area might be fruitful. Cruciani et al. (2004) report that this mutation is found in nine percent of Ethiopians. A more recent study (Trombetta et al. 2015) suggests that the mutation attains a significant frequency among several East African populations speaking languages that belong to either the Afro-Asiatic or Nilo-Saharan language families (see Table 5.5.9). The same study estimates that the mutation evolved around twelve thousand years ago. However, their phylogenetic placement of the mutation within the haplogroup E-M96 hierarchy is substantially different than ISOGG (2017). According to ISOGG, the E-V6 mutation branches directly from E-M35, whereas Trombetta et al. place the mutation much further downstream within E-Z827. Thus based on the ISOGG phylogeny, E-V6 is potentially much older than the estimate provided by Trombetta et al. (2015). Taking this a step further, the E-V6 mutation potentially represents an E-M35 variant that remained in East Africa at a time when E-M78 and E-Z827 left the region. As such, E-V6 may stand as an ancient genetic relic of pre-agricultural East Africa.

Section 7. The Blue Cluster.

In Paper 5.2, Section 3, and the discussion of haplogroup A-M13, a connection was drawn between East African cattle pastoralism and the linguistic prehistory of Nilo-Saharan languages. Additionally, Sections 2 and 3 (above) present a discussion of the evolution of pastoralism in Southwest Asia and the linguistic prehistory of Afro-Asiatic languages. Moreover, Paper 5.3, Section 4 discussed the linguistic prehistory of Niger-Congo languages, the evolution of agriculture in West Central Africa, and the B-M150 mutation. Now the evolution of agriculture in West Central Africa continues with a discussion of blue cluster

mutations, which are variants of E-V38.

The reader is invited, once again, to review Figure 5.5.3. As shown by the figure, and as explained earlier, the E1b1-P2 mutation unites the blue, red, green and purple clusters. The blue cluster represents downstream variants of the E1b1a-V38 mutation that evolved in West Central Africa. The green, red and purple clusters, on the other hand, evolved in East Africa. Trombetta (2015) suggest that the diversification of E-P2 into these west and east variants occurred around 48 thousand years ago. Focusing now on Figure 5.5.6, the reader finds linguistically significant variants of the E-V38 "blue cluster" mutation.

The E-M2 mutation represents a downstream mutation from E-V38. Data from Poznik et al (2016) suggest that E-M2 diverged from E-V38 about 40 thousand years ago. Within the E-M2 phylogeny (see Figure 5.5.6), the E-U174 and E-U175 mutations have been identified as especially strong genetic signatures of the Bantu expansion from the Niger-Congo language family homeland in West Central Africa (e.g. Filippo et al. 2011; Montano et al. 2011; Barbieri et al. 2012; Rowold J. et al. 2016). This is based on frequency data for both mutations. For example, in their 2016 study of Bantu populations in Mozambique, Rowold et al. found that twenty-five percent of the samples belong to E-U174 and thirty-seven percent belong to E-U175. Besides the frequency data, dating estimates taken from West Central African populations also identify E-U174 and E-U175 as the genetic signature of the Bantu expansion. According to Filippo et al. (2011), the E-U175 mutation evolved in West Central Africa about five thousand year ago, and E-U174 evolved in the same region about four thousand years ago. These dating estimates agree with the timeframe for the Bantu expansion as taken from the archaeological record (see discussion in Paper 5.3, Hg. B, Section 4).

In their survey of West Central African populations, Filippo et al. (2011) and Barbieri et al. 2012 found that genetic diversity in Mande speakers and non-Bantoid Atlantic-Congo speakers to be older than in the Bantu populations. Mande and non-Bantoid Atlantic-Congo populations tend to have the orange cluster E-M33 mutation as well as undefined older mutations within E-M2 (see, also, Figure 5.5.3; Table 5.5.10; and Section 8). The Bantus, on the other hand, tend to have, almost exclusively, the E-U174 and E-U175 variants. This observation stands in general agreement with the linguistic evidence that places Mande and non-Bantoid Atlantic-Congo closer towards a theoretical Niger-Congo proto-language.

Filippo et al. (2011) also suggest that E-U174 and E-U175 variation found among Pygmy groups may well be undefined older variants of both mutations, and as such, this questions the extent of Bantu and Pygmy admixture. Perhaps the undefined older variants came from non-Bantoid populations before the Bantu expansion. Thus, additional resolution of E-M2 and its downstream variants is needed in order to further clarify the genetic history of the Bantus, other Niger-Congo speaking populations, and the Pygmies.

Section 8. The Orange Cluster E-M33 Mutation.

According to data from Poznik et al. (2016), E1-P147 diverged from the E-M96 main haplogroup about fifty thousand years ago. Shortly thereafter, about forty-eight thousand years ago, E-M33 diverged from E1-P147. These dating estimates, along with its position within the E-M96 main haplogroup phylogeny (see Figure 5.5.3), reflect that the E-M33 orange mutation evolved shortly after the initial back-to-Africa migration, which occurred around 60 thousand years ago. As such, E-M33 represents a comparatively ancient mutation

that traces its origins close to the initial diversification of E-M96 variation in Africa. As shown by Table 5.5.10, the geographic distribution of E-M33 populations is rather interesting as these populations are found in the Sahel region of Africa, an area that stands as a transition region between the southern border of the Sahara desert and Sub-Saharan Africa. Moreover, this region represents the putative homeland of Nilo-Saharan languages (see discussion in Paper 5.2, Section 3).

Within the Sahel region, Table 5.5.10 reflects that E-M33 attains a moderate frequency among Nilo-Saharan, Niger-Congo and Afro-Asiatic populations. As noted previously in Section 7, the presence of E-M33 among Mande speakers and non-Bantoid Atlantic-Congo speakers confirms what the linguistic evidence suggests, that they are the ancestral populations of the Bantu. Moreover, the E-M33 data supports the idea that Niger-Congo and Nilo-Saharan languages may well share a common linguistic ancestor (see discussion in Paper 5.2, Hg. A, Section 3). Turning now to Afro-Asiatic, the presence of the E-M33 mutation among Chadic speaking populations, such as the Kotoko and Masa of Cameroon (see Bučková et al 2013), and the Hausa of Sudan (see Hassan et al. 2008), might also be significant in that Chadic populations also have a significant frequency of the R1b-V88 mutation (see Paper 5.17, Section 7, and the discussion of haplogroup R1b-V88). This is significant because the discovery of a variant of the R1b-M343 mutation in the Sahel was unexpected and is still difficult to explain (e.g. Cruciani et al. 2010). This marker, otherwise, represents the genetic signature of West Eurasian populations.

Section 9. The Yellow Cluster E-M41 Mutation.

The reader is directed to Figure 5.5.1 and the E-M41 “yellow cluster” mutation. Very little is known about this mutation, including when it diverged from E2-M75. Most of the data comes from a 2010 study (Gomes et al.) that took samples one hundred and eighteen Ng'arkarimojong speakers from three different populations in Uganda, the Dodoth, Jie and Karimojong. Overall, the E-M41 mutation attains a modest frequency of eleven percent in these populations. In their 2005 study Wood et al. report that this mutation attains a frequency of sixty-seven percent among the Alur people of the Democratic Republic of the Congo. However, the sample size was very small (nine men) and ascertainment bias may well have skewed the actual frequency. The only other African population in which E-M41 attains a significant frequency is the Hema of the Democratic Republic of the Congo where the mutation is reported in thirty-nine percent of the men (Wood et al. 2005). However, this population speaks a Niger-Congo Bantoid language. Elsewhere in Africa (e.g. Luis et al. 2004), the M41 mutation attains a very small frequency.

Gomes et al. (2010) suggest that the E-M41 mutation represents a potential marker for understanding the genetic history of Nilo-Saharan speaking populations in East Africa. Indeed, the data suggest that the genetic relics of pre-agricultural Nilo-Saharan speaking populations in Africa are the E-M41 “yellow cluster” mutation, the E-M33 “orange cluster” mutation (see Section 8), and the E-V12 “red cluster” mutation (see Section 5).

Section 10. Conclusions.

Within the E-M96 main haplogroup, thirteen different variants stand as especially informative mutations for deciphering the correlation between linguistic and genetic diversity: E-M34, E-M81, E-M293, E-V12, E-V32, E-V65, E-V13, E-V22, E-V6, E-U174, E-U175, E-M33 and E-M41. The blue cluster E-U174 and E-U175 mutations carry the Bantu expansion southwards from West Central Africa. Proto-Berber and the green cluster E-M81 mutation co-expanded across North-Africa. The red cluster E-V22 and green cluster E-M34 variants represent Afro-Asiatic agriculturalist that entered North and East Africa during the Neolithic. Red cluster E-V12, orange cluster E-M33, and yellow cluster E-M41 are genetic relics of pre-agricultural Nilo-Saharan populations. The lone survivor of pre-agricultural East Africa appears to be the purple cluster E-V6 mutation. E-V32 from the red cluster represents the demic expansion of Nilo-Saharan and/or Afro-Asiatic pastoralists into East Africa. Pastoralism later expanded from this region with the green cluster E-M293 mutation. This expansion triggered a series of language shifts among the herders who joined hunter-gatherer groups. The red cluster E-V13 raises the possibility that some prehistoric Europeans may have spoken a proto-Afro-Asiatic language. Finally, the mysterious presence of E-V65 in North African Arabs requires additional research.

Bibliography for Paper and Tables

Abu-Amero, Khaled K. et al. 2009. "Saudi Arabian Y-chromosome diversity and its relationship with nearby regions." *BioMed Central Genetics* 10:59.

Arredi, Barbara et al. 2004**.** "A predominately Neolithic origin for Y-chromosomal DNA variation in North Africa." *American Journal of Human Genetic*s 75: 338-345.

Bar-Yosef, Ofer 1998. "The Natufian culture in the Levant, threshold to the origins of agriculture." *Evolutionary Anthropology* 6(5): 159-177.

Barbieri**,** Chiara et al. 2012. "Contrasting maternal and paternal histories in the linguistic context of Burkina Faso." *Molecular Biology and Evolution* 29(4): 1213-1223.

Battaglia, Vincenza et al. 2009. "Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe." *European Journal of Human Genetics* 17: 820-830.

Bellwood, Peter 2005. *First farmers: the origins of agricultural societies.* Malden, MA; Oxford, UK; Victoria, Australia: Blackwell Publishing.

Blockley, S.P.E. and R. Pinhasi 2011. "A revised chronology for the adoption of agriculture in the Southern Levant and the role of Late Glacial climatic change." *Quaternary Science Reviews* 30: 98-108.

Bosch, Elena et al. 2001. "High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula." *American Journal of Human Genetics* 68: 1019-1029.

Bučková, Jana et al. 2013. "Multiple and differentiated contributions to the male gene pool of pastoral and farmer populations of the African Sahel." *American Journal of Physical Anthropology* 151: 10-21.

Cadenas, Alicia M. et al. 2008. "Y-chromosome diversity characterizes the Gulf of Oman." *European Journal of Human Genetics* 16: 374-386.

Cruciani, Fulvio et al. 2004. "Phylogenetic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa." *American Journal of Human Genetics* 74: 1014-1022.

Cruciani, Fulvio et al. 2007. "Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosome haplogroups E-M78 and J-M12." *Molecular Biology and Evolution* 24(6): 1300-1311.

Cruciani, Fulvio et al. 2010. "Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages." *European Journal of Human Genetics* 18: 800-807.

Ehret, Christopher et al. 2004. "The Origins of Afroasiatic." *Science* 306: 1680-1681.

Ennafaa, Hajer et al. 2011. "Mitochondrial and Y-chromosome microstructure in Tunisia." *Journal of Human Genetics* 56: 734-741.

Fadhlaoui-Zid, Karima et al. 2011. "Genetic structure of Tunisian ethnic groups revealed by parental lineages." *American Journal of Physical Anthropology* 148: 271-280.

Filippo, Cesare de et al. 2011. "Y-chromosomal variation in Sub-Saharan Africa: Insights into the history of Niger-Congo groups." *Molecular Biology and Evolution* 28(3): 1255-1269.

Fregel, Rosa et al. 2017. "Neolithization of North Africa involved the migration of people from both the Levant and Europe." bioRxiv preprint first posted online Sep. 21, 2017; doi: http://dx.doi.org/10.1101/191569

Gomes, Verónica et al. 2010. "Digging deeper into East African human Y-chromosome lineages." *Human Genetics* 127: 603-613.

Hassan, Hisham Y. et al. 2008. "Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography and history." *American Journal of Physical Anthropology* 137: 316-323.

Henn, Brenna M. et al. 2008. "Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa." *Proceedings of the National Academy of Sciences of the United States of America* 105(31): 10693-10698.

Hetzron, Robert 2009. Afro-Asiatic languages. In: The *World's Major Languages*. Second Edition. Bernard Comrie, Ed. Oxon, UK; New York: Routledge. 545-550.

ISOGG (International Society of Genetic Genealogy). https://isogg.org

Karachanak, Sena et al. 2013. "Y-chromosome diversity in modern Bulgarians: new clues about their ancestry." *Public Library of Science One* 8(3): e56779.

Kuper, Rudolph and Stefan Kröpelin 2006. "Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution." *Science* 313: 803-807

Lazaridis, Iosif et al. 2016. "Genomic insights into the origin of farming in the ancient Near East." *Nature* 25: 536 (7617): 419-424.

Luis, J.R. et al. 2004. "The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations." *American Journal of Human Genetics* 74: 532-544.

Militarev, Alexander 2002. "The prehistory of a dispersal: the Proto-Afrasian (Afroasiatic) farming lexicon." In: *Examining the Farming/Language Dispersal Hypothesis*. Peter Bellwood and Colin Renfrew, Eds. Cambridge: McDonald Institute for Archaeological Research, pp. 135-150.

Montano, Valeria et al. 2011. "The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa." *Molecular Ecology* 20: 2693-2708.

Ottoni, Claudio et al. 2011. "Deep into the roots of the Libyan Tuareg: a genetic survey of their paternal heritage." *American Journal of Physical Anthropology* 145: 118-124.

Pereira, Luísa et al. 2010. "Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel." *European Journal of Human Genetics* 18: 915-923.

Poznik, G. David et al. 2016. "Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences." *Nature Genetics* 48(6): 593-600.

Regueiro, Maria et al. 2012. "High levels of Paleolithic Y-chromosome lineages characterize Serbia." *Gene* 498: 59-67.

Regueiro, Maria et al 2015. "From Arabia to Iberia: a Y chromosome perspective." *Gene* 564: 141-152.

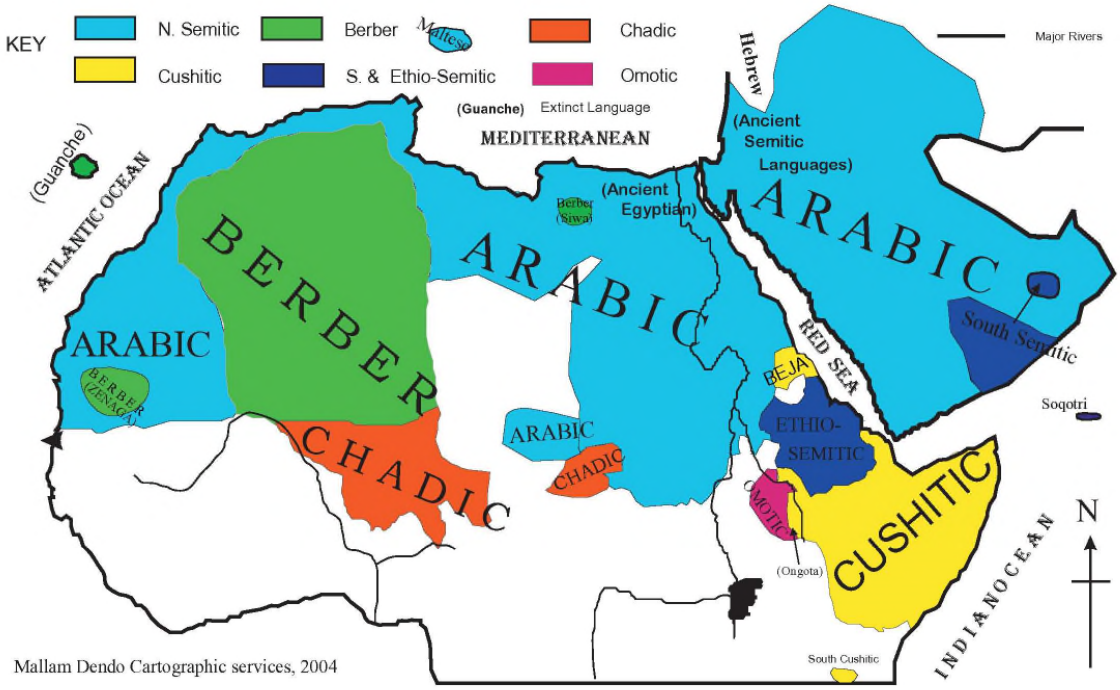
Rowold, Daine J. et al. 2016. "On the Bantu expansion." *Gene* 593: 48-57.

Trombetta, Beniamino et al. 2011. "A new topology of the human Y-Chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms." *Public Library of Science One* 6(1): e16073.

Trombetta, Beniamino et al. 2015. "Phylogeographic refinement and large scale genotyping of human Y chromosome haplogroup E provide new insights into the dispersal of early pastoralists in the African continent." *Genome Biology and Evolution* 7(7): 1940–1950.

Wood, Elizabeth T. 2005. "Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biases demographic processes." *European Journal of Human Genetics* 13: 867-876.

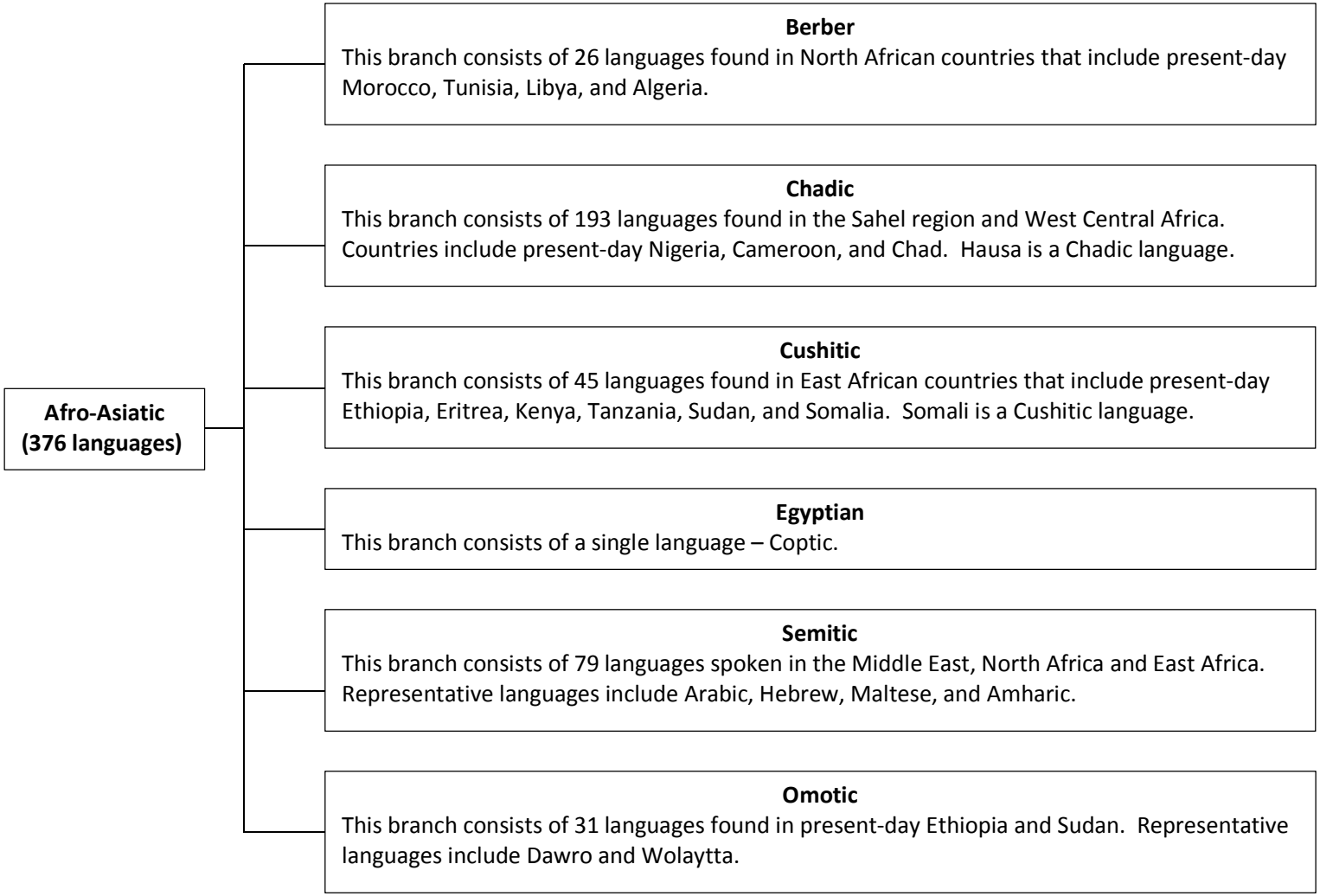
Figure 5.5.1. Geographic Distribution of Afro-Asiatic Languages.



Source: Dr. Roger Blench
McDonald Institute for Archaeological Research, University of Cambridge
Dept. of History, University of Jos
Kay Williamson Educational Foundation
8, Guest Road
Cambridge, CB1 2AL
United Kingdom
http://www.rogerblench.info/Language/Afroasiatic/AASOP.htm

Used with permission.

Figure 5.5.2. Overview of Afro-Asiatic and it Main Branches.

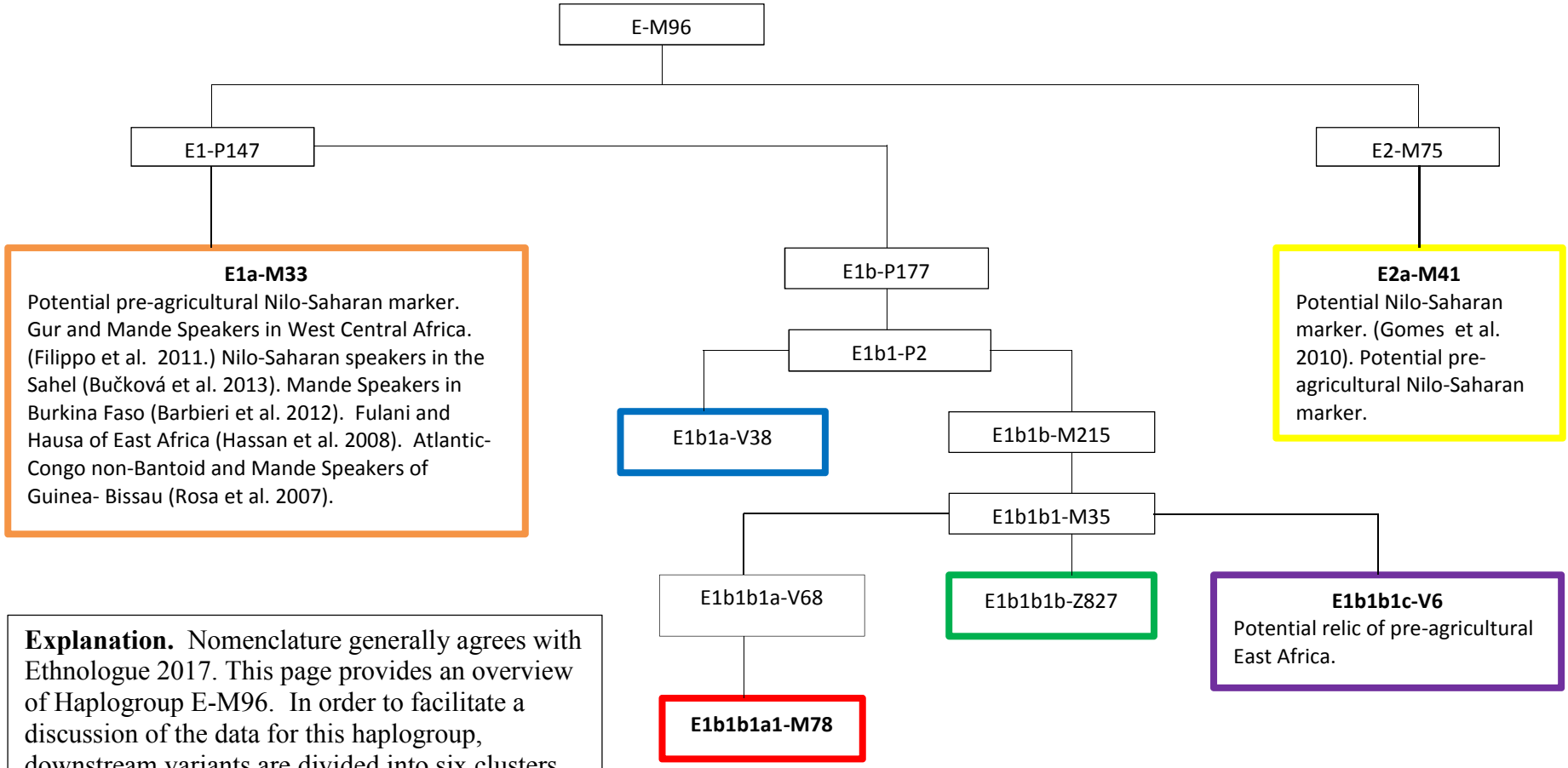


Source: Ethnologue 2017

Paper 5.5. Haplogroup E.

The Genetic-Linguistic Interface.

Figure 5.5.3. Overview of Linguistically Significant Haplogroup E Variants.



Explanation. Nomenclature generally agrees with Ethnologue 2017. This page provides an overview of Haplogroup E-M96. In order to facilitate a discussion of the data for this haplogroup, downstream variants are divided into six clusters that have a color designation: orange, yellow, purple, blue, red and green. Linguistically significant variants within the green, red and blue clusters are presented in Figures 5.5.4, 5.5.5, and 5.5.6 respectively.

Figure 5.5.4. E-Z827 and the Green Cluster Mutations.

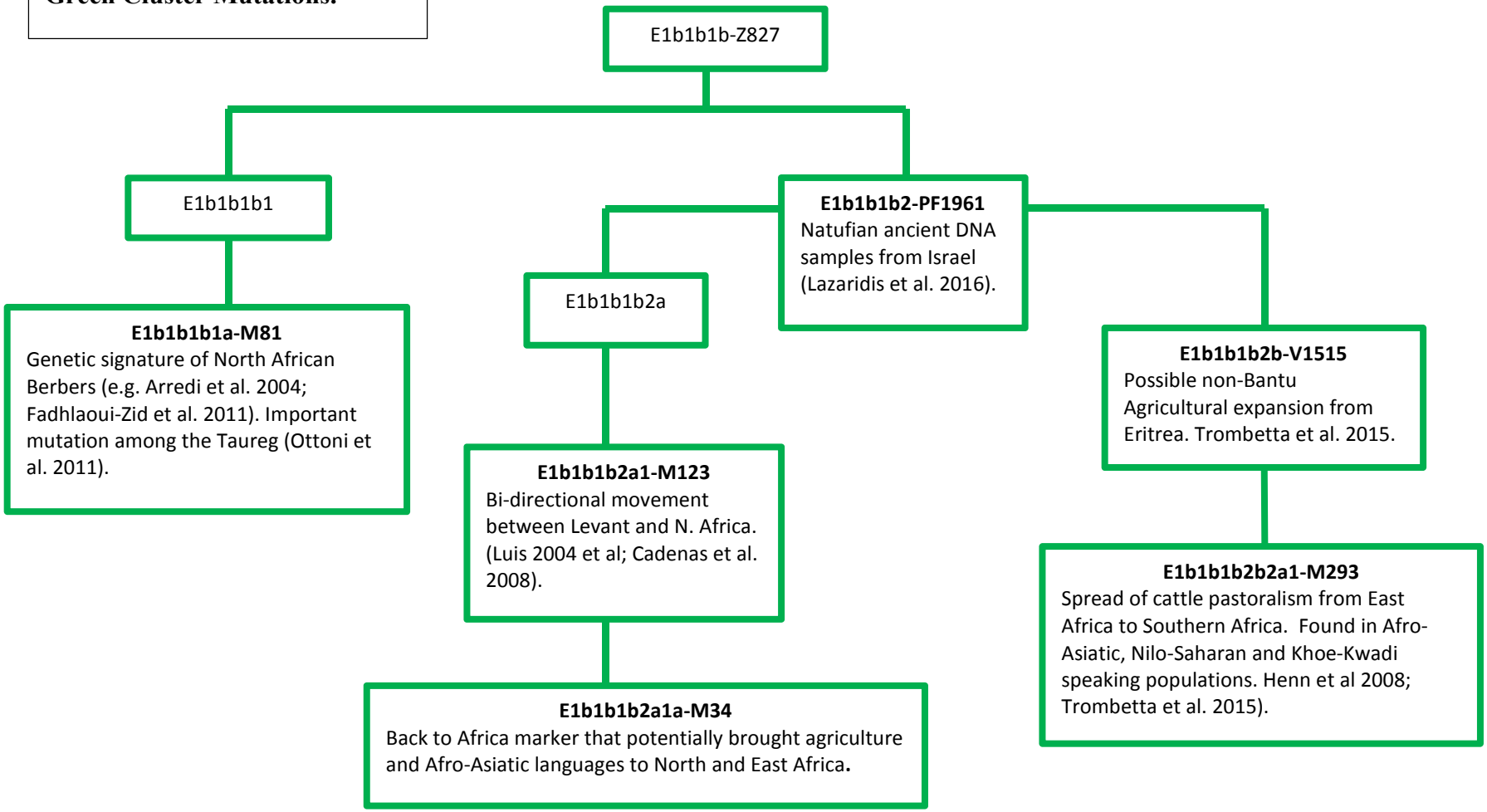


Figure 5.5.5. E-M78 and the Red Cluster Mutations.

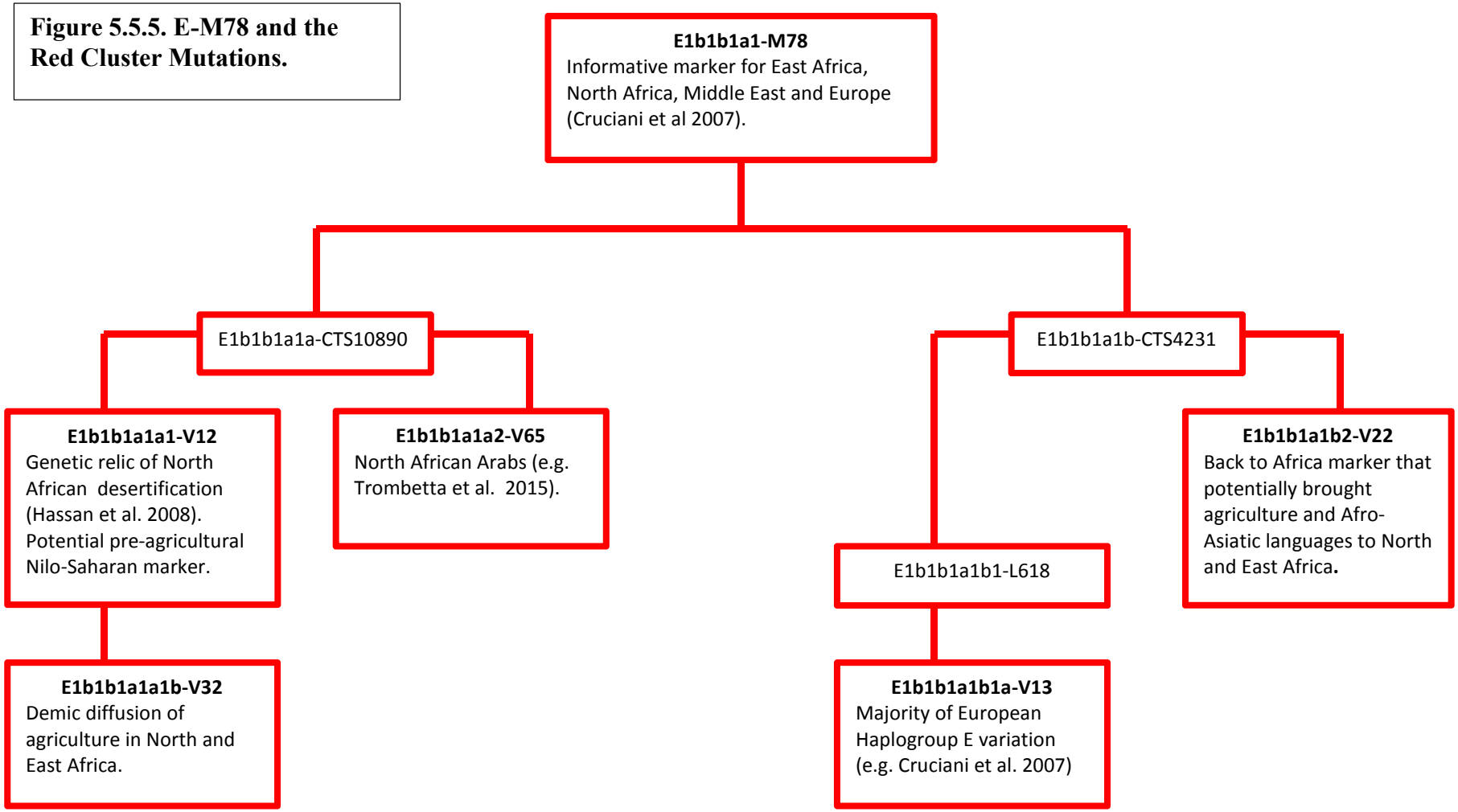


Figure 5.5.6. E-V38 and the Blue Cluster Mutations.

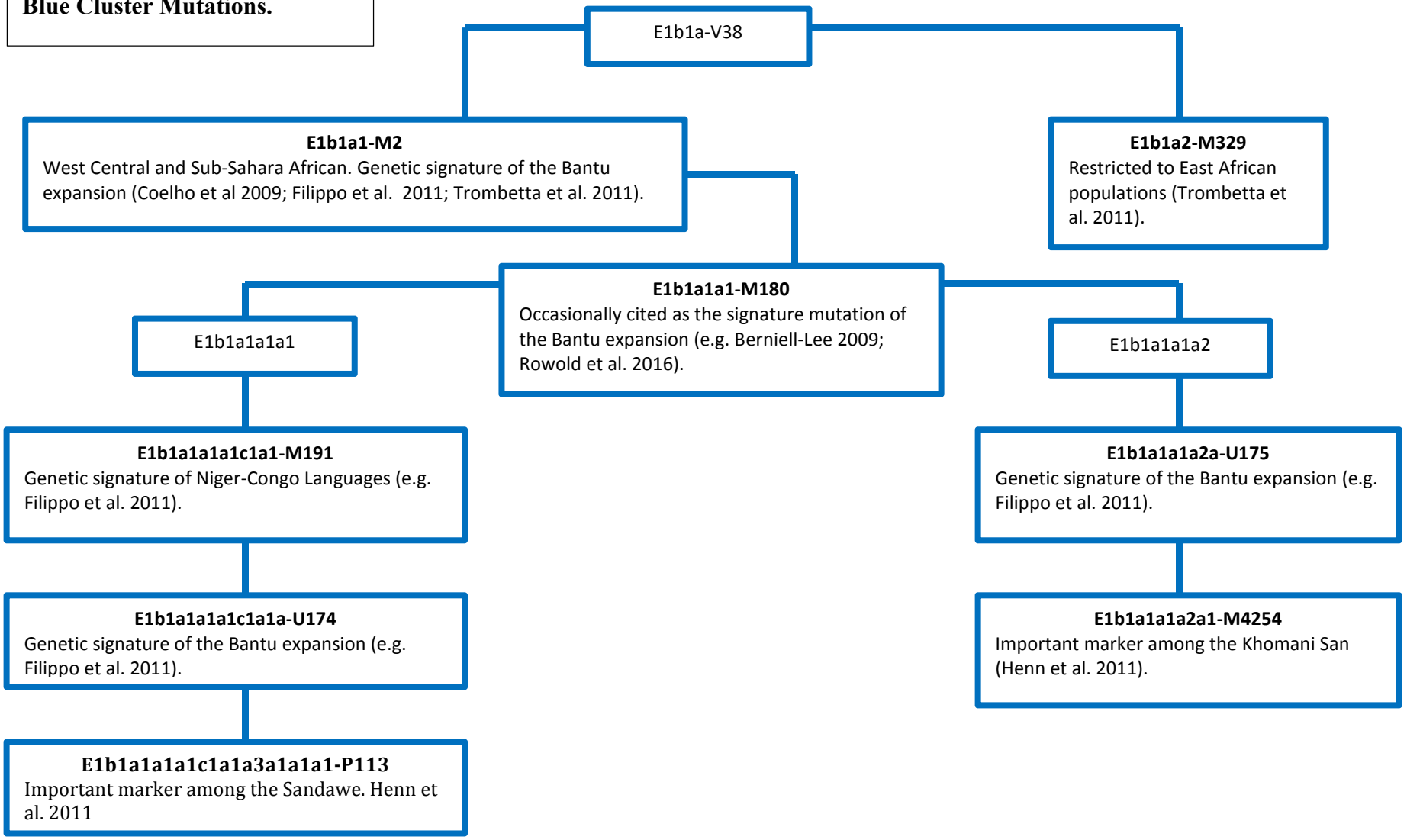


Table 5.5.1. Survey of E-M81 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-M81 (%)
Trombetta et al. 2015	Mozabite Berbers	67	Algeria	Afro-Asiatic	86.6
Trombetta et al. 2015	Asni Berbers	54	Morocco	Afro-Asiatic	79.6
Trombetta et al. 2015	Marrakech Berbers	27	Morocco	Afro-Asiatic	77.8
Trombetta et al. 2015	Bouhria Berbers	67	Morocco	Afro-Asiatic	77.6
Trombetta et al. 2015	Souss Berbers	34	Morocco	Afro-Asiatic	76.5
Trombetta et al. 2015	Moyen Atlas Berbers	69	Morocco	Afro-Asiatic	71.0
Trombetta et al. 2015	Ouarzazate Berbers	31	Morocco	Afro-Asiatic	54.8
Trombetta et al. 2015	Pasangos from Cantabria	56	Spain	Indo-European	41.1
Trombetta et al. 2015	Moroccan Arabs	55	Morocco	Afro-Asiatic	30.9
Trombetta et al. 2015	Libian Arabs	10	Libya	Afro-Asiatic	30.0
Trombetta et al. 2015	Southern Portuguese	49	Portugal	Indo-European	12.2
Trombetta et al. 2015	Tunisian Jews	10	Tunisia	Afro-Asiatic	10.0
Trombetta et al. 2015	Tuareg	22	Niger	Afro-Asiatic	9.1
Trombetta et al. 2015	Turkish Cypriots	46	Turkey	Turkic	8.7
Trombetta et al. 2015	Istanbul Turkish	35	Turkey	Turkic	5.7
Trombetta et al. 2015	Sephardic Turkish	19	Turkey	Turkic	5.3
Trombetta et al. 2015	Egyptians from Baharia	41	Egypt	Afro-Asiatic	4.9
Trombetta et al. 2015	Northern Egyptians	49	Egypt	Afro-Asiatic	4.1
Trombetta et al. 2015	Northern Portuguese	50	Portugal	Indo-European	4.0
Trombetta et al. 2015	Spanish Basques	55	Spain	Basque	3.6
Trombetta et al. 2015	Bedouins	28	Israel	Afro-Asiatic	3.6
Trombetta et al. 2015	French	85	France	Indo-European	3.5

Study	Population	Size	Country	Lang Fam	Frequency of E-M81 (%)
Trombetta et al. 2015	Southwestern Turkish	40	Turkey	Turkic	2.5
Trombetta et al. 2015	Northeastern Turkish	41	Turkey	Turkic	2.4
Trombetta et al. 2015	Asturians	90	Spain	Indo-European	2.2
Trombetta et al. 2015	Southern Spaniards	62	Spain	Indo-European	1.6
Trombetta et al. 2015	Southern Italians	141	Italy	Indo-European	1.4
Trombetta et al. 2015	Northern Italians	80	Italy	Indo-European	1.3
Trombetta et al. 2015	Egyptian Berbers from Siwa	93	Egypt	Afro-Asiatic	1.1
Trombetta et al. 2015	Yemenites	94	Yemen	Afro-Asiatic	1.1
Trombetta et al. 2015	Hungarians	106	Hungary	Uralic	0.9
Trombetta et al. 2015	Central Italians	356	Italy	Indo-European	0.8
Trombetta et al. 2015	Sicilians	153	Italy	Indo-European	0.7
Trombetta et al. 2015	Sardinians	374	Italy	Indo-European	0.3

Table 5.5.2. Survey of E-M34 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-M34 (%)
Flores et al. 2005	Dead Sea Jordanians	45	Jordan	Afro-Asiatic	31.1
Trombetta et al. 2015	Libian Jews	23	Libya	Afro-Asiatic	17.4
Trombetta et al. 2015	Ethiopian Jews	22	Ethiopia	Afro-Asiatic	13.6
Trombetta et al. 2015	Amhara	82	Ethiopia	Afro-Asiatic	13.4
Trombetta et al. 2015	Northern Egyptians	49	Egypt	Afro-Asiatic	10.2
Trombetta et al. 2015	Wolayta	12	Ethiopia	Afro-Asiatic	8.3
Trombetta et al. 2015	Erzurum Turkish	25	Turkey	Turkic	8.0
Trombetta et al. 2015	Omanite	13	Oman	Afro-Asiatic	7.7
Trombetta et al. 2015	Yemenites	94	Yemen	Afro-Asiatic	7.4
Trombetta et al. 2015	Bedouins	28	Israel	Afro-Asiatic	7.1
Trombetta et al. 2015	Sicilians	153	Italy	Indo-European	6.5
Trombetta et al. 2015	Sephardic Turkish	19	Turkey	Turkic	5.3
Trombetta et al. 2015	Sephardic Bulgarians	20	Bulgaria	Indo-European	5.0
Flores et al. 2005	Jordanians	101	Jordan	Afro-Asiatic	5.0
Trombetta et al. 2015	Cunama	20	Eritrea	Nilo-Saharan	5.0
Trombetta et al. 2015	Arabs	41	United Arab Emirate	Afro-Asiatic	4.9
Trombetta et al. 2015	Oromo	62	Ethiopia	Afro-Asiatic	4.8
Trombetta et al. 2015	Southeastern Turkish	24	Turkey	Turkic	4.2
Trombetta et al. 2015	Afar	25	Djibouti	Afro-Asiatic	4.0
Trombetta et al. 2015	Central Italians	356	Italy	Indo-European	3.9
Trombetta et al. 2015	Marrakech Berbers	27	Morocco	Afro-Asiatic	3.7
Trombetta et al. 2015	Druze Arabs	28	Israel	Afro-Asiatic	3.6
Grugni et al. 2012	Iranians	938	Indo-European	Indo-Iranian	3.5

Study	Population	Size	Country	Lang Fam	Frequency of E-M34 (%)
Trombetta et al. 2015	Sardinians	374	Italy	Indo-European	3.5
Trombetta et al. 2015	Palestinians	29	Istrael	Afro-Asiatic	3.4
Trombetta et al. 2015	Central Anatolian	61	Turkey	Turkic	3.3
Trombetta et al. 2015	Continental Greeks	32	Greece	Indo-European	3.1
Trombetta et al. 2015	Istanbul Turkish	35	Turkey	Turkic	2.9
Trombetta et al. 2015	Southern Italians	141	Italy	Indo-European	2.8
Trombetta et al. 2015	Southwestern Turkish	40	Turkey	Turkic	2.5
Trombetta et al. 2015	Turkish Cypriots	46	Turkey	Turkic	2.2
Trombetta et al. 2015	Egyptian Berbers from Siwa	93	Egypt	Afro-Asiatic	2.2
Trombetta et al. 2015	Southern Egyptians	47	Egypt	Afro-Asiatic	2.1
Karachanak et al. 2013	Bulgarians	808	Indo-European	Slavic	1.9
Trombetta et al. 2015	Mozabite Berbers	67	Algeria	Afro-Asiatic	1.5
Trombetta et al. 2015	Corsicans	140	France	Indo-European	1.4
Trombetta et al. 2015	Northern Italians	80	Italy	Indo-European	1.3
Trombetta et al. 2015	Saho	94	Eritrea	Afro-Asiatic	1.1

Table 5.5.3. Survey of E-M293 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-M293* (%)
Henn et al. 2008	Datooga	40	Tanzania	Nilo-Saharan	43.0
Henn et al. 2008	Khwe	26	South Africa	Khoe-Kwadi	31.0
Trombetta et al. 2015	Khwe	26	Namibia	Khoe-Kwadi	30.8
Henn et al. 2008	Burunge	25	Tanzania	Afro-Asiatic	28.0
Trombetta et al. 2015	Maasai	45	Kenya	Nilo-Saharan	24.4
Henn et al. 2008	Sandawe	70	Tanzania	Isolate	24.0
Trombetta et al. 2015	Tutsi	9	Burundi	Niger-Congo	22.2
Henn et al. 2008	Mbugwe	14	Tanzania	Niger-Congo	21.0
Trombetta et al. 2015	Nilotic Western Kenya	11	Kenya	Nilo-Saharan	18.2
Henn et al. 2008	Kenyan Bantu	11	Kenya	Niger-Congo	18.0
Trombetta et al. 2015	Turkana	6	Kenya	Nilo-Saharan	16.7
Henn et al. 2008	Turu	22	Tanzania	Niger-Congo	14.0
Henn et al. 2008	South African Bantu	8	South Africa	Niger-Congo	13.0
Trombetta et al. 2015	Bantu	8	South Africa	Niger-Congo	12.5
Trombetta et al. 2015	Other Bantu	17	Kenya	Niger-Congo	11.8
Trombetta et al. 2015	Kikuyu	9	Kenya	Niger-Congo	11.1
Henn et al. 2008	!Kung	64	South Africa	Khoe-Kwadi	11.0
Henn et al. 2008	Hazda	54	Tanzania	Isolate	11.0
Trombetta et al. 2015	!Kung	64	Angola	Khoe-Kwadi	9.4
Trombetta et al. 2015	Wolayta	12	Ethiopia	Afro-Asiatic	8.3
Trombetta et al. 2015	Luhya	51	Kenya	Niger-Congo	3.9
Henn et al. 2008	Sukuma	30	Tanzania	Niger-Congo	3.0
Trombetta et al. 2015	Oromo	62	Ethiopia	Afro-Asiatic	1.6

Table 5.5.4. Survey of E-V13 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-V13 (%)
Trombetta et al. 2015	Albanians	21	Albania	Indo-European	33.3
Trombetta et al. 2015	Rumanians	30	Romania	Indo-European	26.7
Trombetta et al. 2015	Continental Greeks	32	Greece	Indo-European	25.0
Trombetta et al. 2015	Bulgarians	112	Bulgaria	Indo-European	21.4
Trombetta et al. 2015	Macedonians	99	Macedonia	Indo-European	18.2
Trombetta et al. 2015	Turkish Cypriots	46	Turkey	Turkic	10.9
Trombetta et al. 2015	Druze Arabs	28	Israel	Afro-Asiatic	10.7
Trombetta et al. 2015	Hungarians	106	Hungary	Uralic	9.4
Trombetta et al. 2015	Southern Italians	141	Italy	Indo-European	8.5
Trombetta et al. 2015	Sicilians	153	Italy	Indo-European	7.2
Trombetta et al. 2015	Northern Italians	80	Italy	Indo-European	6.3
Trombetta et al. 2015	Asturians	90	Spain	Indo-European	5.6
Trombetta et al. 2015	Central Italians	356	Italy	Indo-European	5.3
Trombetta et al. 2015	Central Anatolian	61	Turkey	Turkic	4.9
Trombetta et al. 2015	Libian Jews	23	Libya	Afro-Asiatic	4.3
Trombetta et al. 2015	Corsicans	140	France	Indo-European	4.3
Trombetta et al. 2015	Southeastern Turkish	24	Turkey	Turkic	4.2
Trombetta et al. 2015	Southern Portuguese	49	Portugal	Indo-European	4.1
Trombetta et al. 2015	Estonians	74	Estonia	Uralic	4.1
Trombetta et al. 2015	Northern Portuguese	50	Portugal	Indo-European	4.0
Trombetta et al. 2015	Germans	77	Germany	Indo-European	3.9
Trombetta et al. 2015	French	85	France	Indo-European	3.5

Study	Population	Size	Country	Lang Fam	Frequency of E-V13 (%)
Trombetta et al. 2015	Palestinians	29	Istrael	Afro-Asiatic	3.4
Trombetta et al. 2015	Slovenians	104	Slovenia	Indo-European	2.9
Trombetta et al. 2015	Istanbul Turkish	35	Turkey	Turkic	2.9
Trombetta et al. 2015	Danish	35	Denmark	Indo-European	2.9
Trombetta et al. 2015	Polish	40	Poland	Indo-European	2.5
Trombetta et al. 2015	Southwestern Turkish	40	Turkey	Turkic	2.5
Trombetta et al. 2015	Egyptians from Baharia	41	Egypt	Afro-Asiatic	2.4
Trombetta et al. 2015	Northern Egyptians	49	Egypt	Afro-Asiatic	2.0
Trombetta et al. 2015	Bouhria Berbers	67	Morocco	Afro-Asiatic	1.5
Trombetta et al. 2015	Sardinians	374	Italy	Indo-European	1.1

Table 5.5.5. Survey of E-V22 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-V22 (%)
Trombetta et al. 2015	Saho	94	Eritrea	Afro-Asiatic	88.3
Trombetta et al. 2015	Turkana	6	Kenya	Nilo-Saharan	33.3
Hassan et al. 2008	Fula (Fulbe, Fulani)	26	Sudan	Niger-Congo	30.8
Trombetta et al. 2015	Gurage	7	Ethiopia	Afro-Asiatic	28.6
Trombetta et al. 2015	Egyptians from Baharia	41	Egypt	Afro-Asiatic	22.0
Hassan et al. 2008	Fur	32	Sudan	Nilo-Saharan	18.8
Trombetta et al. 2015	Northern Egyptians	49	Egypt	Afro-Asiatic	16.3
Hassan et al. 2008	Masalit	32	Sudan	Nilo-Saharan	15.6
Hassan et al. 2008	Shilluk	15	Sudan	Nilo-Saharan	13.3
Hassan et al. 2008	Dinka	26	Sudan	Nilo-Saharan	11.5
Trombetta et al. 2015	Nilotic Western Kenya	11	Kenya	Nilo-Saharan	9.1
Trombetta et al. 2015	Somali	12	Ethiopia	Afro-Asiatic	8.3
Trombetta et al. 2015	Wolayta	12	Ethiopia	Afro-Asiatic	8.3
Trombetta et al. 2015	Omanite	13	Oman	Afro-Asiatic	7.7
Trombetta et al. 2015	Moroccan Arabs	55	Morocco	Afro-Asiatic	7.3
Trombetta et al. 2015	Palestinians	29	Istrael	Afro-Asiatic	6.9
Trombetta et al. 2015	Maasai	45	Kenya	Nilo-Saharan	6.7
Trombetta et al. 2015	Nara	15	Eritrea	Nilo-Saharan	6.7
Trombetta et al. 2015	Istanbul Turkish	35	Turkey	Turkic	5.7
Trombetta et al. 2015	Cunama	20	Eritrea	Nilo-Saharan	5.0
Trombetta et al. 2015	Sicilians	153	Italy	Indo-European	4.6
Trombetta et al. 2015	Asturians	90	Spain	Indo-European	4.4

Study	Population	Size	Country	Lang Fam	Frequency of E-V22 (%)
Trombetta et al. 2015	Marrakech Berbers	27	Morocco	Afro-Asiatic	3.7
Trombetta et al. 2015	Asni Berbers	54	Morocco	Afro-Asiatic	3.7
Trombetta et al. 2015	Bedouins	28	Israel	Afro-Asiatic	3.6
Trombetta et al. 2015	Southern Spaniards	62	Spain	Indo-European	3.2
Trombetta et al. 2015	Tigrai	32	Eritrea/Ethiopia	Afro-Asiatic	3.1
Trombetta et al. 2015	Northern Italians	80	Italy	Indo-European	2.5
Trombetta et al. 2015	Amhara	82	Ethiopia	Afro-Asiatic	2.4
Trombetta et al. 2015	Arabs	41	United Arab Emirate	Afro-Asiatic	2.4
Trombetta et al. 2015	Turkish Cypriots	46	Turkey	Turkic	2.2
Trombetta et al. 2015	Central Italians	356	Italy	Indo-European	2.0
Trombetta et al. 2015	Southern Italians	141	Italy	Indo-European	1.4
Trombetta et al. 2015	Sardinians	374	Italy	Indo-European	0.8

Table 5.5.6. Survey of E-V12 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-V12* (%)
Trombetta et al. 2015	Southern Egyptians	47	Egypt	Afro-Asiatic	74.5
Hassan et al. 2008	Nuer	12	Sudan	Nilo-Saharan	16.7
Hassan et al. 2008	Copts	33	Sudan	Afro-Asiatic	15.2
Trombetta et al. 2015	Egyptians from Baharia	41	Egypt	Afro-Asiatic	14.6
Hassan et al. 2008	Nubians	39	Sudan	Nilo-Saharan	12.8
Trombetta et al. 2015	Nilotic Western Kenya	11	Kenya	Nilo-Saharan	9.1
Trombetta et al. 2015	Egyptians from Gurna Oasis	34	Egypt	Afro-Asiatic	8.8
Trombetta et al. 2015	Mandenka	16	Senegal	Niger-Congo	6.3
Trombetta et al. 2015	French Basques	16	France	Basque	6.3
Hassan et al. 2008	Beja	42	Sudan	Afro-Asiatic	4.8
Trombetta et al. 2015	Tuareg	22	Niger	Afro-Asiatic	4.5
Trombetta et al. 2015	Northern Egyptians	49	Egypt	Afro-Asiatic	4.1
Trombetta et al. 2015	Erzurum Turkish	25	Turkey	Turkic	4.0
Hassan et al. 2008	Dinka	26	Sudan	Nilo-Saharan	3.8
Trombetta et al. 2015	Marrakech Berbers	27	Morocco	Afro-Asiatic	3.7
Trombetta et al. 2015	Tigrai	32	Eritrea/Ethiopia	Afro-Asiatic	3.1
Trombetta et al. 2015	Egyptian Berbers from Siwa	93	Egypt	Afro-Asiatic	2.2
Trombetta et al. 2015	Yemenites	94	Yemen	Afro-Asiatic	2.1
Trombetta et al. 2015	Central Anatolian	61	Turkey	Turkic	1.6
Trombetta et al. 2015	Mandara	82	Cameroon (North)	Afro-Asiatic	1.2
Trombetta et al. 2015	French	85	France	Indo-European	1.2
Trombetta et al. 2015	Southern Italians	141	Italy	Indo-European	0.7

Study	Population	Size	Country	Lang Fam	Frequency of E-V12* (%)
Trombetta et al. 2015	Sicilians	153	Italy	Indo-European	0.7
Trombetta et al. 2015	Central Italians	356	Italy	Indo-European	0.3
Trombetta et al. 2015	Sardinians	374	Italy	Indo-European	0.3

Table 5.5.7. Survey of V65 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-V65 (%)
Trombetta et al. 2015	Moroccan Arabs	55	Morocco	Afro-Asiatic	32.7
Trombetta et al. 2015	Libian Arabs	10	Libya	Afro-Asiatic	20.0
Trombetta et al. 2015	Moyen Atlas Berbers	69	Morocco	Afro-Asiatic	10.1
Trombetta et al. 2015	Libian Jews	23	Libya	Afro-Asiatic	4.3
Trombetta et al. 2015	Egyptian Berbers from Siwa	93	Egypt	Afro-Asiatic	4.3
Trombetta et al. 2015	Egyptians from Baharia	41	Egypt	Afro-Asiatic	2.4
Trombetta et al. 2015	Mozabite Berbers	67	Algeria	Afro-Asiatic	1.5
Trombetta et al. 2015	Sardinians	374	Italy	Indo-European	1.1
Trombetta et al. 2015	Sicilians	153	Italy	Indo-European	0.7
Trombetta et al. 2015	Central Italians	356	Italy	Indo-European	0.3

Table 5.5.8. Survey of E-V32 Populations.

Study	Population	Size	Country	Lang Fam	Frequency E-V32 (%)
Trombetta et al. 2015	Somali	5	Somalia	Afro-Asiatic	80.0
Trombetta et al. 2015	Borana	9	Kenya	Afro-Asiatic	66.7
Trombetta et al. 2015	Somali	6	Kenya	Afro-Asiatic	66.7
Trombetta et al. 2015	Tigre	5	Eritrea	Afro-Asiatic	60.0
Hassan et al. 2008	Masalit	32	Sudan	Nilo-Saharan	53.1
Hassan et al. 2008	Fur	32	Sudan	Nilo-Saharan	40.6
Hassan et al. 2008	Beja	42	Sudan	Afro-Asiatic	31.0
Trombetta et al. 2015	Somali	12	Ethiopia	Afro-Asiatic	25.0
Trombetta et al. 2015	Somali	40	Djibouti	Afro-Asiatic	25.0
Trombetta et al. 2015	Oromo	62	Ethiopia	Afro-Asiatic	22.6
Trombetta et al. 2015	Tigrai	32	Eritrea/Ethiopia	Afro-Asiatic	21.9
Trombetta et al. 2015	Cunama	20	Eritrea	Nilo-Saharan	20.0
Trombetta et al. 2015	Nara	15	Eritrea	Nilo-Saharan	13.3
Trombetta et al. 2015	Amhara	82	Ethiopia	Afro-Asiatic	11.0
Trombetta et al. 2015	Ethiopian Jews	22	Ethiopia	Afro-Asiatic	9.1
Trombetta et al. 2015	Wolayta	12	Ethiopia	Afro-Asiatic	8.3
Trombetta et al. 2015	Maasai	45	Kenya	Nilo-Saharan	6.7
Trombetta et al. 2015	Luhya	51	Kenya	Niger-Congo	5.9
Trombetta et al. 2015	Moundang	21	Cameroon (North)	Niger-Congo	4.8
Hassan et al. 2008	Fula (Fulbe, Fulani)	26	Sudan	Niger-Congo	3.8
Trombetta et al. 2015	Yemenites	94	Yemen	Afro-Asiatic	3.2
Hassan et al. 2008	Hausa	32	Sudan	Afro-Asiatic	3.1

Study	Population	Size	Country	Lang Fam	Frequency E-V32 (%)
Trombetta et al. 2015	Egyptians from Gurna Oasis	34	Egypt	Afro-Asiatic	2.9
Hassan et al. 2008	Nubians	39	Sudan	Nilo-Saharan	2.6
Trombetta et al. 2015	Northern Egyptians	49	Egypt	Afro-Asiatic	2.0

Table 5.5.9. Survey of E-V6 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-V6 (%)
Trombetta et al. 2015	Afar	25	Djibouti	Afro-Asiatic	56.0
Trombetta et al. 2015	Tigre	5	Eritrea	Afro-Asiatic	20.0
Trombetta et al. 2015	Wolayta	12	Ethiopia	Afro-Asiatic	16.7
Trombetta et al. 2015	Somali	6	Kenya	Afro-Asiatic	16.7
Trombetta et al. 2015	Gurage	7	Ethiopia	Afro-Asiatic	14.3
Trombetta et al. 2015	Nilotic Western Kenya	11	Kenya	Nilo-Saharan	9.1
Trombetta et al. 2015	Amhara	82	Ethiopia	Afro-Asiatic	8.5
Trombetta et al. 2015	Saho	94	Eritrea	Afro-Asiatic	8.5
Trombetta et al. 2015	Nara	15	Eritrea	Nilo-Saharan	6.7
Trombetta et al. 2015	Egyptian Berbers from Siwa	93	Egypt	Afro-Asiatic	6.5
Trombetta et al. 2015	Tigrai	32	Eritrea/Ethiopia	Afro-Asiatic	6.3
Trombetta et al. 2015	Cunama	20	Eritrea	Nilo-Saharan	5.0
Trombetta et al. 2015	Egyptians from Baharia	41	Egypt	Afro-Asiatic	2.4
Trombetta et al. 2015	Oromo	62	Ethiopia	Afro-Asiatic	1.6

Table 5.5.10. Survey of E-M33 Populations.

Study	Population	Size	Country	Lang Fam	Frequency of E-M33 (%)
Bučková et al 2013	Fula Linia	11	Chad	Niger-Congo	45.5
Bučková et al 2013	Kotoko	15	Cameroon	Afro-Asiatic	40.0
Bučková et al 2013	Masa	16	Cameroon	Afro-Asiatic	31.3
Bučková et al 2013	Songhai	23	Mali	Nilo-Saharan	30.4
Barbieri et al. 2012	South Samo	41	Burkina Faso	Niger-Congo	29.0
Bučková et al 2013	Kanuri	28	Nigeria	Nilo-Saharan	25.0
Bučková et al 2013	Fula Tcheboua	22	Cameroon	Niger-Congo	22.7
Rosa et al. 2007	Papel	64	Guinea-Bissau	Niger-Congo	20.3
Bučková et al 2013	Kanembu	11	Chad	Nilo-Saharan	18.2
Barbieri et al. 2012	Marka	33	Burkina Faso	Niger-Congo	18.0
Bučková et al 2013	Fula Tindangou	12	Burkina Faso	Niger-Congo	16.7
Hassan et al. 2008	Hausa	32	Sudan	Afro-Asiatic	15.6
Bučková et al 2013	Fula Banfora	26	Burkina Faso	Niger-Congo	15.4
Filippo et al. 2011	Burkina Faso Mande	152	Burkina Faso	Niger-Congo	14.4
Barbieri et al. 2012	Mandinka	15	Burkina Faso	Niger-Congo	13.0
Bučková et al 2013	Fula Zinder	16	Niger	Niger-Congo	12.5
Rosa et al. 2007	Nalu	17	Guinea-Bissau	Niger-Congo	11.8
Hassan et al. 2008	Fula (Fulbe, Fulani)	26	Sudan	Niger-Congo	11.5
Rosa et al. 2007	Balanta	26	Guinea-Bissau	Niger-Congo	11.5
Bučková et al 2013	Fula Diafarabe	23	Mali	Niger-Congo	8.7
Barbieri et al. 2012	Lyéla	40	Burkina Faso	Niger-Congo	8.0
Barbieri et al. 2012	North Samo	38	Burkina Faso	Niger-Congo	8.0

Study	Population	Size	Country	Lang Fam	Frequency of E-M33 (%)
Bučková et al 2013	Fula Bongor	25	Chad	Niger-Congo	8.0
Bučková et al 2013	Fula Diffa	13	Niger	Niger-Congo	7.7
Bučková et al 2013	Fula Balatungur	13	Niger	Niger-Congo	7.7
Rosa et al. 2007	Fula (Fulbe, Fulani)	59	Guinea-Bissau	Niger-Congo	6.8
Barbieri et al. 2012	Mossi	36	Burkina Faso	Niger-Congo	6.0
Bučková et al 2013	Fula Abalak	21	Niger	Niger-Congo	4.8
Rosa et al. 2007	Bidyogo (Bijago)	21	Guinea-Bissau	Niger-Congo	4.8
Barbieri et al. 2012	Bisa	40	Burkina Faso	Niger-Congo	3.0
Filippo et al. 2011	Burkina Faso Gur	183	Burkina Faso	Niger-Congo	2.7