

Journal: [www.jogg.info](http://www.jogg.info)

Originally Published: Volume 5, Number 1 (Spring 2009)

Reference Number: 51.006

# **Y HAPLOGROUPS, ARCHAEOLOGICAL CULTURES AND LANGUAGE FAMILIES: A REVIEW OF THE POSSIBILITY OF MULTIDISCIPLINARY COMPARISONS USING THE CASE OF HAPLOGROUP E-M35**

Author(s): *Andrew Lancaster*

---

# Y Haplogroups, Archaeological Cultures and Language Families: a Review of the Possibility of Multidisciplinary Comparisons Using the Case of E-M35

Andrew Lancaster

## Abstract

Archaeology, comparative linguistics and population genetics all have something to add to speculation about early human migrations, and the three disciplines often make reference to each other in broad terms. But, in reality, the results are often disappointingly indecisive. This article explores the case of Y haplogroup E-M35 (E1b1b1), which has so far mainly only been mentioned in a passing way in archaeological and linguistic debate, but which, it shall be shown, shows great promise as more detailed information about its phylogenetic structure, and its regional distribution becomes available each year.

## Introduction

Amongst other aims, this article will seek to respond to the following challenge posted in 2004 (actually referring mainly to work done in the 1990s) by Roger Blench (Blench, 2004a; further elaborated in Blench, 2004b):

*If DNA trees and language trees were indeed to correspond, then this would provide striking mutual confirmation for models of human prehistory (e.g., Gibbons, 2001). This plays well on the pages of Nature and hardly at all with most archaeologists and linguists (e.g., Pluciennik, 1996; McEachern [sic], 2000). Partly this is due to innate conservatism and the fact that no academic career points are to be made in being interdisciplinary, where established disciplines have developed internal structures. But it is also because DNA studies have not delivered credible results; linguists are faced with endless trees that show links quite contrary to established results and contradict one another from one paper to the next (cf. Chen et al. 1995; Blench 1999b, for some particularly egregious cases; McEachern, 2000). Claims for a genetic “clock” are endlessly revised and “theoretical” dates seem not to match any actual dates available.*

Blench rounded off his justification of such scepticism about population genetics by comparing it to “the classification of human races by head types, nasal indices, or many another now-forgotten indicator.”

Although population genetics, comparative linguistics, and archaeology all aim to discover the same underlying ancient population movements, these three disciplines tend to work mainly in isolation, using very different sorts of evidence.

For each of these three fields, the evidence is incomplete, and difficult to collect and interpret. So while the above quotation was certainly not without valid grounds, it is striking that Blench’s criticisms are of a type with which both archaeological and linguistic researchers are also often faced.

One thing which is particularly notable about genetics, is that this is the newest and fastest developing of these three types of research.

The present article seeks to review only some of the developments in these fields which seem to show potential for increased multi-disciplinary dialogue. It will specifically look at the much-debated area of Northern African and Middle Eastern contacts before, during and after the Neolithic period in human pre-history in these regions, the period when people learnt to process foods intensively, to herd animals, to make pottery, and perhaps most critically of all, to practice true farming, as opposed to collecting only foods found in the wild.

---

Address for correspondence: Andrew Lancaster,  
andrew.lancaster@skynet.be.

Received: January 17, 2009; accepted: April 14, 2009.

## 1. Defining the subject matter. Y Haplogroup E-M35, the Afroasiatic Language group, and the Neolithic.

### 1.a. The Names for Prehistoric Periods

Discussion about prehistoric humanity tends to be framed in terms of “material cultures” such as the *Palaeolithic* (old stone age), *Mesolithic* (middle stone age), and *Neolithic* (new stone age). These terms describe technologies identified by archaeologists, and not exact dates.

The literature also uses geological or climate-based epoch names, such as *Pleistocene* and *Holocene*. The exact definitions of these can also change with new data. Importantly, at least for any particular region, correspondences are understood to exist between these terminologies, even if debateable and confusing, and so a rough guideline is given in Figure 1.

Note that the full “package” of the Neolithic technology of the Levant and Fertile Crescent, includes pottery. In this region, which was a source of farming technologies

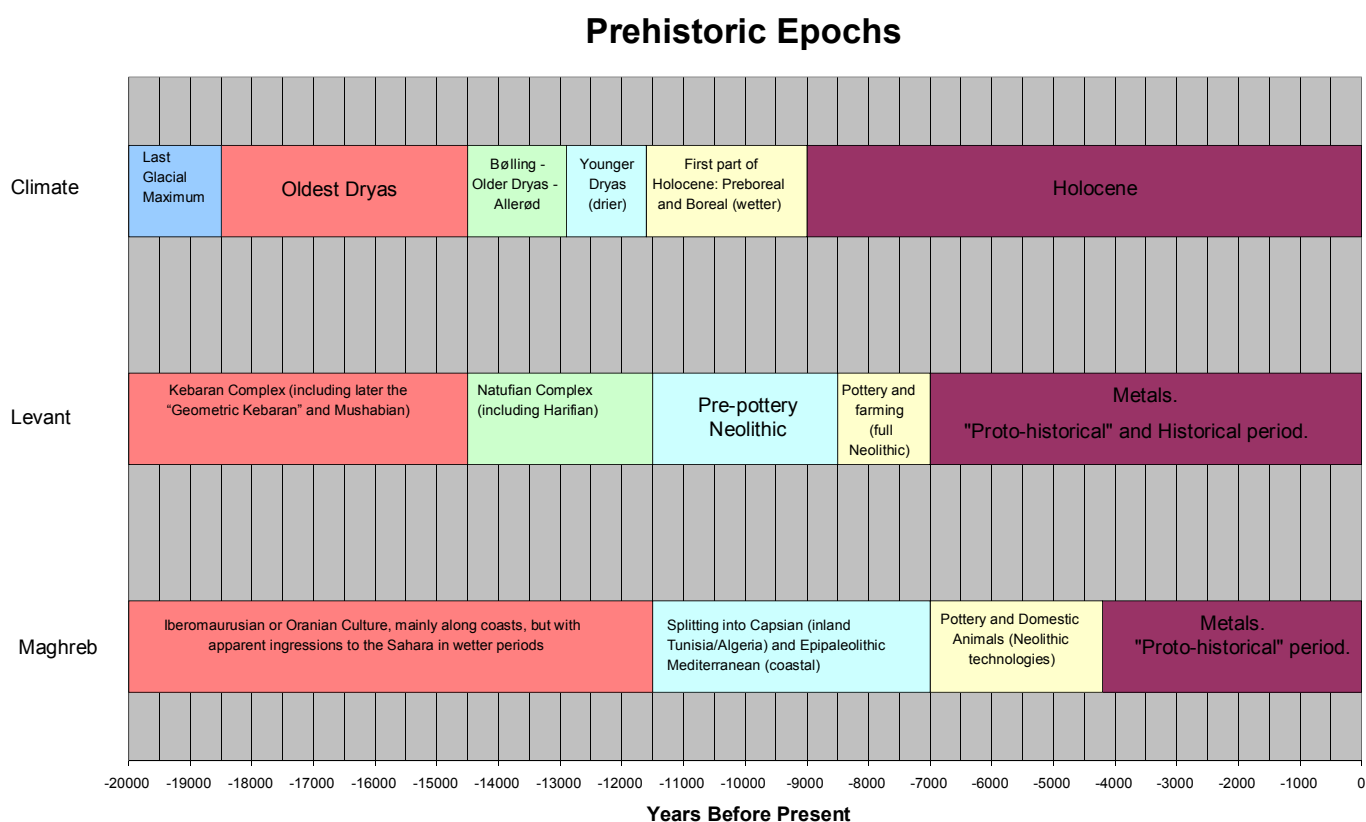


Figure 1. Prehistoric Epochs

for both Africa and Europe, there was a long period with farming, but without pottery. In Northern Africa this is reversed, with pottery appearing in the Saharan and Nilotic regions in the early Holocene, possibly also with “native cattle management” implying that some level of bovine pastoralism may even have developed in Africa first, although there are still doubts about this (Bellwood, 2005, p.99).

It is also important to note that the whole package of Neolithic technology, did not always transfer together to new regions. Farming itself may have entered Africa

later than it did Greece and Italy, and pottery appeared at almost the same time in Greece, Italy, and the Middle East.<sup>1</sup>

### 1.b.i. Y haplogroup E-M35: Definitions

In a series of articles, Cruciani et al. (2002, 2004, 2006, 2007) can be considered to have developed a set of standard theories concerning the most important ancient movements of people reflected in the modern

<sup>1</sup> Bellwood 2005, p.101: “Like the Californian Indians, early Holocene Egyptians did not need agriculture and did not seek it.”

distribution of E-M35 lineages. Until now, other authors in this area (such as Battaglia et al., 2008; Henn et al., 2008; Hassan et al., 2008) have primarily built upon this basis, with relatively small adjustments.

The map shown in Figure 2 is based upon Cruciani et al. (2004, 2007), and was made by the present author for the E-M35 article on the English Wikipedia. It adds some recent information from articles by Hassan et al. (2008, concerning E-V32 in Sudan) and Henn et al. (2008, concerning M293).

The current known phylogenetic structure of this haplogroup, also largely based upon the Cruciani et al. articles, can be seen in an up-to-date form on the ISOGG Y Haplogroup Tree, Haplogroup E page (ISOGG,

2009). Below, in Figure 3, the tree has been simplified and some initial remarks have been added concerning the geographic distribution.

For each clade, the “phylogenetic nomenclature” which can change with each new discovery of a UEP, or clade-defining mutation, can be replaced with a simpler “mutational nomenclature.” For example, E1b1b1 is a name which describes the position in the “family tree” of the branch defined by M35.

Apart from the following presentation of the Haplogroup E phylogeny, mutational nomenclature will be used in this article, except in some quotations, where the current equivalent terminology will be noted in square parentheses.

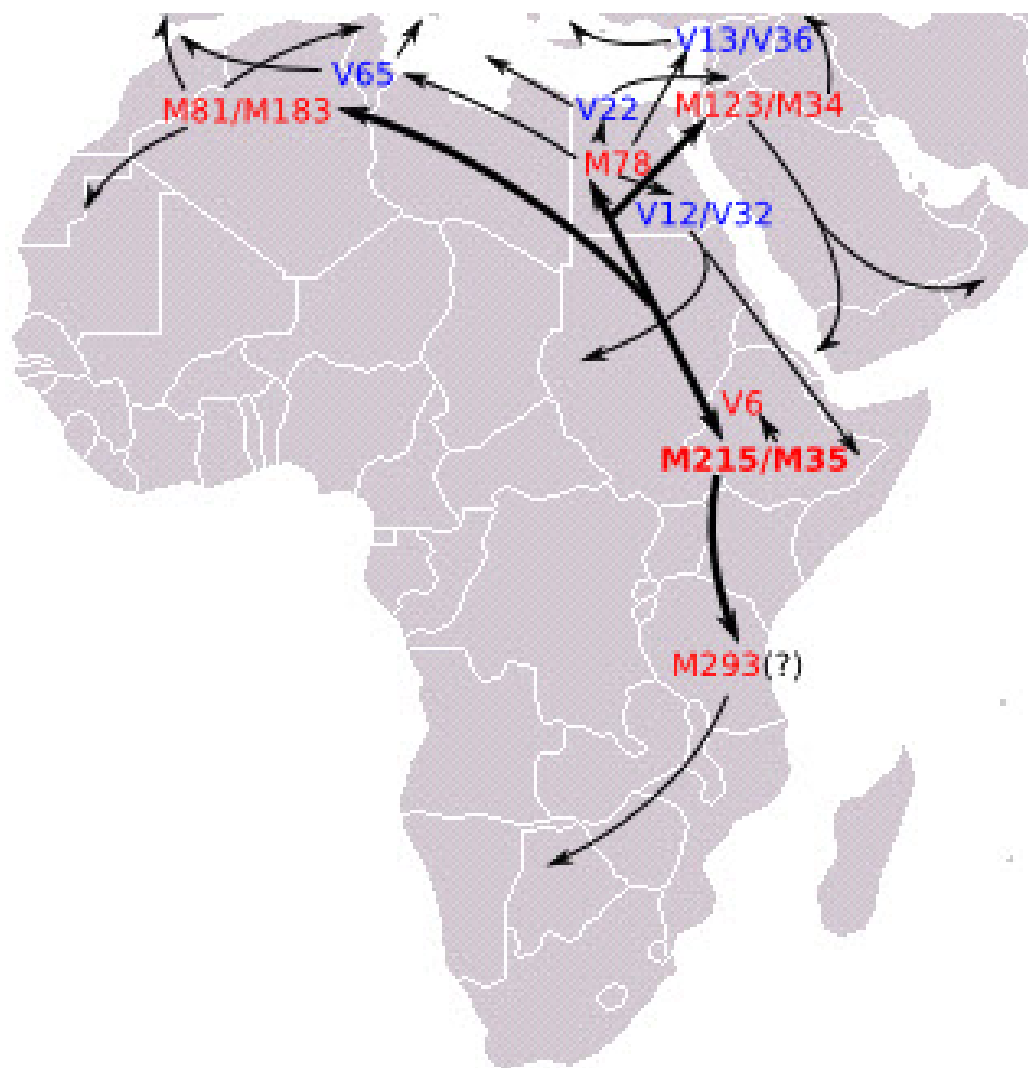


Figure 2. The origin and spread of Haplogroup E-M35.

<b>E1b1b</b>	<b>M215</b>	
<b>E1b1b*</b>		<i>Rare, but some found in Ethiopia (Cruciani et al., 2004) and Yemen (Cadenas et al., 2007).</i>
<b>E1b1b1</b>	<b>M35</b>	<i>The dominant subgroup of E-M215 (Cruciani et al., 2004).</i>
<b>E1b1b1*</b>		<i>Since the discovery of M293 (below) this paraclade appears to be most common in the Horn of Africa (Semino et al., 2004, Henn et al., 2008).</i>
<b>E1b1b1a</b>	<b>M78</b>	<i>Widespread from Egyptian “hub” (Cruciani et al., 2007; Battaglia et al. 2008). See the map in Figure 2.</i>
<b>E1b1b1a*</b>		<i>Rare, but found in small amounts over a wide area (Cruciani et al., 2007, Battaglia et al., 2008).</i>
<b>E1b1b1a1</b>	<b>V12</b>	
<b>E1b1b1a1*</b>		<i>Widespread from Egyptian “hub” (Cruciani et al., 2007).</i>
<b>E1b1b1a1a</b>	<b>M224</b>	<i>Rare (Underhill et al., 2001; Cruciani et al., 2004, 2006; but see also Onofri et al., 2006)</i>
<b>E1b1b1a1b</b>	<b>V32</b>	<i>Horn of Africa (Cruciani et al., 2004, 2006, 2007; Sanchez et al., 2005)</i>
<b>E1b1b1a2</b>	<b>V13, V36</b>	<i>Europe. Most common in parts of the Balkans and Italy, found throughout Europe, but possibly has a Near Eastern origin (Cruciani et al., 2004, 2006, 2007, Battaglia et al., 2008)</i>
<b>E1b1b1a3</b>	<b>V22</b>	<i>Widespread from Egyptian “hub” (Cruciani et al., 2007, Hassan et al., 2008).</i>
<b>E1b1b1a4</b>	<b>V65</b>	<i>Maghreb, for example Libya, Tunisia, Algeria (Cruciani, 2004, 2006, 2007)</i>
<b>E1b1b1a5</b>	<b>M521</b>	<i>Two found in Greece (Battaglia et al. 2008)</i>
<b>E1b1b1b</b>	<b>M81</b>	<i>Maghreb, especially Western Sahara, Morocco and Algeria (Arredi et al., 2004, Cruciani et al. 2004, Luis et al. 2004, Semino et al. 2004, Bosch et al. 2001)</i>
<b>E1b1b1c</b>	<b>M123</b>	<i>Strikingly scattered: Turkey, Oman, Ethiopia, Northern Portugal, Kabyle, Jordan, Jewish populations, etc. Probably originated in Levant or Egypt (Cinnioğlu et al. 2004, Arredi et al. 2004, Luis et al. 2004, Semino et al. 2004, Cruciani et al. 2004, Flores et al. 2005, Cadenas et al. 2007)</i>
<b>E1b1b1c*</b>		<i>Rare (Cruciani et al. 2004, Flores et al. 2005, Cadenas et al. 2007), but found in Northern Portugal (Flores et al. 2004, Gonçalves et al. 2005).</i>
<b>E1b1b1c1</b>	<b>M34</b>	<i>Dominant in E-M123</i>
<b>E1b1b1d</b>	<b>M281</b>	<i>Horn of Africa (Semino et al. 2004)</i>
<b>E1b1b1e</b>	<b>V6</b>	<i>Horn of Africa (Cruciani et al. 2004)</i>
<b>E1b1b1f</b>	<b>P72</b>	<i>(Karafet et al. 2008)</i>
<b>E1b1b1g</b>	<b>M293</b>	<i>Scattered in Eastern and Southern Africa (Henn et al. 2008)</i>

Figure 3. Phylogenetic structure of Haplogroup E-M35 in April 2009 (ISOGG, 2009).

As per convention:

1. An asterisk is used to denote the existence of individuals who have the defining mutation of a clade, but none of the mutations for any of its currently known sub-clades (For example, E-M35\* is the same as E1b1b\* and would indicate a person who has mutation M35, but *not* M78, M81, M123, M281, V6, P72, or M293).
2. Where such a lineage is not tested for all possible sub-clades “x” is used to note which

ones have been tested and excluded. For example “E-M35 (xM78, M81, M123)” means that a person has been tested positive for M35, but was negative for M78, M81, and M123, the most commonly tested sub-clades.

#### 1.b.ii. E-M35, Neolithic Technology and Afroasiatic Languages: Initial Remarks

As shall be shown, there are obvious reasons for considering whether Y Haplogroup E-M35 male lineages may have been present amongst peoples who spread the earliest Afroasiatic languages as well as the

earliest technologies associated with farming and pastoralism in the Middle East, Africa and Europe. This has also been noted in DNA surveys of the last decade. Initially, what was noted was a seeming link to the European Neolithic.

Semino et al. (2000) proposed that in Europe, haplogroup “Eu4” or “Ht-4”, equivalent to E-M35, represented “the male contribution of a demic diffusion of farmers from the Middle East to Europe.”

King and Underhill (2002) went further and showed an association between the distribution of these E-M35 lineages and the distribution of findings of Neolithic painted pottery and figurines, again focussing on diffusion from the Middle East into South-eastern Europe.

Underhill (2002) went still further, mentioning the potential relevance not only of the Middle-eastern link to Europe that this clade showed, but also the clade’s links to Africa:

*Considerable binary haplotype substructure exists within the M35 [E-M35] sub-clade with distinctive lineages in North Africa, the Mediterranean Basin and Western Asia. These provide additional reagents with which to infer specific episodes of population histories associated with the Neolithic agricultural expansion.*

Arredi et al. (2004) focussed on the Maghreb region of Mediterranean Northern Africa, seeing a parallel to what had already been observed for Europe, and similarly proposing that “Y-chromosomal genetic structure observed in North Africa is mainly the result of an expansion of early food-producing societies.” These authors also observed that “this expansion could have involved people speaking a proto-Afro-Asiatic language. These people could have carried, among others, the E3b [E-M35] and J lineages, after which the M81 mutation arose within North Africa and expanded along with the Neolithic population into an environment containing few humans.” They also noted the following options, which correspond to questions we shall discuss below:

*Under the hypothesis of a Neolithic demic expansion from the Middle East, the likely origin of E3b [E-M35] in East Africa could indicate either a local contribution to the North African Neolithic transition (Barker, 2003 [sic, should be 2002]) or an earlier migration into the Fertile Crescent, preceding the expansion back into Africa.*

Ehret et al. (2004) in a short letter to *Science*, perhaps represents the first published remark associating E-M35 with the *earliest origins* of Afroasiatic languages and Neolithic technologies, a subject this article intends to

address in more detail. They argued for a specific interpretation that we’ll discuss in more detail:

*A critical reading of the genetic data analyses ... supports the hypothesis of populations moving away from the Horn [of Africa] or southeastern Sahara northward to the Nile Valley, northwest Africa, the Levant, and Aegean [citing Underhill et al. (2001), Lucotte and Mercier (2003) and Semino et al. (2004)]. The geography of the M35/215 [E-M35] lineage, which is of Horn/East African origin, is largely concordant with the range of Afroasiatic languages. Underhill et al. [2001] state that this lineage was carried from Africa during the “Mesolithic.” The distributions of the Afroasiatic branches and this lineage can best be explained by invoking movements that originated in Africa and occurred before the emergence of food production, as well as after.*

The counter response by Bellwood (2004) was brief on this point: “The genetics papers quoted by Ehret et al. do not settle the matter. The Y chromosome evidence appears to signal complex two-way population movements, with very uncertain chronologies.” Bellwood’s observation is not incorrect, but there is much more to be said about the subject, as we shall seek to show in this article.

Keita (one of the co-authors of the Ehret et al. letter) and Boyce (2005) discussed E-M35 in this context in a little more detail:

*The genetic data, specifically the M35 [E-M35] subclade affiliated with haplotype V in Africa, can be related to the spatial range of much of the Afroasiatic linguistic phylum, which evidence suggests most likely originated in Africa; only one member (Semitic) is found in the Near East.*

With the possible exception of the last two citations, which touch on arguments that will be discussed more below, these comments were not detailed. They wrote of E-M35 as a promising subject for future discussion regarding Neolithic technology or Afroasiatic languages, but did not go far into exploring the possibilities. We shall see that, while proposing ancient links between Y haplogroups, languages, and archaeology in this case is very reasonable, “the devil is in the details.”

It should be noted that the farming “revolution,” while it was indeed of massive importance to humanity, happened over thousands of years, and Afroasiatic is a very old language group. So given the very large time ranges, many possible scenarios need to be considered.

Importantly, we also wish to try to go beyond asserting that E-M35 and Afroasiatic have similar modern

regional distributions. Therefore, this article shall also examine what is known of the larger phylogenetic (family tree) structure within which E-M35 is only one branching, as we must if we are to consider carefully how much genetics can add to debates in linguistics and archaeology.

### 1.c.i. The Afroasiatic Language Family

Afroasiatic is one of the world's largest language groups, including the liturgical languages Arabic, Hebrew, Ge'ez, Coptic, and Aramaic, as well as such well-known languages as Berber, Ancient Egyptian, Hausa, Oromo, and Somali. It is also particularly old, with its earliest branchings so early that they are upon the limits of what can uncontroversially be identified as a single language family. The geographic distribution of the Afroasiatic language family is shown in Figure 4.

It should be mentioned that a range of names are used in the literature for the Afroasiatic language group, often intended to emphasize particular points of view, which can be quite important when it comes to this particular language group. Commonly used alternative terms are "*Hamito-Semitic*" (now sometimes considered to have racist connotations, and therefore out of use in recent literature), the collapsed form "*Afrasian*"; and "*Afrasan*", a name chosen to de-emphasize Asia, and make it more clear that the group is mainly found in Africa.



Figure 4. The Afroasiatic languages. This map is adapted from the English language Wikipedia, with some migration arrows and sub-branch names added in their approximate ancient positions.

The academic situation in the linguistics field is more complex than in population genetics – with more authors having debated over a longer time, and a longer record of trying to integrate findings with those of archaeology. There is more controversy, and more diversity of opinion. The healthy debate is sometimes perhaps supplemented with the strong feelings involved with both the religious and supposed racial divides spanned by this ancient language group.<sup>2</sup>

Linguists dealing with very old language groups must often work with very incomplete and uncertain information. Linguists dealing with modern dialects must constantly deal with the complexities caused by loanwords and "*Sprachbunden*" (groups of neighbouring languages which borrow more than just words from each other). The human genome also results from complex mixing, but linguists do not have the benefit of anything equivalent to the rare genetic mutations known as UEPs which are very useful in defining family trees with a high degree of security, and they also do not have the equivalent of non-recombining Y-chromosome DNA, and mitochondrial DNA, parts of the genome which are passed unmixed to each new generation, allowing their phylogenies to be studied in isolation.

As a result therefore, there is no unanimous phylogenetic tree for Afroasiatic languages. So we shall only name the main distinct branches. These are:

- **Berber.** This language family is native to the Maghreb, an area of North Africa centred around Morocco, Algeria and Tunisia, but also including Libya, Western Sahara and Mauretania, as well as parts of Mali, Niger, and Egypt. Berber languages were dominant in this enormous region before Arabic arrived with the spread of Islamic civilization.
- **Semitic.** Originally Middle Eastern, but now a strong presence in Africa because of the aforementioned spread of Arabic. In the Fertile Crescent Arabic largely replaced various forms of another Semitic language, Aramaic, which had in turn replaced other Semitic languages, such as Hebrew. This pattern of closely related Semitic languages replacing each other in the area goes back at least as far as Akkadian, luckily known to at least some extent from the ancient written records of this region. An ancient branch known as South Semitic is also present in Ethiopia and Eritrea, and this branch has also historically been present in southern parts of the Arabian Peninsula.

<sup>2</sup> Followers of cultural debates as they appear in the popular press will be interested to see Martin Bernal's *Black Athena* in the references for this article, but this particular field is one where it is difficult to avoid some level of controversy, and Bernal's opinions in this subject are unusual, but not particularly controversial.

- **Egyptian**. A modern form of ancient Egyptian barely survives in the form of the Coptic liturgical language of Egypt's native branch of Orthodox Christianity. But in this case we once again have the benefit of very early written records.

The above three were once very widespread and are very clearly defined language groups. They are generally grouped together as a northern branch. There is far more disagreement concerning the diverse southern Afroasiatic languages, the study of which is limited to modern times and modern dialects:

- **Chadic**. A language family including Hausa, is spoken in several countries around Lake Chad, to the south of the Sahara. This is a linguistically diverse and dynamic area. Many non Afroasiatic languages are also found there, but for example Hausa has become a second language for many people in this region.
- **Beja**. Spoken mainly in Sudan, but also Egypt and Eritrea, between the Nile and Red Sea. This language is often said to be a branch of Cushitic.
- **Cushitic**. A major language group of the Horn of Africa and Northern Kenya, including Oromo and Somali.
- **Omoti**. A smaller language family sometimes thought to be related to Cushitic, and again found in the area of the Horn of Africa, scattered in areas such as the Omo river valley in the South West of Ethiopia.

For this article, we'll try to define the agreements and disagreements at the same time as we also begin making multidisciplinary comparisons, from the perspective of an interest in Y chromosomal lineages (patrilineages), especially E-M35.

#### 1.c.ii. Afroasiatic Compared to E-M35: Initial Remarks

There are several relatively uncontroversial proposals concerning the ancient movements of Afroasiatic languages, each of which we can immediately compare to Haplogroup E-M35 and its sub-clades in population genetics:

- Both E-M35 male lineages on the one hand, and Afroasiatic languages on the other, are seen by specialists in the two respective fields as having moved pre-historically within what Cruciani et al. (2007) refer to as a "bi-directional corridor" along the Nile and/or the western coast of the Red Sea,

from the Sinai and Mediterranean, to the Horn of Africa.

- Also in both fields, the population who lived in this "corridor" has obvious ancient connections to a Near Eastern (*Levantine*) corridor spreading into the Fertile Crescent, where such clades as E-V13 and E-M34 may have their origins (Middle Eastern Semitic speaking populations typically also have a smaller presence of E-M35 Eurasian lineages than most African populations of Afroasiatic speakers, with the apparent exception of Chadic. See below).
- There is also in both cases an obvious connection to the West of the Nile, in the direction of the Sahara and along the southern Mediterranean, in the *Maghreb*. The people there are strongly associated with the distinct E-M81 and E-V65 sub-clades of E-M35, and the Berber branch of Afroasiatic languages.
- In both population genetics and linguistics, the Levant branch (re-)entered Africa in historical periods: both in Egypt, which is now Arabic speaking, but was once home to the Egyptian language; and Ethiopia, where Semitic (Southern Semitic) languages now dominate in some areas which are believed to have once been Cushitic speaking. Indeed there must have been much back and forth movement at these two points of contact between Africa and Asia.
- In both fields there are signs of another scattered branch to the south, at least as far south as Tanzania, which has apparently been mostly overrun by later language expansions: both the Bantu expansion (Phillipson, 2002) and the expansion of Nilo-Saharan pastoralist populations southwards into the area (Ehret, 2002b). In the case of languages, these southern remnants are grouped in the so-called Southern Cushitic group. In the case of E-M35 lineages, the remnants that can be seen today are in the form of the distinct E-M293 sub-clade, defined by Henn et al. (2008).

Of course, when we find such similar patterns in both languages and Y lineages, this can give us at least some confidence that the same population movements caused both patterns. However it is important to be cautious. For this region of the world, with only a few possible migration routes avoiding the Red Sea and Sahara, the chances are high that similar migrations have happened many times.

With this in mind, it is proposed that we may at least assume a strong likelihood that E-M35 male lineages were involved in at least many of the migrations and cultural transmissions which caused the present and



historical distributions of the Afroasiatic languages. From this relatively safe starting point, we can tentatively start to explore some of the controversies in the field from a Y chromosome perspective.

### 1.c.iii. Some Controversy Concerning Attempts to Identify Places of Origin

Apart from the shared linguistic and genetic patterns above, it is also striking that both the leading genetics researchers of E-M35 (Cruciani et al., 2002, 2004), and the majority of linguists writing about Afroasiatic propose the Horn of Africa as a homeland for the subject of their discussion. What is even more striking is that in both cases the key argument is that the higher diversity (of genetic haplotypes or languages) found in the Horn of Africa indicates an older presence.

In both linguistics and genetics however, there is an inherent weakness in relying too much upon the argument that the modern area with most diversity must be a homeland. This cause for doubt applies equally in both fields, and in both fields it needs to be kept in mind.

- Deceptively high relative diversity. This can come about if a region is relatively more clannish and divided, something which often happens in highland terrains, populations centred around oases, or if a region is made up of immigrants from many regions, which could for example happen if there were large political, technological or climatic disturbances happening in neighbouring regions.
- Deceptively low relative diversity. This can come about for the opposite reasons. For example in an extreme case when one small isolated community expands rapidly, perhaps because of new food resources being discovered, perhaps through technological advance and/or climate change, some (or even maybe just one) of the founding patrilineages of that original small community can come to dominate simply through chance or “drift.”

There are obvious reasons to believe that both of these misleading effects may apply in this very ancient region of human migration, Northern Africa and the Middle East. Two of the most obvious examples of immigrations for example are the above-mentioned relatively recent entries of Semitic languages and haplotypes from across the Red Sea into Ethiopia, and over the Sinai into Egypt. These specific cases are of course obvious enough to identify. But there could have been other such effects, complicating modern attempts to reconstruct ancient migrations:

- The Nile area is one of the world’s greatest human migration corridors, that has seen peoples moved by empires, both local and foreign, and before there

were empires, by the spread of cultures with new technologies, again both local and foreign. So, Egypt’s genetic diversity must have been pushed and pulled. It is a challenge of any explanation to try to discriminate between the different possible causes of its genetic diversity.<sup>3</sup>

- Much of the Horn of Africa is a highland area with striking ethnic diversity even today, and where the geography plays an obvious role in maintaining that diversity.
- Apart from the obvious likelihood of migrations through Egypt, ancient movements of people from the Nile region, and indeed other surrounding regions, to the Horn of Africa are also very likely, and are indeed part of the standard theory for Cruciani et al. (See for example V12 in the map above, a lineage which dominates Somalia, but is thought to have originated in or near Egypt).
- The Nile area is arid, with fertile areas sometimes scattered around at great distances from each other. So in this region there must have been many pockets of population who were isolated during long periods. Isolated areas with small populations are more susceptible to having decreases in genetic diversity because of “genetic drift.”
- The whole area has been subject to large fluctuations in climate, with extreme dry periods interspersed with much greener periods. In particular, the Sahara runs along the Western side of the entire long “bi-directional corridor” under discussion, and the Sahara is known to have occasionally had a much higher capacity for supporting human cultures during several phases of pre-history (See e.g. Barker, 2000, F Hassan, 2000). Climate in this large region must have inevitably “pushed” and “pulled” populations in and out of the Sahara, as well as the Nile, the Maghreb, the Red Sea coast region, the Lake Chad region, and the Levant—the exact areas where we find E-M35 lineages and Afroasiatic languages today. Because all languages and genetic populations being studied today live during the present dry period in the Sahara, the Sahara’s original diversity is impossible to judge based upon these modern populations alone.

It will of course give an unparsimonious and *ad hoc* impression whenever any researcher feels moved to propose that the Sahara, a place whose ancient population we can hardly know about, is the original home to later cultures, languages and genetic lineages.

<sup>3</sup> “The complexity of the E-M35 fraction in Egypt may have been enhanced by several episodes of backflow, beginning with the introduction of agriculture into Africa, and, later, various historical events, such as the Greek, Roman, and Arab occupations” (Luis et al. 2004).

Nevertheless this is always a possibility that can not be ignored, given what we know of the prehistoric climate and archaeology of the area. For the time being at least, archaeology is the main field of research that has any concrete evidence at all to offer concerning this possibility.

#### 1.c.iv. Afroasiatic and Archaeology

These concerns force us to look for any available complementary sources of information with which we can cross-reference and gain perspective. In comparative linguistics as in population genetics, the chance to study an old population, either because of written language, or a well-preserved DNA sample, is unfortunately rare. Researchers must therefore look to archaeology. In order to compare to archaeology though, we first need to determine approximately how long ago the ancestors of these modern languages and haplogroups began dispersing.

While dating techniques in population genetics are far from uncontroversial, Cruciani et al. (2007) do provide a widely accepted estimation for the time at which E-M35 started spreading, at approximately 22,000 years ago (20.9-23.9 ky).

Glottochronology, the attempt to develop mathematical methods of calculating the age of languages, is notoriously controversial, but in the case of Afroasiatic, researchers are blessed with some of the oldest written records known. Both Egyptian and Akkadian, the oldest attested Semitic language, were already in Egypt and Mesopotamia respectively in the 3<sup>rd</sup> millennium BCE. Furthermore, many linguists believe that Egyptian and Semitic are in the same Northern sub-branch of Afroasiatic (along with Berber), although even when they first appear they are very distinct (e.g. Blench, 2006). This in turn means that they shared a common ancestor a long time before they first appear, and that this common ancestor was a long time after the common ancestor of all Afroasiatic languages. For this reason it is not controversial to propose that Afroasiatic may have even begun dispersing in the period 7,000-10,000 years ago, which brings us to the period of the Natufian culture in the Levant, a culture which preceded the first Neolithic farming technologies in the Middle East.

With the above-mentioned need for archaeological perspectives in mind, it is very notable that most archaeologists believe that in this part of the world, critical Neolithic-associated technologies such as farming and goat or sheep herding, originated not in Africa, but across the Sinai in the Levant and Fertile Crescent, and were distributed from there to Africa. This however leaves open the possibility that the Natufian material technologies which apparently led to the Neolithic in the Levant had "African Roots" or even

African contemporary equivalents and "cousins." This is a common but disputed suggestion, as we shall discuss further below, but unfortunately the archaeology of Africa has not yet been intense enough to allow the sorts of discussion which are possible for Europe, and the Middle East (Barker, 2002; Bellwood, 2005, p.97).

Linguistics and archaeology have a track record of at least some multi-disciplinary debate between the two fields, and there is therefore a body of literature which covers their joint efforts and disputes. As it happens, archaeologists tend to try to explain languages in terms of the material cultures they study, the most obvious one being the Neolithic technology "package." Linguists tend to approach from another direction, comparing reconstructed vocabularies (concerning things like plants, animals, tools, traditions etc) to archaeological evidence. This is a useful approach to try to squeeze more conclusions out of the little evidence available.

From this material we can summarize two broad options most prominently proposed by linguists concerning Afroasiatic, as championed by the linguists who have most addressed themselves to the archaeological question:

1. Alexander Militarev (2002, 2005) argues that Afroasiatic did not disperse from Africa, but rather the Levant. This is a minority position amongst linguists. He would equate proto-Afroasiatic with the pre-Neolithic Natufian culture, making it approximately 10,000 years old, and he would associate its expansion with the later spread of farming technology to Africa.

One distinctive aspect of this theory for linguists is the argument that Militarev sees evidence for a very ancient linguistic connection between Afroasiatic (*Afrasian* in his work, following Diakonoff) and some languages of Eurasia, for example in the Caucasus. Militarev works in the linguistic tradition of trying to find the faint signs of language connections going back before the recognized language families, to an ancient proposed language referred to as "Nostratic." This approach is controversial in linguistics because the seeming links which can be found between different language groups in this way are often so faint that many alternative theories can be found to fit the reconstructions. Furthermore, Martin Bernal (1987) for example apparently accepts this Nostratic connection, but explains it as a language dispersed from Africa, and similarly Keita (2005) proposes that Afroasiatic and Nostratic may be sibling language groups. Other linguists, apparently less convinced about Nostratic, are more willing to consider connections between Afroasiatic and Elamite, in ancient Persia, and even Dravidian,

in the Indian sub-continent (e.g. Blench, 2006). Most generally, even if there is strong evidence linking Afroasiatic to Asian languages, why should the very ancient implied language “super family” not have dispersed from Africa to Asia, rather than the other way around?

Perhaps more importantly, Militarev (2005) also claims that Afroasiatic animal and plant words imply that the language originated in the Levant. In the archaeological field, Militarev finds favour with authors like Peter Bellwood (Diamond and Bellwood, 2003; Bellwood, 2004, 2005) who, in the tradition of Colin Renfrew, see a link between modern languages and the spread of Neolithic early farming and pottery technologies. This is the so-called “farming/language dispersal hypothesis” (Bellwood and Renfrew, 2005 eds).<sup>4</sup>

2. Christopher Ehret similarly equates an early Afroasiatic (*Afrasani* in his work) language with the Natufian culture studied by archaeology, but he believes that this was a proto-Semitic branch of the language family, and that Afroasiatic as a whole had already been in existence long before the Natufian, at least as far back as 15,000 years ago (see Militarev, 2005 for his perspective on these ideas).

In this scenario, Afroasiatic originates in (or near) the Horn of Africa, and was the carrier of a new and more intensive use of plant foods found in the wild, eventually including grass seeds. This way of life was an essential prior step to the Natufian, and in turn to true farming. Pastoralism, a minority proposes, may also have begun very early in Northern Africa, in what is now the Sahara (e.g. Barker, 2002). In terms of archaeological cultures Ehret has a clear proposal, apparently receiving support in comments published by Bar-Yosef (1987) and Keita and Boyce (2005): he believes the Natufian comes out of a mixing of the Mushabian culture, of the Negev and Sinai, with the related Kebaran culture, both of which preceded the Natufian in the Levant (Ehret, 2002, p.38). That the Mashubian blended with the Kebaran, and that the Natufian is descended from this mixture, is not particularly controversial, but it is not unanimously agreed upon within archaeology that either the Natufian or the Mashubian cultures are best explained as being wholly or partly from Africa as Ehret proposes. This is the theory presented in, for example, Bar-Yosef (1987) based upon lithic technologies.<sup>5</sup> Bar-Yosef (personal communication)

<sup>4</sup> See Barker (2002): “A Near Eastern origin obviously chimes best with the archaeological model of Neolithic demic diffusion from the Near East into North Africa.”

<sup>5</sup> In particular, Bar-Yosef mentions the microburin technique and “microlithic forms such as arched backed bladelets and La Mouillah points”. See Barker (2002): “Though linguistic scholars debate whether the language originated in North Africa or the Levant, we

also believes that evidence for the early introduction of the Sycamore Fig, into the Levant around the time of the Natufian gives additional strength to this theory.<sup>6</sup>

Coming to linguistic evidence, Ehret points out that Militarev’s reconstructed proto-Afroasiatic contains no indisputable common vocabulary for farming itself, while the major branches do. He believes that this shows that farming originated after Afroasiatic started to disperse from its original homeland.

So the linguistics literature raises a range of possibilities about the connections of Afroasiatic to the early Neolithic archaeological cultures. All versions of them obviously allow room for at least some branches of Afroasiatic to have played a role in the spread of the very first farming technologies in the Neolithic of both the Levant and Africa. In the Ehret scenario, Afroasiatic was already old when a branch of it entered the Levant, along with people who helped trigger the slow changes which led to the Middle Eastern Neolithic. Militarev makes Afroasiatic younger, but still significantly older than Neolithic technology.

What other scenarios might be worth considering, if any, apart from the two most commonly cited categories above? The majority of linguists accept an African origin for Afroasiatic, but Militarev and Ehret both believe in a relatively old age for Afroasiatic. See, for example, Fakri Hassan (2002) for a cautious summary of the evidence. If we were to doubt this great age postulated by authors like Militarev and Ehret, but accept that proto-Afroasiatic was both pre-agricultural, and African, then the implication would be that Afroasiatic was a hitchhiker that entered very effectively into the whole complex of peoples involved in what became the Neolithic revolution and then proceeded to spread with them, especially in Africa, but also to a significant extent in parts of the Middle East. Is this a realistic scenario? We shall consider below what genetics could add to such a consideration.

## 2. What Can Genetics Add to the Discussion?

At first sight, the relatively young field of population genetics seems to offer only limited assistance.

can at least point to the similarities in the respective archaeological records of the Natufian culture of the Levant and of contemporary foragers in coastal North Africa across the late Pleistocene and early Holocene boundary.”

<sup>6</sup> Bones of the Egyptian fruit bat, *Roussetus Aegyptiacus*, which eats figs, are found in the Levant only from the Natufian onwards. The Sycamore Fig (*Ficus Sycamorus*) appears to have been first introduced to the Nile region from its native habitat much further south in Africa. The Egyptian and Levantine versions of this plant are parthenocarpic, requiring the help of man to reproduce. The closest place where a wild wasp helps fertilize these figs is in Sudan. Stored parthenocarpic fig remains (*Ficus Carica*) have been found in Gilgal I, an early Neolithic village, located in the Lower Jordan Valley, and dating to 11400 to 11200 years ago (Kislev et al., 2006).

However, turning to the latest chronology of E-M35 presented in Cruciani et al. (2004, 2007) the above two theories can both be made to fit with varying degrees of success. The following scenarios are attempts to show multidisciplinary correspondences by the present author.

#### Scenario 1. Very early Afroasiatic, originating in Africa

Using Ehret's scenario, E-M35 and Afroasiatic must have been together from their first dispersals out of the general area of the Horn of Africa. Branches of Afroasiatic fit very well with the branching in E-M35:

1. A large number of E-M35 lineages came north to Egypt, which is where E-M78, for example, is thought to have come into existence. The Nile itself was not always highly populated, depending upon the climate in different eras. The changing rainfall may have pushed and pulled populations towards the Sahara and Levant, and back again.
2. E-V65, for example, is a clade of E-M78 which appears to be native to areas west of Egypt, in the Maghreb. E-M81 is a clade of E-M35 which is more common than E-V65 in the Maghreb, but with its centre of dispersal apparently further west, possibly having split off somewhat earlier from the main "bi-directional corridor" near the Nile.
3. E-M123, for example, is a clade of E-M35 which probably originated in the southern Levant, just to the northeast of Egypt. Many branches of E-M78, from nearby northern Egypt are also present in the Semitic speaking populations of this region. That the Levant is not dominated by E-M35 is not surprising in this scenario, because it is generally accepted that the Natufian culture, while important, only contributed to the bigger changes going on around the Fertile Crescent (Bar-Yosef, 2002), and the Semitic languages may have remained isolated in the Levant until not long after the written record starts in Mesopotamia (Zarins 1990).

#### Scenario 2. Early Afroasiatic, originating in Levant

Fitting Militarev's scenario to the genetic evidence requires more thought, showing how genetic evidence can be very useful for narrowing down the likely alternatives, and also helping show potential weak and strong spots of theories in other disciplines. Because of the extremely strong links between Afroasiatic and E-M35 in such diverse areas as Morocco and Somalia, it is very difficult to imagine any scenario where Afroasiatic and E-M35 did not disperse together from their earliest origins, *at least in Africa*. Outside Africa, the connection is less clear. Although E-M35 is common in parts of the Middle East and Mediterranean, no large

regional non-African population is truly dominated by E-M35 lineages, and hence even amongst the Afroasiatic Semitic speaking populations, such as modern Arabs, the levels of E-M35 found there, could reasonably be explained by theories which require no major events of the types that archaeologists and linguists might easily track.

Here then is a scenario attempting to fit Militarev's scheme to the genetic evidence:

1. E-M35 spread northwards from the Horn of Africa, or nearby, at the right time to be involved in many phases of dispersion of new technologies in Northern and Eastern Africa, certainly at least including technologies which had crossed the Sinai from outside Africa. This implies that it was certainly near the Sinai at an early date, close to the Levant. By the time true farming arrived in Egypt, for example, E-M35 is likely to have been present there for a long time already. Cruciani et al. (2007) propose that its sub-clade E-M78 was founded somewhere near Egypt, very roughly 10,000 - 20,000 years ago.
2. Afroasiatic languages could conceivably have begun to disperse much later than E-M35, and from a different homeland, only later entering into the E-M35 areas of Africa. Such a language expansion would not necessarily require a true migration of significant proportions of the population ("demic diffusion") which would perhaps involve the replacement of E-M35 male lines. Ethnic groups and regions can take up new languages more easily than they take up whole new sets of male lines.
3. In the simplest such scenario then, Afroasiatic languages first entered Africa in a region with a high E-M35 population, then integrated with that population, before dispersing further. In particular, Afroasiatic languages in Africa could have come from the Levant into Egypt, and then started to disperse together further with E-M35 lineages, including E-M78 lineages, from this starting point, more-or-less independently of Semitic languages dispersing outside Africa.

Does this fit with Militarev's own ideas? Militarev writes (Militarev, 2002, p. 135) that the "Proto-Afrasian language, on the verge of a split into daughter languages," meaning, in his scenario, into "Cushitic, Omotic, Egyptian, Semitic and Chadic-Berber", "should be roughly dated to the ninth millennium BC." E-M35 was almost certainly present in Egypt by this period, with perhaps a smaller presence in the Levant already. This is shown not only by looking at the data for E-M78 mentioned above, but also because the E-V12 sub-clade of E-M78 is said by Cruciani et al. (2007) to have come

into existence about 12,500-15,200 years ago in the Egyptian area, and only then moved south to become a very prevalent clade in the Horn of Africa.

In this Militarev-inspired scenario above, the link between Afroasiatic and E-M35 needs a more complex explanation than in the Ehret scenario.

Unless we make ad hoc assumptions that technology could only move in one direction after a certain date, E-M35 might have left Africa and entered the Levant during many periods, before, during or after the Natufian. This, however, raises a question again of whether Afroasiatic might be significantly younger than in the proposals of Militarev. For example Juris Zarins (1990) has proposed that Semitic languages may have their genesis in early pastoralist groups neighbouring the Fertile Crescent farmers. The Harifian culture, a Negev relative of the Natufian, is a likely contributor to this pastoralist culture, and according to Ehret's proposals a recipient of African technology and culture. Might Afroasiatic languages, and indeed E-M35, have remained solely in Africa until this relatively late period?

### Scenario 3. More recent Afroasiatic, originating in Africa.

As a final scenario, we compare the genetic evidence to a synthesis intended to represent the possibility that Afroasiatic (and Semitic) are younger than they are in the well-known proposals of Ehret and Militarev. This seems necessary for completeness, given that, as mentioned above, many linguists appear uncomfortable with such deep chronologies.

1. If the starting point for Afroasiatic would be anywhere within the corridor stretching from the Nile Delta to the Horn of Africa, and if its dispersal started well into the Holocene (more recently than in the scenarios of Ehret and Militarev), then E-M35 is almost certain to have been present in that "proto-Afroasiatic" population.
2. While E-M123 has the appearance of being a subclade of E-M35 that may have originated in the Levant quite early, it is still basically true that the E-M35 diversity found in the Levant today is a branch of the same diversity found in Egypt, which also has a high E-M123 diversity (Luis et al., 2004).
3. This would make Afroasiatic and probably also E-M35, both representatives of the African contribution to the full development of the Neolithic, but not necessarily playing any special early role in the pre-Neolithic.

Examining how well these scenarios work in matching linguistic, archaeological and genetic patterns, some

weaknesses seem to appear which will require more detailed discussion below. For example:

1. While it allows for the entry of E-M35 into the Levant, Scenario 2, based upon Militarev, certainly does not explain it. It requires us to say that Afroasiatic languages and E-M35 moved in opposite directions between the continents.
2. A problem seems to arise when we come to look at the genetic distinctness of the population of the Maghreb. In Scenario 2, and Scenario 3 the ancestor of the Berber languages came from the East (Nile or Levant), in or after the Neolithic. The area has certainly been influenced by the Nile and Levant as we know from more recent millennia, but the genetic impact, while clear, is overwhelmed by a very dominant but also very *distinct* E-M35 population, with E-M81 in particular showing no signs of Holocene common ancestry with other modern E-M35 subclades. This is not an *insurmountable* problem, but it does force us to invoke causes which are difficult to confirm or deny. For example:
  - Despite its large impact, the founding population that brought E-M35 lineages might have been established long before Afroasiatic languages arrived, with the Afroasiatic in this particular region bringing a new language, and possibly new technologies, but not the particular E-M81 lineage that has become so prevalent in the area.
  - A founding population may have brought E-M35 lineages and Berber languages together but this population may have been relatively small, given the isolation of the area, meaning that "founder effects" might make the population there appear more genetically independent than it is from the Nile.

So whether we should favour one of the above scenarios depends partly upon how genetically diverse and old the Maghreb population of today seems to be. This shall be discussed below.

We can already observe that population genetics helps us see the potential weak and strong points of theories, even if only by showing which assumptions are required to fit with the current state of information. It also presents us with ideas about what types of future evidence might favour those scenarios. In a difficult field such as studying ancient populations, any single type of evidence can only normally work in conjunction with other sources of information.

### 3. The African origins of E-M35

In the above we have relied upon the assertion that E-M35 originated in Africa, far from the Levant, as is

posited by Cruciani et al. (2007) and all other peer-reviewed population genetics studies. The oldest geographical location of the E-M35 lineage must be of interest to linguists and archaeologists. This is because, as can be seen above, linguists and archaeologists are very interested in debating precisely the question of how important Africans were in developing and dispersing the technologies which led to the Middle Eastern Neolithic, the African Neolithic, and the Afroasiatic family of languages; and in time to one of humanity's

greatest complexes of civilizations in Egypt, the Fertile Crescent and surrounding areas.

We therefore now turn to deeper examination of the part of the family tree of humanity's male lines within which E-M35 sits, in order to consider how clear the evidence is for African origins. First, we show the phylogeny for Haplogroup E from ISOGG (2009) again, this time to show E-M35 and its ancestral context, in order to assist discussions about its apparent geographic origins, along with those of its closest relatives.

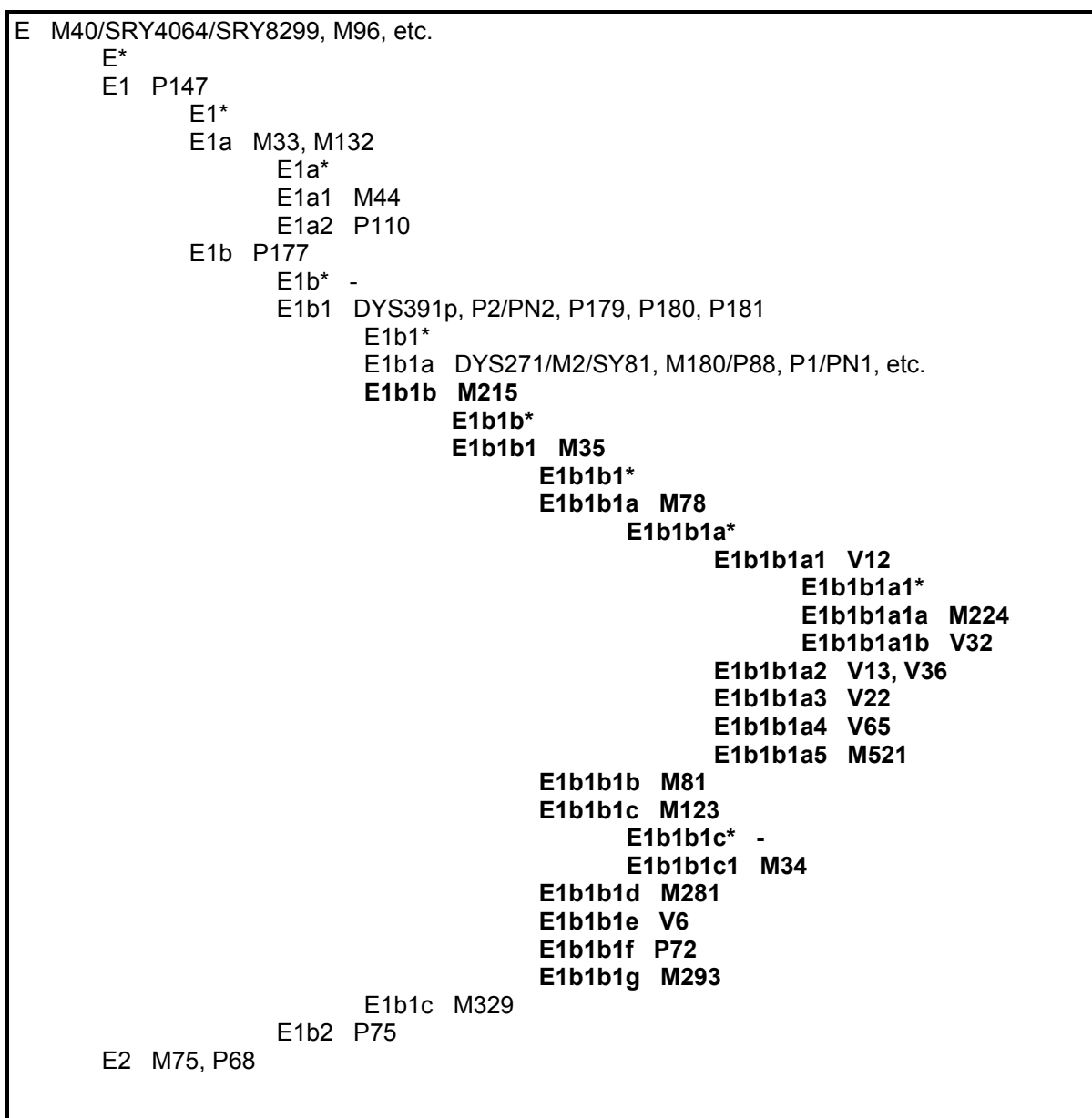


Figure 4. Phylogenetic tree for Haplogroup E with Haplogroup E1b1b (E-M215) and its important subgroup, E1b1b1 (E-M35) showing in bold.

### a. Branches of E-M35

- All branches of E-M35 occur in Africa, and only a few are suggested to have originated just outside Africa in the Levant (E-V13 and E-M123, discussed in more detail below).
- Furthermore, some branches of E-M35 are *only* found in the Horn of Africa, for example E-V6 (Cruciani et al., 2004) and E-M281 (Semino et al., 2004).
- Even more striking, one major clade of E-M35, E-M293 is apparently found only quite far to the *south* of the Horn of Africa (Henn et al., 2008), scattered southward throughout Eastern Africa and well into Southern Africa.
- There also seem to be other branches waiting to be discovered in the Horn of Africa, reflected in the relatively high number of people in that region who are E-M35\*, positive for M35, but not positive for any of the defined sub-clades such as M78, M81, M123, M281, V6, M293 etc. With the discovery of E-M293 in Southern and Eastern Africa, the Horn of Africa now appears to have by far the highest E-M35\* concentration, as can be seen by comparing Henn et al. (2008) to Semino et al. (2004) and Cruciani et al. (2004).<sup>7</sup>

### b. E-M35 as a part of E-M215

The modern population of lineages known as E-M35 has sometimes been referred to as E-M215. Two defining mutations, M35 and M215 are almost always found in the Y chromosomes of the same men. However Cruciani et al. (2004) found 2 Amhara Ethiopians (out of 64) who were E-M215 positive and E-M35 negative, which in the mutational nomenclature makes them E-M215\*. More recently, Cadenas et al. (2007) found one more such individual out of 62 individuals tested in Yemen, across the Red Sea from Ethiopia. Yemenites and Amhara share a common history of speaking languages in the South Semitic language family.

### c. E-M35/E-M215 as a part of E-P2

E-M35, and E-M215, (essentially the same haplogroup) are part of the older E-P2 clade. There are other surviving E-P2 male lines to be found in the modern world.

- E-M2, is also sometimes referred to as E-PN1 or E-P1. E-P1 and E-M35 dominate the modern population of E-P2 lineages. This “sibling” clade to E-M35 is certainly African. It is a large and

<sup>7</sup> But see Adams et al. (2008) concerning Asturias in Spain, where two out of 20 people were tested and found to be “E-M35\*” [E-M35 (xM78, M81, M123)].

widespread clade associated with Western and Central Africa and also with the spread of languages and cultures from that area into Southern Africa, in the Bantu expansion (see for example Underhill et al., 2001). Looking to Europe, Adams et al. (2008) found four E-M2 individuals out of 1140 people tested in Iberia. These were all in the south: Mallorca, Valencia, and Southern Portugal. In Sicily, Di Gaetano et al. (2008) found no E haplotypes apart from E-M35. In the Middle East, Semino et al.’s very large 2004 survey found two in Iraq, out of 218 people tested there. Cadenas et al. (2007) also reports small numbers in the Arabian Peninsula.

- E-M329 is a much smaller sibling clade of E-M215 and E-M2. Semino et al. (2004) found 2 Ethiopian Oromo, out of 78 tested, in a survey of >2400 individuals from many places. Cadenas et al. (2007) found one E-M329 in Qatar, out of 72 people tested there.
- E-P2\*. There must once have been more E-P2 clades, neither E-M35 nor E-M2 nor E-M329, and indeed there still are some E-P2 lineages in existence which fall into none of these sub-clades, with traces in Western Africa, and perhaps more importantly, a relatively significant amount in Ethiopia (Semino et al., 2004; Cruciani et al., 2004).

### d. E-P2 as part of E-P177

Karafet et al. (2008) confirmed that E-P75, originally announced in Hammer et al. (2003), is a sibling to E-P2. Both of these are sub-clades of E-P177. But unfortunately no information seems to be available in either paper concerning where E-P75 is from, or how common it is.

### e. E-P177 as part of E-P147

E-P177 is a sibling to E-M33/E-M132, according to the phylogeny in Karafet et al. (2008), Both are within E-P147.

- E-M33, also known as E1a, is West African. It is found throughout the Sahara, as far north as the Maghreb, but especially in the western areas towards the Atlantic. It is virtually absent outside of this area. Semino et al. (2004) found by far the highest concentration in Mali. Cruciani et al. (2002) found highest levels in the Fulbe and at Tali, both in Cameroon. Looking to Europe, Adams et al. (2008) found three individuals out of 1140 people tested in Iberia. All were in Northern Portugal, an area which shows relatively high levels of haplogroups from Africa (Adams et al., 2009; Gonçalves et al., 2005; Flores et al., 2004; see

below). Cadenas et al. (2007) found none in the Arabian Peninsula.

f. E-P147 (E1) as a part of E-P96 (E)

E-P177 and E-M33/M132 are together in the clade E-P147, also known as E1, and their known surviving “cousin”, the sibling of E-P147, is E-M75, also known as E2.

- E2 is sub-Saharan. Luis et al. (2004) shows highest levels in Bantu speakers from Kenya. Semino et al. (2004) found highest levels in Burkina Faso and Bantu speakers in South Africa. Cruciani et al. (2002) found highest levels in Burkina Faso (Rimaibe) and the Daba people of northern Cameroon. Outside of Africa, this clade is very rare. Cadenas et al. (2007) found 3 individuals out of 72 people tested in Qatar. Luis et al. (2004) found 2 people out of 121 tested in Oman.

In summary, when we look at the siblings and “cousins” of E-M35 in its “family tree” we see very little evidence of origins to the north of the Sahara at all. The Horn of Africa seems to have had stronger prehistoric links with the regions west of it, towards the Nile and the southern edge of the Sahara, than to the Levant. It appears possible that it was only after E-M35 came into being that the E clade became involved in migrations both to the north, *and* to the south.

There also of course appears to be a history of constant but small links with the southern extremities of the Arabian peninsula, across the Red Sea to the East from the Horn of Africa, but these do not seem to be as large as connections within Africa, and there is little evidence that these lineages in the southern Arabian Peninsula are in any way parallel with other parts of the Middle East.

Could Semitic have spread across the Red Sea and from the southern Arabian peninsula to the North? Generally speaking, Semitic languages are thought to have achieved their coverage over the Arabian peninsula starting from the Levant in the north, and spreading to the south. Furthermore, most linguists would accept that Semitic is part of a northern group within Afroasiatic, most closely related to Egyptian and Berber. But Lionel Bender (1997), a leading expert on Ethiopian languages, proposed a scenario upon linguistic grounds wherein Semitic languages originated in Ethiopia and crossed the Red Sea. We can note that although this linguistic theory would be in line with these very particular and unsurprising genetic links between the Horn of Africa and the Southern Arabian Peninsula, it does not correspond to much else in genetics or archaeology, and there is no reason to invoke such a theory in order to explain genetic links between the Horn of Africa and nearby Southern Arabia. As far as the present author is aware, there are no obvious signs,

at least not yet, that either material cultures or major genetic lineages moved from the Southern Arabian peninsula to the North in prehistoric times, and also no sign of further movement onwards, for example to Berber speaking lands in Northwestern Africa. The Southern Arabian peninsula appears to be a genetic outlier, combining lineages from both the Middle East and Africa. Future archaeological research in this region may change our perspective.

Concerning the first origins of E-M35, our examination of the phylogeny of the E haplogroup draws our attention towards the inhabited band along the southern edge of the Sahara desert. There, as shown in Figure 6,



Figure 6. The Sahel. The map is from the English Wikipedia.

we find the so called Sahel, and a Savannah “Parkland” which stretches parallel to and just to the south of the Sahel. These long regions stretch from the Atlantic to the Horn of Africa. Clearly E-M35’s family centers around this band, with E-M35 representing the Eastern or Nile end.

At this point we can refer back to linguistics, and strikingly we find the same pattern. The map of African languages reproduced at the introduction of the present article shows the African continent dominated by three language groups. Apart from Afroasiatic, the other two groups (Nilo-Saharan and Niger-Congo) can be stated



without any controversy to be completely African, and to have dispersed from near the same large band along the southern Sahara.

While these three distinct language groups, apparently all from regions in or near the Sahel, do not have the same Y haplogroup populations, all have high concentrations of E haplotypes. Genetics is able to prove conclusively that they are related to each other and to E-M35, as demonstrated above – showing us links further back in time than the study of languages can.

So there are clear signs in two fields that the ancestors of the population living today in the area to the south of the Sahara, together with the Horn of Africa, formed an ancient core from which cultures, languages and people themselves dispersed further. Despite the enormous size of the area, because of the climate there, migrations and long distance interactions appear to be common. The Fula (Fulbe, Fulani) people for example, are found from the Atlantic all the way to Darfur in Sudan. Indeed it appears likely that the centre of gravity of these peoples has moved around over the millennia, with changing climactic conditions.

Other types of evidence are generally felt to confirm that this region was a major source of human diversity. In the wet early Holocene period an “aquatic tradition” developed in the eastern Sahel and south-eastern Sahara, centred around the lakes, rivers and wetlands of that period. This culture developed intensive resource use and sedentary habits, including the storage of wild grass seeds and the use of pottery. Christopher Ehret proposes indeed that wild grass collection had been introduced by proto-Afroasiatic speakers within this complex, who he believes had already developed this type of resource use to the East of the Nile. Whether this be true or not it seems very likely indeed that this culture near modern Sudan played an important role in E-M35 pre-history. The peoples of this successful culture seem to have ended this way of life and dispersed into neighbouring areas due to the change in climactic conditions (Ehret, 2002a, 2002b).

While it seems very clear that genetics does show that a major and ancient movement of people took place from the Horn of Africa (or in any case from somewhere to the south of the Sahara) to the Semitic speaking areas, it also supplies evidence of at least some migration in the other direction, in the form of lineages within the largely Eurasian haplogroup F (which contains J, R, T etc). However, these haplogroups, which are not the subject of the present article, are far less common in Africa than other clades, and are normally interpreted as a result of more recent (post-Neolithic) “back migration” from Eurasia to Africa – for example during the expansions of

the Arabic speaking, Muslim religion and empire (Luis et al., 2004).<sup>8</sup>

#### 4. Branches of Afroasiatic and E-M35

##### 4.a. Beyond the “Bi-Directional Corridor” Core Regions

We have so far focussed primarily upon the very oldest origins of Haplogroup E-M35, the language family Afroasiatic, and how they might relate to the pre-Neolithic cultures from the Levant and Nile. Now we need to turn to the dispersals of the Neolithic itself, and thereafter. Many of the later regional variations of the languages, haplogroups and archaeological material cultures which this subject potentially involves deserve detailed discussion in their own right. However for the context of this particular article, short summaries will be given which will help demonstrate the potential for further more detailed multidisciplinary work.

We have shown that by the time of the Natufian, E-M35 had dispersed at least as far as Egypt, which means it was at least in contact with the Natufian culture across the Sinai, possibly along the Mediterranean coast. It even seems very likely that both E-M35 and Afroasiatic languages were already in the Levant, and had come from Africa earlier.

There is little doubt about some aspects of what happened next. With the development of farming technologies both Afroasiatic languages and E-M35 spread together from this Nile-delta/Levant “hub.” In one direction, the Middle East, they developed as one part of greater Fertile Crescent complex of cultures where Semitic languages only eventually attained their later prominence (Zarins 1990). In the other direction lay Africa, the homeland of E-M35, and probably also of Afroasiatic. Here too, a complex of successful cultures was formed, merging with local cultures, some of which appear to have come from the Sahara. Peter Bellwood (2005, p.101) writes:

*When agriculturalists first entered the Nile Valley at about 5500 BC, other people who made pottery, collected wild sorghum, and, according to some authorities, herded cattle, had already been living at Nabta Playa and Bir Kiseiba in Egypt's Western Desert for possibly 3,000 years, during periodic and brief phases of wetter climate. There is no clear evidence that such populations were living in the Nile Valley itself at this time, but the possibility must be considered. Mid-Holocene desiccation of the Sahara with retraction of the summer monsoon finally led to virtual abandonment of Nabta Playa and other oasis settlements by about 4000BC, and this might have led to a kind of refuge movement*

<sup>8</sup> Some aspects of the apparent Y-DNA evidence for “back migration” from Eurasia to Africa are discussed below concerning Chadic languages.

*into the Nile Valley, where Saharan people would have met and mixed with the descendants of the Southwest Asian Neolithic population responsible for the introduction of the Southwest Asian agricultural tradition into the Nile Valley about 1,500 years earlier. The resulting amalgam was later to develop into one of the most remarkable civilizations of the ancient world, a true synthesis of Oriental and African.*

In the Middle East E-M35 lineages are common and diverse, but not dominant in many areas. Middle Eastern populations, like those in Egypt, have a complex mixture of Y lineages including many which are thought to be Eurasian in origin, such as the J, G, T and R haplogroup lineages. Only two E-M35 clades are identified in the literature as probably having Levantine origins, indeed the only two which are stated in published literature not to have originated in Africa - E-M123, and E-V13. In parallel, only one family of Afroasiatic languages is non-African, Semitic. Although Semitic is very dominant in this region today, this was not always the case.

It is striking that E-M123 and E-V13, though extrapolated to have Levantine origins, are very widely dispersed. This is however somewhat typical of Middle Eastern Y haplogroups, and that is interesting in itself. It appears that some Fertile Crescent farmers developed colonizing cultures which spread rapidly north, through Anatolia into Europe, and, from the Levant or Anatolia, it seems, west into the Mediterranean basin, colonizing Cyprus very early—which provided “a clear and valuable template for the subsequent diffusion of the Neolithic across the rest of the Mediterranean Basin” at least according to one author (Zeder, 2008, p.11600).

Can we say that in the Fertile Crescent itself there were “waves of dispersal” whereby either languages or Y-DNA were strongly linked to the spread of Neolithic technology? Because Neolithic technology developed in this area in a complex way, the metaphor of a wave seems inappropriate. E-M35 and Afroasiatic appear to have been only parts of a bigger diversity within this region during the Neolithic, both apparently reflecting a contribution from the south-eastern corner of the crescent, with links to Africa. The modern Y DNA shows the diversity of the Middle East more clearly than the modern linguistic situation, which has become dominated by Arabic. The historical dominance of Semitic languages in the Fertile Crescent appears to have only developed after the Neolithic had spread throughout the area, carried by pastoralists who were neighbours to the first farmers (Zarins 1990).

Therefore if there was any radiating wave of Neolithic language and/or genotype dispersal it must be in the areas outside the Fertile Cresecent.

## E-M123 and the Mediterranean Basin

E-M123 is not as common as E-M78 and E-M81, its major sibling clades within E-M35. It is also more difficult to associate with any particular region, being found in relatively high frequencies in such scattered places as Oman, North-western Iberia, Turkey (Cinnioğlu et al., 2004), Tizi-Ouzou in Berber-speaking coastal Algeria (Arredi et al., 2004), and Semitic-speaking Ethiopia. Nevertheless a strong impression of a Middle Eastern origin has been building up.

It was noted by Cruciani et al. (2004), Semino et al. (2004), and Luis et al. (2004) that E-M123 (mostly equivalent to its dominant sub-clade E-M34) appears to be a branch of E-M35 which split off early from the older population through the “Levantine corridor” (Luis et al.). In frequency terms, relatively strong branches appear in Oman and Ethiopia, but these appear to be younger branches from the Levantine and Egyptian populations. Coffman-Levy (2005) noted the presence of the clade in both the Ashkenazi and Sephardic Jewish data for example in Semino et al. (2004), indicating that it might be an ancient Levantine lineage. More recently came the more striking information from Flores et al. (2005) that 14 out of 45 men tested in the Dead Sea area of Jordan are M34 positive (31.1%). Flores et al. specifically chose this area for sampling because they felt the population there was relatively genetically isolated, which can help show ancient genetic influences.

It therefore appears as if E-M123 represents a clade (or clades, for this is an ancient lineage with its own identified sub-clades) which dispersed within the Fertile Crescent pre-historically. It's dominant E-M34 sub-clade is strongly associated with populations with Semitic languages.

Far beyond the Fertile Crescent, E-M123 also appears to have a small but significant and ancient Mediterranean dispersal. For example in Northwestern Iberia (Adams et al., 2009) and in some areas of Sicily (Di Gaetano et al., 2008) as well as in the Albanian speaking community of Cosenza Province in Calabria (Semino et al., 2004). According to the data collected in Adams et al. (2009) the highest M34 levels in Iberia might be on the islands of Minorca and Ibiza.

It is worth remarking, given that we are examining the usefulness of genetic data, that the Adams et al. paper was widely reported in the popular press, based upon its major theme and its title, which involved the following claims (p.732):

*Our admixture approach has identified high mean levels of North African and Sephardic Jewish patrilineal ancestry in modern populations of the Iberian Peninsula and Balearic Islands. We find a*

*mean of 10.6% North African ancestry, somewhat higher than previous ad hoc estimates, and a mean of 19.8% Sephardic Jewish ancestry, a figure that cannot be readily compared with any other study. These findings attest to a high level of religious conversion (whether voluntary or enforced) driven by historical episodes of religious intolerance, which ultimately led to the integration of descendants.*

While the headlines which appear after major DNA surveys constantly place emphasis upon historically famous populations such as Greeks, Romans, Phoenicians, Basques, Celts and Moors, the potential for seeing much older migrations in the data is arguably far more powerful, because after the Neolithic, the population increased relatively dramatically, meaning even large movements of people since that time have had less effect. This style of reporting is popular and pleasant, and geneticists writing in peer-reviewed articles are careful to couch their conclusions in cautious language. But one negative result of the current situation, for example in this case, is that much more interesting results have been *de-emphasized*, possibly discouraging readers from outside the field who might have had something to add. To quote Adams et al. further (p.733):

*An additional factor that could lead to overestimation of Sephardic Jewish ancestry proportions is the effect of other influences on the Iberian Peninsula from eastern Mediterranean populations that might have imported lineages such as G, K\*, and J.*

Note that the authors do not mention E-M123 in the above passage, which is less discussed than G, K (or T) and J in the literature as a Middle Eastern marker, but they did find many Middle Eastern looking lineages. It is important to recall that in this present article we are largely limiting analysis to E-M35, and that other clades can and will substantially add to the picture only being touched upon here. The key observation however, is that north-western Iberia shows a surprisingly high level of Middle Eastern Y DNA, which confuses the picture presented in Adams et al. (2008), and indeed brings it into question. The authors note this but propose to separate the two groups of non-Iberian Y lineages, North African lineages being from historical times and the Middle Eastern ones being much older. However, this should raise the question of how they judged this without making an *ad hoc* decision. As the authors note themselves:

*...the highest mainland proportions of North African ancestry (>20%) are found in Galicia and Northwest Castile, with much lower proportions in Andalusia. The most striking division in North African ancestry proportions is between the western*

*half of the peninsula, where the proportion is relatively high, to the eastern half, where it is relatively low.*

In other words, the Maghrebin lineages are in the wrong part of Iberia to be explained by the Islamic or Phoenician periods. Something much more interesting seems to underlie this pattern.

As it turns out, further examination of the data in this and other studies shows that a diverse range of surprising pockets of Middle Eastern and North African lineages are found in the remote bays and valleys of the East and North of Iberia. E-M123, as it happens, plays a very useful but largely unrecognized role in this pattern. It is useful to summarize some of the most unusual haplotypes of this type, rarely found anywhere in Europe, or in some cases, anywhere at all.

- In a 568-person study in Iberia, Flores et al. (2004) found about 10% of Galicians were E-M34+ (as are most people who are M123+). Perhaps more strikingly, they also found two very rare cases of E-M123\* individuals. Both were in Northern Portugal, out of 109 people tested there. E-M123\* is rare in all places surveyed so far. It might be most common in Northern Portugal. Isolated individuals have been found in Tunis (Arredi et al., 2004), Jordan (Flores et al., 2005), Central Asia (Underhill et al., 2000), and Bulgaria (Cruciani et al., 2004).
- In a 553-person study of Portugal, Gonçalves et al. (2005) (another article which has a title mentioning Berber and Sephardim ancestry, this time focussed on Portugal) found E-M34 mainly in Central Portugal (4 people out of 102 tested there) with one more person found in the Açores. They also found two more rare cases of E-M123\* individuals in Northern Portugal, out of 101 people, as well as 2 in Madeira out of 129 people tested there. Galicia, on the other hand, shows up as an E-M34 enclave, with 2 out of 19 people tested (10.5%).
- In a 292-person study of Galicia, Brion et al. (2004) found 4.11% E-M81 (12 people), 1 person who was E-M123\* and about 1% (3 people) who were E-M34.
- Adams et al (2009) found 9% E-M81 in Galicia (88 people), much higher than the 2% in Eastern Andalusia, for example (95 people). But they also found 10% of E-M35\* in the Asturias (out of 20 people). Such a level of E-M35\* is rare outside of the Horn of Africa (see Cruciani et al., 2004, in conjunction with Henn et al., 2008 concerning the Southern African population, which will be discussed further below).

- In Cruciani et al (2007) 4.44% out of 90 Asturians tested were from the haplogroup E-V22 and 6.25% of 16 French Basques were in E-V12\*. These two sub-clades of E-M78 appear to come from Egypt, and are very rare in Europe outside of Southern Italy.
- Perhaps most well known of all these surprising enclaves, Cruciani et al. (2004) found that no less than about 40% of 56 male lines tested in the Pas valley in mountainous Cantabria were M81+.
- Capelli et al. (2009) confirms Cantabria as an Iberian “hot spot” not only for E-M81, but also E-V65 (equivalent to the “beta cluster” of E-M78 as per Cruciani et al, 2006) and J1, and once again confirms far lower levels of these North African haplotypes in the East of Iberia (Basques, Catalans). No E-V65 or J1-M267 was found in their Andalusian sample.

Looking to other Y haplogroups with similar patterns, as Gonçalves et al. (2005) remark:

*More than 10% of Portuguese Y chromosomes could be classified in haplogroup J. The frequency of this haplogroup in Portugal is significantly higher than among the Spanish ( $p < 0.001$ ) or other West European populations, with the exception of Italy.*

Concerning J lineages, another marker of Eastern Mediterranean origins, we can add from looking at Adams et al. (2008) that they are also relatively frequent compared to the rest of Iberia in the Asturias and Extramadura. And while E and J haplotypes in Europe are often discussed together (e.g. Semino et al., 2004, Cruciani et al., 2007) the pattern of dispersal in Iberia is matched by other Middle Eastern haplogroups, for example G-M201 and K-M9(xM45) (Adams et al., 2008).

This situation, whereby headlines concerning tenuous attempts to reconstruct relatively recent and better-known movements of peoples, using very ancient UEPs, at the expense of studying the Neolithic, is common. A very similar effect happens in the literature concerning Southern Italy. Another recent paper, by Di Gaetano et al. (2008) concerning Sicily, focuses mainly on an attempt to find the traces of the Greek and Phoenician colonies once there. Once again, arguably the most interesting conclusions are not emphasized, but they are mentioned. An significant difference in haplogroups was found distinguishing the west and east of Sicily, but in the “wrong” direction: typical European haplogroups R1b-M269 and I1-M253 are far more common in Western Sicily, the area of highest Tunisian-related (including Phoenician) influence in recent centuries, while the Maghrebin lineages E-V65 and E-M81 are, *once again together with Levantine/Egyptian lineages,*

E-M123, E-V22, E-V12, G, J2-M172, J2-M67, K2-M70, and are found in the East, where many Greek colonies once existed.<sup>9</sup> Once again this pattern does not seem to match the historically famous records from the most recent millennia. Francalacci et al. (2003) confirms at least the Middle Eastern haplotypes being a presence in not only Sicily but also Sardinia and Corsica.

It is hard to avoid concluding when looking at such data, that north-western Iberia and parts of Sicily took part in the same ancient movement of peoples which somehow combined Y lineages that are today associated separately with the Maghrebin Northern Africa, and the Middle East. Given the position of both places, it seems we must be looking for a movement of people around coasts, but was there such a movement that was important enough to leave patterns so significant that they seem to dominate the more recent effects we would have expected from Phoenicians, Greeks, Romans, Berbers and Arabs?

### The Cardial Culture

There was such a culture in the Neolithic, which probably had a very considerable impact upon the population of Southern Europe. This is the so called Cardial culture, named after its frequent use of Cardium cockle shells to imprint their pottery.<sup>10</sup> Zeder’s (2008) summary has their antecedents probably originating “somewhere in the Northern Levant” (approximately modern Lebanon and coastal Syria) and settling Cyprus about 10,500-9,000 years ago, finding their way to their apparent place of origin on the Greek side of the southern Adriatic about 9,000 to 8,000 years ago, entering the “boot heel” of southeastern Italy about 8,000 years ago. In fact, there is no apparent consensus yet concerning the origins of this pottery culture before it appears in the southern Adriatic, and Zeder’s remark represents one of the few speculations on the matter which the present author could find.

Pottery was one of the last elements of the Neolithic package to appear in the Fertile Crescent, apparently beginning about 8000 years ago with the Yarmukian culture in the southern Levant, which was a culture that appears to have arrived in the area at a time of cultural disruption in the area. The culture may have been

<sup>9</sup> It is also worth mentioning that the decision about dividing up samples between “East” and “West” must have had an effect on results. Unlike the above discussion concerning Iberia, the data from Sicily shows no strong East-West pattern, or other simple geographical pattern, and the authors confirm “the general heterogeneous composition of [haplogroups] in our Sicilian data” (DiGaetano, 2009).

<sup>10</sup> The earliest known manifestation of the Cardial culture in Italy and the Adriatic is sometimes referred to as the “Impressa Culture.” The term “Impressed Ware” is a more general term, not always referring to these specific cultures, but it is used by some authors to refer to Cardial Cultures (Binder 2000, Barnett 2000). Battaglia et al (2009) have recently suggested that E-V13 lineages came to Southern Italy from the Balkans along with this early “Neolithic Impressed Ware.” We shall discuss this further below.

brought by migrants. Pottery was significantly older in Africa. Forms of impressed pottery were in the Sahara in the early Holocene although these earliest technologies are not known from the Nile itself, as opposed to the Egyptian Western Desert (Phillipson, 2005, p.156), and it is not clear if they ever reached any of the cultures of the African Mediterranean directly. Instead it appears that pottery may have entered many parts of the Maghreb at roughly the same time as Middle Eastern types of goat or sheep, (ovicaprids), after the Neolithic had arisen in the Middle East. But there are complications that must be considered before it can be confidently proposed that Maghreb pottery had a Levantine in origin. The exact origin of pottery in both these areas is still too unclear. And in some areas of the Maghreb such as the Tangiers region, Cardial pottery, generally associated with Europe, is among the oldest styles to be found in the archaeological sequence (Phillipson, 2005, p.156).

Barich et al. (2006, p.579) remark that the Saharan culture and the Mediterranean Iberomaurusian culture of 10,000 years ago were “entirely related” – “except for the pottery.” These authors state that after 10,000kya, in the early Holocene, pottery arrived in the Iberomaurusian areas, but that there is “currently very little information” about this. Similarly, the origins of Middle Eastern pottery before the Yarmukian, for example whether it has any link to older Saharan pottery, are apparently not a subject about which much can be said yet. This ignorance of the origins and exact links of early Mediterranean pottery cultures impacts directly upon any attempts to explain the origins of the Cardial culture within the greater scheme of Mediterranean pottery.

This Cardial culture was in the “boot” of Italy early, and from here it appears to have leapfrogged up the western coast of Italy to southern France by at least about 7,700 to 7,600 years ago. The exact routes are not yet known, and what happened on the southern side of the Mediterranean in Africa has been a gap in the story.

It is striking, but not often remarked, that pottery first appears at a similar time in Italy, Greece, and the Levant.

Looking to the archaeological literature, Zilhão (2000, 2001) describes the Cardial culture of Northern Portugal as an early Neolithic “enclave” in Western Iberia, defying any simple model of “demic diffusion” from East to West. He proposes “leapfrogging colonization by small seafaring groups of agriculturalists” (Zilhão, 2000). This matches closely what we see with the Middle Eastern DNA in Iberia, including the relatively long distance from any obviously related European enclave of Cardial technology. But can this explain the North African lineages in Iberia being found in similar northern enclaves, it seems, with the Middle Eastern genotypes? For example, could the

coast hoppers who got as far as the Atlantic have taken on an African component in their population before passing the Pillars of Hercules? Indeed, does this explain the large gap between Portugal and Southern France that these colonists seem to have left to other cultures, including other streams of the Cardial culture?

From the archaeological side, this is indeed starting to appear likely. Manen et al. (2007) show that the Zilhão’s maritime “leapfrog” understanding of the origins of the Portuguese Cardial, wherein a rapid movement is understood to have occurred from French Provence to the Atlantic coast of Iberia, leaving very little convincing evidence in between, may owe something to the relative poverty of data that has been available for the Maghreb coastal cultures of this period. The authors showed that the increasing evidence now becoming available indicates that the Portuguese Cardial may have Moroccan antecedents.

### **The Maghreb and Libya: Berber Languages; E-M81 and E-V65**

The pre-Islamic distribution of Berber languages corresponded to populations where E-M81 and E-V65 are dominant, and Berber languages are often seen by linguists as a straightforward offshoot of the same northern branch of Afroasiatic which gave Egyptian and Semitic. According to a very straightforward “wave of advance” or “demic diffusion” hypothesis therefore, the population, and therefore these E-M35 clades, would be expected to come from due east—from the direction of the Nile and Levant. A similar simple hypothesis could be proposed for Berber languages, and also farming technology and the herding of domestic ovicaprids. But this does not mean that all these ancient movements happened at the same time, or followed the same routes. The idea of one simple migration bringing languages, genes and technologies is increasingly difficult to sustain as more complex archaeological data for this region is published.

The most notable patterns concerning E-M35 in the Maghreb can be defined as follows:

Firstly, both haplogroups, especially the more common E-M81, are found elsewhere in small frequencies, into Africa, the Middle East, and Southern Europe, perhaps most surprisingly in Turkish Cypriots (8.7% in Cruciani et al., 2007). But the mixture of E-M35 clades found in the Maghreb is nevertheless best described as strikingly different from those found from the Levant to the Horn of Africa.

Secondly, while there are unsurprisingly significant presences of Middle Eastern and possibly European genetic lineages in the Maghreb (for example Y chromosomal J haplogroups, see Semino et al., 2004),

E-M81 and E-V65 are very dominant and apparently relatively young.

In the literature, ages are currently most often estimated by using STR diversity in order to calculate when the most recent common ancestor lived. For the reasons discussed above, the accuracy of this type of estimate is known to be highly sensitive to pre-historic population fluctuations, but it can be used to help give an idea of how recently a population started expanding.

The trend in the literature so far has been towards younger age estimations for E-M81. Bosch et al. (2001) estimated 15-43 kya for E-M81, but later Luis et al. (2004) estimated 2 kya, while noting that the small amount of Egyptian E-M81 seemed older. Semino et al. (2004) estimated 8.6 kya. Arredi et al. (2004) estimated 4.2 kya. Cruciani et al. (2004) estimated 5.6 kya, but more recently in Cruciani et al. (2007), the authors show concerns at this calculation technique, implying that estimates be revised downwards to less than 5.6 kya. For E-V65 Cruciani et al. (2007) estimated approximately 4 kya.

The male-line ancestors of modern E-M81 and E-V65 men did not always live in the Maghreb. The genetics literature indicates that these lines can be traced back over millennia to E-M35's point of departure near the Horn of Africa. How long this took and which route was taken, is very difficult to determine. For E-M81, as Arredi et al. (2004) mention, the Middle East may even have been on this route. For E-V65 we can at least suggest that Egypt lies on the route, because it is in a sub-clade of E-M78.

E-M81 and E-V65 *may* have quite different pre-histories. Their distribution does not appear to be identical (see the distribution data in Cruciani et al., 2004, 2007, and also Arredi et al., 2004), with E-V65 more common amongst Arab speaking populations, including those to the East of the Maghreb region in Libya, and E-M81 more common as one approaches the Atlantic Ocean.

With the archaeological evidence we have so far discussed, we could expect the presence of E-M35 lineages in the Maghreb to have followed one or more of several likely routes:

- The Sahara itself, lying to the south of modern Maghreb populations, had a higher population during greener periods. These populations may well have contained E-M35 lineages by the start of the Holocene. Part of this population apparently moved north into the higher rainfall areas nearer to the Mediterranean (Barker, 2002).
- The Mediterranean coast, from the Nile/Levant area, via Libya, where as Barker (2002) notes, there was a similar culture, the Iberomaurusian, to the ones found in the Levant before the full Neolithic. Barich et al. (2006, p.579) describe this culture as connected to the Saharan cultures at least until the Holocene.
- The Mediterranean sea, brought by the coast-hopping Cardial Neolithic cultures, presumably via Sicily, or other Mediterranean islands. That Cyprus has a significant presence of E-M81 may yet turn out to be significant. In any case such leads deserve further checking in the near future.

Archaeology gives us reason to doubt the simple scenarios which the linguistic or genetic evidence might imply. While the Maghreb clearly does show links with the Nile, the Sahara and the Mediterranean, there are good reasons to doubt that there was ever any simple replacement of a less advanced Iberomaurusian, by a more advanced Capsian.

This means that it is not as simple as might once have been thought to link either the Maghrebin E-M35 lineages, or the Afroasiatic Berber languages, to any obvious overland migrants. Moreover, it is also becoming increasingly clear that the Cardial culture associated with the Southern European Neolithic dispersal, played a direct role in bringing Neolithic technologies to the Maghreb via coastal routes (Manen et al., 2007; Linstädter, 2008; Daugas and El Idrissi, 2008; see above concerning Iberia). This increases the range of possibilities and complicates things significantly.

Adding to the complexity, despite the domination of E-M35 male (Y DNA) lines in this region of Africa, there is a strong presence of mitochondrial (female line) haplogroups there which appear to come from Europe, and specifically Iberia, e.g., H1, H3 (see Cherni et al., 2008; Ennaffaa et al., 2009). This may represent European links of the pre-Neolithic Iberomaurusian culture. This was the same coastal culture whose descendants eventually appear to have mixed with and absorbed elements of the more advanced Cardial culture. It might appear unlikely that a modern population might have its dominant mitochondrial and Y DNA lineages from different original populations, but imbalanced marriage of a kind which might lead to this has been observed in modern cases of farmer-forager relations, due to the cultural and economic dominance of the farmers (Zvelebil and Little, 2000). So this very imbalance could, in fact, be a sign that E-M35 did enter the area with farming, pastoralism, or some other dominant technology.

From an archaeological point of view, Graeme Barker (2002) proposes that Afroasiatic languages in the

Maghreb need not have arrived with Neolithic technologies from the Levant. He states that the “rapid colonization of the Sahara at the beginning of the Holocene was by foragers, not farmers” (p.157). That these foragers may have spoken an early dialect of Afroasiatic, as the Neolithic farmers from the Levant likely did, can be explained by noting “the similarities in the respective archaeological records of the Natufian culture of the Levant and of contemporary foragers in coastal North Africa” (p.158).

Bender (1997) on the other hand writes that despite assertions “that Berber is a diverse language family, all that I have seen indicates that it is a language cluster, no more internally diverse than (e.g.) Romance.” He would not suggest a split-off from the other Afroasiatic languages earlier than about 8-7,000 years before present. As mentioned above, the most recent estimates of the age of E-M81 and E-V65 seem to agree with such age, at least when considered in terms of its own intra-clade diversity.

Blench (2001) makes a comment concerning Berber languages which parallels the Y DNA evidence strikingly well:

*The problem is as follows:*

- a. *Berber shows surprisingly little internal differentiation, as if it represented a recent expansion.*
- b. *Yet it is very different from its neighbours in Afroasiatic as if it split away a long time ago.*

Blench considers it likely that Berber existed for a very long time in an “equilibrium” of loaning between closely related and highly mobile languages, such as was found amongst the languages of Australian Aborigines. It appears to be probable that the population of Y chromosomes in the Maghreb population shows a low diversity caused by a “founder effect.” When this growth started is not clear yet, but it appears reasonable to suggest that Neolithic technology, and the entry of Afroasiatic languages may have arrived at about the same time. The dominance of E-M81 may have built up in the wake of the ensuing population growth.

#### The Middle Eastern route to the Balkans: E-V13

In Europe, no Afroasiatic languages are considered to be native. Putting aside language correspondences however, in Europe E-M35 and early Neolithic technology do appear to show strong geographical correspondences. Neolithic technology entered the European mainland earliest via the Balkans, and from there it is believed to have travelled westwards into Southern Italy (eventually developing into the coast hopping Cardial culture, discussed above) and northwards to the Danube (eventually developing into the Linearbandkeramik or

“LBK” Culture).<sup>11</sup> But it is the pattern of Neolithic technology’s earliest entry into the Europe, rather than its later dispersals, that has been compared to the modern distribution of E-V13, a sub-clade of E-M78.

E-V13 is a case of an E-M35 lineage with no obvious link to Afroasiatic languages. It is one of the easiest European Y lineages to assign an ancient migration route. E-V13 is found almost entirely in Europe, and within Europe mostly in the Balkans and Italy. Putting the question of timing aside at first, we can say that it is clear that E-V13 almost certainly dispersed to the rest of Europe from somewhere in the Balkans.

What’s more, going further back in time it also seems clear that there are remnants of a less common but more diverse E-V13 population in the Middle East, most notably so far amongst a small set of Druze Arabs reported in Cruciani et al. (2004, 2007) whose exact STR haplotype looked less like European E-V13, and more like other E-M78 sub-clades, E-V22 and E-V12, which Cruciani et al. (2004, 2006, 2007) associated with the area around Egypt. Battaglia et al. (2008) also found that Konya in present day southern Turkey had at least a slightly higher diversity of E-V13 haplotypes than any place they studied in the Balkans.

In summary, E-V13 in Europe dispersed from a Balkan family with deep E-M78 ancestry in the Middle East, and very deep E-M78 ancestry in the area of Egypt. In any of these movements over time, was E-V13 associated with the dispersal of new food production technologies? Cruciani et al. (2004, 2007) have pointed to at least four possible periods suggested by archaeology when major change in population in the Balkans could be envisioned:

- The “post-Last Glacial Maximum expansion (about 20 kya)”
- The “Younger Dryas-Holocene reexpansion (about 12 kya)”<sup>12</sup>
- The “population growth associated with the introduction of agricultural practices (about 8 kya)”
- The “development of Bronze technology (about 5 kya)”

Cruciani et al. (2007) propose the last and most recent of these, the Bronze age, as the period when the modern E-V13 population began *dispersing* to the rest of Europe

<sup>11</sup> See, for example, Price ed. (2000), Bellwood (2005), Zeder (2008).

<sup>12</sup> According to Runnels (2003), by the time Neolithic settlers were colonizing Greece from Asia, there was almost no population living in the area. So any “Younger Dryas-Holocene” entry of E-V13 into the Balkans would have presumably experienced a subsequent population bottleneck before somehow successfully integrating into the successful Neolithic communities who arrived much later..

out of the Balkans. The same paper agrees with earlier papers (e.g. Cruciani et al., 2004; Semino et al., 2004), in suggesting that the clade *entered* Europe with Neolithic technology, the next-youngest of the above four likely periods. Their age estimates of E-V13 as a whole suggest that it *originated* in the early Holocene in an E-M78 population of the Middle East.

Battaglia et al. (2008) propose timing which starts in the same early Holocene period in Egypt, but then moves more rapidly, with E-V13 probably originating in an E-M78 population that reached the Balkans already before the Neolithic arrived in these areas. They reason that the SNP mutation V13 itself may have even happened this far north, pointing to the low frequency of E-V13 in modern central Anatolia and Crete, which they interpret as the areas from which Neolithic technology probably came towards the Balkans.<sup>13</sup> These authors propose that E-V13 later probably dispersed to Southern Italy with the Neolithic Cardial culture. In these two published scenarios, the starting point in Egypt in the early Holocene, and the finishing point of the Neolithic in the Balkans are the same. However, in the period in between Battaglia et al. visualize E-V13 keeping just ahead of the Neolithic wave moving north. Most other authors apparently see it as having been in the wave itself.

Looking at archaeological evidence however, the whole area between Egypt and the Balkans in the early Holocene was populated by the cultures that were developing new food producing technologies. It is not easy to imagine a population integrating into this area from Africa, then leaving in the direction of the Balkans without any of these new technologies. Along the migration path common to both these scenarios is the Levant, which immediately before the Neolithic was the home of the Natufian culture. We have seen that this culture very likely had both E-M35 Y lineages, as well as an Afroasiatic language. Also this population is likely

<sup>13</sup> Some aspects of the Battaglia et al. (2008) article require extra discussion, because it is in some conflict with the explanation given in the present article. (1) According to this argument which draws very strong conclusions about ancient populations from a modern lack of evidence (low E-V13 in Central Anatolia and Crete), it seems we should expect the Middle East and Anatolia to have high levels of E-M78\*. Indeed the authors keenly note the two examples they find in the Balkans. But on the whole it must be said that this argument seems fatally flawed because E-M78\* is not common in the Middle East (or anywhere), and we simply can not expect all ancient clades to still be common where they once were. (2) Battaglia et al. also appear to ignore microsatellite variation data, both their own Konya data, and that of Cruciani et al. (2007), which do show higher implied ages for E-V13 in Asia than Europe. (3) Finally their argument depends to some extent on the assumption that the Neolithic in the Balkans must have arrived from central Anatolia or Crete. However, from the earliest phases the early Balkan Neolithic clearly had “island hopping” capability, and trade networks which reached into Anatolia, so there is no consensus on the route it took. It is in fact quite interesting that not only E-M78 lacks any obvious overland or island route between the Levant and the Balkans. The same geographical gap in the evidence is seen in the archaeological record of Neolithic dispersion (Perlès 2001, Ch. 4).

to be one which is ancestral to the later Neolithic populations of the Fertile Crescent.

Could there be a linguistic aspect to this movement of people from the Middle East to the Balkans? If we take the Neolithic hypothesis, this would correspond to a period when a minority position would hold that Indo-European languages arrived in the Balkans from Asia. As in the case of Afroasiatic languages there are archaeologists such as Colin Renfrew and Peter Bellwood who consistently propose a “farming/language dispersal hypothesis” with regards to major language families. These authors would propose that Indo-European probably arrived in Europe from Anatolia, as part of the technological revolution in food production.

More commonly however, linguistic research, such as once again examining flora and fauna vocabulary, leads to the proposal that the Indo-European languages originated in Europe, and not Asia. Indeed it is widely believed that the speakers of proto-Indo-European descended from a pastoralist steppe people who came into contact with their Neolithic neighbours after they were already established in Europe. Above, it has already been discussed how a similar scenario has been proposed for the spread of Semitic in the Fertile Crescent, as per Zarins (1990). In both cases it seems possible that people with a pastoralist economy eventually came to dominate sedentary farmers, linguistically at least, following a pattern repeated several times throughout history (Mallory 1989). Again, we should keep in mind that while Y lineages will be affected by large population changes such as in the Neolithic, even large populations can change languages relatively easily without dramatic immigration.<sup>14</sup>

Theories about the early geographical origins of Indo-European languages and E-M35 do not match. The earliest apparent geographical dispersals of Indo-European languages and E-M35 thus overlap strongly only in the Balkans, where the Indo-Europeans apparently spread their language to the most technologically advanced part of Europe at that time, but where, at least in Greek (the other old Balkan languages are poorly known), a large percentage of vocabulary seems to be non-Indo-European (Mallory 1989, p.67).

Could Afroasiatic languages have been spoken in the Balkans before Indo-European languages arrived? There seems to be no evidence to support this. Indeed as discussed above, the Afroasiatic languages are at best likely to have been only one of the language groups spoken by populations involved in the Middle Eastern Neolithic. They never seem to have had a major impact in most of Anatolia, let alone Europe, and even in the

<sup>14</sup> And so we may perhaps note that a “farming/Y lineage dispersal hypothesis” might be stronger than a “farming/language dispersal hypothesis”!



Fertile Crescent, Afroasiatic languages were possibly only spread after the Neolithic was established, due to the success of pastoralists living near to farming communities. Indo-European languages may actually have built up their importance in a very similar way, not by being in a Neolithic language dispersal wave, but by being associated with pastoralists in the right place at the right time to have a strong influence on farming neighbours.

In contrast to the case for languages dispersing with the first Neolithic technology, the case for E-M35 having been carried with the first wave of the Neolithic seems strong in Europe.

### Southern Afroasiatic and Southern E-M35.

As has been mentioned, the highest diversity in Afroasiatic language, as well as in E-M35 haplogroup lineages, is found near the Horn of Africa. Below we give short reviews of the probable correspondences, the languages, and the Y lineages, given the state of knowledge at this time.

#### Beja and Cushitic

Cushitic takes its name from the ancient kingdom of "Kush," to the south of Egypt, also known from the Bible. It is a complex family<sup>15</sup> within the greater group of Afroasiatic languages in the Horn of Africa, but as the name tells us, it is sometimes thought to have come from closer to Egypt. Modern researchers also consider this likely, for example Ehret (2002) and Diakonoff (1998).

Beja is either a sibling to the other Cushitic languages, or else its own clade within the Afroasiatic family, but it is generally thought probable that it has been in the same place for a long time—between the Nile and the Red Sea stretching through Sudan into both southern Egypt and Eritrea. An ancient version of this language is thought to have been spoken by the Blemmyes who were involved in occasional conflict in and around the Red Sea side of the Meroitic or Kush kingdom. Unfortunately there does not yet appear to be consensus about the main language used in that kingdom itself, but another ethnic group in and around this empire were the Nubians, who appear to have spoken a non-Afroasiatic Nilo-Saharan language. Ehret (2002b) proposes on the basis of loanwords that Beja and Nilo-Saharan were neighbours on either side of the Nile for a long time. He furthermore proposes that the Cushitic languages moved from a Red Sea homeland into the Ethiopian highlands, whence they dispersed further.

E-M35 appears to show the implied southwards movements well, in the form of E-V12, a sub-clade of E-M78. Cruciani et al. (2007) found highest E-V12\*

frequencies in Southern Egyptians (44.3% of 79 people). Outside of the Egypt-Sudan-Horn axis it is scattered only very thinly, with very low frequencies outside of Africa.<sup>16</sup> Hassan et al. (2008) report a significant presence of E-V12\* in neighbouring Sudan, including 5/39 Nubians, and 5/33 Copts. E-V12\* makes up approximately 20% of the Sudanese E-M78.

Perhaps even more significant, a very distinct sub-clade of E-V12, E-V32, dominates Somalia and is found almost solely in and around the Horn of Africa. Cruciani et al (2007) found "the highest frequencies in the three Cushitic-speaking groups: the Borana from Kenya (71.4%), the Oromo from Ethiopia (32.0%), and the Somali (52.2%)." Sanchez et al. (2005), looking at the same clade using STR information in a study of Somali men in Denmark, stated that "the male Somali population is a branch of the East African population—closely related to the Oromos in Ethiopia and North Kenya (Boranas)" and that their lineages "probably were introduced into the Somali population 4000–5000 years ago." It is therefore quite notable that Hassan et al. (2008) in their study, observed this to be the most common of the sub-clades of E-M78 found in Sudan, especially among the Beja, as well as amongst the Masalit, and the Fur of Darfur.

Hassan et al. (2008) propose that this particular E-M78 presence in Darfur "might have been brought to Sudan from North Africa after the progressive desertification of the Sahara around 6000-8000 years ago" due to sudden climate change.

In summary, E-V12's modern distribution and diversity appear to show that Cushitic speakers descend at least partly from the direction of Southern Egypt, near the modern homeland of the Beja language.

#### Chadic Languages: from the East or North?

Another branch of Afroasiatic which is associated with some controversy is the Chadic branch of the southern Saharan area. This is a group of languages which includes Hausa, and is clustered around the south of Lake Chad. The debate concerns the question of which other Afroasiatic languages this group is closest to—Berber, or Chadic, or perhaps Egyptian. Did its ancestral form arrive in the area from the North (the Sahara) or the East (towards the Horn of Africa) or perhaps from the direction of Egypt to the Northeast?

Results so far seem to show no especially strong relationship to the E-M35 haplotypes found in these three possible source areas. Indeed, Chadic speakers show unusually low levels of any E haplotypes at all, even compared to neighbouring populations. On the other hand they have a remarkably high R haplogroup

<sup>15</sup> E.g., see Bender (1997): "Others question the integrity of Cushitic and see it as comprising in itself up to six distinct independent families."

<sup>16</sup> Cruciani et al. (2007) report highest non-African frequencies in Erzurum in Turkey (4%), and amongst French Basques (6.25%).

frequency. R is the dominant haplogroup of Europe, and one of the major haplogroups of Eurasia generally at least as far as India and Central Asia, but apparently ancient pockets of this Eurasian lineage are found from Sudan down to Cameroon (Luis et al., 2004; Cruciani et al., 2002). We can summarise some of the surprising results found in the literature with particular reference to Chadic speakers.

- Cruciani et al. (2002) tested 54 Chadic speaking men, including 21 Ouldeme (Uldeme) and 18 Daba, and from all these people only two E-M78 people were found. But 38 of the 54 men were in the clade R-M173, including all but one of the Ouldeme men. None of these men were positive for either M269 or M17, whereas most Middle Eastern and European R haplotypes are positive for one of these mutations.
- Hassan et al. (2008) found that Sudanese Hausa have one of the lowest E haplogroup presences in that country. They tested 32 Hausa and found only one E-M78 person that was M35+. On the other hand 13 of these 32 men were R-P25, a sub-clade of R-M173.
- Wood et al. (2005) tested 19 Podokwo, 28 Mandara, and 13 Uldeme (Ouldeme). One of the Podokwe was M78+ and this was the only E-M35 amongst the Chadic group. R-P25 in this group ranged from 61-97%.
- In a study of the Sahel by Bereir et al. (2007), M78 was more common than M81, but amongst the Hausa specifically, E-M78 was about 20% (out of 66 people), whereas it was about 41% amongst other Afroasiatic speakers (out of 81 people), and 26% amongst Nilo-Saharan speakers (out of 90 people). On the other hand 47% of the Hausa men were R-P25, while only 12 men out of the remaining 171 in this study were in this clade.

While the R haplogroup's presence in Africa is not restricted to Chadic speakers, also being found in neighbouring populations, the correspondence between this language group and this haplogroup is striking in the surveys done so far. This gives some quite surprising implications:

- The male lineages found amongst Chadic speakers are very different from those amongst other Afroasiatic language groups. Indeed, they seem genetically, not only geographically, isolated from other Afroasiatic speakers.
- The dominant R lineage found amongst them is very uncommon in most parts of Africa, and outside of this central African region, it is found in small

amounts only, and only in areas near to Europe or the Middle East.

- The R haplotypes found amongst Chadic speakers and their neighbours include a relatively high percentage of unusual cases that are negative for the SNPs M269 and M17, which between them dominate the R haplotypes found in Europe. This perhaps indicates a Middle Eastern origin. Flores et al. (2005) compared the R-M173\* haplotypes of this part of Africa to examples they found in the Dead Sea population in Jordan. They considered the Dead Sea area as a probable example of an isolated community reflecting a sampling of the ancient genetic population of the Levant. As discussed above, it is also very unusual because of its very high E-M34 presence.

Y-DNA evidence therefore seems to suggest that Chadic speaking populations have male line ancestors who arrived from the direction of the Middle East, possibly via the Nile. This is in conflict with proposals that Chadic came from the north, towards the Maghreb, which has been Ehret's proposal (2002b). It is also the account given recently in Tishkoff et al. (2009), in which Ehret is a co-author. Instead, proposals for an eastern origin towards the Nile (e.g., Blench, 1999c) are clearly favoured by the presence of the R-M173\* haplogroup, as well as by the lack of E-M81 and E-V65.

In this context the above-cited remarks of Hassan et al. (2008) concerning evidence of an E-M78-bearing Nilo-Saharan immigration into Sudan from the north, precisely into the area that separates Chadic from the ancient Afro-Asiatic language zones of Egypt, the Beja, and the Horn of Africa, seem significant. We may tentatively ask whether such a migration could have played a role in determining both the geographical and Y-DNA isolation of Chadic speakers from other Afroasiatic speakers.

That there was a movement of both Nilo-Saharan and Cushitic peoples moving south from Egypt around 5500 BCE is not controversial (e.g. Ehret, 2002b). But Blench (1999c) goes so far as to propose that languages "related to present-day Chadic were presumably once spoken in a strip across present-day Sudan but were later eliminated by movements of Nilo-Saharan speakers."

This is part of what Blench calls his "inter-Saharan hypothesis" whereby the Chadic languages descend from the languages of Cushitic pastoralists who moved westwards "from the Nile Valley to Lake Chad, as would the Shuwa Arabs, millennia later." We are led to wonder whether the Y DNA, with an apparently high Eurasian component, possibly entered the ancestral populations of Chadic speakers along with the new "technology" of herding. Indeed, Blench proposes

exactly such changes as part of the complex of events which sent the Chadic language family on a path to its present home, leaving Afroasiatic loanwords for domestic animals in many languages in Sudan.

The case of Chadic is an example which rewards a broader look at different types of DNA (MacEachern, 2007). That the Y DNA of this population, looked at in isolation, shows no obvious similarity to those of other Afroasiatic populations is not surprising, because when we look at Y lineages we are restricted only to a small part of the human genome, and usually the part which is most likely to emphasize the prehistoric geographic movements of new technologies, new ways of life, and languages (Wood et al., 2005). The recent autosomal DNA study of Tishkoff et al. (2009) confirms that on the whole, Chadic speakers are more closely related to their Nilo-Saharan neighbours than to any other Afroasiatic group. Looking at the genome beyond Y-DNA these peoples show far less Eurasian ancestry than the Beja, for example.

However, in contrast to what we see with Y DNA and autosomal DNA though, Černý et al. (2009) propose that mitochondrial DNA links appear to exist in between Chadic speakers and Cushitic speakers far to the East. So as with the Berber population in Northern Africa, the dominant paternal and maternal lines in modern Chadic populations may have quite different origins, with the female lines continuing to show signs of ancient links with distantly related peoples.

#### Omotic: Afroasiatic or not?

Omotic is a remote family of languages found in South-western Ethiopia, including areas near to Southern Sudan and Northern Kenya. In this case debate centres on whether Omotic is Afroasiatic at all. Also in this case, data is still building up.

Amongst Omotic speakers, the present author is only aware of one genetic sample of Wolayta from Ethiopia, who are included in the sample set of Cruciani et al. (2004, 2007). This study showed that out of 12 people tested, no less than 9 were in the E haplogroup:-

E(not E-M35)	2 individuals
E-M35*	2 individuals
E-V32	1 individual
E-V22	1 individual
E-M34	1 individual
E-V6	2 individuals

In other words 7 of the 12 men tested were E-M35+. This result is relatively typical of other regions in Ethiopia in the same Cruciani dataset. In detail we can remark:

- E-V6 is only found in the Horn of Africa and Kenya so far (Cruciani et al., 2004).
- E-M34, a sub-clade of E-M123 discussed above, is not considered likely to have originated in the Horn of Africa (as discussed above), but is found in areas where Semitic languages are spoken, especially including the South Semitic languages of Ethiopia. The Amhara, to the north of the Omotic region, are a dominant ethnic and linguistic group in Ethiopia, and they have high levels of E-M34. This shows unsurprisingly that the Wolayta are not isolated from other Ethiopians. E-M34 seems to have entered the area with Semitic languages, and therefore long after the split between Omotic and the other Afroasiatic languages.
- E-V32, a sub-clade of E-V12, is especially found amongst Cushitic speaking peoples so far, from Ethiopia and Somalia. As discussed above, it appears to have migrated from the area of Southern Egypt (Cruciani et al., 2007) but in Sudan it is mainly found amongst the Northern Beja and in Darfur, and not apparently amongst the Southern Sudanese who are neighbours to the Omotic speaking peoples (Hassan et al., 2008).
- E-V22, like E-V32, probably entered the area in a back migration from the direction of Egypt along the Nile (Cruciani et al., 2007). It is also found throughout neighbouring Sudan in small frequencies (Hassan et al., 2008).

In recent times Hassan et al. (2008) has given us some insight into the Y chromosomes of the peoples to the West of the Omotic languages, in southern Sudan. Amongst the Nilo-Saharan speaking Dinka, Shilluk and Nuer, E-M78 makes up about 25% of the male lines. This includes the above-mentioned northern sub-clades, E-V12\* and E-V22, both also found in similar small amounts in the Wolayta. Unfortunately Hassan et al. did not test for E-M123 or E-M34 or E-V6, but in any case, out of 53 people tested in these three Sudanese groups, only one was in E-M215 who was not in E-M78. Perhaps more importantly, in all three South Sudanese groups the sub-saharan haplogroups A3b2-M13 and B-M60, were both far more common than any other clade including the E haplogroups which dominate Cruciani et al.'s Wolayta data.

The E-M35 evidence therefore agrees with what appears to be a consensus—that the Omotic group is related to other Ethiopian groups, or at least more related to them than to any other neighbouring peoples. However concerning the purely linguistic question about very distantly related languages, it is difficult to dismiss arguments such as that of Orel and Stolbova, that Omotic was in a “*Sprachbund*” with Cushitic, with

much loaning of vocabulary and even grammar, making the phylogenetic tree difficult to reconstruct. Of course, ethnic groups in a Sprachbund are normally bound by genetic ties also, and it appears that Orel and Stolbova accept that the “Cushomotic” Sprachbund may have common Afroasiatic roots.<sup>17</sup>

### South Cushitic and E-M293

That there is a significant amount of E-M35 scattered in Eastern and Southern Africa was known since at least Semino et al. (2004) and Luis et al. (2004), but it was not until Henn et al. (2008) that a defining UEP was discovered that distinguished them from other E-M35 clades. They are therefore now identifiable as mainly corresponding to their own sub-clade of E-M35 named E-M293.

The dominant language family by far in these parts of Africa is Bantu, but not amongst populations who have the highest E-M293 levels. In the most southerly extensions of this genetic presence the languages are Khoisan or “click” languages, generally associated with the ancient hunter gatherers of all these regions before the Bantu migrations. The authors argue that this clade’s dispersal is in any case too old to be associated with the Bantu expansion, and indeed we can point out as discussed above that Bantu populations are not associated with E-M35. Instead the authors propose a link to the introduction of Nilo-Saharan languages and pastoralism into these areas, before the Bantu languages and Iron Age technology became dominant.

Although there is no evidence that Nilo-Saharan languages made it as far as Angola and Namibia, pastoralism itself did of course travel further than the languages, and as the authors discuss, pastoralists and Khoisan hunter gatherers are known in their areas of overlap to have merged in populations in Tanzania, further north, which are also amongst those with the highest E-M293 presence.

As discussed above, the Nilo-Saharan and Cushitic languages have been suggested by Ehret as being ancient pastoralist neighbours from the Eastern Sahara. Ehret (2002b) presents us with a review of the entry of pastoralism from the north into Kenya and Tanzania which also emphasizes the fact that it contained not only Nilo-Saharan groups, but also Cushitic speaking groups.

Henn et al.’s focus upon Nilo-Saharan is dependent upon one population in their study, the Datog, who had the highest frequency and diversity of E-M293 in their 454-person, 13-population study. Two Afroasiatic populations, the Wafiome and Burunge also scored highly, but were not tested as thoroughly. It will be very

interesting to see how this pattern develops as more data is collected for this newly discovered clade.

### Discussion/Conclusions

Review of what is being learnt about the E-M35 haplogroup confirms that despite real difficulties for the new discipline, genetics is quickly becoming a more powerful tool, adding to those already in the hands of researchers studying ancient migrations. Furthermore, some of these growing pains perhaps stem from insufficiently detailed multi-disciplinary efforts in the field so far, debatably leading to an over-emphasis upon movements of people which are relatively well-known and recent.

Advantages of using genetics as a tool include:

- Phylogenetic (“family tree”) relationship structures can be defined exactly even for very old relationships. In linguistics this becomes very difficult at a certain time depth, as in the cases of Omotic and Chadic. In archaeology, unanimous agreement about relationships between different material cultures is rare.
- Techniques for the estimation of ages within such family trees structures, while controversial, are less controversial than in linguistics, and therefore give a useful cross-reference to archaeological dating in such a way that strong indications of relationships between different times and places are also sometimes revealed.
- The scope for further rapid development of knowledge in this field is still extremely large. Looking at Y DNA in particular, compared to mitochondrial and autosomal DNA, it is particularly well-suited to attempts to reconstruct the movements of the most mobile elements of cultures, such as languages and technologies.

Concerning Afroasiatic, this review has been able to show how looking to one Y-chromosome haplogroup *in detail* can increasingly help to add a new perspective in multi-disciplinary discussions, both by narrowing the likely options and helping to propose others for attention. Especially important in this regard is to go beyond simply remarking the similarities of geographical distributions, (such as that of E-M35 and Afroasiatic as a whole) by looking at phylogenetic sub-structure (for example the specific sub-clades in Northern Portugal) and super-structure (such as the evidence that E-M35 has origins quite far from the Levant). Perhaps the most important thing about this, is that the level of phylogenetic detail which can be clearly defined, can increase almost without limit in the future, until even individual families can be identified

<sup>17</sup> This is cited in Bernal (1987) and Blench (2006). Blench remarks “these authors do not go into print with family trees” but he reports his understanding as supplemented by personal communication.

and placed into the greater phylogeny of all men by looking at SNP mutations. This is not yet practical on a large scale, but it seems inevitable, given current trends.

Our review of the E-M35 evidence gives many insights useful for multidisciplinary consideration in both linguistics and archaeology:

- The evidence strongly suggests that the male lineage most strongly associated with Afroasiatic, E-M35, clearly has an origin far from the Levant, in Africa.
- In Africa, the Levant, and the Arabian Peninsula, E-M35 is strongly associated with Afroasiatic languages, with the strongest links being in a great curve from the Maghreb to the Horn of Africa. In Anatolia and Europe this association is not apparent.
- Northern and Eastern Iberia appear to show signs of immigration which combined Middle Eastern and North African roots, and was possibly associated with the Neolithic Cardial culture.
- More generally, pockets of ancient Middle Eastern derived diversity seem to be scattered around the Mediterranean coasts and islands, possibly also due to the Cardial culture, and related Neolithic cultures.
- Berber populations, while overwhelmingly dominated by specific E-M35 male lineages, are not in the same sub-clades as found along the Nile and into the Horn of Africa.
- At least when looking to E-M35 and Y DNA, we can see that Chadic speakers are not only geographically isolated from other Afroasiatic speakers, but also to some extent, genetically isolated from them. This conclusion appears to support the “inter-Saharan hypothesis” of Blench (1999c) as an explanation concerning the origins of Chadic.
- That Cushitic languages came from the North closer to Egypt, is a possibility strongly favoured by the E-M35 evidence.
- Populations speaking Omotic languages, like those languages themselves, are more closely related to other Ethiopians than to nearby Nilo-Saharan speaking populations.

In some areas, looking at E-M35 on its own does not yet give enough fine detail to make a contribution to ongoing debates, and future research will need to involve finer resolution of sub-clades as yet undiscovered and/or the comparison of many different types of genetic data, including other regionally important Y haplogroups

(such as Y haplogroup J), mitochondrial DNA, and autosomal (recombining) DNA.

Very recent advances in the archaeology of the relevant areas such as the Maghreb were also shown to be very important, and this is likely to be a continuing trend which deserves constant attention from genetic and linguistic researchers into these areas.

## Acknowledgments

The author wishes to thank Ofer Ben-Yosef, Dirk Schweitzer, Jörg Lindtstädter, Victor Villareal, Peter Hrechdakian, Steven C. Bird, Sylvestre de Carvalho, Tim Reynolds, Jesus Flavio Martinez Morales, Vincent John Palozzi, Graeme Barker and Matthias Brenzinger for their encouragement, and their assistance in helping to collect and understand a wide range of literature, in many fields. He also thanks the active membership of the E-M35 Phylogeny Project, and “Dienekes Pontikos” of the Dienekes Anthropology Blog, for their thought-provoking discussion, which both inspired and informed this review of the published literature beyond genetics. Thanks must also go to Carlos Flores, and Antonio Brehm for answering questions about their published works.

## Web Resources

Wikipedia (English)

[http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)

ISOGG 2009 Y Phylogenetic Tree, E subpage

[http://www.isogg.org/tree/ISOGG\\_HapgrpE09.html](http://www.isogg.org/tree/ISOGG_HapgrpE09.html)

## References

- Adams SA, Bosch E, Balaesque PL, Ballereau SJ, Lee AC, Arroyo E, López-Parra AM, Aler M, Gisbert Grifo MS, Brion M, Carracedo A, Lavinha J, Martínez-Jarreta B, Quintana-Murci LI, Picornell A, Ramon M, Skorecki K, Behar DM, Calafell F, Jobling MA (2008) *The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula*. *Am J Hum Genet*, 83:725-736.
- Ammerman AJ, Biagi P, eds (2003) *The Widening Harvest. The Neolithic in Europe: Looking Back, Looking Forward*. Archaeological Institute of America, Boston.
- Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C (2004) *A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa*. *Am J Hum Genet*, 75:338-345.
- Bar-Yosef O (1987) Pleistocene connections between Africa and SouthWest Asia: an archaeological perspective. *The African Archaeological Review*; Chapter 5, pg 29-38.
- Bar-Yosef O (1998) *The Natufian Culture in the Levant, Threshold to the Origins of Agriculture*. *Evolutionary Anthropology* 6:159-177.
- Bar-Yosef O (2002) The Natufian Culture and the early Neolithic: social and economic trends in southwestern Asia, in Bellwood P, Renfrew C, eds. (2002).

Barich BB, Garcea EAA, Giraudo C (2006) [Between the Mediterranean and the Sahara: geoarchaeological reconnaissance in the Jebel Gharbi, Libya. \*Antiquity\*, 80:567-582.](#)

Barker G (2002) Transitions to farming and pastoralism in North Africa, in Bellwood P, Renfrew C (2002), pp 151–161.

Barnett WK (2000) Cardial pottery and the agricultural transition in Mediterranean Europe, in Price ed (2000).

Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, Hadziselimovic R, Vidovic S, Drobic K, Durmishi N, Torroni A, Santachiara-Benerecetti AS, Underhill PA, Semino O (2008) Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. [European Journal of Human Genetics](#), on-line publication 24 Dec 2008.

Beleza S, Gusmão L, Lopes A, Alves C, Gomes I, Giouzeli M, Calafell F, Carracedo A, Amorim A (2005) [Micro-phylogeographic and demographic history of Portuguese male lineages. \*Ann Hum Genet\*, 70: 181–194.](#)

Bellwood P, Renfrew C, eds. (2002) *Examining the farming/language dispersal hypothesis*. McDonald Institute for Archaeological Research, Cambridge.

Bellwood P (2004) Response to Ehret et al. (2004) in *Science*, 306:1680-1681.

Bellwood P (2005) *First Farmers*. Blackwell, Oxford.

Bender ML (1997) Upside down Afrasian. *Afrikanistische Arbeitspapiere*, 50:19-34.

Bernal M (1987) *Black Athena: the Afroasiatic Roots of Classical Civilization*, Rutgers University Press, Piscataway, NJ.

Bereir RE, Hassan HY, Salih NA, Underhill PA, Cavalli-Sforza LL, Hussain AA, Kwiatkowski D, Ibrahim ME (2007) Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel. [Eur J Hum Genet](#), 15:1183–1185.

Binder, D (2000) Mesolithic and Neolithic interaction in Southern France and Northern Italy, in Price ed. (2000).

Blench R (1999a) [The languages of Africa: macrophyla proposals and implications for archaeological interpretation in Blench R, Spriggs M, Conroy J, \*Language Change and Cultural Transformation\*, Routledge, Milton Park, UK.](#)

Blench R (1999b) Are the African pygmies an ethnographic fiction? In Biesbrouck K ed (1999) *Central African hunter-gatherers in a multidisciplinary perspective*, Stefan Elders & Gerda Rossel, pp. 41-60. Research School of Asian, African and Amerindian Studies (CNWS). Cited in Blench (2004a).

Blench R (1999c) [The Westward wanderings of Cushitic pastoralists: Explorations in the Prehistory of Central Africa](#), in Baroin C, Boutrais J, eds. (1999) *L'Homme et l'animal dans le bassin du lac Tchad, Actes du colloque du Réseau Méga-Tchad*, Orléans, 15-17 oct. 1997, Éditions IRD, Paris.

Blench RM (2001) [Types of language spread and their archaeological correlates: the example of Berber. \*Origini\*, 23.](#)

Blench R (2004a) [Archaeology and Language: Methods and Issues](#), in Bintliff JL, *A Companion to Archaeology*, Blackwell, Oxford.

Blench R (2004b) [Genetics and linguistics in sub-Saharan Africa](#), presented at SAfA 2004, Bergen, 27 June 2004.

Blench R (2006) [Archaeology, Language, and the African Past](#), Rowman Altamira, Lanham, MD.

Blench R (2008) [Agriculture and phylic dispersals: reevaluating the evidence](#), presented at the meeting, Us and Them: Modelling Past Genetic, Linguistic, and Cultural Boundaries, Bordeaux 15–17 May 2008, and subsequently revised.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between north-western Africa and the Iberian Peninsula. [Am J Hum Genet](#), 68: 1019–1029.

Brion M, Sobrino B., Blanco-Verea A., Lareu MV, Carracedo A (2004) Hierarchical analysis of 30 Y-chromosome SNPs in European populations. [Int J Legal Med](#), 119: 10–15.

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ (2007) Y-chromosome diversity characterizes the Gulf of Oman. [Eur J Hum Genet](#), 16:274-286.

Capelli C, Onofri V, Brisighelli F, Boschi I, Scarnicci F, Masullo M, Ferri G, Tofanelli S, Tagliabracchi A, Gusmão L, Amorim A, Gatto F, Kirin M, Merlitti D, Brion M, Blanco Verea A, Romano V, Cali F, Pascali V (2009) Moors and Saracens in Europe: estimating the medieval North African male legacy in southern Europe. [Eur J Hum Genet](#), advanced publication on-line, 21 Jan 2009.

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. [Hum Genet](#), 114:127-148.

Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, Frigi S, Yacoubi-Loueslati B, Ben Amor M, Slama A, Amorim A, Ben Ammar El Gaaied A, and Pereira L (2008) Post-Last Glacial Maximum Expansion From Iberia to North Africa Revealed by Fine Characterization of mtDNA H Haplogroup in Tunisia. [Am J Phys Anthr](#), on-line publication 17 Dec 2008.

Černý V, Fernandes V, Costa MD, Hájek M, Mulligan CJ, Pereira L (2009) Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. [BMC Evolutionary Biology](#), 9:63.

Chen J, Sokal RR, Ruhlen M (1995) Worldwide analysis of genetic and linguistic relationships of human populations. [Hum Biol](#), 67:595-612. Cited in Blench (2004a).

Coffman-Levy E (2005) A mosaic of people: the Jewish story and a reassessment of the DNA evidence. [J Genet Geneal](#), 1:12-33.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A back migration from Asia to Sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. [Am J Hum Genet](#), 70:1197–1214.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavnhia J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R (2004) Phylogeographic analysis of Haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. [Am J Hum Genet](#), 74: 1014-1022.

Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R (2006) Molecular dissection of the Y chromosome Haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. [Hum Mut](#), Mutation in Brief #906 (on-line only).



- Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon J-M, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh G, Barbujani G, Fuselli S, Vona G, Zografisnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, Scozzari R (2007) Tracing past human male movements in Northern/Eastern Africa and Western Eurasia: New clues from Y-Chromosomal Haplogroups E-M78 and J-M12. *Mol Biol Evol*, 24:1300-1311.
- Daugas J-P, El Idrissi A (2008) Le Néolithique ancien au Maroc septentrional: données documentaires, sériation typochronologique et hypothèses génétiques. *Bulletin de la Société Préhistorique Française*, 105:787-812.
- Di Gaetano C, Cerutti N, Crobu F, Robino C, Inturri S, Gino S, Guarrera S, Underhill PA, King RJ, Romano V, Cali F, Gasparini M, Matullo G, Salerno A, Torre C, Piazza A (2008) Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur J Hum Genet*, 17:91-99. Epub 2008 Aug 6.
- Diamond J, Bellwood P (2003) Farmers and Their Languages: The First Expansions. *Science*, 300:596-603.
- Diakonoff I (1998) The earliest Semitic society: linguistic data. *J Semitic Studies*, 43:209-219.
- Ehret C (1995) *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary*. University of California Press, Berkeley.
- Ehret C (2002a) *The Civilizations of Africa: A History to 1800*, James Currey Publishers, Oxford.
- Ehret C (2002b) Language family expansions: broadening our understandings of cause from an African perspective, in Bellwood and Renfrew eds. (2002).
- Ehret C, Keita SOY, Newman P (2004) The Origins of Afroasiatic a response to Diamond and Bellwood (2003) in the Letters of *SCIENCE* 306, no. 5702, p. 1680.
- Ennafaa H, Cabrera VM, Abu-Amro KK, González AM, Amor MB, Bouhaha R, Dzimir N, Elgaaied AB, Larruga JM (2009) Mitochondrial DNA Haplogroup H structure in North Africa. *BMC Genetics*, 10:8.
- Flores C, Maca-Meyer N, González AM, Oefner PJ, Shen P, Pérez JA, Rojas A, Larruga JM, Underhill PA (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *European Journal of Human Genetics*, 12:855-863.
- Flores C, Maca-Meyer N, Larruga JM, Cabrera VM, Karadsheh N, Gonzalez AM (2005) Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. *J Hum Genet*, 50: 435-441.
- Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, Paoli G, Tofanelli S, Calò CM, Ghiani ME, Varesi L, Memmi M, Vona G, Lin AA, Oefner P, Cavalli-Sforza LL (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am J Phys Anthro*, 121:270-279.
- Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T, Brehm A (2005) Y-chromosome Lineages from Portugal, Madeira and Açores Record Elements of Sephardim and Berber Ancestry. *Ann Hum Genet*, 69: 443-454.
- Hammer MF, Blackmer F, Garrigan D, Nachman MW, and Wilder JA (2003) Human population structure and its effects on sampling Y chromosome sequence variation. *Genetics*, 164: 1495-1509.
- Hassan FA (2002) Archaeology and linguistic diversity in North Africa, in Bellwood and Renfrew, eds. (2002 ).
- Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME (2008) Y-Chromosome Variation Among Sudanese: Restricted Gene Flow, Concordance With Language, Geography, and History. *Am J Phys Anthro*, 137:316-323.
- Henn BM, Gignoux C, Lin AA, Oefner PJ, Shen P, Scozzari R, Cruciani F, Tishkoff SA, Mountain JL, Underhill PA (2008) Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proc Nat Acad Sci (U.S.)*, 105:10693-10698. Supplementary data available at: <http://www.pnas.org/cgi/content/full/0801184105/DCSupplemental>
- Keita SOY (2005) Early Nile Valley farmers from El-Badary, aboriginals or "European" Agro-Nostratic Immigrants? Craniometric affinities considered with other data. *J Black Studies*, 36:191-208.
- Keita SOY, Boyce AJ (2005) Genetics, Egypt, and History: Interpreting Geographical Patterns of Y Chromosome Variation. *History in Africa*, 32:221-246.
- King RJ, Ozcan SS, Carter T, Kalfo lu E, Atasoy S, Triantaphyllidis C, Kouvatsi A, Lin AA, Chow CE, Zhivotovsky LA, Michalodimitrakis M, Underhill PA (2008) Differential Y-chromosome Anatolian Influences on the Greek and Cretan Neolithic. *Annals of Human Genetics*, 72:205-214.
- King R, Underhill PA (2002) Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. *Antiquity*, 76:707-14.
- Kislev ME, Hartmann A, Bar-Yosef O (2006) Early domesticated fig in the Jordan Valley. *Science*, 312:1372-1374.
- Linstädter J (2008) The Epipalaeolithic-Neolithic-Transition in the Mediterranean region of Northwest Africa. *Quartär*, 55: 41-62.
- Lucotte G, Mercier G (2003) Brief communication: Y-chromosome haplotypes in Egypt. *Am J Phys Anthro*, 121:63-66. Cited in Ehret et al. (2004).
- Luis JR, Rowold DJ, Regueiro M, Caciari B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet*, 74: 532-544. Also see Errata.
- MacEachern S (2007) Where in Africa does Africa start? Identity, genetics and African studies from the Sahara to Darfur. *J Social Archaeology*, 7:393-412. Cited in Blench (2004a).
- Mallory J (1989) *In Search of the Indo-Europeans. Language Archaeology and Myth*, Thames and Hudson, New York.
- Manen C, Marchand G, Carvalho AF (2007) Le Néolithique ancien de la péninsule Ibérique: vers une nouvelle évaluation du mirage africain? *Actes du 26e congrès préhistorique de France*, 133-151.
- Martinez L, Underhill PA, Zhivotovsky LA, Gayden T, Moschonas NK, Chow CE, Conti S, Mamolini E, Cavalli-Sforza LL, Herrera RJ (2007) Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *Eur J Hum Genet*, 15:485-493.
- Militarev A (2002) The prehistory of a dispersal: the proto-Afrasian (Afroasiatic) farming lexicon, in Bellwood and Renfrew, eds. (2002).
- Militarev A (2005) Once more about glottochronology and comparative method: the Omotic-Afrasian case. In (Aspects of

comparative linguistics - 1), FS S. Starostin, *Orientalia et Classica II* (Moscow), p. 339-408.

Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Foren Sci Int*, 157:23-35.

Perlès C, Montheil G (2001) The Early Neolithic in Greece: The First Farming Communities in Europe. Cambridge University Press, Cambridge.

Phillipson DW (2002) Language and Farming Dispersals in Sub-Saharan Africa, with Particular Reference to the Bantu-speaking Peoples, in Bellwood and Renfrew, eds. (2002).

Phillipson DW (2005) *African Archaeology*, Cambridge University Press, Cambridge, third ed.

Plucienik M (1996) Genetics, archaeology and the wider world, *Antiquity* 70:13-14. Cited in Blench (2004a).

Price TD, ed. (2000) *Europe's First Farmers*, Cambridge University Press, Cambridge.

Runnels C (2003) The origins of the Greek Neolithic: a personal view, in Ammerman and Biagi (2003 eds).

Sanchez JJ, Hallenberg C, Børsting C, Hernandez A, Morling N (2005) High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. *Eur J Hum Genet*, 13: 856-866.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science*, 290:1155-1159.

Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet*, 70: 265-268.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, Diffusion, and Differentiation of Y-

Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. *Am J Hum Genet*, 74:1023-1034.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, Oefner PJ (2004) Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mutat*, 24: 248-260.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Korze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM (2009) The Genetic Structure and History of Africans and African Americans. *Science*, published on-line, 30 April 2009.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon-Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet*, 65:43-62.

Underhill PA (2002) Inference of Neolithic population histories using Y-chromosome haplotypes, in Bellwood and Renfrew, eds. (2002).

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassman BI, Soodyall H, Hammer MF (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet*, 13:867-876.

Zarins J (1990) Early Pastoral Nomadism and the Settlement of Lower Mesopotamia, *Bulletin of the American Schools of Oriental Research*, No. 280, pp. 31-65.

Zeder MA (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Nat Acad Sci (U.S.)*, 105:11597-11604.

Zilhão J (2000) From the Mesolithic to the Neolithic in the Iberian Peninsula, in Price ed. (2000).

Zilhão J (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Nat Acad Sci (U.S.)*, 98,(24): 14180-14185.

Zvelebil M, Little M (2000) Transition to agriculture in eastern Europe, in Price (2000 ed.)