To request an email subscription to DNA Tribes® Digest, email digest@dnatribes.com with the subject “Subscribe.” To unsubscribe from DNA Tribes® Digest, email digest@dnatribes.com with the subject “Unsubscribe.”

Table of Contents:

Introduction ..................................................................................................................................... 1
Old Europes: Genetic Evidence of Indo-Europeans and Substrates in the West ......................... 2
Part One: Linguistic Background and Identification of
the Central Indo-European (CIE) Genetic Zone ............................................................................. 3
  Native Language Families of Europe ............................................................................................. 3
  Identification of the Central Indo-European (CIE) Genetic Zone .......................................... 5
  Central Indo-European (CIE) Genetic Contributions in Europe ............................................. 8
Announcements for July 2009: New Populations and Expanded Europa Analysis ....................... 9

Introduction

Hello, and welcome to the July 2009 issue of DNA Tribes® Digest. This month’s Digest will feature part one of a two part analysis to examine the relationship between languages and genetic structure among European sub-regions based on our new, more detailed DNA Tribes® Europa analysis. This analysis will apply linguistic models for a more in depth description of European genetic structure1.

Our analysis will begin with the Proto-Indo-Europeans, the ancient cattle-rustlers and horseback2 conquerors of Eurasia whose wagon trains left traces in a vast territory between Ireland and western China. For more information, a readable and engaging introduction to the subject is In Search of the Indo-Europeans: Language, Archaeology and Myth by J.P. Mallory.

Best regards and I hope to speak with you soon,
Lucas Martin
DNA Tribes

---

2 The archaeological record is unclear as to whether horses were first ridden or used to pull chariots. However, later Indo-European peoples used both cavalry and chariotry, and steppe cultures eventually came to rely more on cavalry than on chariotry in warfare.
Old Europes: Genetic Evidence of Indo-Europeans and Substrates in the West

Most Europeans living today speak languages of the Indo-European (IE) language family. Neighboring these predominant Indo-European languages are several unrelated tongues highlighted in Figure 1 below: the Basque languages (spoken in the Basque Europa sub-region) and the Uralic languages (spoken in the Finnic and Urals Europa sub-regions). However, division between Indo-Europeans and their neighbors are not so simple: Indo-European tongues spoken today are permeated with substrates, which are archaic elements of speech absorbed from other cultures.

This two-part article will explore the relationships between European languages and European genetic structure. Part One will discuss European language structure and identify a Central Indo-European (CIE) genetic zone. On the basis of this CIE genetic zone, Part Two3 will then identify genetic substrates present throughout Europe.

Figure 1: Genetic sub-regions of Europe identified by the newly expanded DNA Tribes® Europa analysis. Sub-regions where non-Indo-European languages are spoken are highlighted in red. More information about DNA Tribes® Europa is available at: http://dnatribes.com/dnatribes-europa.html.

---

Part One: Linguistic Background and Identification of the Central Indo-European (CIE) Genetic Zone

Native Language Families of Europe

Most languages spoken in Europe today belong to the Indo-European (IE) family of languages, which includes European languages such as English, Spanish, and Russian as well as Asian languages such as Armenian, Persian, and Hindi. Less well known is that Indo-European languages were once widely spoken in what is today the Turkic speaking world, including Iranian languages of the Scythians and other steppe nomads, Anatolian languages once spoken south of the Black Sea, and the Tocharian languages once spoken in the Tarim Basin. Figure 2 illustrates IE languages spoken today, as well as some IE languages that have fallen into disuse or been replaced by other languages in their former territories.

Figure 2: Map of Indo-European (IE) languages spoken around the world today. IE languages that are extinct or no longer in widespread use are listed in parentheses where they were once spoken.

4 The Indo-European and Turkic cultures of Central Asia are often described as categorically separate. However, the historical record is less clear cut, and present day Turkic populations retain substantial European genetic characteristics in addition to East Asian genetic characteristics. This suggests the possibility that Turkic cultures arose relatively recently (perhaps not long before the Göktürks entered history in the 6th century AD) from a fusion of early [Pre-Proto-]Mongolic cultures with preceding Indo-European Tocharians and Iranians, also absorbing nearby Paleo-Siberian, Uralic, and Yeniseian cultures. More detailed genetic analysis of Turkic populations is available in “Generals of the Steppes” at http://dnatribes.com/dnatribes-digest-2009-12-29.pdf.
However, Europe is also home to several languages of other families (illustrated in Figure 3). Uralic languages of Fenno-Scandinavia, such as Finnish (Suomi) and Sami, are related to languages native to eastern parts of European Russia and also to Hungarian (Magyar), which entered Central Europe from further east in the ninth century AD.

Another non-IE language spoken today in parts of present day Spain and France is Basque (Euskara). However, Southern Europe has been home to several other ancient non-IE cultures, including the Etruscans who founded an urban civilization in central Italy that predated IE Rome. In ancient Greece, the Indo-European Hellenes displaced and absorbed native Pelasgian cultures, whose non-Greek languages are sometimes thought to have been non-IE. Other early non-European languages of Southern Europe are known only from scattered fragments, such as Iberian and Raetic.

In regions neighboring Europe, other non-IE tongues are also spoken, including the Berber languages of North Africa, the extinct Hattic languages of Anatolia (partly absorbed by the IE Hittite language), the Semitic languages of the Levant, and the varied languages spoken in the Caucasus. More recently, Turkic languages have also been introduced to the steppe lands of Eastern Europe. These non-IE languages spoken in and near Europe can suggest potential sources for substrates in IE populations, by

---

Figure 3: Map of non-Indo-European (non-IE) languages native to Europe and neighboring lands. Languages with uncertain classifications are noted with a question mark (?). Some languages introduced in relatively recent times are also shown where they are spoken today, such as Hungarian (Magyar) and the Turkic languages.

identifying cultures that could have mediated gene flow between Indo-Europeans and non-Indo-Europeans.

**Identification of the Central Indo-European (CIE) Genetic Zone**

**Background:** The Indo-European languages are thought to derive from an ancient culture known as the Proto-Indo-Europeans (PIE). The PIE culture is known partly from linguists, who have compared similar words shared between IE languages spoken today and recorded in ancient texts (such as the Rigveda of India) to reconstruct a common root vocabulary. For instance: similarities between the English word *father*, Old Irish *athir*, Latin *pater*, Greek *pater*, Armenian *hayr*, and Sanskrit *pitar*, are used to reconstruct the Proto-Indo-European root *phater*.6

Compared to this reconstructed PIE root language, all IE languages spoken today are thought to have changed over time, in part due to the absorption of non-IE speech patterns (substrates). However, the modern languages thought to be most similar to the PIE root language are Baltic languages (Latvian and Lithuanian) and Slavic languages (such as Polish and Russian). Although they too have diverged from PIE in some ways, these Baltic and Slavic languages are considered the least altered from the original PIE tongue in terms of grammar, pronunciation, and vocabulary.

These linguistic reconstructions are complemented by the work of archaeologists such as Marija Gimbutas, who documented a mobile horse riding Kurgan culture (named for a type of burial mound) that spread from a base in present day southern Russia7. This Kurgan culture corresponds closely to the picture of early Indo-Europeans depicted in literary works such as the Indian *Rigveda* and the Irish *Táin Bó Cúailnge* (“Cattle Raid of Cooley”): highly mobile societies emphasizing pastoralism (particularly cattle raising) more than agriculture, patrilineal descent, and a military mode of life. According to Gimbutas’ Kurgan hypothesis, elements of this martial horse riding culture conquered and “kurganized” neighboring societies of Northern Europe and then spread outward to lands of the Mediterranean and Central Asia.

However, Gimbutas’ Kurgan hypothesis has been contested, and several alternate theories of PIE origins have been proposed (a summary of some notable PIE Homeland theories is provided in Table 1). For instance, Colin Renfrew’s Anatolian hypothesis would place PIE origins in early farming cultures of Turkey, where the earliest written records of an IE language are attested. Similarly, the Out of India theory would trace PIE origins to the Indian Subcontinent, home of the Sanskrit language that amazed European scholars with its similarity to classical Greek and Latin. The Corded Ware culture that grew out of local Northern European traditions (which Gimbutas considered antithetical to Kurgan cultures) has also been traditionally associated with the IE cultures. Along similar lines, Mario Alinei has proposed a Paleolithic Continuity Theory tracing PIE origins to Stone Age hunting cultures resident in Europe since the first settlements of modern humans, long before the spread of agriculture or the domestication of the horse.

---

6 Examples are from *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World* by J. P. Mallory and D.Q. Adams. Words are written here in the ordinary English alphabet for clarity and ease of reading.

7 This Kurgan staging area roughly corresponds to the Scythian sub-region identified in DNA Tribes® Europa analysis.
DNA Tribes® Digest July 29, 2009

Theory | Proponents | PIE Homeland
--- | --- | ---
Anatolian hypothesis | Colin Renfrew | Turkey
Kurgan hypothesis | Marija Gimbutas | Southern Russia
Out of India theory | Lachhmi Dhar Shastri Kalla | India
Paleolithic Continuity theory | Mario Alinei | Europe

Table 1: Several of the many theories of PIE origins (also known as *Urheimat* or “Original Homeland” theories).

**Genetic analysis**: In order to assess the multiple PIE homelands proposed by linguists and archaeologists, a European Index was constructed to identify the regions and sub-regions where genetic patterns are most specifically European. The European Index was calculated as a percentage of general European ancestry, as distinguished from (related) Near Eastern ancestry and ancestry from other continents. European Index scores were calculated for all European sub-regions. European Index scores were also calculated for all non-European world regions where Indo-European languages are spoken (the Mesopotamian, North India, Eastern India, and South India regions) or have been spoken in the past (the Altaian region). Results are listed in Table 2 and illustrated in Figure 4.

<table>
<thead>
<tr>
<th>Europa Sub-Region or World Region</th>
<th>European Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germanic (IE)</td>
<td>1.000</td>
</tr>
<tr>
<td>Polish (IE)</td>
<td>1.000</td>
</tr>
<tr>
<td>Balkan (IE)</td>
<td>1.000</td>
</tr>
<tr>
<td>Norse (IE)</td>
<td>0.990</td>
</tr>
<tr>
<td>Belgic (IE)</td>
<td>0.987</td>
</tr>
<tr>
<td>Celtic (IE)</td>
<td>0.977</td>
</tr>
<tr>
<td>Russian (IE)</td>
<td>0.977</td>
</tr>
<tr>
<td>Scythian (IE)</td>
<td>0.971</td>
</tr>
<tr>
<td>Finnic (non-IE)</td>
<td>0.880</td>
</tr>
<tr>
<td>Basque (non-IE)</td>
<td>0.835</td>
</tr>
<tr>
<td>Portuguese (IE)</td>
<td>0.807</td>
</tr>
<tr>
<td>Spanish (IE)</td>
<td>0.792</td>
</tr>
<tr>
<td>Thracian (IE)</td>
<td>0.784</td>
</tr>
<tr>
<td>Urals (non-IE)</td>
<td>0.676</td>
</tr>
<tr>
<td>Italian (IE)</td>
<td>0.653</td>
</tr>
<tr>
<td>Greek (IE)</td>
<td>0.526</td>
</tr>
<tr>
<td>Altaian (former IE)</td>
<td>0.302</td>
</tr>
<tr>
<td>Mesopotamian (some IE)</td>
<td>0.196</td>
</tr>
<tr>
<td>North India (IE)</td>
<td>0.122</td>
</tr>
<tr>
<td>Eastern India (some IE)</td>
<td>0.000</td>
</tr>
<tr>
<td>South India (some IE)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2: European Index scores for Europa sub-regions and non-European world regions where Indo-European (IE) languages are spoken or were spoken in the past.

Discussion: Three European sub-regions were assigned the highest European Index scores: the Balkan, Germanic, and Polish sub-regions. This indicates that of all surveyed regions, genetic patterns in these three sub-regions are most specifically European. These scores do not imply that this zone has been isolated from world patterns of gene flow in any absolute sense, but do indicate that it is relatively characteristic of Europeans. These three sub-regions are located in adjacent territories of Central Europe (mapped in Figure 4), which will be termed the Central Indo-European (CIE) genetic zone in subsequent portions of this article. These results indicate that the highest European Index scores were observed in geographically central locations of Europe where Indo-European languages are spoken, including linguistically conservative Balto-Slavic languages. However, this CIE zone does not include the Scythian region where the Kurgan cultures (central to Gimbutas’ Kurgan hypothesis) are thought to have

---

8 Large scale relationships between Europe, the Near East, and India are explored in more detail in “Macro Relationships in West Eurasia” at [http://dnatribes.com/dnatribes-digest-2010-03-31.pdf](http://dnatribes.com/dnatribes-digest-2010-03-31.pdf).
originated. Instead, the CIE genetic zone is more similar to the Corded Ware horizon, which developed from local cultural traditions of North-Central Europe (such as the Funnelbeaker or TRB culture).

Of all theories of PIE origins discussed above, these observations are most consistent with the basic premise of Alinei’s Paleolithic Continuity Theory: that PIE cultures emerged from an older background of indigenous Stone Age cultures resident in Europe long before the development of agriculture or pastoralism. This would be consistent with a slow changing and locally rooted aspect of European language and genetic structure, in contrast to more swiftly changing technological and political developments. This need not contradict Gimbutas’ basic description of Kurgan developments on the Pontic-Caspian steppe and transmission to other parts of Europe, but suggests this “kurganization” was a later process impacting an older CIE culture indigenous to North-Central Europe.

It is also notable that substantial European Index scores are observed for non-European regions where Indo-Europeans are thought to have expanded in ancient times: in particular, the Altaian, Mesopotamian, and North India regions⁹. This would be consistent with substantial population expansions from Europe associated with the spread of IE languages in Asia, facilitated by the mobile and militaristic Kurgan culture.

Figure 4: Map of the Central Indo-European (CIE) zone (highlighted in blue) identified on the basis of European Index scores listed in Table 2.

Central Indo-European (CIE) Genetic Contributions in Europe

**Genetic analysis:** Genetic contributions from the Central Indo-European (CIE) genetic zone were then identified for all IE speaking European sub-regions. Results are illustrated in Figure 5 below.

![Figure 5: Map of Central Indo-European (CIE) contributions to Indo-European speaking sub-regions of Europe.](image)

**Discussion:** Results indicate substantial contributions from the CIE zone to all IE speaking sub-regions of Europe. Contributions are highest in the Germanic (96.8%), Polish (95.8%), and (to a lesser extent) Norse (90.6%) regions (corresponding to the Funnelbeaker and later Corded Ware archaeological horizons). However, contributions to the Scythian sub-region (corresponding to the Kurgan staging area) are substantially lower (80.8%). As discussed earlier, this suggests that perhaps the Kurgan culture was an extension of an older North-Central European CIE culture that expanded to the Pontic-Caspian steppe, developed the mobile and martial Kurgan mode of life, then transmitted it back to North-Central Europe.

Now that the CIE genetic zone has been identified, it can be used as a point of reference to identify the remainder of European genetic structure. In Part Two of our analysis, we will examine European sub-regions to identify the substrates that underlie the CIE contributions that predominate where IE languages are spoken.

---

Announcements for July 2009: New Populations and Expanded Europa Analysis

We are pleased to announce the addition of several new populations to our database. Numbers in parentheses indicate the number of individuals included in each sample:

New East Asian Populations:
- Evenki (Inner Mongolia, China) (90)
- Han (Chengdu, China) (210)
- Han (Liaoyang City, China) (2211)
- Han (North China) (597)
- Han (Shandong, China) (200)
- Han (Xian, China) (100)
- Kathmandu, Nepal (77)
- Korean (Mudanjiang, China) (98)
- Lhasa, Tibet (196)
- Maonan (Guangxi, China) (143)
- Newar (Nepal) (66)
- Temang (Nepal) (45)
- Tibet (153)

New European Populations:
- Bulgarians (Bulgaria) (103)
- Karakachani (Bulgaria) (102)
- Wallachia, Romania (1910)

New Latin American Populations:
- Central West Brazil (1,232)
- North Brazil (2,128)
- Northeast Brazil (6,136)
- South Brazil (13,580)
- Southeast Brazil (2,658)

Newly Expanded DNA Tribes® Europa Analysis:

DNA Tribes® Europa add-on analysis has now been expanded to identify seventeen genetic sub-regions among European populations, for an enhanced comparison to European genetic structure.


Updates:

Previous DNA Tribes customers who would like to update their results to include these new populations and our most up to date algorithms can order using the $24.99 "Update Your Analysis" option through our secure online checkout at: [http://dnatribes.com/order.html](http://dnatribes.com/order.html).