THE INDO-EUROPEAN LANGUAGE FAMILY: SOLVING A 200-YEAR-OLD ENIGMA?

Paul Heggarty & Cormac Anderson
MPI for the Science of Human History (Jena)

heggarty@shh.mpg.de — http://eva-mpg.academia.edu/PaulHeggarty
The Indic, Iranian and Armenian branches of Indo-European.

(not Dravidian)
Language lineage of more of the world’s population and land area than any other.
THREE LEVELS: THE INDO-EUROPEAN ENIGMA

• When?
  – Short chronology: c. 6000 BP.
  – Long chronology: c. 9000 BP.

• Where?
  – Pontic-Caspian Steppe (Ukraine).
  – Central-East Anatolia (Turkey).

• Why?
  – Nomadic pastoralism and technologies: domestication of horse, riding, wheel.
  – Farming, demographic expansion.

• Not (quite) the only games in town ...
  – Indigenous to India? Armenia? Hybrids?
Sprung from a Common Genetic Source?

- Uralic
- Balto-Slavic
- Scandinavian
  - Germanic
  - (all or part)

[Genetic figures from Haak et al. (2015)]
How (Not) to Do "Language, Archaeology and Genes"

Language ≠ Culture ≠ Genes

1 language family ≠ 1 archaeological culture ≠ 1 genetic profile

• But the same forces, ‘processes’ impacted on human societies and populations, and shaped all records (linguistic, archaeological, genetic) of the human past.

  e.g. Language families = expansive, divergent processes.

Track a single common genetic component within different overall patterns.
ANCIENT DNA

—

ASKING THE RIGHT

‘INDO-EUROPEAN QUESTION’
Massive migration from the steppe was a source for Indo-European languages in Europe

Wolfgang Haak, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Balkir Bañón and Roderik Poinar

*et al.*

Received 29 December 2014  |  Accepted 12 February 2015  |  Published online 02 March 2015

Haak *et al.* (2015)

“a steppe origin of at least some of the Indo-European languages of Europe.”
Population genomics of Bronze Age Eurasia

... responsible for shaping major parts of present-day demographic structure in both Europe and Asia.

Our genomic evidence for the spread of Yamnaya people from the Pontic-Caspian steppe to both northern Europe and Central Asia ... corresponds well with the hypothesized expansion of the Indo-European languages

Allentoft et al. (2015: 167, 170)

Our findings are consistent with the hypothesized spread of Indo-European languages during the Early Bronze Age.

• No. N. Europe and C. Asia = only a minority subset of those languages.
ASKING THE *WRONG* ‘INDO-EUROPEAN’ QUESTION?

These results provide support for the theory of a steppe origin of at least some of the Indo-European languages of Europe.

Haak *et al.* (2015: 3, abstract)

- The ‘Steppe hypothesis’ is **not** …
  - a “theory of a steppe origin of at least some of the Indo-European languages of Europe”.

- The ‘Steppe hypothesis’ is …
  - a theory of a steppe origin of **all** the Indo-European languages of Europe…
  - **and** of all Indo-European languages **elsewhere too**: Anatolia, Iran, India …
WHAT IS THE ‘INDO-EUROPEAN’ QUESTION?

Is the Yamnaya → Corded Ware movement …

• Sufficient to explain all Indo-European in Europe, incl. in Mediterranean?

• Compatible with Indo-European outside Europe also originating in Yamnaya?

• Too recent to explain the time-depth of the whole family, at just 4500 BP?

• Or is Yamnaya → Corded Ware just one part of a larger, deeper story…
  … just as pastoralism is part of the wider spread of food production?
What The Language Data Say
Reconstruct Culture? Horse, Wheel and ... King?

naïveté [that] seems to enjoy the status of high acumen” (Pulgram 1958: 145)

The \(*H₃rég\) has no clothes

Nor a wheel, nor a domesticated horse, etc.
CALLING THE BLUFF ON ‘LINGUISTIC PALAEONTOLOGY’

It looks as if ‘wheel’ was not in the proto-lexicon and the various words for it were created independently after the dispersal. [Coleman 1988: 450]

the whole doctrine of making cultural inferences from linguistic evidence, known as linguistic palaeontology, has rarely enjoyed particularly high repute [Anttila 1989: 373]

in practice, however, there are major pitfalls ... and the reliability of the approach is questioned by many linguists [Trask 2000: 198]

the method is so problematic or limited in reliability, and treated sceptically or rejected even by most present-day Western linguists [Bryant (2001): 240]

If Renfrew were able to convince ... that the first farmers were the only possible bearers of PIE, then philologists could probably explain away all the shared vocabulary that has seemed to imply later phases of civilization. [Sims-Williams 1998: 510]
LANGUAGE DATING BY AMOUNT OF CHANGE?

Mallory's dating, which presupposes that Proto-Anatolian, Proto-Indo-Iranian, Greek and other descendant languages could have diverged from each other for a mere 2000 years, is absolutely inconceivable. time frame is "utterly unrealistic"

[Dolgopolosky 1990: 239]

• Lithuanian: ‘wonder’ at retention of case endings since Proto-Indo-European?
  – But if no change in 6000 years, why not … none in 9000 years?

• Avestan and Vedic ‘so close’, a matter of centuries?
  – Cognacy in basic lexicon (COBL), entered independently by language experts = only 55-60%.
# How Fast Do Sounds ‘Decay’?

## In c. 2000 years:

<table>
<thead>
<tr>
<th>Latin</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>[akʷam]</td>
<td>[o] &lt;eau&gt;</td>
</tr>
<tr>
<td>[altʊm]</td>
<td>[o] &lt;haut&gt;</td>
</tr>
<tr>
<td>[ad illʊs]</td>
<td>[o] &lt;aux&gt;</td>
</tr>
<tr>
<td>[habitʊm]</td>
<td>[y] &lt;eu&gt;</td>
</tr>
<tr>
<td>[kalidʊm]</td>
<td>[ʃo] &lt;chaud&gt;</td>
</tr>
<tr>
<td>[kaballos]</td>
<td>[ʃfo] &lt;chevaux&gt;</td>
</tr>
<tr>
<td>[stellam]</td>
<td>[etwal] &lt;étoile&gt;</td>
</tr>
<tr>
<td></td>
<td>[steauə] &lt;stjauǎ&gt;</td>
</tr>
</tbody>
</table>

## In c. 2500 years:

<table>
<thead>
<tr>
<th>Proto-Germanic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[augoːn]</td>
<td>[aɪ]</td>
</tr>
<tr>
<td>[auzoːn]</td>
<td>[ɪə]</td>
</tr>
</tbody>
</table>
How slowly do sounds ‘decay’?

In c. 6000-9000 years...

From [bʰréh₂t(ē)r] brother in Proto-Indo-European:

- → [bʌðə] <brother> English
- → [fʁɛʁ] <frère> French
- → [bhā'ī] <भाई> Hindi
- → [braut] <brawd> Welsh
- → [brat] <brat>  <брат> Russian
THE LEAST BAD WAY TO DATE ...

- Comparative data points that a language can iteratively change.
  - Not sounds or structures that may disappear and cannot then change again.

- Cognate status, in basic word concept slots.
  = do languages use a related, inherited word for a given meaning.
  
  *e.g.* German *Arm* cognate with English *arm* — but *Bein* not cognate with *leg*.

- Indo-European languages ‘descend with modification’ from common ancestor.

- Bayesian, probabilistic phylogenetic analysis, and ‘relaxed clock’:
  - No constant rate: rate variability throughout the tree.
  - No single rate: separate rate (and distribution) for each target concept.
  - Calibrate with as many known ancient languages as possible.
  - No single, all-or-nothing ‘perfect phylogeny’.
  - Results expressed as distributions of most highly probable date ranges.
DESPERATE FOR A (STEPPE) DATE?

Chang et al. (2015)
A CONSISTENT DATA-SET

Cognacy in Basic Lexicon in Indo-European:  www.cobl.info
CoBL — Cognacy in Basic Lexicon in Indo-European

With Cormac Anderson

— and 70+ language and Indo-European specialists (e.g. Martin Kümmel, Jena)

• 175 target concepts, strictly defined.
• Consistency policies, for quantitative applications.
• Completely new data, by language experts.
• Improved handling of loanwords.
• Expanded coverage, more (and corrected) historical language calibration points.
• New tools for qualitative and quantitative analysis.
• Fixes errors and Bayesian model mis-specifications.
• Multi-state analysis, for data by word meaning.
• Fine-tree structure much more accurate in known historical cases.
**Yamnaya: A Secondary Phase in Indo-European Divergence**

Each gridline = 1000 years

recent results from Bayesian phylogenetic analysis
BACK TO THE ANCIENT DNA
Haak *et al.* (2015) Fig. 3: Admixture proportions in modern DNA

- **Haak *et al.* (2015):** Corded Ware population a ‘replacement’ from Steppe.
- **Allentoft *et al.* (2015: 171):** “admixture with the local Neolithic people”.

---

**Europe: North(-East) vs. Mediterranean**

Linguistic and historical origins

<table>
<thead>
<tr>
<th>Language</th>
<th>Uralic</th>
<th>Balto-Slavic</th>
<th>Scandinavian</th>
<th>Germanic (all or part)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norwegian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithuanian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Icelandic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scottish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belarusian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ukrainian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orcadian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French_South</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Croatian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish_North</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulgarian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuscan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basque</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bergamo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albanian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sardinian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Early Neolithic (LBK_EN)**
- **Western European Hunter-Gatherer (Loschbour)**
- **Yamnaya**
**SOME INDO-EUROPEAN — SOME NOT?**

- Whole of Indo-European split well before 4500 BP.
- Balto-Slavic split: compatible with 4500 BP — and with Uralic.
- ‘Eastern’ Indo-European characteristics, contact with early Uralic?

![Map of Europe showing linguistic regions](image)

- Slavic (+Germanic?) much more recent: replaced unknown languages.
- Celtic? No mention in Olalde *et al.* (2017) on Bell Beaker aDNA.
“Mycenaeans ... had at least three-quarters of their ancestry from the first Neolithic farmers of western Anatolia and the Aegean, and most of the remainder from ancient populations related to those of the Caucasus and Iran.”

“the Mycenaeans ... deriving additional [= less than half of one quarter] ancestry ... introduced via a proximal source related to the inhabitants of either the Eurasian steppe or Armenia ... ”

“Continuity but not isolation in the history of populations of the Aegean, before and after the time of its earliest civilizations.”
WHERE DID YAMNAYA PASTORALISM COME FROM?

• Food production on Steppe ... ultimately from the Fertile Crescent.

• In Steppe environment, specialises to pastoralism.

• Haak et al. (2015: SI 124): “the Yamnaya ... had additional ancestry of a farmer-related population ... perhaps from the Caucasus or Near East”.

• Incoming farmers (bring Indo-European?), admix with locals (Uralic?).

• Yamnaya → Corded Ware = secondary within the spread of food production.
WHAT ABOUT THE INDO-?

• Bellwood adamant: cultures of subcontinent do not go back to the Steppe.

• No ancient DNA yet published for India — preservation limited.

• BMAC archaeological samples are being processed, however — often proposed as the candidate ‘Indo-Iranic’ culture that brought Indo-Iranic to Iran & India.
Upper Palaeolithic genomes reveal deep roots of modern Eurasians


Nature Communications 6, Article number: 8912 | doi:10.1038/ncomms9912
Received 20 July 2015 | Accepted 15 October 2015 | Published 16 November 2015

We extend the scope of European palaeogenomics by

CHG genomes significantly contributed to the Yamnaya steppe herders who migrated into Europe ~3,000 BC, supporting a formative Caucasus influence on this important Early Bronze age culture. CHG left their imprint on modern populations from the Caucasus and also central and south Asia possibly marking the arrival of Indo-Aryan languages.

Jones et al. (2015)
Early Neolithic genomes from the eastern Fertile Crescent

Farnaz Broushaki, Mark G Thomas, Vivian Link, Saioa López, Lucie van Dorp, Karola Kirsanow, Zuzana

The genetic structure of the world’s first farmers

Iosif Lazaridis1,2,3, Dani Nadel3, Gary Rollefson4, Deborah C. Merrett5, Nadin Rohland1, Swan Mallick1,2,6, Daniel Fernandes7,8, Mario Novak7,9, Beatriz Gamarra7, Kendra Sirak7,10, Fu1,12,13, Gloria16, Fanny Bocquentin17,18, Ahuva-Sivan19,20, Bejenaru23, Matthias24, Philippe Froguel28,29,20,31,32,33, Darren34, Niall O'Reilly35, Martin B.

First farmers’ motley roots

Unrelated groups adopted farming at about the same time in different parts of the Fertile Crescent

By Ann Gibbons

The genetics of an early Neolithic pastoralist from the Zagros, Iran


The agricultural transition profoundly changed human societies. We sequenced and analysed the first genome (1.39x) of an early Neolithic woman from Ganj Dareh, in the Zagros Mountains of Iran, a site with evidence for an economy based on goat herding, ca. 10,000 BP. We show that Western Iran was inhabited by a population genetically most similar to hunter-gatherers from the Caucasus, but distinct from the Neolithic Anatolian people who later brought food production into Europe. The inhabitants of Ganj Dareh made little direct genetic contribution to modern European populations.
... BECOMES THE ‘EASTERN FERTILE CRESCENT’ COMPONENT?

“Early Neolithic genomes from the Zagros region of Iran (eastern Fertile Crescent), where some of the earliest evidence for farming is found ...”

“... show affinities to modern day Pakistani and Afghan populations, but particularly to Iranian Zoroastrians. We conclude that multiple, genetically differentiated hunter-gatherer populations adopted farming in SW-Asia, that components of pre-Neolithic population structure were preserved as farming spread into neighboring regions, and that the Zagros region was the cradle of eastward expansion.”

Broushaki et al. (2016)

Lazaridis et al. (2016)
A GENETIC COMPONENT COMMON TO BOTH INDO- AND EUROPEAN ...

- **Eastern Fertile Crescent:**
  - Yes, in Indo-Iranic.

- **North Eurasia component:**
  - Not in Indo-Iranic.

- **Yamnaya = mix of both.**

- **Main component shared in Europe and N. India:**
  - Indigenous Steppe component? No.
  - Eastern Fertile Crescent component? Yes.
How and When Did Indo-European Reach Iran & India?

Modern Middle Easterners and South Asians appear to possess mixed ancestry from ancient Iranian and Steppe populations (tables S19 and S20). However, Steppe-related ancestry may also have been acquired indirectly from other sources (7) and it is not clear if this is sufficient to explain the spread of Indo-European languages from a hypothesized Steppe homeland to the region where Indo-Iranian languages are spoken today. On the other hand, the affinities of Zagros Neolithic individuals to modern populations of Pakistan, Afghanistan, Iran, and India is consistent with a spread of Indo-Iranian languages, or of Dravidian languages (which includes Brahui), from the Zagros into southern Asia, in association with farming (19).

Broushaki et al. (2016)

• Contra Lazaridis et al. (2016): both Iran and (Bronze Age?) Steppe?
When Steppe Genes — But Not Languages — Reached India?
INDO-EUROPEAN ORIGINS: MORE COMPLEX, MORE REALISTIC?

• Ancient DNA not fully compatible with either Steppe or farming hypothesis.
• Both part-right, both part-wrong? Various possible ‘hybrid’ hypotheses...

For more on these themes, see:

• Other papers at: http://eva-mpg.academia.edu/PaulHeggarty


http://doi.org/10.1126/science.353.6296.207


