Archaeological temporal constraints make phylogenetic methods support the steppe homeland theory of Indo-European

Igor Yanovich, Armin Buch, Johannes Dellert, Marisa Delz, Fabrício Marcel Ferraz Gerardi, Roland Mühlenbernd, Johannes Wahle and Gerhard Jäger

Universität Tübingen / Carnegie Mellon University

Capturing Phylogenetic Algorithms for Linguistics © Lorentz Center, Leiden

October 29, 2015
Indo-European homeland: steppe or Anatolia?

Many linguists and archaeologists: steppe
Some linguists and archaeologists: Anatolia

Previous phylogenetic studies: surely, Anatolia

Our study: well, we got support for steppe
We used [Bouckaert et al., 2012]’s methods and added two prehistoric constraints.

Bottom line: the issue is still not settled
Plan

1. The archaeology of Indo-European
2. Two prehistoric calibration constraints
3. Our results
4. Quick comparison with other recent results
### Steppe and Anatolia: what is the timing?

<table>
<thead>
<tr>
<th>Study</th>
<th>Anatolia</th>
<th>Steppe</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Ryder and Nicholls, 2011]</td>
<td>ca. 6500 BC</td>
<td>4500-4000 BC</td>
</tr>
<tr>
<td>[Bouckaert et al., 2012]</td>
<td>7500-6000 BC</td>
<td>4000-3000 BC</td>
</tr>
<tr>
<td>Present paper</td>
<td>6500-6000 BC</td>
<td>4700-4200 BC</td>
</tr>
</tbody>
</table>

![Temporal constraints chart](chart.png)
Timing IE under the Anatolian (=farming) theory

- Colin Renfrew formulated his Anatolian theory on much less data than now.
- In the last decade, it transpired that there were two waves of migration from Anatolia into the Balkans [Özdoğan, 2014]:
  - First wave 6500-6400 BC: rather small; Fikirtepe; both pure-immigrant and merged-with-locals sites; possibly only got as far as Turkish Thrace.
  - Second wave 6100-5900 BC: massive; Aşağı Pınar, pre-Karanovo, etc.; covered the Balkans with hundreds of new farming sites similar to each other.
- NB: the migrants didn’t take the material culture of a single place. They were a diverse bunch. [Perlès, 2005], [Özdoğan, 2010]
Timing IE under the steppe theory

Two candidates for the first IE split:

- Around 4400–4200 BC, Suworovo-Novodanilovka culture emerges, expressed in burials. The Novodanilovka group is in the steppes, near the Dnieper. The Suworovo group is mostly in the Danube delta. Suworovo is roughly simultaneous with massive abandonment of old tell settlements in the area.

- Suworovo is likely connected to the Sredni Stog culture of ordinary settlements on the Dnieper. Sredni Stog is formed from the local component and a component from the Khvalynsk culture on the Volga (pottery, new burial ritual [Anthony, 2007], craniometry [Potekhina, 1999], [Kazarnitsky, 2014]). Later Khvalynsk will also participate in the formation of the Yamnaya culture.

Fig. 5 from [Kazarnitsky, 2014]

10: Khvalynsk I; 12: Sredni Stog

Baltic (Zvejnieki), and the Upper Volga (Sakhtysh). Their shared features are very broad and somewhat flattened face and nose. Another Chalcolithic group, Khvalynsk II,
### Good and bad

<table>
<thead>
<tr>
<th>Anatolia</th>
<th>Steppe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Some of the good points</strong></td>
<td></td>
</tr>
<tr>
<td>Farming language dispersals exist</td>
<td>The steppe did give rise to widely and rapidly distributed cultures</td>
</tr>
<tr>
<td>Good deal of time for “core PIE” after Anatolians depart</td>
<td>Good deal of time for “core PIE” after Anatolians depart</td>
</tr>
<tr>
<td><strong>Some of the bad points</strong></td>
<td></td>
</tr>
<tr>
<td>No obvious base left in Anatolia to develop the Anatolian branch</td>
<td>Perhaps Sredni Stog and Suvorovo are different cultures after all (not chiefs and commoners) [Rassamakin, 1999]</td>
</tr>
<tr>
<td>Problems with long presence in Anatolia</td>
<td>Not clear what happened to Suvorovo later in the Balkans</td>
</tr>
<tr>
<td>No secure link to the steppes later</td>
<td>No secure link into Anatolia</td>
</tr>
</tbody>
</table>
Getting Anatolians right for the Anatolian theory

- Does not need to explain how Anatolian speakers got to Anatolia as a whole. But a story about movement within Anatolia is still needed.
- Before the farming migration into the Balkans (and long after it, too), Anatolia was not a single unified cultural area. Hittite and Luvian are in central Anatolia, while the population movement to the Balkans cannot be traced back to those exact locations.
- [Melchert, 2011]: if Anatolian speakers were in Anatolia all along, why aren’t there any Sumerian or Akkadian borrowings into Proto-Anatolian? (=common to different branches)
- Possible answer: evidence for significant contact with Mesopotamia in the 4th millennium is present only in the east [Sagona and Zimansky, 2009]. Wheel-made pottery makes it to the west of Anatolia only in mid-3rd mill. BC.

If Anatolian speakers were in the west, perhaps they were spared direct linguistic influences from Mesopotamia. But we then need late language spread events to the east...
Getting Anatolians right for the steppe theory

- Needs to trace the chain from Suvorovo-Novodanilovka and/or Sredni Stog in the late 5th millennium BC to Central and Eastern Anatolia in the 2nd millennium BC.
- No conspicuous chain emerges. Suvorovo → Cernavoda → Ezero → Troy I → further Anatolia?
- Ezero → Troy I?
  (cf. [Anthony, 2013])
- Ezero and the early Troy did have contacts, manifested by e.g. pottery, [Menkova, 2000]. But it’s not clear at all whether they are indicative of population exchange, or that they specifically favor the Ezero→Troy direction.
  Alleged imports in architecture were destroyed by a careful analysis of Troy in [Ivanova, 2013]: she shows Troy had autochthonous developments.
  ⇒ can’t be ruled out, but Ezero→Troy is a just-so story at the moment.
State of the evidence

- Both the steppe and the Anatolian theories are sensible.
- But both have serious explanatory gaps.
- To most linguists and archaeologists, those problems appear greater for Anatolia. But it is a matter of subjectively weighting the conflicting evidence.
- Phylogenetics:
  - May support one of the theories, but this won’t close all the explanatory gaps.
  ⇒ Cannot give a definitive answer on its own.
- Genetics:
  - May boost the plausibility of some population histories, closing some explanatory gaps.
  E.g. [Haak et al., 2015] show there was a massive population replacement event in Europe in the 3rd mill. BC — expected if the steppe hypothesis is true. But this does not disprove the Anatolian theory (contrary to Haak et al.’s claim.)
  ⇒ Cannot give a definitive answer on its own.
Phylogenetics: Historic and prehistoric constraints

- To derive age estimates, we need calibrations.
- Calibrations from historical sources are reasonably certain.
- For prehistory, the current analyses only use very broad constraints.

- **Our contribution**: archaeologically justified narrow prehistoric constraints.
  - In fact there is agreement between Anatolian and steppe theories’ proponents on the timing of some prehistoric events.

- Divergence of Indic and Iranian: $\approx 1900-1600$ BC (quite certain)
- Split of Tocharian from the “core IE”: $\approx 3300-3000$ BC (*terminus post quem*)

- **Finding**: predicted root age is highly dependent on the calibration constraints in the prehistoric part of the tree.
  - (Not very surprising given UCLN relaxed clock...
Indo-Iranian divergence

- Indic speakers are recorded in Syria in the 15th century BC (Mitanni), and in India very late in the 1st millennium BC. The steppes are filled with Iranian speakers in the 1st millennium BC.

(map from [Mallory, 1989])

- Both the steppe ([Anthony, 2007]) and Anatolian ([Renfrew, 1999]) theorists place Indo-Iranians in the European and/or West Asian steppes. [Kuz’mina, 2007] identifies the Andronovo cluster of cultures (2nd millennium BC) of northern Kazakhstan with the Indo-Iranians. Some Andronovan pastoralists, Indic speakers, move south into the Central Asia (≈1900-1800 BC), where pastoralists and autochthonous agriculturalists coexist (≈1800-1600 BC, dates by [Anthony, 2007]). (BUT: What happens next is unfortunately extremely unclear. Partly because scholarship is scarce.)
Not all archaeologists accept Andronovans as Indo-Iranians: [Lamberg-Karlovsky, 2002], [Frachetti, 2008].

But those who don’t do not accept any deep identifications of archaeological cultures and languages at the present stage of research. They’d be even greater skeptics about the Indo-European homeland.

Why be a skeptic? Scholars who defend Andronovo=Indo-Iranian:

- often talk exclusively in terms of migrations
- emphasize a greater cultural unity where in fact many distinct local variants exist

Skeptics emphasize diffusion of material culture through contact, and shifting contact configurations on the steppes.
They reject the notion that all Andronovans necessarily spoke the same language.
Indo-Iranian skeptics

Why not to be a skeptic:

- If we are concerned with language diffusion, it’s fine if Andronovans weren’t linguistically uniform. It suffices that some were Indo-Iranians.

- The steppes are the only place they could have come from according to both steppe and Anatolia theories.

- If they were in the steppes anyway, and must later go to India and Iran from there, the Andronovan scenario seems most likely:
  - in 15th cent. BC, an Indic Mitanni language is already observed, so the divergence must be earlier.
  - No earlier good candidates for a massive culture spread across the steppes before the Andronovans.

Dating Indo-Iranian divergence to 1900-1600 BC or closeby is fairly certain.
Tocharian split

- Historical Tocharians lived in Xinjiang (Western China) in mid-1st mill. CE.

- Tocharian languages are a good candidate for the second IE split (to happen after the first split, by Anatolian languages).
- Not established when and how exactly Tocharians got to Xinjiang (＝Tarim Basin).
- Similarly for their split from IE.
Tocharian split

- For the steppe theory, Tocharians must have separated (eastward) from the IE speakers in the Pontic steppes.
- *Terminus post quem* for that: the creation of the Afanasyevo culture in the Altai, with material similarities to Repin and Yamnaya cultures on the Volga.
- With recent radiocarbon dates, the Afanasyevo migration is dated to 3300–3000 BC [Anthony, 2013].

![Map of Afanasyevo culture migration](image)

- Tocharians could have split *later* than that. But *hardly earlier*. Choosing this earliest possible split favors the Anatolian theory.
Tocharian split

- For Renfrew, Tocharians separate from the ‘Old Europe’ cultural area in the Balkans into the Pontic steppes. They get to Asia at some unspecified later.

- Initially Renfrew suggested an early separation into the steppes. But there is no support for that.

- Cucuteni-Tripolye and the steppe cultures did interact, and eventually the former disappeared.

  One story is that Cucuteni-Tripolye was suppressed, and steppe cultures took its habitat. Another story is that Cucuteni-Tripolye people colonized the steppes, [Manzura, 2005].

- It is the second story that is compatible with Renfrew’s Anatolian theory. According to Manzura, such colonization would have happened in the second half of the 4th mill. BC.
  
  ≈ the Afanasyevo migration at the end of the 4th mill. BC

  ⇒ we can use the same Tocharian constraint to cover both cases.
Our results

- We did a replica of [Bouckaert et al., 2012] with and without constraints.

- Indo-Iranian divergence: \( \approx 1900-1600 \) BC

- Tocharian split from the rest: \( \approx 3300-3000 \) BC

- We only considered root age estimates, discarding geographical predictions.
  
  On the one hand, we believe geography may be useful even in this model, as modern related languages are often not far from each other.

  On the other, currently geographical inferences about the past are completely unreliable. E.g., [Bouckaert et al., 2012] fail to predict Iranians in the Kazakhstan steppes where they lived in 1st mill. BC–1st mill. CE.

- Replica results: median 5569 BC, \( BF_{\text{steppe/Anatolia}} \approx 1.0 \)

- Results with new constraints: median 4986 BC, \( BF_{\text{steppe/Anatolia}} \approx 4.0 \)
Prior probabilities were not obviously affected by the constraints, while the posterior was.
Comparison to other analyses

Results by [Chang et al., 2015]

- [Chang et al., 2015] introduce ancestry constraints (recall Remco’s talk) and other innovations, which drive estimates down ca. half a millennium.

- Chang et al. claim strong support for the steppe. But they use a dataset with 2 out of 3 Anatolian languages excluded. This alone drives the age down by a millennium.

- Now, those languages, Luvian and Lycian, have very sparse data. But simply throwing them away makes it too easy for the steppe to win. The internal evidence from the Anatolian branch is very important as it is directly adjacent to the root.

⇒ a different method from ours also drives age estimates down considerably. But this is not decisive on its own.
Remco presented new results on Monday, with a different model. Without the new constraints, the peak density was around 6000 BC — ideal for the new dates of the Anatolian theory. With our constraints, the peak went down ca. half a millennium, just as in our replica.

It is not clear what phylogenetics would show in the end.

But it makes sense to improve our methods — and our datasets. In particular, adding prehistoric data can affect age estimates greatly. This means we have a lot of work to do for the prehistoric parts!
*The Horse, the Wheel, and Language. How Bronze Age Riders from the Eurasian Steppes Shaped the Modern World.*  
Princeton University Press.

*Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism.*  
*Journal of Language Relationship, 9*(1-21).

Mapping the origins and expansion of the Indo-European language family.  

Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis.  
*Language*.

*Pastoralist landscapes and social interaction in Bronze Age Eurasia.*  
University of California Press, Berkeley.

Language-tree divergence times support the Anatolian theory of Indo-European origin.  

Massive migration from the steppe was a source for Indo-European languages in Europe.  
*Nature, 522*:207–211.

*Yanovich et al. (Uni Tübingen & CMU) Archaeological temporal constraints on IE*
Domestic architecture in the Early Bronze Age of western Anatolia: the row-houses of Troy I.  
_Anatolian Studies_, 63:17–33.

On the biological distinctness of the Pit-Grave (Yamnaya) people in the Northwestern Caspian: cranial evidence.  
_Archaeology Ethnology and Anthropology of Eurasia_, 42(1):142–150.

_The Origin of the Indo-Iranians_.  
Brill.

Archaeology and language. the Indo-Iranians.  

_In Search of the Indo-Europeans: Language Archaeology and Myth_.  
Thames and Hudson.

Steps to the steppe: or, how the North Pontic region was colonised.  

Indo-europeans.  

Time and space in Ezero culture interactions: the Early Bronze Age.  

Westward expansion of the Neolithic way of life: sorting the Neolithic package into distinct packages.
Archaeological evidence on the westward expansion of farming communities from Eastern Anatolia to the Aegean and the Balkans. 
*Current Anthropology*, 52:S415–S430.

A new look at the introduction of the Neolithic way of life in Southeastern Europe. changing paradigms of the expansion of the Neolithic way of life. 
*Documenta Praehistorica*, 41:33–50.

From the Near East to Greece: let’s reverse the focus. cultural elements that didn’t transfer. 

*Naselenie Ukrainy v epoxi neolita i rannego eneolita po dannym antropologii.* 
Naukova Dumka, Kyiv.

The Eneolithic of the Black Sea steppe: dynamics of cultural and economic development 4500–2300 BC. 
In *Late prehistoric exploitation of the Eurasian steppe*, chapter 3, pages 59–182. McDonald Institute for Archaeological Research.

Time depth, convergence theory, and innovation in Proto-Indo-European: ‘Old Europe’ as a PIE linguistic area. 
*Journal of Indo-European Studies*, 27(3-4):257–293.

Ryder, R. J. and Nicholls, G. K. (2011).
Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. 