Modelling sound change in relation to time-depth and geography: a case study on the Indo-European and Tupían language families

GERD CARLING, SANDRA CRONHAMN, NIKLAS JOHANSSON, JOOST VAN DE WEIJER

SCIHS, UC Berkeley, May 31, 2014

Why analyse sound change?

- Many current computational cladistic analyses are based on the method of **basic vocabulary comparison** (BVC, see Hammarström 2014, 60f.) or **typology** (e.g., Muysken & O'Connor 2014, 1ff.).
- Traditional comparative method rests on three legs (e.g., Meillet 1924, 11; Campbell 2014, 107ff.).
  - **Vocabulary**
  - **Phonology** (Morphonology)
  - **Morphology** (Morphosyntax)
- Conventional subgrouping (tree-models, wave-models, combinations) based on common innovations or isoglosses (Gamkrelidze & Ivanov 1984, Garrett 2006, Josephson 2013)

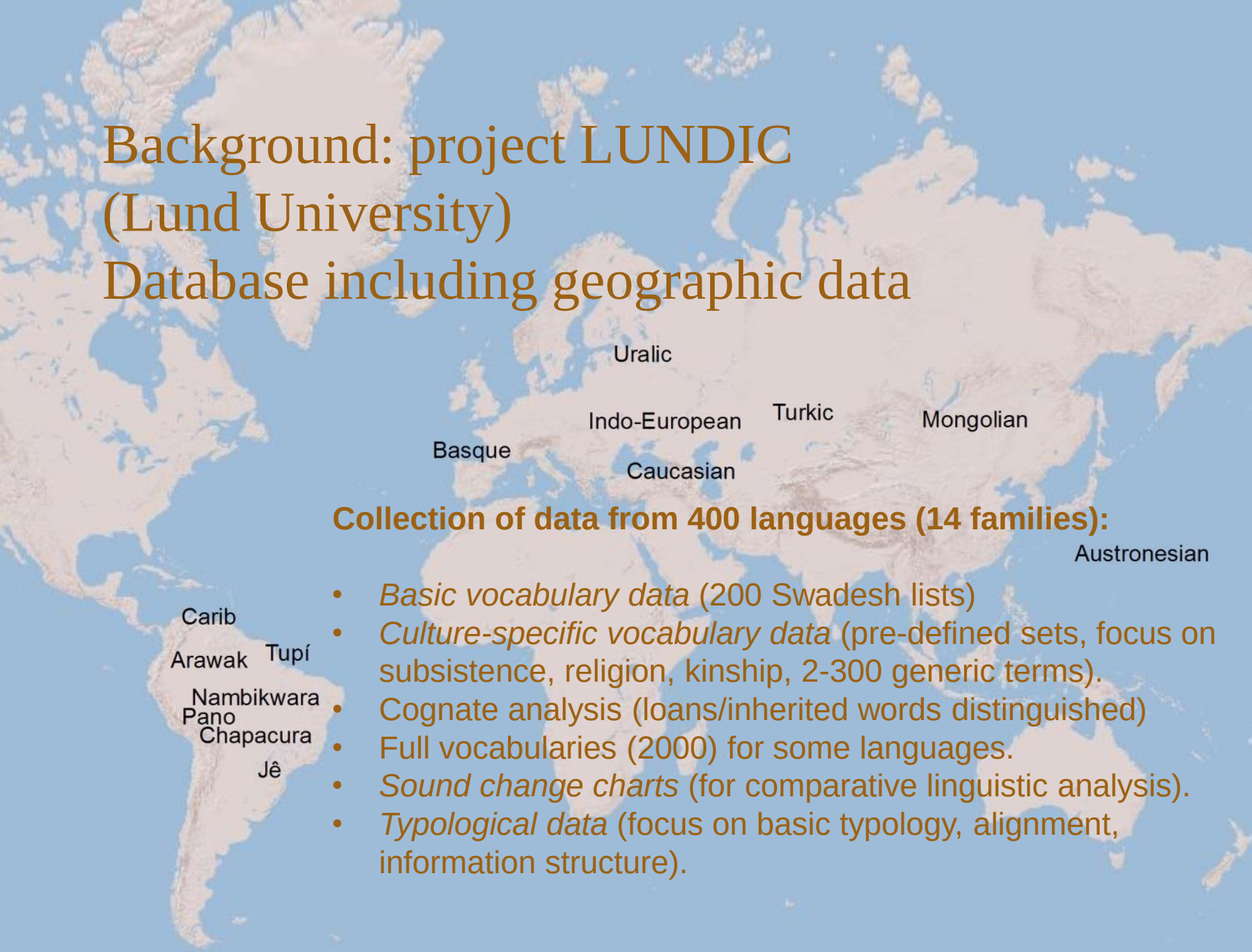
LUND UNIVERSITY

Why Indo-European and Tupí?

- Indo-European language family
  = sound change well attested and investigated (Kümmel 2007)

- Tupí (like all South American languages families)
  = very little data
  - No ancient sources
  - Sound change → not well investigated
  - Vocabularies → sparse

Background: project LUNDIC (Lund University) Database including geographic data



Collection of data from 400 languages (14 families):

- *Basic vocabulary data* (200 Swadesh lists)
- *Culture-specific vocabulary data* (pre-defined sets, focus on subsistence, religion, kinship, 2-300 generic terms).
- Cognate analysis (loans/inherited words distinguished)
- Full vocabularies (2000) for some languages.
- *Sound change charts* (for comparative linguistic analysis).
- *Typological data* (focus on basic typology, alignment, information structure).

Basic questions

- How does sound change relate to:
  - Language diversity and change
  - Time depth
  - Geographic spread

- Aims:
  1) Testing quantification of sound change against basic vocabulary
  2) Testing the results against time depth and geography

Tools

- For clustering and subgrouping:

Program	Analysis	Description
R	Biplots (Principal Component Analysis)	Demonstrates the amount of variance of samples based on a relative similarity between objects
MrBayes	Cladograms	Creates cladistic trees based on a hierarchy of individual elements

- For geographic analysis:

Program	Analysis	Description
ArcGIS	(Layered) maps	Demonstrates the geographic spread of variance and clusters

LUND UNIVERSITY

Theoretical framework: sound change, time depth and geography

- Critical changes are **parallel** and **independent** (Meillet 1924:380).

- The reconstructed state is a **linguistic structural unity**, which we should expect to correspond to a linguistic unity reminiscent of a spoken language.

- The connection of a reconstructed unity to an exact time and place (”nation indo-européenne”) in history must remain uncertain (Meillet 1924: 375).

LUND UNIVERSITY

Basic model of quantification, sound change

- Sound change charts, model of, e.g., Meillet (1924), Krahe (1966).

- Focus on critical and common innovations, as derived from comparative method.

- Reconstructed phoneme systems used as a basis.

LUND UNIVERSITY

Model of quantification, sound change

	*ū	ū-ȳ/[i-mut]	ū-ȳ/[i-mut]-i:	ū-ʉ:	ū-aʊ	ū-i:	ū-i:-ɪ
PIE	1						
Proto-Germanic	1						
Old Norse	1	1					
Swedish	1	1		1			
Old English	1	1					
English	1	1	1		1		
Gothic	1						
Proto-Celtic	1						
Old Irish	1						
Irish	1						
Old Welsh	1					1	
Welsh	1					1	1
Gaulish	1						
Proto-Italic	1						
Latin	1						
Old Italian	1						
Italian	1						
Old French	1						
French	1						
Oscan	1						

Our analysis

I.-E.	SKR.	ZD	V. PERSE.	V. SL.	LIT.	ARM.	GR.	LAT.	IBL.	V. H. A.
*ei	e¹	aē, ōi	ai	i	ē², ei	(ə)	ει	ī³	ē, īa	ī
*eu	o¹	ao, əu	au	ju	iaū	oy	ευ	ū³	ō, ūa	eo, iu
*er	ar	ar	ar	rə⁵	ēr	er⁶	ερ	er	er	er
*el	ar¹	ar	ar	lə⁵	ēl	el	ελ	ul	el	el
*en	an	an	a(n)	ę	eñ	in	εν	en	(en)	in
*em	am	əm	am	ę	eñ	im	εμ	em	(em)	im
*oi	e¹	aē, ōi	ai	ə(i)⁴	ē², ai	e	οι	ū³	oe	ai, ei, ē
*ou	o¹	ao, əu	au	u	aū	oy	ου	ū³	ō, ūa	au, ou, ō
*or	ar	ar	ar	ra⁵	ar	or⁶	ορ	or	or	ar
*ol	ar¹	ar	ar	la⁵	al	ol	ολ	ul	ol	al
*on	an	an	a(n)	ǫ	añ	un	ον	on	(on)	an
*om	am	əm	am	ǫ	am	um	ομ	um	(om)	am
*ai	e¹	aē, ōi	ai	ə(i)⁴	ē², ai	ay	αι	ae	ae	ai, ei, ē
*au	o¹	ao, əu	au	u	au	aw	αυ	au	ō, ūa	au, ou, ō
*ar	ar	ar	ar	ra⁵	ar	ar⁶	αρ	ar	ar	ar
*al	ar¹	ar	ar	la⁵	al	al	αλ	al	al	al
*an	an	an	a(n)	ǫ	an	an	αν	an	an	an
*am	am	əm	am	ǫ	am	am	αμ	am	am	am

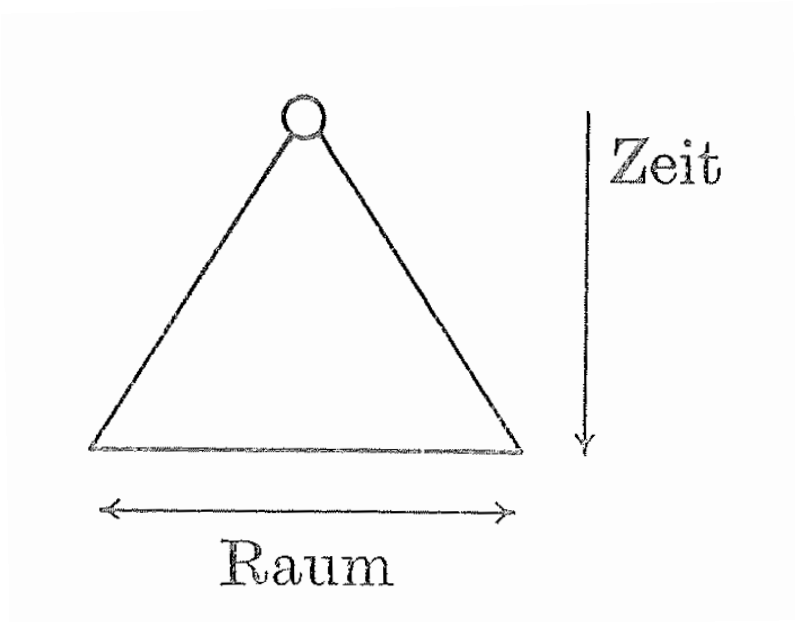
Notes :

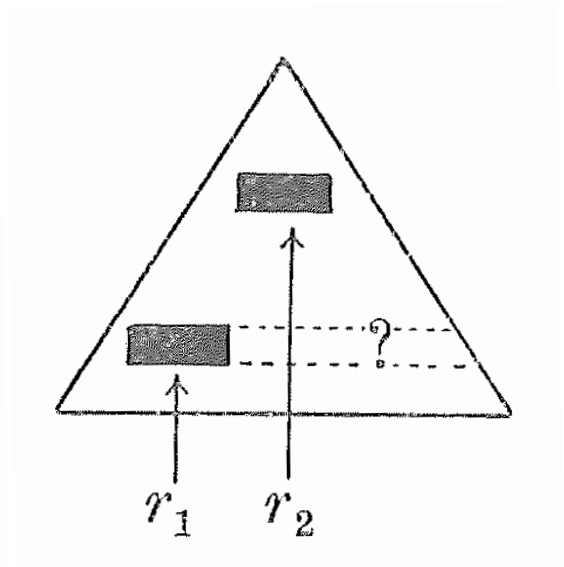
1° Skr. *e* et *o* sont des longues issues d'anciennes diphtongues indo-iraniennes *ai*, *au* conservées en vieux perse ; le fait qu'elles représentent des diphtongues est reconnaissable en sanskrit même et a été vu par les grammairiens indigènes. — Les diphtongues

Meillet (1924)

Model of quantification: time and space

- Language states (including reconstructed) quantified as layers, connected to time periods and coordinates.
- The greater time-depth, the more unstable the basis for a reconstruction, both in time and space (space-time model of Meid 1975).



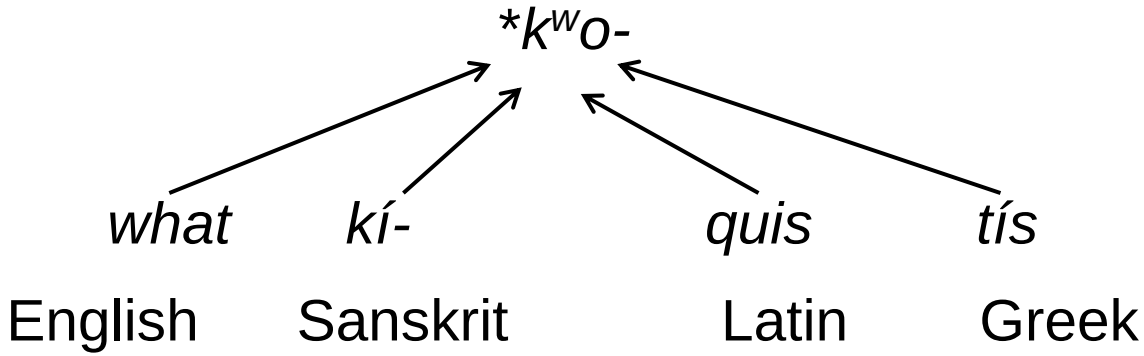


LUND UNIVERSITY

Methodology: sound change data

- Point of departure: established reconstructed phoneme inventory of Proto-Indo-European (5000-8000 BP) and Proto-Tupí (5000-6000 BP), e.g. (simplified),

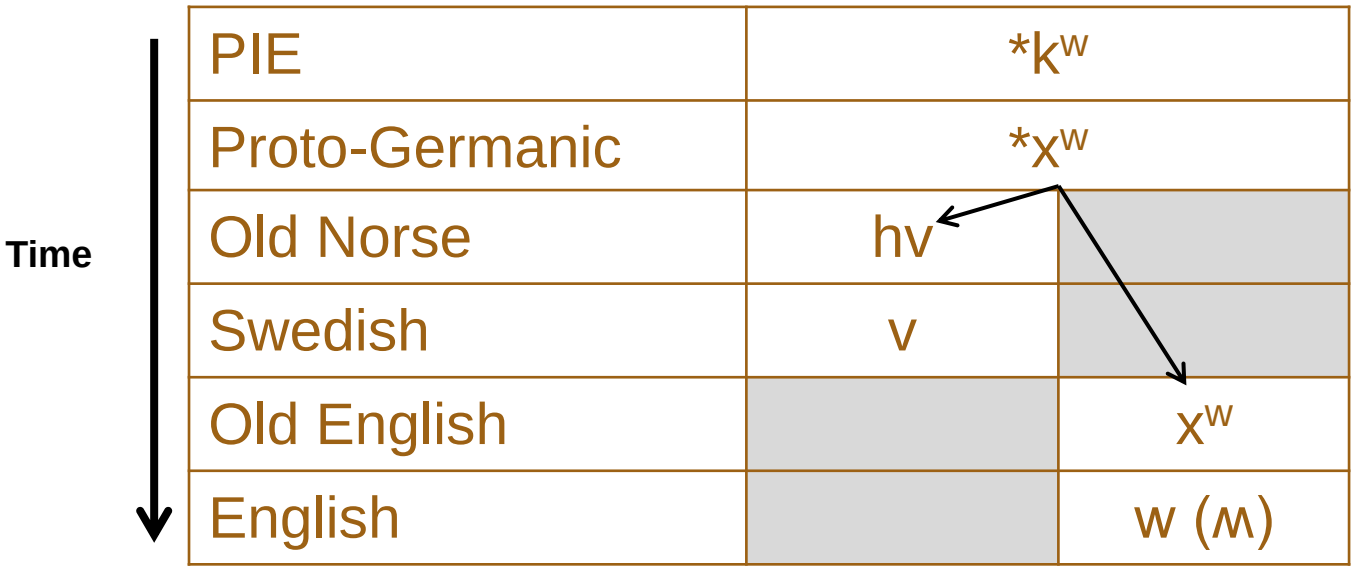
Proto-Indo-European (PIE) *k^w in *$k^w o$- 'what'



LUND UNIVERSITY

Methodology: sound change data

- The evolution of the reconstructed phonemes is followed to the reconstructed, historically attested and contemporary language states, e.g. (simplified),



Methodology

- Each change is quantified by value 1
- Earlier stages of change are inherited by their descendants, e.g., Modern English has inherited all values from Old English, Proto-Germanic, etc.
  → Indirect relative chronology
- **kʷo* → Eng. ***wh****at* & Swe. ***v****ad*
  → Hittite ***ku****i-*

Innovations

PIE	*kʷ	*kʷ→*xʷ	*kʷ→*xʷ→hv	*kʷ→*xʷ→hv→v	*kʷ→*xʷ→w
Proto-Germanic	1	1	0	0	0
-Old Norse	1	1	1	0	0
-Swedish	1	1	1	1	0
-Old English	1	1	0	0	0
-English	1	1	0	0	1
Proto-Anatolian	1	0	0	0	0
-Hittite	1	0	0	0	0

Data availability and comparability

	Indo-European	Tupí
Primary branches	~10	~10
Reconstructed languages	12 ⟷	19
Historical languages	18 ⟷	3
Contemporary languages	13 ⟷	40
Total languages	**43**	**62**

LUND UNIVERSITY

Selection of changes in current set

- **Indo-European** (sound changes well established) → Focus on critical innovations
  - Unconditioned changes:
    - » Regular sound changes
    - » Complete phoneme deletions
  - Conditioned changes: selection of changes generally affecting more than one language and/or more than one phoneme
    - » If no unconditioned change exists for a phoneme → all conditioned changes included
    - » Relevant sound laws; Grimm's, Osthoff's, Ruki-rule etc.
    - » Important conditioned changes, such as palatalizations, velarizations, labializations

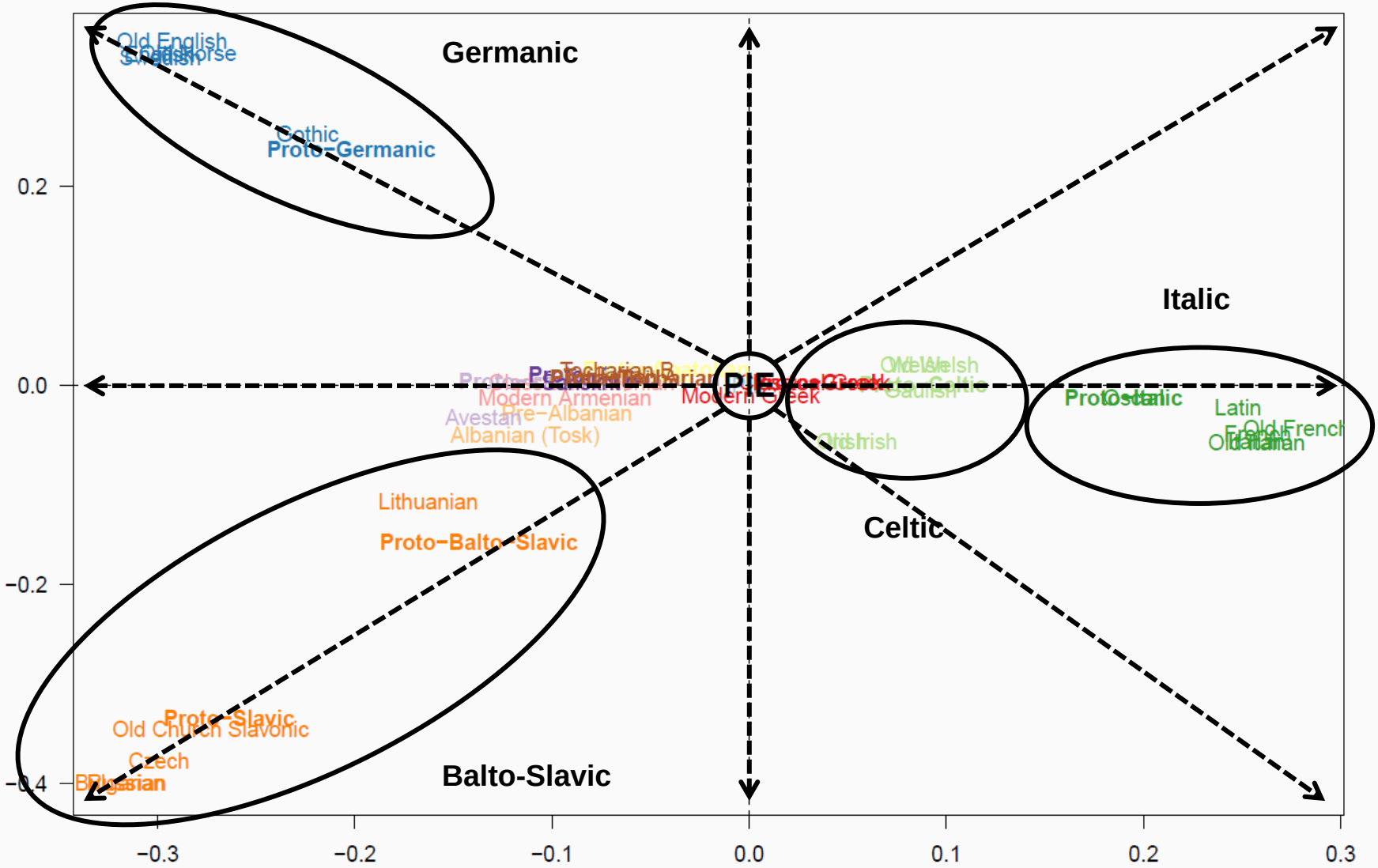
Selection of changes in current set

- **Tupí** (sound changes less well established)
  - Based on the reconstructed phonological system by Rodrigues & Cabral (2012)
  - Sparse information about sound changes → everything is included
  - Different branches are differently well-documented → possible effect on the results

- **Total number of changes in data sets:**
  - Tupí 237
  - Indo-European 891

LUND UNIVERSITY

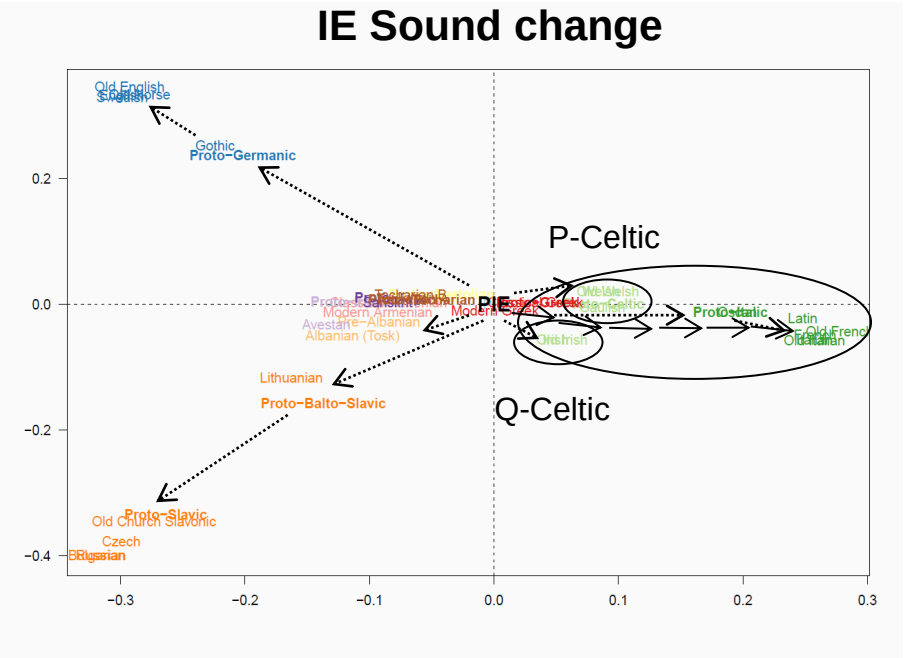
Biplot: Indo-European sound change

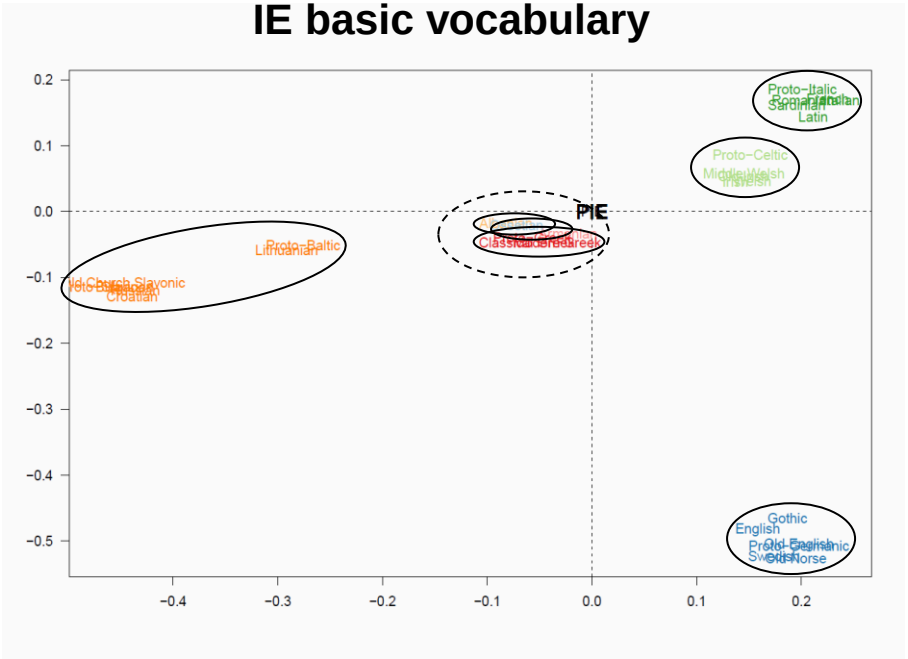


Numbers: relative distance to 0.0 = Proto-Indo-European (all sounds value 1)
Colours: traditional subgroupings

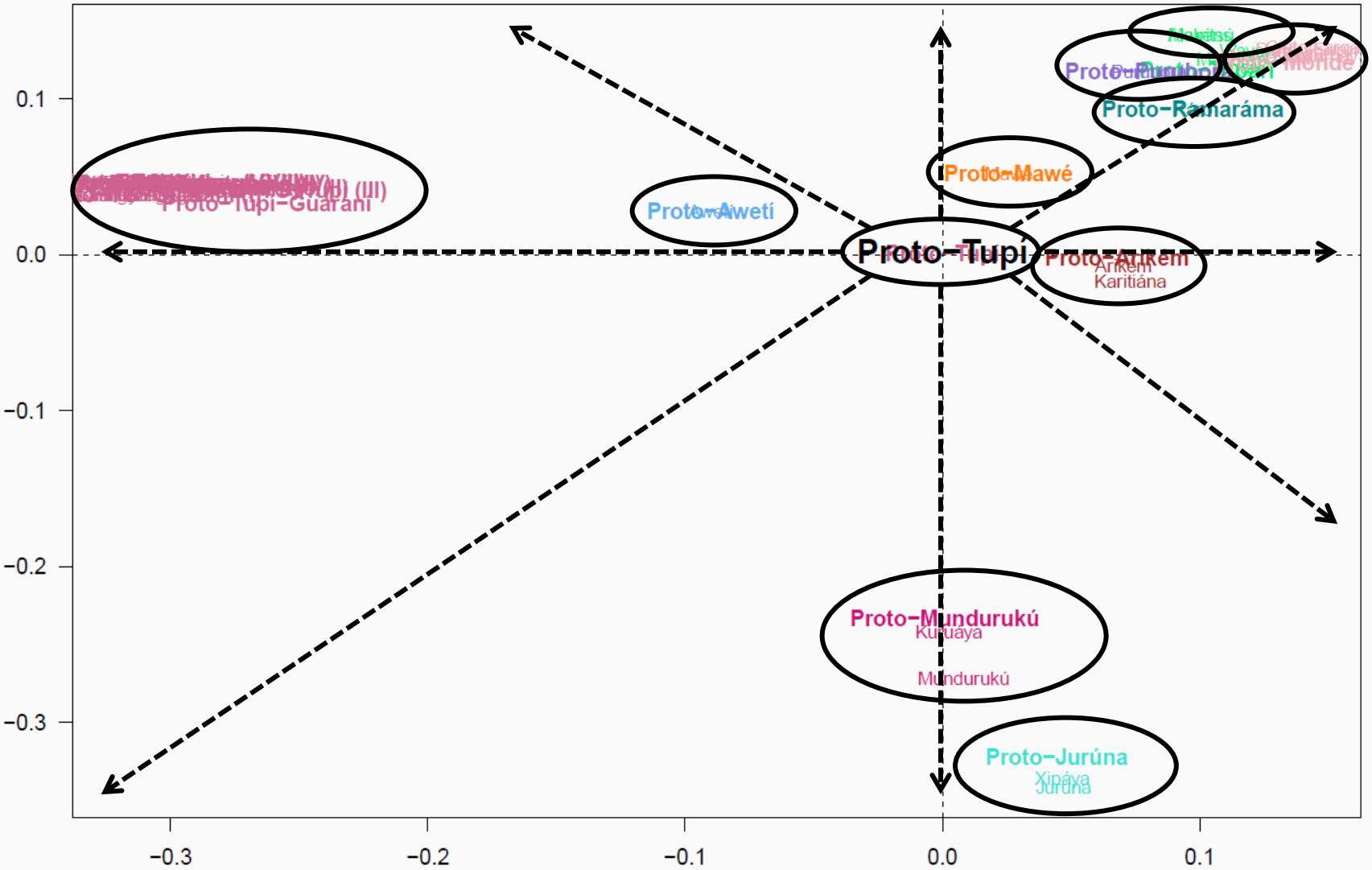
Comparison: sound change vs. basic vocabulary (Indo-European)

- Basic vocabulary → clustered subgrouping
- Sound change → gradient subgrouping





Biplot: Tupí sound change

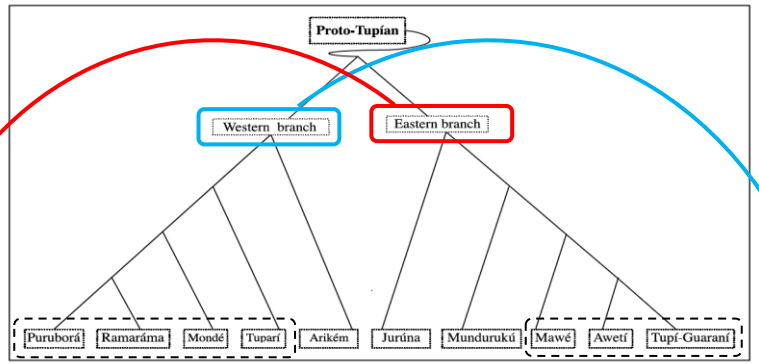


Numbers: relative distance to 0.0 = Proto-Tupí (all sounds value 1)
Colours: traditional subgroupings

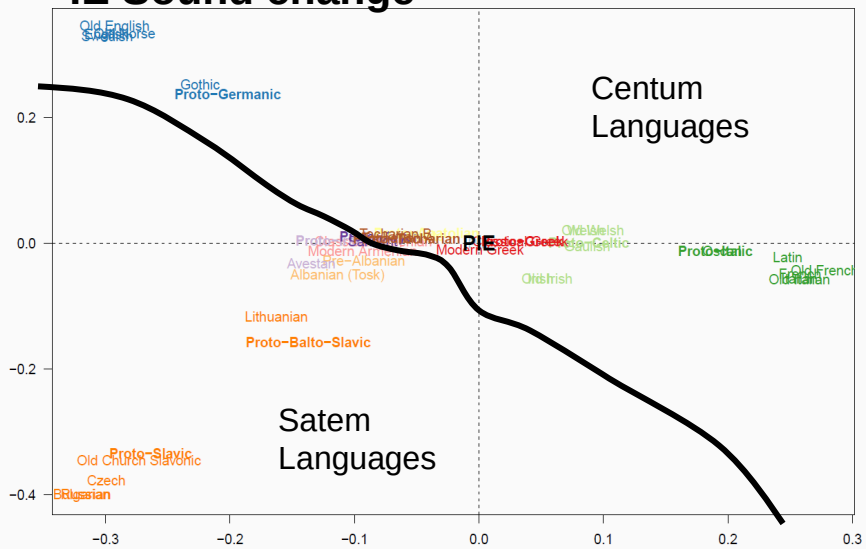
Comparison: Indo-European vs. Tupí sound change

- Indo-European = more gradient
- Tupí = more clustered
- Possibly due to the difference in data granularity
- Matches traditional subgrouping

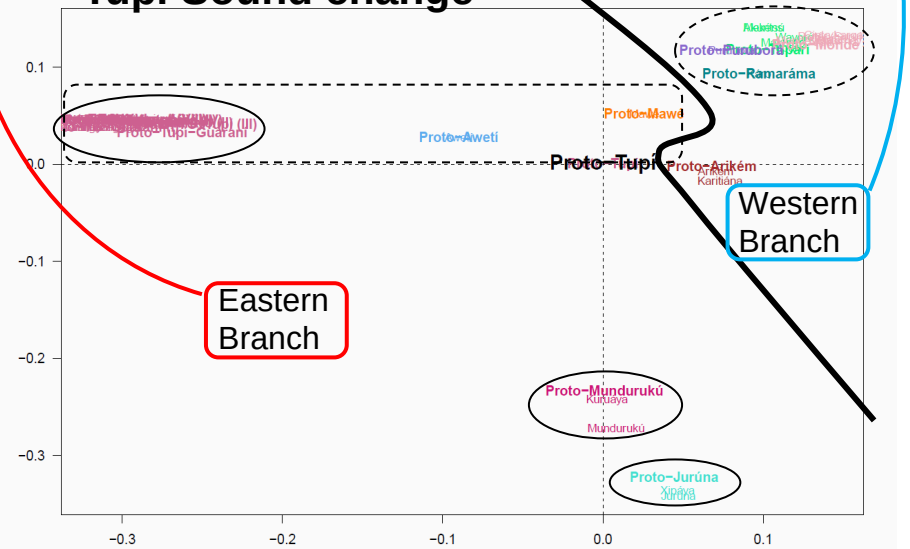
Figure 1 Family tree of the Tupian linguistic stock



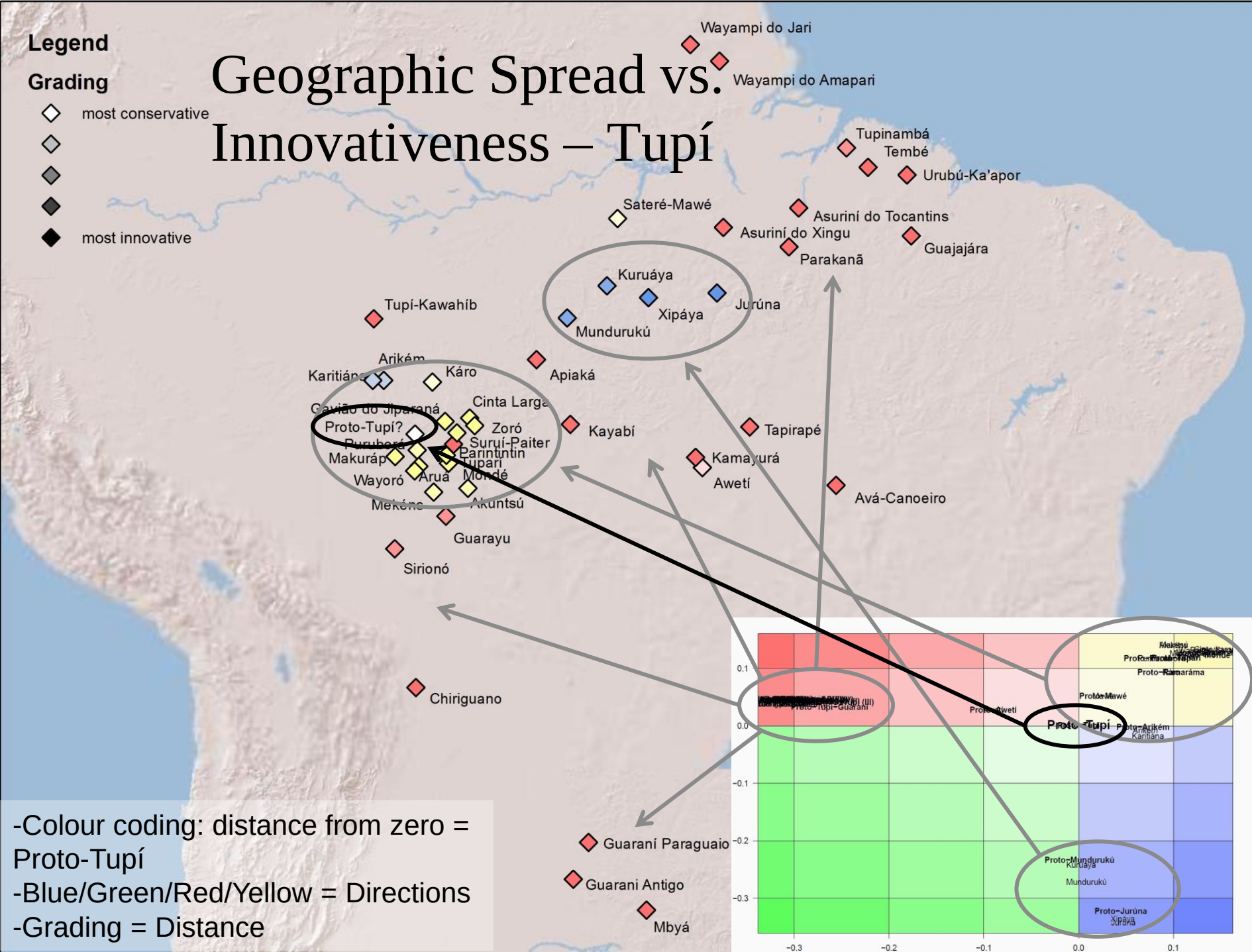
IE Sound change



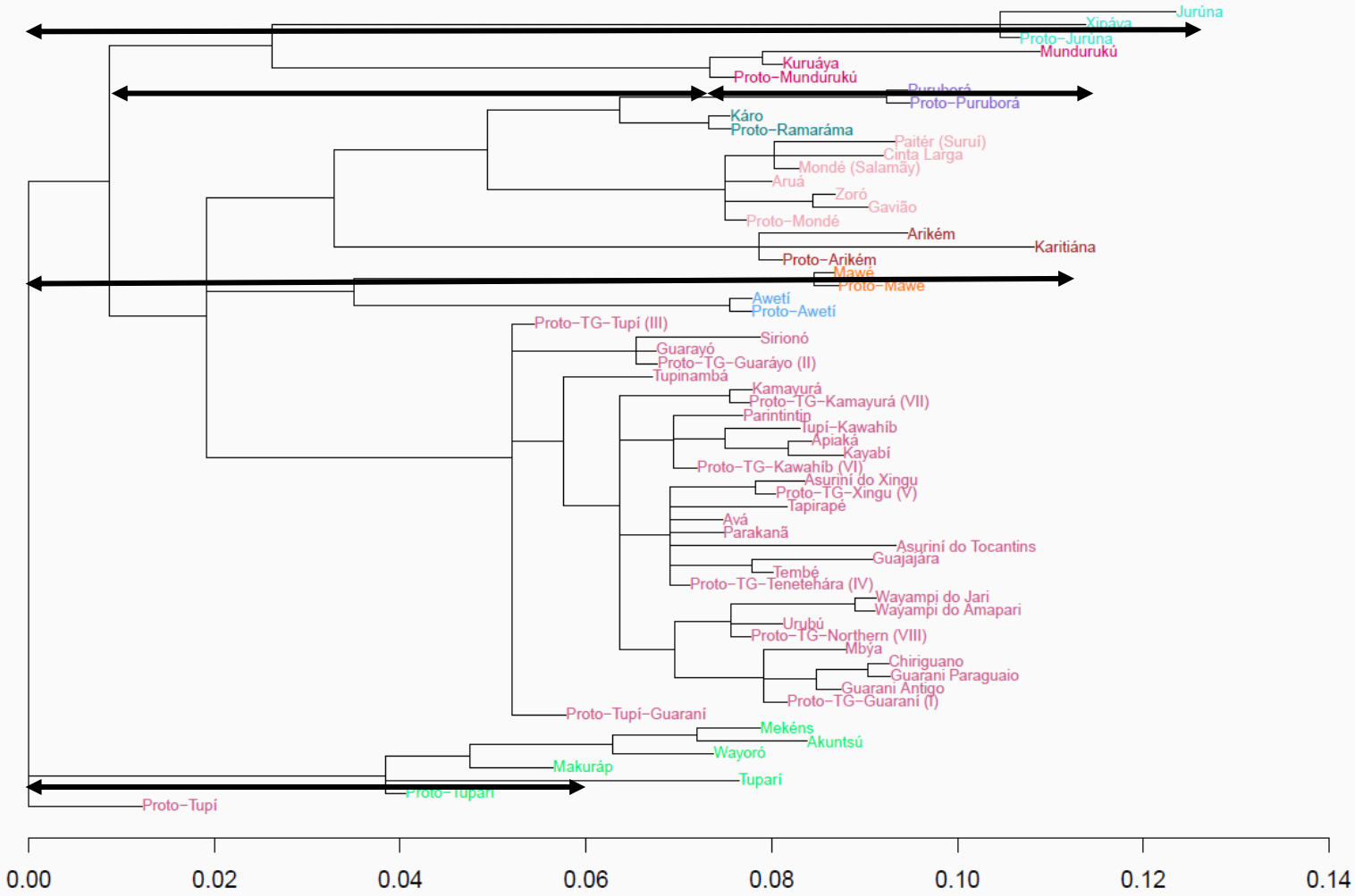
Tupí Sound change



Geographic Spread vs. Innovativeness – Tupí

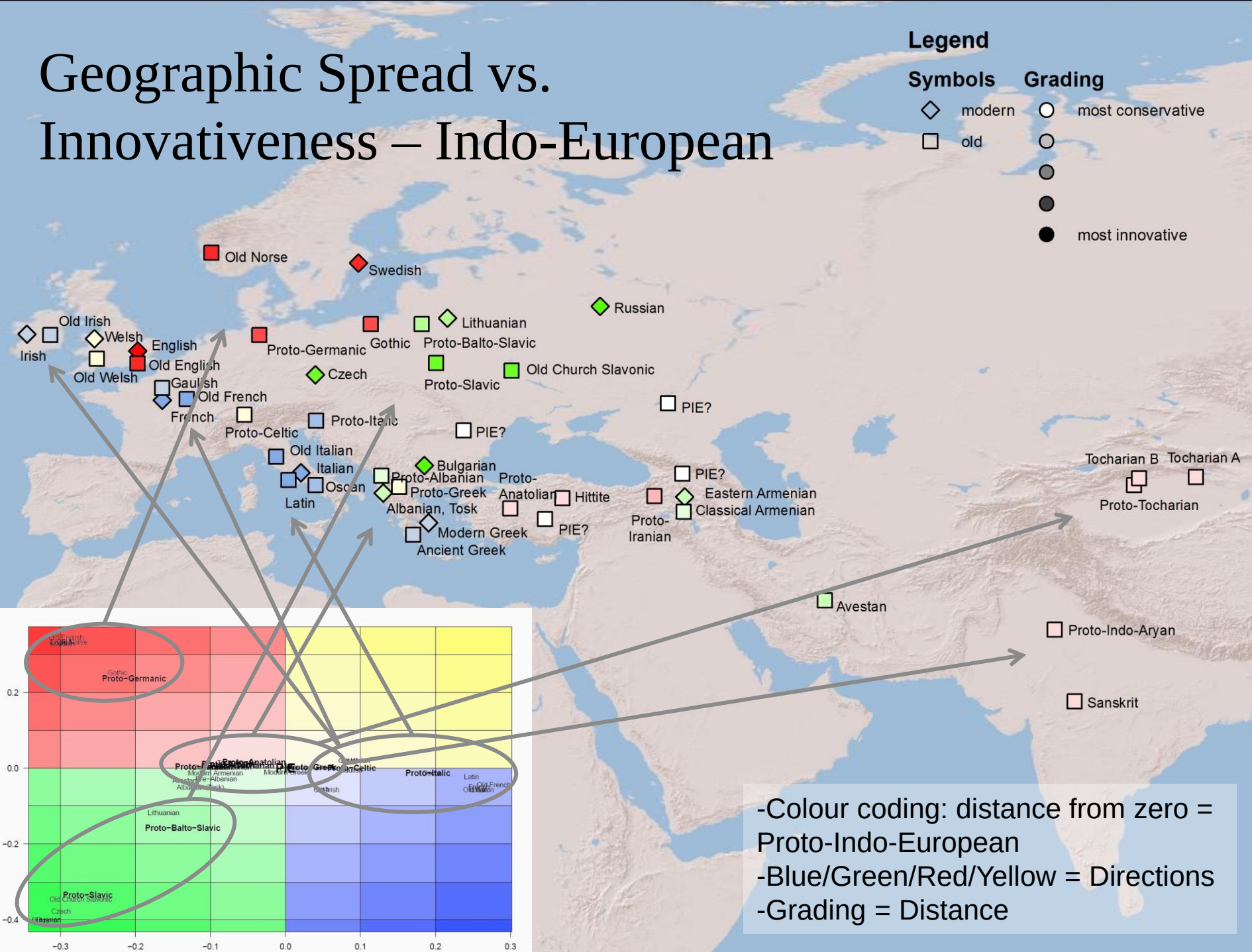


Cladogram: Tupí



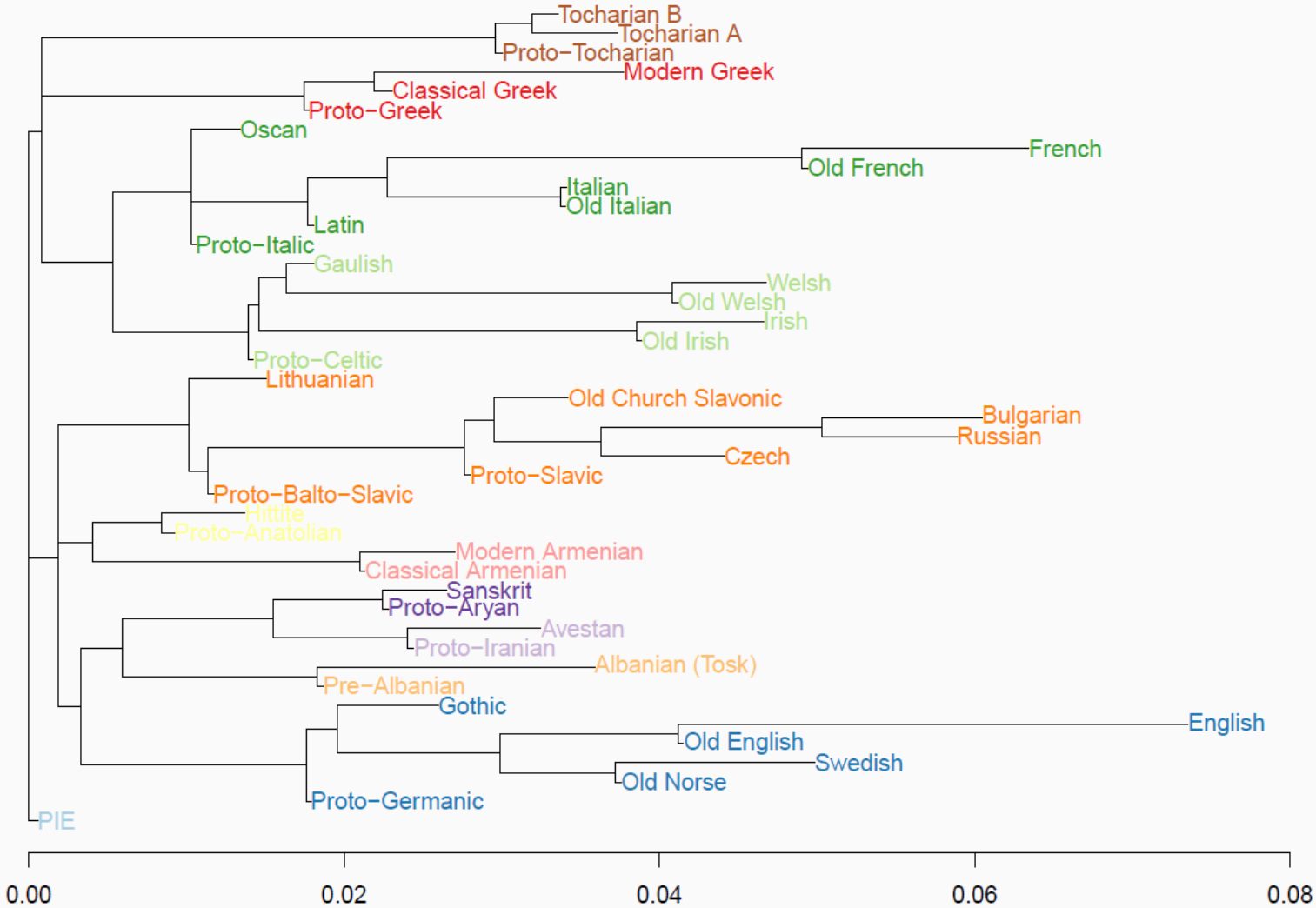
Scale: relative distance from zero Colour coding: traditional subgrouping

Geographic Spread vs. Innovativeness – Indo-European



-Colour coding: distance from zero = Proto-Indo-European
-Blue/Green/Red/Yellow = Directions
-Grading = Distance

Cladogram: Indo-European



Scale: relative distance from zero Colour coding: traditional subgrouping

Conclusion, method in general

Advantages

- Results correspond to conventional subgroupings.
- Methodology embraces common innovations (independent change) as well as shared similarities (parallel evolution)
- Methodology suitable for measuring innovativeness (branch length)

Limitations

- Requires a solid preparatory work of comparative linguistic analysis (disadvantage for undescribed language families)
- Data granularity plays some role, but not for the overall results

LUND UNIVERSITY

Results and perspectives

Results:

- Phylogenetic models should be based on mixed data sets.
- Areal effects (with results from both trees and biplots), both in subgrouping and linguistic distance
- Supports a geographical diffusion model, related to convergence.

LUND UNIVERSITY

Results and perspectives

Perspectives:

- Can possibly be useful for the discussion on proto-language time depth?
- Data sets can be used for, e.g., measuring directionality of sound change?

Thank You!

Funding from:

Marcus and Amalia Wallenberg Foundation

Crafoord Foundation

Special thanks to:

Anne Goergens for creating maps

Love Eriksen for providing geographic data for Tupí

Ana Suelly Cabral, Niklas Burenhult, Rob Verhoeven & Filip Larsson for valuable remarks

LUND UNIVERSITY

Key References

- Carling, G., Eriksen, L., Holmer, A. & van de Weijer, J. (*2013*). Contrasting linguistic and archaeology in the matrix model: GIS and cluster analysis of the Arawakan languages. In Saxena, A. & Borin, L. (eds.). *Comparing Approaches to Measuring Linguistic Differences*. Berlin: Mouton de Gruyter, 29-58.
- Josephson, Folke (2013). How aberrant are the divergent Indo-European subgroups? In Saxena, A. & Borin, L. (eds.). *Comparing Approaches to Measuring Linguistic Differences*. Berlin: Mouton de Gruyter, 83-106.
- Meid, Wolfgang, 1975. *Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen*. In: H. Rix (ed.), *Flexion und Wortbildung.* Wiesbaden: Reichert, 204-219
- Meillet, Antoine (1924). Introduction à l'étude comparative des langues indo-européennes. Paris: Libraire Hachette

LUND UNIVERSITY

Primary Sources

Indo-European

Baldi, Philip (1983). *An introduction to the Indo-European languages: y Philip Baldi*. Carbondale: Southern Illinois univ. Press

Ball, Martin John & Fife, James (red.) (1993). *The Celtic languages*. London: Routledge

Beekes, Robert S. P. (1995). *Comparative Indo-European linguistics: an introduction*. Philadelphia: Benjamins

Clackson, James (2007). *Indo-European linguistics: an introduction*. Cambridge, UK: Cambridge University Press

Krahe, Hans & Meid, Wolfgang (1966). *Germanische Sprachwissenschaft. 1, Einleitung und Lautlehre*. Berlin:

Kümmel, Martin J. (2007). *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert

Lewis, Henry & Pedersen, Holger (1937). *A concise comparative Celtic grammar*. Göttingen: Vandenhoeck & Ruprecht

Melchert, H. Craig (1994). *Anatolian historical phonology*. Amsterdam: Rodopi

Palmer, Leonard Robert (1954). *The Latin language*. London: Faber and Faber

Pinault, Georges-Jean (1992). Introduction au Tokharien; in LALIES: *Actes des sessions de linguistique et de littérature,* 11; Presses de L'École Normale Supérieure, Paris

Pokorny, Julius (1994). *Indogermanisches etymologisches Wörterbuch*. 3. Aufl. Tübingen: Francke

Prokosch, Eduard (1939). *A comparative Germanic grammar*. Philadelphia, Penn.

Orel, Vladimir E. (2000). *A concise historical grammar of the Albanian language: reconstruction of Proto-Albanian*. Boston: Brill

Szemerényi, Oswald (1980). *Einführung in die vergleichende Sprachwissenschaft*. 2. überarb. Aufl. Darmstadt: Wissenschaftl. Buchgesellschaft

Sussex, Roland & Cubberly, Paul (2006). *The Slavic languages*. Cambridge: Cambridge University Press

Willis, D (2009). Old and Middle Welsh. In: *The Celtic languages*, edited by Martin Ball and Nicole Müller. London: Routledge, 117-60.

Tupí

Dietrich, W. (2013). Correspondências fonológicas e lexicais entre Karitiána (Arikém, Tupí) e Tupí-Guaraní. In Rodrigues, A. D. I. & Cabral, A. S. A. C. (eds.), *Revista Brasileira de Linguística Antropológica*,*1* (2), 191-214. Brasília: Laboratório de Línguas Indígenas, Instituto de Letras, Universidade de Brasília.

Gabas Jr, N. (1997). Genetic relationship within the Ramaráma family of the Tupí stock (Brazil). In *Ensayos sobre lenguas indígenas de las tierras bajas de Sudamérica. Contribuciones al 49º Congreso Internacional de Americanistas en Quito* (71-82).

Hanke, W., Swadesh, M., & Rodrigues, A. D. I. (2013). Notas de fonologia Mekens. In Rodrigues, A. D. I. & Cabral, A. S. A. C. (eds.), *Revista Brasileira de Linguística Antropológica*, *3*(2), 173-196. Brasília: Laboratório de Línguas Indígenas, Instituto de Letras, Universidade de Brasília.

Mello, A. A. S. (2000). *Estudo histórico da família lingüística Tupí-Guaraní: aspectos fonológicos e lexicais.* PhD dissertation. Florianópolis: Universidade Federal de Santa Catarina.

Rodrigues, A. D. I. (2011). Relações internas na família linguística Tupí-Guaraní. In Rodrigues, A. D. I. & Cabral, A. S. A. C. (eds.), *Revista Brasileira de Linguística Antropológica*, *3* (2), 233-252. Brasília: Laboratório de Línguas Indígenas, Instituto de Letras, Universidade de Brasília.

Rodrigues, A. D. I. (2011). Tupí-Guaraní e Mundurukú: evidências lexicais e fonológicas de parentesco genético. In Rodrigues, A. D. I. & Cabral, A. S. A. C. (eds.), *Revista Brasileira de Linguística Antropológica*, *3*(2), 153-165. Brasília: Laboratório de Línguas Indígenas, Instituto de Letras, Universidade de Brasília.

Rodrigues, A. D. I. & Cabral, A. S. A. C. (2012). Tupían. In Campbell, L. & Grondona, V. (eds.), *The indigenous languages of South America*. 59-166, Berlin: De Gruyter Mouton.

LUND UNIVERSITY

LUND
UNIVERSITY