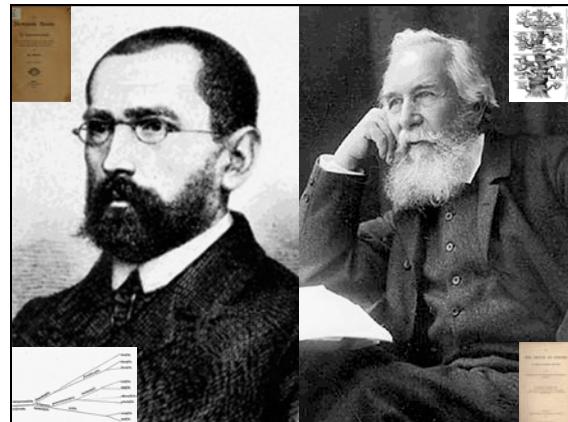


The big, bright future of cultural evolution and linguistics



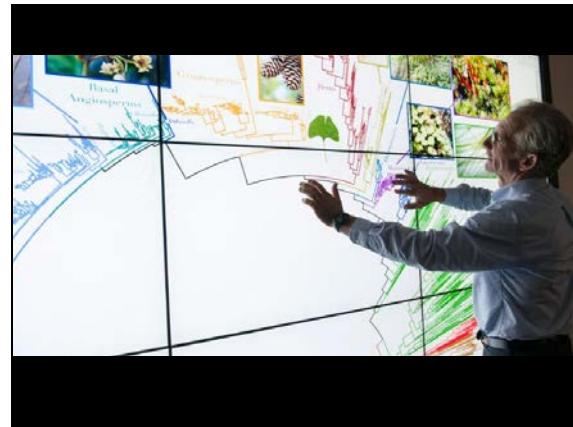
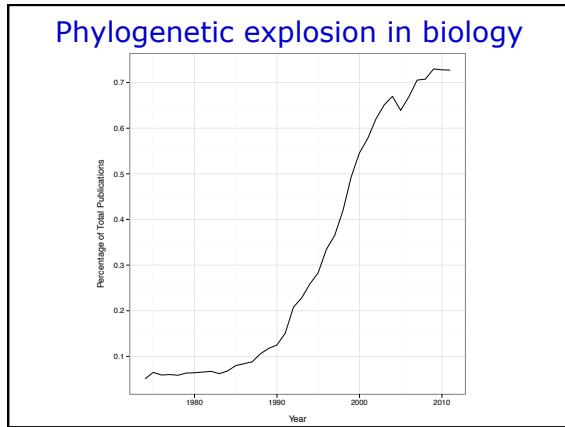
Max Planck Institute for the Science of Human History



Plan – scaling up...think big!

1. Big data
2. Better methods
3. Big questions
4. Big interdisciplinary teams





Linguistics today

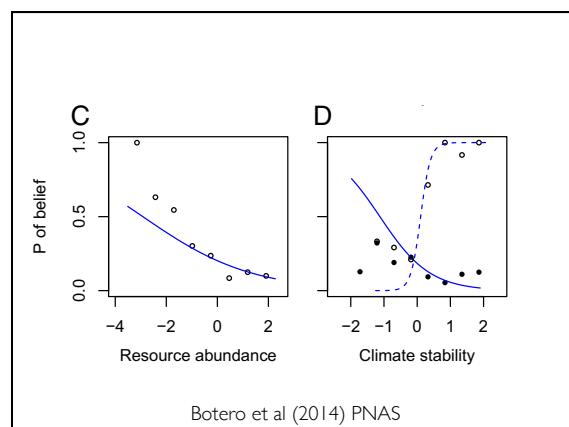
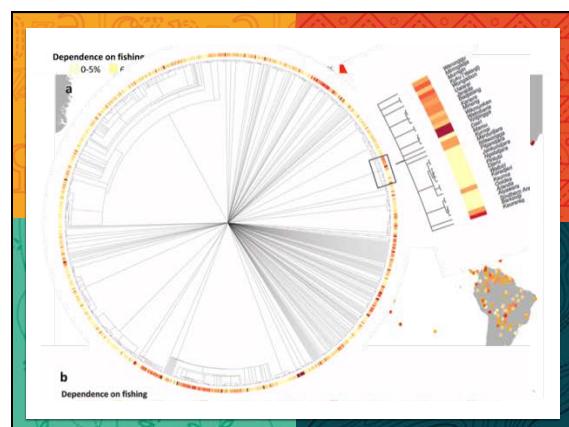
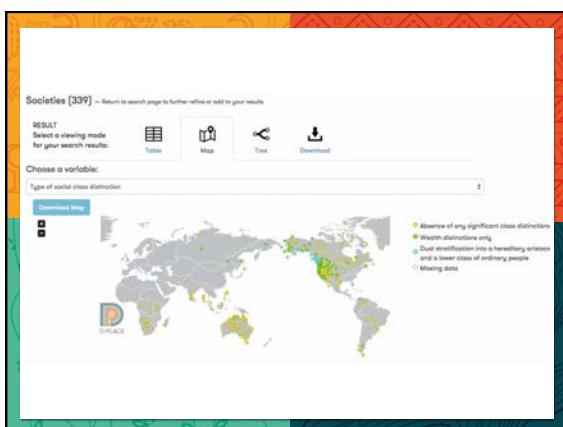
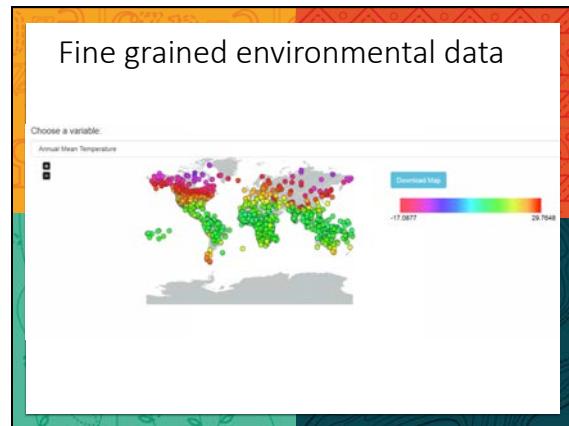
Database issue

1. Standard coding (ISO, Glottocode, IPA etc)
2. Accessible (no more private hobby databases)
3. Extensible
4. Correctable
5. Linked

GLOTTOBANK: WORLD-SCALE LINGUISTIC DATABASES

D-PLACE
A Global Database of Cultural, Linguistic & Environmental Diversity

PLACES LANGUAGE CULTURE ENVIRONMENT



Religious beliefs are surprisingly predictable!

0.91

Botero et al (2014) PNAS

PULOTU
DATABASE OF PACIFIC RELIGIONS

Home About Cultures Compare Cultures Glossary



PULOTU
DATABASE OF PACIFIC RELIGIONS

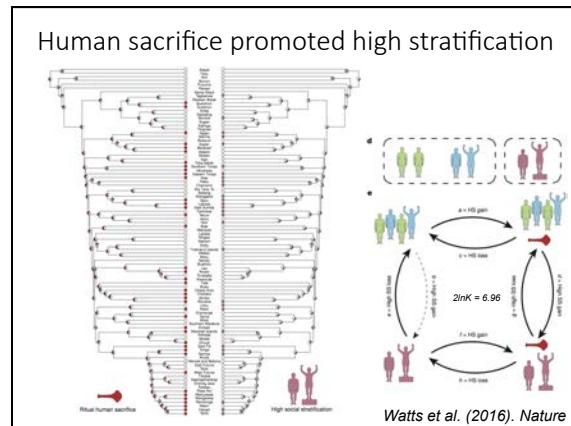
Pulotu, meaning abode of the gods, is a database of Austronesian religious beliefs and practices. You will find information on 118 cultures from the Moken of mainland Asia to the Maori of New Zealand.

PLOS ONE

<https://pulotu.shh.mpg.de>



Joseph Watts



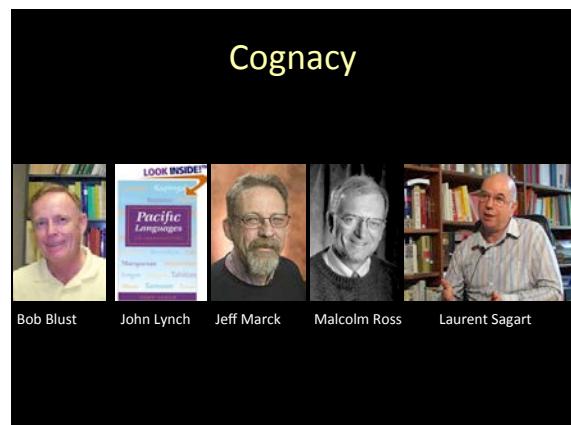
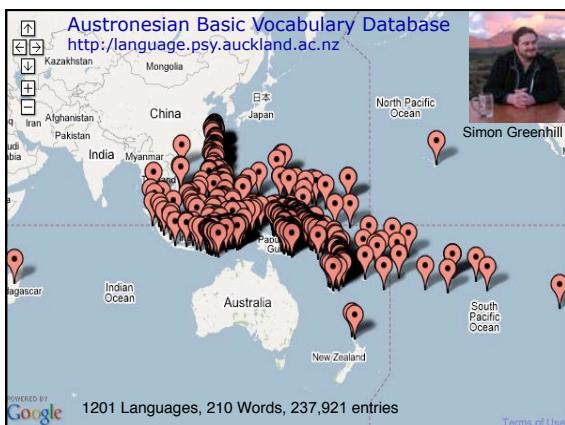
Plan – scaling up...

1. Big data
2. Better methods



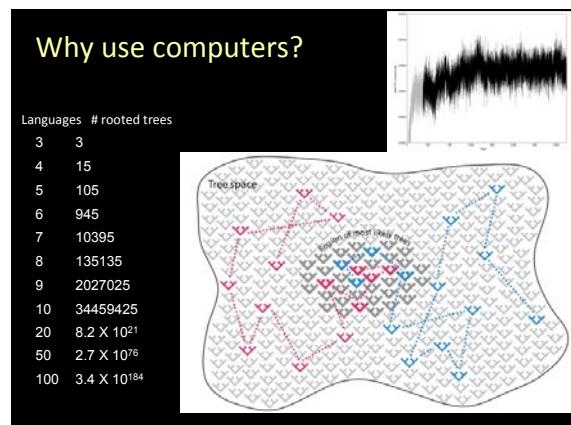


WAEREN 200-WORD BASIC VOCABULARY (MODIFIED)	
<i>Completed by: John Lynch Language/Dialect: Lenape Date: October, 1980</i>	
1. hand	awlin-
2. left	mwl-
3. right	nmawatu
4. leg/foot	nalk-
5. to walk	alivuk
6. road/path	suutu
7. to come	va
8. to turn (e.g. to turn the wheel, etc.)	atilim
9. to swim	aik
10. dirty	amk'awak
11. dust	nmawoul
12. skin	navig-
13. back (body part)	nmawinfina-
14. belly	netp-
15. bone	aswuu naka'kalii-
16. intestine	aswaga-
17. liver	nakem-pn-
18. heart	neha-
19. shoulder	nauvik-
20. to know (shape), hination	be knowledgeable
21. to think	nhin-n
22. to fear	oyp-
23. blood	nataka-
24. head	kapwa-
25. nose	-
26. hair (of the head)	neuvantu
27. nose	aspungnhaag-
28. to breathe	amig-
29. to sniff, smell	areu*
30. mouth	nhul-
31. tooth	ne-
32. eye	nam-
33. to laugh	ahmehla
34. to cry	akala
35. to vomit	zuk-
36. to spit	agh
37. to eat	kan
38. to chew (e.g. general term than to chew hotel	chawai
39. to speak (e.g. general term than to talk food)	viiin
40. to eat/mouth	chawumuw
41. to bite	ka-
42. to drink	ta-
43. ear	namukately-
44. to hear	areu*
45. eye	nam-
46. to see	amih-
47. to yell	ouang-that (e open- end)
48. to sleep	apul
49. to lie down (to sleep)	amal



Cognate coding – from wordlists to binary matrix

Language	"father"	cognacy	binary
Paiwan	tjama	1	10
Ibayaten	qamaq	1	10
Motu	tama-na	1	10
Fijian (Bau)	tama-na	1	10
Tongan	tama i	1	10
Rarotongan	metua	2	01
Maori	matua	2	01



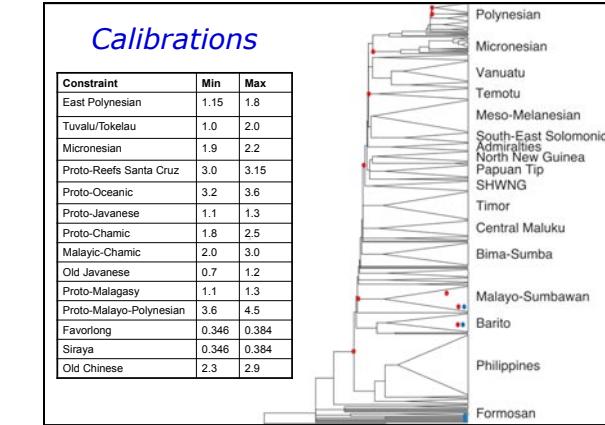
"linguists don't do dates"

April & Robert McMahon (2006)

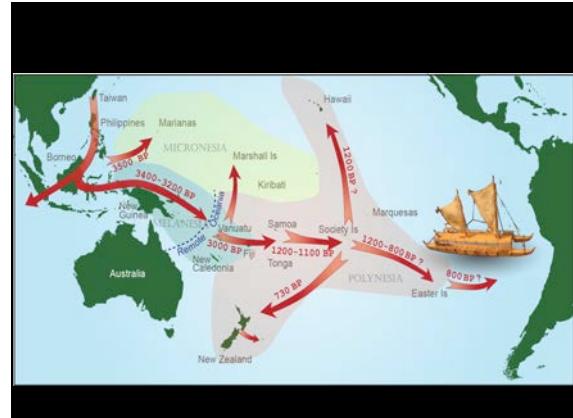
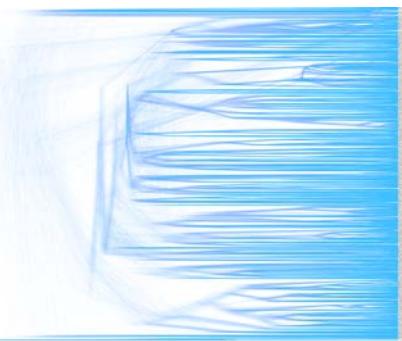


Calibrations

Constraint	Min	Max
East Polynesian	1.15	1.8
Tuvalu/Tokelau	1.0	2.0
Micronesian	1.9	2.2
Proto-Reefs Santa Cruz	3.0	3.15
Proto-Oceanic	3.2	3.6
Proto-Javanese	1.1	1.3
Proto-Chamic	1.8	2.5
Malayic-Chamic	2.0	3.0
Old Javanese	0.7	1.2
Proto-Malagasy	1.1	1.3
Proto-Malayo-Polynesian	3.6	4.5
Favorlong	0.346	0.384
Siraya	0.346	0.384
Old Chinese	2.3	2.9

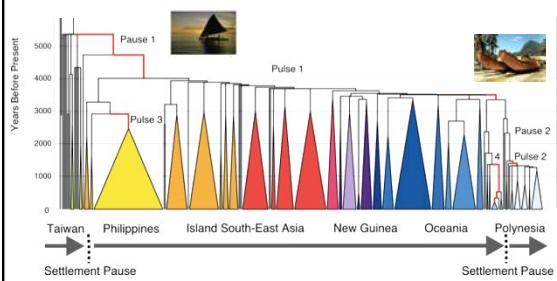


Posterior Tree Distribution



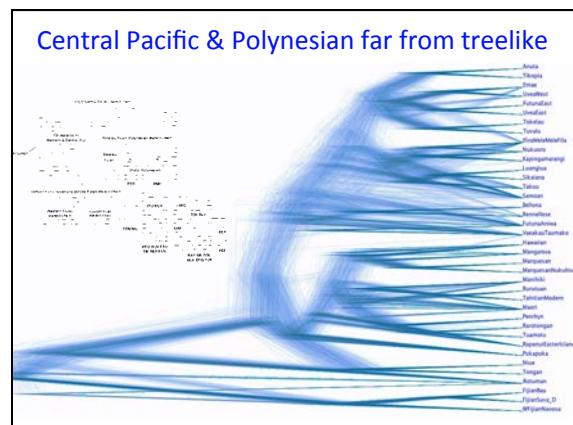
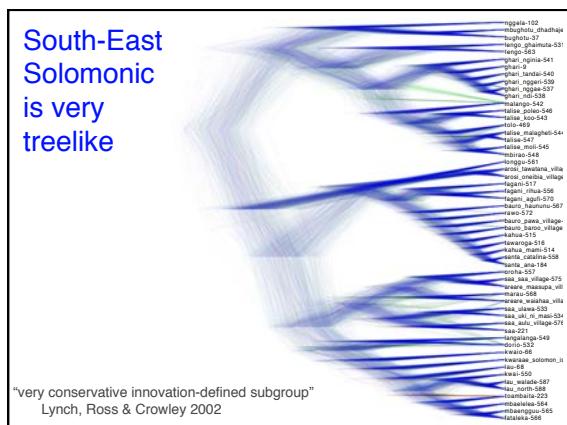
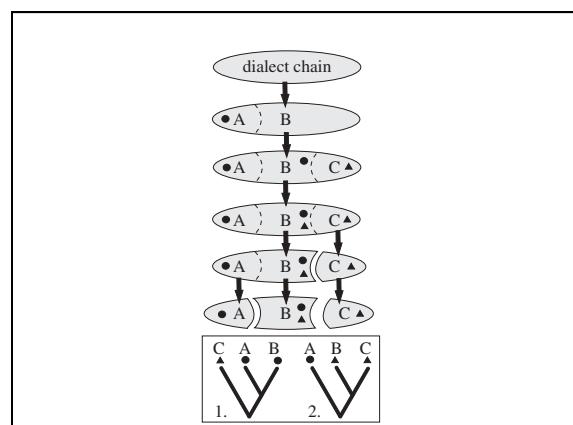
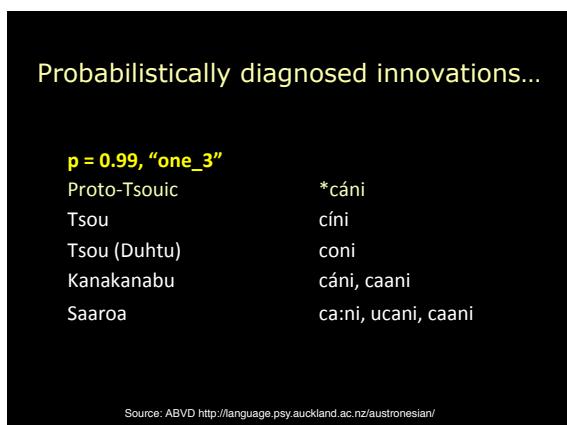
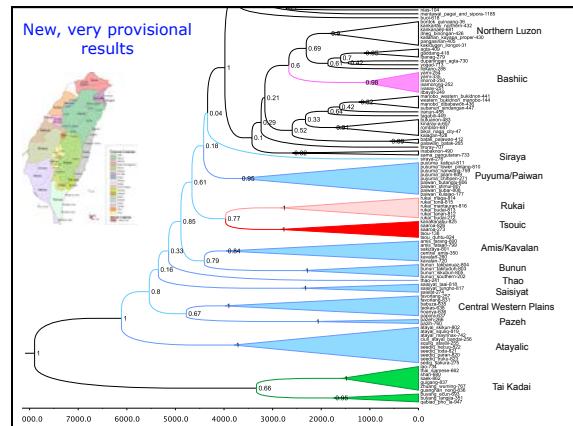
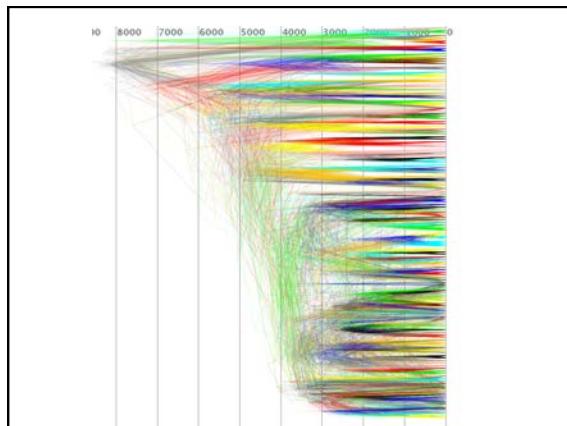
3 predictions:

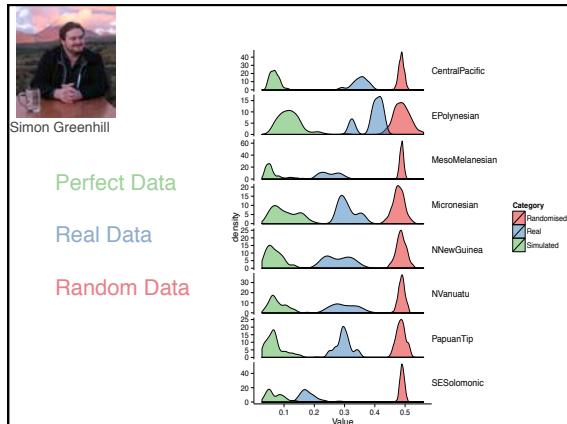
Sequence, timing, pulses and pauses
Gray, Drummond & Greenhill. 2009. *Science*, 323, 479-483.



Jena 2015 Austronesian working group



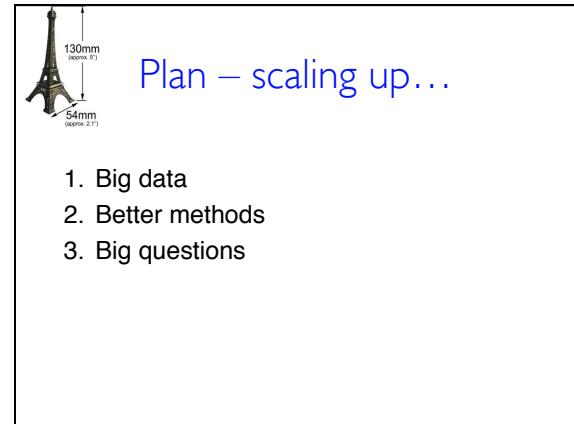




Perfect Data

Real Data

Random Data



What are the Hilbert problems in linguistics?

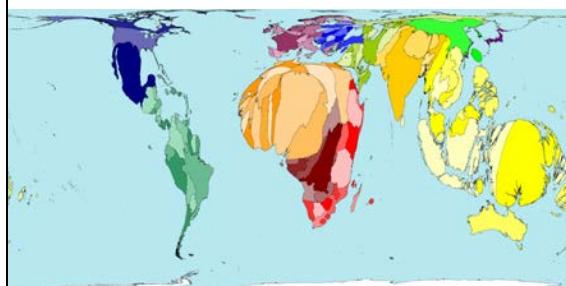


David Hilbert



<https://www.youtube.com/watch?v=X4OaN39eNAI&feature=youtu.be>

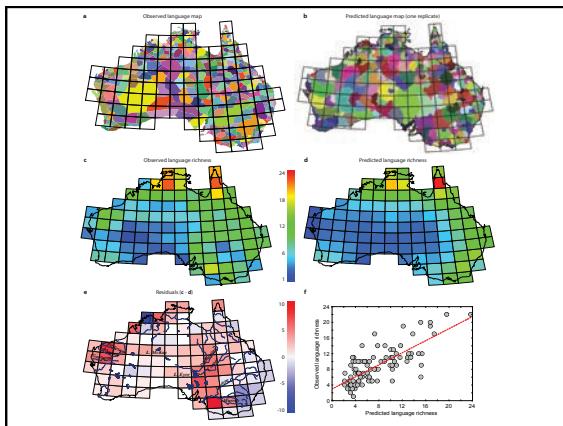
Explain this!



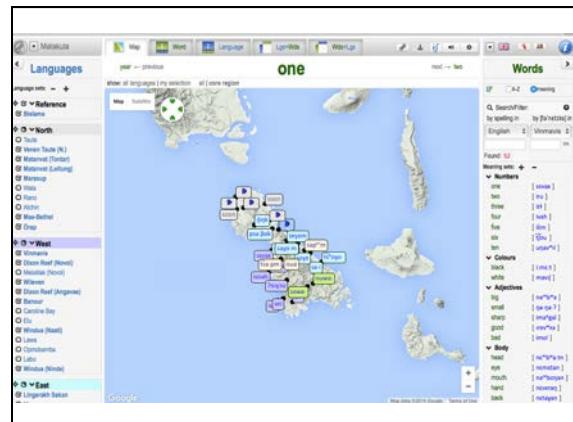
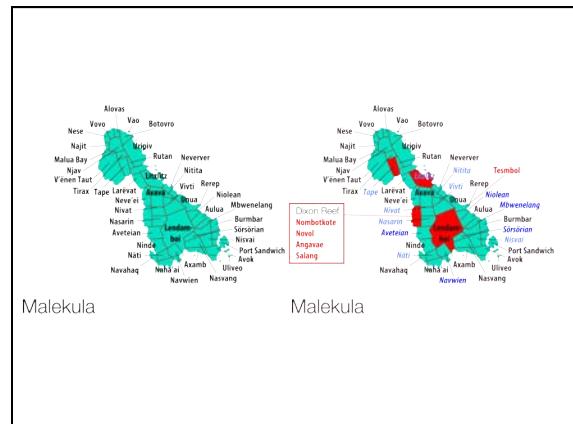
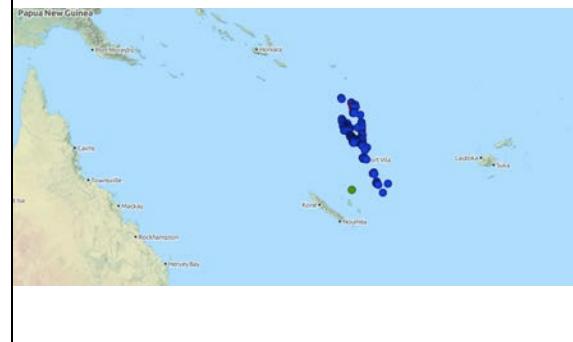
Some suggestions...

1. Why are there approximately 7000 languages?
2. Why is language diversity distributed so patchily?
3. When did spoken language evolve?
4. How far back can we push the time barrier for detecting language relationships?





Vanuatu – the Galapagos of linguistic and cultural evolution



Plan – scaling up...

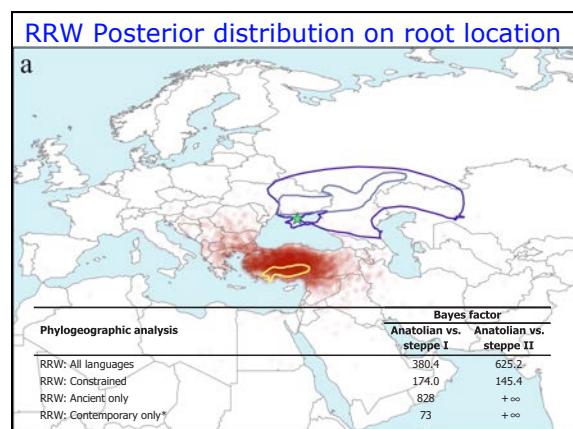
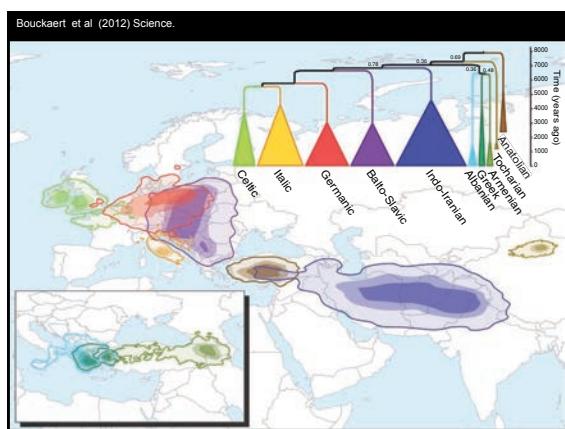
1. Big data
2. Better methods
3. Big questions
4. Big interdisciplinary teams

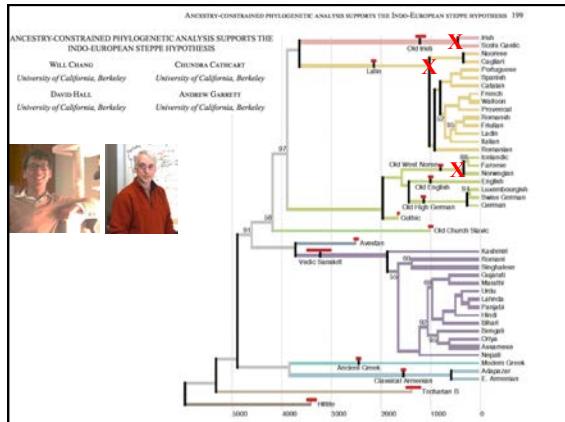


The origin of Indo-European languages

"the most intensively studied, yet still most recalcitrant, problem of historical linguistics"

Diamond and Bellwood, *Science*, 2003





Big problem – inconsistent data coding
(Swadesh policy of most commonly used lexeme not consistently followed)

Meaning	Orthographic	Transliteration	Not Lex. Sust. exst.	Lex. Ism.	nCog	Root Form	Root Language	Lex. Ext.
BREATHE	uxah lo	uxah lo						
BURN	jíle	jíle						
CHILD	ápli	ápli						
CHILD	jiput	jiput						
CHILD	sau	sau						
CHILD	lará	lará						
CHILD	késut	késut						
CHILD	sowilli	sowilli						

This screenshot shows a database interface for Assamese words. The table lists meanings like 'BREATHE', 'BURN', and various forms of 'CHILD'. The 'Orthographic' column shows Assamese characters, while the 'Transliteration' column shows their Romanized equivalents. The 'Root Form' and 'Root Language' columns are empty, which is noted as a problem in the caption.

Welcome to the CoBL website.

The database contains:

- Languages
- Meanings
- Words
- Cognate Sets
- Coded Characters

Language int: 2016-02-17_144
Wordset: all

This is the homepage of the CobL (Cognate-Oriented Basic Lexicon) website. It displays a grid of small portraits of people, likely researchers, and a search bar at the top. The main content area is titled 'The database contains' and lists categories such as Languages, Meanings, Words, Cognate Sets, and Coded Characters.

COBL – Ancient Greek

Meaning	Orthographic	Transliteration	Not Loan Root	Loan Root	Root Form	Root Language	nCog	PhoneTic	PhoneMic	PhoneTic Gloss	Notes	Dict ID	Dubt
ANT	παῦρον	paúron											
ASH	ἄσης	ásēs											
BACK	ἀγέρε	ágere											
BAD	ἀδεῖς	ádeis											
BARK	ἀγλαῖς	áglaȋs											
BELLY	ἀγόρης	ágorēs											
BIG	ἀριτίς	áritis											
BIRD	ἄρνης	árñēs											
BITE	ἀπίστη	ápitísi											
BITTER	ἄπικης	ápikēs											

This screenshot shows a database interface for Ancient Greek words. The table lists meanings like 'ANT', 'ASH', and various parts of the body. The 'Orthographic' column shows Greek characters, while the 'Transliteration' column shows their Romanized equivalents. The 'Root Form' and 'Root Language' columns are empty.

