The phylogeny of word meanings
Inferring the directionality of semantic change from word lists

Gerhard Jäger
joint work with Alla Münch and Johannes Dellert

Seminar für Sprachwissenschaft, Tübingen

Wissenschaftskolleg Berlin

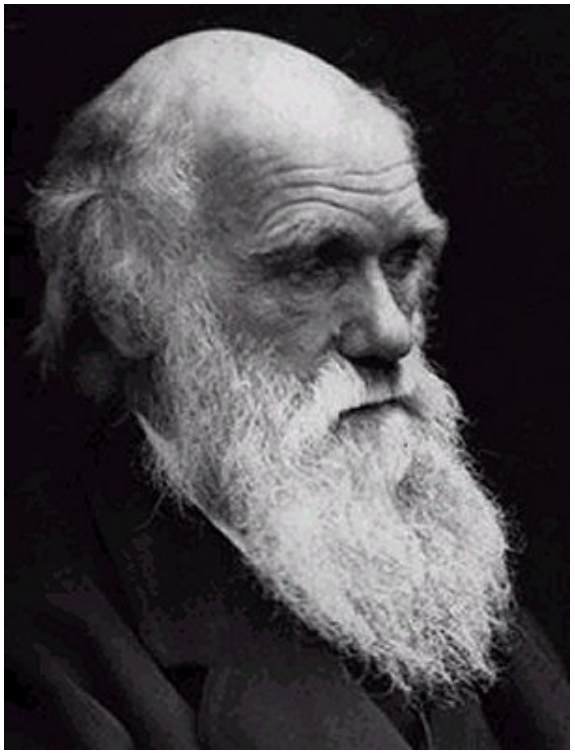
January 26, 2016

WORDS BONES GENES TOOLS
Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past
EBERHARD KARLS UNIVERSITÄT TÜBINGEN
DFG
erc
European Research Council
Established by the European Commission

Evolution and language change

"The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel. [...] We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation. The manner in which certain letters or sounds change when others change is very like correlated growth. [...] The frequent presence of rudiments, both in languages and in species, is still more remarkable. [...]
Languages, like organic beings, can be classed in groups under groups; and they can be classed either naturally according to descent, or artificially by other characters. Dominant languages and dialects spread widely, and lead to the gradual extinction of other tongues."

(Darwin, The Descent of Man)



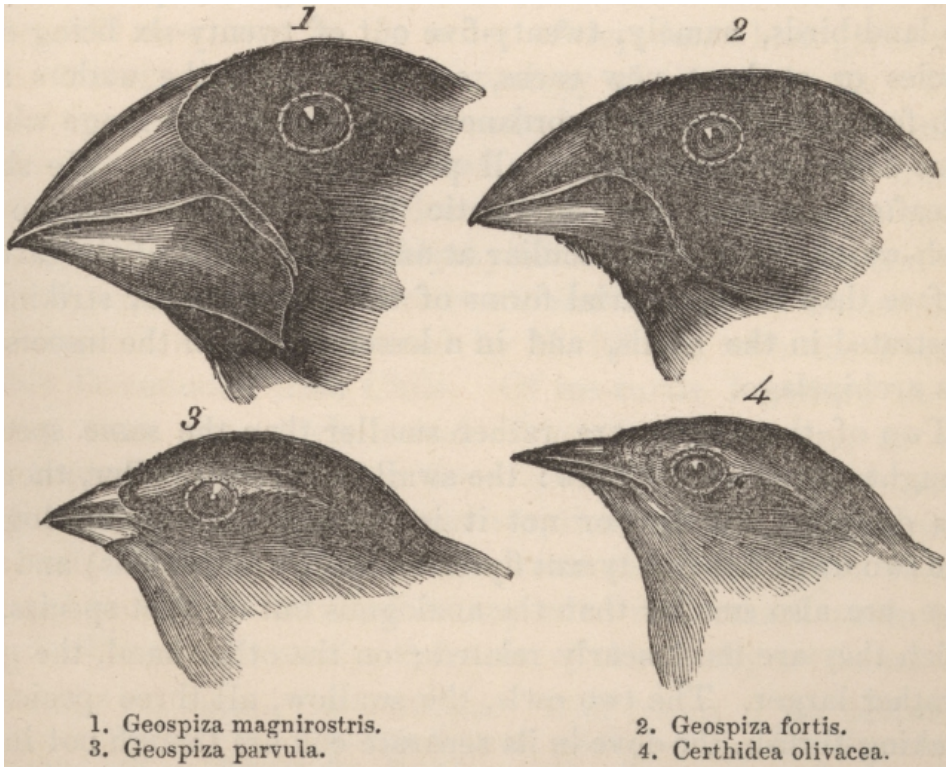
Evolution and language change

Vater Unser im Himmel, geheiligt werde Dein Name

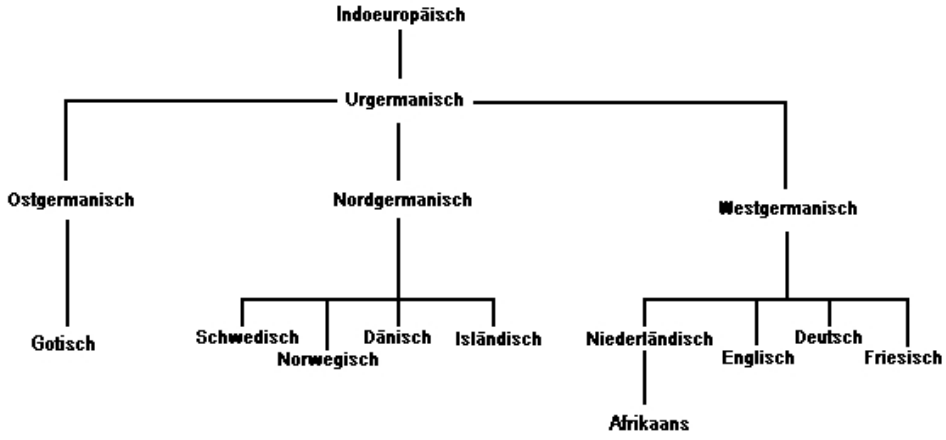
Onze Vader in de Hemel, laat Uw Naam geheiligd worden

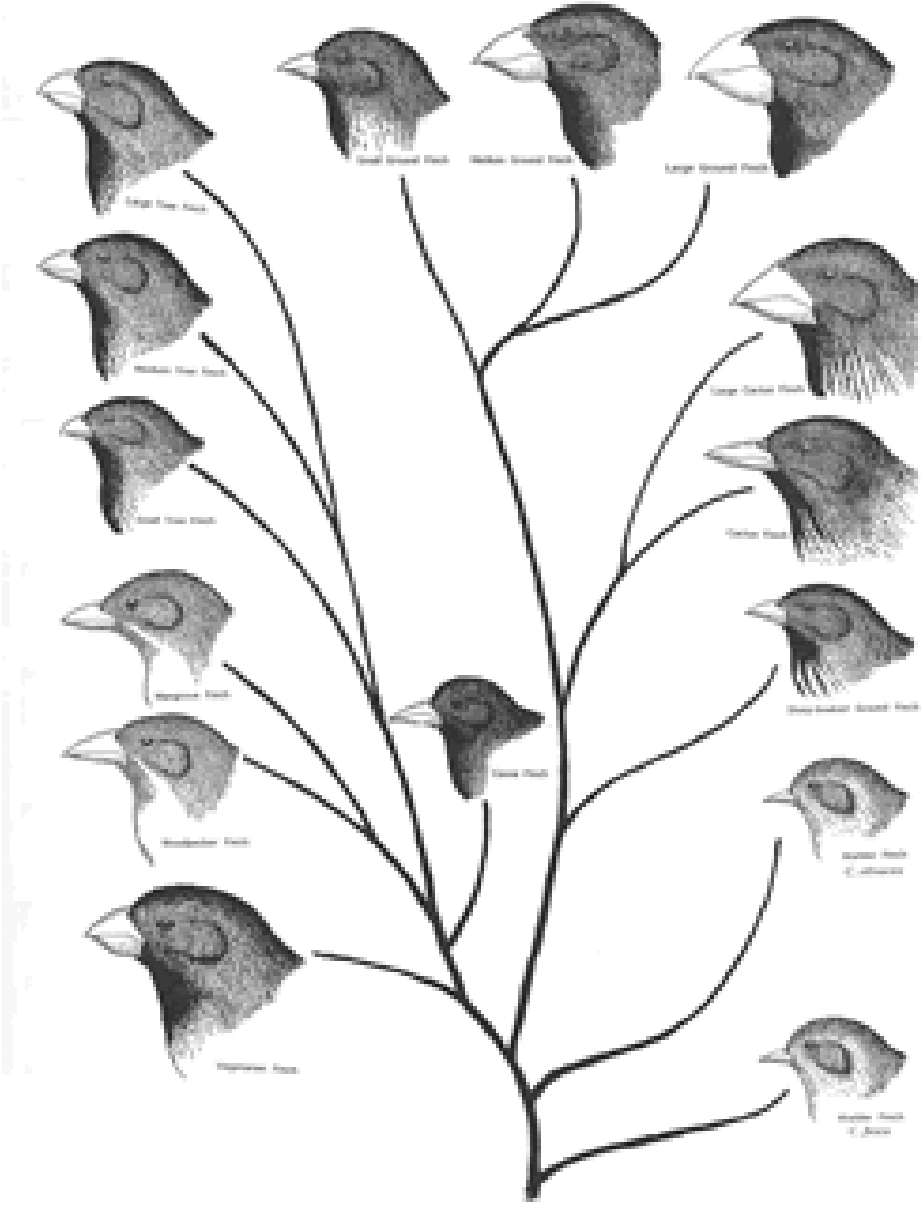
Our Father in heaven, hallowed be your name

Fader Vor, du som er i himlene! Helliget vorde dit navn



Evolution and language change



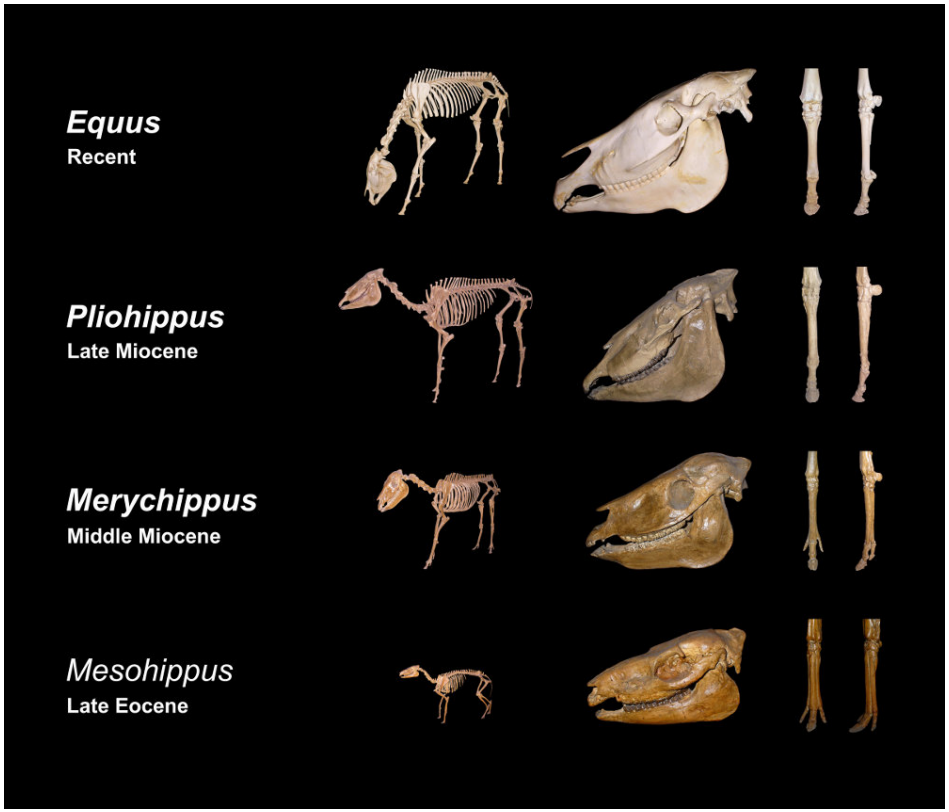


Evolution and language change

Middle High German:
Got vater unser, dâ du bist in dem himelrîche gewaltic alles des dir ist, geheiliget sô werde dîn nam

Old High German:
Fater unser thû thâr bist in himile, si giheilagôt thîn namo

Gothic:
Atta unsar þu in himinam, weihnai namo þein



Genetic language relationships

- In most cases, we do not have written records of earlier stages
- Regular **sound correspondences** provide evidence for genetic relationship though
  - Correspondences indicate common ancestor + different sound shifts
  - The more **cognates** two languages share and the fewer sound shifts separate them, the closer they are related

Example: High German vs. West Germanic

Erste bzw. Germanische Lautverschiebung (Indoeuropäisch → Germanisch)	Phase	Zweite bzw. Hochdeutsche Lautverschiebung (Germanisch→Althochdeutsch)	Beispiele (Neuhochdeutsch)	Jahrhundert	Dialektgebiete
G: /*b/→/*p/	1	**/*p/→/ff/→/f/**	niederdeutsch: sla**p**en, englisch: slee**p** → schla**f**en; niederdeutsch und englisch: Schi**pp**, shi**p** → Schi**ff** niederdeutsch: scher**p**, englisch: shar**p** → scha**rf**	4/5	oberdeutsch und mitteldeutsch
	2	**/*p/→/pf/**	niederdeutsch: **P**eper, englisch: **p**epper → **Pf**effer; niederdeutsch: **P**lauch, englisch: **p**lough → **Pf**lug; niederdeutsch: scher**p**, englisch: shar**p**, althochdeutsch: scar**ph**, mittelhochdeutsch: scha**rpf**	6/7	oberdeutsch
G: /*d/→/*t/	1	**/*t/→/ss/→/s/**	niederdeutsch: da**t**, wa**t**, e**t**en; englisch: tha**t**, wha**t**, ea**t** → da**s**, wa**s**, e**ss**en	4/5	ober- und mitteldeutsch[1]
	2	**/*t/→/ts/**	niederdeutsch: **T**iet, englisch: **t**ide (Gezeiten), schwedisch: **t**id → **Z**eit; niederdeutsch: ver-**t**ellen, englisch: **t**ell → er-**z**ählen; **T**immermann → **Z**immermann	5/6	ober- und mitteldeutsch
G: /*g/→/*k/	1	**/*k/→/xx/→/x/**	niederdeutsch: i**k**, altenglisch: i**c** → i**ch**; niederdeutsch und englisch: ma**k**en, ma**k**e → ma**ch**en; niederdeutsch: au**k** → au**ch**	4/5	ober- und mitteldeutsch[2]
	2	**/*k/→/kx/**	**K**ind → bairisch: **Kch**ind	7/8	südbairisch, hoch- und höchstalemannisch
G: /*b^h/→/*b/ V: /*p/→/*b/	3	**/*b/→/p/**	**B**erg, ***b**ist* → *bairisch: **p**erg, **p**ist*	8/9	teilweise bairisch und alemannisch
G: /*d/→/*đ/→/*d/ V: /*t/→/*đ/→/*d/	3	**/*d/→/t/**	niederdeutsch: **D**ag oder **D**ach, englisch: **d**ay → **T**ag; niederfränkisch: *va**d**er* → Va**t**er	8/9	oberdeutsch
G: /*g^h/→/*g/ V: /*k/→/*g/	3	**/*g/→/k/**	**G**ott → bairisch: **K**ott	8/9	teilweise bairisch und alemannisch
G: /*t/→/þ/ [ð]	4	**/þ/→/d/** **/ð/→/d/**	englisch: **th**orn, **th**istle, **th**rough, bro**th**er → **D**orn, **D**istel, **d**urch, Bru**d**er	9/10	gesamtes deutsches Dialektkontinuum

Example: Polynesian languages

- Taken from Crowley & Bowern (2010)

How historical linguists do it

TABLE 5.1 Data from Four Polynesian Languages

	Tongan	Samoan	Rarotongan	Hawaiian	
1.	tapu	tapu	tapu	kapu	'forbidden'
2.	pito	pute	pito	piko	'navel'
3.	puhi	feula	puʔi	puhi	'blow'
4.	tafaʔaki	tafa	taʔa	kaha	'side'
5.	taʔe	tae	tae	kae	'feces'
6.	taŋata	taŋata	taŋata	kanaka	'man'
7.	tahi	tai	tai	kai	'sea'
8.	malohi	malosi	kaʔa	ʔaha	'strong'
9.	kalo	ʔalo	karo	ʔalo	'dodge'
10.	aka	aʔa	aka	aʔa	'root'
11.	ʔahu	au	au	au	'gall'
12.	ʔulu	ulu	uru	poʔo	'head'
13.	ʔufi	ufi	uʔi	uhi	'yam'
14.	afi	afi	aʔi	ahi	'fire'
15.	faa	faa	ʔaa	haa	'four'
16.	feke	feʔe	ʔeke	heʔe	'octopus'
17.	ika	iʔa	ika	iʔa	'fish'
18.	ihu	isu	putaŋio	ihu	'nose'
19.	hau	sau	ʔau	hau	'dew'
20.	tafuafi	siʔa	ʔika	hiʔa	'firemaking'
21.	hiku	siʔu	ʔiku	hiʔu	'tail'
22.	hake	aʔe	ake	aʔe	'up'
23.	huu	ulu	uru	komo	'enter'
24.	maŋa	maŋa	maŋa	mana	'branch'
25.	maʔu	mau	mau	mau	'constant'
26.	maa	mala	mara	mala	'fermented'
27.	naʔa	faʔaŋa	maninia	naa	'quieten'
28.	nofo	nofo	noʔo	noho	'sit'
29.	ŋalu	ŋalu	ŋaru	nalu	'wave'
30.	ŋutu	ŋutu	ŋutu	nuku	'mouth'
31.	vaka	vaʔa	vaka	waʔa	'canoe'
32.	vaʔe	vae	vae	wae	'leg'
33.	laho	laso	raʔo	laho	'scrotum'
34.	lohu	lou	rou	lou	'fruit-picking pole'

Guidelines for reconstruction

- Only establish sound correspondences if you are reasonably sure the words are cognate
- Assume sound shifts that are plausible (are known to occur frequently)
- Assume as few sound changes as possible for reconstructing a proto-language
- The reconstructed proto-language should have a typologically plausible sound system

Polynesian example

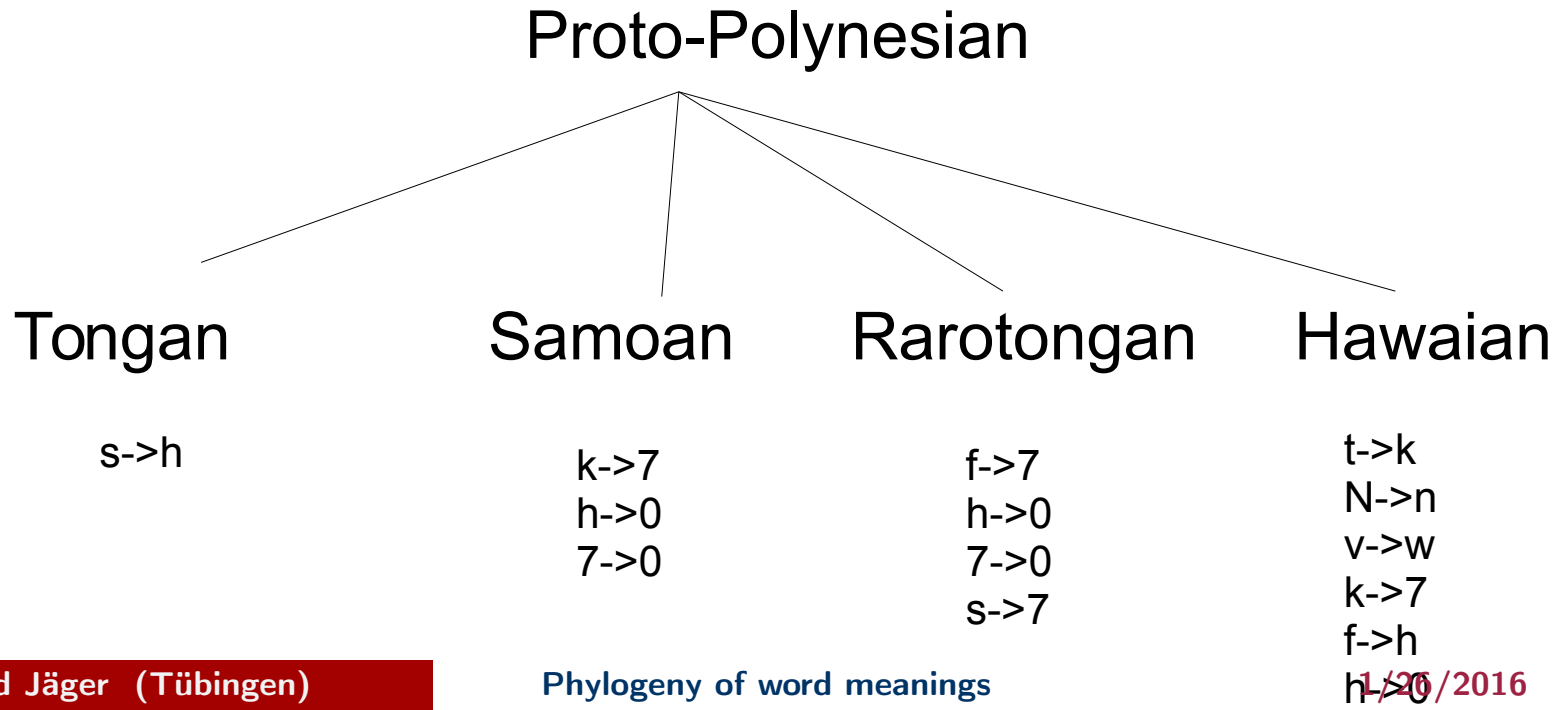
- Vowels in Proto-Polynesian are unchanged in daughter languages (otherwise we would stipulate unnecessary sound shift)
- Likewise, *p*, *m* and *n* are unchanged
- Majority rule:
  - pp. **t, *N, *v* → hw. *k, n, w*
- lenition is more likely than fortition
- also, Proto-Polynesian has *p* and *t*, so it should also have a *k*, hence:
  - pp. **k* → sm., hw. *7* (rather than **7* → tg./rg. *k*)

Polynesian example

- majority rule:
  - pp. **f* → rg. *7*, hw. *h*
- not enough data to reconstruct the *l* and *r*
- majority rule:
  - pp. **h, *7* → sm., rg., hw. *0*
- change *s* → *h* is known to be more common than *h* → *s*, hence (against majority rule):
  - pp. **s* → tg./hw. *h*, rg. *7*

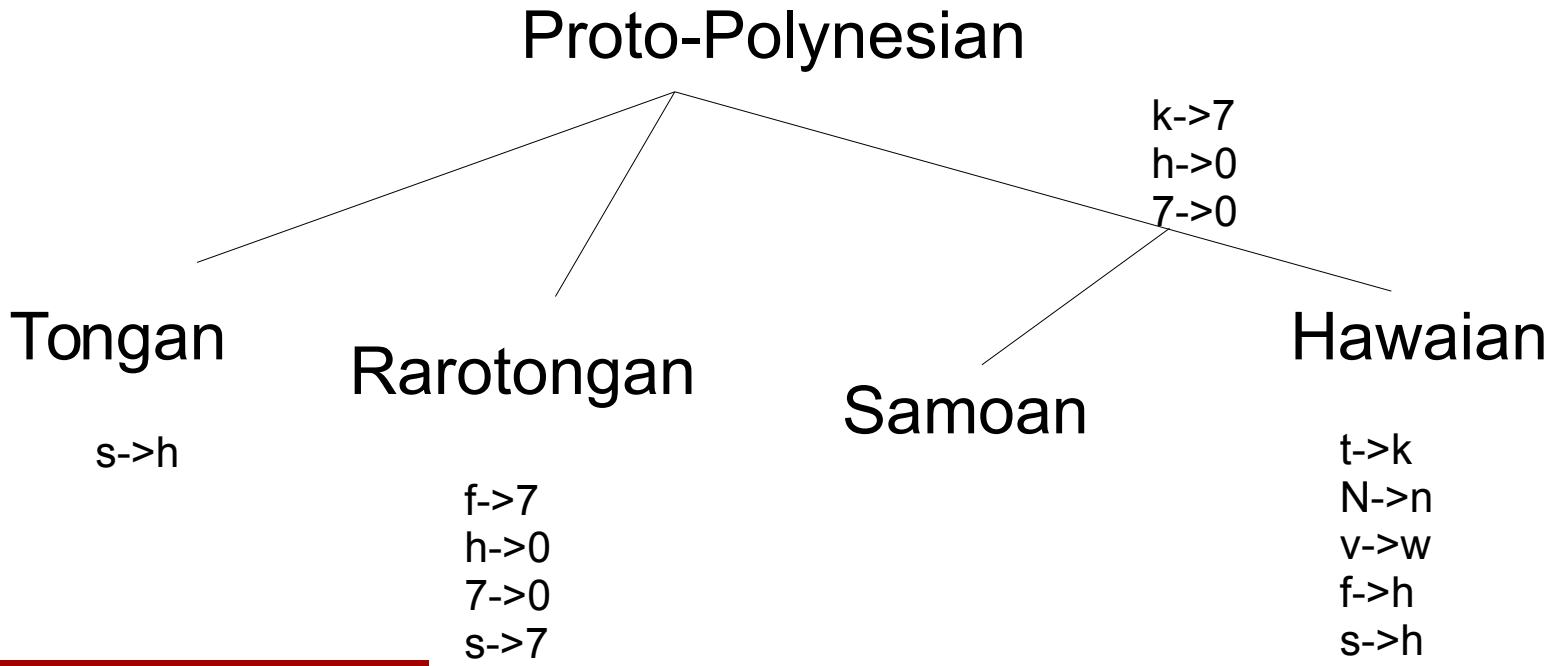
Polynesian example

- constructing a tree

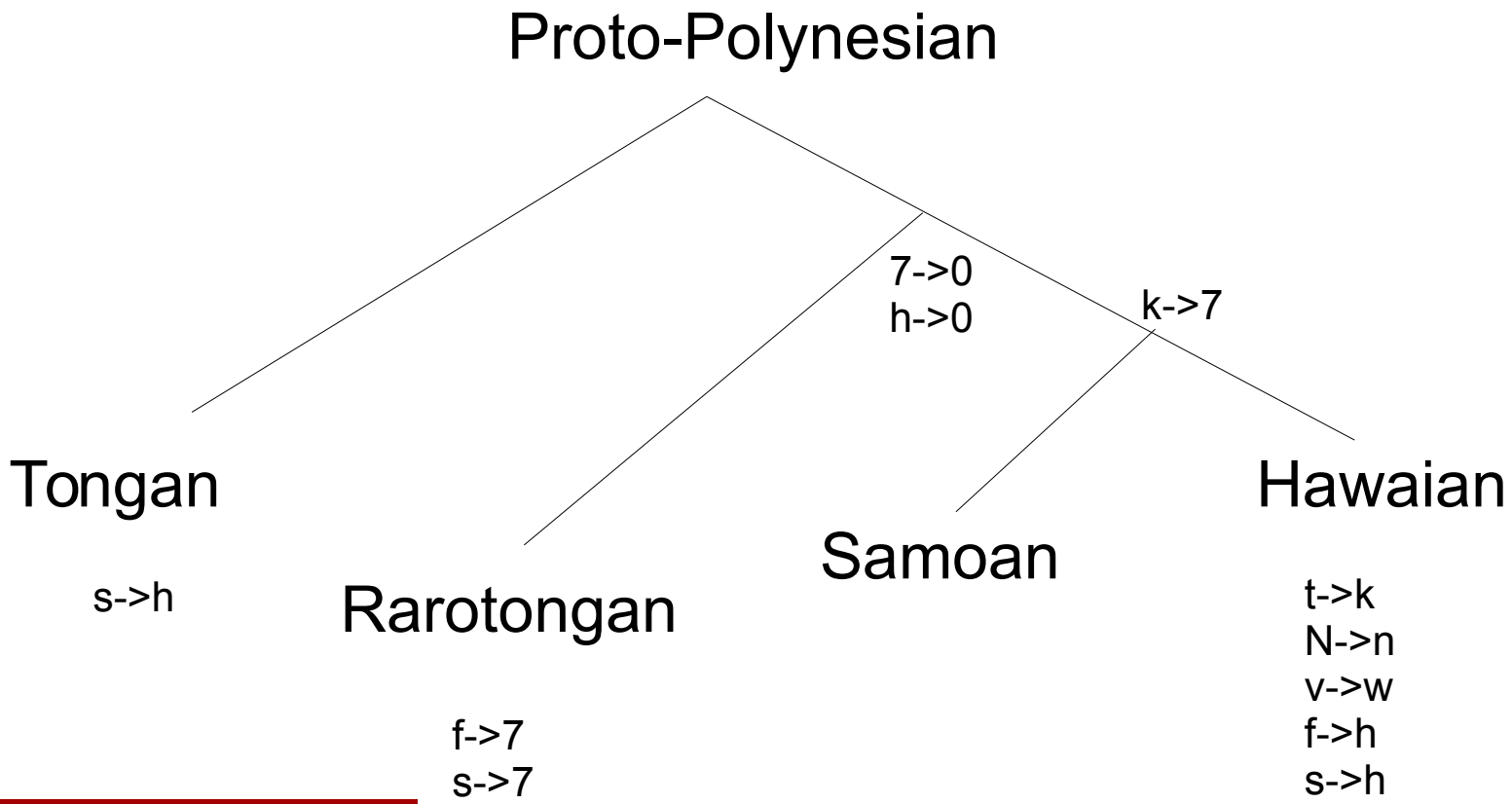


Polynesian example

- constructing a tree



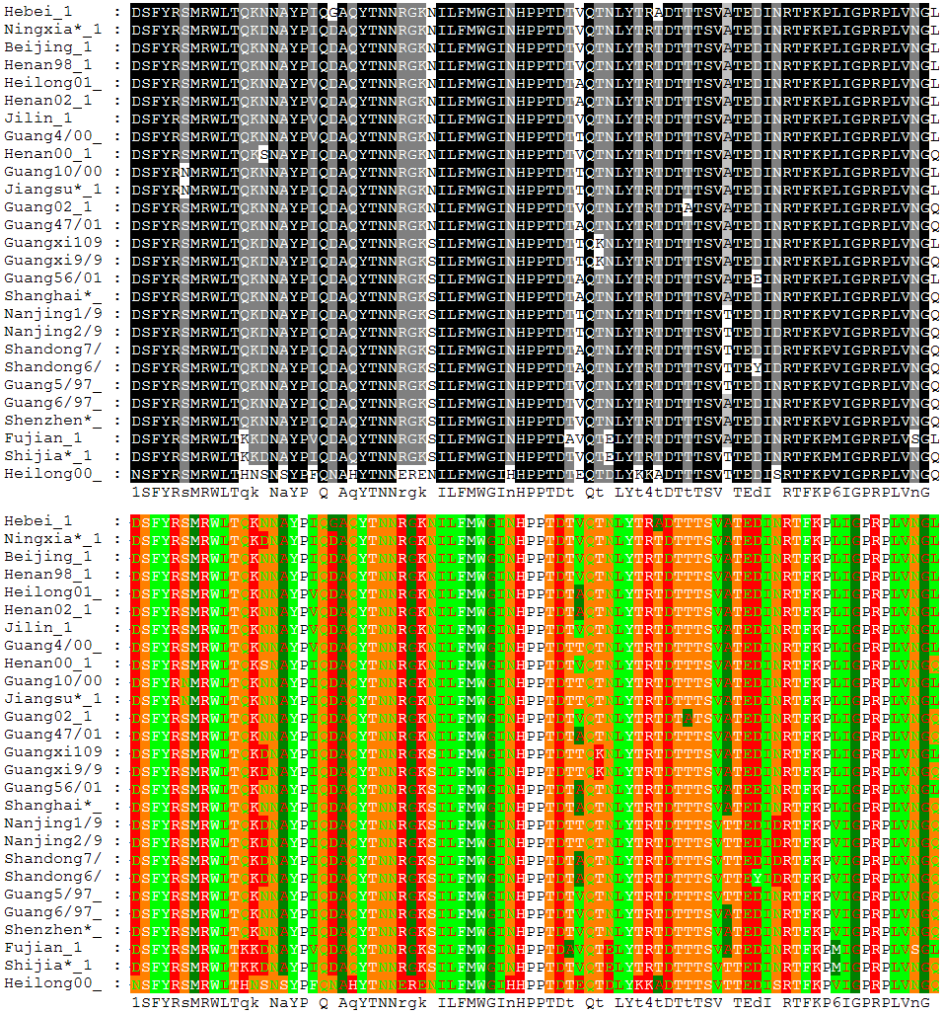
Polynesian example



Polynesian example

- reconstruction seems reasonable because
  - only one shift is assumed twice (s->7), and this type is known to occur frequently
  - reconstruction assumes (pull-) chain shifts
    - Rarotongan and Proto-Samoan/Hawaian restore the lost 7
    - Hawaiian additionally restores the lost k and h
- this procedure started from a reconstructed proto-language; usually tree construction and reconstructon of ancestral forms go hand in hand

How computational biologists do it



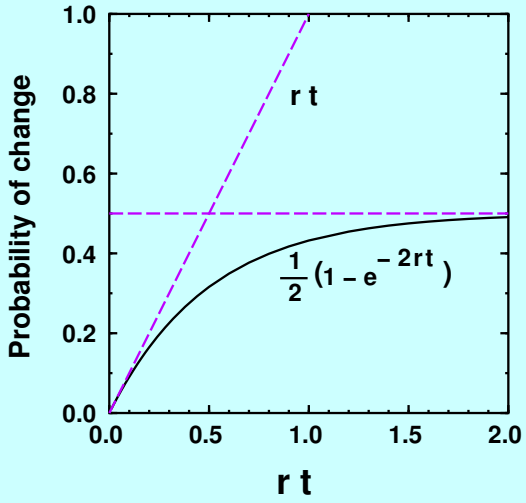
How computational biologists do it

Transition probabilities in a two-state model

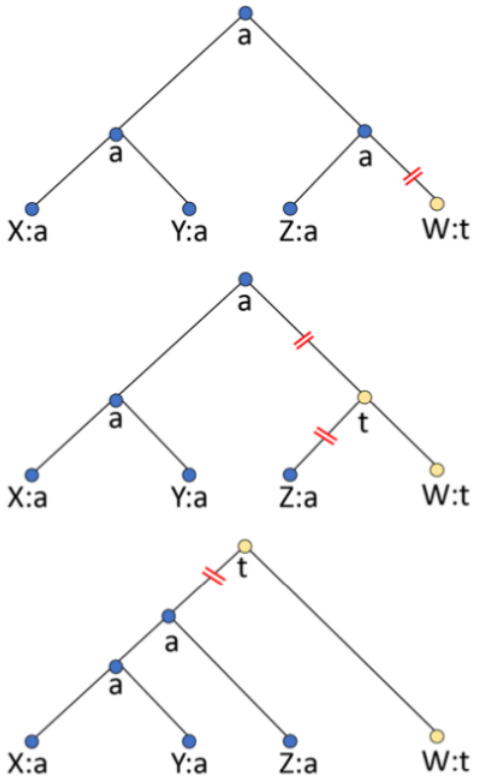
In a symmetric two-state model with rate of change r per unit time, where

$$0 \underset{r}{\overset{r}{\rightleftharpoons}} 1$$

we have $\mathrm{Prob}\,(1 \mid 0, t, r) \;=\; \frac{1}{2}\,(1 - e^{-2\,r\,t})$



When rt is small then to very good approximation the probability of the events in the branch is rt or $1 - rt$. The latter is nearly 1.



How computational biologists do it

Rooted, bifurcating, labelled trees

species	number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

Genome 570, Phylogenetic Inference – p.33/45

How computational biologists do it

- computer doggedly
  - looks at one damned tree after another
  - computes the likelihood of the data given that tree, and
  - moves on to the next tree
- exhaustive search is impossible, given the astronomical number of different trees
- no guarantee to find the single best tree, but there are good heuristics
- highly informative results:
  - estimates of branch lengths
  - confidence values for branches of a tree
  - probabilistic reconstructions of ancestral states
  - ...

How computational biologists do it

English	Dutch	French
two	twee	deux
three	drie	trois
mountain	berg	mont
dog	hond	chien
tree	boom	arbre
tooth	tand	dent
cold	koud	froid

$\Rightarrow$

English	Dutch	French
A	A	A
A	A	A
A	B	A
A	B	B
A	B	C
A	A	A
A	A	B

- Different cognate classes for a given meaning are treated just like different alleles of a gene.
- Once the cognacy data are in this format, the full gamut of computational phylogenetics can be unleashed.

Spectacular results

Skeptically received by traditional historical linguists

Asya Pereltsvaig (historical linguist), blog post
The Malformed Language Tree of Bouckaert and His Colleagues

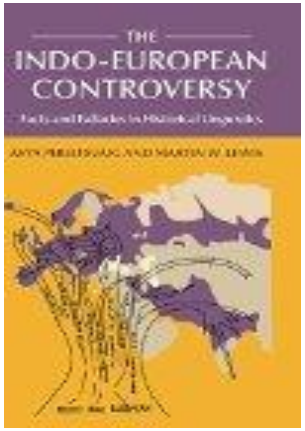
The study also groups Frisian with Flemish and Dutch, rather than English, which many Germanic scholars find objectionable.
…
It should also be pointed out that the authors of the Science article make a critical mistake in that they do not distinguish shared innovation from shared retention.
…
Another unsupportable configuration of the Science model concerns the internal grouping of the Slavic languages.
…
Yet again, we see that accepting the model proposed by Bouckaert and his colleagues requires one to believe not just three but actually dozens of impossible things before breakfast (with apologies to Lewis Carroll).

Language Log Home About Comments policy

Ultraconserved words? Really??
May 8, 2013 @ 12:05 pm · Filed by Sally Thomason under Linguistic history

« previous post | next post »

On the web site of the *Proceedings of the National Academy of Sciences,* in the "Early Edition" section, is an article by Mark Pagel, Quentin D. Atkinson,



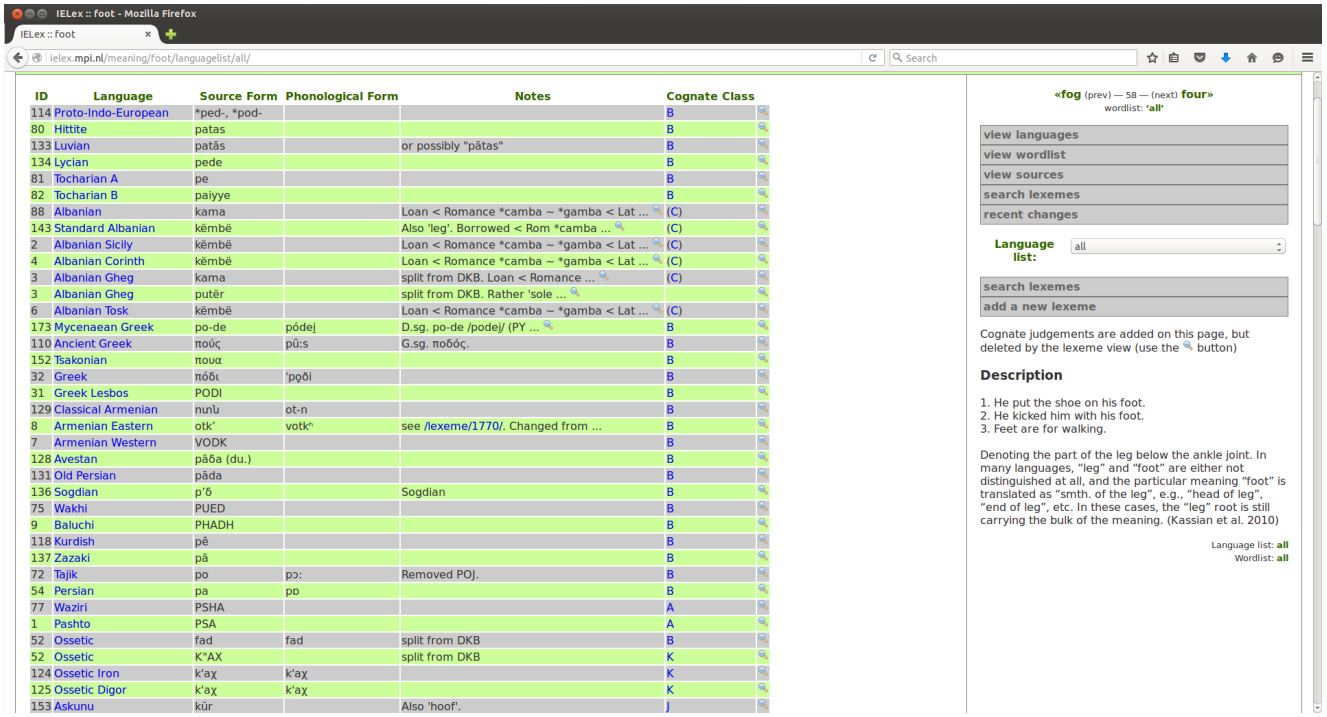
Martin Lewis (geographer), blog post
Mismodeling Indo-European Origin and Expansion: Bouckaert, Atkinson, Wade and the Assault on Historical Linguistics

In the Science piece, the painstaking work of generations of historical linguists who have rigorously examined Indo-European origins and expansion is shrugged off as if it were of no account, even though the study itself rests entirely on the taken-for-granted work of linguists in establishing relations among languages based on words of common descent (cognates).

While purporting to offer a truly scientific approach, Bouckaert et al. actually forward an example of scientism, or the inappropriate and overweening application of specific scientific techniques to problems that lie beyond their own purview.

A closer look at cognate classes

- based on expert judgments
- usually cover ca. 200 concepts (Swadesh list)
- publicly available for very few language families (mainly Indo-European and Austronesian)

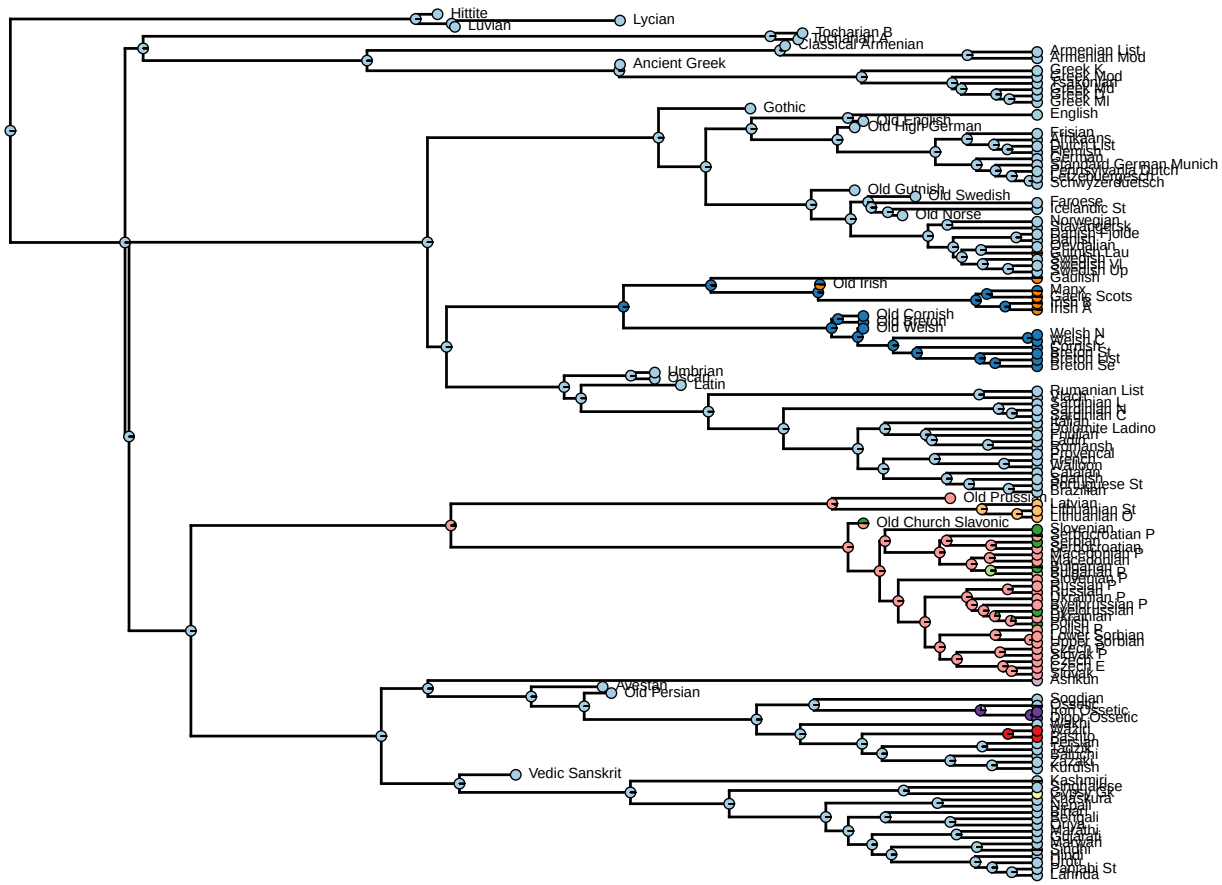


ID	Language	Source Form	Phonological Form	Notes	Cognate Class
114	Proto-Indo-European	*ped-, *pod-			B
80	Hittite	patas			B
133	Luvian	patās		or possibly "pātas"	B
134	Lycian	pede			B
81	Tocharian A	pe			B
82	Tocharian B	paiyye			B
88	Albanian	kama		Loan < Romance *camba ~ *gamba < Lat ...	(C)
143	Standard Albanian	këmbë		Also 'leg'. Borrowed < Rom *camba ...	(C)
2	Albanian Sicily	këmbë		Loan < Romance *camba ~ *gamba < Lat ...	(C)
4	Albanian Corinth	këmbë		Loan < Romance *camba ~ *gamba < Lat ...	(C)
3	Albanian Gheg	kama		split from DKB. Loan < Romance ...	(C)
3	Albanian Gheg	putër		split from DKB. Rather 'sole ...	
6	Albanian Tosk	këmbë		Loan < Romance *camba ~ *gamba < Lat ...	(C)
173	Mycenaean Greek	po-de	pódej	D.sg. po-de /podej/ (PY ...	B
110	Ancient Greek	πούς	pū́:s	G.sg. ποδός.	B
152	Tsakonian	πουα			B
32	Greek	πόδι	'pǫði		B
31	Greek Lesbos	PODI			B
129	Classical Armenian	nuṅu	ot-n		B
8	Armenian Eastern	otk'	votkʰ	see /lexeme/1770/. Changed from ...	B
7	Armenian Western	VODK			B
128	Avestan	pāδa (du.)			B
131	Old Persian	pāda			B
136	Sogdian	p'δ		Sogdian	B
75	Wakhi	PUED			B
9	Baluchi	PHADH			B
118	Kurdish	pê			B
137	Zazaki	pā			B
72	Tajik	po	pɔ:	Removed POJ.	B
54	Persian	pa	pɒ		B
77	Waziri	PSHA			A
1	Pashto	PSA			A
52	Ossetic	fad	fad	split from DKB	B
52	Ossetic	K"AX		split from DKB	K
124	Ossetic Iron	k'aχ	k'aχ		K
125	Ossetic Digor	k'aχ	k'aχ		K
153	Askunu	kūr		Also 'hoof'.	J

http://ielex.mpi.nl/

Phylogenetic inference with cognate classes

- defines a classification of languages into discrete states for each Swadesh concept
- example: Indo-European, concept *foot*



Phylogenetic inference with cognate classes

- cognate classes are mathematically treated like alleles of a gene, or different DNA letters
- evolution as continuous time Markov chain:
  - evolution along continuous time line (no discrete generations)
  - a language spontaneously, at each point in time
    - split into two daughter languages, or
    - change it's state for any concept
  - state transition probability density described by *rate matrix*
- for a given tree structure (with branch lengths) and rate matrix, each distribution of observed states at the leaves has a well-defined *likelihood*

Phylogenetic inference with cognate classes

- Inference methods:
  - **Maximum Likelihood:**
    - find the tree structure/rate matrix combination maximizing the likelihood of observed distribution of cognacy classes over languages
  - **Bayesian inference:**
    - assume prior distributions over tree structures and rate matrices
    - may incorporate historical, archaeological, linguistic etc. information
    - generate posterior distribution over trees

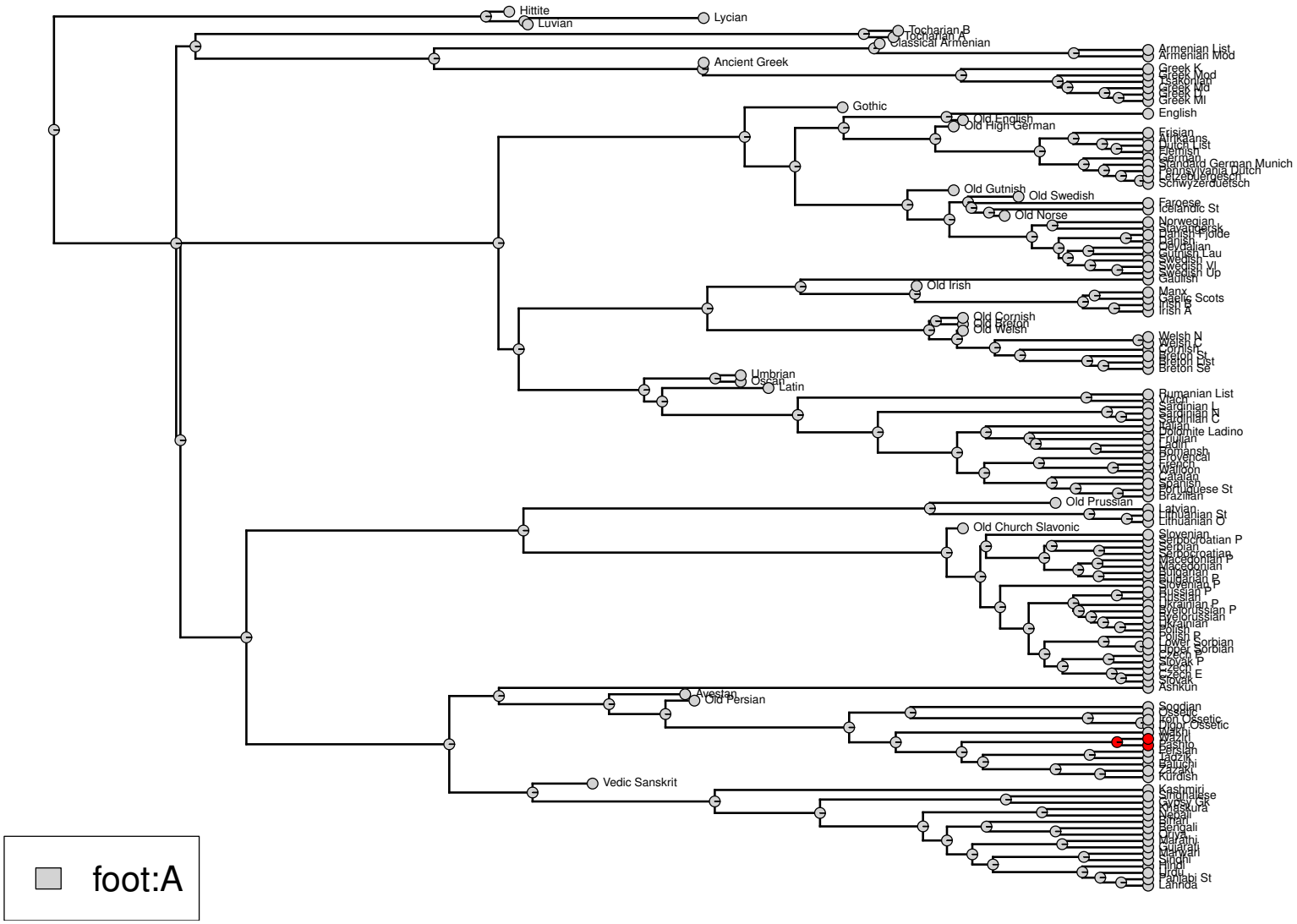
Practical issues

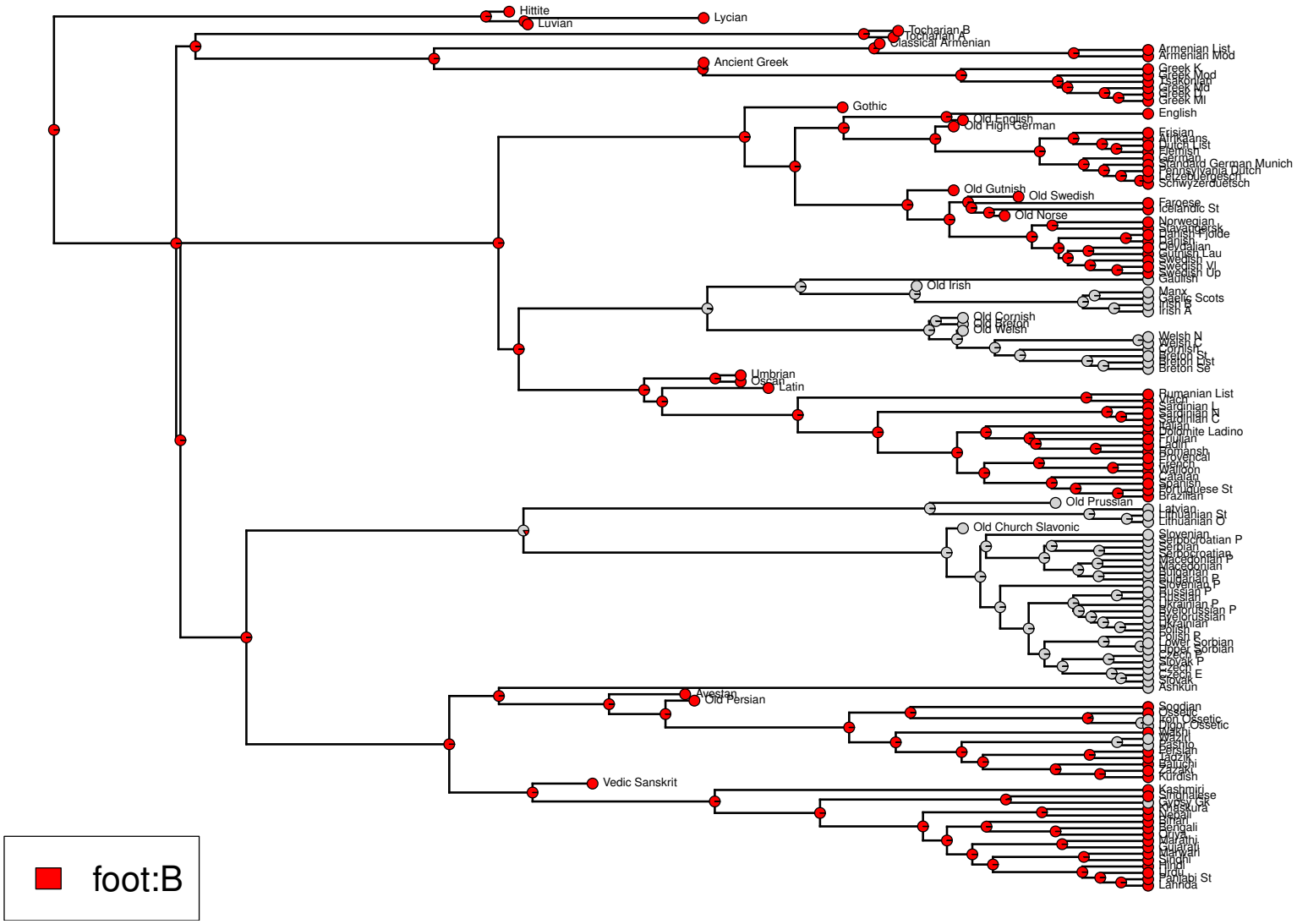
- for n cognate classes, $n^2 - n$ transition rates have to be fitted $\Rightarrow$ often insufficient data
- therefore *binary recoding* of cognate classification:
  - each cognate class is a separate feature
  - values are 1 (present in a language) and 0 (absent)
- only two rates need to be fitted per concept

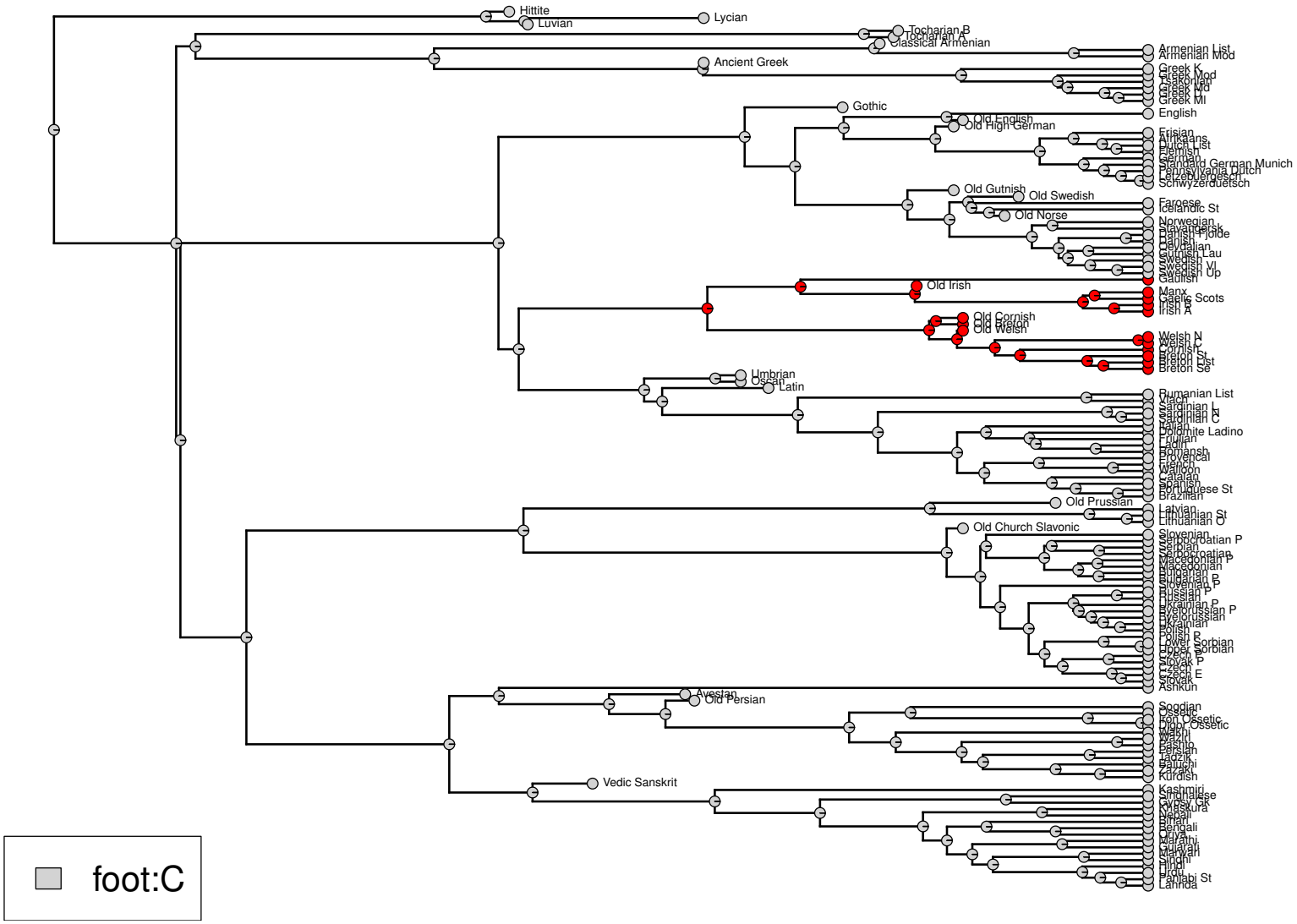
Practical issues

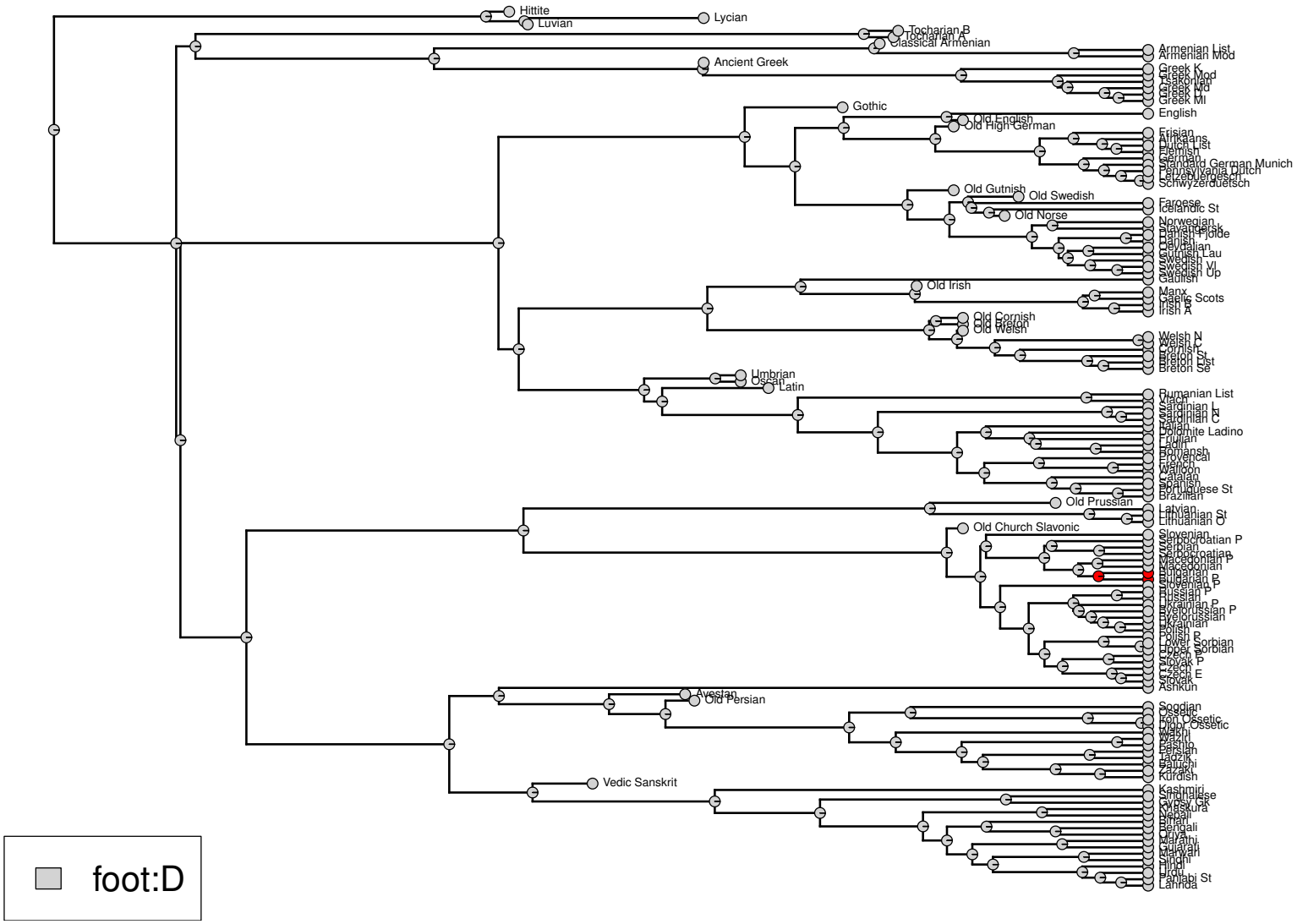
language	*cognate class*	*orthographic form*
Italian	A	*cane*
Spanish	-	*perro*
French	A	*chien*
Portuguese	A	*cão*
Romanian	A	*câine*
Greek	A, G	*κων, skilos*
English	-	*dog*
German	A	*Hund*
Danish	A	*hund*
Icelandic	A, Q	*hundur, rakki*
Norwegian	A, Q, R	*hund, rakke, bikkje*
Bulgarian	E, K	*PES, kuce*
Serbocroation	E	*pas*
Polish	E	*pies*
Russian	A	*собака*
Irish	A, I	*cú, madra*
Marathi	H	*kutrā*
Hindi	H	*kutta*
Latin	A	*canis*
Classical Greek	A	*κύων*
Gothic	A	*hunds*
Old English	A	*hund*

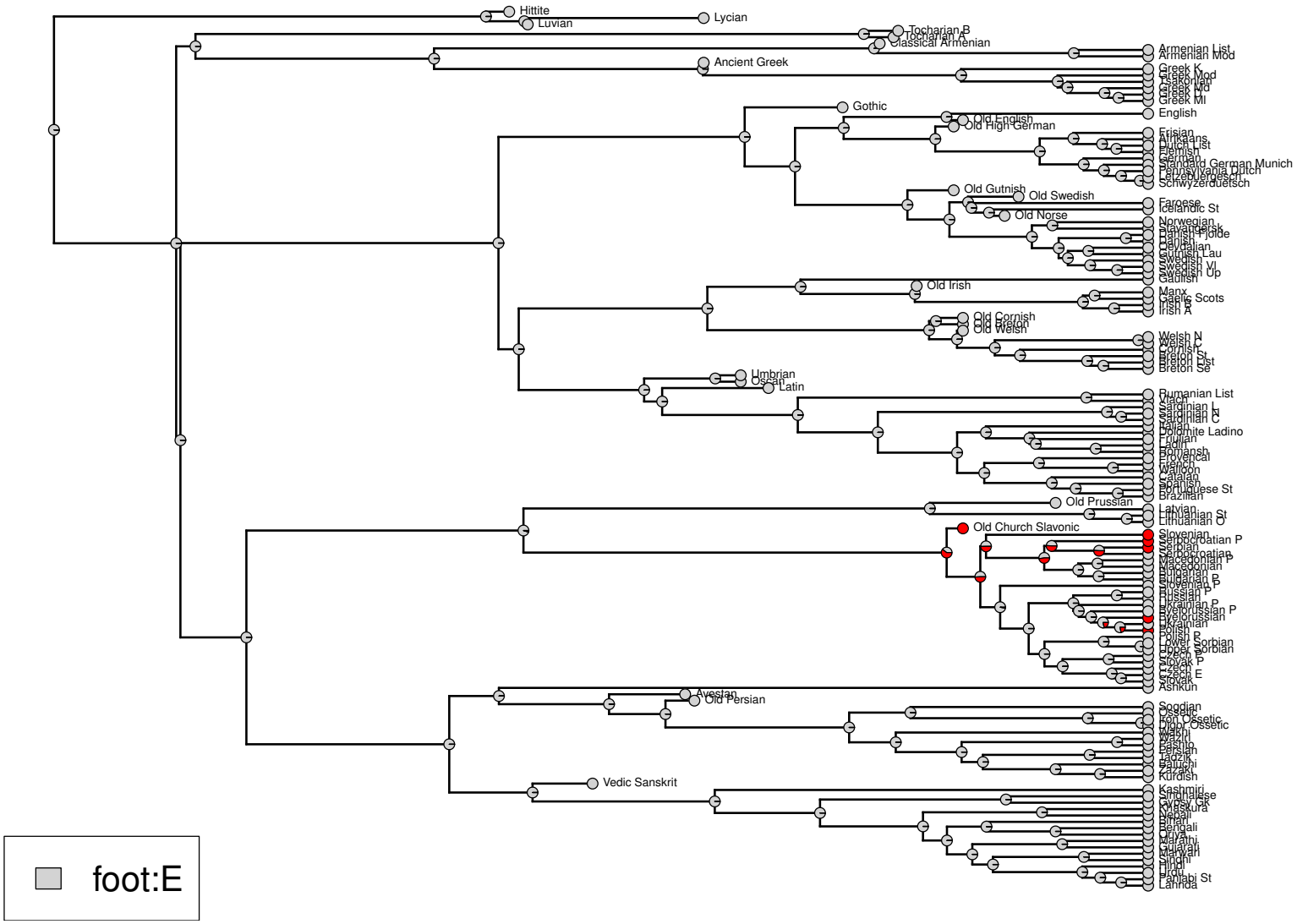
	A	*E*	*G*	*H*	*I*	*K*	*Q*	*R*
Italian	1	0	0	0	0	0	0	0
Spanish	0	0	0	0	0	0	0	0
French	1	0	0	0	0	0	0	0
Portuguese	1	0	0	0	0	0	0	0
Romanian	1	0	0	0	0	0	0	0
Greek	1	0	1	0	0	0	0	0
English	0	0	0	0	0	0	0	0
German	1	0	0	0	0	0	0	0
Danish	1	0	0	0	0	0	0	0
Icelandic	1	0	0	0	0	0	1	0
Norwegian	1	0	0	0	0	0	1	1
Bulgarian	0	1	0	0	0	0	0	0
Serbocroatian	0	1	0	0	0	1	0	0
Polish	0	1	0	0	0	0	0	0
Russian	1	0	0	0	0	0	0	0
Irish	1	0	0	0	1	0	0	0
Marathi	0	0	0	1	0	0	0	0
Hindi	0	0	0	1	0	0	0	0
Latin	1	0	0	0	0	0	0	0
Classical Greek	1	0	0	0	0	0	0	0
Gothic	1	0	0	0	0	0	0	0
Old English	1	0	0	0	0	0	0	0

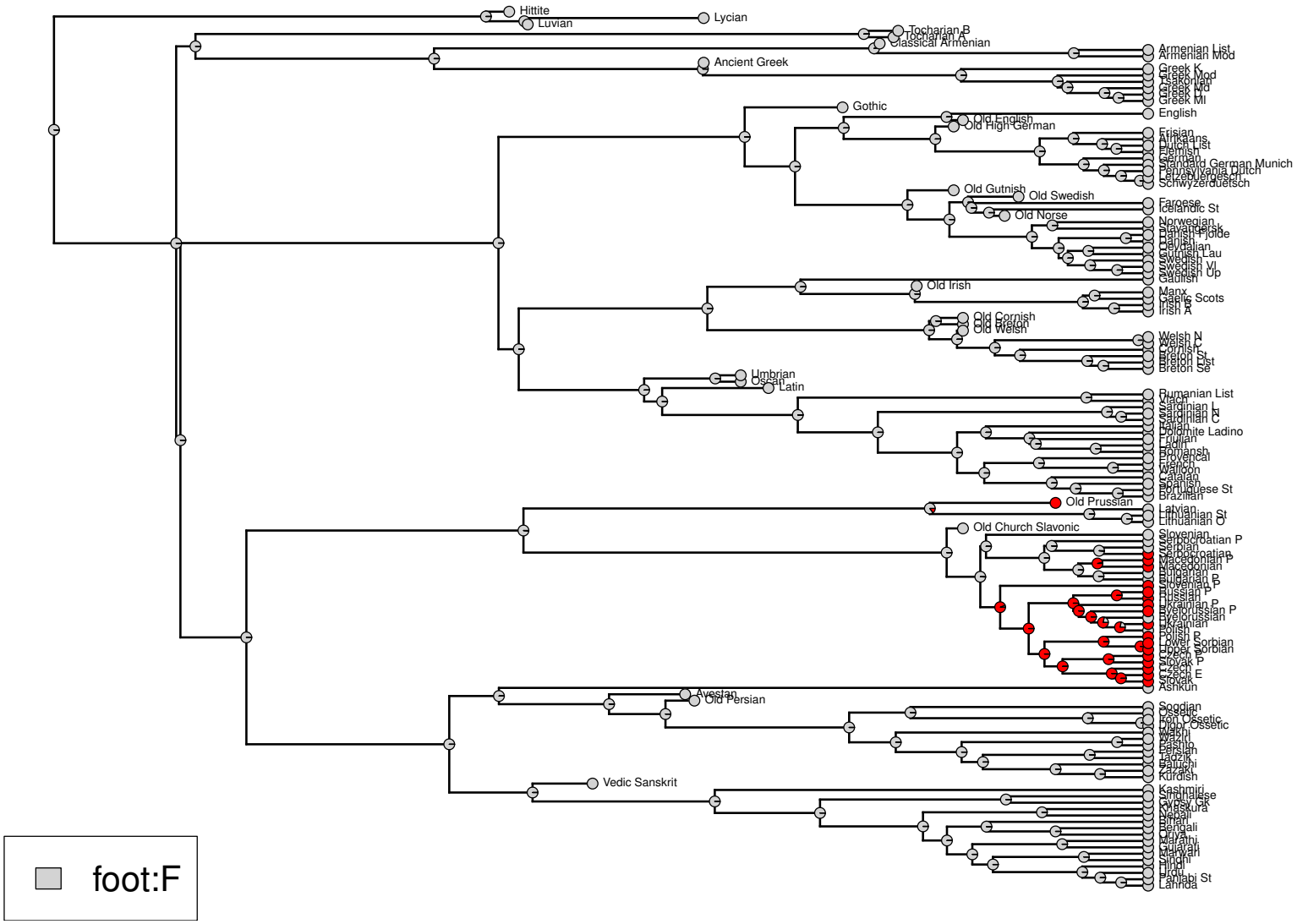

Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek A
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Dalecarlian
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Ukrainian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech P
Slovak P
Czech
Czech E
Slovak
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Ormuri
Tajik
Tajik
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:A

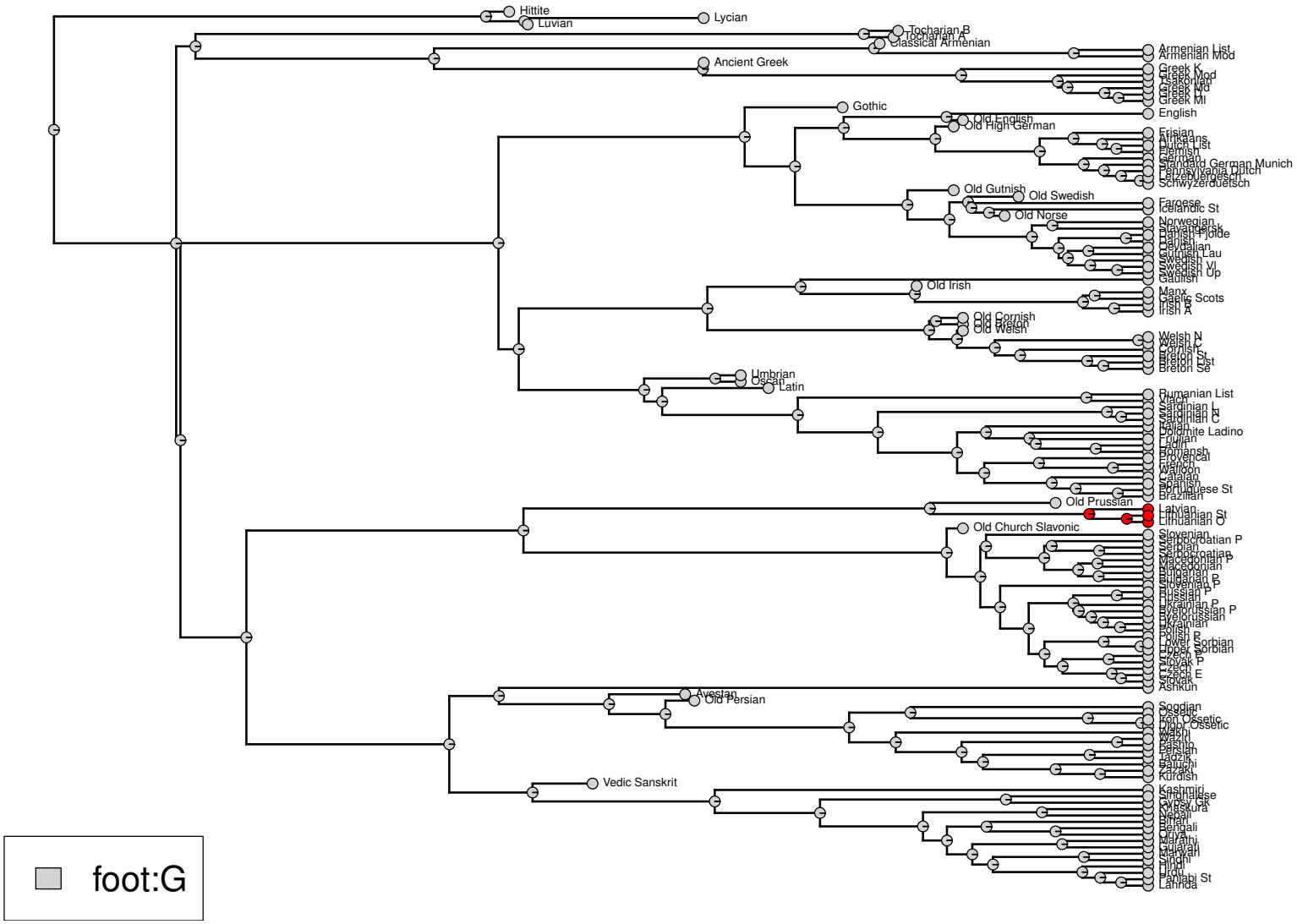

Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Ukrainian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech
Slovak P
Czech P
Czech E
Slovak
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Tadzik
Tadzik
Baluchi
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:B

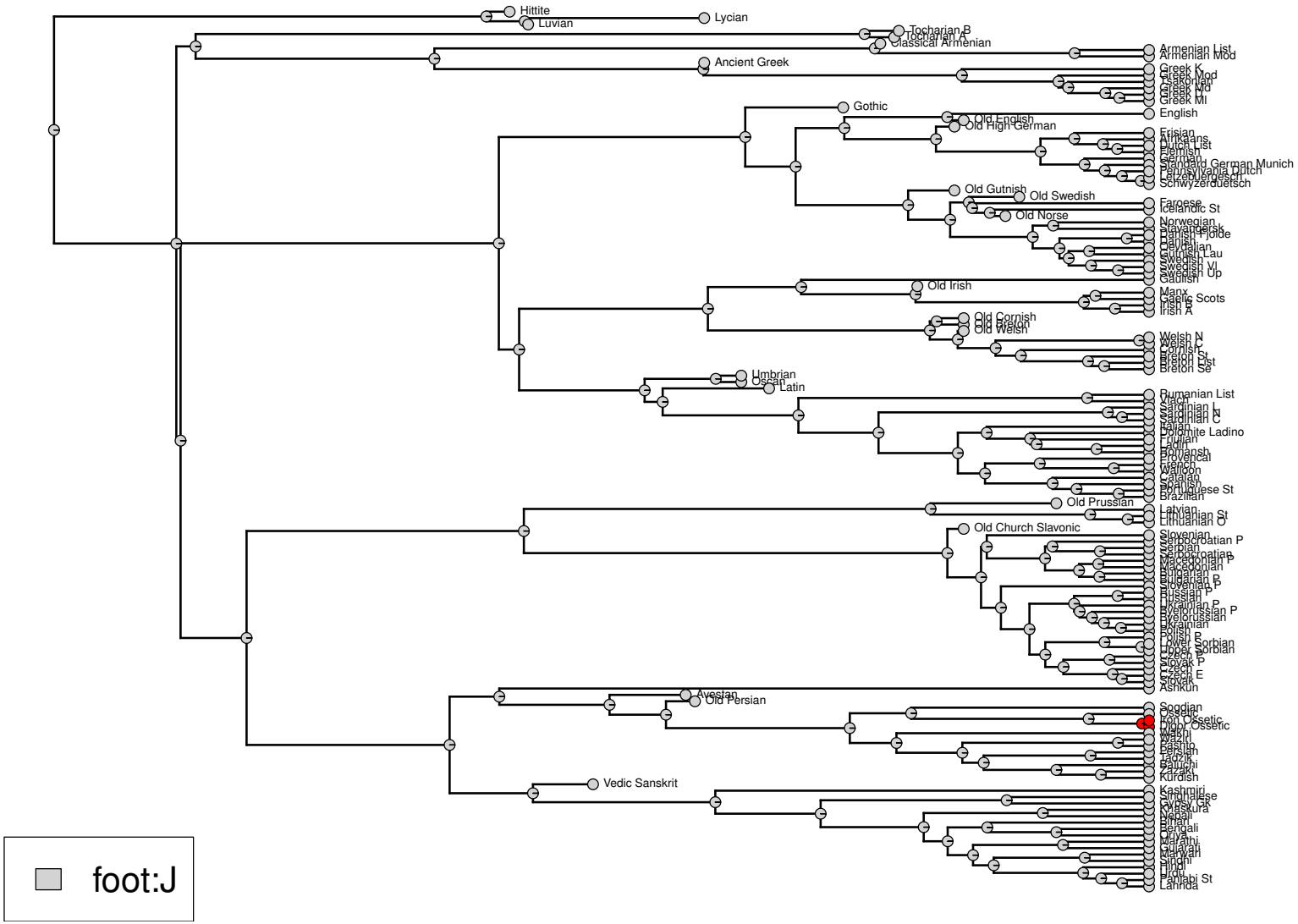

Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek A
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Dalecarlian
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Ukrainian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech
Slovak P
Czech P
Czech E
Slovak
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Gaulish
Tadzik
Tadzik
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:C


Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Gothic
Old English
Old High German
Old Gutnish
Old Swedish
Old Norse
Old Irish
Old Cornish
Old Breton
Old Welsh
Umbrian
Oscan
Latin
Old Prussian
Old Church Slavonic
Avestan
Old Persian
Vedic Sanskrit
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek A
Greek Ml
English
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Gutnish
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Manx
Gaelic Scots
Irish B
Irish A
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Latvian
Lettgalian St
Lithuanian O
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Ukrainian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech
Slovak P
Slovak
Czech E
Sorbian
Ashkun
Soggdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Gaulish
Tajik
Tajik
Baluchi
Zazaki
Kurdish
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Nepali
Bihari
Bengali
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:D


Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek A
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Dalecarlian
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Bykaulian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish B
Lower Sorbian
Upper Sorbian
Czech P
Slovak P
Czech
Czech E
Slovak
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Ormuri
Tajik
Tajik
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:E


Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Dalecarlian
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Kajkavian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech
Slovak P
Slovak
Czech E
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Ormuri
Tajik
Tajik
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:F

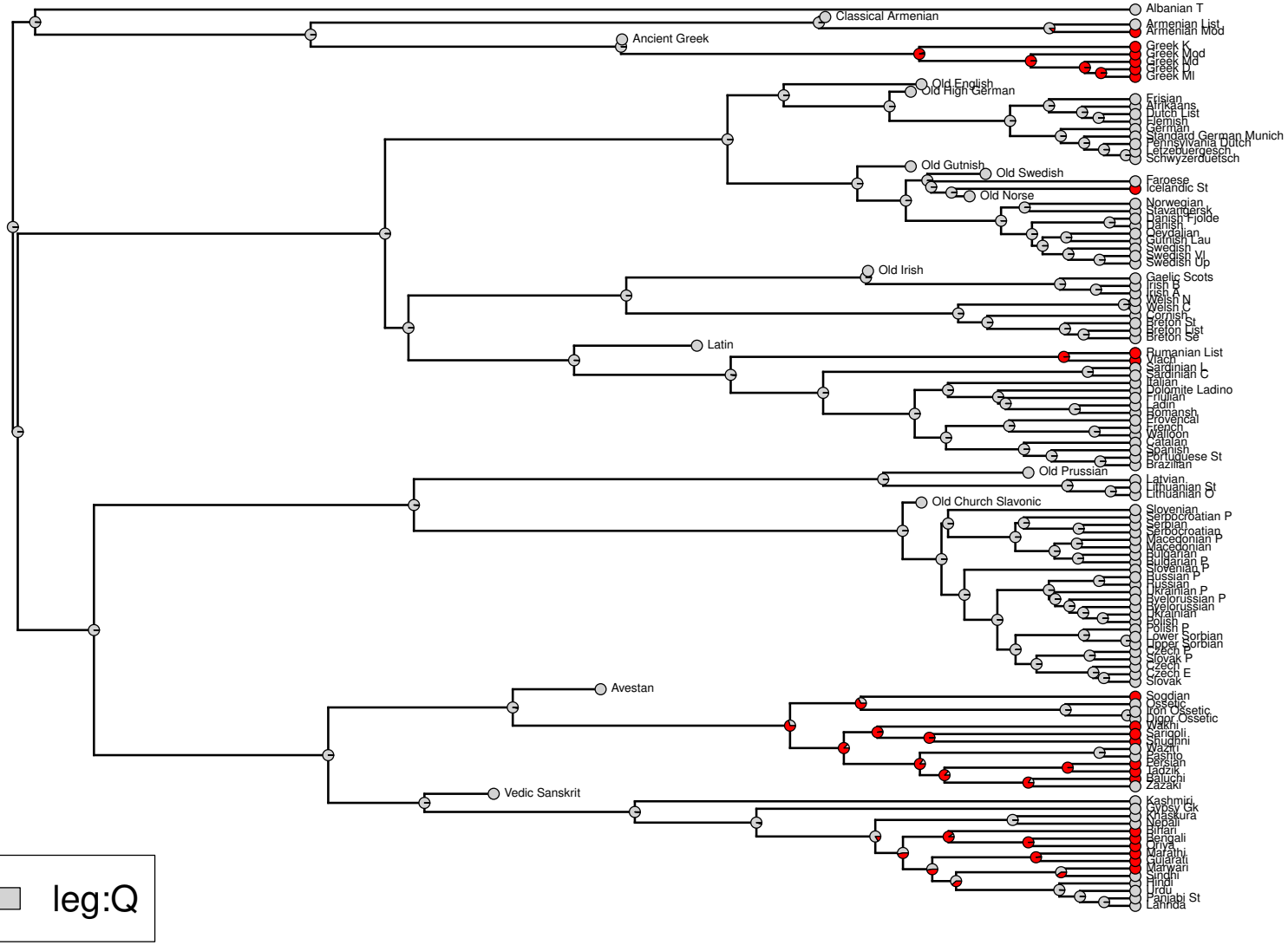

Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek L
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdutsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Dalecarlian
Gutnish Lau
Swedish L
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Kajkavian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech
Slovak P
Czech E
Slovak
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Gaulish
Tajik
Tajik
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:G


Hittite
Luvian
Lycian
Tocharian B
Tocharian A
Classical Armenian
Ancient Greek
Armenian List
Armenian Mod
Greek K
Greek Mod
Tsakonian
Greek Md
Greek A
Greek Ml
Gothic
English
Old English
Old High German
Frisian
Afrikaans
Dutch List
Flemish
German
Standard German Munich
Pennsylvania Dutch
Luxembourgish
Schwyzerdütsch
Old Gutnish
Old Swedish
Old Norse
Faroese
Icelandic St
Norwegian
Nynorsk
Danish Fjolde
Danish
Dalecarlian
Gutnish Lau
Swedish
Swedish Vl
Swedish Up
Gaulish
Old Irish
Manx
Gaelic Scots
Irish B
Irish A
Old Cornish
Old Breton
Old Welsh
Welsh N
Welsh C
Cornish
Breton St
Breton List
Breton Se
Umbrian
Oscan
Latin
Rumanian List
Vlach
Sardinian L
Sardinian N
Sardinian C
Italian
Dolomite Ladino
Friulian
Ladin
Romansh
Provençal
French
Walloon
Catalan
Spanish
Portuguese St
Brazilian
Old Prussian
Latvian
Lithuanian St
Lithuanian O
Old Church Slavonic
Slovenian
Serbocroatian P
Serbian
Serbocroatian
Macedonian P
Macedonian
Bulgarian
Bulgarian P
Slovenian P
Russian P
Russian
Ukrainian P
Byelorussian P
Byelorussian
Ukrainian
Polish
Polish P
Lower Sorbian
Upper Sorbian
Czech P
Slovak P
Czech
Czech E
Slovak
Ashkun
Avestan
Old Persian
Sogdian
Ossetic
Iron Ossetic
Digor Ossetic
Wakhi
Wakhi
Pashto
Pashto
Ladakh
Tajik
Tajik
Baluchi
Zazaki
Kurdish
Vedic Sanskrit
Kashmiri
Sinhalese
Romany Gk
Romany
Nepali
Bihari
Bengali
Oriya
Marathi
Gujarati
Marwari
Sindhi
Hindi
Urdu
Panjabi St
Lahnda
foot:J

Parallel evolution

- by definition, each cognate class can emerge only once
- transition rate $0 \rightarrow 1$ should thus be exceedingly small
- no parallel evolution $0 \rightarrow 1$ to be expected
- only expected exception: undetected borrowings
- However (see next slide...)

Parallel evolution



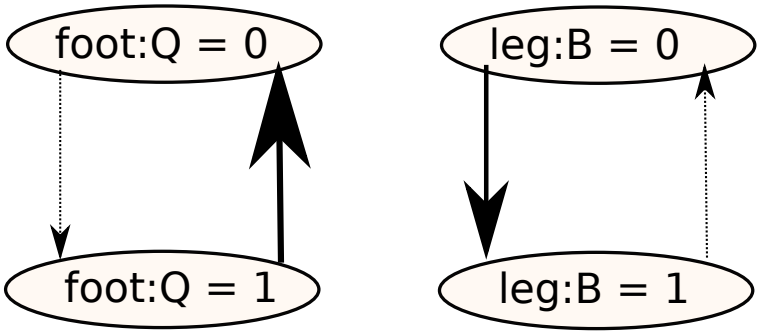
Parallel evolution

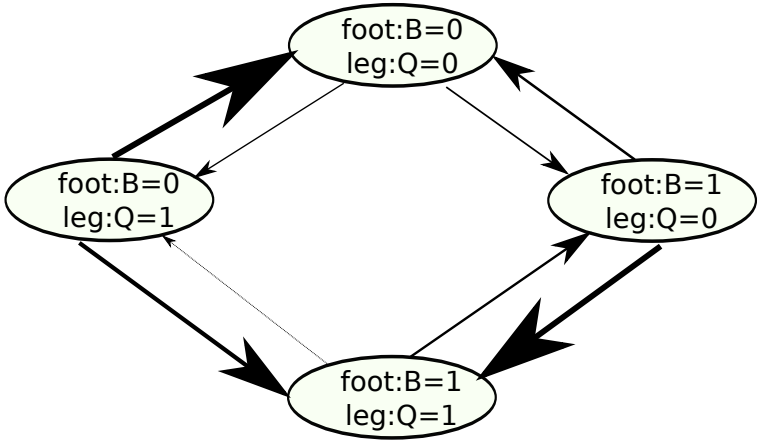
- apparently independent emergence of *leg:Q* in Greek, Icelandic, Romanian and Proto-Indo-Iranian
- traditional scholarship:
  - members of class *leq:Q* are cognate to members of *foot:B*
  - both cognate classes descend from Proto-Indo-European **ped-*
  - parallel semantic change *foot* $\rightarrow$ *leg* in several branches of Indo-European

Testing for correlated evolution

Following Pagel and Meade (2006) (see also Dunn et al. 2011), using software BayesTraits (`http://www.evolution.rdg.ac.uk/BayesTraits.html`)

- model comparison





- log Bayes Factor $\approx$ 243 in favor of second model $\Rightarrow$ strong evidence
- downsides:
  - strong evidence in favor of dependence, but weak evidence in favor of directionality
  - information about Swadesh entries of a single family does not afford scaling up

Polysemy networks

- A polysemy network (cf. Dellert and Münch 2015, see also List et al. 2013) is
  - a graph over concepts (expressed by glosses)
  - where each link represents the fact that at least one language has a lexical item which can denote both concepts (or rather: can be translated by both glosses)
  - if we count the number of instances in different language families, link strength measures cross-linguistic colexification between concepts
- Polysemy networks are a potential source of evidence for plausibility of semantic shifts:
  - intuition: every instance of semantic shift needs to pass an intermediate stage where the word in question is polysemous
  - a massively cross-linguistic polysemy network provides us with a snapshot of semantic evolution in action

  $\Rightarrow$ Polysemy networks might be a good model for modeling semantic change in computational historical linguistics!

Polysemy networks

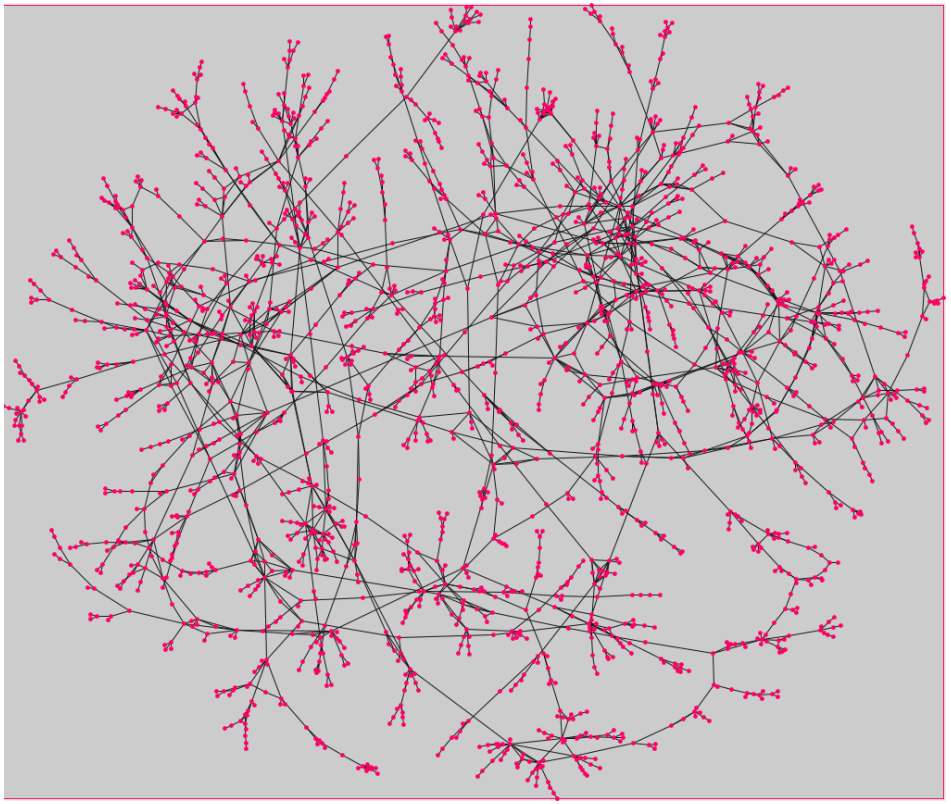
- **Question:** How well does a polysemy network model attested semantic shifts?
- **Data:**
  - a massively cross-linguistic polysemy network (TUE) developed in Tübingen within EVOLAEMP (Dellert, 2014)
  - the Catalogue of Semantic Shifts (ZAL) maintained by the Russian academy of sciences (Zalizniak, 2008; Zalizniak et al., 2012)
- **Method:** for each attested change, determine the length of the shortest paths between the relevant concepts in the network

Tübingen Polysemy Network (TUE)

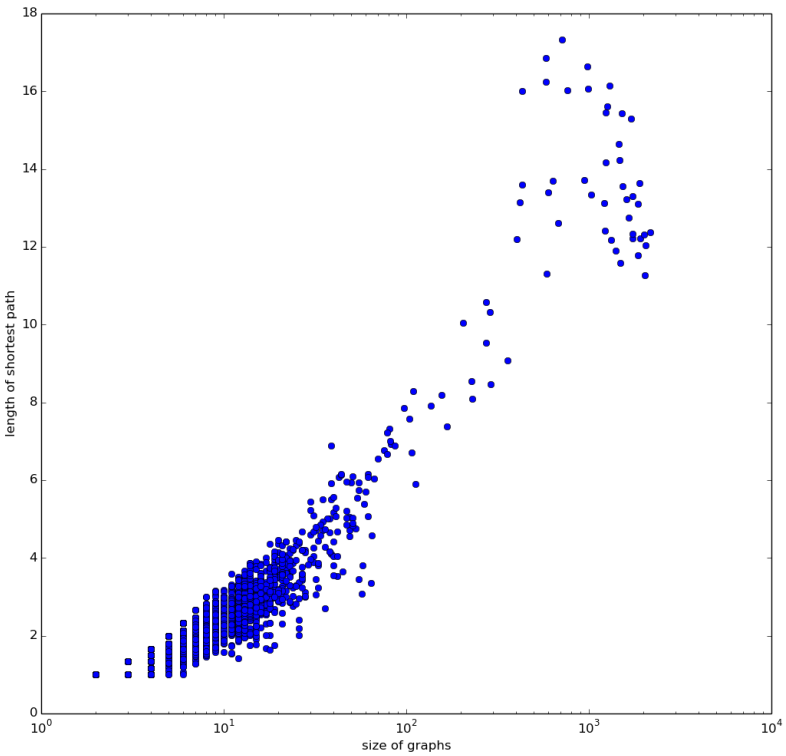
- is derived from a large German-based dictionary database containing more than 750,000 entries in 114 languages
- has near-complete coverage of a 1,000-item list of basic concepts in 88 languages
- is built on more than 10,000 translations each for 31 languages from 10 primary language families of Eurasia
- is available to other researchers upon request

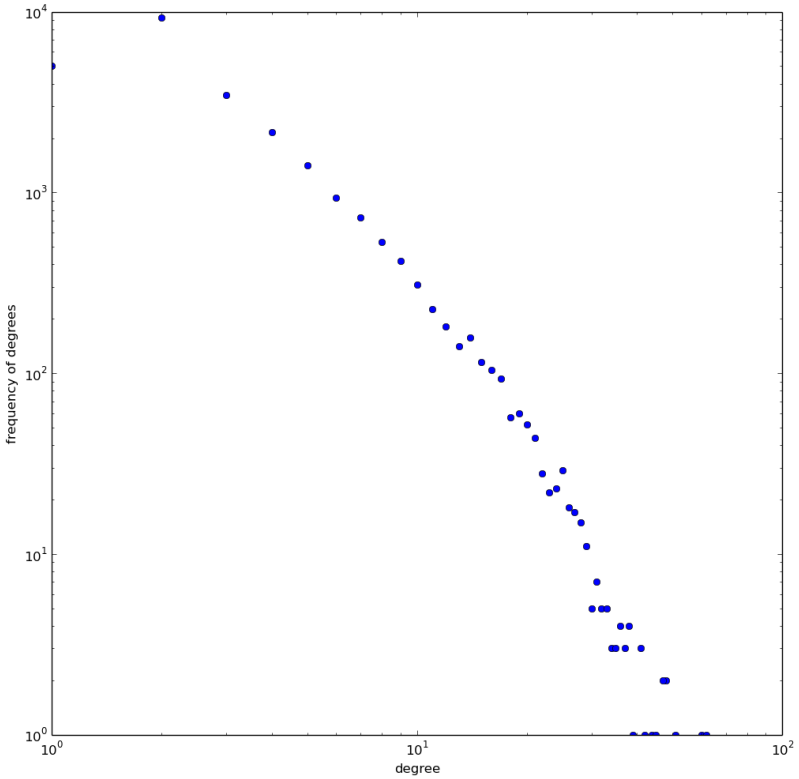
Tübingen Polysemy Network (TUE): Some data

- 32,653 concepts and 47,647 links
- a central connected component of 13,073 nodes (!)
- 1,640 smaller components and 14,391 unconnected islands
- an average of 2.4 neighbors for each node
- a typical shortest path length of 8



Small world characteristics





The Catalogue of Semantic Shifts

- crosslinguistically recurring shifts collected manually by experts
- no restrictions or preferences in the range of meanings or languages (de facto: focus on Eurasian)
- contained 3,650 semantic shifts at the time of publication
- classified into six types, each supported by up to 40 realizations
- publicly available via a web interface: `http://semshifts.iling-ran.ru/`

Procedure

- English version of the catalogue was extracted from the website
- considerable semi-automated cleanup work
- the total number of shift pairs in the result: 6,174

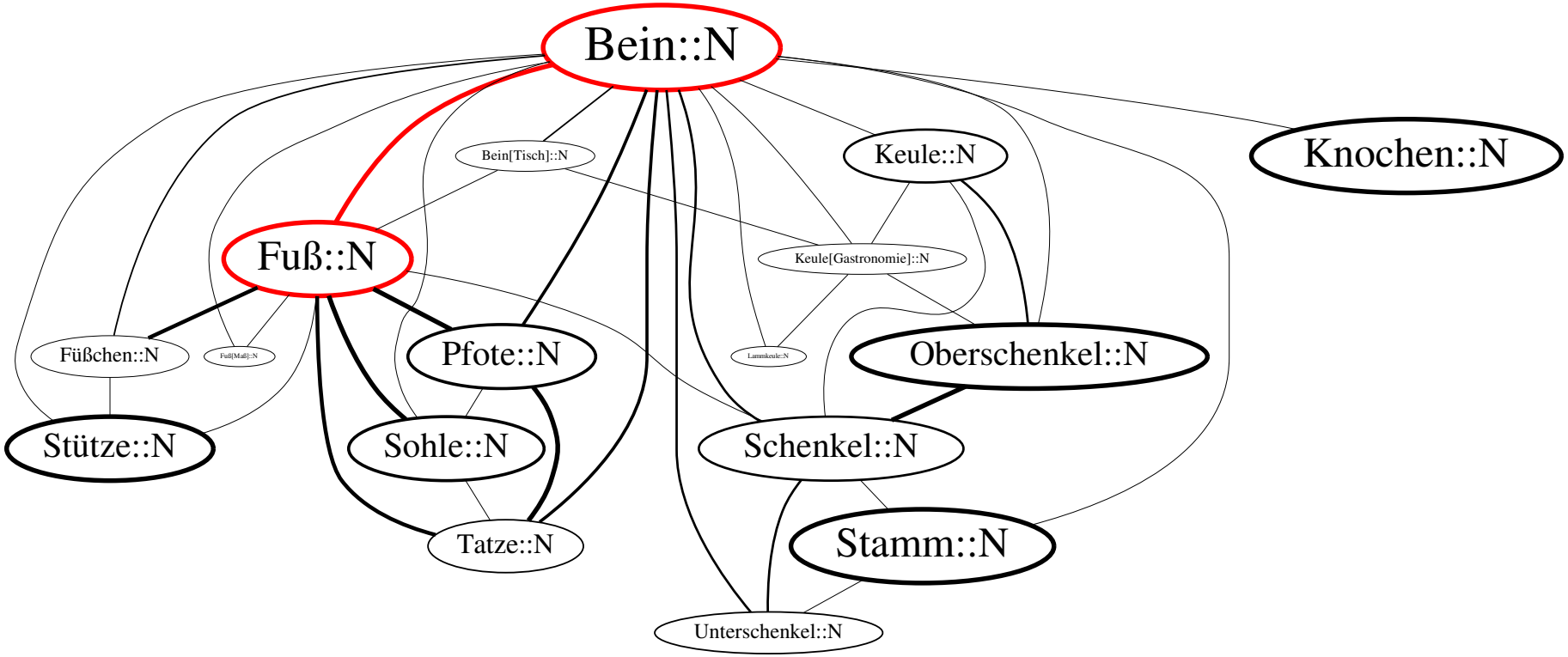
Preperatory steps

- metalanguages of the catalogue: English and Russian
- primary language of the Tübingen Polysemy Network: German
- electronic English-German dictionary needed to be useed as an intermediary for comparing the links in both resources

Experiment

- evaluate pairs of English glosses in the catalogue against the network by determining **the shortest paths in the network between any pair of German translation equivalents**
- do this separately for each type, and compare recall

Example


Bein::N
Knochen::N
Keule::N
Bein[Tisch]::N
Fuß::N
Keule[Gastronomie]::N
Füßchen::N
Fuß[Maß]::N
Pfote::N
Lammkeule::N
Oberschenkel::N
Stütze::N
Sohle::N
Schenkel::N
Tatze::N
Stamm::N
Unterschenkel::N

Results

Types of Semantic Shifts in the Catalogue

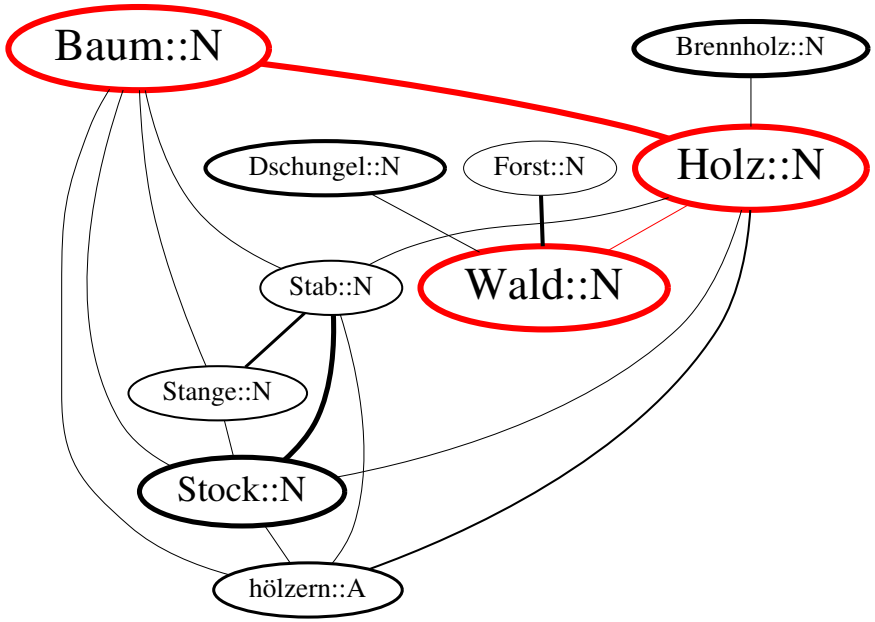
- Polysemy
- Diachronic Semantic Evolution
- Syncretism
- Cognates
- Borrowings
- Morphological Derivation

Only Diachronic Semantic Evolution, Cognates, and Borrowings contain instances of semantic change in the stricter sense.

Results

Polysemy

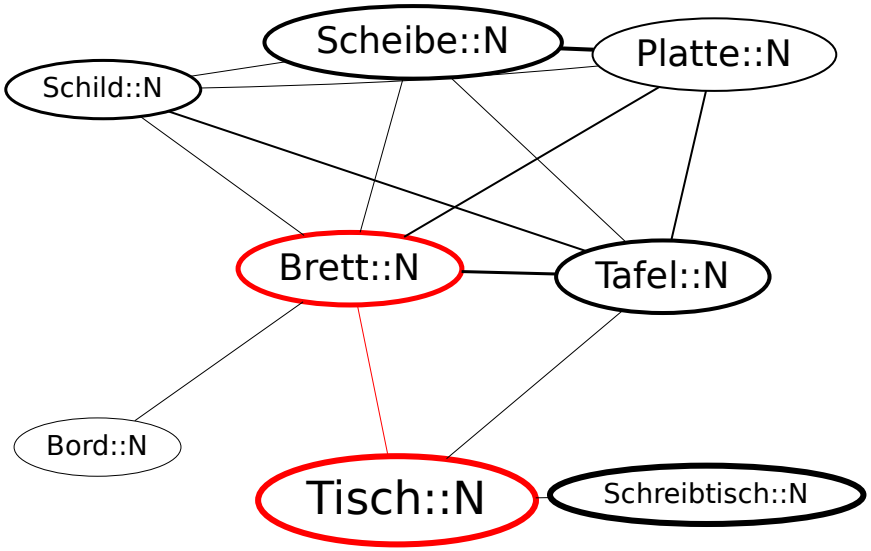
- **synchronic polysemy:** A and B are meanings of a polysemous word
- ZAL: tree — forest / TUE: Baum ‘tree’ — Holz ‘wood’ — Wald ‘forest’
- Example: Tuvan *yjash* ‘tree; forest’



Results

Diachronic semantic evolution

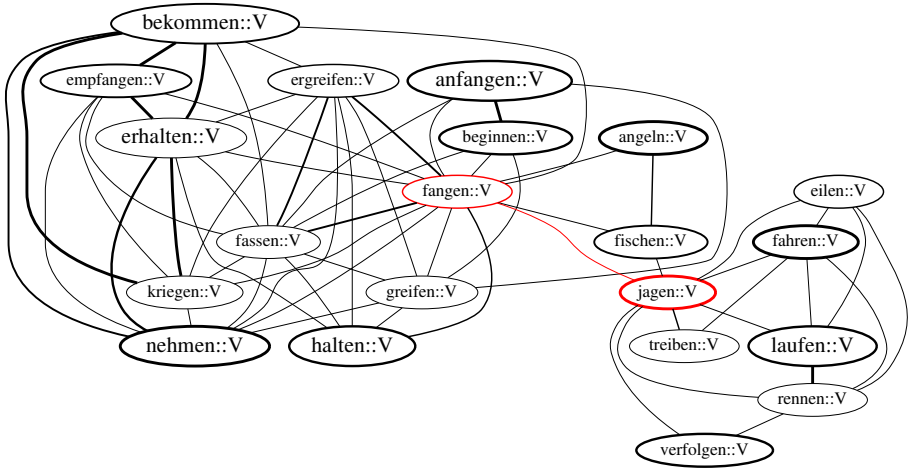
- Diachronic semantic evolution of a word in one language or from an ancestor language to a descendant language
- ZAL: board, plank — table, desk / TUE: Brett 'board' — Tisch 'table'
- Example: Latin *tabula* 'plate, tablet' → Italian *tavola* 'table'



Results

Diachronic semantic evolution

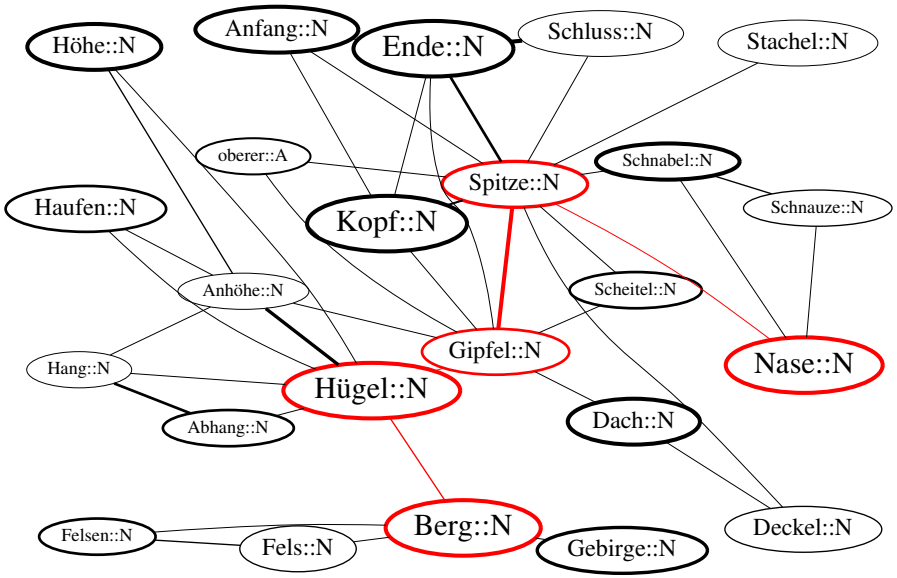
- Diachronic semantic evolution of a word in one language or from an ancestor language to a descendant language
- ZAL: catch — hunt / TUE: jagen ‘hunt’ — fangen ‘catch’
- Example: Latin *capto* ‘various meanings related to catching’ → Italian *cacciare* ‘to hunt’



Results

Diachronic semantic evolution

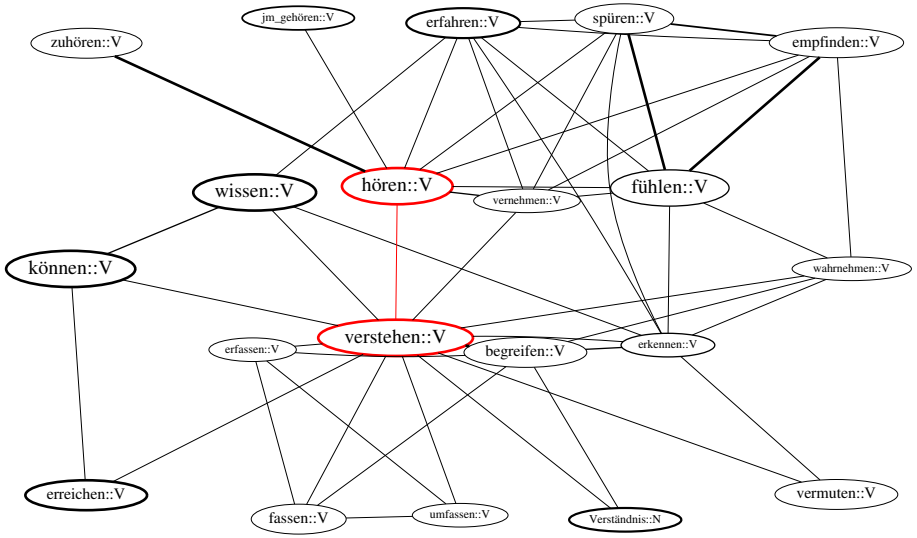
- Diachronic semantic evolution of a word in one language or from an ancestor language to a descendant language
- ZAL: mountain — nose / TUE: Berg ‘mountain’ — Hügel ‘hill’ — Gipfel ‘summit’ — Spitze ‘tip’ — Nase ‘nose’
- Example: Hungarian *orr* ‘nose’ ← ‘top, mountain’



Results

Cognates

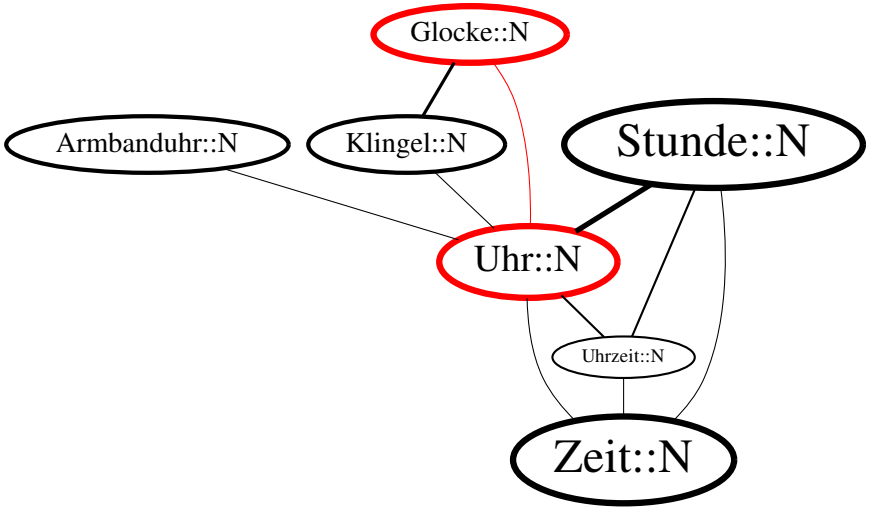
- meanings A and B belong to words of two sister languages which have developed from the same root in their common ancestor
- ZAL: hear — understand / TUE: hören ‘hear’ - verstehen ‘understand’
- Example: French *entendre* ‘to hear’ — Spanish *entender* ‘to understand, to be knowledgeable’



Results

Borrowings

- B is the meaning of a borrowed word, while A is the meaning of its source in the donor language
- ZAL: bell — watch/clock / TUE: Glocke 'bell' - Uhr 'watch/clock'
- Example: Medieval Latin *clocca* 'bell' → English *clock* 'clock, watch'



Results

Table: Shifts in the catalogue covered by the polysemy network.

Semantic shift type	Number of instances	No path	Path length 1 or 2	Path length 3 and more
Polysemy	2315	20.6 %	35.0 %	44.4 %
Semantic Evolution	107	26.2 %	33.6 %	40.2 %
Morphological Derivation	597	28.5 %	29.0 %	42.5 %
Syncretism	43	25.6 %	55.8 %	18.6 %
Borrowing	58	31.0 %	41.4 %	27.6 %
Cognates	393	23.7 %	37.9 %	38.4 %

Towards detecting directional biases in meaning change

- **Assumptions**
  - meaning change proceeds along the edges of the polysemy graph
  - for each possible transition, there is a fixed transition rate, across language families and times $\rightarrow$ ultimately founded in cognition
- **Research questions**
  - Can we estimate those rates from data?
  - Can we detect directional biases?
- **Requirement**
  - Many data — across language families and beyond the Swadesh list

The NorthEuraLex database

- Massive data collection effort of the Tübingen EVOLAEMP project
- (currently) translations of 1,017 concepts into 103 (mostly) Northern Eurasian languages (cf. Dellert, 2015)
- everything transcriped in IPA
- (so far) no manual cognate coding

	Auge::N	Ohr::N	Nase::N	Mund::N	Zahn::N	Zunge::N	Lippe::N	Wange::N \
iso								
fin	silmæ	kɔrʊa	nɛnæ	su:	ham:as	kiɛli	hu:li	pɔski
krl	silmæ	kɔrva	nɛnæ	su:	ham:as	kiɛli	hu:li	nʊala
olo	silmʏ	kɔrvʊ	nɛnæ	su:	ham:as	kiɛli	hu:li	ʃɔk:ʊ
vep	silʲm	kɔrv	nɛna	sʊ	pi-hambaz	kɛlʲ	hʊlʲ	mɔdpɔliʃk
ekk	sʲilm	kɤrv	nʲina	su:	ham:as	ke:l	hu:l	pɤsk

	Gesicht::N	Stirn::N	...	malen::V \
iso			...	
fin	na:ma-kasʊɔt	ɔt͡sa	...	ma:lata
krl	nækœ-mʊɔtɔ	ɔt͡ʃ:a	...	mʊalata
olo	nægœ	ɔt͡ʃ:ʊ	...	rɛsʊija-pi:rʊstʊa
vep	mɔd	ɔt͡s	...	pirta
ekk	nægu	otsmik	...	ma:lʲima

	zeichnen::V	schreiben::V	besitzen::V	kaufen::V	verkaufen::V \
iso					
fin	pi:rtæ:-pi:rʊsta:	kirjɔit:a:	ɔmista:	ɔsta:	my:dæ
krl	pi:rʊstʊa	kirjʊt:ʊa	ɔmistʊa	ɔstʊa	myyvvæ
olo	rɛsʊija-pi:rʊstʊa	kirjʊt:ʊa	ɔmistʊa	ɔstʊa	mʏvvæ
vep	pirta	kirjʊtada	ɔmiʃtada	ɔst:a	mæda
ekk	jo:nʲistama	kirjutama	omama	ostma	my:ma

	bezahlen::V	zahlen::V	beherrschen::V	ertragen::V
iso				
fin	maksa:	maksa:	hal:ita	kɛstæ:
krl	maksʊa	maksʊa	isæn:œija	kɛstyæ-siɛtyæ
olo	maksʊa	maksʊa	iʒændœijæ	kærziɛ-kɛstiæ
vep	maksta	maksta	valdɔita	kant:a
ekk	maksma	maksma	valʲitsʲema	taluma

Automatic cognate classification

- clustering of the words for a given concept
- heuristic technique based on phonetic similarity
- so far only rough approximation to expert classification

Language	Word	Cognate class
Dra.Southern Dravidian.Kannada	paada	Fuß:12
IE.Baltic.Lithuanian	peeda	Fuß:12
IE.Greek.Greek	po8i	Fuß:12
IE.Indic.Hindi	pEEr	Fuß:12
IE.Indic.Hindi	paaw	Fuß:12
IE.Iranian.Ossetic	fad	Fuß:12
IE.Iranian.Pashto	psa	Fuß:12
IE.Iranian.Persian	paa	Fuß:12
IE.Romance.French	pye	Fuß:12
IE.Romance.Italian	piede	Fuß:12
IE.Romance.Portuguese	pE	Fuß:12
IE.Romance.Spanish	pye	Fuß:12
Kor.Korean.Korean	pal	Fuß:12
Ura.Permic.Udmurt	p3d	Fuß:12
Alt.Tungusic.Manchu	pothxo	Fuß:6
Brs.Burushaski.Burushaski	ut	Fuß:6
Brs.Burushaski.Burushaski	utis	Fuß:6
IE.Armenian.Armenian	votkh	Fuß:6
IE.Germanic.Danish	foo78	Fuß:6
IE.Germanic.Dutch	vut	Fuß:6
IE.Germanic.English	fut	Fuß:6
IE.Germanic.German	fuus	Fuß:6
IE.Germanic.Icelandic	foutir	Fuß:6
IE.Germanic.Norwegian	fuut	Fuß:6
IE.Germanic.Swedish	fuut	Fuß:6

Getting a handle on meaning change

- two cognate classes A and B, for concepts c_A and c_B, are merged if
  - c_A and c_B are neighbors in the polysemy graph
  - in at least one language, A and B are instantiated by the same phonetic string

Language	Word	Cognate class
Dra.Southern Dravidian.Kannada	paada	foot
IE.Baltic.Lithuanian	peeda	foot
IE.Greek.Greek	po8i	foot
IE.Indic.Hindi	pEEr	foot
IE.Indic.Hindi	paaw	foot
IE.Iranian.Ossetic	fad	foot
IE.Iranian.Pashto	psa	foot
IE.Iranian.Persian	paa	foot
IE.Romance.French	pye	foot
IE.Romance.Italian	piede	foot
IE.Romance.Portuguese	pE	foot
IE.Romance.Spanish	pye	foot
Kor.Korean.Korean	pal	foot
Ura.Permic.Udmurt	p3d	foot
Alt.Tungusic.Nanai	b3gdi	Bein:9
IE.Greek.Greek	po8i	Bein:9
IE.Indic.Bengali	pa	Bein:9
IE.Iranian.Pashto	psa	Bein:9
IE.Iranian.Persian	paa	Bein:9
Krt.Kartvelian.Georgian	phExi	Bein:9
Ura.Permic.Udmurt	p3d	Bein:9

[0]see also (Dellert, 2016) for a different approach

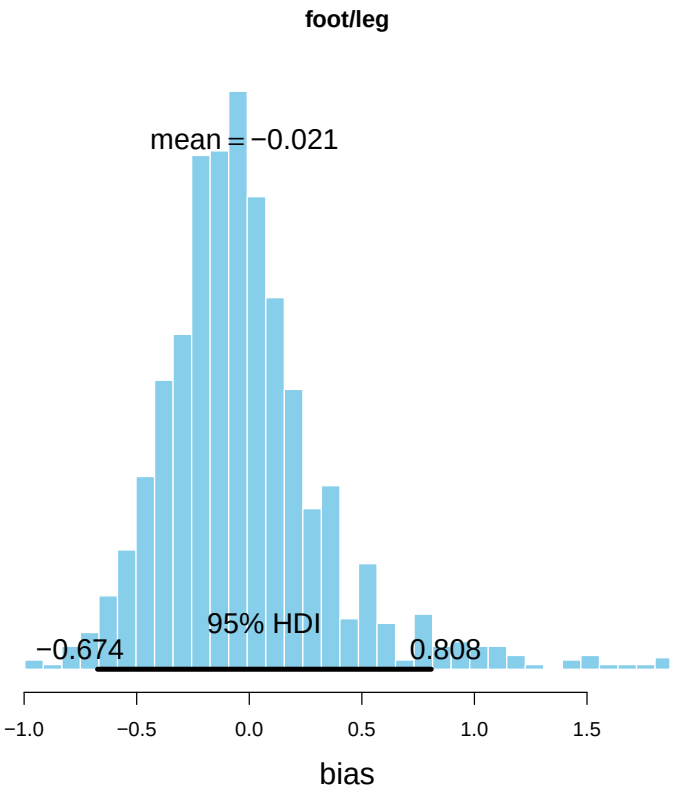
Getting a handle on meaning change

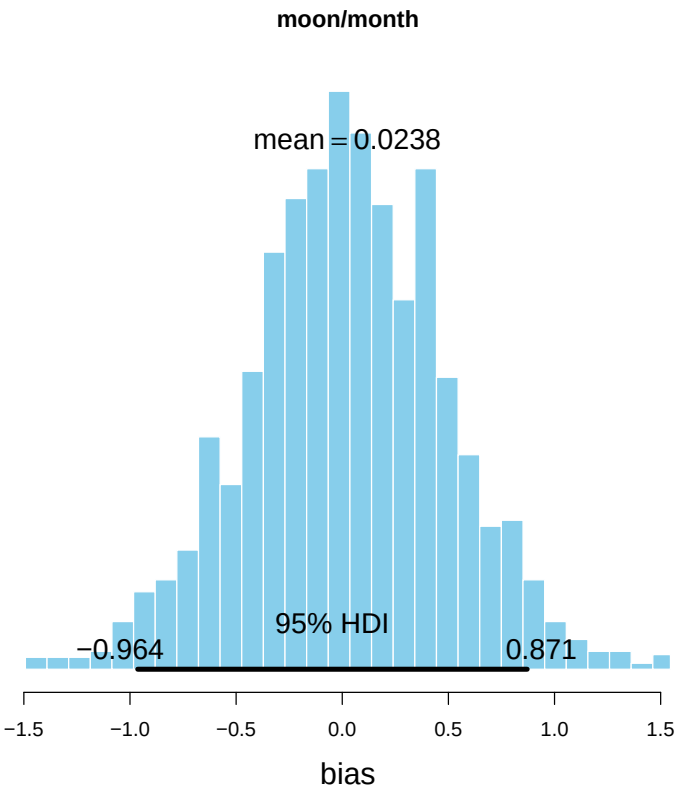
- treat each merged cognate class as feature and meaning as value
- fit a model with two parameters:
  - transition rate *foot* → *leg*
  - transition rate *leg* → *foot*

Language	Word	meaning	
Alt.Tungusic.Nanai	b3gdi		leg
Dra.Southern Dravidian.Kannada	paada	foot	
IE.Baltic.Lithuanian	peeda	foot	
IE.Greek.Greek	po8i	foot	leg
IE.Indic.Bengali	pa		leg
IE.Indic.Hindi	pEEr/paaw	foot	
IE.Iranian.Ossetic	fad	foot	
IE.Iranian.Pashto	psa	foot	leg
IE.Iranian.Persian	paa	foot	leg
IE.Romance.French	pye	foot	
IE.Romance.Italian	piede	foot	
IE.Romance.Portuguese	pE	foot	
IE.Romance.Spanish	pye	foot	
Kor.Korean.Korean	pal	foot	
Krt.Kartvelian.Georgian	phExi		leg
Ura.Permic.Udmurt	p3d	foot	leg

Results

- no evidence of directional bias for
  - *foot* ↔ *leg*
  - *moon* ↔ *months*

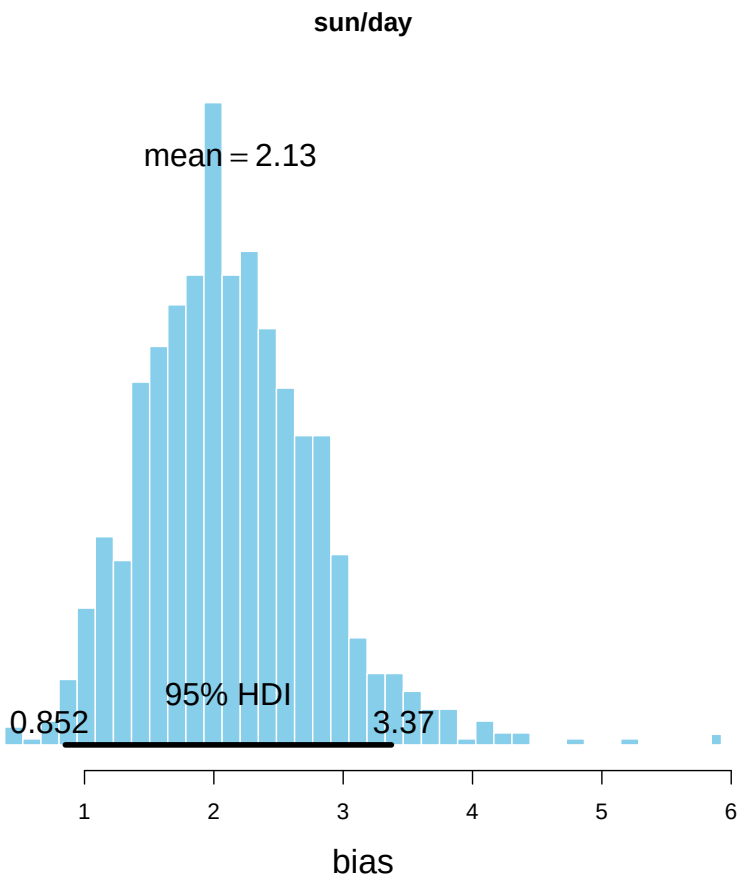


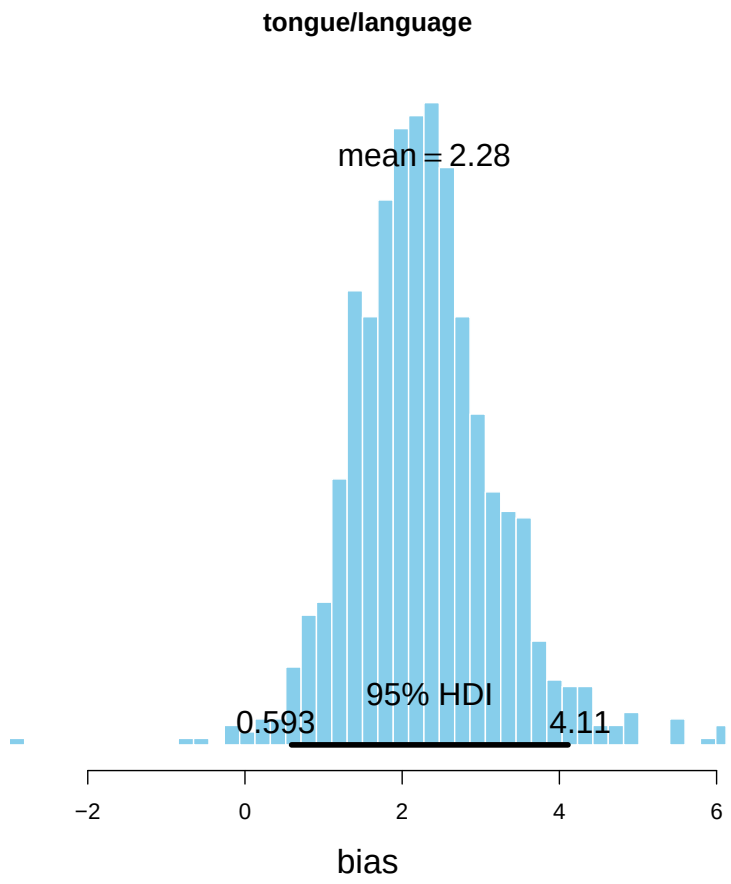


[0]model fitted using BayesTraits (Pagel and Meade, 2014); tree sample generated with BayesPhylogenies (Pagel and Meade, 2015)

Results

- clear evidence of directional bias for
  - *sun* ↔ *day*
  - *tongue* ↔ *language*





Conclusion

- method suffers from poor quality of cognate classification
- results therefore to be taken with a pound of salt
- still, proof of concept that crosslinguistic lexical data provide a handle on studying laws of meaning evolution

Great potential for combining cognitive with phylogenetic methods!

Atkinson, Q. D. and R. Gray (2005). Curious parallels and curious connections — phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, **54**(4):513–526.

Bouckaert, R., P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, **337**(6097):957–960.

Crowley, T. and C. Bowern (2010). *An introduction to historical linguistics*. Oxford University Press, Oxford.

Dellert, J. (2014). Evaluating cross-linguistic polysemies as a model of semantic change for cognate finding. Proceedings of the Workshop on semantic technologies for research in the humanities and social sciences (STRiX). November 24-25, Gothenburg, Sweden.

Dellert, J. (2015). Compiling the Uralic dataset for NorthEuraLex, a lexicostatistical database of Northern Eurasia. Proceedings of the First International Workshop on Computational Linguistics for Uralic Languages. January 16, Tromsø, Norway.

Dellert, J. (2016). Using causal inference to detect directional tendencies in semantic evolution. In *11th International Conference on the Evolution of Language (EvoLang 2016)*. New Orleans.

Dellert, J. and A. Münch (2015). Evaluating the potential of a large-scale polysemy network as a model of plausible semantic shifts. In H. Baayen, G. Jäger, M. Köllner, J. Wahle, and A. Baayen-Oudshoorn, eds., *Proceedings of QITL-6*. University of Tübingen, Tübingen.

Dunn, M., S. J. Greenhill, S. Levinson, and R. D. Gray (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, **473**(7345):79–82.

Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Inc. Publishers, Sunderland.

List, J.-M., A. Terhalle, and M. Urban (2013). Using network approaches to enhance the analysis of cross-linguistic polysemies. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pp. 347–353. Association for Computational Linguistics, Potsdam, Germany.

Pagel, M., Q. D. Atkinson, A. S. Calude, and A. Meade (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences*, **110**(21):8471–8476.

Pagel, M. and A. Meade (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist*, **167**(6):808–825.

Pagel, M. and A. Meade (2014). BayesTraits 2.0. software distributed by the authors.

Pagel, M. and A. Meade (2015). BayesPhylogenies 2.0. software distributed by the authors.

Pereltsvaig, A. and M. W. Lewis (2015). *The Indo-European Controversy*. Cambridge University Press.

Zalizniak, A. A. (2008). A catalogue of semantic shifts: Towards a typology of semantic derivation. In M. Vanhove, ed., *From polysemy to semantic change. Towards a typology of lexical semantic associations*, pp. 217–232. John Benjamins, Amsterdam and Philadelphia.

Zalizniak, A. A., M. Bulakh, D. Ganenkov, I. Gruntov, T. Maisak, and M. Russo (2012). The catalogue of semantic shifts as a database for lexical semantic typology. *Linguistics*, **50**(3):633–669.