

Wie die Bioinformatik hilft, Sprachgeschichte zu rekonstruieren

Gerhard Jäger

SfS, Tübingen

24. November 2011

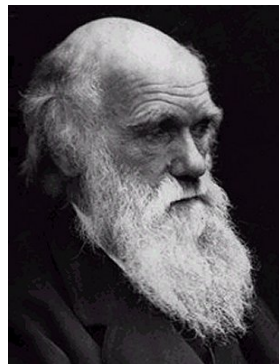


Sprachwandel und Evolution

„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)



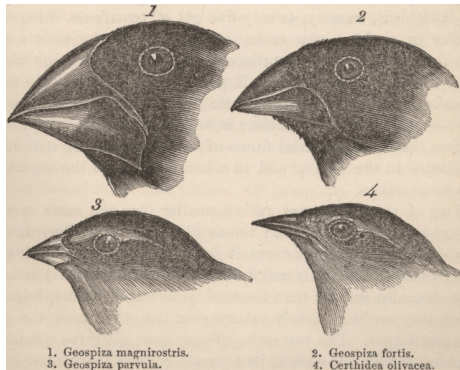
Sprachwandel und Evolution

Vater Unser im Himmel, geheiligt
werde Dein Name

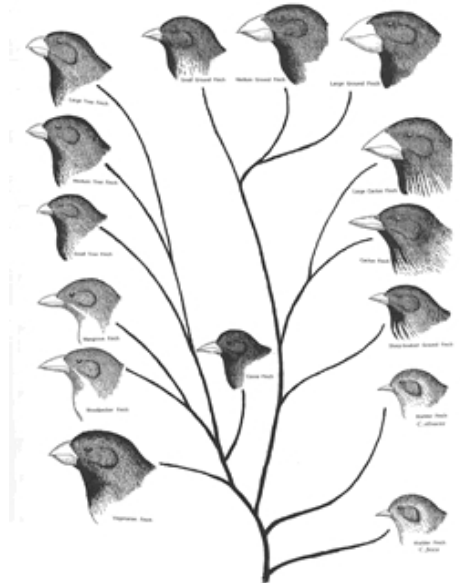
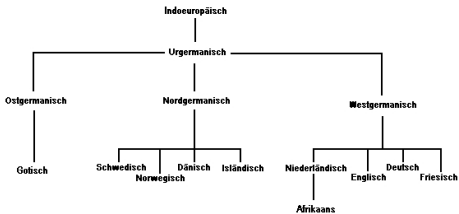
Onze Vader in de Hemel, laat Uw
Naam geheiligt worden

Our Father in heaven, hallowed be
your name

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



Sprachwandel und Evolution



Sprachwandel und Evolution

Mittelhochdeutsch:

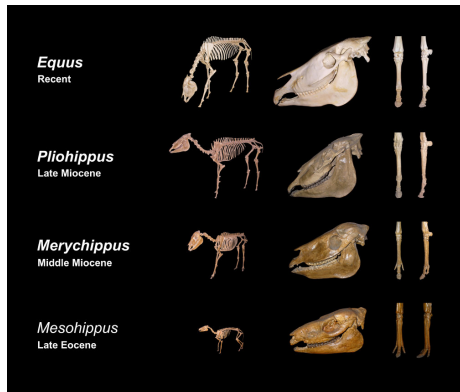
Got vater unser, dâ du bist in dem
himelrîche gewaltic alles des dir ist,
geheiliget sô werde dîn nam

Althochdeutsch:

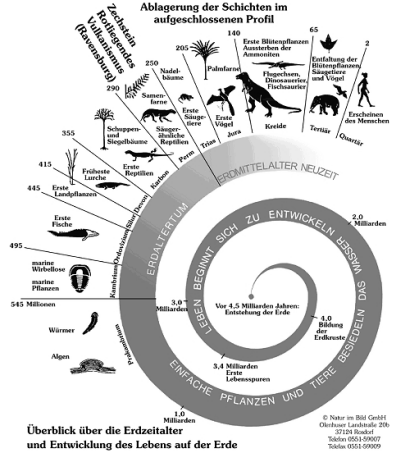
Fater unser thû thâr bist in himile, si
giheilagôt thîn namo

Gotisch:

Atta unsar þu in himinam, weihnai
namo þein



Sprachwandel und Evolution



Höherentwicklung im Sprachwandel

Pidgin- und Kreolsprachen

- eine Indianerin zu einem weißen Verehrer in Pidgin-English:

You silly. You weak. You baby-hand. No catch horse. No kill buffalo. No good but for sit still—read book.

- Satz aus dem Sranan, einer Englisch-basierten Kreolsprache aus Surinam:

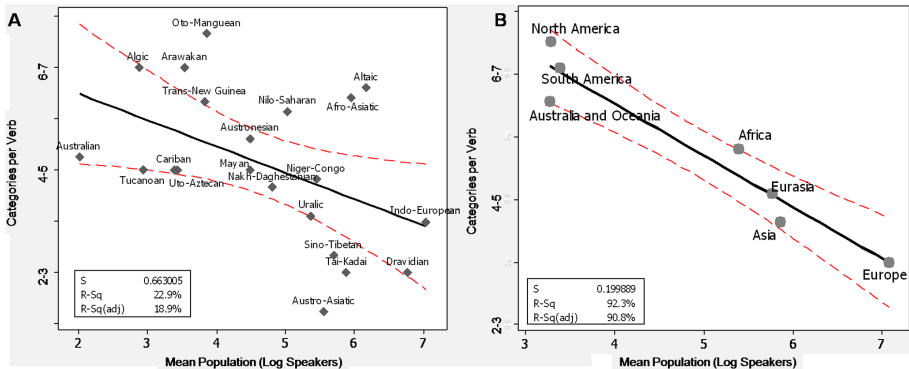
A hondiman datai ben bai wan oso gi en mati.

‘Der Jäger, der ein Haus gekauft hat, gab es seinem Freund.’

(aus John McWhorter, 2003, *The Power of Babel*)

Höherentwicklung im Sprachwandel

Anpassung der Grammatik an soziale Gegebenheiten

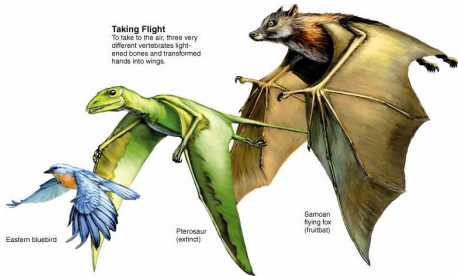


(aus G. Lyupan & R. Dale, 2010, PLoS ONE 5(1))

Konvergente Evolution

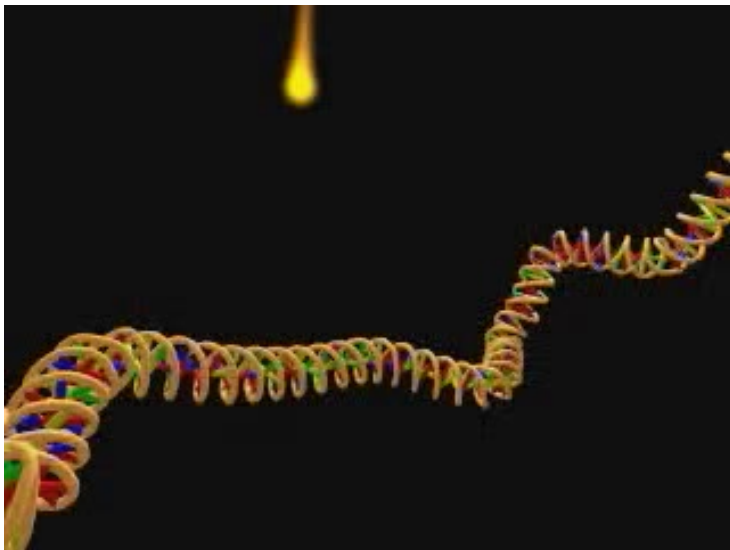
Taking Flight

To take to the air, three very different vertebrates lightened bones and transformed hands into wings.



- Altenglisch *docga* > Englisch *dog* ('Hund')
- Proto-Paman **gudaga* > Mbabaram *dog* ('Hund')

Evolution via Mutation in der Biologie



Lautgesetze

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: slapen , englisch: sleep → schlafen ; niederdeutsch und englisch: Schipp , ship → Schiff niederdeutsch: scherp , englisch: sharp → scharf	4/5	oberdeutsch und mitteldeutsch
	2	/*p/ → /pf/	niederdeutsch: Peper , englisch: pepper → Pfeffer ; niederdeutsch: Plauch , englisch: plough → Pflug ; niederdeutsch: scherp , englisch: sharp , althochdeutsch: scarph , mittelhochdeutsch: scharpf	6/7	oberdeutsch
G: /*d/ → /*t/	1	/*t/ → /ss/ → /s/	niederdeutsch: dat , wat , eten ; englisch: that , what , eat → das , was , essen	4/5	ober- und mitteldeutsch ¹
	2	/*t/ → /ts/	niederdeutsch: Tiet , englisch: tide (Gezeiten), schwedisch: tid → Zeit ; niederdeutsch: ver-tellen , englisch: tell → er-zählen ; Timmermann → Zimmermann	5/6	ober- und mitteldeutsch
G: /*g/ → /*k/	1	/*k/ → /xx/ → /x/	niederdeutsch: ik , altenglisch: ic → ich ; niederdeutsch und englisch: maken , make → machen ; niederdeutsch: auk → auch	4/5	ober- und mitteldeutsch ²
	2	/*k/ → /kx/	Kind → bairisch: Kchind	7/8	südbairisch, hoch- und höchstalemannisch
G: /*bʰ/ → /*b/ V: /*p/ → /*b/	3	/*b/ → /p/	Berg , bist → bairisch: perg , pist	8/9	teilweise bairisch und alemannisch
G: /*d/ → /*d/ → /*d/ V: /*t/ → /*d/ → /*d/	3	/*d/ → /t/	niederdeutsch: Dag oder Dach , englisch: day → Tag ; niederfränkisch: vader → Vater	8/9	oberdeutsch
G: /*gʰ/ → /*g/ V: /*k/ → /*g/	3	/*g/ → /k/	Gott → bairisch: Kott	8/9	teilweise bairisch und alemannisch
G: /*t/ → /b/ [ð]	4	/b/ → /d/ /ð/ → /d/	englisch: thorn , thistle , through , brother → Dorn , Distel , durch , Bruder	9/10	gesamtes deutsches Dialektkontinuum

Lautgesetze

- Lautgesetze sind spezifisch für eine bestimmte Sprachwandel-Periode
- gelten nahezu universell für alle Instanzen des betroffenen Lautes in der betroffenen Sprache
- im Idealfall gibt es schriftliche Zeugnisse der älteren und der jüngeren Sprachstufe (z.B. Latein/romanischen Sprachen, Althochdeutsch/Mittelhochdeutsch)
- meistens müssen Lautgesetze durch systematischen Vergleich verwandter Sprachen identifiziert werden
- erlaubt partielle Rekonstruktion der gemeinsamen Ursprungssprache

Sprachrekonstruktion durch die komparative Methode

The Indo-European language family

- William Jones 1786:
„The Sanskrit Language, whatever be its antiquity, is of wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either; yet bearing to both of them a stronger affinity both in the roots of verbs and the forms of grammar, than could possibly have been produced by accident; so strong indeed that no philologer could examine them at all without believing them to have sprung from some common source, which perhaps no longer exists: there is similar reason, so not quite so forcible, for supposing that both the Gothic and the Celtic, though blended with a different idiom, had the same origin with the Sanskrit; and the old Persian might be added to the same family, if this were the place for discussing any question concerning the antiquities of Persia.“

Sprachrekonstruktion durch die komparative Methode

- erste erfolgreiche Anwendung auf *Indo-europäisch* im 19. Jhd.

	Griechisch	Vedisch	Awestisch	Latein	Wallsisch	Gotisch	Armenisch	Tocharisch A	A.K.Slawisch	Litauisch	Indogermanisch (rekonstruiert)
1	<i>heis</i> (< *hens < *sems)	<i>éka</i>	<i>aéuua</i>	<i>ūnus</i> (Altlatein: oinos)	<i>un</i>	<i>ains</i>	<i>mi</i>	<i>sas</i>	<i>inū</i>	<i>vianas</i>	*oyno-, oyko-, sem-
2	<i>dúō</i>	<i>dvá</i>	<i>duua</i>	<i>duō</i>	<i>dau</i>	<i>twai</i>	<i>erkow</i>	<i>wu</i>	<i>dūva</i>	<i>dū</i>	*duwóh ₁
3	<i>treis</i>	<i>tri</i>	<i>θrāilō</i>	<i>trēs</i>	<i>tri</i>	<i>preis</i>	<i>erek'</i>	<i>tre</i>	<i>trije</i>	<i>trýs</i>	*tréyes
4	<i>téttares</i>	<i>catváras</i>	<i>caθuuārō</i>	<i>quattuor</i>	<i>pedwar</i>	<i>fidwor</i>	<i>čork'</i>	<i>štwar</i>	<i>četyre</i>	<i>keturi</i>	*k ^w etwóres
5	<i>pénte</i>	<i>pánca</i>	<i>panca</i>	<i>quinque</i>	<i>pump</i>	<i>flmf</i>	<i>hing</i>	<i>pāñ</i>	<i>peťi</i>	<i>penki</i>	*pénk ^w e
6	<i>héks</i>	<i>šát</i>	<i>xšuaaš</i>	<i>sex</i>	<i>chwech</i>	<i>saihs</i>	<i>več</i>	<i>šák</i>	<i>šesti</i>	<i>šeši</i>	*swéks
7	<i>heptá</i>	<i>saptá</i>	<i>hapta</i>	<i>septem</i>	<i>saith</i>	<i>sibun</i>	<i>ewt`n</i>	<i>špāt</i>	<i>sedmí</i>	<i>septyni</i>	*septrij
8	<i>októ</i>	<i>aštá</i>	<i>ašta</i>	<i>octō</i>	<i>wyth</i>	<i>ahtau</i>	<i>owt`</i>	<i>okāt</i>	<i>osmí</i>	<i>aštuoni</i>	*októ
9	<i>ennéa</i>	<i>náva</i>	<i>nauua</i>	<i>novem</i>	<i>naw</i>	<i>niun</i>	<i>inn</i>	<i>ñu</i>	<i>deveťi</i>	<i>devyni</i>	*néwn
10	<i>déka</i>	<i>dáša</i>	<i>daša</i>	<i>decem</i>	<i>deg</i>	<i>taihun</i>	<i>tasn</i>	<i>šák</i>	<i>deseťi</i>	<i>dėšimt</i>	*dék ^w
20	<i>wikati</i> (dorisch)	<i>vimsati</i>	<i>visaiti</i>	<i>viginti</i>	<i>ugeint</i> (Mittelwallsisch)		<i>k`san</i>	<i>wiki</i>			*wikmťi
100	<i>hekatón</i>	<i>šátám</i>	<i>satam</i>	<i>centum</i>	<i>cant</i>	<i>hund</i>		<i>kānt</i>	<i>sūto</i>	<i>šimtās</i>	*k ^w mtóm

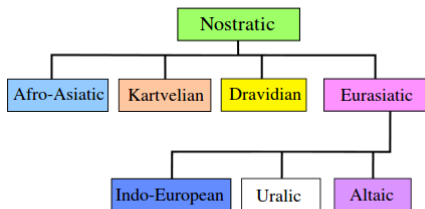
Grenzen der komparativen Methode

- Zeittiefe beschränkt auf 2 000 bis 8 000 Jahre



Tiefe Sprachverwandtschaften

- Vielzahl von Vorschlägen für Meta-Familien
 - Nostratisch:
 - erstmals von Pedersen (1903) vorgeschlagen
 - ursprünglicher Vorschlag: Indo-europäisch, Finno-ugrisch, Samoyedisch, Turk-Sprachen, Mongolisch, Manchu, Yukaghir, Eskimo, Semitisch und Hamitisch
 - weiterentwickelt durch „Moskauer Schule“ in den 1960ern
 - Versuch der Rekonstruktion von Wortschatz



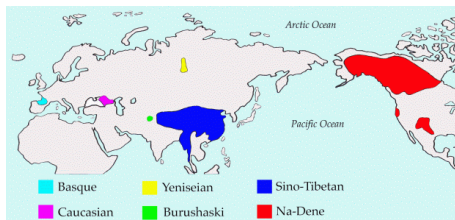
Tiefe Sprachverwandtschaften

- Vielzahl von Vorschlägen für Meta-Familien
 - Eurasiatisch
 - vorgeschlagen von Greenberg (2000)
 - umfasst Indo-europäisch, Uralisch-Yukaghirisch, Altaisch, Tschuktscho-Kamtschadalisch, Eskimo-Aleutisch, Koreanisch-Japanisch-Ainu, Gilyak, Etruskisch
 - diverse Argumente, v.a. Morphologie und Phonologie



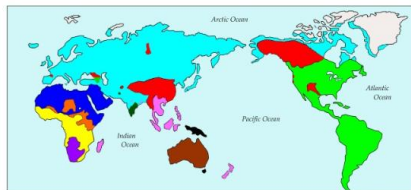
Tiefe Sprachverwandtschaften

- Vielzahl von Vorschlägen für Meta-Familien
 - Dene-Kaukasisch
 - umfasst Ne-Dene, Kaukasisch, Sino-Tibetisch, Jenniseisch, Burushaski, manchmal auch Baskisch



Tiefe Sprachverwandtschaften

- Vielzahl von Vorschlägen für Meta-Familien
 - Amerindisch
 - vorgeschlagen von Greenberg (1987)
 - umfasst alle Indianersprachen außer Na-Dene



Khoisan	Dravidian	Austric
Niger-Kordofanian	Kartvelian	Indo-Pacific
Nilo-Saharan	Eurasianic	Australian
Afro-Asiatic	Dene-Caucasian	Amerind

Language Families of the World (after Greenberg)

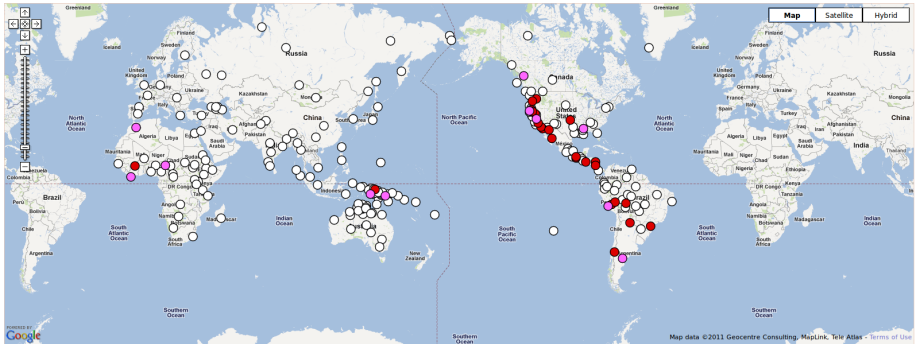


Tiefe Sprachverwandtschaften

- Merritt Ruhlen, ein Schüler von Greenberg, behauptet sogar, „Proto-World“ z.T. rekonstruieren zu können, z.B. das Wort *akwa* für Wasser (das sich faszinierenderweise von Adam und Eva über Cicero bis zu Umberto Eco im Indoeuropäisch/Italisch/Lateinisch/Italienischen Zweig nicht verändert hat)
- derartige Vorschläge basieren häufig auf geographischen Häufungen einzelner Merkmale, wie z.B. Pronominalformen

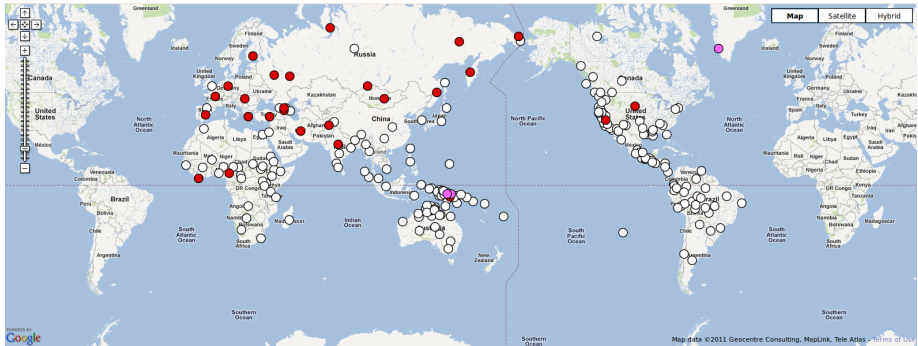
Tiefe Sprachverwandtschaften

- N/M-Pronomina



Tiefe Sprachverwandtschaften

● M/T-Pronomina



Phylogenetische Rekonstruktion in der Bioinformatik

Sequenzalinerung

- Algorithmus findet optimale Alinerung zwischen Sequenzen
- Anzahl der Mutationen wird somit abgeschätzt
- ergibt Abschätzung des evolutionären Abstands zwischen den entsprechenden Organismen

```

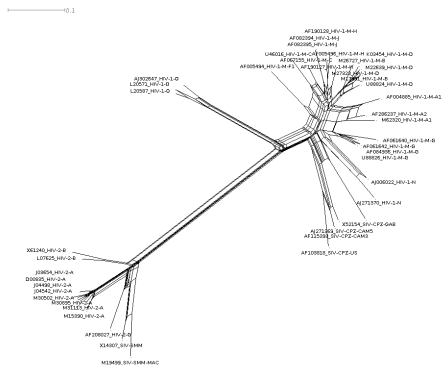
Hebei_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Ningxia*_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Beijing_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Henan98_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Heilong01 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Henan02_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Jilin_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang4/00 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Henan*_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang10/00 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Jiangsu*_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang02_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang47/01 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guangxi109 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guangxi9/9 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang56/01 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Shanghai*_ : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Nanjing1/9 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Shandong7/ : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Shandong6/ : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang5/97 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Guang6/97 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Shenzhen*_ : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Fujian_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Shijia*_1 : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
Heilong00_ : DSFFRMRMLCKFNNAVFCQAAYTNNRNGILFPMGIIHPPDTCQLMLVRDDTTSVTEIIRFFFPFLIGPRLVWGL
18FY88MRMLTqk NaYF Q AqYTNRngk ILFPMGIIHPPDTC QL LYT4cDTTTSV TEI0 RFFFP6IGPRLVWG

Hebei_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Ningxia*_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Beijing_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Henan98_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Heilong01 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Henan02_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Jilin_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang4/00 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Henan0*_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang10/00 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Jiangsu*_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang02_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang47/01 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guangxi109 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guangxi9/9 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang56/01 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Shanghai*_ : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Nanjing1/9 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Shandong7/ : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Shandong6/ : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang5/97 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Guang6/97 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Shenzhen*_ : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Fujian_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Shijia*_1 : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
Heilong00_ : SFFYRSMW X X G Y T R R R R 794 P P T T T Y R R S T T R T R R F F F P F L I G P R L V W G L
18FY88MRMLTqk NaYF Q AqYTNRngk ILFPMGIIHPPDTC QL LYT4cDTTTSV TEI0 RFFFP6IGPRLVWG
    
```


Phylogenetische Rekonstruktion in der Bioinformatik

Phylogenetische Bäume

- statistische Verfahren zur Rekonstruktion des wahrscheinlichsten Stammbaums
- häufig konfligierende Information wegen:
 - konvergenter Evolution
 - Rück-Mutation
 - lateraler Gen-Transfer
- Darstellung alternativer Rekonstruktionen in Netzwerk-Strukturen



SplitsTree Software, Huson & Bryant, MatNat-Fakultät

Phylogenetische Rekonstruktion in der Bioinformatik

Alternative: Cluster-Karten

- Organisation aller Datenpunkte (=Molekularsequenzen) in 2- oder 3-dimensionalen Raum
 - größere Ähnlichkeit entspricht (simulierter) physikalischer Anziehungskraft und umgekehrt
 - Algorithmus findet Energie-Minimum
- ⇒ verwandte Sequenzen bilden Cluster

Frickey & Lupas, MPI für Entwicklungsbiologie

Die Daten des Automated Similarity Judgment Project

- Projekt am MPI EVA in Leipzig um Sören Wichmann
- erfasst inzwischen über 5 000 Sprachen
- für jede Sprache Grundwortschatz von 40 Wörtern in (vereinfachter) phonetischer Umschrift
- frei elektronisch verfügbar

verwendete Konzepte: *I, you, we, one, two, person, fish, dog, louse, tree, leaf, skin, blood, bone, horn, ear, eye, nose, tooth, tongue, knee, hand, breast, liver, drink, see, hear, die, come, sun, star, water, stone, fire, path, mountain, night, full, new, name*

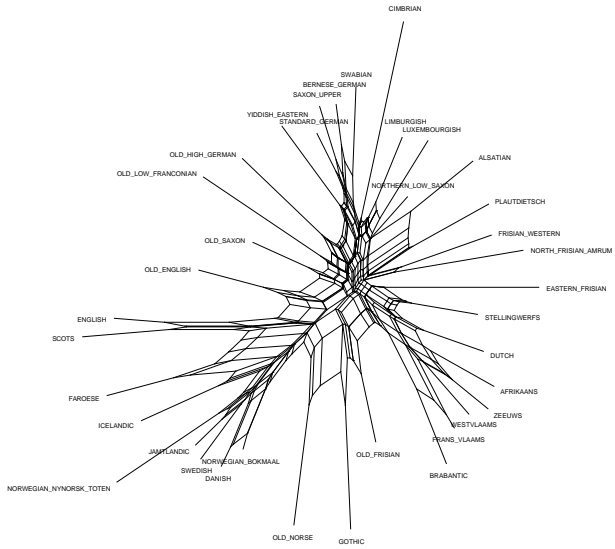
Automated Similarity Judgment Project

<i>Konzept</i>	Deutsch	Englisch
<i>I</i>	iX	Ei
<i>you</i>	du	yu
<i>we</i>	vir	wi
<i>one</i>	ains	8is
<i>two</i>	cvai	8Et
<i>person</i>	mEnS	pers3n
<i>fish</i>	fiS	fiS
<i>dog</i>	hunt	dag
<i>louse</i>	laus	laus
<i>tree</i>	baum	tri
<i>leaf</i>	blat	lif
<i>skin</i>	haut	skin
<i>blood</i>	blut	bl3d
<i>bone</i>	knoX3n	bon
<i>horn</i>	horn	horn
<i>ear</i>	XXX	ir
<i>eye</i>	aug3	Ei

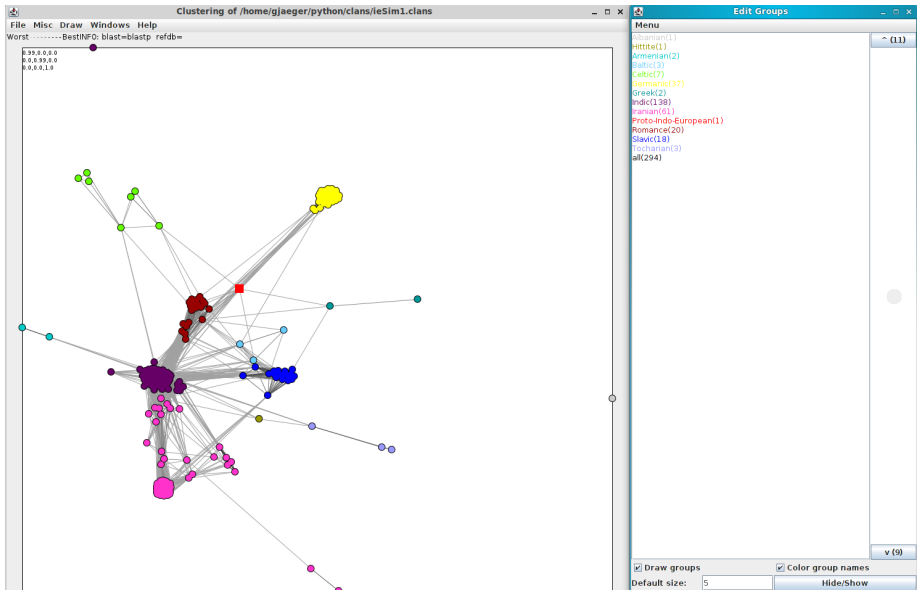
<i>Konzept</i>	Deutsch	Englisch
<i>nose</i>	naz3	nos
<i>tooth</i>	ch an	tu8
<i>tongue</i>	ch uN3	t3N
<i>knee</i>	kni	ni
<i>hand</i>	hant	hEnd
<i>breast</i>	brust	brEst
<i>liver</i>	leb3r	liv3r
<i>drink</i>	triNk3n	drink
<i>see</i>	ze3n	si
<i>hear</i>	her3n	hir
<i>die</i>	Sterb3n	dEi
<i>come</i>	kh om3n	k3m
<i>sun</i>	zon3	s3n
<i>star</i>	StErn	star
<i>water</i>	vas3r	wat3r
<i>stone</i>	Stain	ston
<i>fire</i>	foia	fEir

Einfache Sequenz-Alinierung

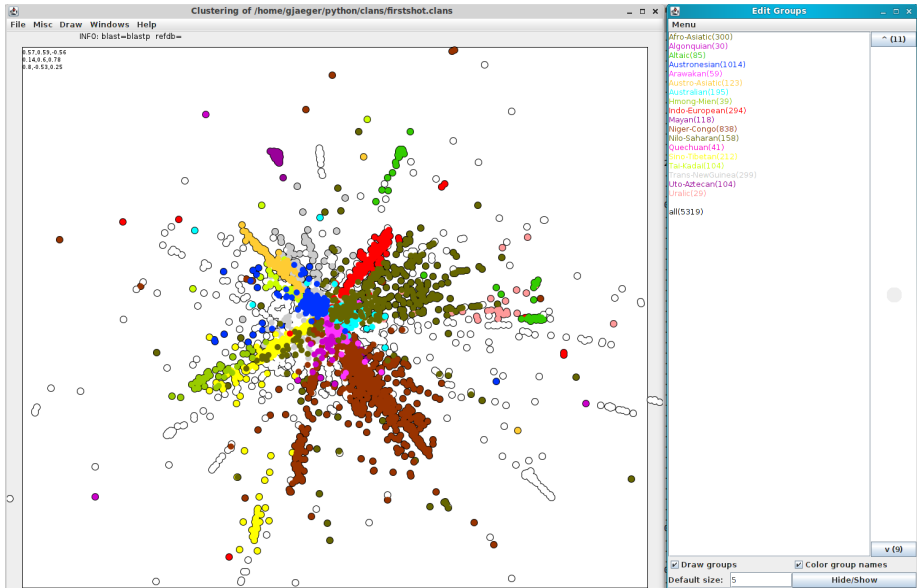
11,0



Einfache Sequenz-Alinierung

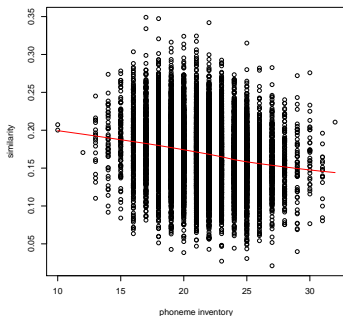


Einfache Sequenz-Alinierung



Einfache Sequenz-Alinierung

- Störeffekt: bei Sprachen mit kleineren Lautinventaren ergeben sich mehr Zufallsähnlichkeiten also bei Sprachen mit vielen verschiedenen Lauten
- Daher erscheinen Sprachen mit wenigen Lauten einander ähnlicher, als sie es tatsächlich sind.

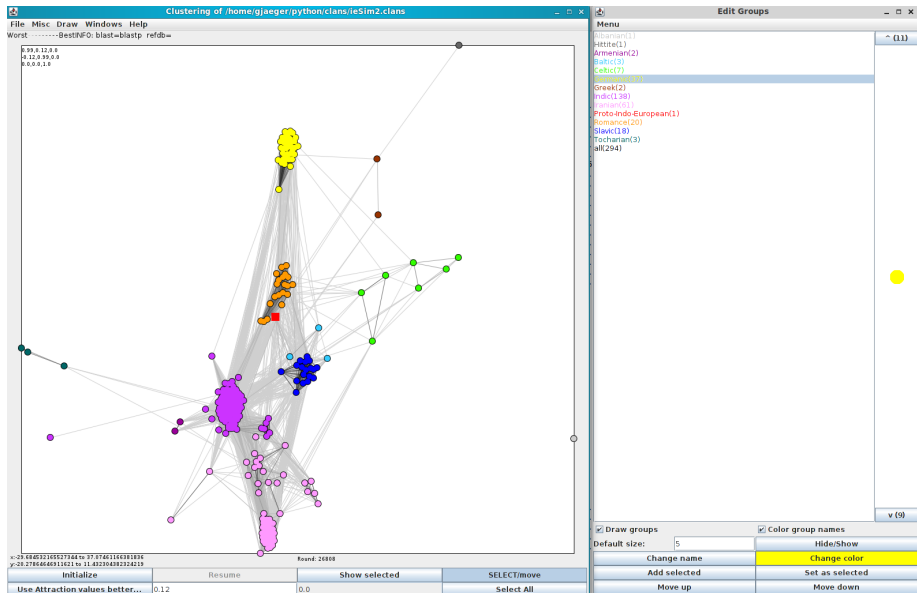


Kalibrierte Alinierung

Ähnlichkeit des Deutschen zu:

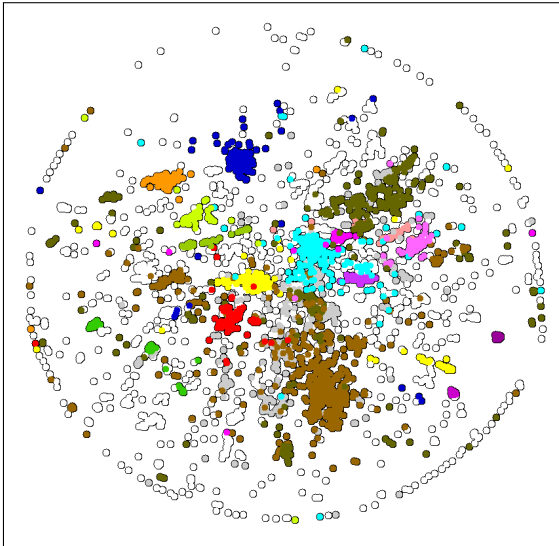
- Niederländisch: 35,4
- Englisch: 17,7
- Ur-Indoeuropäisch: 11,0
- Latein: 6,4
- Spanisch: 1,7
- Russisch: 3,3
- Türkisch: 0,5
- Ungarisch: 0,3

Kalibrierte Alinierung



Kalibrierte Alinierung

Alle 5 000 Sprachen:



Edit Groups

Menu

- Afro-Asiatic(300)
- Algonquian(30)
- Altaic(85)
- Austronesian(1023)
- Awasan(5)
- Austra-Asiatic(123)
- Australian(198)
- Hmong-Mien(39)
- Indo-European(294)
- Mayan(118)
- Niger-Congo(838)
- Nilo-Saharan(159)
- Quechuan(41)
- Sino-Tibetan(212)
- Takadim(10)
- Trans-NewGuinea(299)
- Uto-Aztecan(104)
- Uralic(29)

all(5381)

v (9)

Draw groups Color group names

Default sl... 5	Hide/Show
Change name	Change color
Add selected	Set as selected
Move up	Move down
Update	Delete

Kalibrierte Alinierung

- etablierte Sprachfamilien bilden stabile Cluster
- keine darüber hinausgehenden sichtbaren Muster

Kalibrierte Alinierung

- Methode ist relativ grobkörnig

h	a	n	t	h	a	n	t
h	E	n	d	m	a	n	o

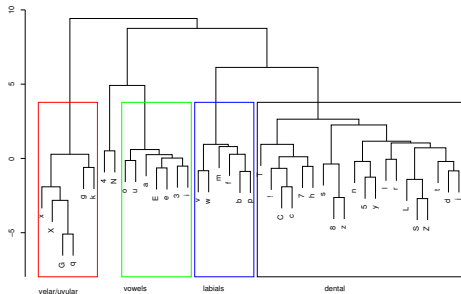
- Ähnlichkeit ist in beiden Fällen 50%
- Korrespondenz $a \sim E$, $t \sim d$ sind nach linguistischen Kriterien viel natürlicher als $h \sim m$ or $t \sim o$
- Deutsch/Englisch und Deutsch/Spanisch erscheinen hier äquidistant, obwohl die Ähnlichkeit zwischen Deutsch und Englisch intuitiv viel größer ist

Needleman-Wunsch-Algorithmus

- Analogie zur Bioinformatik: Mutationen zwischen verschiedenen Aminosäuren-Paaren sind unterschiedlich wahrscheinlich
- Algorithmus sucht *wahrscheinlichste* Übereinstimmungen zwischen Sequenzen
- $a \sim E, d \sim t$ sind im Sprachwandel wahrscheinlicher als $t \sim o$

Gewichtete Alinierung

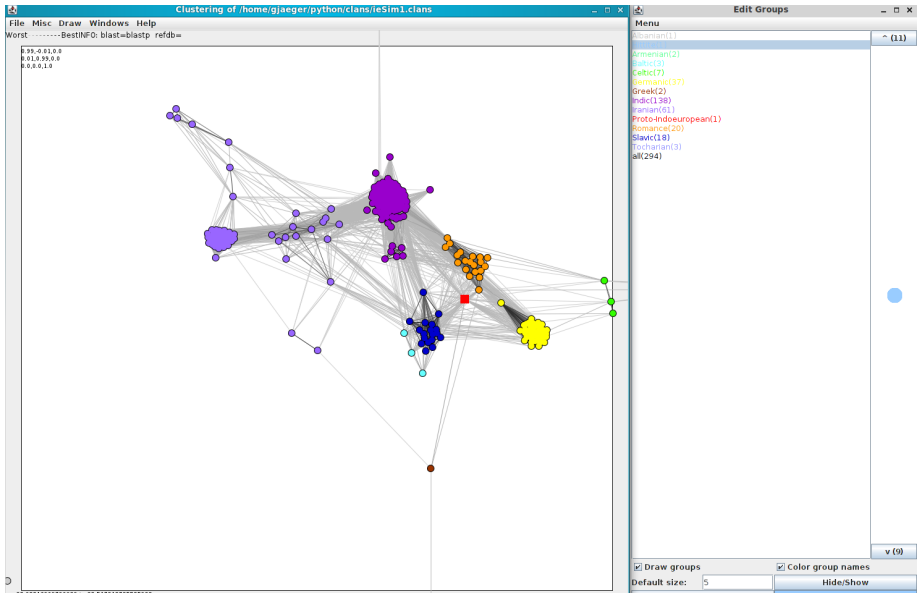
- automatisch bestimmte Gewichte:
 - $d \sim t$: 0.69
 - $a \sim E$: 0.07
 - $h \sim m$: -0.61
 - $t \sim o$: -0.80



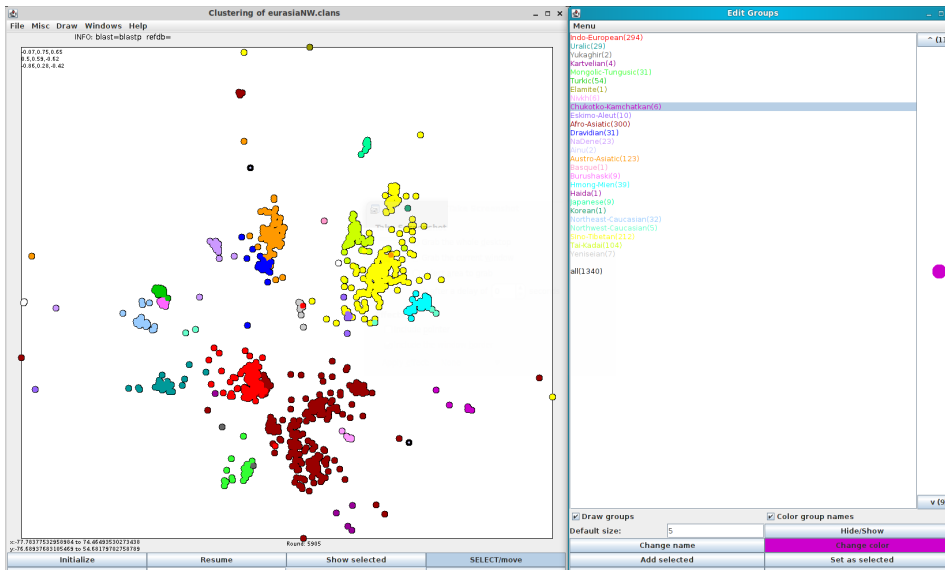
Gewichtete Alinierung

- Ähnlichkeit des Deutschen zu:
 - Dutch 35.4 / 38.3
 - English: 17.7 / 20.5
 - Proto-Indoeuropean: 11.0 / 14.6
 - Latin: 6.4 / 12.4
 - Spanish: 1.7 / 1.8
 - Russian: 3.3 / 6.5
 - Turkish: 0.5 / 0.6
 - Hungarian: 0.3 / 2.1

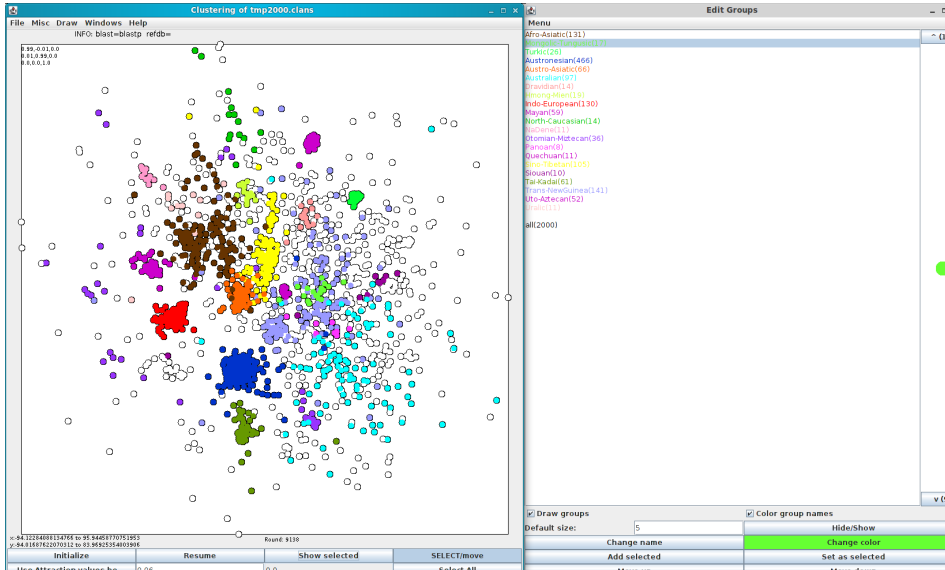
Gewichtete Alinierung



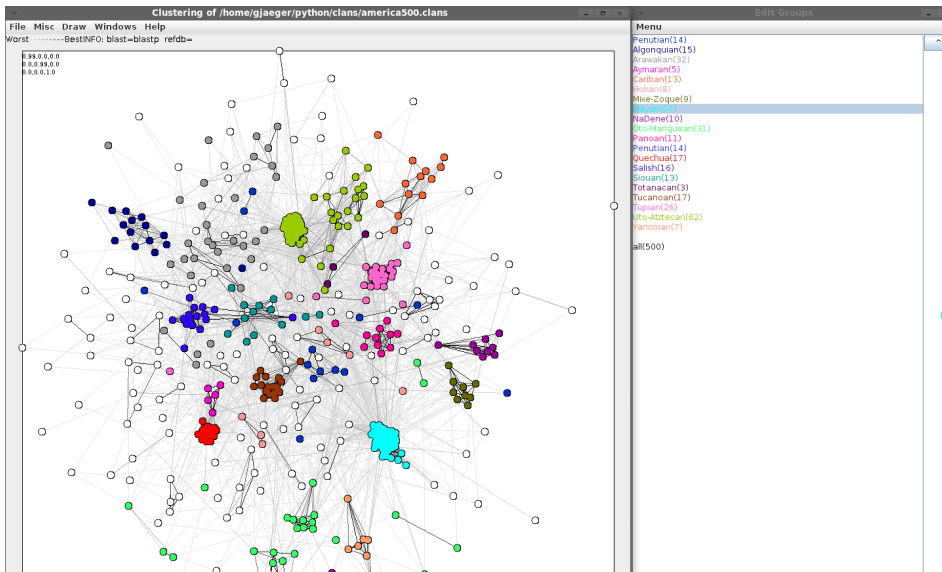
Gewichtete Alinierung



Gewichtete Alinierung

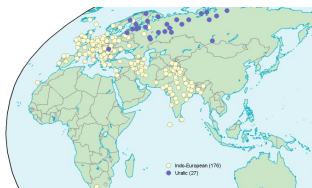


Gewichtete Alinierung



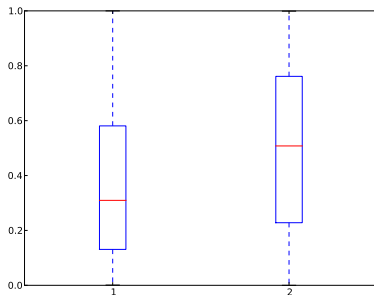
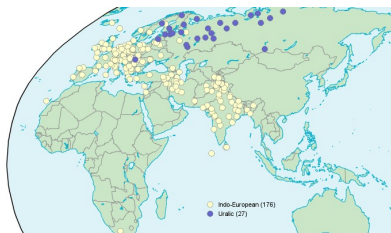
Gewichtete Alinierung

- einige interessante Meta-Verwandtschaften werden sichtbar, v.a.
 - Indo-europäisch/Uralisch
 - Austronesisch/Tai-Kadai



Indo-europäisch / Uralisch

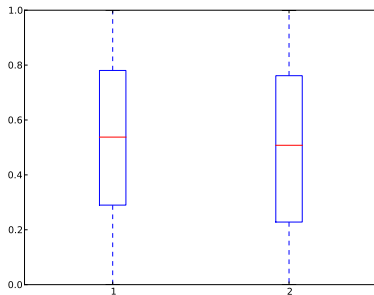
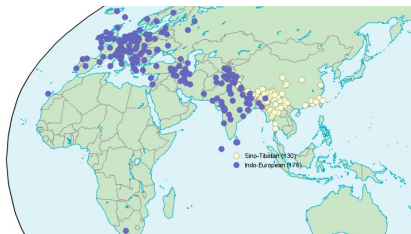
- p -Werte für Vergleich Ähnlichkeiten IE/Ura vs. Zufallspaarungen



p -Wert: $1,5 \times 10^{-20}$

Indo-europäisch / Uralisch

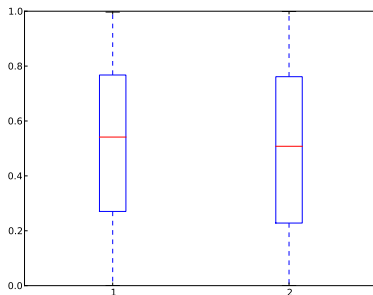
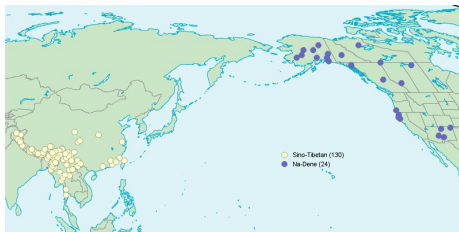
- p -Werte für Vergleich Ähnlichkeiten IE/Sino-Tibetisch vs. Zufallspaarungen



p -Wert: 1

Indo-europäisch / Uralisch

- p -Werte für Vergleich Ähnlichkeiten Sino-Tibetisch/Na-Dene vs. Zufallspaarungen



p -Wert: 1

Indo-europäisch / Uralisch

- p -Werte für Vergleich Ähnlichkeiten Tai-Kadai/Austronesisch vs. Zufallspaarungen



p -Wert: 5×10^{-5}

