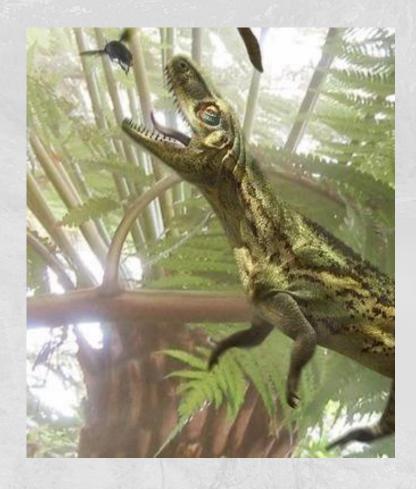
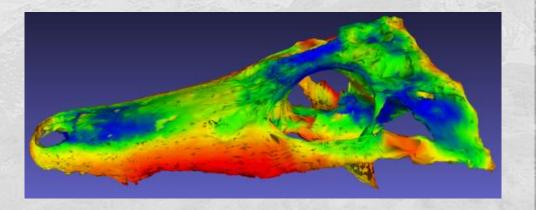


Me

- Background in Triassic archosaurs and their phylogenetics Currently working on inferring phylogeny from 3D morphological data Previously some work (and interest in!) language phylogeny





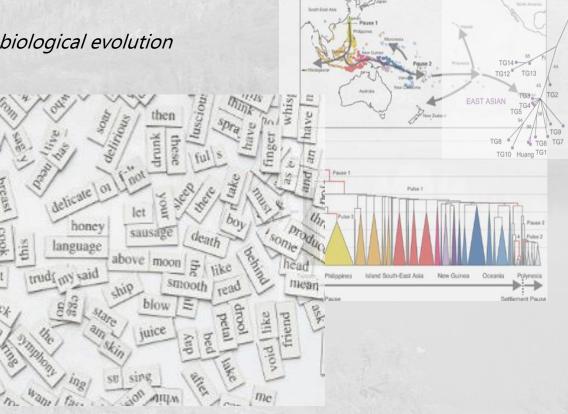






Language and culture evolve too!

- Descent with modification also occurs in culturolinguistic systems
- Human languages can be grouped into clades
 Many aspects of culture can also be grouped in this way
- Lecture plan:
 - Some history
 - Methods
 - Similarities and differences with biological evolution
 - **Applications**





History

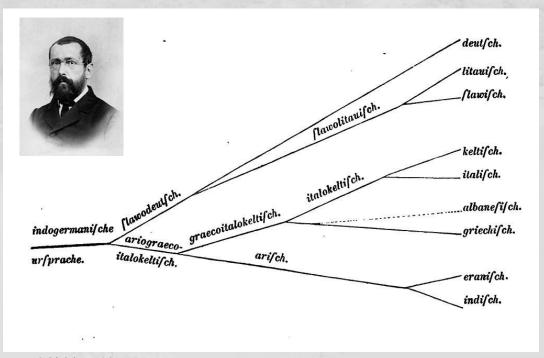
- At least to William Jones (Bengal, 1780)
- Some of the oldest "phylogenetic" trees were linguistic
- Schleicher developed the *Stammbaumtherorie* around the same time as Darwin
- Mentioned by Darwin



"[Sanskrit bears to Latin and Greek] ...a stronger affinity... than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to **have sprung from some common source**, which, perhaps, no longer exists"



"the proper or even the only possible arrangement would ...be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and recent, by the closest affinities, and would give the filiation and origin of each tongue." (ch. 13: 422)

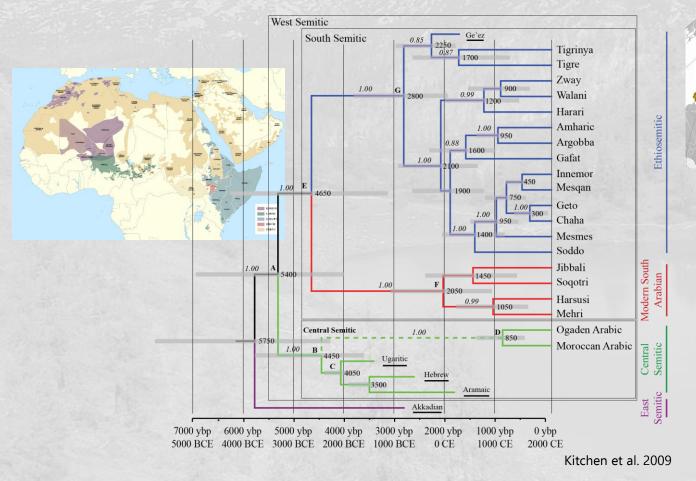


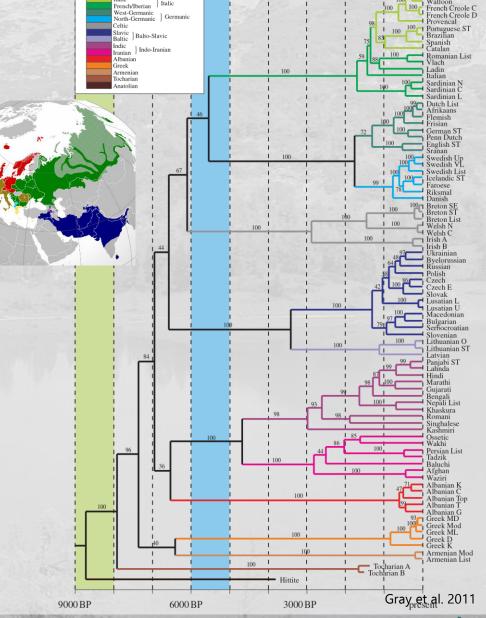
From Schleicher 1861 Compendium der vergleichenden Grammatik der indogermanischen Sprachen



History

 Languages are indeed grouped into clades using trees today, e.g. Indo-European, Semitic (with Afroasiatic)







water

Fundamental basis is the idea of **cognates**

=Homologies in biology, i.e. **features shared due to common origin Words** can be cognates, but may change their meaning and sound

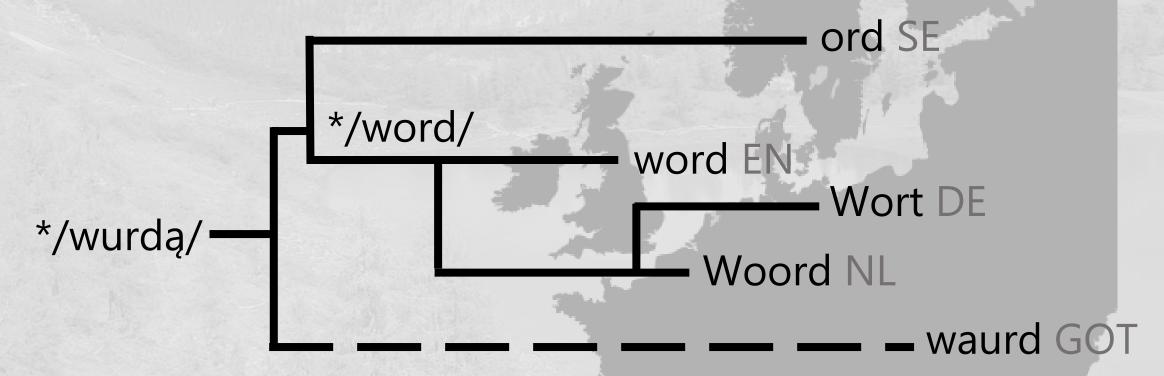
dat

Sounds can be corresponding phonemes (=homologues) even though they differ Shared words due to common origin should all experience the same sound shifts – if not, borrowing (=hybridisation) may have occurred

DE	Wasser	das(s)	Butter	butyrum LAT
EN	water	that	butter	

boeter

- Classically "comparative method"
 - Identified cognate (=homologue) words (and sounds phonemes)
 - Reconstructed proto language (manually)
 - = ancestral state reconstruction

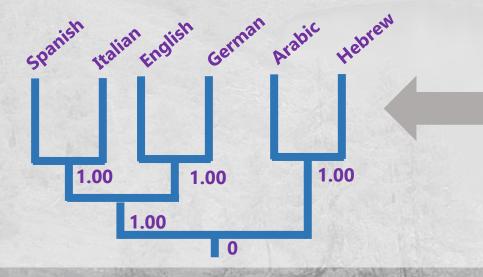




- Today similar methods to biological phylogenetics:
 - Common/fundamental words (less subject to borrowing) coded into matrix of "presence" or "absence"
 - Bayesian, ML, parsimony, neighbour joining networks
 - Some specific models (e.g. no regain after loss) can be used

English	house	water	fish	I
German	Haus	Wasser	Fisch	ich
Italian	casa	acqua	pesce	io
Spanish	casa	agua	pez	yo
Arabic	بیت bayt	māʾ مَاء	سمكة samaka	ana أنا
Hebrew	báyit בַּיִת	máyim מַיִּם	dag דג	'ani אָנֹכִי

	house	casa	bayt	fish	samaka	dag	I	ana
English	1	0	0	1	0	0	1	0
German	1	0	0	1	0	0	1	0
Italian	0	1	0	1	0	0	1	0
Spanish	0	1	0	1	0	0	1	0
Arabic	0	0	1	0	1	0	0	1
Hebrew	0	0	1	0	0	1	0	1













Also possible to code using multitstate approaches, but rarely done

English	house	water	fish	I
German	Haus	Wasser	Fisch	ich
Italian	casa	acqua	pesce	io
Spanish	casa	agua	pez	yo
Arabic	بیت bayt	māʾ مَاء	سمكة samaka	ana أنا
Hebrew	báyit בַּיִת	máyim מַיִם	dag דג	'ani אָנֹכִי

	house	casa	bayt/báyit	fish	samaka	dag	I	ana
English	1	0	0	1	C	0	1	0
German	1	0	0	1	C	0	1	0
Italian	0	1	0	1	C	0	1	0
Spanish	0	1	0	1	C	0	1	0
Arabic	0	0	1	0	1	0	0	1
Hebrew	0	0	1	0	C	1	0	1

Binary

	house	water	fish	Ι
English	0	0	0	0
German	0	0	0	0
Italian	1	1	0	0
Spanish	1	1	0	0
Arabic	2	2	1	1
Hebrew	2	2	1	1

Multistate



- Phonemes can also be coded as characters
 - Rarely done, because very labile
- "Morphological" features, i.e. grammatical structure (cases, word order etc.)
 - Difficult to comparably code characters, but may be useful for deep phylogeny



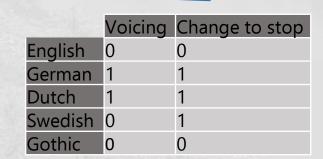
	"th" in "thorn"
English	thorn /θ/
German	Dorn /d/
Dutch	doorn /d/
Swedish	torn /t/
Gothic	thaurnus */θ/



	/θ/	/d/	/t/
English	1	0	0
German	0	1	0
Dutch	0	1	0
Swedish	0	0	1
Gothic	1	0	0



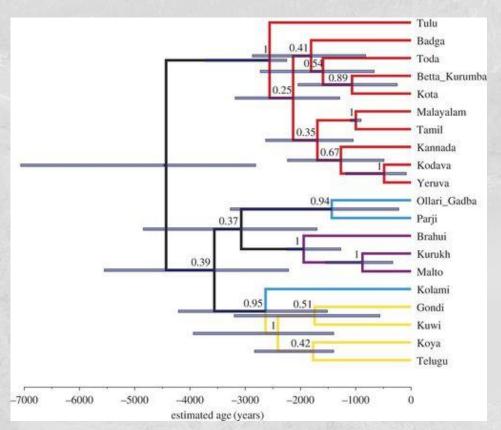
English	0
German	1
Dutch	1
Swedish	2
Gothic	0



Gender	Masculine		
Indefinite	en hest a horse		
Definite	hest en horse.DEF		
Double definite	den hesten that horse.DEF		
Adjective	en fin hest a nice horse		
Possessive	min hest/hesten min my horse/horse.DEF		
	my		



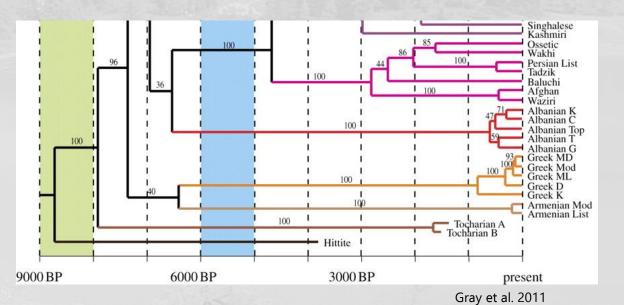
- Tree dating can be incorporated, as in biological phylogeny
- E.g. minimum divergence dates can be based on manuscript ages "Ancestral" languages usually treated as separate tips, as in biology



na ingern dazini. bed cynnig buyin se prumon hude cepetingal elle the medon, oft feeld (cerus feether buencim mone zu miechum mendo feel of tenh estade empl syddan agreet par per scenez runder he har morne sel peox under polenim people myndum bah of him whole bana your fixen dua open Inion pade hopan feelde sombarr goldan finer god cyning. diem eurona par efter conned scong in secundum bone tod fende folce zofhorhe tohor fente on year place of dungon aldon age lange hoile him her lu pour pulsor por local popold ape pop sair being par buen bled pide figurant feel by course feels landom in Speciel and Speciel

MFILSEVEROPIOPER COPONTIFICMAXIM IL:ANTONINO:NVGI OPTIMISFORTISS

> ॥ ॥श्रीमरीय्रायमनः॥स्त्रः। भातप्रजा।स्त्रारीरीयम इलिखनक्रमः गणिकागीरीपद्रग्राचिमधासविज्ञीविजयाजया देवसनास्वधास्त्राहामतरोज्ञाकः गारः शतःपृष्टिसात्रात्रियातः नःकलदेवता धतिघोडशमाताः एगधिपाः क्रमेएफिस्तेणसं स्याप्यः त्रीनर्भवः एताचा इश्मातः समागाधियाः समागछ इति संस्थाय्यप्रतिष्ठामंत्रेणसंस्थाय अग्रेहेहत्यादिएतासंबीदशमा त्रलांसगरणियरां असककर्म लिनिमतंत्रजामहंकी खे इतिवृतिज्ञागगंधव्यसियनेवस् विप्रजागवसाधामंत्रः



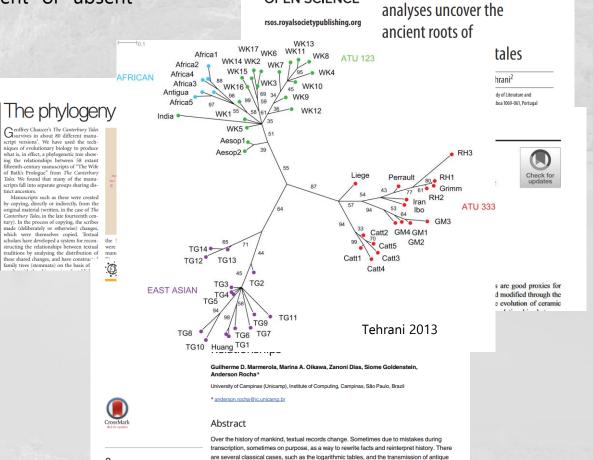
Cur ne de de de de

Kolipakam et al. 2018



- Can also code other cultural data
 - For example, folk tales
 - Characters and elements of stories coded as "present" or "absent"
 - Other examples: pots, manuscripts





G OPEN ACCESS

ROYAL SOCIETY OPEN SCIENCE

Comparative phylogenetic



Similarities and differences with biological evolution

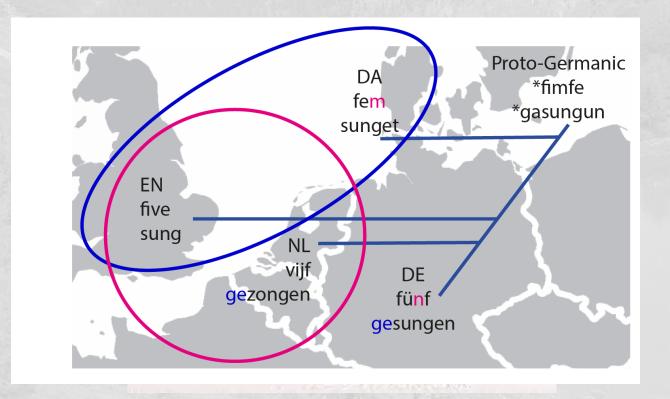
- Many aspects of culture, especially core parts of language, are inherited There is continued borrowing, even long after initial divergence (*never* complete separation)
- In this way, more similar to plant or bacterial evolution

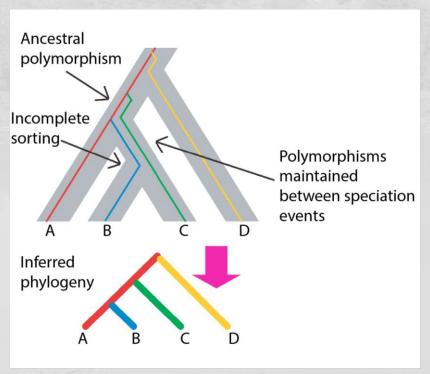




Similarities and differences with biological evolution

- "Wave theory" (Wellentheorie opposing Stammbaumtherie) of Schmidt identified waves of radiation of language characteristics effectively spread of alleles within population
- Because new features can arise in particular areas of a diverging language area, they can conflict with main signal and fit geography – not only due to "hybridisation" - effectively incomplete lineage sorting

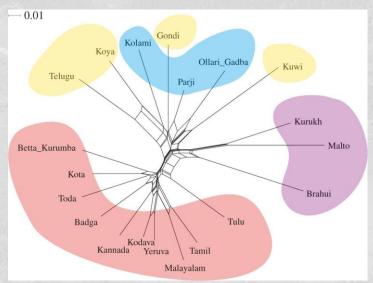




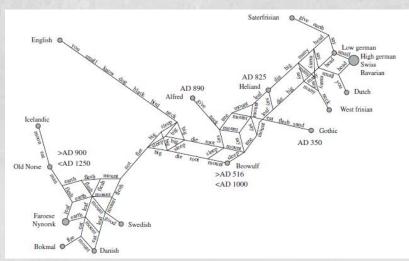


Similarities and differences with biological evolution

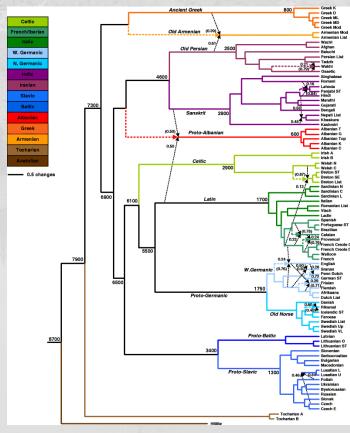
- Network graphs can be more appropriate, especially for very "labile" aspects like phonetics Incorporating hybridisation events is becoming possible, and commonly used



Kolipakam et al. 2018



Heggarty et al. 2010



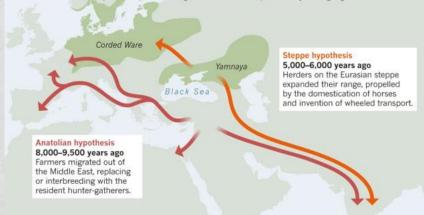
Willems et al. 2016

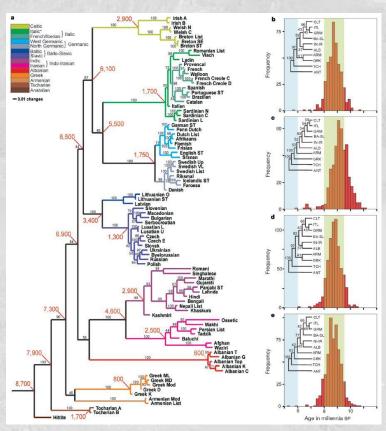


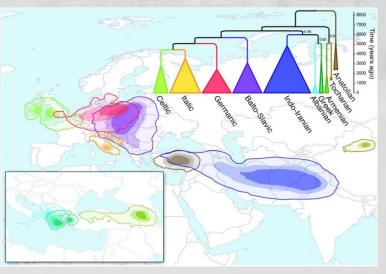
- Dating population divergence Placing population divergence geographically E.g. the Anatolian versus Kurgan hypothesis of agriculture

STEPPE IN TIME

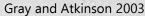
An ancient-DNA study links the Corded Ware culture of northern Europe with the Yamnaya culture of the Eurasian steppe. It points to a mass migration northwest that would support the Steppe hypothesis, one of two theories that compete to explain the origins of the Indo-European family of languages.







Bouckaert et al. 2012





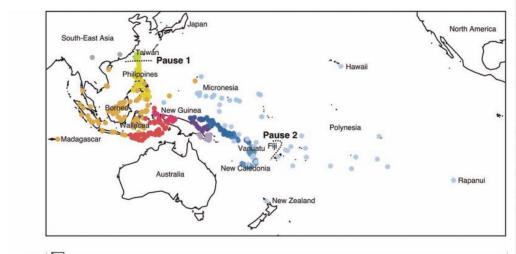


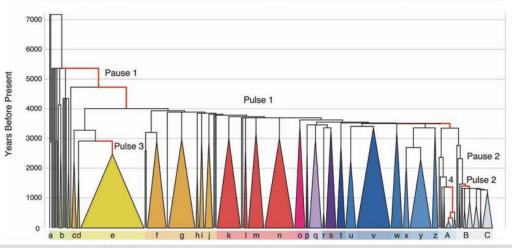
Bouckaert et al. 2012



- Another example:
 - Origin and expansion of Austronesian people
 - "Slow boat" from Wallacea v. "pause-pulse" recent from Taiwan



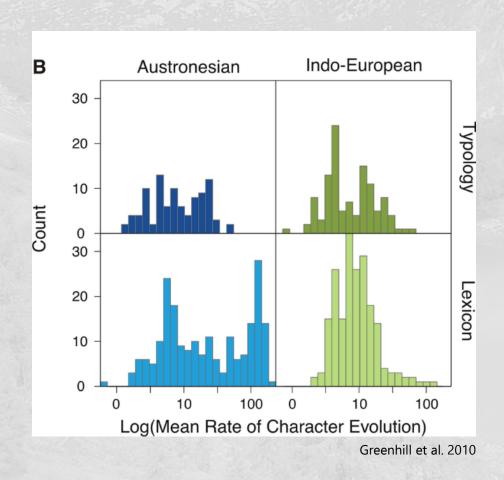


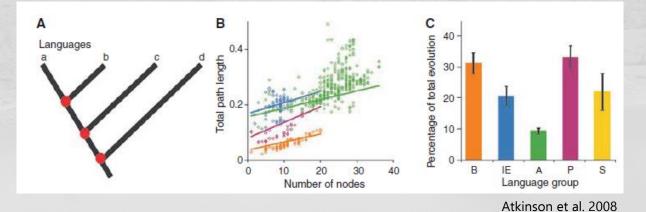


Gray et al. 2009

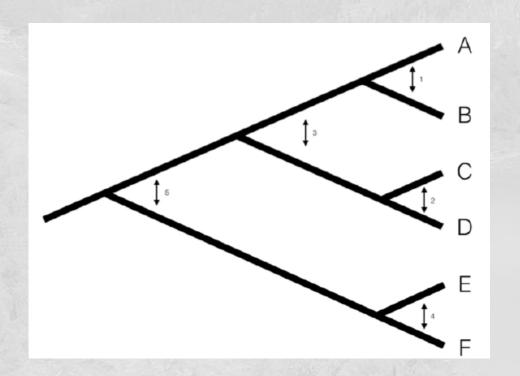


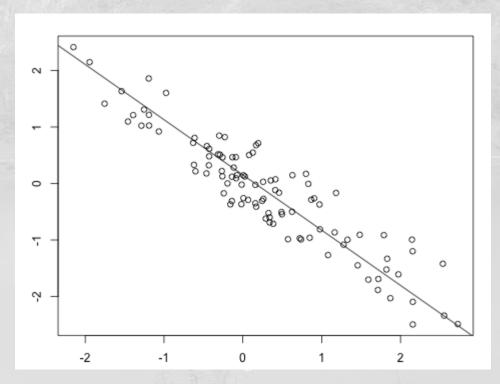
Examining rates of evolution, and how culturolinguistic systems evolve





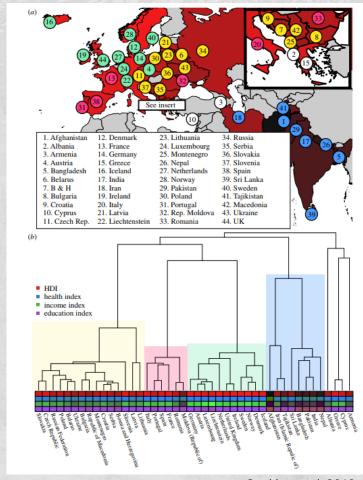
- Language phylogenies can be used for examining other aspects of cultural evolution allow **independent contrasts** and **phylogenetically informed regressions**Can map any kind of social traits onto phylogeny, and look at how they evolve



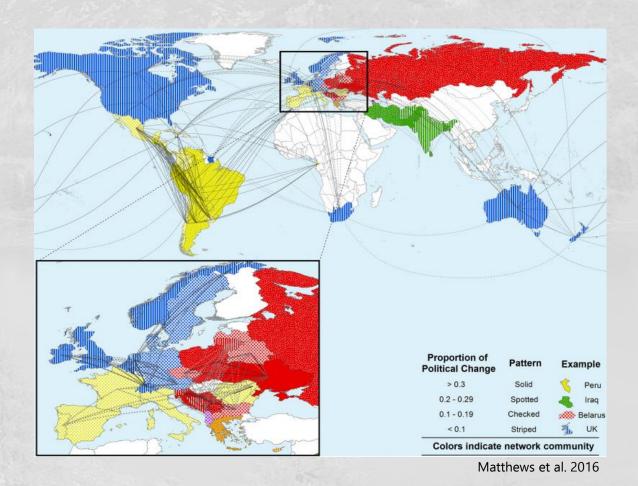




Example: economic and social indicators



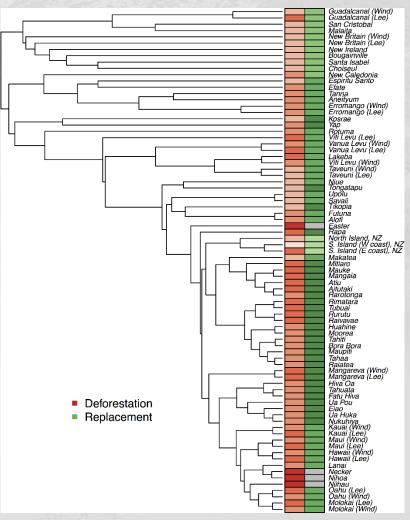
Sookias et al. 2018





Further examples: forest use patterns in polynesia



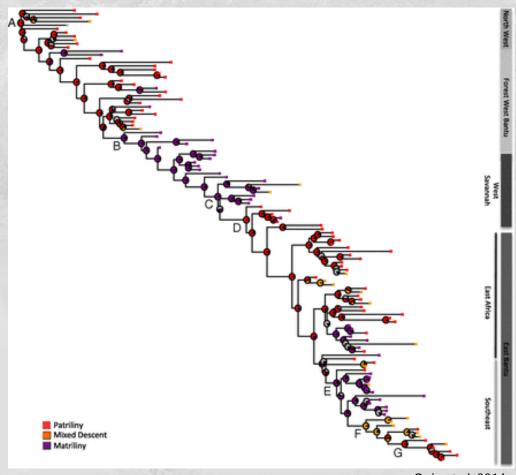


Atkinson et al. 2016



Further examples: how tradiational marriage residence rules (e.g. matri/patrilocal) affect other social structures

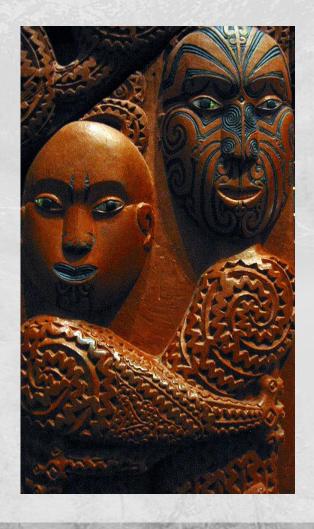


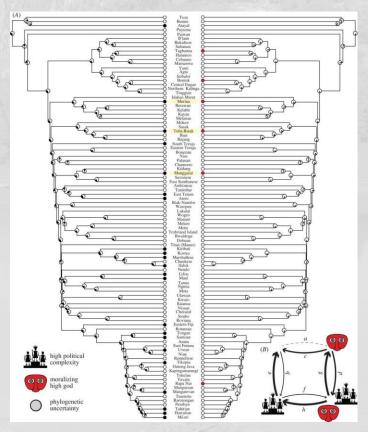


Opie et al. 2014



• Further examples: do you need high "moralizing high gods" for complex societies? No...

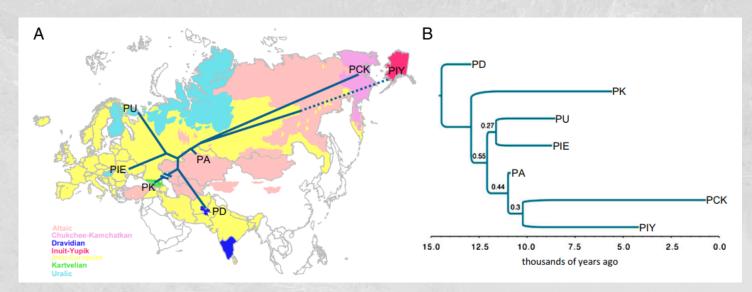




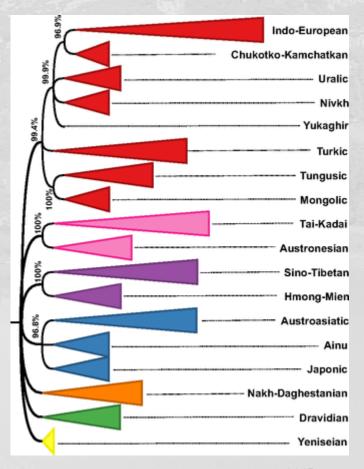
Watts et al. 2015



- Deep linguistic phylogeny Are certain features more strongly conserved?
- Mass lexical comparison



Pagel et al. 2013

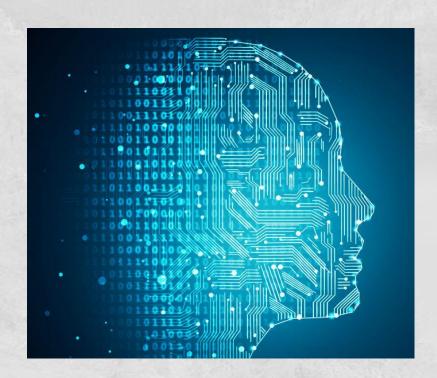


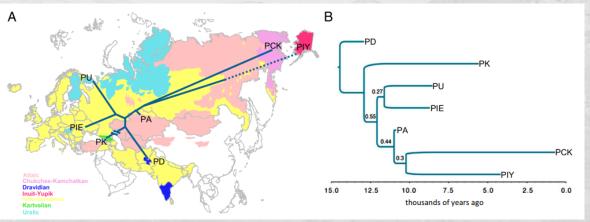
Jäger 2015



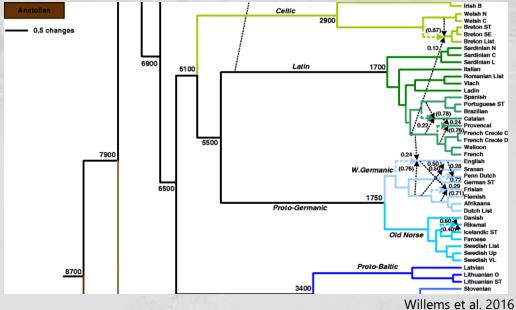
Ways forward

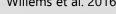
- Expanding knowledge of deep phylogeny Automatic cognate judgement AI? Broad incorporation of hybridisation in models More and more varied questions!





Pagel et al. 2013

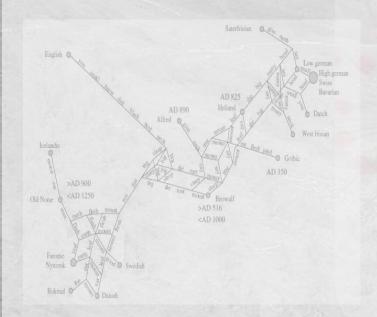


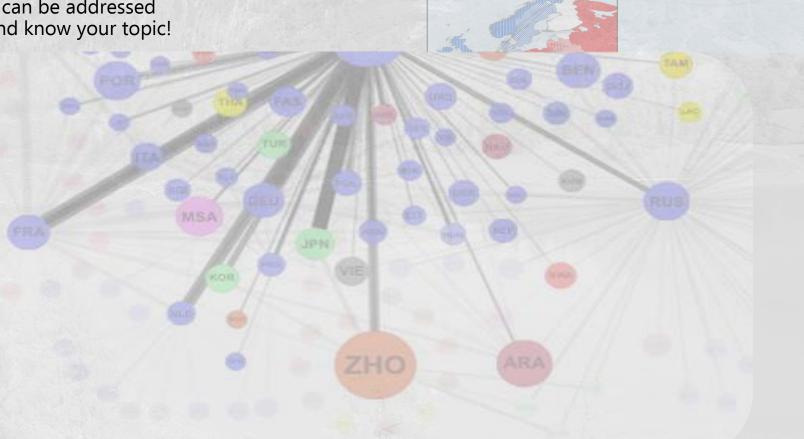




Take-aways

- Language and culture (also non-human) are evolving systems, at least partially inherited
 Very similar methods and concepts to biological evolution
 Huge number of different questions can be addressed
 Be careful can be very sensitive, and know your topic!



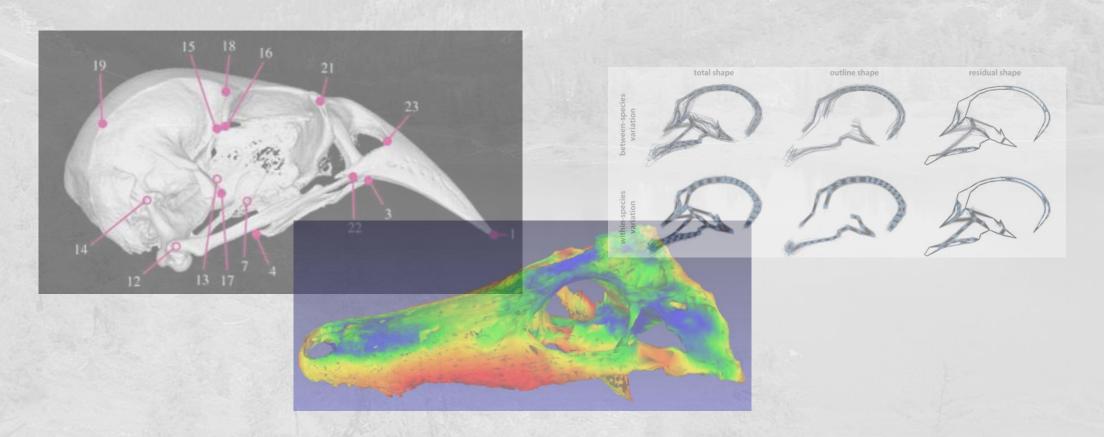






Plan

- Why continuous data?
 What is continuous data and geometric morphometric data?
 Methods of using this for phylogeny
 The effect of ecomorphology





Why use continuous data for phylogeny?

- Life is not discrete!
- Continuous variation, and morphology especially is not easily delimited
- Manual delimitation is often very subjective

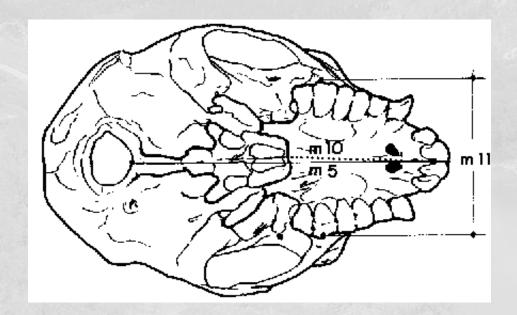
Manual delimitation is very time-consuming

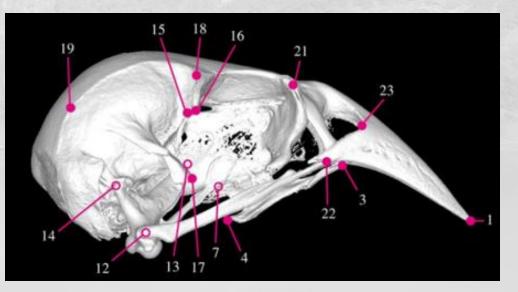




Continuous data: what is it?

- Any measurement data (=standard morphometrics)
 Also geometric morphometric (GMM) data (=coordinate data)





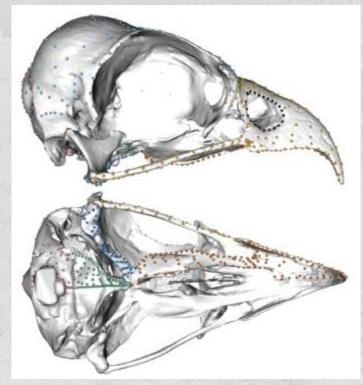
Abzhanov & James 2016



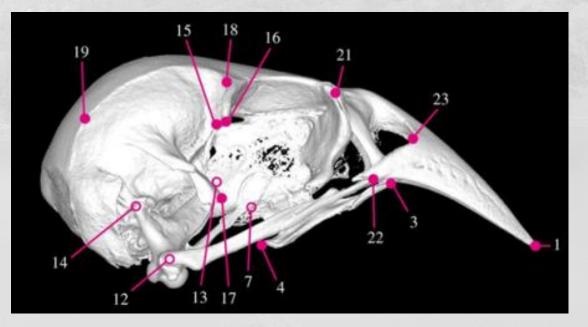
Continuous data: geometric morphometrics

- Typical way of capturing shape variation a lecture in itself!
 Points put on homologous or functionally homologous points in 2D or 3D
- Semilandmark series along sutures, edges
- Semilandmark patches across surfaces

 Can capture more of overall shape variation than traditional morphometrics, and easier to undertake



Felice & Goswami 2018



Abzhanov & James 2016



Continuous data: geometric morphometrics

- Points rotated and scaled (Procrustes analysis) to fit mean shape Often ordinated (PCA) subsequently, concentrating variation into first few axes

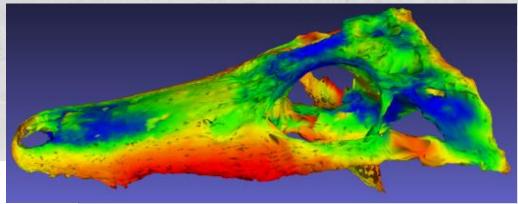




Continuous data: whole surface data

- Subset of GMM data
- Generalized procrustes surface analysis (Pomidor 2016)
- Rotates and scales 3D shapes appropriately, then places landmarks across entire surface automatically Data can be used raw or ordinated just like other GMM data



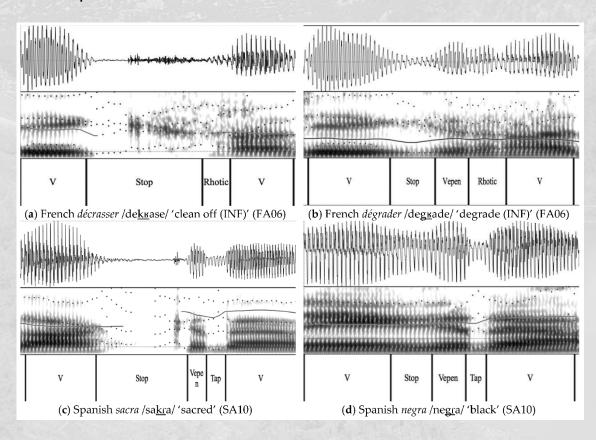


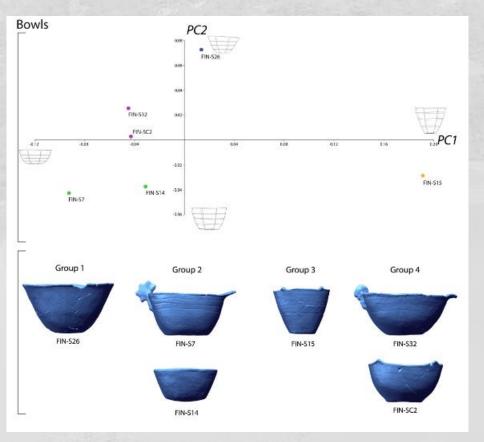




Continuous data: culturolinguistic data

- Phonemes
- Human development indices
- Pot shapes



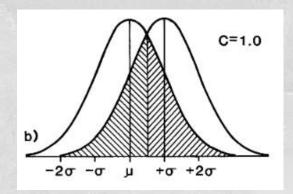


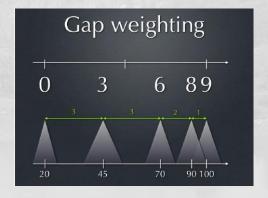
Selden et al. 2014



Methods: automatic discretization

- Several methods developed based on sample variance
 - Gap coding: new state if separation between means greater than one standard deviation
 - Gap weighting: also weights size of gap
 - Not well tested (e.g. against known or molecular phylogeny)
 - Not further discussed here, but would be interesting to test further





Arbitrary discretization into certain number of states, e.g. maximum number of states in programme





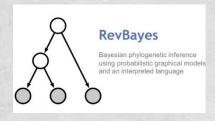


Methods: direct use of continuous data

- Continuous characters can be analyzed directly using distance-based, maximum likelihood or Bayesian methods
- Some evidence that this can be successful

RAXML







Morphometrics and hominoid phylogeny: Support for a chimpanzee-human clade and differentiation among great ape subspecies

Charles A. Lockwood***5, William H. Kimbel**, and John M. Lynch*1

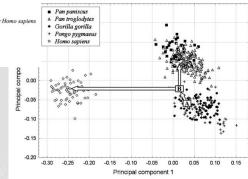
*Institute of Human Origins, †Department of Anthropology, and ¶Barrett Honors College, Arizona State University, Tempe, AZ 85287; and †Department of Anthropology, University College London, London WC1E 6BT, United Kingdom

Edited by David Pilbeam, Harvard University, Cambridge, MA, and approved January 7, 2004 (received for review September 27, 2003)

Taxonomic and phylogenetic analyses of great apes and humans have identified two potential areas of conflict between molecular and morphological data; phylogenetic relationships among living species and differentiation of great ape subspecies. Here we address these problems by using morphometric data. Three-dimensional landmark data from the hominoid temporal bone effectively quantify the shape of a complex element of the skull. Phylogenetic analysis using distance-based methods corroborates the molecular consensus on African ape and human phylogeny, strongly supporting a Pan-Homo clade. Phenetic differentiation of great ape subspecies is pronounced, as suggested previously by mitochon-

also evaluate geometric morphometric methods of studying shape to determine whether they improve on the qualitative and quantitative data that have failed to obtain the hominoid molecular tree (8). Improvement would be expected for several reasons. Landmark data are repeatable to a greater degree than qualitative character assessment, and 3D relationships among landmarks offer greater resolution of shape differences than do chord measurements and angles. We chose to focus on great ape and human temporal bones because this anatomical region (i) presents a complex surface with numerous replicable landmarks and (ii) reflects the influ-





advantage of an underutilized source of skeletal evidence, but we

Use of Continuous Traits Can Improve Morphological Phylogenetics

CAROLINE PARINS-FUKUCHI*

¹Department of Ecology and Ecolutionary Biology, University of Michigan, 830 N. University, Ann Arbor, MI 48109, USA

*Correspondence to be sent to: Department of Ecology and Ecolutionary Biology, University of Michigan, 830 N. University, Ann Arbor, MI 48109, USA:

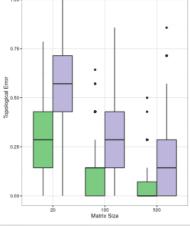
E-mail: effaction/disturbinable.

Received: 05 February 2017; revised 21 August 2017; accepted 30 August 2017 Associate Editor: Luke Harmon

Abstract.—The recent surge in enthusiasm for simultaneously inferring relationships from extinct and extant species has reinvigorated interest in statistical approaches for modeling morphological evolution. Current statistical methods use the Mk model to describe substitutions between discrete character states. Although representing a significant step forward, the Mk model presents challenges in biological interpretation, and its adequacy in modeling morphological evolution has not been well explored. Another major hurdle in morphological phylogenetics concerns the process of character coding of discrete characters. The often subjective nature of discrete character coding can generate discordant results that are rooted in

individual researchers' subjective interpretations. Employees some of these issues. Although not widely used in the characters have been well examined, and their statistica the substantial ambiguity often associated with the ass to determine whether use of continuous characters is a phylogeny. I compare relative reconstruction accuracy characters. These tests demonstrate significant promis as compared to reconstruction from discrete characters equal performance when simulated under an Ornstein well in the presence of covariance between sites. I ar be benefit efforts to maximize phylogenetic informatic space compared to many discretization schemes. I al reconstruction may alleviate potential concerns of dis-further study in this area. This study provides an initi in phylogenetic inference. [Phylogenetics; morphology

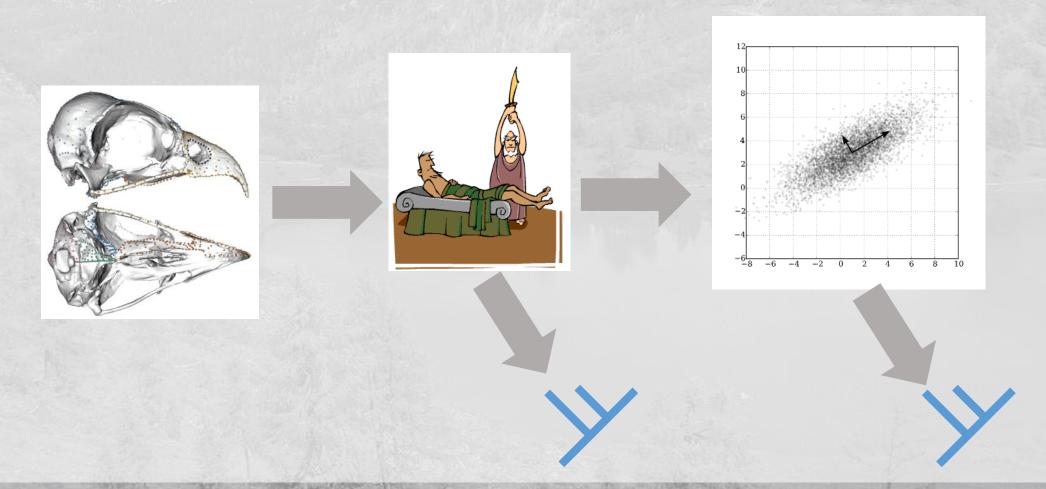
The development and widespread adoption statistical phylogenetic methods has revolution disparate disciplines in evolutionary biology, ep miology, and systematics. Studies utilizing maxim likelihood (ML) and Bayesian approaches have bee the preferred means to analyze molecular data, lar





Methods: geometric morphometrics

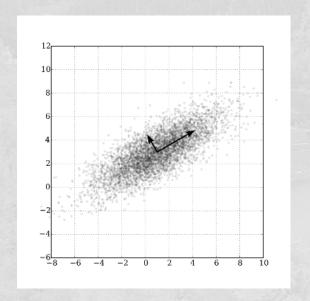
Either usable following Procrustes (coordinates still) or ordination (specimen values)

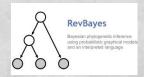




Methods: GMM ordination axis values as characters

- Ordination axis values for taxa can be used as continuous characters
- Should perhaps be weighted by variance on that axis







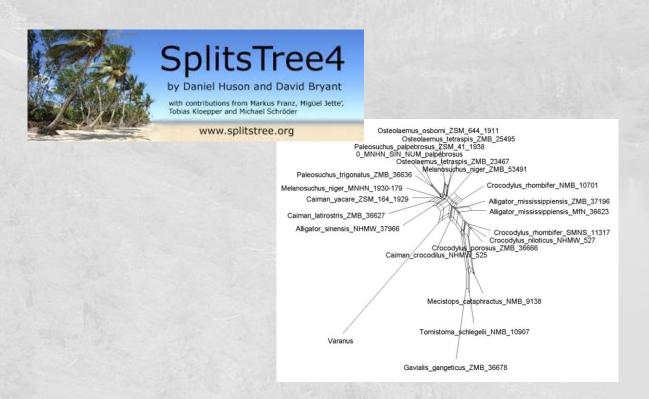


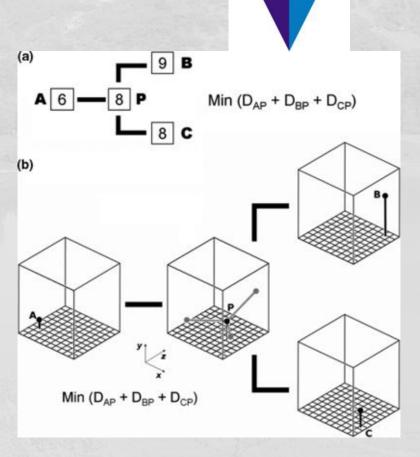




Methods: raw GMM data (after Procrustes)

- Distance-based methods (neighbour joining, UPGMA etc.)
- "Phylogenetic morphometrics" (Catalano et al. 2010), directly implemented in TNT
 - Tree with minimum distances between ancestor-descendant points
 - Analogous to Farris optimization in parsimony
 - Some controversy methodologically

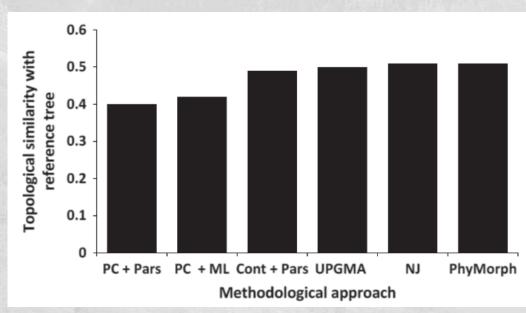




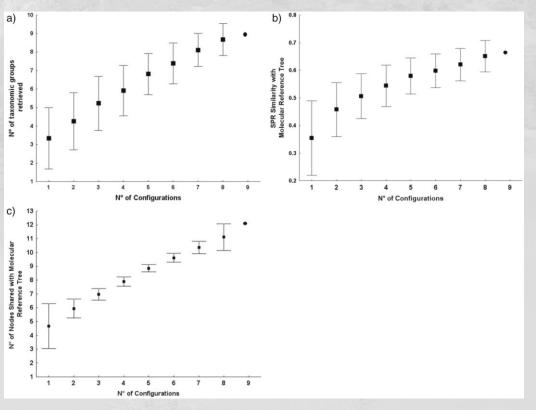


Methods: what's best for GMM data?

- For morphology, raw data seem better than PC values
- Neighbour joining almost as good (and much quicker) than phylogenetic morphometrics Accuracy increases with more skeletal elements



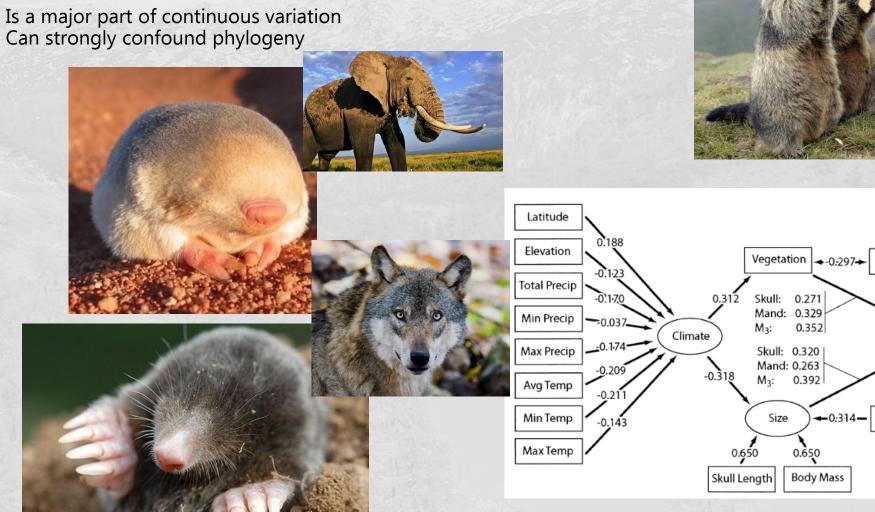
Catalano and Torres 2016



Catalano et al. 2016

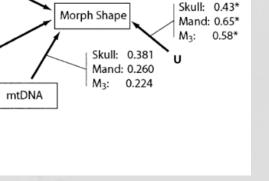


Ecomorphological signal





Diet



Skull: 0.502 Mand: 0.325

M₃: 0.304

Caumul and Polly 2005



Removing ecomorphological signal?

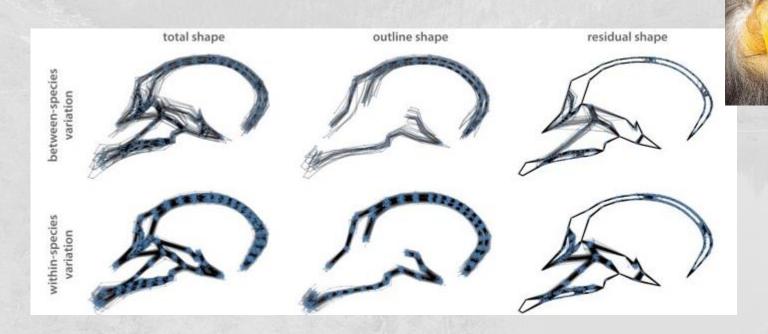
Regress out habitat or behaviour?

Look at specific aspects of shape?

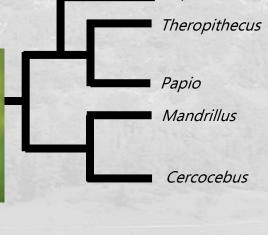
Grunstra et al. (2021) take out overall (=outline) shape

Remaining compositional/structural and local shape show stronger phylogenetic

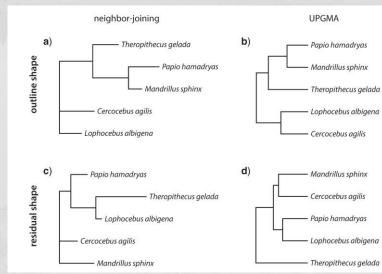
signal and yield better tree



Grunstra et al. 2021



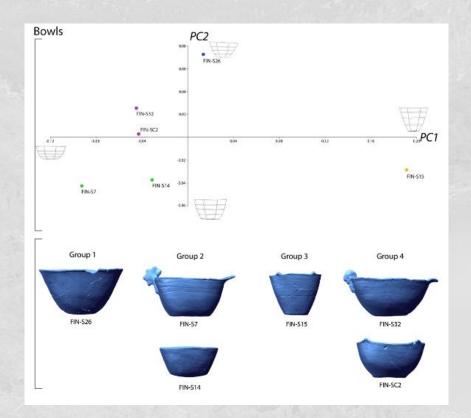
Lophocebus

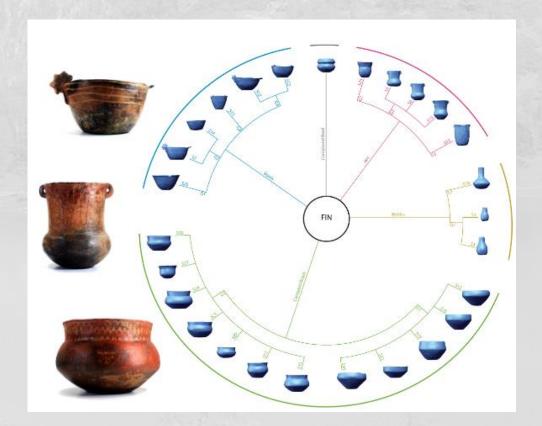


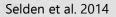


Continuous cultural data

- Little explored, but all same approaches should apply...
 However: very labile and high hybridisation, so may not be as useful if phylogenetic signal is the main interest





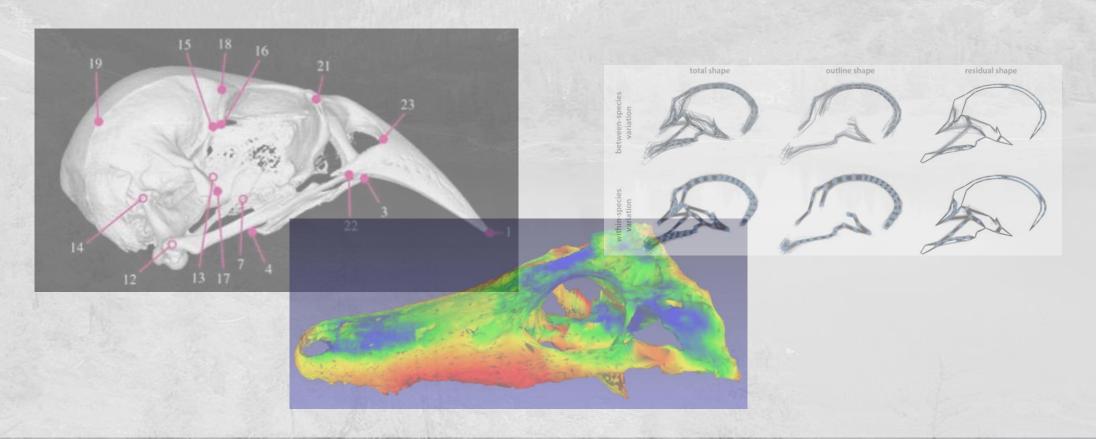




Take-aways

- Continuous data can be used both directly and through discretization

- GMM data can be used directly and after ordination, and has shown some promise Ecological signal is very strong, but can possibly be removed by looking at local/structural shape **Caution:** the field is still new, and attracted much controversy previously due to theoretical concerns over homology





Any questions? Also welcome to email later! sookias.r.b@gmail.com rsookias.info/teaching-resources

EXTRA SLIDES

Discretization from distance trees

- Recently attempted approach (Celik, Phillips)
 Make distance tree from GM data for one strcture
- Basically draw line between major branches to score character

