

Q&A: What is human language, when did it evolve and why should we care?

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Pagel, M. ORCID: <https://orcid.org/0000-0001-7287-8865>
(2017) Q&A: What is human language, when did it evolve and why should we care? BMC Biology, 15 (1). 64. ISSN 1741-7007 doi: 10.1186/s12915-017-0405-3 Available at <https://centaur.reading.ac.uk/71865/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1186/s12915-017-0405-3>

Publisher: BioMed Central

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

QUESTION AND ANSWER

Open Access



Q&A: What is human language, when did it evolve and why should we care?

Mark Pagel

Abstract

Human language is unique among all forms of animal communication. It is unlikely that any other species, including our close genetic cousins the Neanderthals, ever had language, and so-called sign 'language' in Great Apes is nothing like human language. Language evolution shares many features with biological evolution, and this has made it useful for tracing recent human history and for studying how culture evolves among groups of people with related languages. A case can be made that language has played a more important role in our species' recent (circa last 200,000 years) evolution than have our genes.

What is special about human language?

Human language is distinct from all other known animal forms of communication in being *compositional*. Human language allows speakers to express thoughts in sentences comprising subjects, verbs and objects—such as 'I kicked the ball'—and recognizing past, present and future tenses. Compositionality gives human language an endless capacity for generating new sentences as speakers combine and recombine sets of words into their subject, verb and object roles. For instance, with just 25 different words for each role, it is already possible to generate over 15,000 distinct sentences. Human language is also *referential*, meaning speakers use it to exchange specific information with each other about people or objects and their locations or actions.

What is animal 'language' like?

Animal 'language' is nothing like human language. Among primates, vervet monkeys (*Chlorocebus pygerythrus*) produce three distinct alarm calls in response to the presence of snakes, leopards and eagles [1]. A number of parrot species can mimic human sounds, and

some Great Apes have been taught to make sign language gestures with their hands. Some dolphin species seem to have a variety of repetitive sound motifs (clicks) associated with hunting or social grouping. These forms of animal communication are symbolic in the sense of using a sound to stand in for an object or action, but there is no evidence for compositionality, or that they are truly generative and creative forms of communication in which speakers and listeners exchange information [2].

Instead non-human animal communication is principally limited to repetitive instrumental acts directed towards a specific end, lacking any formal grammatical structure, and often explainable in terms of hard-wired evolved behaviours or simple associative learning [2]. Most ape sign language, for example, is concerned with requests for food. The trained chimpanzee Nim Chimpsky's longest recorded 'utterance', when translated from sign language, was 'give orange me give eat orange me eat orange give me eat orange give me you' [3]. Alarm calls such as observed in the vervet monkeys often evolve by kin-selection to protect one's relatives, or even selfishly to distract predators away from the caller. Hunting and social group communications can be explained as learned coordinating signals without 'speakers' knowing why they are acting as they are.

When did human language evolve?

No one knows for sure when language evolved, but fossil and genetic data suggest that humanity can probably trace its ancestry back to populations of anatomically modern *Homo sapiens* (people who would have looked like you and me) who lived around 150,000 to 200,000 years ago in eastern or perhaps southern Africa [4–6]. Because all human groups have language, language itself, or at least the capacity for it, is probably at least 150,000 to 200,000 years old. This conclusion is backed up by evidence of abstract and symbolic behaviour in these early modern humans, taking the form of engravings on red-ochre [7, 8].

Correspondence: m.pagel@reading.ac.uk
School of Biological Sciences, University of Reading, Reading RG6 6UR, UK

The archaeological record reveals that about 40,000 years ago there was a flowering of art and other cultural artefacts at modern human sites, leading some archaeologists to suggest that a late genetic change in our lineage gave rise to language at this later time [9]. But this evidence derives mainly from European sites and so struggles to explain how the newly evolved language capacity found its way into the rest of humanity who had dispersed from Africa to other parts of the globe by around 70,000 years ago.

Could language be older than our species?

Ancient DNA reveals us to be over 99% identical in the sequences of our protein coding genes to our sister species the Neanderthals (*Homo neanderthalensis*) [10]. The Neanderthals had large brains and were able to inhabit much of Eurasia from around 350,000 years ago. If the Neanderthals had language, that would place its origin at least as far back as the time of our common ancestor with them, currently thought to be around 550,000 to 750,000 years ago [10, 11].

However, even as recently as 40,000 years ago in Europe, the Neanderthals show almost no evidence of the symbolic thinking—no art or sculpture for example—that we often associate with language, and little evidence of the cultural attainments of *Homo sapiens* of the same era. By 40,000 years ago, *Homo sapiens* had plentiful art, musical instruments and specialized tools such as sewing needles. Neanderthals probably didn't even have sewn clothing, instead they would have merely draped themselves with skins [12]. And, despite evidence that around 1–5% of the human genome might be derived from human–Neanderthal matings [13], the Neanderthals went extinct as a species while we flourished.

Can genetic evidence help to decide when language evolved?

Yes. Modern humans and Neanderthals share a derived version of a transcription factor gene known as *FOXP2* that differs from the chimpanzee version by two amino acid replacements [14]. *FOXP2* influences the fine-motor control of facial muscles required for the production of speech. Indeed, inserting this derived form into mice causes them to squeak differently [15]! However, in spite of having identical primary sequences to Neanderthals, modern humans have acquired changes to the regulation of their *FOXP2* genes that seem likely to cause their *FOXP2* to be expressed differently to that of the Neanderthals [16], and these expression differences are pronounced in brain neurons. Combining these genetic hints with the differences in symbolic and cultural behaviour that are evident from the fossil record suggests language arose in our lineage sometime after our split

from our common ancestor with Neanderthals, and probably by no later than 150,000 to 200,000 years ago.

Was there a single origin of language?

This question has parallels in biological evolution. Did life evolve once or many times? The presence of the same RNA and DNA in all organisms and homologies in the machinery of DNA transcription and translation suggest that at least all current life on Earth has a common origin. It is possible that life evolved more than once but all descendants of these other origins went extinct and left no fossil or other traces.

With language the inference is harder to make because features such as vocabulary and grammar change too rapidly to be able to link all of the world's languages to a common original mother tongue. On the other hand, all human languages rely on combining sounds or 'phones' to make words, many of those sounds are common across languages, different languages seem to structure the world semantically in similar ways [17], all human languages recognize the past, present and future and all human languages structure words into sentences [18]. All humans are also capable of learning and speaking each other's languages (some phones are unique to some language families—such as the famous 'click' sound of some San languages of Southern Africa—but these are probably within the capability of all human speakers if they are exposed to learning that sound at the right time of life).

These considerations suggest that the anatomical, neurological and physiological underpinnings of language are shared among all of humanity. If the capacity for language did evolve more than once, all traces of it seem to have been lost. This conclusion is buttressed by the *FOXP2* evidence (all humans share the same derived gene) and by the fact that genetic data point to all modern humans descending from a common ancestor [19].

Is language evolution like biological evolution?

Darwin observed that “The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously the same” (page 59 in [20]). He also asserted that “The survival and preservation of certain favoured words in the struggle for existence is natural selection.” (pages 59–60 in [20]).

Darwin was right on both counts. Linguists have known from at least the late 18th century [21]—about 100 years before Darwin—that languages predominantly evolve by a process of descent with modification from earlier ancestral languages, just as biological species descend from earlier ancestral forms. An example is differences observed between the ancient Greek vocabulary in Homer's *Iliad* from around 750 BCE and modern Greek

vocabulary (Table 1) [22]: some words have merely changed their pronunciation while others have been replaced by new unrelated words.

Regarding Darwin's assertions that certain words are favoured in the 'struggle for existence', it is useful to remember that there is seldom any connection between a sound (a word) and its meaning. This means that selection is reasonably free to choose among words and so features of the words we actually use might reveal its actions. The simplest example is that words that are used more often—such as *I*, *he*, *she*, *it*, *the*, *you*—tend to be shorter, and consequently easier to pronounce, than less frequently used words, such as *obstreperous* or *catafalque* [23]. This is an example of a form of natural selection except here instead of biological individuals competing in the physical environment to survive and reproduce, words compete for space in the environment of the human mind. Our minds give preference to shorter versions of the frequently used words, presumably to reduce effort [23]. This pressure is relaxed among the less frequently used words, allowing them to be longer. It might also be the case that once the frequently used words have occupied the space of possible short words, there are fewer opportunities for the less frequently used words [24].

Is it possible to reconstruct the history of a group of languages like we do with species?

Yes. Using common lists of words that are found in all or nearly all languages, linguists can identify shared sets of *cognate* words—words that descend from common ancestral words—just as it is possible to identify *homologous* genes that share a common ancestral gene. For instance the Spanish *mano* ('hand') and the French *main* descend from the earlier Latin *manus*, while the English and German words *hand* do not. A cognate set identifies

groups of related languages. In the example here *mano* and *main* identify the so-called Romance languages (Spanish, French, Italian, Portuguese) and *hand* and *hand* identify the Germanic languages (Fig. 1). By combining the information in many different cognate sets with appropriate statistical models [25, 26], it is possible to infer detailed family histories or phylogenetic trees of language families, such as has been done for the Indo-European languages (Fig. 1). These phylogenies are directly analogous to phylogenies of biological species.

What other evolutionary features do genes and language share?

Linguistic and biological evolution share features beyond descent with modification and selection, including mechanisms of mutation and replication, speciation, drift and horizontal transfer (Table 2). At a deeper level, both genes and languages can be represented as digital systems of inheritance, built on the transmission of discrete chunks of information—genes in the case of biological organisms, and words in the case of language. Genes in turn comprise combinations of the four bases or nucleotides (A, C, G, T) while words can be modelled as comprising combinations of discrete sounds or phones (in fact, phones or sounds vary in a continuous space but languages are commonly represented as expressing a particular set of discrete phonemes).

These similarities mean that we can—and should—think of language as a system for the transmission of information that is tantamount to 'aural DNA'. Even the peculiar phenomenon of concerted evolution in genetics—where a nucleotide replacement at a specific site in one gene is quickly followed by the same nucleotide replacement at the same site in other, typically related, genes—is also observed in language. Known as *regular sound change*, a specific phone or sound changes over a relatively short period of time to the same other phone in many words in the lexicon [27, 28]. A well-known example is the *p* → *f* sound change in the Germanic languages where an older Indo-European *p* sound was replaced by an *f* sound, such as in *pater* → *father*; or *pes*, *pedis* → *foot*.

Can changes to language be used to trace human history?

There are currently about 7000 languages spoken around the world, meaning that, oddly, most of us cannot communicate with most other members of our species! Even this number is probably down from the peak of human linguistic diversity that was likely to have occurred around 10,000 years ago, just prior to the invention of agriculture [29]. Before that time, all human groups had been hunter-gatherers, living in small mobile tribal societies. Farming societies were demographically

Table 1 Linguistic descent with modification spanning nearly three millennia^a

Word	Homeric Greek	Modern Greek
One	eis (εἷς)	ena (ένα)
Day	emar (ἡμαρ)	emera (ἡμέρα)
Two	dyo (thuo) (δύο, δύο)	dyo (δύο)
Father	pater (πατήρ)	pateras (πατέρας)
To eat	etho (έδω)	faei (φάει)
Bird	ornis (όρνις)	pouli (πούλι)

^aHomeric Greek of the Iliad dates from ~750 BCE. Words in bold have been replaced in Modern Greek

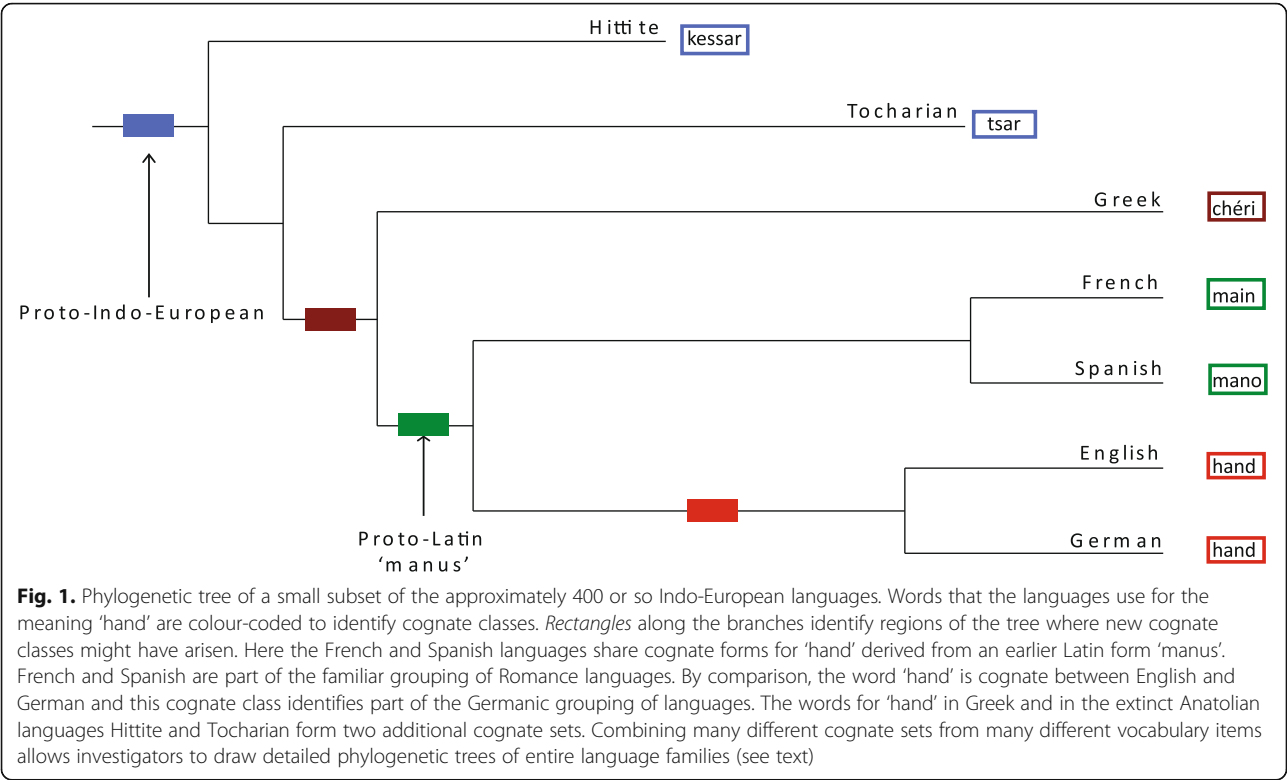


Table 2 Some parallels between biological and linguistic evolution

Biological evolution	Language evolution
Discrete heritable units (for example, nucleotides, amino acids and genes)	Discrete heritable units (for example, words, phonemes and syntax)
DNA copying	Teaching, learning and imitation
Mutation (for example, many mechanisms yielding genetic alterations)	Innovation (for example, formant variation, mistakes, sound changes, and introduced sounds and words)
Homology	Cognates
Natural selection	Social selection and trends
Drift	Drift
Speciation	Language or cultural splitting
Concerted evolution	Regular sound change
Horizontal gene transfer	Borrowing
Hybridization (for example, horse with zebra and wheat with strawberry)	Language Creoles (for example, Surinamese)
Geographic clines	Dialects and dialect chains
Fossils	Ancient texts
Extinction	Language death

more prosperous and group sizes were larger than among hunter-gatherers, so the expansion of agriculturalists likely replaced many smaller linguistic groups. Today, there are few hunter-gatherer societies left so our linguistic diversity reflects our relatively recent agricultural past.

Phylogenies of languages can be used in combination with geographical information or information on cultural practices to investigate questions of human history, such as the spread of agriculture. Phylogenies of language families have been used to study the timing, causes and geographic spread of groups of farmers/fishing populations, including the Indo-Europeans [30–33]; the pace of occupation of the Pacific by the Austronesian people [34]; and the migration routes of the Bantu-speaking people through Africa [35, 36].

Linguistic phylogenies are also used to investigate questions of human cultural evolution, including the evolution and spread of dairying [37–39], relationships between religious and political practices [40], changing political structures [41] and the age of fairy tales [42], and have even supplied a date for Homer’s Iliad [22].

What role has language played in our species’ success?

Language has played a prominent and possibly pre-eminent role in our species’ history. Consider that where all other species tend to be found in the environments

their genes adapt them to, humans can adapt at the cultural level, acquiring the knowledge and producing the tools, shelters, clothing and other artefacts necessary for survival in diverse habitats [12, 43]. Thus, chimpanzees are found in the dense forests of Africa but not out on the savannah or in deserts or cold regions; camels are found in dry regions but not in forests or mountaintops, and so on for other species. Humans, on the other hand, despite being a species that probably evolved on the African savannahs, have been able to occupy nearly every habitat on Earth. Our behaviour is like that of a collection of biological species [43]. Why this striking difference?

It is probably down to language. Possessing language, humans have had a high-fidelity code for transmitting detailed information down the generations. Many, if not most, of the things we make use of in our everyday lives rely on specialized knowledge or skills to produce. The information behind these was historically coded in verbal instructions, and with the advent of writing it could be stored and become increasingly complex.

Possessing language, then, is behind humans' ability to produce sophisticated cultural adaptations that have accumulated one on top of the other throughout our history as a species. Today as a result of this capability we live in a world full of technologies that few of us even understand. Because culture, riding on the back of language, can evolve more rapidly than genes, the relative genetic homogeneity of humanity in contrast to our cultural diversity shows that our 'aural DNA' has probably been more important in our short history than genes.

Acknowledgements

An Advanced Investigator Award 268744 to M. Pagel from the European Research Council has supported most of my recent research on language evolution.

Competing interests

The author declares that he has no competing interests.

Published online: 24 July 2017

References

1. Seyfarth RM, Cheney DL, Marler P. Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim Behav*. 1980;28(4):1070–94.
2. Fitch WT. *The evolution of language*. Cambridge University Press; 2010.
3. Terrace HS. How Nim Chimsky changed my mind. *Psychol Today*. 1979; 13(6):65.
4. Fleagle JG, Assefa Z, Brown FH, Shea JJ. Paleanthropology of the Kibish Formation, southern Ethiopia: introduction. *J Hum Evol*. 2008;55(3):360–5.
5. Poznik GD, Henn BM, Yee M-C, Sliwerska E, Euskirchen GM, Lin AA, et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science*. 2013;341(6145):562–5.
6. Brown FH, McDougall I, Fleagle JG. Correlation of the KHS Tuff of the Kibish Formation to volcanic ash layers at other sites, and the age of early Homo sapiens (Omo I and Omo II). *J Hum Evol*. 2012;63(4):577–85.
7. Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GA, et al. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science*. 2002;295(5558):1278–80.
8. Henshilwood CS, Dubreuil B. Reading the artefacts: gleanings language skills from the Middle Stone Age in southern Africa. *Cradle Language*. 2009;261–92.

9. Klein RG. *The human career: human biological and cultural origins*. University of Chicago Press; 2009.
10. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 2014;505(7481):43–9.
11. Meyer M, Arsuaga J-L, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*. 2016;531(7595):504–7.
12. Pagel M. *Wired for culture: origins of the human social mind*. WW Norton & Company; 2012.
13. Kuhlwillm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature*. 2016;530(7591):429–33.
14. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, et al. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*. 2002; 418(6900):869–72.
15. Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T, Somel M, et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*. 2009;137(5):961–71.
16. Maricic T, Günther V, Georgiev O, Gehre S, Čurlin M, Schreiwies C, et al. A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Mol Biol Evol*. 2013;30(4):844–52.
17. Youn H, Sutton L, Smith E, Moore C, Wilkins JF, Maddieson I, et al. On the universal structure of human lexical semantics. *Proc Natl Acad Sci U S A*. 2016;113(7):1766–71.
18. Greenberg JH. Some universals of grammar with particular reference to the order of meaningful elements. In: Greenberg JH, editor. *Universals of language*. Cambridge: MIT Press; 1963. p. 73–113.
19. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovskiy LA, et al. Genetic structure of human populations. *Science*. 2002; 298(5602):2381–5.
20. Darwin CR. *The descent of man and selection in relation to sex: in two volumes: with illustrations*. J. Murray; 1871.
21. Jones W. *Discourses delivered before the Asiatic Society: and miscellaneous papers, on the religion, poetry, literature, etc., of the nations of India*. CS Arnold; 1824.
22. Altschuler EL, Calude AS, Meade A, Pagel M. Linguistic evidence supports date for Homeric epics. *BioEssays*. 2013;35(5):417–20.
23. Zipf GK. *Human behaviour and the principle of least-effort*. Cambridge MA edn. Reading: Addison-Wesley; 1949.
24. Pagel M, Meade A. The deep history of the number words. *Phil Trans R Soc B*. in press.
25. Pagel M. Human language as a culturally transmitted replicator. *Nat Rev Genet*. 2009;10(6):405–15.
26. Pagel M. Darwinian perspectives on the evolution of human languages. *Psychonomic Bull Rev*. 2017;24(1):151.
27. Crowley T, Bower C. *An introduction to historical linguistics*. Oxford University Press; 2010.
28. Hruschka DJ, Branford S, Smith ED, Wilkins J, Meade A, Pagel M, et al. Detecting regular sound changes in linguistics as events of concerted evolution. *Curr Biol*. 2015;25(1):1–9.
29. Pagel M. The history, rate and pattern of world linguistic evolution. In: Knight C, Studdert-Kennedy M, Hurford J, editors. *The evolutionary emergence of language: social function and the origins of linguistic form*. Cambridge University Press; 2000.
30. Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, et al. Mapping the origins and expansion of the Indo-European language family. *Science*. 2012;337(6097):957–60.
31. Chang W, Cathcart C, Hall D, Garrett A. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*. 2015; 91(1):194–244.
32. Gray RD, Atkinson QD. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*. 2003;426(6965):435–9.
33. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015;522(7555):207–11.
34. Gray RD, Drummond AJ, Greenhill SJ. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*. 2009; 323(5913):479–83.
35. Currie TE, Meade A, Guillon M, Mace R. Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proc R Soc B*. 2013;280(1762):20130695.

36. Grollemund R, Branford S, Bostoen K, Meade A, Venditti C, Pagel M. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc Natl Acad Sci U S A*. 2015;112(43):13296–301.
37. Holden C, Mace R. Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum Biol*. 2009;81(5/6):597–619.
38. Holden CJ, Mace R. Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proc R Soc Lond B Biol Sci*. 2003;270(1532):2425–33.
39. Mace R, Jordan F, Holden C. Testing evolutionary hypotheses about human biological adaptation using cross-cultural comparison. *Comp Biochem Physiol A Mol Integr Physiol*. 2003;136(1):85–94.
40. Watts J, Greenhill SJ, Atkinson QD, Currie TE, Bulbulia J, Gray RD. Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proc Biol Sci*. 2015; 282(1804):20142556.
41. Currie TE, Greenhill SJ, Gray RD, Hasegawa T, Mace R. Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature*. 2010; 467(7317):801–4.
42. Da Silva SG, Tehrani JJ. Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *R Soc Open Sci*. 2016;3(1):150645.
43. Pagel M, Mace R. The cultural wealth of nations. *Nature*. 2004;428(6980):275–8.