



HAL
open science

Coevolution of languages and genes

Brigitte Pakendorf

► **To cite this version:**

Brigitte Pakendorf. Coevolution of languages and genes. *Current Opinion in Genetics and Development*, 2014, 29, pp.39-44. 10.1016/j.gde.2014.07.006 . halshs-01178752

HAL Id: halshs-01178752

<https://shs.hal.science/halshs-01178752>

Submitted on 16 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coevolution of languages and genes

[article published as: Pakendorf, Brigitte (2014): Coevolution of languages and genes. *Current Opinion in Genetics and Development* 29: 39-44; <https://doi.org/10.1016/j.gde.2014.07.006>]

Brigitte Pakendorf

Laboratoire Dynamique du Langage, UMR5596, CNRS & Université Lyon Lumière 2

postal address:

ISH-DDL

14 avenue Berthelot

69363 Lyon Cedex 07

FRANCE

Brigitte.Pakendorf@cnrs.fr

Tel. (office): +33-4-72 72 64 10

Tel. (mobile): +49-1578-779 61 91

Abstract

The evolution of languages shares certain characteristics with that of genes, such as the predominantly vertical line of transmission and the retention of traces of past events such as contact. Thus, studies of language phylogenies and their correlations with genetic phylogenies can enrich our understanding of human prehistory, while insights gained from genetic studies of past population contact can help shed light on the processes underlying language contact and change. As demonstrated by recent research, these evolutionary processes are more complex than simple models of gene-language coevolution predict, with linguistic boundaries only occasionally functioning as barriers to gene flow. More frequently, admixture takes place irrespective of linguistic differences, but with a detectable impact of contact-induced changes in the languages concerned.

Introduction

Ever since Darwin [1] it has been assumed that the genetic and linguistic evolution of humans should be largely correlated [2–4]. The reasons for this lie in the perceived parallelism of genes and languages: genes are obligatorily passed on from parents to their offspring, and in general the first language children learn to speak is that of their parents, so that languages tend to be inherited in a vertical line as well. Similarly, genepools diverge when populations become substructured and increasingly isolated from each other; such reproductive isolation would also lead to communicative isolation and thus linguistic substructuring and the development of new languages out of an erstwhile common ancestor. Furthermore, both genes and languages can retain traces of past demographic events such as contact, detectable via genetic admixture on the one hand and linguistic changes (loanwords or structural borrowings) on the other.

Thus, the correlations (as well as the lack thereof) between genetic and linguistic relationships can help shed light on the (pre)history of human populations and enrich our understanding of the processes that shape both genetic and linguistic diversity [5]. This relatively young but burgeoning field of interdisciplinary research can be approached from three different angles of investigation: 1) the coevolution of genes and languages, 2) prehistoric population contact and its effect on language evolution and change, and 3) the demographic history of language families to shed light on the prehistory of the peoples speaking these languages. I here review each of these approaches in turn for readers with no background in linguistics, covering to a large extent the last five years. Since the review by Pagel [6] focusses largely on mechanisms of language evolution and thus does not cover all the aspects touched upon here, some references to the older background literature are also included. However, I do not cover the question of how human language may have evolved in the first place; for some recent discussion see [7] and [8].

Coevolution of genes and languages

The investigation of language-gene coevolution was first undertaken in the late 1980s and early 1990s, when sufficient allele frequency data for a large number of human populations had been collected to make such research feasible [9]. Major questions of interest concern the extent to which linguistic differences present barriers to gene flow and thus shape genetic diversity [4], as suggested by the results of early studies [2,3,10,11], as well as whether the coevolution of genes and languages follows a branching model marked by successive splits and isolation, or rather an isolation-by-distance model with decreasing genetic and linguistic exchange over increasing geographical distances [12]. While some studies have found that genetic structure indeed correlates with linguistic affiliation [4,13], implying that language can represent a barrier to gene flow, others have found that at regional levels gene flow and language contact have erased such patterns of phylogenetic splits [12,14].

In order to investigate the coevolution of genes and languages, correlations between matrices of linguistic and genetic distances are most commonly investigated. This poses a methodological problem: while the genetic distances are based on actual empirical data that are tailored to the markers used, until recently the linguistic distances were arbitrary values assigned on the basis of a (controversial) phylogeny of languages compiled by Ruhlen [15]. In such approaches, languages belonging to different linguistic phyla are assigned the largest value and languages at lower levels in the phylogeny receive smaller values, e.g. [4,16]; the validity of results obtained by such comparisons is of course rather questionable. More recent studies have circumvented the methodological problem of defining appropriate linguistic distances by calculating these with empirical data – frequently the number of retained cognate items (i.e. words going back to a common ancestor) in word lists [13,14,17].

However, calculating linguistic distances from lexical cognates is restricted to relatively closely related languages, as sound changes and replacement of words reduce the number of detectable cognates with time; it is commonly assumed that the time-depth for the establishment of genealogical relationships based on lexical cognates is 7,000-10,000 years [18,19]. Since genetic data are not subject to such temporal constraints, it is necessary to find linguistic measures that are amenable to comparison even across very distinct language families if one wants to investigate gene-language coevolution at a global scale. Structural data, i.e. abstract grammatical features such as the order of subject, verb, and object or the presence/absence of definite or indefinite articles, have been suggested as potentially more suitable for the investigation of genealogical relationships at deeper time depth [19–21]. Thus, using such structural features Dunn et al. [22] were able to reconstruct a phylogeny of Papuan languages of Island Melanesia which are unclassifiable using lexical data.

A further such attempt at avoiding the temporal restriction of linguistic data is the "Parametric Comparison Method" [23–25]. This is based on so-called parameter settings, which are abstract feature values at a varying number of syntactic features supposedly "... predefined by our invariant language faculty, Universal Grammar ..." [24: 1684]. Since this approach assumes that these parameters are part of the innate Universal Grammar, they should be found, and hence comparable, across all languages irrespective of their degree of genealogical relationship [23,24] – making them perfectly comparable to genetic data and avoiding the problems inherent in the use of lexical data mentioned above. Therefore this approach appears

to be the ideal solution for the investigation of genetic and linguistic coevolution at a global scale. However, there are several issues that diminish its value. The biggest problem concerns the list of supposedly universal parameters – which is yet to be defined. Even the proponents of the idea admit that "UG parameters number at least in the hundreds, although we are too far from being able to make precise estimates" [24: 1687], while a survey of the relevant literature was unable to find more than seven parameters that were mentioned by more than one author, none of which were uncontroversial among specialists [26]. In addition, the ascertainment scheme for the parameters is heavily biased: most are proposed on the basis of data from only individual language families or even subfamilies, or at most on a comparison of two very distinct languages, e.g. English and Japanese [26]. Furthermore, Longobardi and colleagues propose to restrict their investigation to the domain of nominal arguments, such as *Mary*, *Mary's book*, *the person I spoke to*, etc [23,24], in order to circumvent the problem of the potentially large number of as yet unidentified parameters for which insufficient data are available across numerous languages. At last count their dataset comprised only 56 parameters, not all of which are independent of each other – restricting the applicability of phylogenetic tools that assume independence of data [25]. Thus their linguistic distance measures are based on a very small set of features ascertained in a small number of languages and taken from a very limited domain of grammar (comparable to a genetic study of global diversity limiting itself to a small number of partly linked SNPs ascertained in only a few populations), which casts considerable doubt on the validity of the results that can be obtained with this method.

While initial analyses of language-gene coevolution simply assessed the degree of correlation between linguistic and genetic diversity, increasingly such studies are used to address specific hypotheses, e.g. [12,14,27]. For instance, de Filippo et al. [14] used correlations between genetic distances, linguistic distances calculated from cognate lexemes in word lists, and model-based geographic distances to investigate which route would have been followed by the expanding speakers of Bantu languages in sub-Saharan Africa. They find that the data fit a model of a relatively late split of eastern and western Bantu languages, although the signal of this split would have subsequently been diminished by gene flow and language contact (Figure 1). At a broader geographic scale, Hunley et al. [27] undertook a worldwide comparison of the diversity of autosomal microsatellites and that of inventories of distinctive sounds (phonemes) to elucidate whether serial founder events would have shaped both systems equally. With this study they partly addressed the claim by Atkinson [28] that phoneme inventories show evidence of a serial founder event correlating with the Out-of-Africa dispersal of modern humans. In contrast to Atkinson's results, Hunley et al. find no evidence for serial founder events in the phoneme data. This different result is most probably due to the very different datasets the studies were based on: Atkinson [28] made use of a published dataset [29] in which phoneme inventories of 504 languages were classified into size bins (e.g. 'small', 'medium', and 'large' for vowel inventories). In contrast, Hunley et al. [27] based their study on 908 phonemes coded as present or absent for 725 languages.

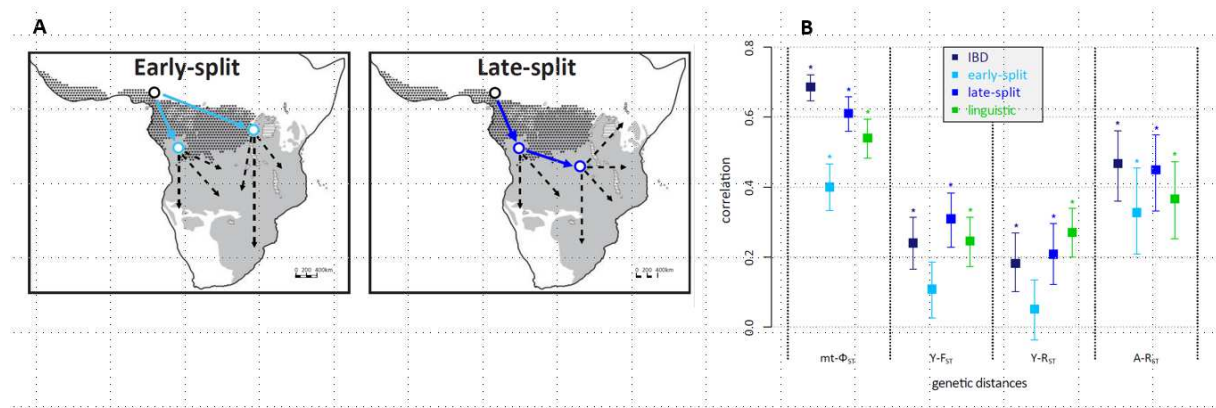


Figure 1: Using genetic and linguistic data to test models of the Bantu expansion. A: Schematic of the expansion models tested. B: Correlations between genetic distances and model-based geographic distances (isolation-by-distance, early-split, late-split) as well as linguistic distances; asterisks mark statistically significant correlations. IBD = isolation by distance; mt = mitochondrial DNA; Y = Y-chromosome; A = autosomal microsatellites. (Slightly modified from de Filippo et al. 2012; their Figure 1 and Figure 6.)

Prehistoric population contact and its effect on language evolution and change

As discussed in the preceding section, it is becoming increasingly clear that at small geographical scales isolation by distance – involving gene flow and the transfer of linguistic features – plays a large role in shaping linguistic and genetic diversity. In order to better understand the processes underlying these isolation-by-distance effects, molecular anthropological investigations can be coupled with fine-scaled linguistic studies. Several studies of African populations have highlighted the role played by female-biased intermarriage in the adoption and spread of foreign linguistic features. For instance, a study of Bantu-speaking populations from southwestern Zambia [30] was able to demonstrate that large-scale incorporation of autochthonous women by the Fwe population is likely to have triggered the incorporation of click consonants into the Fwe language. These highly salient consonants are characteristic of three language families indigenous to southern Africa collectively called 'Khoisan'. The spread of these consonants in Fwe from words borrowed from a Khoisan language to some words of Bantu origin [31] could have been a means to flag the new identity of a community characterized by a large proportion of maternal ancestors of Khoisan origin.

A study of populations from central Burkina Faso speaking languages belonging to two very distinct branches of the Niger-Congo phylum, Mande and Gur [32] demonstrated the complex role played by linguistic differences in the structuring of genetic diversity. On the one hand, there is absolutely no differentiation among these populations in the maternal lineages, irrespective of which language they speak, so that linguistic differences clearly are no barrier at all to gene flow in the maternal line. In contrast, the paternal lineages are discernibly structured by linguistic affiliation, with a highly significant F_{ST} value of 9.4% when comparing Mande-speakers with Gur-speakers, as opposed to 8.8% variance within linguistic groups. These results are in good accordance with the strictly patrilocal social structure of

these populations, with women moving to their husbands' homestead after marriage, where they have to speak the language of the homestead. This high mobility of women across linguistic boundaries is most probably the reason for the spread of shared grammatical features across several languages of the region [33], since women would be transmitting their native language patterns to their children.

Genetic studies of Khoisan populations [34–36] have further demonstrated the complex relationship between linguistic affiliation and genetic affinity. While on the one hand populations speaking languages of the Kx'a family are found on both sides of a deep genetic split dated to approximately 30,000 years ago, on the other hand some populations speaking languages belonging to all three accepted Khoisan language families [37] are closely related [34] (Figure 2). The shared genetic affinities of these populations are partly reflected in contact-induced changes in their languages [38–40]. This again demonstrates that gene flow can easily take place across linguistic boundaries, even when these involve very distinct languages belonging to different language families, and that such genetic admixture can have an impact on the languages in contact.

Other studies, too, have demonstrated considerable variation in the degree to which genetic variation of human populations is structured by their linguistic affiliation. In the Pacific there are generally no consistent genetic differences between populations speaking Papuan languages and those speaking Austronesian languages [41–43], whereas a study of autosomal SNPs in Ethiopian populations found considerable structure along linguistic boundaries [44], as did a study of Y-chromosomal variation in the North Caucasus [13].

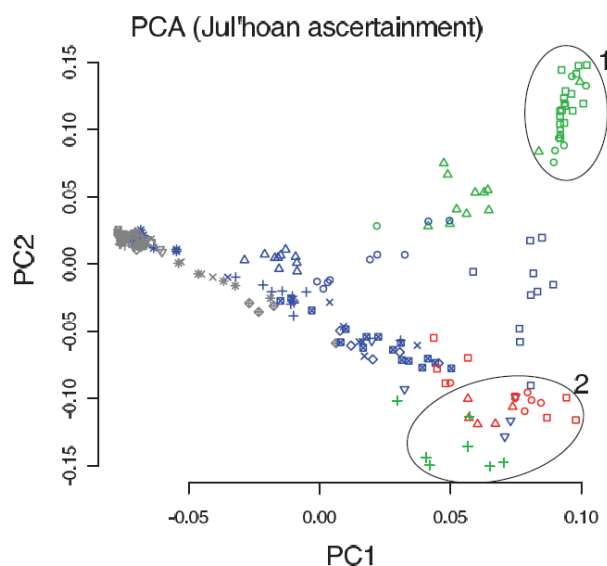


Figure 2: Languages and genes do not necessarily correlate. Principal Components Analysis on autosomal SNPs in individuals from southern Africa. The colour and shape of the symbols designate the ethnolinguistic affiliation of each individual, with red, green, and blue representing the three different Khoisan language families Tuu, Kx'a, and Khoe-Kwadi, respectively, and grey representing the Bantu language family. While genetic and linguistic relationships sometimes go hand-in-hand (see for instance the genetically closely related individuals (circle 1), who speak closely related languages belonging to the Kx'a family), this is not always the case: thus, the individuals clustering in the bottom right corner (circle 2)

speak languages belonging to all three Khoisan language families. Note also that populations speaking different languages of the Kx'a family (green symbols) are genetically distinct. (Slightly modified from Pickrell et al. 2012, their Figure 1b.)

Language family phylogenies and speaker prehistory

While the studies surveyed so far have combined both genetic and linguistic data, in this last section I will briefly touch upon work based solely on linguistic data. This fruitful area of research was initiated only in the new millennium, when innovative studies addressed the evolution of the Austronesian, Bantu, and Indo-European language families [45–47]. Since then, ever more studies are being conducted that explicitly model migrations and test alternative hypotheses, e.g. [21,48–51]. For instance, Lee & Hasegawa [51] use Bayesian analysis of lexical data from 19 Ainu varieties with geographic coordinates to investigate two competing models of Ainu prehistory, namely whether the Ainu are the direct descendants of the prehistoric Jomon people who would have migrated to Japan from Southeast Asia approximately 10,000 years ago, or whether they represent the result of admixture between the Jomon people and the so-called Okhotsk people who would have migrated into northern Japan from the Amur river basin 900-1600 years ago. Their results provide strong support for the admixture model, both from a temporal perspective (with the tree of Ainu varieties rooting 1288 years before present) as well as from a geographical perspective, with strong support for a diffusion of Ainu linguistic diversity from the north. Thus, the linguistic data help shed light on the origins of the Ainu and complement archaeological and genetic studies.

While studies of language family phylogenies are mostly based on lexical data, investigations of deeper relationships have been attempted with structural data [20–22], as mentioned in the section on language-gene coevolution. For example, Sicoli & Holton [21] use Bayesian analysis of structural data from 37 Na-Dene languages of North America and two Yeniseian languages from Siberia to distinguish between two hypotheses concerning the migration of the putative shared ancestor [52,53] of these languages, namely 1) a migration from Eurasia to the New World, or 2) a bidirectional migration out of Beringia. Their results consistently contradict an early split-off of the Yeniseian languages, as implicit in the Eurasian hypothesis [52], and provide strong support for the Beringian origin of these languages. This is in good accordance with previous mtDNA analyses which also showed a bidirectional migration out of Beringia to Siberia and the New World [54].

Conclusions

As demonstrated by the diverse studies reviewed here, the history of languages can sometimes, but often does not, parallel the genetic history of their speakers. Thus, adding insights obtained from linguistic data to molecular anthropological investigations will help shed further light on the complex processes at play in the demographic history of modern humans. It is to be expected that further advances will be made in all three approaches

outlined above, although finding a suitable measure of linguistic distances comparable to the genetic measures remains a challenge.

Acknowledgments

I thank Mark Stoneking for helpful discussion. Much of the research presented in the section on "Prehistoric population contact and its effect on language evolution and change" was undertaken in the framework of the Max Planck Research Group on Comparative Population Linguistics (2007-2011), which was generously funded by the Max Planck Society.

References

1. Charles Darwin. *On the Origin of Species. The illustrated edition edited by D. Quammen.* Sterling; 2008 [1859].
2. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J: **Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data.** *Proc. Natl. Acad. Sci. U. S. A.* 1988, **85**:6002–6006.
3. Sokal RR: **Genetic, geographic, and linguistic distances in Europe.** *Proc. Natl. Acad. Sci. U. S. A.* 1988, **85**:1722–1726.
4. Belle EMS, Barbujani G: **Worldwide analysis of multiple microsatellites: Language diversity has a detectable influence on DNA diversity.** *Am. J. Phys. Anthropol.* 2007, **133**:1137–1146.
5. Pakendorf B: **Historical linguistics and molecular anthropology.** In *The Routledge Handbook of Historical Linguistics.* Edited by Bowerman C, Evans B. Routledge; 2014:627–641.
6. Pagel M: **Human language as a culturally transmitted replicator.** *Nat. Rev. Genet.* 2009, **10**:405–415.
7. Dediu D, Levinson SC: **On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences.** *Front. Psychol.* 2013, **4**:397.
8. Hauser MD, Yang C, Berwick RC, Tattersall I, Ryan M, Watumull J, Chomsky N, Lewontin R: **The mystery of language evolution.** *Front. Psychol.* 2014, [no volume].
9. Barbujani G: **Walking with Robert Sokal.** *Hum. Biol.* 2012, **84**:481–488.
10. Barbujani G, Sokal RR: **Zones of sharp genetic change in Europe are also linguistic boundaries.** *Proc. Natl. Acad. Sci. U. S. A.* 1990, **87**:1816–1819.
11. Cavalli-Sforza LL, Minch E, Mountain JL: **Coevolution of genes and languages revisited.** *Proc. Natl. Acad. Sci. U. S. A.* 1992, **89**:5620–5624.

- *12. Hunley K, Dunn M, Lindstrom E, Reesink G, Terrill A, Healy ME, Koki G, Friedlaender FR, Friedlaender JS: **Genetic and linguistic coevolution in Northern Island Melanesia.** *PLoS Genet* 2008, **4**:e1000239.

One of the first studies to use both genetic and empirical linguistic data from Island Melanesia to investigate the processes underlying the coevolution of languages and genes.

13. Balanovsky O, Dibirova K, Dybo A, Mudrak O, Frolova S, Pocheshkhova E, Haber M, Platt D, Schurr T, Haak W, et al.: **Parallel evolution of genes and languages in the Caucasus region.** *Mol. Biol. Evol.* 2011, **28**:2905–2920.

- **14. De Filippo C, Bostoen K, Stoneking M, Pakendorf B: **Bringing together linguistic and genetic evidence to test the Bantu expansion.** *Proc. Biol. Sci.* 2012, **279**:3256–3263.

An innovative study that used genetic, linguistic, and model-based geographic distances to investigate competing hypotheses about the route of expansion taken by Bantu agriculturalists in sub-Saharan Africa.

15. Ruhlen M: *A Guide to the World's Languages. Volume 1: Classification.* Stanford University Press; 1991.

16. Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L: **Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics.** *Am J Hum Genet* 1997, **61**:1015–35.

- *17. Lansing JS, Cox MP, Downey SS, Gabler BM, Hallmark B, Karafet TM, Norquest P, Schoenfelder JW, Sudoyo H, Watkins JC, et al.: **Coevolution of languages and genes on the island of Sumba, eastern Indonesia.** *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:16022–16026.

One of the first studies to use both genetic and linguistic empirical data at a small geographical scale to elucidate the process of settlement and the evolution of Austronesian languages on Sumba island.

18. Hock HH, Joseph BD: *Language history, language change and language relationship. An introduction to historical and comparative linguistics.* de Gruyter; 2009.

19. Dediu D, Levinson SC: **Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages.** *PloS One* 2012, **7**:e45198.

- **20. Reesink G, Singer R, Dunn M: **Explaining the linguistic diversity of Sahul using population models.** *PLoS Biol.* 2009, **7**:e1000241.

A very thorough study that applied phylogenetic methods to structural linguistic data from 121 languages to investigate potential deep linguistic relationships in the Pacific.

- *21. Sicoli MA, Holton G: **Linguistic phylogenies support back-migration from Beringia to Asia.** *PloS One* 2014, **9**:e91722.

This study used structural linguistic data to investigate the potential route of migration taken by the ancestors of populations speaking Yeniseic and Na-Dene languages (now settled in central Siberia and northwest America, respectively).

- *22. Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC: **Structural phylogenetics and the reconstruction of ancient language history**. *Science* 2005, **309**:2072–2075.

The first study to use structural linguistic data to elucidate the potential relationships among the Papuan languages of Island Melanesia.

23. Colonna V, Boattini A, Guardiano C, Dall'ara I, Pettener D, Longobardi G, Barbujani G: **Long-range comparison between genes and languages based on syntactic distances**. *Hum. Hered.* 2010, **70**:245–254.
24. Longobardi G, Guardiano C: **Evidence for syntax as a signal of historical relatedness**. *Lingua* 2009, **119**:1679–1706.
25. Longobardi G, Guardiano C, Silvestri G, Boattini A, Ceolin A: **Toward a syntactic phylogeny of modern Indo-European languages**. *J. Hist. Linguist.* 2013, **3**:122–152.
26. Haspelmath M: **Parametric versus functional explanations of syntactic universals**. In *The limits of syntactic variation*. Edited by Biberauer T. Benjamins; 2008:75–107.
- *27. Hunley K, Bower C, Healy M: **Rejection of a serial founder effects model of genetic and linguistic coevolution**. *Proc. Biol. Sci.* 2012, **279**:2281–2288.

A study that examined the hypothesis of a serial founder effect model of language expansion with empirical linguistic data and compared it to genetic data.

28. Atkinson QD: **Phonemic diversity supports a serial founder effect model of language expansion from Africa**. *Science* 2011, **332**:346–349.
29. Haspelmath M, Dryer M, Gil D, Comrie B: *The World Atlas of Language Structures*. Oxford University Press; 2005.
- *30. Barbieri C, Butthof A, Bostoen K, Pakendorf B: **Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia**. *Eur. J. Hum. Genet. EJHG* 2013, **21**:430–436.

This study used molecular anthropological data to elucidate the possible sociocultural processes at play in the borrowing of click consonants into a Bantu language.

31. Bostoen K, Sands B: **Clicks in south-western Bantu languages: contact-induced vs. language-internal lexical change**. In *Proceedings of the 6th World Congress of African Linguistics Cologne 2009*. Edited by Brenzinger M. Rüdiger Köppe Verlag; 2012:129–140.
- *32. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B: **Contrasting maternal and paternal histories in the linguistic context of Burkina Faso**. *Mol. Biol. Evol.* 2012, **29**:1213–1223.

This molecular anthropological study of populations speaking very different languages in central Burkina Faso demonstrated that social practices such as patrilocality play a role in determining to what extent linguistic affiliation will structure genetic diversity.

33. Beyer K, Schreiber H: **Intermingling speech groups: Morpho-syntactic outcomes of language contact in a linguistic area in Burkina Faso, West Africa.** In *The Interplay of Variation and Change in Contact Settings*. Edited by Léglise I, Chamoreau C. John Benjamins; 2013:107–134.

*34. Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, Kure B, Mpoloka SW, Nakagawa H, Naumann C, et al.: **The genetic prehistory of southern Africa.** *Nat. Commun.* 2012, **3**:1143.

This detailed study of autosomal genetic diversity in 'Khoisan' populations demonstrated that languages and genes do not necessarily correlate.

35. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MGB, et al.: **Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History.** *Science* 2012, doi:10.1126/science.1227721.

36. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B: **Unraveling the complex maternal history of Southern African Khoisan populations.** *Am. J. Phys. Anthropol.* 2014, **153**:435–448.

37. Güldemann T: **'Khoisan' linguistic classification today.** In *Beyond "Khoisan". Historical relations in the Kalahari Basin*. Edited by Güldemann T, Fehn A-M. John Benjamins; to appear 2014:1–40.

38. Traill A, Nakagawa H: **A historical!Xóo-|Gui contact zone: linguistic and other relations.** In *The State of Khoesan Languages in Botswana*. Edited by Batibo HM, Tsonope J. Tasalls Publishing and Books; 2000:1–17.

39. Güldemann T, Loughnane R: **Are There "Khoisan" Roots in Body-Part Vocabulary? On Linguistic Inheritance and Contact in the Kalahari Basin.** *Lang. Dyn. Change* 2012, **2**:215–258.

40. Gerlach L, Berthold F: **N!aqraxe (ǀ'Amkoe) spatial terms from a genealogical and areal perspective.** In *Beyond "Khoisan". Historical relations in the Kalahari Basin*. Edited by Güldemann T, Fehn A-M. John Benjamins; to appear 2014:209–232.

41. Wilder JA, Cox MP, Paquette AM, Alford R, Satyagraha AW, Harahap A, Sudoyo H: **Genetic continuity across a deeply divergent linguistic contact zone in North Maluku, Indonesia.** *BMC Genet.* 2011, **12**:100.

42. Delfin F, Myles S, Choi Y, Hughes D, Illek R, van Oven M, Pakendorf B, Kayser M, Stoneking M: **Bridging near and remote Oceania: mtDNA and NRY variation in the Solomon Islands.** *Mol. Biol. Evol.* 2012, **29**:545–564.

43. Van Oven M, Brauer S, Choi Y, Ensing J, Schiefenhövel W, Stoneking M, Kayser M: **Human genetics of the Kula Ring: Y-chromosome and mitochondrial DNA**

variation in the Massim of Papua New Guinea. *Eur. J. Hum. Genet. EJHG* 2014, doi:10.1038/ejhg.2014.38.

- *44. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, et al.: **Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool.** *Am. J. Hum. Genet.* 2012, **91**:83–96.

A detailed study of genetic diversity of Ethiopian populations that demonstrated an impact of linguistic affiliation on genetic structure.

45. Gray RD, Jordan FM: **Language trees support the express-train sequence of Austronesian expansion.** *Nature* 2000, **405**:1052–1055.
46. Holden CJ: **Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis.** *Proc. Biol. Sci.* 2002, **269**:793–799.
47. Gray RD, Atkinson QD: **Language-tree divergence times support the Anatolian theory of Indo-European origin.** *Nature* 2003, **426**:435–439.
48. Kitchen A, Ehret C, Assefa S, Mulligan CJ: **Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East.** *Proc. Biol. Sci.* 2009, **276**:2703–2710.
- **49. Gray RD, Drummond AJ, Greenhill SJ: **Language phylogenies reveal expansion pulses and pauses in Pacific settlement.** *Science* 2009, **323**:479–483.

A ground-breaking study that elucidated the process of expansion of Austronesian speakers over the Pacific using linguistic data.

50. Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, Gray RD, Suchard MA, Atkinson QD: **Mapping the origins and expansion of the Indo-European language family.** *Science* 2012, **337**:957–960.
- *51. Lee S, Hasegawa T: **Evolution of the Ainu language in space and time.** *PloS One* 2013, **8**:e62243.

An interesting study that used linguistic data to investigate two competing hypotheses about the origins of the Ainu.

52. Ruhlen M: **The origin of the Na-Dene.** *Proc. Natl. Acad. Sci.* 1998, **95**:13994–13996.
53. Vajda EJ: **A Siberian link with Na-Dene languages.** In *The Dene-Yeniseian Connection: Bridging Asia and North America*. Edited by Kari J, Potter BA. Alaska Native Language Center; 2010:33–99.
54. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, et al.: **Beringian standstill and spread of Native American founders.** *PloS One* 2007, **2**:e829.