

Bayesian phylogeography of the Arawak expansion in lowland South America

Robert S. Walker and Lincoln A. Ribeiro

Proc. R. Soc. B 2011 **278**, 2562-2567 first published online 19 January 2011
doi: 10.1098/rspb.2010.2579

Supplementary data

["Data Supplement"](#)

<http://rsjb.royalsocietypublishing.org/content/suppl/2011/01/12/rspb.2010.2579.DC1.html>

References

[This article cites 26 articles, 10 of which can be accessed free](#)

<http://rsjb.royalsocietypublishing.org/content/278/1718/2562.full.html#ref-list-1>

Subject collections

Articles on similar topics can be found in the following collections

[taxonomy and systematics](#) (360 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Proc. R. Soc. B* go to: <http://rsjb.royalsocietypublishing.org/subscriptions>

Bayesian phylogeography of the Arawak expansion in lowland South America

Robert S. Walker^{1,*} and Lincoln A. Ribeiro²

¹Department of Anthropology, University of Missouri, MO, USA

²Universidade Estadual de Goiás-GICLI, Anápolis, Brazil

Phylogenetic inference based on language is a vital tool for tracing the dynamics of human population expansions. The timescale of agriculture-based expansions around the world provides an informative amount of linguistic change ideal for reconstructing phylogeographies. Here we investigate the expansion of Arawak, one of the most widely dispersed language families in the Americas, scattered from the Antilles to Argentina. It has been suggested that Northwest Amazonia is the Arawak homeland based on the large number of diverse languages in the region. We generate language trees by coding cognates of basic vocabulary words for 60 Arawak languages and dialects to estimate the phylogenetic relationships among Arawak societies, while simultaneously implementing a relaxed random walk model to infer phylogeographic history. Estimates of the Arawak homeland exclude Northwest Amazonia and are bi-modal, with one potential homeland on the Atlantic seaboard and another more likely origin in Western Amazonia. Bayesian phylogeography better supports a Western Amazonian origin, and consequent dispersal to the Caribbean and across the lowlands. Importantly, the Arawak expansion carried with it not only language but also a number of cultural traits that contrast Arawak societies with other lowland cultures.

Keywords: Amazonian languages and cultures; population expansions; phylogenetic trees; ancient histories

1. INTRODUCTION

The dynamics of prehistoric agricultural expansions of human populations around the world are interpretable in light of contemporary ethnolinguistic and geographic distributions [1]. Homelands are often inferred using the heuristic approach of locating a linguistic outgroup or delimiting the geographical region of highest linguistic diversity [2,3], but a model-free approach gives limited insight into the actual spatial dynamics of expanding populations. In contrast, a Bayesian framework for inference of phylogeography offers the opportunity to fully reconstruct likely ancestral histories while accounting for phylogenetic uncertainty [4], and can also provide fundamental understanding of the evolutionary dynamics of human culture [5].

Arawak is a geographically dispersed language family scattered across lowland South America from Argentina to the Bahamas and from the mouth of the Amazon River to the foothills of the Andes (see map in electronic supplementary material, figure S1). Arawak societies are encountered in a diverse array of ecologies, including tropical forests, Andean foothills, Caribbean coasts and islands, dry forests of central Brazil, and the savannahs of Colombia/Venezuela and Bolivia. Arawak forms an outgroup to the other major lowland language families Je, Carib and Tupi, according to genetic [6], linguistic [7] (but see [8]) and cultural [9,10] evidence.

Phylogenetic trees based on language have proven to be an important analytical tool for reconstructing human population and cultural histories (e.g. Austronesia [11],

Bantu [12], Indo-European [13], Semitic [14]). Although contested (e.g. [15–17]), these studies fruitfully use a Bayesian statistical approach on the systematic codings of linguistic cognates [18]. Here we extend this method to lowland South America by modelling the continuous spatial dispersal of the Arawak expansion with a relaxed random walk (RRW) model. The RRW model was recently developed for the spread of viral outbreaks and has the advantage of accommodating branch-specific variation in dispersal rates across time and space [19]. Our goal is to test hypotheses concerning the homeland of the Arawak expansion and pursue a deeper understanding of Amazonian prehistory and contemporary ethnolinguistic variation.

2. METHODS

We compiled Swadesh [20] lists of 100 common vocabulary items and scored cognate sets across 60 Arawak languages and dialects representing all the major branches of the Arawak language family (see electronic supplementary material, table S1). It was not deemed necessary to exclude dialects given that phylogenetic analyses simply clump dialects together. An advantage of the Swadesh list is that it contains lexical terms (e.g. simple nouns, adjectives, numbers and pronouns) that are relatively resistant to borrowing (although there are clearly exceptions, e.g. [21]) and demonstrate low rates of change [16,22]. We relied heavily on Payne's [23] cognate reconstructions of shared lexical retentions. The words in a cognate set are derived from a single common ancestral form that was present in an ancestral language. However, we acknowledge that it is not always easy to distinguish loanwords from true cognates, and our data probably include unidentified borrowed words that may affect our results.

* Author for correspondence (walkerro@missouri.edu).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2010.2579> or via <http://rsob.royalsocietypublishing.org>.

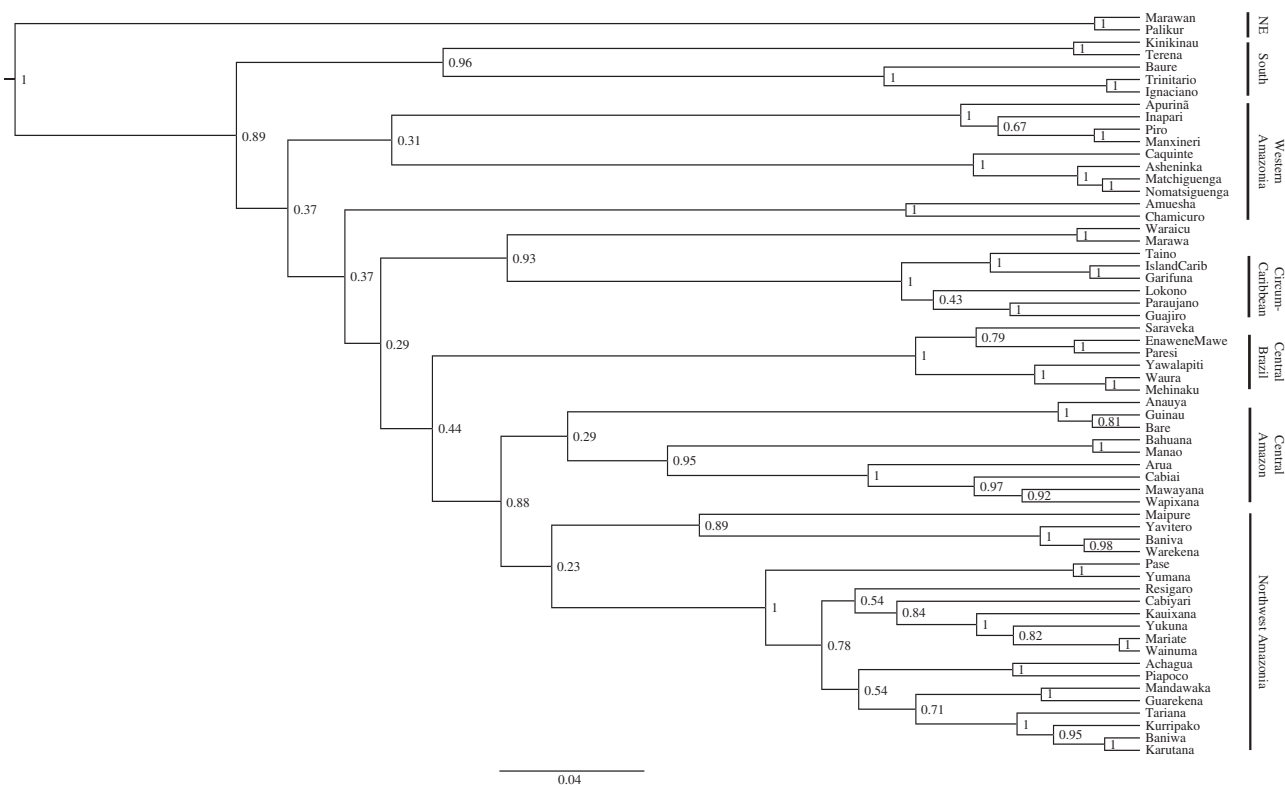


Figure 2. Maximum clade credibility tree from 1000 Bayesian Markov chain Monte Carlo trees. Nodes are labelled with posterior probabilities representing the proportion of trees that support the formation of a particular clade. Clade labels represent geographical regions. ‘NE’ represents Northeast.

inferring phylogeny is that trees are sampled in proportion to their likelihood and phylogenetic certainty is represented by the proportion of trees in which certain clades emerge (i.e. posterior probabilities or clade credibilities). The continuous RRW model generates a posterior probability distribution of potential homeland locations that are overlaid on a map of South America.

3. RESULTS

(a) *Arawak phylogeny*

The Arawak maximum clade credibility tree presented here (figure 2; for phylogram see electronic supplementary material, figure S2) is broadly consistent with expert classifications by linguists [8,23,26–31], at least near the tips. The main point of departure is that some linguists have suggested a deep North–South split in the Arawak language family [29] or other ancient clades [30]. Instead, our results support, with a posterior probability of 0.89, a deep divergence at the base separating the Marawan and Palikur which are located in the Northeast near the Atlantic seaboard north of the mouth of the Amazon River. The other deeper divergences in our phylogeny have considerable uncertainty in terms of their order of divergence (posterior probabilities range from 0.29 to 0.44). Clades are fairly well formed (i.e. posterior probabilities > 0.89) into the following geographical regions in the order of divergence: (i) Northeast, (ii) South (Bolivia and southern Brazil), (iii) three clades in Western Amazonia (Purus River basin and two clades in the Andean foothills), (iv) Circum-Caribbean, (v) Central Brazil, (vi) two clades in Central Amazon, and (vii) two clades in and around Northwest

Amazonia. These latter two geographic regions—Central Amazon and Northwest Amazonia—cluster together deep in the phylogeny with a posterior probability of 0.88.

Interestingly, the Marawa and Waraicu languages located near the Amazon River form a deep clade with Circum-Caribbean languages (posterior probability of 0.93). These languages do not show especially high conservatism (figure 1 and electronic supplementary material, figure S2). This suggests that this ancient clade may have originated around the main branch of the Amazon River, with a potential migration up the Rio Branco towards the Guyanas and later the Caribbean, and not a migration originating from Northwest Amazonia. In fact, Northwest Amazonia is the last clade to diverge according to our phylogeny, making it an unlikely candidate for the Arawak homeland, despite its numerosity.

(b) *Continuous phylogeographic dispersal*

Figure 3 plots the homeland estimates based on the RRW model overlaid onto a map of South America. While there is considerable variation across the 1000 samples, the Bayesian chains commonly visited two potential homelands in a bi-modal geographical distribution. One potential homeland is the Atlantic seaboard around the recent location of the Marawan and Palikur languages. These languages represent the Northeast clade, the earliest divergence in the phylogeny (figure 2). However, Bayesian chains more frequently sampled a large area of Western Amazonia, at about twice the frequency of the potential Northeast homeland (figure 3). The centre of this Western Amazonian scatter is approximately the present-day location of the Apurinã language in the Purus region. The geographical regions within the

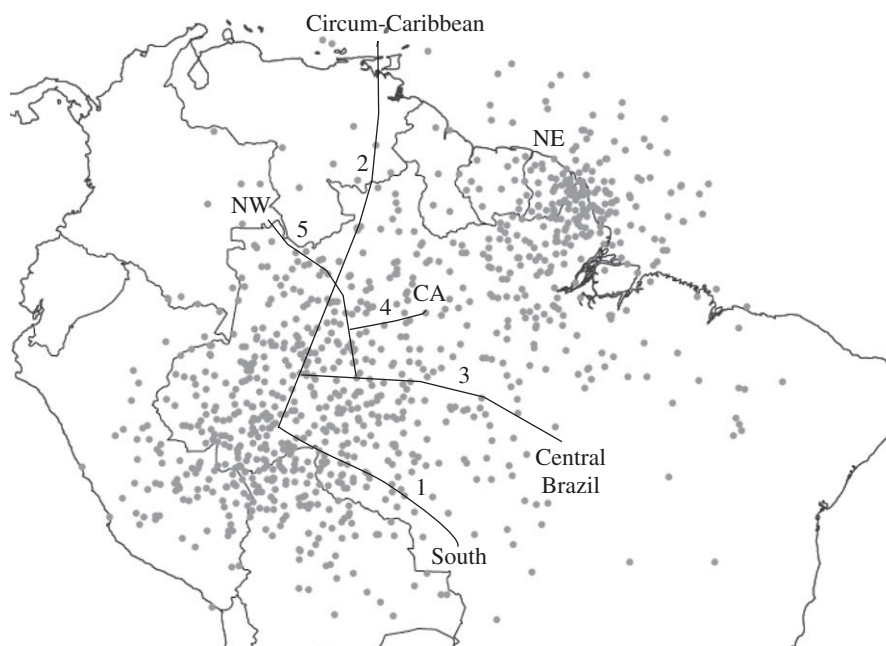


Figure 3. RRW model estimates of the geographical homeland of the Arawak expansion (grey circles) overlaid on a map of South America. Results are bi-modal with one potential homeland on the Atlantic seaboard and another more likely originate somewhere in Western Amazonia. Northwest Amazonia (NW, or the ‘dog’s head’ of northwest Brazil) is not well supported as a potential homeland. Lines show a likely diffusion scenario assuming a Western Amazonian homeland and the ordering of splits from the phylogenetic analysis (although these have low posterior probabilities; figure 2). An early migration may have been to the northeast (NE) down the Amazon River (line not shown). Subsequent diffusions are numbered in order. Central Amazon is denoted CA.

95 per cent high probability density (HPD) of the homeland include both Central Amazon and Western Amazonia clades. The HPD excludes Northwest Amazonia (i.e. the ‘dog’s head’ region of northwest Brazil) as a purported Arawak homeland. Other regions excluded from the HPD are South, Central Brazil and Circum-Caribbean, all unlikely Arawak homelands.

(c) *Discrete phylogeographic dispersal*

To better evaluate among potential homeland alternatives, we also implemented a phylogeographic model of dispersal as a discrete process among our seven geographical regions. We used a gamma-distributed reversible-jump hyperprior (RJHP) in BAYESTRAITS [18] to reconstruct a discrete dispersal process onto the phylogenies. With RJHP, Markov chains explore the model and parameter space to automatically discover whether fewer transition rates can adequately explain the data. The total number of possible (reversible) transition rates among the seven regions is 42. The most common transition rate (non-zero in 93% of all samples) is from Central Amazon to Northwest Amazonia, tentatively suggesting movement up the Rio Negro from the main branch of the Amazon River. However, all of the 42 transition rates commonly appear in our model at least 76 per cent of the time. In other words, we do not have the ability to correctly reconstruct a likely sequence of discrete migrations owing to considerable phylogenetic uncertainty in the middle of the Arawak phylogeny after the divergence of the Northeast clade.

The discrete RJHP model does produce a pattern of likely homelands. RJHP takes the root estimate of each sample to indicate a likely homeland, just as ancestral biological or cultural traits are reconstructed when mapped onto phylogenies. Western Amazonia emerged as the

most likely homeland with a mean posterior probability of 44 per cent. Both Northeast and South had probabilities of 16 per cent. Central Amazon received 12 per cent support, Caribbean and Central Brazil both 4 per cent, and Northwest Amazonia was the least-supported potential homeland with a posterior probability of only 3 per cent.

4. DISCUSSION

Phylogeographic reconstructions are essential for evaluating plausible scenarios for Arawak and other language family expansions. Our results generally point to a Western Amazonian homeland for the Arawak language family. While our phylogeny indicates that the deepest divergence in the Arawak tree is the Northeast clade, this region was less frequently sampled by the continuous phylogeographic dispersal model in comparison with the Western Amazonian homeland. We find the Western Amazonian homeland scenario to be more probable because it is supported more by both the continuous (RRW) and discrete (RJHP) phylogeographic dispersal models. The continuous RRW also includes Central Amazon as a potential homeland, but this region is not well supported by the discrete RJHP model. Importantly, despite the fact that Northwest Amazonia contains the largest number of Arawak languages, this region was not supported by the phylogeny nor either of the dispersal models.

While there is considerable uncertainty in both the location of the Arawak homeland and the exact orderings of early linguistic splits, we propose the following scenario as the most likely given our phylogeographic results. A Western Amazonian homeland implies that an early migration was probably down the Madeira or Purus

River to the main Amazon River and then down to the mouth near the present-day Palikur and Marawan in the northeast. An overland migration or diffusion to the South is likely to have occurred next, followed by a northern riverine migration that ended up in and around the Caribbean. Later migration or diffusion to Central Brazil may have also been terrestrial, followed by riverine colonization of Central Amazon and finally up the Rio Negro to Northwest Amazonia (figure 3).

A Western Amazonian homeland for Arawak reopens the question of manioc cultivation as a potential driver for population expansion [32]. Manioc is derived from a single wild South American progenitor found in transitional forests in the ecotone between southern lowland Amazonian rainforests, the Bolivian savannahs and the dry forests of Central Brazil [33,34]. This region includes the southern border of the homeland estimates for Western Amazonia around the Brazilian–Bolivian border (figure 3 and electronic supplementary material, figure S1). This general area harbours a number of other early horticultural domesticates [35], as well as many prehistoric earthworks including geoglyphs, canals, causeways and raised fields [36,37].

Perhaps the most significant aspect of the Arawak expansion is that it carried with it not only language but also a number of cultural traits that contrast many Arawak-speaking peoples with most other lowland societies. Ethnographers have identified a set of cultural practices considered together as being distinctively Arawak, albeit with considerable variation. These cultural traits include socio-political alliances with ethnolinguistically distant peoples; strong emphasis on religion, class and descent; and limited endo-warfare that contrasts with cycles of vendetta and within-group warfare common for many lowland Amazonians [9,10,38–40]. At one extreme were Arawak chiefdoms in the Circum-Caribbean, the Taíno and Lokono, with strict hierarchical orderings, divine ancestry and chiefly elite lineages. Social hierarchy was also strongly present in Northwest Amazonia, Central Amazon, Central Brazil and the South [9,10]. Arawak identity was apparently based in many places on control of fertile floodplains and trade routes along major rivers at the expense of their neighbours [9]. Post-marital residence in Arawak societies tends more towards virilocality (whereby females transfer) in contrast to the common lowland pattern of uxorilocality (whereby males transfer). Paternity beliefs in Arawak societies often involve only one father, or singular paternity (although there are at least two exceptions), whereas most other Amazonian societies have a common belief in partible paternity where multiple men are purported to be the co-genitors of one child [41]. The co-occurrence of virilocality and singular paternity may be related to the common pattern of patrilineal descent of Arawak leadership and social power.

Arawak cultural traits (in particular the elevated level of social hierarchy), although present to some extent in some non-Arawaks and absent in some Arawak societies, are found in all the major clades of the Arawak linguistic tree, with the possible exceptions of the Northeast and Western Amazonia (both Purus and Andean foothills, although the Amuesha were a hierarchical society), the two potential homelands in our analysis. This suggests that, while elevated social hierarchy may or may not have been present in proto-Arawak, it probably developed before most of the

major Arawak dispersals. A parsimonious explanation of this pattern may be one of vertical transmission stemming from a common ancestral Arawak demographic pump that spread ethnolinguistic traits across the lowlands in multiple migration events. Santos-Granero [39] underscores the importance of vertical transmission of Arawak culture but combines it with horizontal exchange that led to the emergence of some trans-ethnic groups. Zucchi [42] suggests that the Arawak expansion stemmed from a combination of vertical (migration as the result of demographic growth) and horizontal (aggregation through regional alliances and trade networks) transmission processes. Regardless, whether the success of the Arawak expansion resulted from a technological innovation (e.g. canoes), manioc cultivation, the trade network system itself or perhaps the hierarchical social system remains to be seen. Explicit phylogeographic models are ideally suited to help answer these questions as further research better integrates data from archaeology, plants, genes, languages and cultures.

Special thanks to Gláucia Vieira Cândido, Michael Cysouw, Søren Wichmann, David Payne, and Harold and Diana Green for much linguistic assistance. We also thank several anonymous reviewers, Michael Gurven, Marcus Hamilton and Briggs Buchanan for excellent suggestions to improve this paper. Financial support was provided by the Max Planck Society.

REFERENCES

- Bellwood, P. & Diamond, J. 2003 Farmers and their languages: the first expansions. *Science* **300**, 597–603. (doi:10.1126/science.1078208)
- Sapir, E. 1916 *Time perspective in aboriginal American culture, a study in method*. Geological Survey Memoir 90: No. 13, Anthropological Series. Ottawa, Canada: Government Printing Bureau.
- Wichmann, S., André, M. & Viveka, V. 2010 Homelands of the world's language families: a quantitative approach. *Diachronica* **27**, 247–276. (doi:10.1075/dia.27.2.05wic)
- Bloomquist, E. W., Lemey, P. & Suchard, M. A. 2010 Three roads diverged? Routes to phylogeographic inference. *Trends Ecol. Evol.* **25**, 626–632. (doi:10.1016/j.tree.2010.08.010)
- Mace, R., Holden, C. & Shennan, S. (eds) 2005 *The evolution of cultural diversity: a phylogenetic approach*. Walnut Creek, CA: Left Coast Press.
- Salzano, F., Hutz, M., Salamoni, S., Roh, r.P. & Callegari-Jacques, S. 2005 Genetic support for proposed patterns of relationship among lowland South American languages. *Curr. Anthropol.* **46**, S121–S129. (doi:10.1086/497667)
- Rodrigues, A. 1985 Evidence for Tupi-Carib relationships. In *South American Indian languages: retrospect and prospect* (eds H. Klein & L. Stark), pp. 54–94. Austin, TX: University of Texas Press.
- Loukotka, C. 1968 *Classification of South American Indian languages*. Los Angeles, CA: Latin American Center, University of California.
- Hornborg, A. 2005 Ethnogenesis, regional integration, and ecology in prehistoric Amazonia. *Curr. Anthropol.* **46**, 589–620. (doi:10.1086/431530)
- Hill, J. & Santos-Granero, F. (eds) 2002 *Comparative Arawakan histories: rethinking language family and culture area in Amazonia*. Champaign, IL: University of Illinois Press.
- Gray, R., Drummond, A. & Greenhill, S. 2009 Language phylogenies reveal expansion pulses and pauses in Pacific

- settlement. *Science* **323**, 479–483. (doi:10.1126/science.1166858)
- 12 Holden, C. & Mace, R. 2003 Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proc. R. Soc. Lond. B* **270**, 2425–2433. (doi:10.1098/rspb.2003.2535)
- 13 Gray, R. & Atkinson, Q. 2003 Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426**, 435–439. (doi:10.1038/nature02029)
- 14 Kitchen, A., Ehret, C., Assefa, S. & Mulligan, C. 2009 Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proc. R. Soc. B* **276**, 2703–2710. (doi:10.1098/rspb.2009.0408)
- 15 Donohue, M. & Denham, T. 2010 Farming and language in Island Southeast Asia: reframing Austronesian history. *Curr. Anthropol.* **51**, 223–256. (doi:10.1086/650991)
- 16 McMahon, A. & McMahon, R. 2005 *Language classification by numbers*. Oxford, UK: Oxford University Press.
- 17 Nakhleh, L., Warnow, T., Ringe, D. & Evans, S. N. 2005 A comparison of phylogenetic reconstruction methods on Indo-European dataset. *Trans. Phil. Soc.* **103**, 171–192. (doi:10.1111/j.1467-968X.2005.00149.x)
- 18 Pagel, M. & Meade, A. 2006 Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825. (doi:10.1086/503444)
- 19 Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. 2010 Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885. (doi:10.1093/molbev/msq067)
- 20 Swadesh, M. 1952 Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proc. Am. Phil. Soc.* **96**, 452–463.
- 21 Payne, D. L. 1985 The genetic classification of Resigaró. *Int. J. Am. Linguist.* **51**, 222–231. (doi:10.1086/465867)
- 22 Pagel, M., Atkinson, Q. & Meade, A. 2007 Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717–721. (doi:10.1038/nature06176)
- 23 Payne, D. 1991 A classification of Maipuran (Arawakan) languages based on shared lexical retentions. In *Handbook of Amazonian languages* (eds D. Derbyshire & D. G. Pullum), pp. 355–499. Berlin, Germany: Mouton de Gruyter.
- 24 Bryant, D. & Moulton, V. 2004 Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265. (doi:10.1093/molbev/msh018)
- 25 Drummond, A. J. & Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214. (doi:10.1186/1471-2148-7-214)
- 26 Mason, J. 1950 The languages of South American Indians. *Handb. South Am. Indians, Bur. Am. Ethnol. Bull.* **143**, 157–317.
- 27 Noble, K. 1965 *Proto-Arawakan and its descendants*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.
- 28 Matteson, E. 1964 Algunas afiliaciones de la familia Arawak. In *XXXV Congreso Internacional de Americanistas, México 1962 actas y memorias*, vol. 2, pp. 519–523. Mexico: Editorial Libros de México.
- 29 Aikhenvald, A. 1999 Arawak languages. In *Amazonian languages* (eds R. Dixon & A. Aikhenvald), pp. 487–517. Cambridge, UK: Cambridge University Press.
- 30 Ramirez, H. 2001 *Línguas Arawak da Amazônia setentrional: comparação e descrição*. Manaus, Brazil: Editora da Universidade do Amazonas.
- 31 Gordon, R. & Grimes, B. (eds) 2005 *Ethnologue: languages of the world*. Dallas, TX: SIL International.
- 32 Lathrap, D. W. 1970 *The upper Amazon: ancient peoples and place*. London, UK: Thames & Hudson.
- 33 Olsen, K. M. 2002 Population history of Manihot esculenta (Euphorbiaceae) inferred from nuclear DNA sequences. *Mol. Ecol.* **11**, 901–911. (doi:10.1046/j.1365-294X.2002.01493.x)
- 34 Olsen, K. M. & Schaal, B. A. 2001 Microsatellite variation in cassava and its wild relatives: further evidence for a southern Amazonian origin of domestication. *Am. J. Bot.* **88**, 131–142. (doi:10.2307/2657133)
- 35 Pearsall, D. M. 2008 Plant domestication. In *Encyclopedia of archaeology* (ed. D. M. Pearsall). New York, NY: Academic Press.
- 36 Heckenberger, M. & Neves, E. G. 2009 Amazonian archaeology. *Annu. Rev. Anthropol.* **38**, 251–266. (doi:10.1146/annurev-anthro-091908-164310)
- 37 Mann, C. 2008 Ancient earthmovers of the Amazon. *Science* **321**, 1148–1152. (doi:10.1126/science.321.5893.1148)
- 38 Heckenberger, M. 2002 Rethinking the Arawakan diaspora: hierarchy, regionality, and the Amazonian formative. In *Comparative Arawakan histories: rethinking language family and culture area in Amazonia* (eds J. D. Hill & F. Santos-Granero), pp. 99–122. Urbana, IL: University of Illinois Press.
- 39 Santos-Granero, F. 2002 The Arawakan matrix: ethos, language, and history in native South America. In *Comparative Arawakan histories: rethinking language family and culture area in Amazonia* (eds J. D. Hill & F. Santos-Granero), pp. 25–50. Urbana, IL: University of Illinois Press.
- 40 Vidal, S. M. 2000 Kuwé Duwákalmi: the Arawak sacred routes of migration, trade, and resistance. *Ethnohistory* **47**, 635–667. (doi:10.1215/00141801-47-3-4-635)
- 41 Walker, R. S., Flinn, M. & Hill, K. 2010 The evolutionary history of partible paternity in lowland South America. *Proc. Natl Acad. Sci. USA* **107**, 19 195–19 200. (doi:10.1073/pnas.1002598107)
- 42 Zucchi, A. 2002 The Arawakan matrix: ethos, language, and history in native South America. In *Comparative Arawakan histories: rethinking language family and culture area in Amazonia* (eds J. D. Hill & F. Santos-Granero), pp. 199–222. Urbana, IL: University of Illinois Press.