# The riddle of Tasmanian languages

## Claire Bowern*

*Department of Linguistics, Yale University, 370 Temple Street, New Haven, CT 06511, USA*

Recent work which combines methods from linguistics and evolutionary biology has been fruitful in discovering the history of major language families because of similarities in evolutionary processes. Such work opens up new possibilities for language research on previously unsolvable problems, especially in areas where information from other sources may be lacking. I use phylogenetic methods to investigate Tasmanian languages. Existing materials are so fragmentary that scholars have been unable to discover how many languages are represented in the sources. Using a clustering algorithm which identifies admixture, source materials representing more than one language are identified. Using the Neighbor-Net algorithm, 12 languages are identified in five clusters. Bayesian phylogenetic methods reveal that the families are not demonstrably related; an important result, given the importance of Tasmanian Aborigines for information about how societies have responded to population collapse in prehistory. This work provides insight into the societies of prehistoric Tasmania and illustrates a new utility of phylogenetics in reconstructing linguistic history.

**Keywords:** linguistics; evolutionary anthropology; Tasmania

## 1. INTRODUCTION

The indigenous people of Tasmania were severely affected by European settlement in the nineteenth century [1]. Although it is known from ethnographic sources and early reports [2] that Indigenous Tasmanians comprised 48 bands in nine tribes [3,4] (figure 1), the number of languages and their internal phylogenetic relationships have remained a mystery. Previous work [5–9] has identified anywhere from a single language [5] to as many as 12 [6]. Despite the dearth of information about them, Tasmanian Aborigines have long held an important place in anthropology [10–12]. Their toolkit, for example, was the simplest of any attested group in the nineteenth century, and they are often cited as an example of how population collapse may also lead to technological collapse and societal decomplexification [10,11] (the so-called 'Tasmanian effect').

Information from language has thus far been underused in studying Tasmanian society; nonetheless, it provides an important window on Tasmanian internal diversity. The linguistic information may, indeed, be the only investigable source for Tasmanian heterogeneity at the level of the whole island. The anthropological, archaeological and genetic data are all insufficient here. Ethnographically, Ryan [4] describes Tasmanian tribes as a single culture bloc with extensive shared practices and beliefs (such as 'star gods' and the evil spirit *Wrageowrapper*), and a common toolkit. Jones [13] provides evidence for a strong cultural boundary between eastern and western Tasmania, but also notes many shared practices across the island. While recognizing nine distinct tribes, both Ryan & Jones [3] focus on the documentation of exchange networks and seasonal travel which reinforce reciprocal links across the island. Other work assumes a monolithic view of Tasmania without

discussion [14]. The archaeological record is patchy, with few Pleistocene sites [15]; moreover, Tasmanians did not have a rich material culture and the Tasmanian climate is not conducive to long-term preservation of wooden artefacts. There is, however, some evidence of internal diversity in the archaeological record (for example, the abandonment of rainforest sites after the Late Pleistocene [16] and the expansion of people down the western coast over the last 3000 years [13]). Within genetics, there is not sufficient genetic information to be able to determine any differences between Tasmanian populations, and subsequent history has led to sufficient European admixture that such work is not possible. There is, however, work which compares genetic data from Tasmanians with other populations [17,18], including those from Australia, such as Presser *et al.* [19], who find evidence of mitochondrial DNA links between Tasmania and the mainland.

It is known that the Tasmanian population underwent a population crash following the flooding of Bass Strait at the end of the Last Glacial Maximum [13], approximately 12 000 years ago. The population remained well below carrying capacity and was only recovering at the time of European colonization. Presumably, the Early Holocene population collapse led to a reduction in linguistic diversity on the island. It is not known, however, whether rates of language diversification were rapid enough to obliterate any evidence of a bottleneck, or whether the current languages and families show a common ancestor which predates the flooding of Bass Strait. If the data show that Tasmanian languages most probably belong to a single family, this would provide good evidence for slow rates of change in small societies, since there is no evidence for population replacement during the Holocene. However, it is also possible that any evidence for a linguistic bottleneck would be obliterated by subsequent linguistic diversification. This would have implications for our interpretation of the closeness of linkages between Tasmanian groups, since populations require isolation for linguistic diversification. I return to these points below.

*\*claire.bowern@yale.edu

Figure 1. The island of Tasmania, showing tribal boundaries inferred from ethnographic sources.

Given the paucity of island-wide research into genetic, archaeological and ethnographic diversity, language may provide us with the best opportunity of inferring change in prehistory. However, records of Tasmanian languages are poor [5,6]. The 44 known wordlists were recorded between 1777 and 1847. Vocabularies were recorded opportunistically, often with very little information about speakers or locations of recording. They vary in length from a single word to nearly 1040 items and originate from all over the island. In five cases, there is no information about provenance. Backhouse and Walker, for example, recorded vocabularies on Flinders Island from displaced persons of unknown tribal affiliation [5]. Other sources combined, or 'admixed', vocabulary from multiple locations, as evidenced both by the number of synonyms given in the lists and from comments from compilers. Other lists contain only general or ambiguous location information.

Previous attempts [1,6,9] to discover the linguistic history of Tasmania are rife with equivocations and are internally irreconcilable, despite being based on identical source material. Roth [1] was convinced that there was a single language, despite quoting considerable ethnographic evidence to the contrary. Walker [20] follows Robinson in arguing for four languages, but does not provide any evidence for this conclusion. Schmidt [9] found two languages, one with three dialects; O'Grady [7] also found 'at least two languages' (but not the same two as Schmidt), while Crowley & Dixon [6] argued that source materials are too poor to determine the number of languages, but are detailed enough to tentatively reject a single family. Most authors have not provided objective measures of degrees of cognacy of materials, relying instead on inspection of isolated forms to gauge the extent to which the varieties represent related languages or dialects.

Here I present a new analysis of the extant Tasmanian sources. The aims are to identify the number and composition of discrete linguistic units (languages) in the data, to determine whether they are demonstrably related, and if so, in how many families, to classify the wordlists without information about provenance, and to determine levels of admixture within wordlists. I use several methods

from evolutionary biology to systematically investigate the Tasmanian corpus. Such concepts and approaches have proved useful in shedding light on linguistic prehistory [21–23]. Although linguistic traits evolve more rapidly than genetic data, they exhibit many of the same properties and problems [24,25]. Some of these methods have been used on historical records already [23,26], though this work is the first to investigate Tasmanian data in this way, and the first to use methods in population biology to test for admixture in old sources.

## 2. RESULTS AND DISCUSSION

Previous work on Tasmanian languages did not estimate the degree of source mixture within vocabularies. This was achieved with STRUCTURE [27], a Bayesian clustering algorithm designed to identify admixture. Figure 2 shows the results of the structure analysis for $K = 2-5$ hypothesized groups, with $K = 5$ the highest likelihood value. Mean-likelihood values and further discussion are given in the electronic supplementary materials. STRUCTURE provides both information of admixture levels and tentative assignments of wordlists of unknown origin to clusters. These include the Norman vocabulary, which appears to group with the northeast languages, and the Lhotsky and Backhouse vocabularies, which are a separate language within the northeastern cluster. The Fisher vocabulary appears to be western, and the vocabularies designated by Plomley as 'southern' and 'northern' belong predominantly to the western group, with southeastern and northern mixture, respectively. Furthermore, several vocabularies identified as belonging to the Oyster Bay region show substantial mixing with the northeastern region. This mixture is probably what has led previous authors to posit relatedness between the languages of the east coast. Since this pattern is confined to only two sources, it is likely to represent source mixing rather than a genuine fact about the languages.

To investigate language relations further, 26 vocabularies of more than 100 items which did not show evidence of admixture were coded for similarity, and a Neighbor-Net [28] was produced using SplitsTree [29] (figure 3). This network shows five primary clusters: the same groups recovered in the STRUCTURE analysis. Furthermore, these clusters nest well with the tribal groups identified by Jones [3] and shown in figure 1. The southeastern/Bruny Island cluster corresponds to the Bruny tribe. The Oyster Bay cluster contains the Oyster Bay and Big River tribes. The northeastern cluster comprises vocabularies associated with the northeastern, Ben Lomond and north midlands tribes. The western cluster includes the northwestern and southwestern tribes, and the northern cluster includes the northern tribe and vocabularies designated by Milligan [30,31] as 'western' and 'northwestern'; these vocabularies were recorded on the Flinders Island mission. Electronic supplementary materials provide further discussion. There is no evidence here to group the southeastern, Oyster Bay and northeastern languages into a single macro-group, contra Schmidt [9] and O'Grady [3,7].

Linguists consider the boundary between languages and dialects often difficult to define in absolute terms. In this respect, languages are similar to biological species [32]. There are no single arbitrary cut-off points for
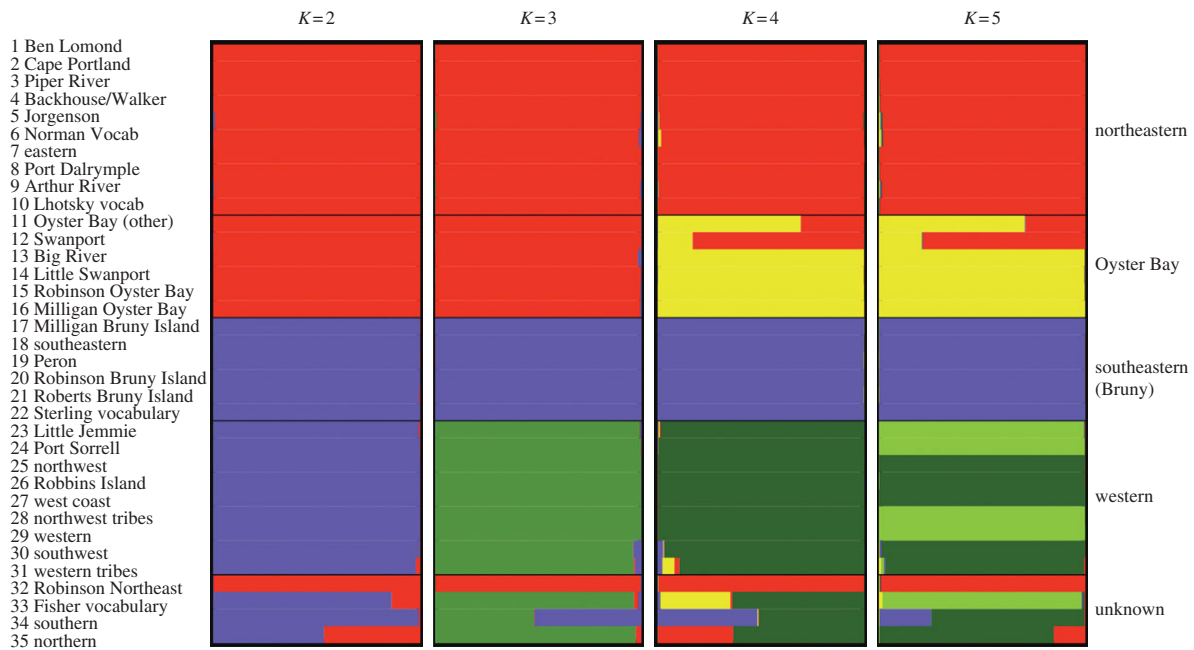
Figure 2. STRUCTURE analysis of wordlist data (35 individual vocabularies, five populations and 3339 characters), grouped by location of wordlist origin.
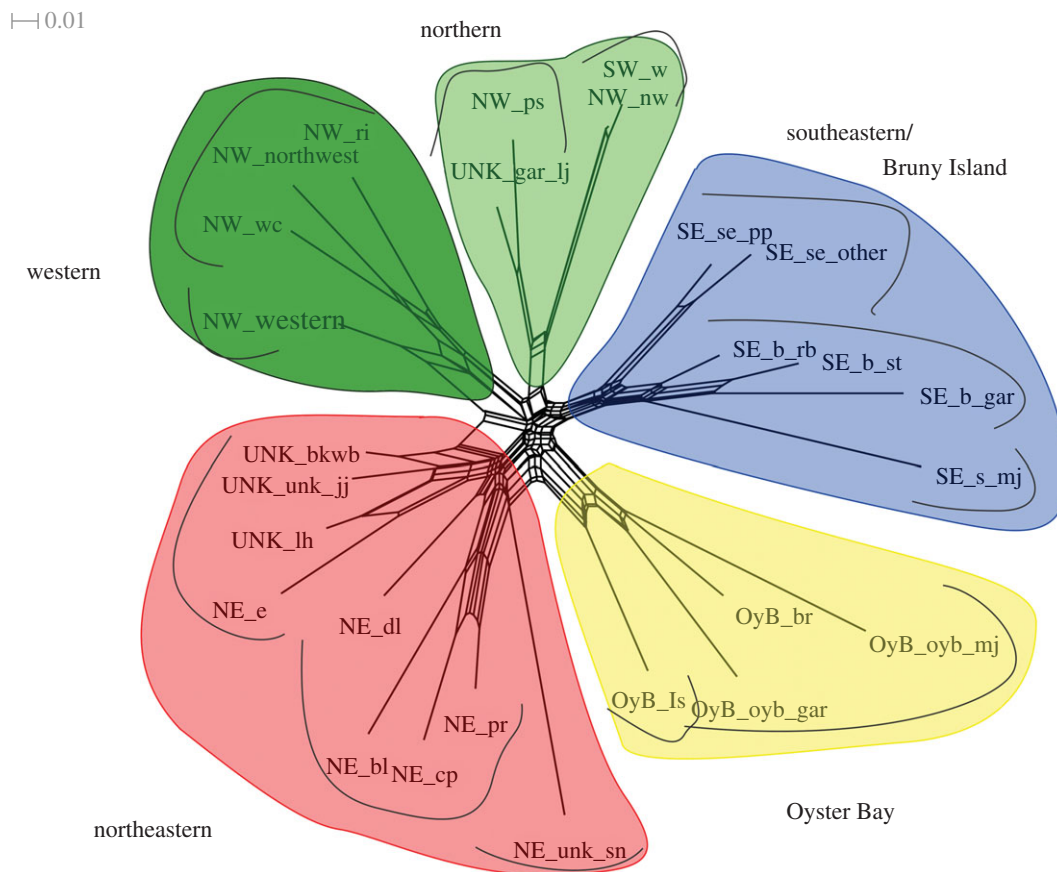


Figure 3. Neighbor-Net of wordlists with high levels missing data and admixture removed, showing five language clusters.

measurements of similarity between languages and dialects and measures based on the percentage of shared vocabulary are unreliable [33]. In the absence of speaker intuitions, however, we are reliant on the use of arbitrary measures of similarity. Here I group together as single 'language' vocabularies which have an uncorrected $p$ distance of less than 0.15. These languages are marked

on figure 3 within each cluster. They accord well with geographical placement, where known. This measure suggests that 12 languages are represented in the source materials. Two of the languages are represented exclusively by vocabularies recorded on the Flinders Island mission. These are the Milligan vocabularies (SW_W and NW_nw) in the northern cluster and the Milligan Bruny Island
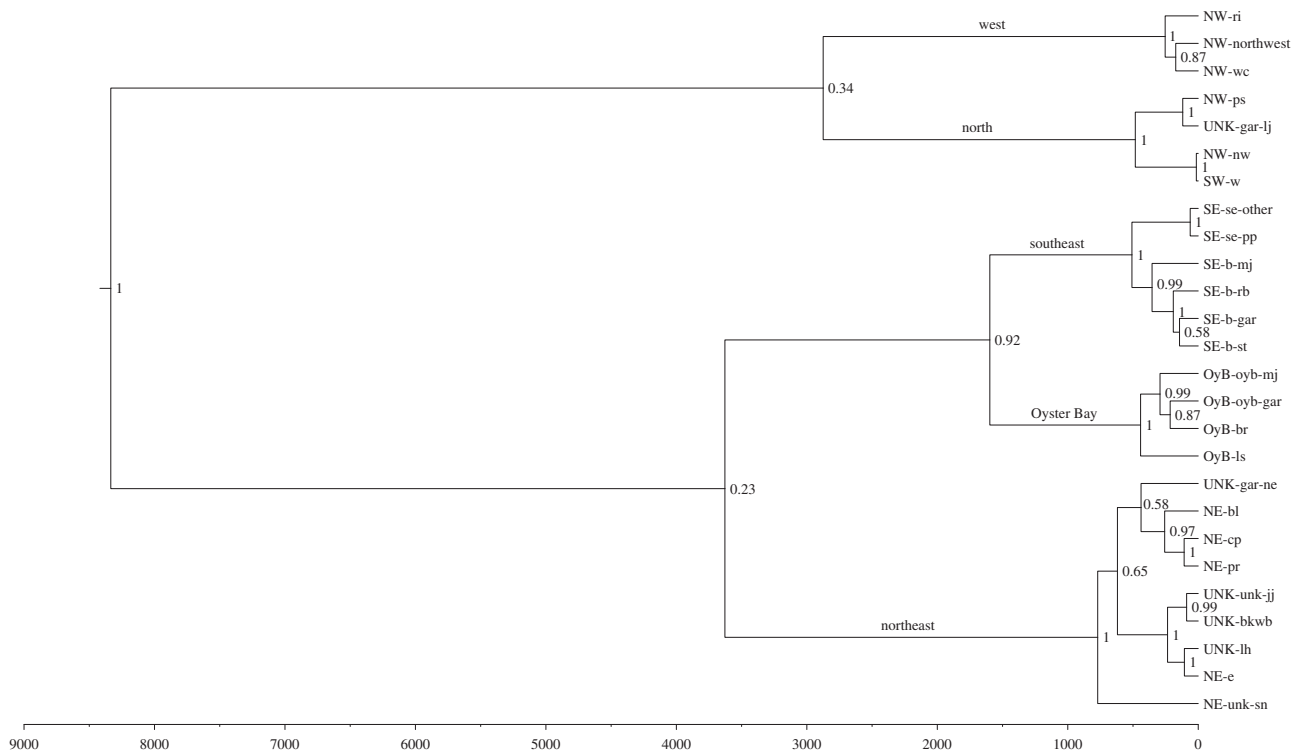
Figure 4. Fifty per cent consensus tree based showing posterior probabilities of internal nodes, based on a Bayesian maximum-likelihood analysis of 2777 data points from 26 non-admixed vocabularies.

vocabulary (SE_s_mj). The Milligan Oyster Bay vocabulary clusters with other Oyster Bay sources and the other late Flinders Island vocabulary (UNK_bkwb), recorded by Backhouse and Walker, clusters with materials recorded earlier. Plomley [5, p. 19] has suggested that Milligan's vocabularies are heavily influenced by the Tasmanian pidgin [8] in use on Flinders Island, and this is most likely (though they could also represent an otherwise unrecorded language).

Neighbor-Nets assume relatedness between all taxa under analysis. Moreover, they cannot be used to date divergences in lineages, or to quantify support for groups except informally. To further explore these topics, a Bayesian maximum-likelihood analysis was conducted in BEAST [34], using a covarion model with base frequencies estimated empirically (figure 4). While the same clusters identified in the previous methods were also identified here with strong support, more remote groups had only very weak support, including the grouping of northern and western vocabularies. The exception is the node joining the Bruny Island and Oyster Bay clusters, which had strong support. Therefore, we can, with a high degree of confidence, recover four Tasmanian macro-families; though there is weak evidence that the languages of the north and west are remotely related, as are, perhaps, the three families of the eastern coast (see the electronic supplementary materials for discussion of cognate words). There is linguistic support here for Jones' [13] archaeological division between eastern and western Tasmania, and the clusters here correspond to groups previously labelled as 'dialects' or 'languages' in work such as Schmidt [9].

Evidence for a single Tasmanian macro-family, however, is non-existent. Only 24 words (out of 3412) are found in all main branches, and most of those are either terms for recently introduced items, such as cattle, or

are cultural or mythological terms and thus likely to be borrowed. If the languages were all ultimately descended from a common ancestor, that ancestor was spoken too far in the past to be recoverable. Calibrating the divergence of the western and northern group to Jones' identified spread of people down the west coast of Tasmania 3000 years ago [13] leads to a root age estimate of approximately 8500 years BP. However, root estimate confidence limits are very poor (see electronic supplementary material, figure S5); a better interpretation of the tree is that there is very weak support for any relationship beyond the five clusters already identified.

## 3. CONCLUSIONS

The findings here have implications for wider prehistory. Previous work regarding Tasmanian external relationship [35–37]—whether to Australian languages or as part of an Indo-Pacific family—assumes the unity of a Tasmanian language family, and in fact requires such an assumption. The data, however, do not support this view. There is no evidence at this stage that Tasmanian languages are all related to one another.

Likewise, there is no evidence here that 'Tasmanian' is related to the attested Indigenous languages on the Australian mainland. The languages closest to Tasmania in mainland Australia are Pama-Nyungan, a family that most probably spread from northern Australia in the Mid-Holocene [38–41], well after Tasmania was separated from the Australian mainland. If we were to look for Tasmania's nearest linguistic relatives on the mainland, we would perhaps find them among the groups who were replaced by Pama-Nyungan speakers in southern Victoria. However, those languages have been lost without trace.

The results here—12 languages in five clusters—are plausible. Linguistic populations will diversify due to

drift once isolated from one another [42]. Thus, for Tasmanian groups to have maintained one or two languages over thousands of years would have required long-standing cohesive social connections over a population numbering several thousands. But what the ethnographic literature clearly describes is a collection of mobile hearth groups of *ca* 50 people with strong local affiliations (in 'bands') and a much looser set of regional affiliations ('tribes') with seasonal contact and reciprocal exchange rights [1,3,4,43]. Jones infers clan exogamy but tribal endogamy, which would also reinforce localist [44] rather than island-wide stances and would lead to boundary enforcement between tribes. Unless the ethnographic practices and principles of social organization documented at European settlement arose in the recent past prior to Tasmanian colonization, we would expect to find multiple languages, and—depending on the length of time those social patterns had been in place—multiple language families. The results presented here are thus consistent with the anthropological record, while a single language or language family is not.

It is established here that the Tasmanian linguistic scene was diverse at the time of European settlement. The age of that diversity is difficult to establish. It is possible that the Early Holocene demographic collapse led to a reduction in linguistic diversity. Bayesian dating using calibration from divergence along the western coast implies a common ancestor of 8000 years BP, around the time of the demographic collapse. However, root age estimates from the linguistic data are so poor that such a conclusion should be treated with extreme caution. Alternatively, the attested families could represent diversity preserved from a time previous to the demographic collapse. This would imply that language diversity may persist among small groups, even where cultural diversity is low. The former view (that the diversity is more recent) is more likely, however, since modern attested responses to demographic collapse [45] show clear substantial reductions in linguistic diversity.

## 4. METHODS

All vocabulary materials in this study come from N. J. B. (Brian) Plomley's [5] compilation of manuscript and published materials on the languages of Tasmania. The materials are categorized by recorder or compiler, geographical area, and occasionally by speaker (where that information is known). Plomley grouped words from all sources by a standard gloss, and within a gloss, provided language 'headwords'; that is, words from different vocabularies which he argued to represent the same form-meaning pair. The headword sets are closely parallel to a character set in computational phylogenetics [46–48]. The electronic supplementary materials provide further discussion and illustration.

To determine which sources likely contained admixture, STRUCTURE [27] was used. This is a Bayesian clustering procedure that uses multilocus genotype data both to model population structure and to assign individuals to populations. Individuals may be assigned to more than one population, in which case the proportionate membership in each population is inferred. The algorithm identifies populations which are characterized by specific allele frequencies at the loci under study. In linguistic terms, the algorithm identifies clusters of vocabularies with cognate items. It has previously been used

in linguistics by Reesink *et al.* [21] to infer ancient population clusters in Sahul. Following Reesink *et al.*, each wordlist is treated as an individual, and the character values are analogically equivalent to alleles (see electronic supplementary material information for discussion). In the STRUCTURE analysis, data were treated as haploid and no linkage was inferred. The model was a simple admixture model [21]. Run parameters included 35 individuals—the vocabularies under consideration here—with 559 loci. Each locus is an English gloss (examples are in the electronic supplementary material, table S2). Each 100 independent runs were completed for *K*-values varying between 2 and 20, with a 10 000 iteration burn-in followed by 25 000 reps. Electronic supplementary material, figure S1 (compiled with DISTRUCT v. 1.1 [49]) gives the distribution of log-likelihood scores.

For the Bayesian maximum-likelihood analysis conducted with the software program BEAST [34], 26 non-admixed vocabularies were used. Data were 2777 Plomley headwords (out of 3412 possible), coded as present or absent. Headwords which were absent, but for which there was another headword recorded with the same gloss, were coded as 0. Items for which no headword was recorded in that gloss were recorded as missing. The excluded data were those with greater than 70 per cent missing data and terms which refer to acculturation terms (that is, recently introduced items such as guns and cattle). Shared terms for such items reflect rather post-contact loan patterns rather than shared ancestry. Results reported here are from the covarion model; other models are discussed in the electronic supplementary materials.

## REFERENCES

1 Roth, H. L. 1899 *The aborigines of Tasmania*. Halifax, UK: F. King and Sons.

2 Plomley, N. J. B. (ed.) 1966 *Friendly mission; the Tasmanian journals and papers, 1829–1834*. Hobart, Australia: Tasmanian Historical Research Association 1966.

3 Jones, R. M. 1971 *Rocky Cape and the problem of the Tasmanians*. PhD Thesis. University of Sydney, Sydney.

4 Ryan, L. 1981 *The aboriginal Tasmanians*. St. Lucia, Australia: University of Queensland Press.

5 Plomley, N. J. 1976 *A word-list of the Tasmanian Aboriginal languages*. Hobart, Australia: N. Plomley in association with the Government of Tasmania.

6 Crowley, T. & Dixon, R. M. W. 1981 *Handbook of Australian languages* (Tasmanian), vol. 2, pp. 394–427.

7 Wurm, S. A. 1971 Classifications of Australian languages, including Tasmanian. In *Current trends in linguistics* (ed. T. A. Sebeok), pp. 721–778. The Hague: Mouton.

8 Crowley, T. 1996 Early language contact in Tasmania. In *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas* (eds S. A. Wurm, P. Mühlhäusler & D. T. Tryon), p. 25. Berlin, New York: Mouton de Gruyter.

9 Schmidt, W. 1952 *Die Tasmanischen Sprachen*. Utrecht, The Netherlands: Spectrum.

10 Powell, A., Shennan, S. & Thomas, M. G. 2009 Late Pleistocene demography and the appearance of modern

human behavior. *Science* **324**, 1298. (doi:10.1126/science.1170165)

11 Henrich, J. 2004 Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *Am. Antiquity* **69**, 197–214. (doi:10.2307/4128416)

12 Bowdler, S. 1980 Fish and culture: a Tasmanian polemic. *Mankind* **12**, 334–340.

13 Jones, R. 1995 Tasmanian archaeology: establishing the sequences. *Annu. Rev. Anthropol.* **24**, 423–446. (doi:10.1146/annurev.an.24.100195.002231)

14 Pardoe, C., Bowdler, S., Brace, C. L., Plomley, N., Turner, C. G. & Wolfpoff, M. H. 1991 Isolation and evolution in Tasmania (and comments and reply). *Curr. Anthropol.* **32**, 1–21. (doi:10.1086/203909)

15 Bowdler, S. 2010 The empty coast: conditions for human occupation in southeast Australia during the late Pleistocene. *Terra Australis* **32**, 177–185.

16 Thomas, I. 1993 Late pleistocene environments and aboriginal settlement patterns in Tasmania. *Aust. Archaeol.* **36**, 1–11.

17 van Holst Pellekaan, S. 2011 Genetic evidence for the colonization of Australia. *Quat. Int.* **30**, 1–13. (doi:10.1016/j.quaint.2011.04.014)

18 O'Connell, J. & Allen, J. 2004 Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. *J. Archaeol. Sci.* **31**, 835–853. (doi:10.1016/j.jas.2003.11.005)

19 Presser, J. C., Stoneking, M. & Redd, A. J. 2002 Tasmanian aborigines and DNA. *Papers Proc. R. Soc. Tasmania* **136**, 35–38.

20 Walker, G. W., Walker, J. B. & Tasmania, R. S. 1898 *Notes on the Aborigines of Tasmania.* Hobart, Australia: William Grahame, Jun., Govt. Printer.

21 Reesink, G., Singer, R. & Dunn, M. 2009 Explaining the linguistic diversity of sahul using population models. *PLoS Biol.* **7**, e1000241. (doi:10.1371/journal.pbio.1000241)

22 Gray, R. D., Drummond, A. J. & Greenhill, S. J. 2009 Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479–483. (doi:10.1126/science.1166858)

23 Jones, C. & Laffan, S. 2008 Lexical similarity and endemism in historical wordlists of Australian Aboriginal languages of the greater Sydney region. *Trans. Philol. Soc.* **106**, 456–486. (doi:10.1111/j.1467-968X.2008.00209.x)

24 Hunley, K., Bowern, C. & Healy, M. 2012 Rejection of a Serial Founder Effects Model of Genetic and Linguistic Coevolution. *Proc. R. Soc. B* **279**, 2281–2288. (doi:10.1098/rspb.2011.2296)

25 Atkinson, Q. D. & Gray, R. D. 2005 Curious parallels and curious connections: phylogenetic thinking in biology and historical linguistics. *Syst. Biol.* **54**, 513–26. (doi:10.1080/10635150590950317)

26 Nash, D. 2002 Historical linguistic geography of south-east Western Australia. In *Language in native title* (eds J. Henderson & D. Nash), pp. 205–230. Canberra, Australia: Aboriginal Studies Press.

27 Pritchard, J. K., Stephens, M. & Donnelly, P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945.

28 Holland, B. R., Huber, K. T., Moulton, V. & Lockhart, P. J. 2004 Using consensus networks to visualize contradictory evidence for species phylogeny. *Mol. Biol. Evol.* **21**, 1459–1461. (doi:10.1093/molbev/msh145)

29 Huson, D. H. & Bryant, D. 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267. (doi:10.1093/molbev/msj030)

30 Milligan, J. 1857 *Vocabulary of dialects of aboriginal tribes in Tasmania.* Hobart, Australia: James Barnard, Government printer.

31 Milligan, J. 1890 *Vocabulary of the dialects of some of the aboriginal tribes of Tasmania.* Hobart, Australia: Government Printer.

32 Coyne, J. A. & Orr, H. A. 2004 *Speciation.* Sinauer, MA: Sinauer Associates.

33 Dyen, I., Kruskal, J. B. & Black, P. 1992 An Indo-European classification: a lexicostatistical experiment. *Trans. Am. Phil. Soc.* **82**, iii–132. (doi:10.2307/1006517)

34 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)

35 Ruhlen, M. 1994 *On the origin of languages: studies in linguistic taxonomy.* Palo Alto, CA: Stanford University Press.

36 Cavalli-Sforza, L. L. 2001 *Genes, peoples, and languages.* Berkeley, CA: University of California Press.

37 Greenberg, J. H. 1971 The Indo-Pacific hypothesis. *Current trends in linguistics* **8**, 809–871.

38 O'Grady, G. & Hale, K. L. 2004 The coherence and distinctiveness of the Pama-Nyungan language family within the Australian linguistic phylum. In *Australian languages: classification and the comparative method* (eds C. Bowern & H. Koch), pp. 69–92. Amsterdam, The Netherlands: John Benjamins.

39 Hale, K. 1964 Classification of Northern Paman languages, Cape York Peninsula, Australia: a research report. *Oceanic Linguist.* **3**, 248–265. (doi:10.2307/3622881)

40 Evans, N. & Jones, R. 1997 The cradle of the Pama-Nyungans: archaeological and linguistic speculations. In *Archaeology and linguistics: Aboriginal Australia in global perspective* (eds P. McConvell & N. Evans), pp. 385–417. Melbourne, Australia: Oxford University Press.

41 Evans, N. & McConvell, P. 1997 The enigma of Pama-Nyungan expansion in Australia. In *Archaeology and language II*, pp. 174–191. Abingdon, UK: Routledge.

42 Foley, R. A. 2004 The evolutionary ecology of linguistic diversity in human populations. In *Traces of ancestry: studies in honour of Colin Renfrew* (ed. M. Jones). Cambridge, UK: McDonald Institute for Archaeological Research.

43 Walker, J. 1897 Some notes on the tribal divisions of the Aborigines of Tasmania. *Pap. Proc. R. Soc. Tasmania* **1897**, 176–187.

44 Hill, J. H. 1978 Language contact systems and human adaptations. *J. Anthropol. Res.* **34**, 1–26.

45 Nettle, D. & Romaine, S. 2000 *Vanishing voices: the extinction of the world's languages.* Oxford, UK: Oxford University Press.

46 Gray, R. D., Greenhill, S. J. & Ross, R. M. 2007 The pleasures and perils of darwinizing culture (with phylogenies). *Biol. Theory* **2**, 360–375. (doi:10.1162/biot.2007.2.4.360)

47 Greenhill, S. J., Blust, R. & Gray, R. D. 2008 The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evol. Bioinform. Online* **4**, 271.

48 McMahon, A. & McMahon, R. 2005 *Language classification by numbers.* Oxford, UK: Oxford University Press.

49 Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. & Feldman, M. W. 2002 Genetic structure of human populations. *Science* **298**, 2381–2385. (doi:10.1126/science.1078311)