

**Science**

 AAAS

**Language Phylogenies Reveal Expansion Pulses  
and Pauses in Pacific Settlement**

R. D. Gray, *et al.*

*Science* **323**, 479 (2009);

DOI: 10.1126/science.1166858

***The following resources related to this article are available online at  
www.sciencemag.org (this information is current as of January 22, 2009 ):***

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/323/5913/479>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/323/5913/479/DC1>

This article **cites 31 articles**, 5 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/323/5913/479#otherarticles>

This article appears in the following **subject collections**:

Evolution

<http://www.sciencemag.org/cgi/collection/evolution>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

# Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement

R. D. Gray,<sup>1</sup> A. J. Drummond,<sup>2</sup> S. J. Greenhill<sup>1</sup>

Debates about human prehistory often center on the role that population expansions play in shaping biological and cultural diversity. Hypotheses on the origin of the Austronesian settlers of the Pacific are divided between a recent “pulse-pause” expansion from Taiwan and an older “slow-boat” diffusion from Wallacea. We used lexical data and Bayesian phylogenetic methods to construct a phylogeny of 400 languages. In agreement with the pulse-pause scenario, the language trees place the Austronesian origin in Taiwan approximately 5230 years ago and reveal a series of settlement pauses and expansion pulses linked to technological and social innovations. These results are robust to assumptions about the rooting and calibration of the trees and demonstrate the combined power of linguistic scholarship, database technologies, and computational phylogenetic methods for resolving questions about human prehistory.

A fundamental goal of the human sciences is to understand the major factors that have shaped the diversity of our species. At one extreme, innovationist models argue that advances in technology and social organization have driven population expansions and shaped the patterns of cultural and biological diversity (1, 2). At the other extreme, diffusionist/wave models (3) argue that innovations and population expansions are not critically linked, and new technologies diffuse between societies. The settlement of the Pacific ocean by Austronesian speakers (hereafter we will use the term “Austronesian” to refer to these people) is one of the most remarkable prehistoric human expansions. The innovationist “pulse-pause” scenario posits that the Austronesians originated in Taiwan around 5500 years ago and spread through the Pacific in a sequence of expansion pulses and settlement pauses (2, 4–6). According to this scenario, the first pause occurred after the settlement of Taiwan and was followed by a rapid expansion pulse as the Austronesians spread over 7000 km from the Philippines to Polynesia in less than 1200 years. As the Austronesians spread through these regions, they integrated with existing populations and innovated new technologies, including the Lapita cultural complex (5). The archaeological evidence suggests the Austronesians reached the previously uninhabited islands of the Reefs/Santa Cruz around 3000 to 3200 years before the present (B.P.) (7), New Caledonia, and Vanuatu around 3000 years B.P., and Tonga, Samoa and Fiji in Western Polynesia in the period between 2900 to 3200 years B.P. (8, 9). This initial rapid pulse was followed by a second pause in Western

Polynesia coinciding with the development of pre-Polynesian society (6, 10), before a second expansion phase into Eastern Polynesia between 1200 and 1800 years B.P., settling Tahiti, the Cook Islands, Tuamotu, Marquesas, Hawaii, Rapanui, and New Zealand.

In contrast, proponents of the slow-boat scenario argue that the Austronesians emerged from an extensive sociocultural network of maritime exchange in Wallacea (in the region of modern day Sulawesi and the Moluccas) around 13,000 to 17,000 years B.P. based on the dating of mitochondrial lineages (11, 12). This Wallacean slow-boat scenario differs from an alternate slow-boat model that, in agreement with the pulse-pause scenario, postulates an East Asian/Taiwanese origin (13, 14). According to the Wallacean slow-boat scenario, the spread of the Austronesians was driven by the submerging of the Sunda shelf at the end of the last ice age (15). These floods triggered population expansions from the Austronesian homeland in Wallacea in a two-pronged expansion. One of these prongs moved north through the Philippines and into Taiwan. The second expansion prong spread east along the New Guinea coast and into Oceania and Polynesia (following the same route described for the pulse-pause scenario). The pulse-pause and slow-boat scenarios differ substantially in where they locate the Austronesian homeland, in the expansion sequence they postulate, and in the age and timing of this expansion. Genetic studies of Pacific settlement (13, 16–18) have been hampered by problems in separating ancient from recent admixture (19) and difficulties in precisely dating the mitochondrial and Y chromosome haplogroups found in the Pacific (20, 21).

We used phylogenetic analyses of languages to trace the history of human populations because language is linked to other cultural traits (22), contains large amounts of information (23), and

evolves at a rapid rate (24). Gray and Jordan’s (25) previous parsimony analysis of Austronesian lexical data found support for the expansion sequence predicted by the pulse-pause scenario but limitations of the data and methods used meant that the predictions about the timing of Pacific settlement could not be tested.

**Lexical data.** The Austronesian language family is the one of the largest in the world, with around 1200 languages spread from Taiwan to New Zealand and Madagascar to Easter Island. We have constructed a large database of Austronesian basic vocabulary (23, 26), which stores 210 items of basic vocabulary from each language, including words for animals, kinship terms, simple verbs, colors, and numbers. Basic vocabulary is both relatively stable over time and generally less likely to be borrowed between languages (27). From this database, a team of linguists identified the sets of homologous words (“cognates”) following the linguistic comparative method (28). We extracted the cognate sets for 400 well-attested languages for analysis. These languages comprise a third of the entire family and include a representative sample of each recognized Austronesian subgroup. We included two non-Austronesian languages as outgroups to “root” the trees: an archaic variant of the Sino-Tibetan language Chinese that was spoken between 2300 and 2900 years B.P. and the Tai-Kadai language Buyang (28). These languages are not traditionally part of the Austronesian family, but a number of cognates have been identified (29). The cognate sets for all 210 meanings across these 400 languages were encoded into a binary matrix. Identified “borrowings” between languages were removed from further analyses. Simulation studies have shown that the amount of undetected borrowing needs to be very substantial (>20%) to substantially bias either the tree topology or the date estimates (30). The resulting matrix contained a total of 34,440 characters (twice the length of whole mitochondrial genomes), and 6436 of these characters were parsimony informative.

**Language tree topology.** To test the predictions about the origin, sequence, and timing of the Austronesian expansion, we constructed trees using Bayesian phylogenetic methods under a number of models of cognate evolution (28). The best-performing model had a single parameter for cognate gains and losses and modeled character-specific rate variation using a covarion approach where characters could switch between fast and slow rates at different branches on the tree (31).

Early attempts to estimate Austronesian language relationships using lexicostatistical methods (32) produced trees that were dramatically different from those obtained by linguists using the comparative method (33). In contrast, the Bayesian phylogenetic trees (Fig. 1 and fig. S5) we obtained from our basic vocabulary data were congruent with the traditional subgroups identified by phonological and morphological evidence, such as the loss of the Proto-Oceanic uvular trill \*R

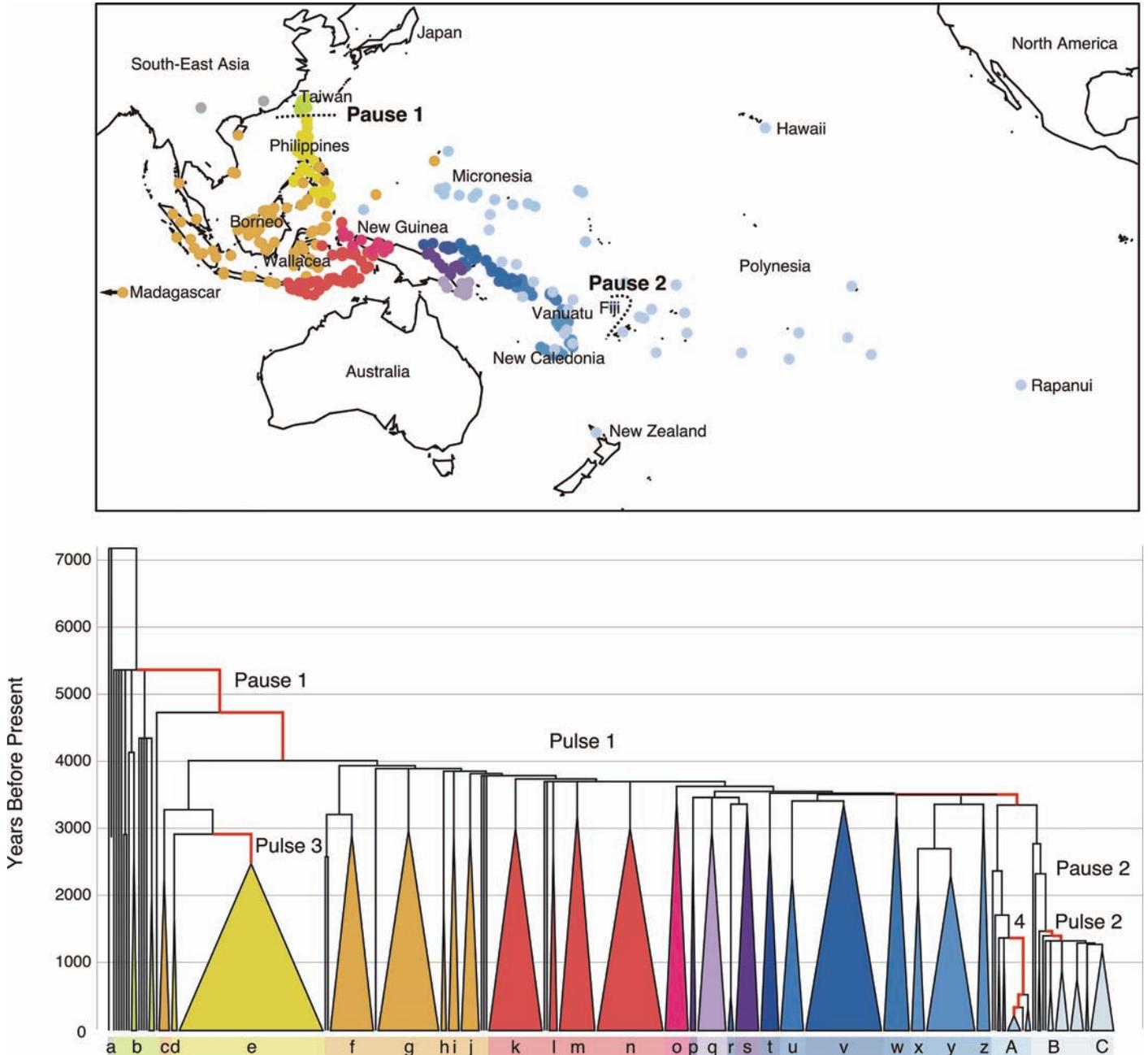
<sup>1</sup>Department of Psychology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. <sup>2</sup>Department of Computer Science, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.

in the Central Pacific subgroup (34) or the lowering of high vowels in morphemes identifying Central-Eastern Malayo-Polynesian (35). The trees support 26 of the 34 putative Austronesian language subgroups and linkages discussed in (28). Of the remaining seven unsupported groups, two are linkages that lack exclusively shared innovations (Central and Western Malayo-Polynesian), and one is only supported by a single sound

change (East Formosan). The remaining five (Western Oceanic, Malayo-Chamic, Greater Central Philippines, Greater Barito, and Barrier Islands/North Sumatra) may be obscured in our analyses because of conflicting signals caused by undetected borrowing between neighboring languages. Our results place the Formosan languages of Taiwan at the base of the trees immediately after the outgroups (Fig. 1). Fol-

lowing these are the languages of the Philippines, Borneo/Sulawesi, Central Malayo-Polynesia, South Halmahera/West New Guinea, and the Oceanic languages. This chained topology is precisely the structure predicted by the pulse-pause scenario.

One potential problem with this analysis is that Old Chinese may be too distantly related to Austronesian to reliably root the tree, whereas



**Fig. 1.** Map and maximum clade credibility tree of 400 Austronesian languages. The tree shows four major expansion pulses and two pauses in Pacific settlement. Branches colored red are those identified as having significant increases in language diversification rates. Major language subgroups are color-coded and labeled as follows: a, Outgroups (Buyang, Old Chinese); b, Formosan; c, Sama-Bajaw; d, Gorontalo-Mongondowic; e, Philippine; f, Barito; g, Malayo-Sumbawan; h, Greater South Sulawesi; i,

Sangiric; j, Celebic; k, Bima-Sumba; l, Yamdena-North Bomberai; m, Central Maluku; n, Timor; o, South Halmahera-West New Guinea; p, Schouten (North New Guinea); q, Papuan Tip; r, Willaumez (Meso-Melanesian); s, North New Guinea; t, Admiralties; u, South-East Solomonic; v, Meso-Melanesian; w, Temotu; x, South Vanuatu; y, North Vanuatu; z, Loyalties/New Caledonia; A, Micronesian; B, Polynesian; and C, Eastern Polynesian.

Buyang and the other Tai-Kadai languages may be a sister group to Malayo-Polynesian (29). To investigate the reliability of the Taiwanese rooting, we conducted a separate analysis using a

stochastic-Dollo model of cognate evolution (28), which can estimate where the root should be without specifying outgroups (36). This additional analysis placed Old Chinese and Buyang at

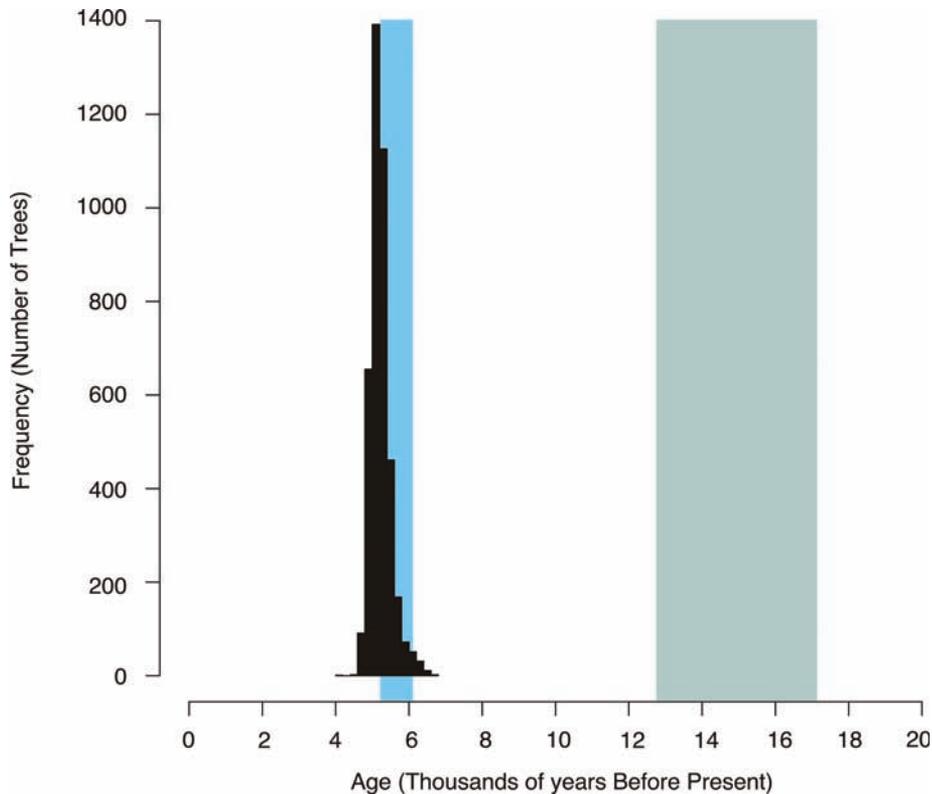
the base of the tree, followed by the Formosan languages with 100% posterior probability.

**Age of Austronesian.** The second key difference between the settlement scenarios is the age of the Austronesian language family. The pulse-pause scenario predicts an origin of Austronesian between 5000 to 6000 years B.P., whereas the Wallacean slow boat scenario predicts an older age of between 13,000 to 17,000 years B.P. To test between these predictions, we estimated the age of Proto-Austronesian by using a penalized likelihood rate-smoothing approach (37). Rather than assuming constant rates of lexical replacement, this method uses calibrations to smooth the observed rates of character change across the trees. We calibrated 10 nodes on the trees with archaeological date estimates and known settlement times (28). The languages Old Chinese and Old Javanese were calibrated to the ages when they were spoken, and Favorlang and Siraya were calibrated to their data collection times. All other languages were treated as contemporaneous.

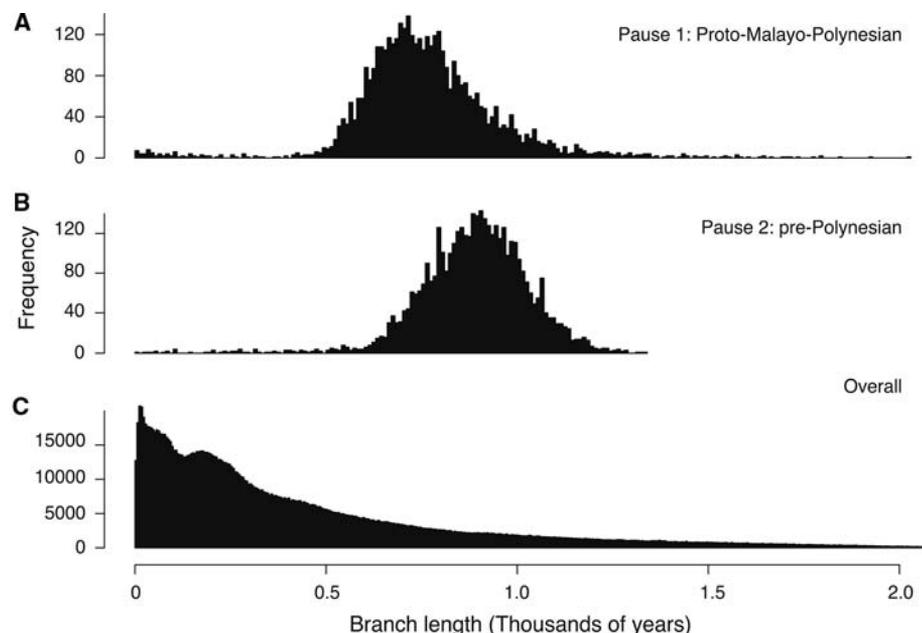
The divergence time estimates for the age of the Austronesian language family support the pulse-pause scenario (Fig. 2). The estimated root age of Austronesian across all the post-burn-in trees has a mean of 5230 years [95% highest posterior density (HPD) interval, 4750 to 5800 years B.P.]. The divergence time estimates were robust across a range of calibrations and different models (28). In particular, estimating the root age of Austronesian without the Proto-Malayo-Polynesian constraint had a trivial effect on the estimated age (mean, 5230 years; 95% HPD, 4730 to 5790 years B.P.). To thoroughly assess the impact of different calibrations, we estimated the age of Proto-Austronesian on the Maximum Clade Credibility tree for all possible calibration combinations. The resulting distribution of the age of Proto-Austronesian had a median of 5110 years (28).

Our estimates for the age of the Austronesian expansion are considerably younger than the deep age estimates of the slow-boat scenario (11, 12, 15). One possibility is that these deep estimates are artifacts due to problems with accurately dating genetic change. There is increasing evidence that rates of genetic change estimated over thousands of years are substantially higher than the long-term substitution rate (21). This violation of the molecular clock leads to the systematic overestimation of recent divergence times. The difficulties of obtaining accurate molecular dates are probably compounded by the use of the error-prone *rho* dating method (20), especially when it is applied to sequences of high-rate heterogeneity such as hypervariable region 1 (HVR-1).

Another possibility is that the genes and languages have quite different histories. However, both Austronesian expansion scenarios envisage a coupling between genetic and linguistic histories. In the pulse-pause scenario, the considerable diversity of Formosan languages reflects the Taiwanese origin of Austronesian. In contrast, the Wallacean slow-boat scenario argues that Taiwan



**Fig. 2.** Histogram of the estimated age of the Austronesian language family. The light blue bar shows the age range predicted by the pulse-pause scenario (5000 to 6000 years B.P.), and the grey bar shows that predicted by the slow-boat scenario (13,000 to 17,000 years B.P.).



**Fig. 3.** Histograms of the branch length distributions. (A) The distribution of the Proto-Malayo-Polynesian pause, (B) the distribution of the pre-Polynesian pause, and (C) the overall branch-length distribution.

is an “Austronesian backwater” (12) and that the initial diversity of Malayo-Polynesian languages has been obscured by language-leveling as a result of extensive socioeconomic networks (11). Recent genetic studies of complete mitochondrial sequences (18) and genome-wide autosomal markers (14, 16) also show that, despite considerable admixture in Near Oceania (38), there is a clear signature linking Austronesian speakers from Taiwan to Polynesia. Even at a fine geographic scale on the east Indonesian island of Sumba, there are strong correlations between languages, genes, and geography (39). The Austronesian expansion has therefore produced a close initial coupling between genes and languages that has subsequently broken down in some regions such as Near Oceania (38).

**Pulses and pauses.** If language diversification (cladogenesis) is linked to population expansions, then expansion pulses should leave a series of short branches in the phylogenies because there will be little time for linguistic changes to accumulate before speech communities fragment. In contrast, when the geographic spread of cultures is constrained by physical or social boundaries, the rate of linguistic diversification should decrease leading to longer branches (anagenesis). The pulse-pause scenario predicts the existence of two settlement pauses: the first occurring before the settlement of the Philippines and corresponding with the development of the Proto-Malayo-Polynesian language around 3800 to 4500 years B.P. (4, 6), and the second occurring after the settlement of Western Polynesia by 2800 years B.P., before the expansion into Central and Marginal Eastern Polynesia (4, 6, 10). This Western Polynesian settlement coincides with the development of the temporally brief Proto-Central Pacific dialect network in Fiji, Tonga, and Samoa (10), with the Polynesian languages emerging from the eastern part of this dialect network some-time later. This second pause is therefore harder to place cleanly on a tree, but should correlate with the development of a pre-Polynesian stage.

To test for the predicted signature of settlement pauses, we extracted all the internal branch lengths from the posterior distribution of the dated trees. We compared the branches corresponding to the Proto-Malayo-Polynesian and pre-Polynesian stages with all other internal branches in the trees (Fig. 3). This is a conservative test because the pauses may be spread across a number of branches (Fig. 1). The Proto-Malayo-Polynesian and pre-Polynesian branches were longer than 81 and 85%, respectively, of a random sample of branches from the overall branch-length distribution (28). A rank-sum test suggests a low probability ( $P = 0.057$ ) of obtaining these ranks or higher by chance.

If these settlement pauses were followed by expansion pulses, then the trees should also show increases in language diversification rates after the pauses. To test this possibility, we modeled diversification rates as a change-point process down the estimated language phylogeny. At each

branch in the tree, we used an indicator variable to model whether the rate of language diversification changed below the branch and a relative rate variable to specify the new rate of diversification relative to the rate on the parent branch. If no change is indicated on a given branch, the diversification rate of the parent language is inherited. We employed a full language-diversification model in which the number, phylogenetic location, and magnitude of the changes in diversification rate were all estimated directly from the data by using Bayesian stochastic variable selection within a Markov chain Monte Carlo method (28).

The posterior estimate of the number of changes in diversification was 4.3 (95% credible set: 1 to 7) with a total of 10 branches showing strong evidence of changes [Bayes Factor (BF) > 20] in diversification rates (Fig. 2 and fig. S4). The pulse with the highest posterior probability occurred in three of the branches leading to Proto-Malayo-Polynesian (BF = 397, 79, and 33, respectively). The second identified pulse occurred in two of the branches after the pre-Polynesian stage (BF = 29 and 36). The location of these two pulses is in agreement with those predicted by the pulse-pause scenario. Changes in diversification rate were also evident in three other locations. The third pulse was found in the branch leading to the Philippines languages (BF = 38) after another lengthy pause. Our results place the age of this pulse around 2500 years B.P. This is consistent with arguments that the Greater Central Philippines subgroup expanded at the expense of other lineages between 2000 and 2500 years B.P., reducing linguistic diversity in the region (40). A fourth pulse was evident in three of the branches leading to the Micronesian languages (BF = 66, 29, and 23). Within this Micronesian group the Trukic languages contain the fastest (single-population) rates in the entire family. The final branch to show a significant increase in diversification rates is that leading to both the Micronesian and Polynesian subgroups and suggests that there might be a common underlying factor between the subsequent pulses into Polynesia and Micronesia (Fig. 1 and fig. S4).

**Discussion.** Our results show that the diversification of Austronesian languages was closely coupled with geographic expansions. The availability of appropriate social and technological resources probably determined the timing of the expansion pulses and settlement pauses. The first pause between the settlement of Taiwan and the Philippines may have been due to the difficulties in crossing the 350-km Bashi channel between Taiwan and the Philippines (4, 6). The invention of the outrigger canoe and its sail may have enabled the Austronesians to move across this channel before spreading rapidly over the 7000 km from the Philippines to Polynesia (4). This is supported by linguistic reconstructions showing that the terminology associated with the outrigger canoe complex can only be traced back to Proto-Malayo-Polynesian and not Proto-Austronesian (41). One possible reason for the second long

pause in Western Polynesia is that the final pulse into the far-flung islands of Eastern Polynesia required further technological advances. These might have included the ability to estimate latitude from the stars, the ability to sail across the prevailing easterly tradewinds, and the use of double-hulled canoes with greater stability and carrying capacity (4, 42). Alternatively, the vast distances between these islands might have required the development of new social strategies for dealing with the greater isolation found in Eastern Polynesia (42). These technological and social advances in Eastern Polynesia may also underlie the fourth pulse into Micronesia.

The language phylogenies reveal the rapidity of major cultural development in the Pacific. As the Austronesians spread along New Guinea and into the Solomons, they developed the Lapita cultural complex through interaction with the existing populations in Near Oceania (5, 10). This complex includes distinctive and often elaborately decorated pottery, adzes/axes, shell ornaments, tattooing, and bark-cloth (10). The phylogenies show that there was only a very small time-window for this complex to develop. Based on the age of the Eastern Malayo-Polynesian clade the Austronesians entered the South Halmahera/West New Guinea region at around 3680 years B.P. (95% HPD, 3640 to 3,710 years B.P.), and had reached Remote Oceania by 3575 years B.P. (95% HPD, 3560 to 3590 years B.P.). The high levels of male-biased admixture detected in Polynesian genetic studies (13, 14) must either have occurred over this very short time span (approximately four generations), with Melanesian males actively incorporated into the Austronesian expansion, or there was extended post-settlement contact between Near Oceania and Polynesia. The results presented here show the combined power of Bayesian phylogenetic methods and large lexical databases to resolve questions about human prehistory. Just as molecular phylogenies provide the fundamental framework for studies of biological evolution, language phylogenies open up the exciting possibility of a Darwinian approach to cultural evolution based on rigorous phylogenetic methods (43).

#### References and Notes

1. J. Diamond, *Guns, Germs, and Steel* (Norton, New York, 1997).
2. J. Diamond, P. Bellwood, *Science* **300**, 597 (2003).
3. R. L. Welsch, J. Terrell, J. A. Nadolski, *Am. Anthropol.* **94**, 568 (1992).
4. R. Blust, in *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, E. Zeitoun, P. J. K. Li, Eds. (Symposium Series of the Institute of Linguistics, Academia Sinica, Taipei, Taiwan, 1999), vol. 1, pp. 31–94.
5. R. C. Green, in *Pacific Archaeology: Assessments and Prospects. Proceedings of the International Conference for the 50th Anniversary of the First Lapita Excavation, Kone-Noumea 2002*, C. Sand, Ed. (Département Archéologie, Service des Musées et du Patrimoine., Noumea, Nouvelle-Calédonie, 2003), pp. 95–120.
6. A. Pawley, in *Examining the Farming/Language Dispersal Hypothesis*, P. Bellwood, C. Renfrew, Eds. (McDonald Institute for Archaeological Research, Cambridge, 2002), pp. 251–274.

7. R. C. Green, M. Jones, P. Sheppard, *Archaeol. Ocean.* **43**, 49 (2008).
8. D. V. Burley, W. R. Dickinson, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11829 (2001).
9. P. D. Nunn *et al.*, *Archaeol. Ocean.* **39**, 139 (2004).
10. P. Kirch, R. Green, *Hawaiki, Ancestral Polynesia: An Essay in Historical Anthropology* (Cambridge Univ. Press, Cambridge, UK, 2001).
11. C. Hill *et al.*, *Am. J. Hum. Genet.* **80**, 29 (2007).
12. S. Oppenheimer, M. Richards, *Sci. Prog.* **84**, 157 (2001).
13. M. Kayser *et al.*, *Mol. Biol. Evol.* **23**, 2234 (2006).
14. M. Kayser *et al.*, *Am. J. Hum. Genet.* **82**, 194 (2008).
15. P. Soares *et al.*, *Mol. Biol. Evol.* **25**, 1209 (2008).
16. J. S. Friedlaender *et al.*, *PLoS Genet.* **4**, e19 (2008).
17. T. Melton *et al.*, *Am. J. Hum. Genet.* **57**, 403 (1995).
18. J. A. Trejaut *et al.*, *PLoS Biol.* **3**, e247 (2005).
19. M. E. Hurler *et al.*, *Am. J. Hum. Genet.* **72**, 1282 (2003).
20. M. P. Cox, *Hum. Biol.* **80**, 335 (2008).
21. S. Y. Ho, B. Shapiro, M. J. Phillips, A. Cooper, A. J. Drummond, *Syst. Biol.* **56**, 515 (2007).
22. R. Mace, M. Pagel, *Curr. Anthropol.* **35**, 549 (1994).
23. S. J. Greenhill, R. Blust, R. D. Gray, *Evol. Bioinform.* **4**, 271 (2008).
24. M. Pagel, Q. D. Atkinson, A. Meade, *Nature* **449**, 717 (2007).
25. R. D. Gray, F. M. Jordan, *Nature* **405**, 1052 (2000).
26. Available at <http://language.psy.auckland.ac.nz/austronesian>.
27. S. M. Embleton, *Statistics in Historical Linguistics* (Studienverlag Brockmeyer, Bochum, 1986).
28. Materials and methods are available as supporting material on Science Online.
29. L. Sagart, in *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, L. Sagart, R. Blench, A. Sanchez-Mazas, Eds. (Routledge-Curzon, London, 2005), pp. 161–176.
30. Q. Atkinson, G. Nicholls, D. Welch, R. D. Gray, *Trans. Philolog. Soc.* **103**, 193 (2005).
31. D. Penny, B. J. McComish, M. A. Charleston, M. D. Hendy, *J. Mol. Evol.* **53**, 711 (2001).
32. I. Dyen, in *Indiana University Publications in Anthropology and Linguistics: Memoir 19 of the International Journal of American Linguistics* (Indiana University, Bloomington, IN, 1965).
33. G. W. Grace, *Ocean. Linguist.* **5**, 361 (1964).
34. J. Lynch, M. Ross, T. Crowley, *The Oceanic Languages* (Curzon, Richmond, UK, 2002).
35. R. Blust, *Ocean. Linguist.* **32**, 241 (1993).
36. G. K. Nicholls, R. D. Gray, J. Roy, *Stat. Soc. B Met.* **70**, 545 (2008).
37. M. J. Sanderson, *Mol. Biol. Evol.* **14**, 1218 (1997).
38. K. Hunley *et al.*, *PLoS Genet.* **4**, e1000239 (2008).
39. S. J. Lansing *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16022 (2007).
40. R. Blust, in *Current Issues in Philippine Linguistics and Anthropology: Parangal kay Lawrence A. Reid, H. Liao, C. R. Galvez Rubino*, Eds. (The Linguistic Society of the Philippines and Summer Institute of Linguistics, Manila, 2005), pp. 31–68.
41. A. Pawley, M. Pawley, in *Austronesian Terminologies: Continuity and Change*, A. Pawley, M. Ross, Eds. (Australian National University, Canberra, 1994), pp. 329–361.
42. G. Irwin, *J. Polyn. Soc.* **107**, 111 (1998).
43. R. D. Gray, S. J. Greenhill, R. M. Ross, *Biol. Theory* **2**, 360 (2007).
44. We thank R. Blust, J. Lynch, J. Marck, C. Nickelsen, M. Ross, L. Sagart, and all contributors to the Austronesian Basic Vocabulary Database for their assistance with lexical data and cognate judgments. We gratefully acknowledge BeSTGRID ([www.bestgrid.org](http://www.bestgrid.org)) for Web hosting and A. Meade, M. Pagel, and the Centre for Advanced Computing and Emerging Technologies at the University of Reading for computational support. R. Blust, M. Corballis, R. Green, F. Jordan, M. Pagel, R. Ross, and M. Suchard commented on the manuscript. Funding was provided by the Royal Society of New Zealand Marsden Fund and a Bright Futures doctoral scholarship to S.G.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/323/5913/479/DC1](http://www.sciencemag.org/cgi/content/full/323/5913/479/DC1)  
Materials and Methods  
SOM Text  
Figs. S1 to S5  
Tables S1 to S3  
References

6 October 2008; accepted 10 December 2008  
10.1126/science.1166858

## REPORTS

# An Entanglement Filter

Ryo Okamoto,<sup>1,2\*</sup> Jeremy L. O'Brien,<sup>3\*</sup> Holger F. Hofmann,<sup>4</sup> Tomohisa Nagata,<sup>1,2</sup> Keiji Sasaki,<sup>1</sup> Shigeki Takeuchi<sup>1,2†</sup>

The ability to filter quantum states is a key capability in quantum information science and technology, in which one-qubit filters, or polarizers, have found wide application. Filtering on the basis of entanglement requires extension to multi-qubit filters with qubit-qubit interactions. We demonstrated an optical entanglement filter that passes a pair of photons if they have the desired correlations of their polarization. Such devices have many important applications to quantum technologies.

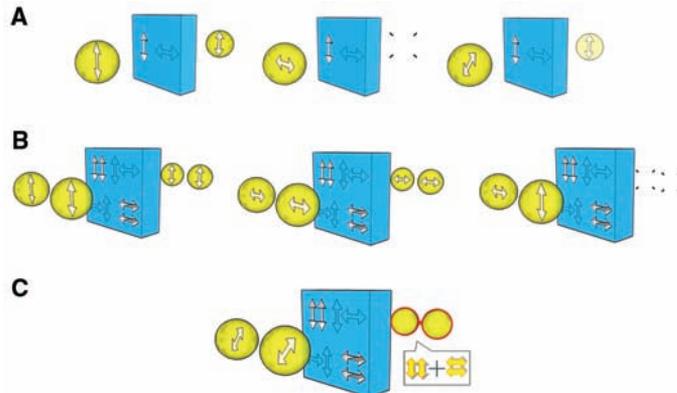
**F**ilters, which pass the desired and reject the unwanted (material, signal, frequency, or polarization, for example), are one of the most important scientific and technological tools available to us. Quantum information science and technology is concerned with harnessing quantum mechanical effects to gain exponential improvement of and new functionality for particular tasks in communication (1), computation (2), measurement (3), and lithography (4). Perhaps the most distinctive of these quantum mechanical features is entanglement. Filters that act on the

quantum correlations associated with entanglement must operate nonlocally on multiple quantum systems, typically two-level qubits. Such a device has been proposed for photonic qubits (5); however, the technical requirements to build such a device, an optical circuit with two ancillary

photons and multiple quantum gates, requiring both quantum interference and classical interference in several nested interferometers, have been lacking.

We demonstrate an entanglement filter made by combining two key recent technological approaches: a displaced-Sagnac architecture (6) and partially polarizing beam splitters (PPBSs) (7–9). The entangling capability of the filter was verified, distinguishing it from classical ones. Because our entanglement filter acts on photonic qubits, it is promising for quantum technologies; photons are the logical choice for communication (1, 10), metrology (6, 11), and lithography (12) and are a leading approach to information processing (13). The filter can be used for the creation as well as the purification of entanglement (14, 15), which will be important in realizing quan-

**Fig. 1.** The function of a polarization filter and an entanglement filter. (A) Polarization filters pass only the certain polarization component of single photons. (B) Entanglement filters pass a pair of photons only if they share the same horizontal or vertical polarization. (C) Because the quantum coherence between these two possibilities is preserved during the process, the output state is entangled when two diagonally polarized photons are input.



<sup>1</sup>Research Institute for Electronic Science, Hokkaido University, Sapporo 060-0812, Japan. <sup>2</sup>The Institute of Scientific and Industrial Research, Osaka University, Mihogaoka 8-1, Ibaraki, Osaka 567-0047, Japan. <sup>3</sup>Centre for Quantum Photonics, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK. <sup>4</sup>Graduate School of Advanced Sciences of Matter, Hiroshima University, Hiroshima 739-8530, Japan.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: [takeuchi@es.hokudai.ac.jp](mailto:takeuchi@es.hokudai.ac.jp)