



Structural Phylogenetics and the Reconstruction of Ancient Language History

Michael Dunn, *et al.*
Science **309**, 2072 (2005);
DOI: 10.1126/science.1114615

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 3, 2007):

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/309/5743/2072>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/309/5743/2072/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/309/5743/2072#related-content>

This article **cites 11 articles**, 1 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/309/5743/2072#otherarticles>

This article has been **cited by** 8 article(s) on the ISI Web of Science.

This article has been **cited by** 2 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/309/5743/2072#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>

approximately 2:1 were tested in a closed-system Y-tube olfactometer (25) against undamaged wild-type *Arabidopsis* plants of the same age. The predatory mites (*Phytoseiulus persimilis*) highly significantly preferred the volatiles emitted by CoxIV-FaNES1 plants to those of wild-type plants (binomial test, $P < 0.001$; Fig. 3A). An infestation with spider mites (*T. urticae*) that did not result in emission of (3S)-(E)-nerolidol and (E)-DMNT did not make wild-type *Arabidopsis* attractive to predatory mites, the natural enemies of the spider mites (Fig. 3A).

Because CoxIV-FaNES1 plants emitted both (E)-DMNT and (3S)-(E)-nerolidol, we assessed which of the two volatiles attracts the predators. (E)-DMNT was previously shown to attract *P. persimilis* (2, 31) (Fig. 3A). However, CoxIV-FaNES1 plants that only emitted (3S)-(E)-nerolidol and no (E)-DMNT were also attractive to *P. persimilis* (Fig. 3A). We then tested the attraction of *P. persimilis* to racemic (E)-nerolidol and found that the predators were significantly attracted. Although nerolidol is often reported as a component in the volatile blend induced by herbivory, to our knowledge, attraction of *P. persimilis* or any other carnivorous arthropod to (3S)-(E)-nerolidol has not been reported previously. Thus, the introduction of a mitochondrially targeted FaNES1 into *Arabidopsis* resulted in the emission of two terpenoids that both attract the predatory mite *P. persimilis*. These signaling molecules, (E)-DMNT and (3S)-(E)-nerolidol, are known to be induced by *P. persimilis*' prey in several plant species (15, 17, 18), but not in wild-type *Arabidopsis* (Fig. 2B).

Attraction of predators to CoxIV-FaNES1 plants was also tested, using plants in soil under more natural conditions, in an octagon setup (Fig. 3B). In this open setup, the odor spreads through diffusion rather than by directing the odor of enclosed plants through a closed container with an air stream. In 10 independent experiments, we found that the majority of the predatory mites made their first visit to the CoxIV-FaNES1 plants, which demonstrates a clear preference ($P < 0.001$) for the undamaged transgenic plants that emit (E)-DMNT and (3S)-(E)-nerolidol (Fig. 3B).

We have shown that genetic engineering of *Arabidopsis*, resulting in plants that emit one or two novel volatiles, provides a novel tool to investigate the role of signaling compounds in mediating tritrophic interactions. This is especially true for compounds that are not commercially available and not easy to synthesize in enantiomer-pure form, such as sesquiterpenoids [e.g., (3S)-(E)-nerolidol] and homoterpenes [e.g., (E)-DMNT]. The levels of the sesquiterpene alcohol (3S)-(E)-nerolidol as well as the homoterpene (E)-DMNT that were emitted by the transgenic plants are the highest reported so far, indicating that FPP is readily available in the mitochondria for metabolic

engineering. Emission of these signaling chemicals from engineered plants demonstrated that these volatiles influence bodyguard behavior in vivo. Our results show that the transgenic approach holds considerable promise for improving crop protection through a transgenic approach (e.g., by exploiting herbivore-inducible promoters coupled to genes responsible for biosynthesis of signaling compounds), so that crop plants can be generated that more effectively recruit biological control agents after infestation with arthropod pests.

References and Notes

1. A. Kessler, R. Halitschke, I. T. Baldwin, *Science* **305**, 665 (2004).
2. M. Dicke *et al.*, *J. Chem. Ecol.* **16**, 381 (1990).
3. T. C. J. Turlings, J. H. Tumlinson, W. J. Lewis, *Science* **250**, 1251 (1990).
4. M. A. Birkett *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9329 (2000).
5. M. Dicke, M. W. Sabelis, in *Causes and Consequences of Variation in Growth Rate and Productivity of Higher Plants*, H. Lambers, M. L. Cambridge, H. Konings, T. L. Pons, Eds. (SPB, The Hague, 1989), pp. 341–358.
6. M. E. Fritzsche-Hoballah, T. C. J. Turlings, *Evol. Ecol. Res.* **3**, 553 (2001).
7. J. J. A. Van Loon, J. G. De Boer, M. Dicke, *Entomol. Exp. Appl.* **97**, 219 (2000).
8. R. M. P. Van Poecke, M. A. Posthumus, M. Dicke, *J. Chem. Ecol.* **27**, 1911 (2001).
9. R. M. P. Van Poecke, M. Dicke, *J. Exp. Bot.* **53**, 1793 (2002).
10. P. W. Paré, J. H. Tumlinson, *Nature* **385**, 30 (1997).
11. N. Dudareva, E. Pichersky, J. Gershenzon, *Plant Physiol.* **135**, 1893 (2004).
12. W. Boland, Z. Feng, J. Donath, A. Gäbler, *Naturwissenschaften* **79**, 368 (1992).
13. G. Arimura, D. P. W. Huber, J. Bohlmann, *Plant J.* **37**, 603 (2004).
14. M. Dicke, *J. Plant Physiol.* **143**, 465 (1994).

15. H. J. Bouwmeester, F. W. A. Verstappen, M. A. Posthumus, M. Dicke, *Plant Physiol.* **121**, 173 (1999).
16. F. Chen *et al.*, *Plant Cell* **15**, 481 (2003).
17. J. Degenhardt, J. Gershenzon, *Planta* **210**, 815 (2000).
18. K. Ament, M. R. Kant, M. W. Sabelis, M. A. Haring, R. C. Schuurink, *Plant Physiol.* **135**, 2025 (2004).
19. T. M. Hohn, J. B. Ohlrogge, *Plant Physiol.* **97**, 460 (1991).
20. T. E. Wallaart, H. J. Bouwmeester, J. Hille, L. Poppinga, N. C. A. Majers, *Planta* **212**, 460 (2001).
21. A. Aharoni *et al.*, *Plant Cell* **15**, 2866 (2003).
22. N. Cunillera *et al.*, *J. Biol. Chem.* **271**, 7774 (1996).
23. N. Cunillera, A. Boronat, A. Ferrer, *J. Biol. Chem.* **272**, 15381 (1997).
24. R. H. Kohler, W. R. Zipfel, W. W. Web, M. R. Hanson, *Plant J.* **11**, 613 (1997).
25. See supporting data on Science Online.
26. A. Aharoni *et al.*, *Plant Cell* **16**, 3110 (2004).
27. J. Donath, W. Boland, *J. Plant Physiol.* **143**, 473 (1994).
28. W. Boland, J. Hopke, J. Donath, J. Nueske, F. Blublitz, *Angew. Chem. Int. Ed. Engl.* **34**, 1600 (1995).
29. R. Karban, I. T. Baldwin, *Induced Responses to Herbivory* (Univ. of Chicago Press, Chicago, 1997).
30. R. M. P. Van Poecke, M. A. Posthumus, M. Dicke, unpublished data.
31. J. G. De Boer, M. A. Posthumus, M. Dicke, *J. Chem. Ecol.* **30**, 2215 (2004).
32. We thank I. Small for the Cox-GFP construct, W. Boland for the synthetic (E)-DMNT and (E,E)-TMTT, and R. Hall, J. van Loon, and R. Mumm for critically reading the manuscript. Supported by Netherlands Organization for Scientific Research (NWO) Technology Foundation grant WPB 5479 (I.F.K., L.L.P.L.) and VICI grant 865.03.002 (M.D.), by Centre for BioSystems Genomics (CBSG) part of Netherlands Genomics Initiative (NGI) project A-1 (A.A.), and by European Commission contract MC-RTN-CT-2003-504720 "ISONET" (M.D.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5743/2070/DC1
 Materials and Methods
 References

17 June 2005; accepted 23 August 2005
 10.1126/science.1116232

Structural Phylogenetics and the Reconstruction of Ancient Language History

Michael Dunn,^{1*} Angela Terrill,^{1,2} Ger Reesink,^{1,2} Robert A. Foley,³ Stephen C. Levinson^{1,2}

The contribution of language history to the study of the early dispersals of modern humans throughout the Old World has been limited by the shallow time depth (about 8000 ± 2000 years) of current linguistic methods. Here it is shown that the application of biological cladistic methods, not to vocabulary (as has been previously tried) but to language structure (sound systems and grammar), may extend the time depths at which language data can be used. The method was tested against well-understood families of Oceanic Austronesian languages, then applied to the Papuan languages of Island Melanesia, a group of hitherto unrelatable isolates. Papuan languages show an archipelago-based phylogenetic signal that is consistent with the current geographical distribution of languages. The most plausible hypothesis to explain this result is the divergence of the Papuan languages from a common ancestral stock, as part of late Pleistocene dispersals.

The linguistic comparative method used to construct language family trees relies on recognizing “cognate sets”: words in different

languages that are related in meaning and form because they can be shown to have the same ultimate source in an ancestor language. The

comparative method has helped define the major linguistic family groups that are recognized today. Unfortunately, because of the continual process of linguistic change, the method is limited to a time depth of approximately 8000 ± 2000 years (1). However, it is probable that a considerable portion of linguistic diversification occurred at earlier dates, associated with later Pleistocene human dispersals. Alternative attempts to reach further back and link the world's ~300 language families (2) into larger taxonomic units are controversial (3–5).

One example of this older diversification may be found in Island Melanesia. Radiocarbon dating for Island Melanesia has demonstrated Pleistocene occupation more than 35,000 years ago (6, 7) (Fig. 1). Evidence suggests high levels of inter- and intrapopulational genetic variation (8, 9), with no simple relationship with linguistic patterns. The languages spoken in the area are of two groups: (i) over 100 languages belonging to four groups of the well-established Austronesian family, which probably originated in the area close to Taiwan and spread to this region about 4000 years ago (10); and (ii) 23 “Papuan” languages, which are not known to have any phylogenetic relation to one another and are of much greater antiquity in the region.

The lexical evidence for relationships between Papuan languages is minimal. Apart from shared Austronesian loans, there are few plausible cognate candidates found in comparisons of pairs of words from Papuan vocabularies (Fig. 2) [see, however, (11)]. Assuming that the rate of vocabulary loss in the Papuan languages is similar to rates observed elsewhere, these languages are either unrelated or have been separated at least since the early Holocene or late Pleistocene. These languages do, however, show a high degree of structural similarity, distinguishing them as a group from their Austronesian neighbors, which has led scholars to propose genealogical (or near-genealogical) groupings (12, 13). In the absence of identifiable lexical cognates, we have used computational cladistic analysis of these features of linguistic structure to test whether a phylogenetic signal can be identified beyond the resolution of lexical form-based methods [for other cladistic methods using lexicons, see (14–21)]. The structural features of a language, like the lexicon, are subject to processes of decay over time and can also be borrowed or exchanged across languages. However, such

exchange usually only occurs under special conditions of prolonged and intensive contact, and it is at least plausible that where the lexical signal has been lost, a faint structural signal might still be discernible. Linguistic structure—that is, grammar rather than vocabulary—has previously been used in historical linguistics to show statistical evidence for ancient links between languages from different parts of the world (1, 2, 22, 23) but not directly to reconstruct phylogenetic relationships.

A questionnaire-based database was constructed, in which linguistic structural features were coded for their presence or absence in each of the target languages. These characters were abstract (coded without respect to their formal expression) and were selected to provide broad typological coverage, reflecting the known linguistic variation of the region (24), as well as to be features that would typically be described in a published sketch grammar. Traits invariant in the region (either entirely absent, such as polysynthesis or proximate/obviative case distinctions; or present in all the languages, such as the existence of a word class

of verbs) were not coded. Characters that show strong implicational correlations were excluded, although characters with weaker tendencies to covariance were not excluded where the current state of linguistic typological knowledge does not allow us to systematically distinguish functionally motivated covariance from phylogenetic or areal patterns. The completed data matrix contained 125 binary features coded for 15 Papuan and 16 Austronesian languages spoken in an overlapping region. The Papuan database was mostly compiled by linguists with field experience in the language and was supplemented from published and unpublished sources where available. The Austronesian database was constructed from published sources (25). All sets of data were checked by a second coder to ensure consistency.

The binary-coded linguistic features allowed us to treat these as character traits distributed among taxonomic units (languages) and thus to apply cladistic algorithms (maximum parsimony or NeighborNet) to determine potential phylogenetic relationships among them (26).

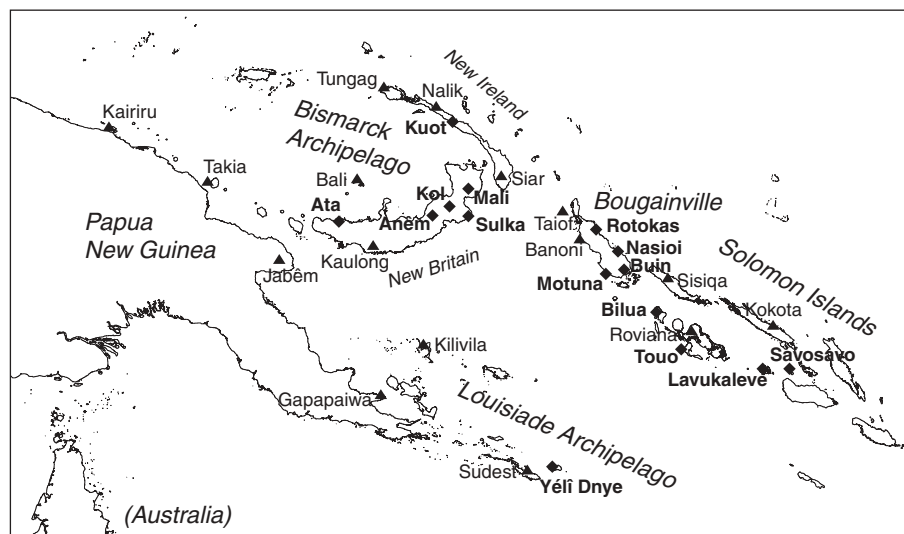


Fig. 1. Island Melanesia, showing the distribution of the Western Oceanic (Austronesian) (triangles) and Papuan (diamonds) languages used in the sample.

Fig. 2. The transparency of cognates in three dispersed Austronesian versus four close Papuan languages (Austronesian cognates/loanwords are shown in italics). The three Papuan languages have an apparent level of 3 to 5% shared vocabulary in a standard 200-word list (29). Using a scrambling test, the word list for each language was randomly reordered, and apparent lexeme correspondences were recounted. The level of apparent cognacy on this random list was exactly the same as on the correctly sorted list, demonstrating that the amount of apparently shared lexicon between any pair of Papuan languages is not greater than chance.

	<i>hand/arm</i>	<i>father</i>	<i>eye</i>
Motu (AN) (mainland PNG)	<i>ima</i>	<i>tama-</i>	<i>mata-</i>
Gela (AN) (central Solomon Islands)	<i>lima</i>	<i>tama-</i>	<i>mata-</i>
Samoan (AN) (Samoa)	<i>lima</i>	<i>tama</i>	<i>mata</i>
Bilua (Pap) (western Solomon Islands)	ngase	<i>mama</i>	vilu
Touo (Pap) (western Solomon Islands)	obi	yae	bero
Lavukaleve (Pap) (central Sol. Islands)	tau vegome	kalem	lemi
Savosavo (Pap) (central Sol. Islands)	kakau	mau	nito

¹Max Planck Institute for Psycholinguistics, Post Office Box 310, 6500 AH Nijmegen, Netherlands.

²Center for Language Studies, Radboud University, Post Office Box 9102, 6500 HC Nijmegen, Netherlands. ³Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK.

*To whom correspondence should be addressed. E-mail: michael.dunn@mpi.nl

The hypothesis that grammatical structure retained a phylogenetic signature was first tested among 16 languages belonging to the Meso-Melanesian, Papuan Tip, and North New Guinea linkages, three sister clades within the Western Oceanic subgroup of Austronesian, the relationship of which has been established by the comparative method (10, 27) {although not completely unambiguously, because there is lexical evidence in particular that the Papuan Tip and the North New Guinea linkages had a period of shared history after their separation from Meso-Melanesian [(10), p. 101]}. We carried out a parsimony analysis on the structural data from these languages, from which we obtained a consensus tree [tree length, 224 steps; consistency index (CI) = 0.42; rescaled consistency index (RC) = 0.19; retention index (RI) = 0.46]. When this tree (Fig. 3, right) is compared with the classification based on the comparative method (Fig. 3, left), there is a close match. In the consensus tree, the Meso-Melanesian group forms a major branch. Papuan Tip and North New Guinea together form a clade, with the North New Guinea linkage nested as a subclade within it. This is consistent with uncertainties in the linguistic reconstruction. The internal structure of the Meso-Melanesian group is quite flat, but all except one of the clades posited by the comparative method are congruently represented in the consensus tree. These results show that cladistically analyzed grammatical structure can preserve a signal that is consistent with a known phylogeny derived by traditional lexical techniques.

On the basis of this result, we applied the same method to a set of languages in which

lexical similarities are not present. Taking 15 Papuan languages for which we have full structural data and applying the same methods, we obtained a consensus tree of the most parsimonious cladograms for the bootstrapped data set (Fig. 4). This tree has a tree length of 349 steps, CI = 0.35, RC = 0.14, and RI = 0.39. The results show a remarkably geographically consistent pattern: The major clades represent archipelagos, and within each archipelago nearest neighbors tend to form sister clades, despite a nearly complete absence of lexical relatedness.

Interpretation is problematic, because there are no generally accepted independent linguistic criteria for assessing the Papuan trees. One possibility is that these trees reflect contact with local Austronesian neighbors, providing an areal rather than phylogenetic signal. In experiments, combined Austronesian-Papuan consensus trees were in some cases intermeshed, but the result was statistically weak (28). Because Papuan and Austronesian are very unlikely to be genuine sister clades, a high degree of homoplasy can be the result of either contact or chance convergence, and combined trees of very remotely related families are likely to be less robust than those where there are good grounds for assuming monophyly. A second possibility is the null hypothesis of no relatedness between the Papuan languages. In that case, we would not expect the orderly and geographically consistent phylogenetic signal that does emerge from the data. This signal is consistent with migration followed by divergence through local isolation. A further possibility is that the geographically consistent tree reflects recent areal contact among Papuan

speakers, but most of these languages are not currently spoken in contiguous regions. Because these languages may have been contiguous in the past, regional diffusion also may account for the phylogenetic signal observed, a possibility that we cannot test without more detailed archaeological information.

We therefore suggest that this method reveals evidence of large-scale genealogical clustering of the Island Melanesian languages; the lack of putative lexical cognates dates these relationships considerably before the Austronesian arrival, in line with the radiocarbon dates from the later Pleistocene, when humans entered Island Melanesia from mainland Papua New Guinea.

There remain important issues to resolve. The first is methodological; bootstrap values, especially in the deeper branches, are low by comparison with biological systems, and further work is required to determine whether this reflects rates of convergence, trait covariation, or processes other than phylogenesis alone. Second, the branching sequence does not fit the generally expected dispersal path. A priori, Island Melanesian Papuan languages should show a general west-to-east pattern of diversification, with the center of diversity in the west. The results of our data are more complex. In particular, the position of the Solomons languages is anomalous, located in the tree between the Bismarcks clade and the Bougainville clade, in violation of geographic expectation [because Bougainville is the natural way-station on the route from mainland New Guinea to the Solomons (Fig. 1)]. During the late Pleistocene, Bougainville and the Solomons were

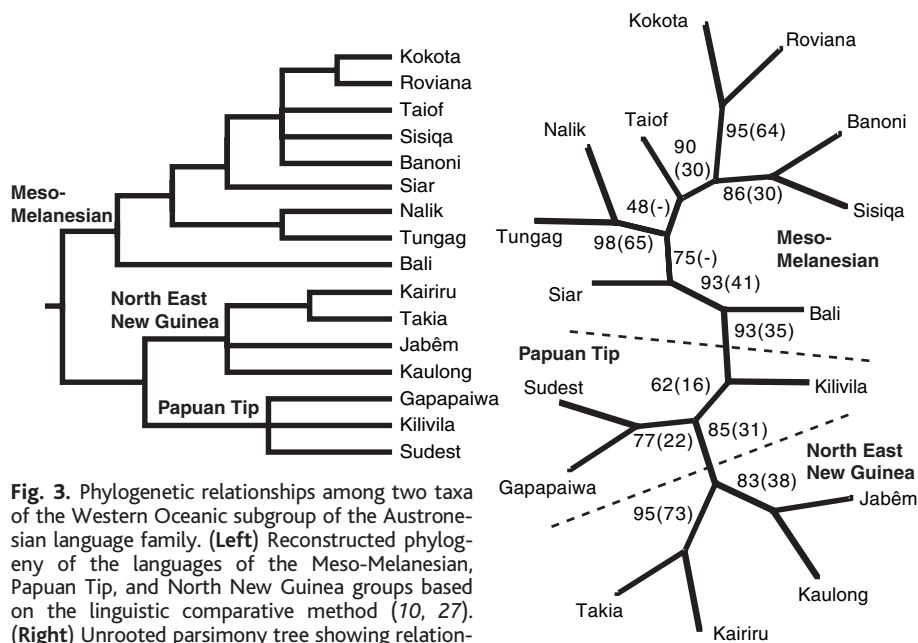


Fig. 3. Phylogenetic relationships among two taxa of the Western Oceanic subgroup of the Austronesian language family. (Left) Reconstructed phylogeny of the languages of the Meso-Melanesian, Papuan Tip, and North New Guinea groups based on the linguistic comparative method (10, 27). (Right) Unrooted parsimony tree showing relationships among the Meso-Melanesian and Papuan Tip groups based on grammatical traits only (that is, discarding abundant lexical evidence) (the figure shows reweighted and raw bootstrap values). The two trees show a high degree of concordance, with monophyly in both major taxa and the similar geographical structuring of within-taxon diversity.

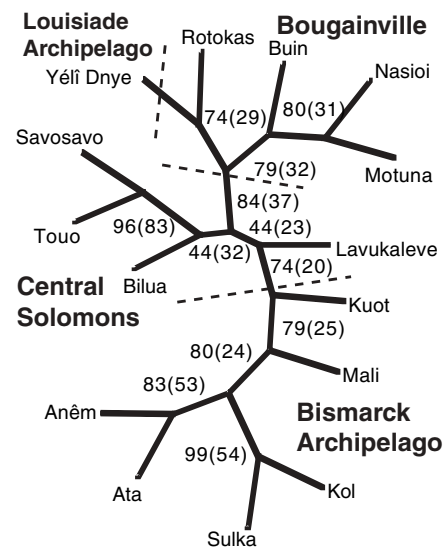


Fig. 4. Maximum parsimony tree of Island Melanesian Papuan languages with reweighted and raw bootstrap values. The tree shows a high level of geographic patterning by island group. Solomon Island languages are intermediate between Bougainville and Bismarck Archipelago languages, which is in violation of geographic progression.

united into a single island, from which the Bismarcks were always separate. A plausible interpretation of the Papuan language tree is thus that the two language groups now located on the Solomons and Bougainville separated from a common ancestor. This could have happened while they could still freely migrate on a common landmass, a time depth (~10,000 years) in accord with that required to erode traces of common vocabulary. This population history hypothesis will require further testing with both linguistic and genetic data.

If grammatical structures can retain a phylogenetic signal beyond the current temporal ceiling on the reconstruction of language history, then the possibility is opened up of finding relationships between others of the world's 300 or so existing language families and isolates.

References and Notes

1. J. Nichols, *Linguistic Diversity in Space and Time* (Univ. of Chicago Press, Chicago, 1992).
2. J. Nichols, in *The Handbook of Historical Linguistics*, B. D. Joseph, R. D. Janda, Eds. (Blackwell, Oxford, 2003), pp. 283–310.
3. J. H. Greenberg, *Language in the Americas* (Stanford Univ. Press, Stanford, CA, 1987).
4. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
5. D. Bolnick, B. Shook, L. Campbell, I. Goddard, *Am. J. Hum. Genet.* **75**, 519 (2004).
6. P. Kirch, *The Lapita Peoples* (Blackwell, London, 1997).

7. M. Spriggs, *The Island Melanesians* (Blackwell, London, 1997).
8. D. A. Merriwether et al., in *Genomic Diversity: Applications in Human Population Genetics*, S. S. Papiha, R. Deka, R. Chakraborty, Eds. (Kluwer Academic/Plenum, New York, 1999), p. 153.
9. M. Kayser et al., *Am. J. Hum. Genet.* **72**, 281 (2003).
10. J. Lynch, M. Ross, T. Crowley, *The Oceanic Languages* (Curzon Press, Richmond, UK, 2002).
11. M. Ross, in *The Boy from Bundaberg: Studies in Melanesian Linguistics in Honour of Tom Dutton*, A. Pawley, M. Ross, D. Tryon, Eds. (Pacific Linguistics, Canberra, Australia, 2001), pp. 301–321.
12. S. A. Wurm, *Papuan Languages of Oceania* (Gunter Narr Verlag, Tübingen, Germany, 1982).
13. J. H. Greenberg, in *Current Trends in Linguistics*, vol. 8, *Linguistics in Oceania*, T. A. Sebeok, Ed. (Mouton and Co., the Hague, 1971), pp. 807–871.
14. D. Ringe, T. Warnow, A. Taylor, *Trans. Philol. Soc.* **100**, 59 (2002).
15. P. Forster, A. Toth, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9079 (2003).
16. R. D. Gray, Q. D. Atkinson, *Nature* **426**, 435 (2003).
17. R. D. Gray, F. M. Jordan, *Nature* **405**, 1052 (2000).
18. C. J. Holden, R. Mace, *Proc. R. Soc. London Ser. B* **270**, 2425 (2003).
19. C. J. Holden, *Proc. R. Soc. London Ser. B* **269**, 793 (2001).
20. K. Rexová, D. Frynta, D. J. Zrzavý, *Cladistics* **19**, 120 (2003).
21. A. McMahon, R. McMahon, *Trans. Philol. Soc.* **101**, 7 (2003).
22. J. Nichols, in *Historical Linguistics 1993. Papers from the Eleventh International Conference on Historical Linguistics*, H. Andersen, Ed. (John Benjamins, Amsterdam, 1995), pp. 337–356.
23. J. Nichols, in *The Origin and Diversification of Language*, N. Jablonski, L. C. Aiello, Eds. (California Academy of Sciences, San Francisco, 1998), pp. 127–170.

24. M. Dunn, G. Reesink, A. Terrill, *Ocean. Linguist.* **41**, 28 (2002).
25. See sources of language data in the supporting online material.
26. Materials and methods and a description of linguistic characters are available as supporting material on Science Online.
27. M. Ross, *Proto Oceanic and the Austronesian Languages of Western Melanesia* (Pacific Linguistics C-98, Canberra, Australia, 1988).
28. This intermeshing of trees does reflect long-term contact in New Britain (26).
29. D. Tryon, B. Hackman, *Solomon Island Languages: An Internal Classification* (Pacific Linguistics C-72, Canberra, Australia, 1983).
30. This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO); the Arts and Humanities Research Council, UK; and the EC Sixth Framework Programme under contract no. ERAS-CT-2003-980409. Additional fieldwork data used in this study were provided by E. Lindström, S. Robinson, T. Stebbins, W. Thurston, and C. Wegener; assistance with coding from published sources was provided by S. Nordhoff and V. Rodrigues. We thank M. Mirazon-Lahr, E. Lindström, and G. Senft for discussion.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5743/2072/DC1

Materials and Methods

Figs. S1 and S2

References

Sources of language data

Linguistic characters

Data file

9 May 2005; accepted 15 August 2005

10.1126/science.1114615

Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments

Edo Kussell* and Stanislas Leibler

Organisms in fluctuating environments must constantly adapt their behavior to survive. In clonal populations, this may be achieved through sensing followed by response or through the generation of diversity by stochastic phenotype switching. Here we show that stochastic switching can be favored over sensing when the environment changes infrequently. The optimal switching rates then mimic the statistics of environmental changes. We derive a relation between the long-term growth rate of the organism and the information available about its fluctuating environment.

Organisms adapt readily to regularly varying environments, for instance, by adjusting to the daily light cycles by using internal circadian clocks. Real problems arise when environmental fluctuations are irregular. Organisms can adapt to sudden changes in chemical composition, local temperature, or illumination by sensing the changes and responding appropriately, for example, by switching phenotype or

behavior. But there is a cost: each individual must maintain active sensory machinery.

Population diversity offers an alternate way to adapt to randomly fluctuating environments. Different subsets of the total population may be well-adapted to different types of environments. In genetically clonal populations, phenotypic diversity is generated by stochastic phenotype-switching mechanisms (1–9). Examples include flagellin phase variation in *Salmonella enterica* (6); microsatellite length variation (slipped-strand mispairing), controlling the expression of contingency genes in *Haemophilus influenzae* (2, 4); and swarming motility in *Bacillus subtilis* (8). The persistence

mechanism in *Escherichia coli*, by which cells switch spontaneously and reversibly to a phenotype exhibiting slower growth and reduced killing by antibiotics (9), allows cells to survive prolonged exposure to antibiotics (10). Many other switching mechanisms are known in diverse bacteria (2, 7), fungi (1–3), and slime molds (1).

The idea that randomization of phenotype can be advantageous in fluctuating environments is well established in the ecology and population genetics literature (where it is known as bet-hedging). This idea has found applications in diverse contexts (11), and it was previously analyzed in several theoretical and computational studies (12–18).

We consider two extreme types of phenotype switching: responsive switching (*R*), occurring as a direct response to an outside cue detected by a sensing mechanism, and spontaneous stochastic switching (*S*), occurring without any direct sensing of the environment. Within a theoretical model, we address several questions. First, under which circumstances should each mechanism be used? For instance, if the detection of a sudden unfavorable environmental change, or the subsequent response, would be too slow, then it could be advantageous to have a subpopulation ready in an appropriate phenotype, before the environmental change.

Second, what determines parameters such as the switching rates? Random environmental

Laboratory of Living Matter and Center for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, Box 34, New York, NY 10021-6399, USA.

*To whom correspondence should be addressed. E-mail: kussele@rockefeller.edu