



An Aboriginal elder, speaking one of the remaining Pama-Nyungan languages, explains ancient customs to boys at Mudjawakalal in Australia.

SOLVING AUSTRALIA'S LANGUAGE PUZZLE

Methods borrowed from evolutionary biology show how people spread across the continent, giving birth to new languages

By **Michael Erard**

The first person to set foot on the continent of Australia was a woman named Warramurrungunji. She emerged from the sea onto an island off northern Australia, and then headed inland, creating children and putting each one in a specific place. As she moved across the landscape, Warramurrungunji

told each child, "I am putting you here. This is the language you should talk! This is your language!"

This myth, from the Iwaidja people of northwestern Australia, has more than a grain of truth, for the peopling and language origins of Australia are closely entwined, says linguist Nicholas Evans of Australian National University (ANU) in

Canberra. But researchers have long puzzled over both. When Europeans colonized Australia 250 years ago, the continent was home to an estimated half-million to 2 million people who were organized into about 700 different groups and spoke at least 300 languages.

Linguists have struggled to work out how these languages were related and when they

emerged. Each was spoken by relatively few people, and as cultures were wiped out by disease and violence, many languages vanished before they could be studied. Researchers prioritized gathering information from the few remaining speakers over deciphering ancient language relationships. But in recent years, researchers borrowing methods used in biology to derive evolutionary trees have begun to unravel the Australian linguistic puzzle. And this week, the approach takes a major step forward, with a combined genetic and linguistic study of the largest Australian language family.

The paper, published in this week's issue of *Nature* along with two other genomic studies of the peopling of Australia (see p. 1352), offers a modern version of

could open a window on how language itself emerged among small social groups in the distant human past. "We need to look at places like Australia, which offer models of language diversification closest to the earliest state that shaped humankind," Evans says.

BACK IN 1963, linguist Ken Hale of the Massachusetts Institute of Technology in Cambridge identified what he considered to be a new Australian language family. He named it Pama-Nyungan ("pama-nahyoongan") for two distinct words for "person," drawn from the geographical extremes of the family's range, which extends across most of Australia (see map, p. 1359). If Hale was right, then Pama-Nyungan, with more than 200 identified languages, would be one

Indo-European word ("génu.")

But Australian languages have few cognates. For example, the sentence "you eat fish" in the Aboriginal languages Iwaidja and Gundjeihmi shares only one cognate element, a grammatical particle that marks the tense of verbs. In Russian ("ty esh rybku") and Elizabethan English ("thou eatest fish"), the sentence shares three—"ty" and "thou," "e-" with "eat," and "-sh" with "est." Yet Moscow and London are much farther apart than the areas where the two Aboriginal languages are spoken.

Perhaps because of these puzzling patterns, linguists have diverged sharply over basic questions such as whether and how Australian languages are related to each other and to languages in nearby New Guinea, likely the source of the first settlers. Some suggested that the Pama-Nyungan family, if it exists, entered the continent in a separate migration, whereas others argued that it split off from other Aboriginal languages only a few thousand years ago.

Now, a new generation of researchers is attacking the problem, and a small but growing group is taking its cue from evolutionary biology, which relies on genetic clues to decipher relationships between organisms. They are using computers to sort giant databases of cognates and generate millions of possible family trees based on assumptions about, say, how quickly languages split. The method, called computational Bayesian phylogenetics, forces researchers to explicitly quantify the uncertainty in the models, says linguist Claire Bowers of Yale University, a pioneer of the approach and co-author of the new study. "That's useful in Pama-Nyungan," she explains, "because you don't have good data, and you have to rely on single authors who may not be that familiar with the languages." Based on a set of parameters, researchers can winnow millions of trees into groups of the most plausible ones.

The first such computational efforts, done by biologists borrowing linguistic data, drew harsh responses from many linguists (*Science*, 19 September 2014, p. 1443). "Most look exclusively at words, seen as something like the equivalent of the gene as a unit of analysis in genetics," says Lyle Campbell, a historical linguist at the University of Hawaii, Manoa. But linguists traditionally determined historical relationships through sounds and grammar, which are more stable parts of language.

Bowers counters that the "instability" of words can actually be a boon, serving as a tracer for how languages change over time. In 2012, she and Quentin Atkinson, a biologist at the University of Auckland in New Zealand, constructed a family tree for the



Warramurrungunji came out of the ocean and walked across the land of Australia, planting languages as she went, as depicted in this illustration of an Aboriginal myth.

Warramurrungunji's story. It paints a picture of how people entered and spread across the continent, giving birth to new languages as they went. It's "a major advance," says Peter Hiscock, an archaeologist at the University of Sydney in Australia. "It presents evidence for an elaborate population history in Australia, spanning 50 millennia." The study, led by evolutionary geneticist Eske Willerslev of the University of Copenhagen, also marks a milestone in collaboration between geneticists and linguists, who for years stayed in their separate camps.

The 25 Aboriginal languages still being passed to new generations make up one of the last and most diverse great hunter-gatherer linguistic groups left. So understanding how they and their extinct relatives diversified

of the world's largest language families—larger than Indo-European and almost as large as Sino-Tibetan.

Not everyone agrees that Pama-Nyungan is one family, however, for, like other Australian language families, it presents a puzzling pattern of similarities and differences. Linguists had long noted that most languages across Australia draw from the same set of sounds, and that their verbs and pronouns share similar patterns of construction.

Given these similarities, linguists would expect the languages to share many cognates, or words derived from a common ancestor. (The English word "knee," ancient Greek "gónu," and Sanskrit "jānu" are all cognates, descended from the Proto-

elusive Pama-Nyungan, using a massive database of 600,000 words to compensate for the low number of cognates. They analyzed 36,000 words from 195 Pama-Nyungan languages and compared the loss and gain of cognate words in 189 meanings through time.

This initial work found that Pama-Nyungan has a deep family tree with four major divisions tied to the southeastern, northern, central, and western regions of the continent. For the study published in *Nature*, Bowers drew from an expanded database of 800,000 words, which contains 80% of all Australian language data ever published, and looked at cognates from 28 languages across 200 meanings. Then she compared her tree with genomic data from Willerslev's new survey.

Willerslev's team sequenced complete genomes from 83 Aboriginal Australians as well as 25 Highland Papuans, and combined those data with published genomes. Using genetic changes as a molecular clock, they conclude that Papuan and Aboriginal Australian ancestors diverged perhaps 37,000 years ago, long before Australia and New Guinea were separated by rising seas. That suggests that people separated into distinct groups while still living on the ancient continent of Sahul, which included modern Australia, New Guinea, and Tasmania. The genetic analysis also found no evidence of multiple migrations into Australia, suggesting that Pama-Nyungan languages must have diversified on the continent.

To the researchers' amazement, the genetic pattern mirrored the linguistic one. "It's incredible that those two trees match. None of us expected that," says paleoanthropologist Michael Westaway of Griffith University, Nathan, in Australia, a co-author on the Willerslev paper. "But it's confusing: The [genetic splits] date to 30,000 years ago or more but the linguistic divisions are only maybe 6000 years old."

Willerslev says he first thought the languages must be much older than thought. "But the linguists told me, 'no way.'"

Both types of data also show that the population expanded from the northeast to the southwest. This migration occurred within the last 10,000 years and likely came in successive waves, Bowers says, in which existing languages were overlaid by new ones. This expansion also seems to correspond with a stone tool innovation called a backed

Tracking a linguistic expansion

Pama-Nyungan is spoken across 90% of Australia. Linguists conclude that the family originated in northeastern Australia and spread to the southwest over millennia.



edge blade. But the accompanying gene flow was just a trickle, suggesting that only a few people had an outside cultural impact, Willerslev says. "It's like you had two men entering a village, convincing everyone to speak a new language and adopt new tools, having a little sexual interaction, then disappearing," he says. Then the new languages continued to develop, following the older patterns of population separation. "It's really strange but it's the best way we can interpret the data at this stage."

When it comes to languages, the Pama-Nyungan tree "gives us the first and only hypothesis of the higher-level branching of the Pama-Nyungan family," says Harold Koch, a historical linguist at ANU who was not involved in the *Nature* study, although he was Bowers' undergraduate adviser. "No one else has tried to answer this question, not because we don't believe there was such a grouping, but because the task seemed too hard. This makes the contribution of huge significance." With his field's usual care, Koch says he'd like to see the model tested with other types of linguistic evidence.

Bowers hopes to also mine the cognate database for insights into pronouns, color terms, and changes of meaning that may give clues to ancient ways of life when climate conditions changed or trading intensified. Last fall in a paper in the *Proceedings of the Royal Society B*, for example, she used the database to analyze how languages gain and lose numbers. One finding was that acquir-

ing a word for "five" often tipped a language into accumulating words for even higher numbers, a change that may have reflected new trade relations that required the ability to count higher (see <http://scim.ag/CountLang>).

Not all linguists embrace Bowers' method or results. Linguist R.M.W. Dixon of James Cook University, Cairns, in Australia, who made his name in the 1960s and 1970s doing fieldwork on Aboriginal languages, says these languages are so unique that new theories of linguistic change must be invented to explain them. In his view the best model of Pama-Nyungan family relations is the parallel lines of a rake, not a tree, and the many similarities in these languages can mainly be accounted for by diffusion—in which language A gets word X from language B because the speakers interact or many people speak both languages. (That's why the word "taco" diffused from Spanish

into English, for example.)

Other linguists argue that the computational models, built for genes that can only be inherited, deal poorly with languages that spread by diffusion. "Borrowings don't really tell us anything about language relatedness," says Asya Pereltsvaig, an independent linguist in Santa Clara, California. "They only obscure it."

Bowers counters that the phylogenetic methods are actually ideal for investigating borrowing, because you can test models with different rates of borrowing and see how well the resulting trees match known facts. Worldwide, about 5% to 10% of languages' vocabularies are borrowed from other languages; Bowers estimates the Pama-Nyungan rate to be 9%. That suggests that Pama-Nyungan languages developed much as other world languages did, rather than being a rarefied case, she argues.

The Aboriginal stories suggest as much, describing the birth of languages much the way Bowers thinks it happened. In 2004, Evans recorded an Iwaidja speaker, Brian Yambikbik, explaining how his language might be related to the one spoken on distant islands. "We used to speak the same language as them, but then the sea came up and we drifted apart, and now our languages are different." ■

Michael Erard is the author of Babel No More: The Search for the World's Most Extraordinary Language Learners.



Solving Australia's language puzzle

Michael Erard (September 22, 2016)

Science **353** (6306), 1357-1359. [doi:

10.1126/science.353.6306.1357]

Editor's Summary

This copy is for your personal, non-commercial use only.

Article Tools Visit the online version of this article to access the personalization and article tools:

<http://science.sciencemag.org/content/353/6306/1357>

Permissions Obtain information about reproducing this article:

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

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.