# Letter to the Editor

*Am. J. Hum. Genet. 75:519–523, 2004*

## Problematic Use of Greenberg's Linguistic Classification of the Americas in Studies of Native American Genetic Variation

*To the Editor:*

In recent years, there has been a burgeoning interest in comparisons of genetic and linguistic variation across human populations. This synthetic approach can be a powerful tool for reconstructing human prehistory, but only when the patterns of genetic and linguistic variation are accurately represented (Szathmary 1993). If one or both patterns are inaccurate, the resulting conclusions about human prehistory or gene-language correlations may be incorrect. Here, we present evidence that comparisons of genetic and linguistic variation in the Americas are problematic when they are based on Greenberg's (1987) classification of Native American languages, for these very reasons.

Greenberg (1987) argued that all Native American languages, except those of the "Na-Dene" and Eskimo-Aleut groups, are similar and can be classified into a single linguistic unit, which he called "Amerind." His tripartite classification (Amerind, Na-Dene, and Eskimo-Aleut) was based on the method of multilateral comparison, which examines many languages simultaneously to detect similarities in a small number of basic words and grammatical elements (Greenberg 1987). Greenberg (1987) also suggested that his three language groupings represent three separate migrations to the Americas, and Greenberg et al. (1986) interpreted their synthesis of the linguistic, dental, and genetic evidence as supportive of this three-migration hypothesis.

Over the past 18 years, this three-migration model has become entrenched in the genetics literature as the hypothesis against which new genetic data are tested (e.g., Torroni et al. 1993; Merriwether et al. 1995; Zegura et al. 2004), and Greenberg's linguistic classification has been the primary scheme used in studies comparing genetic and linguistic variation in the Americas. Of 100 studies of Native American genetic variation published between 1987 and 2004, 61 cite Greenberg (1987) or Greenberg et al. (1986), and at least 19 others were influenced by his tripartite classification (15 studies use the Amerind, Na-Dene, and Eskimo-Aleut groupings, and 4 others use the similar language groupings of Greenberg's student M. Ruhlen.)

Whereas Greenberg's classification has been widely and uncritically used by human geneticists, it has been rejected by virtually all historical linguists who study Native American languages. There are many errors in the data on which his classification is based (Goddard 1987; Adelaar 1989; Berman 1992; Kimball 1992; Poser 1992), and Greenberg's criteria for determining linguistic relationships are widely regarded as invalid. His method of multilateral comparison assembled only superficial similarities between languages, and Greenberg did not distinguish similarities due to common ancestry (i.e., homology) from those due to other factors (which other linguists do). Linguistic similarities can also be due to factors such as chance, borrowing from neighboring languages, and onomatopoeia, so proposals of remote linguistic relationships are only plausible when these other possible explanations have been eliminated (Matisoff 1990; Mithun 1990; Goddard and Campbell 1994; Campbell 1997; Ringe 2000). Greenberg made no attempt to eliminate such explanations, and the putative long-range similarities he amassed appear to be mostly chance resemblances and the result of misanalysis—he compared many languages simultaneously (which increases the probability of finding chance resemblances), examined arbitrary segments of words, equated words with very different meanings (e.g., excrement, night, and grass), failed to analyze the structure of some words and falsely analyzed that of others, neglected regular sound correspondences between languages, and misinterpreted well-established findings (Chafe 1987; Bright 1988; Campbell 1988, 1997; Golla 1988; Goddard 1990; Rankin 1992; McMahon and McMahon 1995; Nichols and Peterson 1996).

Consequently, empirical studies have shown that "the method of multilateral comparison fails every test; its results are utterly unreliable. Multilateral comparison is worse than useless: it is positively misleading, since the patterns of 'evidence' that it adduces in support of proposed linguistic relationships are in many cases mathematically indistinguishable from random patterns of chance resemblances" (Ringe 1994, p. 28; cf. Ringe 2002). Because of these problems, Greenberg's methodology has proven incapable of distinguishing plausible

**Table 1**

**Populations and Language Classifications Used in AMOVAs**

| Population | Language Classification | | Reference |
|---|---|---|---|
| | Greenberg (1987) | Campbell (1997) | |
| Cheyenne/Arapaho | Amerind | Algic | Zegura et al. 2004; D. A. Bolnick and D. G. Smith, unpublished data |
| Chippewa | Amerind | Algic | D. A. Bolnick and D. G. Smith, unpublished data |
| Fox | Amerind | Algic | D. A. Bolnick and D. G. Smith, unpublished data |
| Kickapoo | Amerind | Algic | D. A. Bolnick and D. G. Smith, unpublished data |
| Shawnee | Amerind | Algic | D. A. Bolnick and D. G. Smith, unpublished data |
| ORC Cherokee | Amerind | Iroquoian | D. A. Bolnick and D. G. Smith, unpublished data |
| Stillwell Cherokee | Amerind | Iroquoian | D. A. Bolnick and D. G. Smith, unpublished data |
| Omaha | Amerind | Siouan | D. A. Bolnick and D. G. Smith, unpublished data |
| Sioux | Amerind | Siouan | D. A. Bolnick and D. G. Smith, unpublished data |
| Ingano | Amerind | Quechuan | Bortolini et al. 2003 |
| Paacas Novos | Amerind | Chapacuran | Bortolini et al. 2003 |
| Wayuu (Guajiro) | Amerind | Maipurean | Bortolini et al. 2003 |
| Waiapi (Wayampi) | Amerind | Tupian | Bortolini et al. 2003 |
| Ache | Amerind | Tupian | Bortolini et al. 2003 |
| Asurini | Amerind | Tupian | Bortolini et al. 2003 |
| Cinta-Larga | Amerind | Tupian | Bortolini et al. 2003 |
| Guarani | Amerind | Tupian | Bortolini et al. 2003 |
| Parakana | Amerind | Tupian | Bortolini et al. 2003 |
| Urubu-Kaapor | Amerind | Tupian | Bortolini et al. 2003 |
| Tiriyo | Amerind | Cariban | Bortolini et al. 2003 |
| Yukpa | Amerind | Cariban | Bortolini et al. 2003 |
| Huitoto | Amerind | Witotoan | Bortolini et al. 2003 |
| Yagua | Amerind | Yaguan | Bortolini et al. 2003 |
| Barira (Barí) | Amerind | Chibchan | Bortolini et al. 2003 |
| Warao | Amerind | Warao | Bortolini et al. 2003 |
| Gorotire (Kayapó) | Amerind | Jêan | Bortolini et al. 2003 |
| Kaingang | Amerind | Jêan | Bortolini et al. 2003 |
| Kraho | Amerind | Jêan | Bortolini et al. 2003 |
| Mekranoti (Kayapó) | Amerind | Jêan | Bortolini et al. 2003 |
| Xikrin (Kayapó) | Amerind | Jêan | Bortolini et al. 2003 |
| Ticuna | Amerind | Ticuna | Bortolini et al. 2003 |
| Chickasaw | Amerind | Muskogean | D. A. Bolnick and D. G. Smith, unpublished data |
| Choctaw | Amerind | Muskogean | D. A. Bolnick and D. G. Smith, unpublished data |
| Creek | Amerind | Muskogean | D. A. Bolnick and D. G. Smith, unpublished data |
| Seminole | Amerind | Muskogean | D. A. Bolnick and D. G. Smith, unpublished data |
| Chipewyan | Na-Dene | Eyak-Athabaskan | Bortolini et al. 2003 |
| Greenland Inuit | Eskimo-Aleut | Eskimo-Aleut | Bosch et al. 2003 |

proposals of linguistic relationships from implausible ones, such as Finnish-Amerind (Campbell 1988). Thus, specialists in Native American linguistics insist that Greenberg's methodology was so flawed that it completely invalidates his conclusions about the unity of Amerind, and Greenberg himself estimated that 80%–90% of linguists agreed with this assessment (Lewin 1988).

Given this, the use of Greenberg's (1987) classification can confound attempts to understand the relationship between genetic and linguistic variation in the Americas. Many studies of Native American genetic variation continue to use this classification (e.g., Bortolini et al. 2002, 2003; Fernandez-Cobo et al. 2002; Lell et al. 2002; Gomez-Casado et al. 2003; Zegura et al. 2004). However, Hunley and Long (2004) recently showed that there is a poor fit between Greenberg's classification and the patterns of Native American mtDNA variation. On the basis of their findings, we believe that Greenberg's groupings should no longer be used in analyses of mtDNA variation.

To further evaluate how the use of this classification influences our understanding of the relationship between genetic and linguistic variation in the Americas, we examined how well different linguistic classifications "explain" the patterns of Native American Y-chromosome variation. Data were compiled on the Y-chromosome haplogroups of 523 Native Americans, representing 36 populations (table 1). We compared hierarchical analyses of molecular variance (AMOVAs), using Greenberg's (1987) classification and a more conservative one (Campbell 1997) that is widely accepted by specialists

in historical linguistics of Native American languages (Golla 2000; Hill and Hill 2000). The AMOVAs were based on population frequencies of the haplogroups known to be pre–European contact Native American lineages (Q-M19, Q-M3*, Q-M242*, and C-M130). All calculations were performed by Arlequin 2.000 (Schneider et al. 2000).

The AMOVAs show that differences among Greenberg's three groups could account for some genetic variance ($\Phi_{CT} = 0.319$; $P = .027$), but the more generally accepted linguistic classification (as given in Campbell [1997]) of the same populations (17 groups) explains a greater proportion of the total genetic variance ($\Phi_{CT} = 0.448$; $P < .001$). The magnitude of $\Phi_{CT}$ increases 40.4% when the accepted language classification is used, which indicates that it is important to consider language classifications other than that of Greenberg (1987) when evaluating the relationship between genes and language in the Americas. Other factors, such as geography, have likely influenced patterns of genetic variation more than language, but accepted language groupings should, nonetheless, be used when exploring these relationships.

Thus, in future studies comparing genetic and linguistic variation in the Americas, we recommend use of the consensus linguistic classification, as given in Campbell (1997), Goddard (1996), and Mithun (1999), rather than Greenberg's tripartite classification (Greenberg et al. 1986; Greenberg 1987). In addition, since there is no legitimate reason to believe that "Amerind" is a unified group (linguistic or otherwise), it has been essentially abandoned in linguistics and should not be used in genetic analyses. Finally, because synthetic studies provide such important insights into human prehistory, we advocate continued collaboration between geneticists and linguists (and other anthropologists) to ensure accurate comparisons of genetic, linguistic, and cultural variation.

## Acknowledgments

We thank David Glenn Smith, Stephen Ousley, Keith Hunley, Mark Grote, and two anonymous reviewers for valuable discussions and/or helpful comments on the manuscript.

Deborah A. (Weiss) Bolnick,[1]
Beth A. (Schultz) Shook,[1]
Lyle Campbell,[2,3] and Ives Goddard[4]
[1]*Department of Anthropology, University of California, Davis;* [2]*Department of Linguistics, University of Canterbury, Christchurch, New Zealand;* [3]*Department of Linguistics, University of Utah, Salt Lake City; and* [4]*Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC*

## References

Adelaar WFH (1989) Review of *Language in the Americas,* by Joseph H. Greenberg. Lingua 78:249–255

Berman H (1992) A comment on the Yurok and Kalapuya data in Greenberg's *Language in the Americas*. Int J Am Ling 58:230–233

Bortolini M-C, Salzano FM, Bau CHD, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Bedoya G, Ruiz-Linares A (2002) Y-chromosome biallelic polymorphisms and Native American population structure. Ann Hum Genet 66:255–259

Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524–539

Bosch E, Calafell F, Rosser ZH, Norby S, Lynnerup N, Hurles ME, Jobling MA (2003) High levels of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum Genet 112:353–363

Bright W (1988) Review of *Language in the Americas* by Joseph H. Greenberg. In: American reference books annual 19. Libraries Unlimited, Englewood, CO, p 440

Campbell L (1988) Review of *Language in the Americas* by Joseph H. Greenberg. Language 64:591–615

——— (1997) American Indian languages: the historical linguistics of Native America. Oxford University Press, New York

Chafe WL (1987) Review of *Language in the Americas* by Joseph H. Greenberg. Curr Anthropol 28:652–653

Fernandez-Cobo M, Agostini HT, Britez G, Ryschkewitsch CF, Stoner GL (2002) Strains of JC virus in Amerind-speakers of North America (Salish) and South America (Guarani), Na-Dene-speakers of New Mexico (Navajo), and modern Japanese suggest links through an ancestral Asian population. Am J Phys Anthropol 118:154–168

Goddard I (1987) Review of *Language in the Americas* by Joseph H. Greenberg. Curr Anthropol 28:656–657

——— (1990) Review of *Language in the Americas* by Joseph H. Greenberg. Linguistics 28:556–558

——— (1996) Introduction. In: Goddard I (ed) Languages: handbook of North American Indians. Vol 17. Smithsonian Institution, Washington, DC, pp 1–16

Goddard I, Campbell L (1994) The history and classification of American Indian languages: what are the implications for the peopling of the Americas? In: Bonnichsen R, Steele DG (eds) Method and theory for investigating the peopling of the Americas. Center for the Study of the First Americans, Oregon State University, Corvallis, pp 189–207

Golla V (1988) Review of *Language in the Americas* by Joseph H. Greenberg. Am Anthropol 90:434–435

——— (2000) Review of *American Indian languages: the historical linguistics of Native America*. Lang Soc 29:150–153

Gomez-Casado E, Martinez-Laso J, Moscoso J, Zamora J, Martin-Villa M, Perez-Blas M, Lopez-Santalla M, Lucas

Gramajo P, Silvera C, Lowy E, Arnaiz-Villena A (2003) Origin of Mayans according to HLA genes and the uniqueness of Amerindians. Tissue Antigens 61:425–436

Greenberg JH (1987) Language in the Americas. Stanford University Press, Stanford

Greenberg JH, Turner CG II, Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. Curr Anthropol 27:477–497

Hill JH, Hill KC (2000) American Indian languages. Am Anthropol 102:161–163

Hunley K, Long JC (2004) Does Greenberg's linguistic classification predict patterns of New World genetic diversity? Paper presented at the Annual Meeting of the American Association of Physical Anthropologists, Tampa, April 14–17

Kimball G (1992) A critique of Muskogean, "Gulf," and Yukian material in *Language in the Americas*. Int J Am Ling 58:447–501

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC (2002) The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet 70:192–206

Lewin R (1988) American Indian language dispute. Science 242:1632–1633

Matisoff JA (1990) On megalo-comparison: a discussion note. Language 66:106–120

McMahon A, McMahon R (1995) Linguistics, genetics and archaeology: internal and external evidence in the Amerind controversy. Trans Philol Soc 93:125–225

Merriwether DA, Rothhammer F, Ferrell RE (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98:411–430

Mithun M (1990) Studies of North American Indian languages. Ann Rev Anthropol 9:309–330

——— (1999) The languages of native North America. Cambridge University Press, Cambridge

Nichols J, Peterson DA (1996) The Amerind personal pronouns. Language 72:336–371

Poser WJ (1992) The Salinan and Yurumanguí data in *Language in the Americas*. Int J Am Ling 24:174–188

Rankin RL (1992) Review of *Language in the Americas* by Joseph H. Greenberg. Int J Am Ling 58:324–351

Ringe D (1994) Multilateral comparison: an empirical test. Paper presented at the Annual Meeting of the American Association for the Advancement of Science, San Francisco, February 18–23

——— (2000) Some relevant facts about historical linguistics. In: Renfrew C (ed) America past, America present: genes and languages in the Americas and beyond. McDonald Institute for Archaeological Research, Cambridge, pp 139–162

——— (2002) Review of Joseph L. Greenberg, *Indo-European and its closest relatives: the Eurasiatic language family. Vol. 1: grammar.* J Ling 38:415–420

Schneider S, Roessli D, Excoffier L (2000) Arlequin version 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Szathmary EJE (1993) mtDNA and the peopling of the Americas. Am J Hum Genet 53:793–799

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Zegura SL, Karafet TM, Zhivotosky LA, Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164–175

Address for correspondence and reprints: Deborah A. Bolnick, Department of Anthropology, University of California, Davis, CA 95616. E-mail: daweiss@ucdavis.edu


0002-9297/2004/7503-0020$15.00