Am. J. Hum. Genet. 75:514–517, 2004

Are Sequence Family Variants Useful for Identifying Deletions in the Human Y Chromosome?

To the Editor:

We read with interest the report of a novel deletion of part of the azoospermia factor c (*AZFc* [MIM 415000]) region of the human Y chromosome (Fernandes et al. 2004). This article reported that the deletion is found only in branch N of the Y-chromosome genealogical tree, occurs through one mutational pathway, is ~2.2 Mb in size, and has no effect on spermatogenesis. We, too, recently reported this deletion, which Fernandes et al. termed the "g1/g3" deletion and which we termed the "b2/b3" deletion (Repping et al. 2004). Our findings, however, differed from those of Fernandes et al. in several important particulars: (1) our screening of 1,563 men demonstrated that this deletion is not confined to branch N and that it has at least four independent origins; (2) our analysis revealed two mutational pathways, rather than one, that can generate the deletion, and we confirmed the existence of the inverted *AZFc* organizations that are the intermediate steps in these pathways; (3) on the basis of the reference sequence of the Y chromosome, we concluded that the size of the deletion is 1.8 Mb, rather than ~2.2 Mb; (4) using interphase FISH, we confirmed the amplicon organization that was postulated in the deletion and also identified three instances of duplication subsequent to the deletion; and (5) because of the possibility of a compensatory factor on Y chromosomes in branch N and because of the limited number of deletions outside this branch, we concluded that a possible effect of this deletion on risk of spermatogenic failure cannot be excluded (Repping et al. 2004).

Beyond these differences, however, the characterizations of this and other partial deletions of *AZFc* (Repping et al. 2003) highlight a more important question. At issue is the relative utility of sequence family variants (Saxena et al. 2000), compared with that of plus/minus STSs, for identification and differentiation of deletions involving *AZFc*. *AZFc* is composed entirely of amplicons—repeat units 115–678 kb in length that only differ by ~1 nt per 3,000 bp. These rare differences are called "sequence family variants" (SFVs). We previously relied on SFVs to map and sequence the *AZFc* region of one man's Y chromosome (Kuroda-Kawaguchi et al. 2001). The report by Fernandes et al. (2004) emphasized the use of SFVs in identification of the novel deletion, whereas our analysis relied on plus/minus STSs for identification of the deletion, followed, in most instances, by confirmation with FISH.

Two observations led us to ask whether SFVs, as opposed to plus/minus STSs, offer the simpler and more robust means of detecting and distinguishing deletions in *AZFc*. First, figures 1 and 4 in the report by Fernandes et al. (2004) indicated that negative results at the plus/minus STS sY1192 or 50f2/C combined with positive

Table 1

Plus/Minus STS Results Distinguishing Different Types of Deletions Involving *AZFc*

	Result at STS[a]					
Deletion	sY142	sY1197	sY1191, sY1192, and/or 50f2/C	sY1291	sY1206	sY1201
b2/b3[b]	+	+	−	+	+	+
gr/gr	+	+	+	−	+	+
b1/b3	+	−	−	−	+	+
b2/b4[c]	+	+	−	−	−	+
None	+	+	+	+	+	+

Note.—See Kuroda-Kawaguchi et al. (2001), Repping et al. (2003), Skaletsky et al. (2003), Fernandes et al. (2004), Repping et al. (2004), and GenBank for STSs.

[a] + = present; − = absent.

[b] Termed the "g1/g3" deletion by Fernandes et al. (2004).

[c] "Classical" *AZFc*.

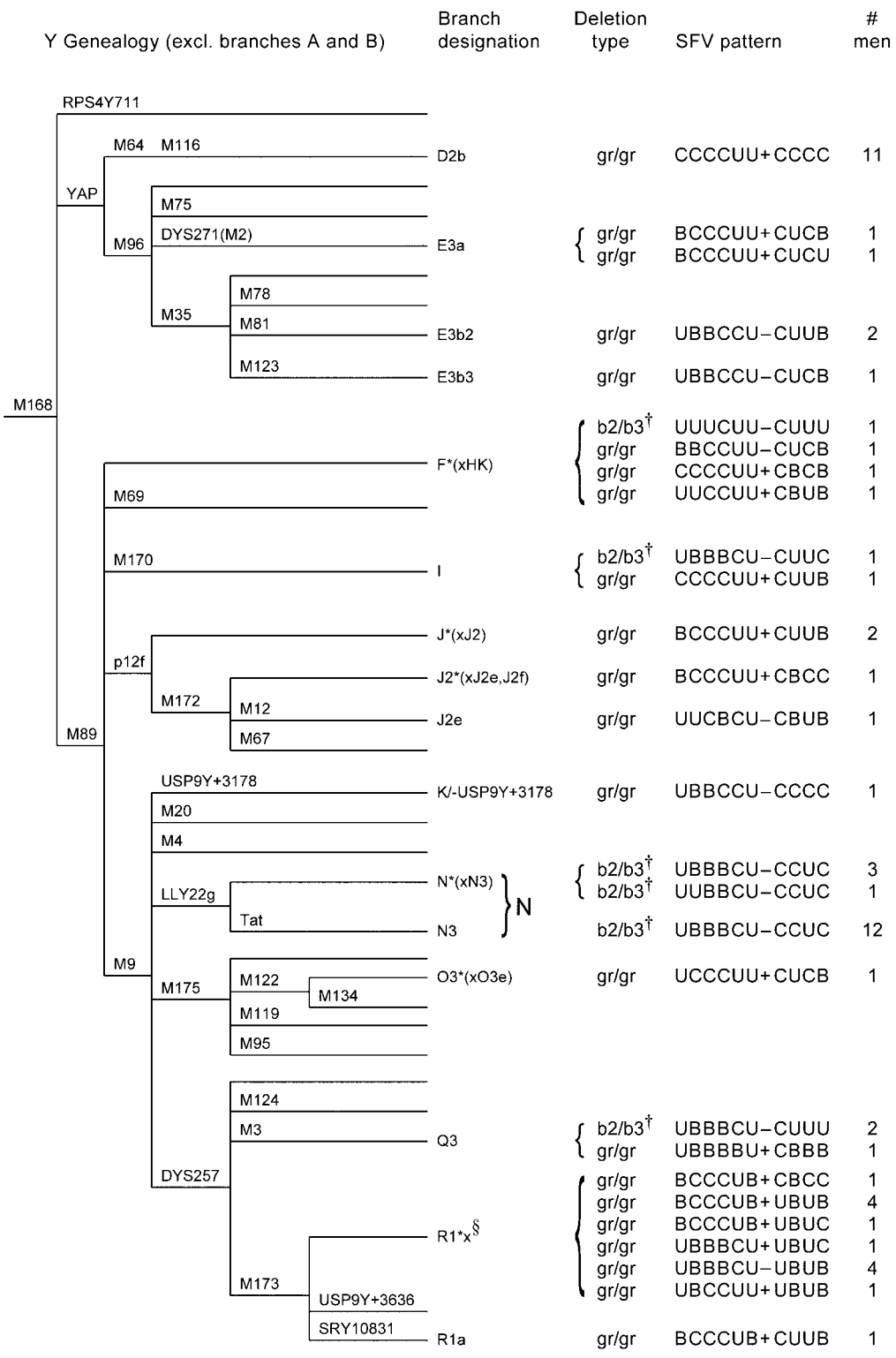

Figure 1 Genealogical analysis of SFV patterns associated with b2/b3 and gr/gr deletions. In the SFV patterns, "C" indicates the cut variant described by Fernandes et al. (2004), "U" indicates the uncut variant, "B" indicates both variants, and + and − indicate the presence or absence, respectively, of the Y-DAZ3 variant. The order of SFVs is as shown in table 2 in the work of Fernandes et al. (2004): DAZ-SNV I, DAZ-SNV II, sY586 (DAZ-SNV III), DAZ-SNV IV, sY587 (DAZ-SNV V), DAZ-SNV VI, *AZFc* SFV 18 (assayed by Y-DAZ3), TTY4-SNV I, BPY2-SNV, GOLY-SNV I, and *AZFc* SFV 20 (AZFc-P1-SNV I) (Saxena et al. 2000; Kuroda-Kawaguchi et al. 2001 [Web table E]; Fernandes et al. 2002, 2004). The genealogical tree of extant human Y chromosomes and the branch designations are from the studies by Underhill et al. (2000) and the Y-Chromosome Consortium (2002). §, R1*x is an abbreviation for R1*(xR1a,R1/-USP9Y+3636). †, Termed "g1/g3" by Fernandes et al. (2004).

results at flanking STSs are sufficient to detect the deletion (table 1). Moreover, the b2/b3 deletion and other types of deletions involving *AZFc* can be distinguished by their plus/minus signatures, without the use of SFVs (table 1).

Second, table 2 in the report by Fernandes et al. (2004) showed that the SFV patterns of undeleted chromosomes vary considerably among different branches of the Y-chromosome genealogy and that the patterns also vary among individuals within branches. These observations suggested that the link between SFV patterns and particular types of deletions would likely not be consistent across the worldwide diversity of Y chromosomes.

The diversity of SFV patterns in undeleted chromosomes is not surprising, since *AZFc* is subject to large inversions, deletions, and duplications caused by ectopic homologous recombination between amplicons (Kuroda-Kawaguchi et al. 2001; Repping et al. 2003, 2004). Such events would rearrange the locations of particular variants and would blur the association between SFV patterns and particular types of deletions. The association would likely be further blurred by gene conversion, which frequently erases small sequence differences (i.e., SFVs) between amplicon copies on the Y chromosome (Rozen et al. 2003).

We experimentally investigated the consistency of SFV patterns in different types of deletions involving *AZFc*. First, using the SFVs employed by Fernandes et al. (2004), we typed 20 men reported elsewhere to have the b2/b3 deletion (Repping et al. 2004) (see GenBank Web site for SFV assays). These men represented branch N and three other branches of the Y-chromosome genealogy (fig. 1). Second, using the same SFVs, we typed 40 men reported elsewhere to have the gr/gr deletion, the other common partial *AZFc* deletion (Repping et al. 2003). These men represented 14 branches of the Y-chromosome genealogy (fig. 1).

The b2/b3 deletions outside branch N showed diverse SFV patterns, and the gr/gr deletions showed even greater diversity (fig. 1). This greater diversity was likely due to the larger number of independent gr/gr deletions studied. Two branches, F*(xHK) and R1*x, contained numerous deletions and a high diversity of SFV patterns (fig. 1). In these branches, multiple independent deletion events probably account for the high diversity. By contrast, two other branches, D2b and N, contained numerous deletions but uniform SFV patterns. This uniformity is explained by the fact that all chromosomes in these branches descended from deleted founders (Repping et al. 2003, 2004; Fernandes et al. 2004). Thus, the chromosomes in each of these branches represent a single deletion event.

Our data also showed that the SFV patterns of b2/b3 and gr/gr deletions are not distinct from each other. For example, the b2/b3 pattern UUUCUU−CUUU (branch F*[xHK]) is more similar to the gr/gr pattern UUCCUU+CBUB (branch F*[xHK], four differences [*underlined*]) than to the b2/b3 pattern UBBBCU−CCUC (branch N, six differences). In another example, the gr/gr pattern UBBBCU−UBUB (branch R1*x) is more similar to the b2/b3 pattern UBBBCU−CUUC (branch I, three differences) than to the gr/gr pattern BCCCUB+CBCC (branch R1*x, 10 differences).

In conclusion, the SFV patterns of b2/b3 and gr/gr deletions vary widely and are not clearly distinct. SFVs can offer insight only if one knows the common SFV organizations in the genealogical branches represented by the Y chromosomes being tested. However, SFV organizations across the Y-chromosome genealogical tree are largely unknown, and SFV patterns vary even among individuals in the same branch. Just as important is that a large number of two-step assays are needed for SFV typing and for determining the Y-chromosome branch. By contrast, six simple plus/minus STSs distinguish between the deletions involving *AZFc* (table 1). Thus, plus/minus STSs provide a straightforward means of identifying and distinguishing the deletions of part of *AZFc,* whereas, in most situations, SFVs do not.

Acknowledgments

We thank J. Lange and H. Skaletsky, for comments, and C. Disteche, B. Gilbert, K. Keppler, T. Kuroda-Kawaguchi, and T. Ogata, for samples. This work was supported by the National Institutes of Health, the Howard Hughes Medical Institute, and the Academic Medical Center.

Sjoerd Repping,[1,2] Cindy M. Korver,[2] Robert D. Oates,[3] Sherman Silber,[4] Fulco van der Veen,[2] David C. Page,[1] and Steve Rozen[1]

[1]*Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge;* [2]*Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam;* [3]*Department of Urology, Boston University Medical Center, Boston; and* [4]*Infertility Center of St. Louis, St. Luke's Hospital, St. Louis*

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for STSs 50f2/C [accession number Y07728], sY142 [accession number G38345], sY1191 [accession number G73809], sY1192 [accession number G67166], sY1197 [accession number G67168], sY1201 [accession number G67170], sY1206 [accession number G67171], and sY1291 [accession num-

ber G72340] and for SFV assays DAZ-SNV I [accession number G73167], DAZ-SNV II [accession number G73166], sY586 [accession number G63907], DAZ-SNV IV [accession number G73168], sY587 [accession number G63908], DAZ-SNV VI [accession number G73169], Y-DAZ3 [accession number G73170], TTY4-SNV I [accession number BV012731], BPY2-SNV [accession number BV012732], GOLY-SNV I [accession number BV012733], and *AZFc* SFV 20 [AZFc-P1-SNV I] [accession number G73351])

Nature Genetics, http://www.nature.com/ng/journal/v29/n3/extref/ng757-S6.doc (for *AZFc* SFVs 18 and 20 in Web table E in Kuroda-Kawaguchi et al. 2001)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for *AZFc*)

References

Fernandes S, Huellen K, Goncalves J, Dukal H, Zeisler J, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH (2002) High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 8:286–298

Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes *DAZ3/DAZ4* and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180–187

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001) The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 29:279–286

Repping S, Skaletsky H, Brown L, van Daalen SKM, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JWA, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247–251

Repping S, van Daalen SKM, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004) A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83:1046–1052

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston JH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873–876

Saxena R, de Vries JWA, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JMN, Page DC (2000) Four *DAZ* genes in two clusters found in the *AZFc* region of the human Y chromosome. Genomics 67:256–267

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza L, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Y-Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348

Address for correspondence and reprints: Dr. Steve Rozen, Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142. E-mail: rozen@wi.mit.edu

0002-9297/2004/7503-0018$15.00

Am. J. Hum. Genet. 75:517–519, 2004

Reply to Repping et al.

To the Editor:

We welcome the enormous contribution that Repping and colleagues have made to the elucidation of the DNA sequence and organization of the Y chromosome, but many questions remain unanswered after the sequencing of the Y chromosome of one man (Skaletsky et al. 2003). It was appreciated, almost a decade ago, that the structure of the *AZFc* region is particularly variable; six independent deletion events and four duplications that affect one short section of this region (50f2/C), together representing ~8% of normal men, were identified by Jobling et al. (1996), and this study could have detected only a small proportion of the total *AZFc* variation. Yet, it provides a useful benchmark for an assessment of our current understanding. We can now define the molecular basis of one of the deletions described in 1996, the haplogroup-12 "small" 50f2/C deletion (Fernandes et al. 2004; Repping et al. 2004*b*), and possibly a second (if the "small" 50f2/C deletion in haplogroup 2 [Jobling et al. 1996] corresponds to the b2/b3 deletion in YCC haplogroup F*[xHK] or I [Repping et al. 2004*b*]), but the 50f2/C duplications all fall on haplotypic backgrounds different from those of the b2/b4 duplications (which include 50f2/C) described so far (Repping et al. 2003). Thus, researchers have still not accounted for at least 8 of 10 rearrangements reported in 1996. It seems that our current methods, whether based on SNVs/SFVs or on plus/minus STSs, allow us to describe only a small proportion of the variation present in this region.

Are plus/minus STSs, nevertheless, more useful than SNVs/SFVs for characterizing *AZFc* variation (Repping et al. 2004*a* [in this issue])? It is a matter of opinion. Even for the best-characterized variants, the gr/gr and g1/g3 (also known as "b2/b3") deletions, it is unclear whether the independent deletions on different lineages

represent true recurrent mutations—taking place at the same recombination site each time—or whether the recombination events have occurred in different locations within the amplicons on different occasions. In the latter case, conflation of different structures—which could have different gene contents—by plus/minus STSs would be a weakness of this classification scheme, and discrim ination between them by SNVs, a strength (fig. 1 of Repping et al. [2004*a*] [in this issue]). It would, however, seem rash to rely on either of these two methods alone—FISH, used by Repping et al. (2003); Southern blotting, used by Fernandes et al. (2002, 2004); and quantitative PCR can all be helpful in defining the structures. But most important of all, this work highlights the importance of an evolutionary understanding of the Y chromosome, and we particularly welcome Repping et al.'s acceptance of this evolutionary approach.

Evolutionary interpretations must, however, be made with caution—we should avoid the "fallacy of the contemporary ancestor" (Jobling et al. 2004). Modern inverted Y chromosomes (see fig. 3 of Repping et al. [2004*b*]) are not the ancestors of haplogroup-N chromosomes, and their frequencies do not indicate which mutational pathway was followed. The best guide to the pre-N structure may be provided by haplogroup-O chromosomes, a sister clade to N in the current Y phylogeny (Jobling and Tyler-Smith 2003) and thus the closest known outgroup. The b2/b3 inversion has indeed been reported in haplogroup O (Repping et al. 2004*b*); if it was present in the common ancestor of the two lineages, the haplogroup-N deletion would result from a g1/g3 deletion following this b2/b3 inversion, rather than a b2/b3 deletion following a g1/g3 inversion. If so, the conclusions from the SNV-based study (Fernandes et al. 2004) would be more accurate than those from the plus/minus STS–based one (Repping et al. 2004*b*).

The present discussion can take place only because our methods for characterizing *AZFc* structures are pitifully inadequate. Rather than behaving like the proverbial group of blind men who encounter an elephant from different sides and insist on describing it from their own favorite partial perspectives, we should assume that all the inversions, duplications, and deletions that are permitted by the sequence will occur, limited only by the winnowing of natural selection. The resulting structures may differ, by many rounds of rearrangement, from the modern haplogroup-R GenBank sequence, but use of the SNP-based phylogeny (Jobling and Tyler-Smith 2003) may allow us to understand the relationship between these structures. It would be even better to develop radically improved ways of elucidating the entire structure so that we can obtain a reasonably complete view of this complex and evolutionarily labile region.

Susana Fernandes,[1,2] Silvia Paracchini,[3] L. H. Meyer,[1] Giovanna Floridia,[4] Chris Tyler-Smith,[5] and Peter H. Vogt[1]

[1]*Section of Molecular Genetics & Infertility, Department of Gynecological Endocrinology & Reproductive Medicine, University of Heidelberg, Heidelberg, Germany;* [2]*Department of Human Genetics, Faculty of Medicine, University of Porto, Porto, Portugal;* [3]*The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford;* [4]*Department of Cell Biology and Neuroscience, Istituto Superiore di Sanità, Rome; and* [5]*The Wellcome Trust Sanger Institute, Hinxton, United Kingdom*

References

Fernandes S, Huellen K, Goncalves J Dukal H, Zeisler J,, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH (2002) High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 8:286–298

Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes *DAZ3/DAZ4* and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180–187

Jobling MA, Hurles ME, Tyler-Smith C (2004) Human evolutionary genetics. Garland Science, New York and Abingdon, United Kingdom

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, Tyler-Smith C (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 5:1767–1775

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4: 598–612

Repping S, Korver CM, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004*a*) Are sequence family variants useful for identifying deletions in the human Y chromosome? Am J Hum Genet 75:514–517 (in this issue)

Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247–251

Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004*b*) A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83: 1046–1052

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al (2003) The male-

specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837

Address for correspondence and reprints: Dr. P. H. Vogt, Section of Molecular Genetics & Infertility, Department of Gynecological Endocrinology & Reproductive Medicine, University of Heidelberg, D-69115 Heidelberg, Germany. E-mail: peter_vogt@med.uni-heidelberg.de

0002-9297/2004/7503-0019$15.00

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 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%–

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

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 Green-

berg'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 Φ_{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

Am. J. Hum. Genet. 75:523–524, 2004

The Phylogeography of Mitochondrial DNA Haplogroup L3g in Africa and the Atlantic Slave Trade

To the Editor:

From the 16th to the 19th century, ~4 million slaves were transported from sub-Saharan Africa to Brazil. With the use of historical records, it is possible to estimate that ~65% of them were Bantus from west-central Africa, ~30% originated from western Africa, and ~5% came from southeastern Africa (Klein 2002).

Salas et al. (2004) have compared the phylogeography of mtDNA haplogroups in Africa with available data on Brazilians and have concluded that their results agree with these historical estimates. However, they were careful to point out that the west-central African contribution to Brazil, signaled by a high frequency of haplogroups L1c and L3e (Bortolini et al. 1997; Alves-Silva et al. 2000; Bandelt et al. 2001), derives largely from an area (the Congo basin) that, thus far, has not been thoroughly analyzed for mtDNA variation. Also, Salas et al. (2002, 2004) called attention to the presence of the haplogroup L3g—which they had only encountered in eastern African populations—in three Brazilians (among the 92 African mtDNA haplotypes that were characterized), one Colombian, and one African American individual. On the basis of this observation, they proposed the occurrence of either direct slave trade from eastern Africa to America or hitherto undetected gene flow from eastern Africa into western or southeastern Africa and then into America.

Intrigued by this proposal, we tried to identify historical evidence of direct slave trade from eastern Africa to Brazil and, indeed, found some anecdotal reports but certainly not enough to explain the significant frequency of L3g seen among African mtDNA haplogroups in Brazil (proportion, 3.3%; 95% CI, 0.7%–9.2%). We then tried to identify other potential sources for the Brazilian L3g mtDNAs. While studying the mtDNA haplotypes of 10 individuals from Cameroon (described in Da Silva

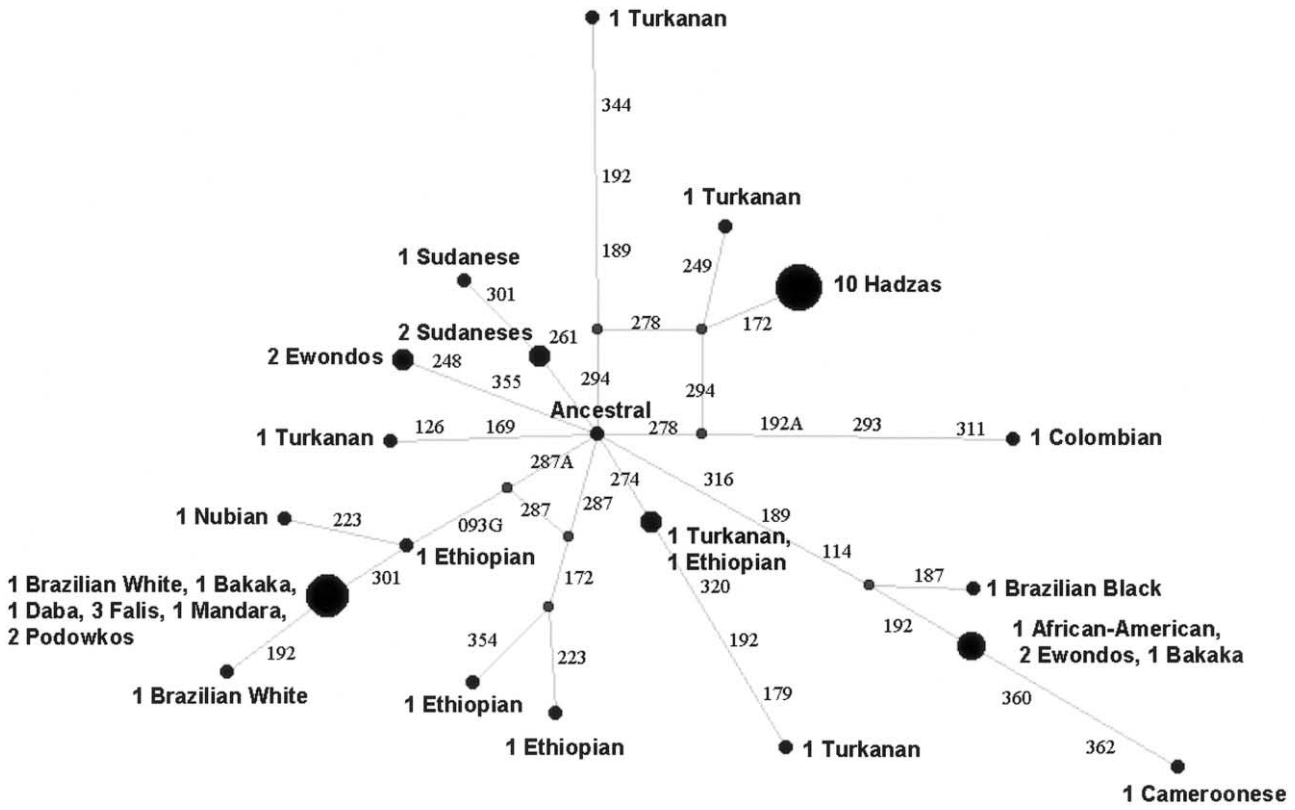

Figure 1 Median network of 20 different L3g lineages identified in populations from sub-Saharan Africa and America. The ancestral haplotype (16223T-16293T-16311C-16355T-16362C) was identified in a single Sudanese individual by Salas et al. (2002). The branches of the network show the additional mutations that identify each haplotype. Positions are indicated as variants from the reference sequence, minus 16,000; a letter next to the position indicates a transversion. The data were compiled from the following groups: Ethiopians, Colombians, and African Americans (Salas et al. 2002); Hadzas (Vigilant et al. 1991); Turkanans (Watson et al. 1997); Nubians (Krings et al. 1999); Ewondos (Destro-Bisol et al. 2004); Bakaka, Bassa, Daba, Falis, Podowkos, and Mandara (see Web site of the Laboratory of Molecular Anthropology, University of Rome "La Sapienza"); Brazilian whites (Alves-Silva et al. 2000); Brazilian blacks (Bortolini et al. 1997); and Camaroonese (present study).

et al. 1999), we identified 1 individual of undisclosed ethnic origin whose mtDNA unequivocally belonged to the L3g haplogroup (fig. 1). This finding stimulated us to search for further mtDNA data from Cameroon, and we came across an article by Destro-Bisol et al. (2004) in which they reported 4 instances of the L3g haplogroup among 53 Ewondo individuals. Moreover, the same authors described (on the Laboratory of Molecular Anthropology Web site) another 11 instances of the L3g haplogroup in several ethnic groups (Bakaka, Bassa, Ewondo, Daba, Fali, Podowko, and Mandara) from different geographical regions of Cameroon. The 14 mtDNA sequences from Cameroon belonged to only four different haplotypes.

We incorporated the four L3g lineages from Cameroon with those from eastern Africa and obtained the median-joining network shown in figure 1. One Brazilian haplotype was identical to the most common Cameroonese haplotype (seen in 8 of the 14 cases), whereas another was closely related. Both the third Brazilian and the single African American haplotype also clustered with Cameroonese sequences. A noteworthy feature of the network was that there did not appear to exist any clear segregation of the Cameroonese L3g haplotypes from the eastern African counterparts. This feature, plus the fact that the putative ancestral haplotype was seen in an individual from Sudan and that there is a much smaller haplotype diversity in Cameroon (0.602) as compared with eastern Africa (0.911, excluding the Hadza), suggests that the Cameroonese L3g lineages might have originated from eastern Africa by transcontinental gene flow, as put forward by Salas et al. (2004) in one of their two possible scenarios. At any rate, it appears that the L3g lineages seen in America probably have their immediate origin in Cameroon or in neighboring regions and not in eastern Africa.

Acknowledgments

We are grateful to Dr. Giovanni Destro-Bisol, Dr. Valentina Coia, and Dr. Gabriella Spedini from the Laboratory of Molecular Anthropology, Department of Animal and Human Biology, University of Rome "La Sapienza," who very kindly gave us access to their data and allowed us to use their data in this letter. Our research was supported by the Conselho Nacional de Pesquisas of Brazil and the Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior.

MARIA CÁTIRA BORTOLINI,[1]
WILSON A. DA SILVA, JR.,[2]
MARCO ANTÔNIO ZAGO,[3] JACQUES ELION,[4]
RAJAGOPAL KRISHNAMOORTHY,[4]
VANESSA F. GONÇALVES,[5] AND SÉRGIO D. J. PENA[5]
[1]*Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil;* [2]*Departamento de Genética and* [3]*Departamento de Clínica Médica, Universidade de São Paulo, Ribeirão Preto, Brazil;* [4]*INSERM UMR 458, Biochimie Génétique, Hôpital Robert Debré, Paris; and* [5]*Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*

Electronic-Database Information

The URL for data presented herein is as follows:

Laboratory of Molecular Anthropology, University of Rome "La Sapienza," http://www.scienzemfn.uniroma1.it/labantro/

References

Alves-Silva J, da Silva Santos M, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SDJ, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461

Bandelt H-J, Alves-Silva J, Guimarães PEM, Santos MS, Brehm A, Pereira L, Coppa A, Larruga JM, Rengo C, Scozzari R, Torroni A, Prata MJ, Amorim A, Prado VF, Pena SDJ (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65:549–563

Bortolini MC, Zago MA, Salzano FM, Silva-Junior WA, Bonatto SL, da Silva MC, Weimer TA (1997) Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. Hum Biol 69:141–159

Da Silva WA Jr, Bortolini MC, Meyer D, Salzano FM, Elion J, Krishnamoorthy R, Schneider MPC, De Guerra DC, Layrisse Z, Castellano HM, Weimer TA, Zago MA (1999) Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. Am J Phys Anthropol 109:425–437

Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, Spedini G, Calafell F (2004) The analysis of variation of mtDNA hypervariable region 1 suggests that eastern and western pygmies diverged before the Bantu expansion. Am Nat 163:212–226

Klein HS (2002) As origens Africanas dos escravos brasileiros. In: Pena SDJ (ed) *Homo brasilis:* aspectos genéticos, linguísticos, históricos e socioantropológicos da formação do povo brasileiro. FUNPEC, Ribeirão Preto, Brazil, pp 93–112

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet 64:1166–1176

Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo Á (2002) The making of the African mtDNA landscape. Am J Hum Genet 71: 1082–1111

Salas A, Richards M, Lareu M-V, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo Á (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of mitochondrial DNA. Science 253:1503–1507

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704

Address for correspondence and reprints: Dr. Sérgio D. J. Pena, Departamento de Bioquímica, ICB-UFMG, Caixa Postal 486, 30161-970 Belo Horizonte, MG, Brazil. E-mail: spena@dcc.ufmg.br

0002-9297/2004/7503-0021$15.00

Am. J. Hum. Genet. 75:524–526, 2004

Reply to Bortolini et al.

To the Editor:

The availability of new mtDNA data from central Africa has allowed Bortolini et al. (2004 [in this issue]) to evaluate two alternative scenarios, formulated by Salas et al. (2004), regarding the source of the L3g mtDNAs carried from Africa to America by the Atlantic slave trade. Bortolini et al. proposed that the American L3g haplotypes have an Atlantic African provenance, rather than a direct eastern African origin, and that their most likely source was Cameroon or the neighboring regions.

On the basis of the extensive amount of new data that can be added to the L3g phylogeny (fig. 1), we are wholeheartedly in favor of this scenario. Of particular interest are three mtDNA sequences belonging to the L3g1 haplogroup that we observed in southwestern Africa (Angola and Cabinda), this region being the second most important source for the Atlantic slave trade (Thomas 1997), with an important demographic impact in Bra-

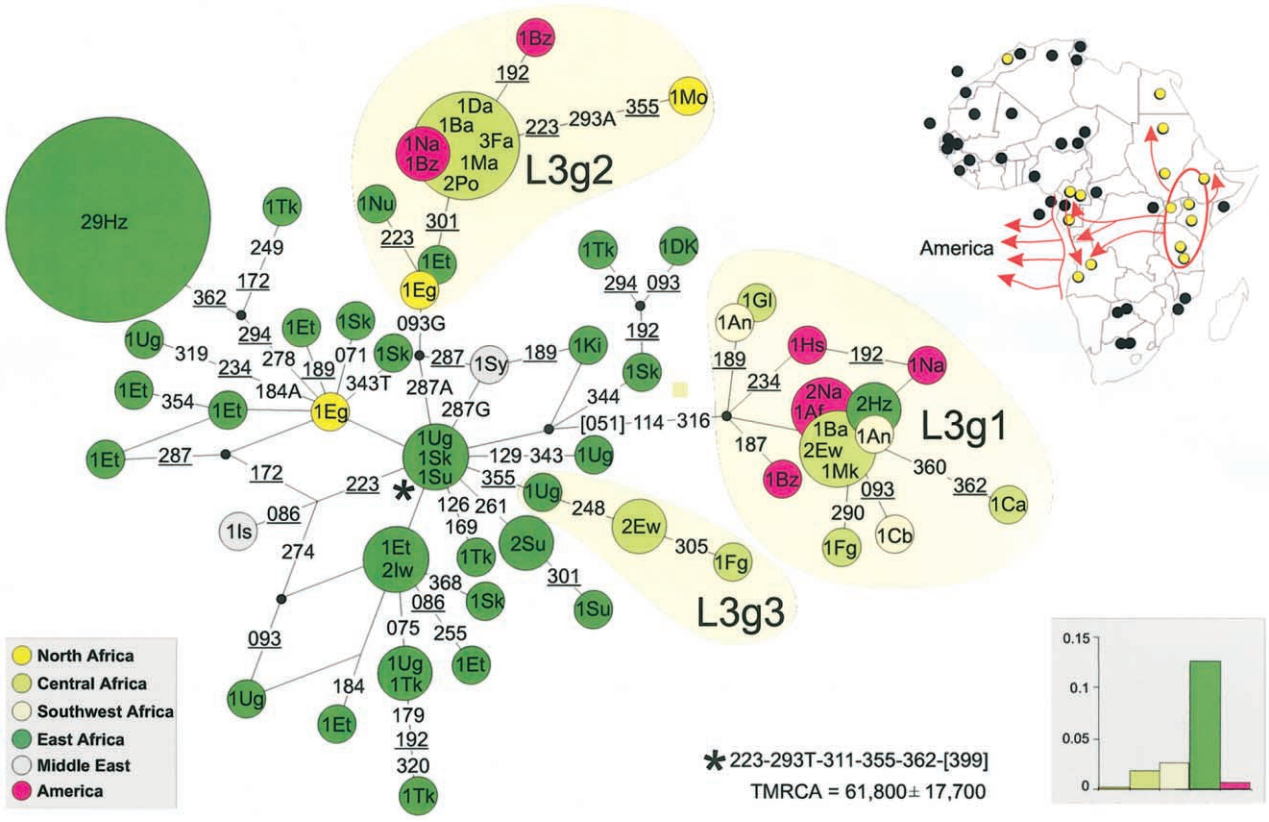

Figure 1 Median-joining network (Bandelt et al. 1999) of L3g mtDNAs. The circle sizes are proportional to the haplotype frequency in the sample. Positions are indicated as variants from the revised Cambridge reference sequence, minus 16,000; a letter next to the position indicates a transversion. Parallel mutations are underlined and diagnostic positions outside of the common mtDNA segment analyzed (from 16060 to 16362) are in brackets. We here define three sublineages of L3g: L3g1, characterized by 16051-16114-16189-16316; L3g2, characterized by 16093G-16287A; and L3g3, which lacks transition 16355 from the root. The histogram (*bottom right*) illustrates the frequency of L3g in different African regions and in America, and the map (*top right*) shows the diffusion pattern of L3g within Africa. A yellow dot indicates the presence of L3g in the population sample, whereas a black dot indicates its absence. Time estimates are computed as in Salas et al. 2002. Note that the Colombian mtDNA included in Bortolini et al.'s (2004 [in this issue]) network has been excluded from this phylogeny, whereas the Cameroonese data is included here. Eastern Africa: Tz = Tanzania, Hz = Hadza (Tanzania), Dk = Dakota (Tanzania), Sk = Sukuma (Tanzania), Iw = Iraqw (Tanzania), Ki = Kikuyu (Kenya), Tk = Turkana (Kenya), Ug = Uganda, Et = Ethiopia, Su = Sudan, Nu = Nubia (Sudan/Egypt). Northern Africa: Eg = Egypt, Mo = Berber (Morocco). Central Africa: Ew = Ewondo (Cameroon), Ba = Bakaka (Cameroon), Da = Daba (Cameroon), Fa = Falis (Cameroon), Ma = Mandara (Cameroon), Po = Podowkos (Cameroon), Ca = Cameroon, Fg = Fang (Gabon), Mk = Makina (Gabon), Gl = Galoa (Gabon). Southwestern Africa: Cb = Cabinda, An = Angola. America: Bz = Brazil, Hs = "Hispanic" (North America), Na = North America. Middle East: Sy = Syria, Is = Israel. TMRCA = time to the most common recent ancestor. Details of the L3g sequences will be supplied by the corresponding author on request.

zil. Overall, L3g1 appears to reflect the contribution of southwestern Africa—and probably central Africa also—to America (66% of the American L3g sequences), whereas the American L3g2 mtDNAs might be of predominantly central African origin. Because L3g3 is present in central and southwestern Africa, it might be expected that L3g3 sequences will also be found in future surveys of American populations of recent African descent. Thus, Brazilian types (which occur within L3g1 and L3g2) can be of either central or southwestern African origin.

We can now briefly reconstruct a plausible history of L3g (fig. 1). Both the phylogeography and the time depth (TMRCA, 61,800 years ago ± 17,700 years) of L3g clearly testify to its eastern African origin. Indeed, the root type (16223-16293T-16311-16355-16362-16399) is found in Sudan, Uganda, and Tanzania, and L3g displays the highest divergence in Tanzania, Uganda, Kenya, and Ethiopia (with a strong founder event distinguishing the click-language isolate of the Hadza) (fig. 1). Diversity on the Atlantic coast of Africa is, by contrast, restricted to a few mtDNAs at the tips of the network. This may have been the result of interactions established after contact between southerly dispersing western and eastern Bantu speakers who spread from the Cameroon region and the Great Lakes, respectively

(Phillipson 1993). In the wake of this interaction, some L3g lineages may have been diffused towards the Atlantic west coast (Cameroon, Gabon, and Angola). The recent arrival of L3g on the Atlantic coast (during or subsequent to the initial Bantu dispersals) likely explains its low diversity in this region, in contrast with the high diversity in eastern Africa (e.g., a single L3g2 type accounts for most [~57%] of the central African L3g mtDNAs). Long-term networks established between central and southwestern Africa after the initial long, gradual, and intermittent western Bantu expansion (Vansina 1995) would have contributed to its subsequent diffusion. The Bantu expansion would also explain the distribution of other central African haplogroups (e.g., L1c) and the lack of strong genetic drift in southwestern Africa (which is detected in the southeast in some Bantu lineages [Salas et al. 2002]). More recently, these haplogroups would have been carried to America during the slave-trade period.

From this view, we can safely rule out the Atlantic coast of western Africa as an important source for American L3g, since this haplogroup has not been detected at present in a large sample (>1,200 mtDNAs) that includes individuals from, among other places, Cabo Verde, Senegal, Sierra Leone, and Nigeria. Some diffusion into northern Africa (Egypt, with signatures in Sudan and Nubia) as well as into the Middle East (Syria, Israel, and Palestine) has been detected, probably reflecting the haplogroup's greater antiquity in eastern Africa.

In conclusion, we can now extend the putative area of origin of the American L3g to the Atlantic fringe that runs from Cameroon to Angola and can probably rule out a direct eastern African origin. The latter surmise also agrees with historical documentation. Important regions, however, remain uncharacterized, such as the Congo basin and the Central African Republic.

Acknowledgments

We would like to thank Giovanni Destro-Bisol, Valentina Coia, and Gabriella Spedini (Laboratory of Molecular Anthropology, Department of Animal and Human Biology, University of Rome "La Sapienza") and Sandra Beleza, Leonor Gusmão, and Antonio Amorim (Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Portugal), for allowing us the use of their unpublished and valuable data.

ANTONIO SALAS,[1] ANTONIO TORRONI,[2] MARTIN RICHARDS,[3] LLUIS QUINTANA-MURCI,[4] CATHERINE HILL,[5] VINCENT MACAULAY,[6] AND ÁNGEL CARRACEDO[1]

[1]*Unidad de Genética, Instituto de Medicina Legal, Universidad de Santiago de Compostela, A Coruña, Galicia, Spain;* [2]*Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy;* [3]*School of Biology, University of Leeds, Leeds, United Kingdom;* [4]*Centre National de la Recherche Scientifique, Institut Pasteur, Paris;* [5]*Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield, United Kingdom; and* [6]*Department of Statistics, University of Glasgow, Glasgow, United Kingdom*

References

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bortolini MC, Da Silva WA Jr, Zago MA, Elion J, Krishnamoorthy R, Gonçalves VF, Pena SDJ (2004) The phylogeography of mitochondrial DNA haplogroup L3g in Africa and the Atlantic slave trade. Am J Hum Genet 75:523–524 (in this issue)

Phillipson DW (1993) African archaeology. Cambridge University Press, Cambridge

Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo Á (2002) The making of the African mtDNA landscape. Am J Hum Genet 71: 1082–1111

Salas A, Richards M, Lareu M-V, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo Á (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465

Thomas H (1997) The slave trade: the history of the Atlantic slave trade, 1440–1870. Macmillan, London

Vansina J (1995) New linguistic evidence and the "Bantu expansion." J Afr Hist 36:173–195

Address for correspondence and reprints: Dr. Antonio Salas, Unidad de Genética Forense, Universidad de Santiago de Compostela, Galicia, Spain. E-mail: apimlase@usc.es

0002-9297/2004/7503-0022$15.00

Am. J. Hum. Genet. 75:526–528, 2004

Beckwith-Wiedemann Syndrome and IVF: A Case-Control Study

To the Editor:

A recent series of observations has suggested a link between in vitro fertilization (IVF) and imprinting disorders, such as Beckwith-Wiedemann syndrome (BWS [MIM 130650]) and Angelman syndrome (MIM 105830). BWS is a model imprinting disorder and is

Table 1

Clinical Features of Four Patients Diagnosed with BWS Who Were Conceived Using IVF

	Finding in Patient			
Clinical Feature	1	2	3	4
Intracytoplasmic sperm injection	No	No	No	Yes
Frozen embryo	Yes	No	Yes	Yes
Day of transfer	2	2	2	2
Sex	Female	Male	Male	Female
Gestation (wk)	40	33	38	37
Macrosomia	Yes	Yes	Yes	No
Hypoglycemia	No	Yes	No	Yes
Macroglossia	Yes	Yes	Yes	Yes
Ear anomalies	Yes	No	Yes	Yes
Abdominal-wall defects	Exomphalos	No	Exomphalos	No
Hemihypertrophy	No	Yes	No	No
Isolated loss of methylation at *KVDMR1/LIT1*	Yes	Not performed	Yes	Yes

characterized by prenatal and/or postnatal overgrowth, macroglossia, abdominal-wall defects, neonatal hypoglycemia, hemihypertrophy, ear abnormalities, and an increased risk of embryonal tumors (DeBaun et al. 2002). An analysis of BWS registries from three centers has shown the proportion of individuals with BWS conceived using IVF to be 3/65 (DeBaun et al. 2003), 6/149 (Maher et al. 2003), and 6/149 (Gicquel et al. 2003). These data suggest that ~4% of individuals with BWS are conceived using IVF, a figure greater than the generally accepted usage of IVF in these centers. Further interpretation of these results has been limited because of a reliance by these studies on case records and questionnaire data to determine the method of conception in BWS cases, a lack of the use of appropriate controls, and a statistical significance that was either borderline (Gicquel et al. 2003; Maher et al. 2003) or not mentioned (DeBaun et al. 2003). A recent review of the epidemiology and molecular biology behind these and other related studies has highlighted the need for case-control studies in this area (Niemitz and Feinberg 2004). We report here the results of what we believe is the first case-control study done to test the null hypothesis that there is no difference between the rate of IVF in BWS cases and that in non-BWS controls, in an Australian population.

The present study was possible because the State of Victoria, Australia, is serviced by a single clinical genetics service and laboratory providing molecular tests for BWS. This allowed complete ascertainment of children born in Victoria between 1983 and 2003 and diagnosed with BWS by a clinical geneticist. Only cases meeting the DeBaun criteria (DeBaun and Tucker 1998) were included in this study. Appropriate controls were obtained using data from the Victorian Perinatal Data Collection Unit, which registers all births of >19-wk gestation. For each BWS case, four live-born controls were randomly selected from babies born within 1 mo of that case, in which parity was 1 and the maternal age was within 1 year of the risk-set case. Manual record linkage was then used to determine if the BWS cases and the controls were recorded in the databases of the providers of IVF services in Victoria, with the use of maternal names and the dates of birth of mothers and babies. Ethics approval was obtained from all sites providing data. Statistical significance of differences in proportions between groups was assessed using Epi Info, with results expressed as odds ratios (ORs) and as Fisher's-exact-test two-sided P values to account for cell sizes <5.

Among ~1,316,500 live births in Victoria between 1983 and 2003 (2003 data were estimated, as they were known to be very similar to 2002 data), 37 cases of BWS were detected, giving an overall BWS prevalence of ~1/35,580 live births for this period. The average maternal age for BWS cases was 27.0 years. Record linkage of the 37 BWS cases and 148 matched controls identified IVF as the method of conception in 4 BWS cases (10.81%) and in 1 control (0.67%), giving an OR of 17.8 (95% CI 1.8–432.9), and Fisher's-exact-test two-sided $P = .006$. The clinical and molecular features of the four patients with BWS conceived using IVF are listed in table 1, and the reasons for the use of IVF were varied (two unexplained infertility, one egg donation, and one oligospermia). Our results indicate that if a child has BWS, the odds that the child was conceived using IVF is ~18 times greater than that for a child without BWS, although the magnitude of this OR should be cautiously interpreted, given the wide CI. During the study period (1983–2003), 14,894 babies were born as a result of an IVF procedure (excluding gamete intrafallopian transfer). Using our population-based data, we can then estimate the absolute risk of having a live-born baby with BWS when IVF is used as the means of conception to be 4/14,894.

This study demonstrates that children conceived by IVF are significantly more likely to have BWS, compared with children conceived naturally. Our study design with a control group matched by maternal age has ensured that the rate of IVF procedures in the control (non-BWS) population is accurate for the entire study period, which encompasses a time from infrequent use of IVF (0.2% of pregnancies in 1983) to more frequent use (3% in 2003). We can quantify, for the first time, the risk of BWS in our IVF population as ~1/4,000, or 9 times greater than in the general population. The mechanisms underlying this increased risk remain unclear, but this study and previous studies (DeBaun et al. 2003; Gicquel et al. 2003; Maher et al. 2003) have shown that patients with BWS conceived by IVF consistently show isolated hypomethylation at the maternal *KVDMR1/LIT1* locus at 11p15.5. By comparison, this molecular mechanism is observed in only 46% of our overall BWS population, with the remainder of BWS cases resulting from uniparental disomy of chromosome 11 (16%), biparental methylation of *H19DMR* (7%), or an unidentified mutation (31%). The preponderance of BWS cases conceived by IVF that show hypomethylation of maternal *KVDMR1/LIT1* suggests that collection of in vitro cultures might disturb methylation in the oocyte or early embryo, predisposing to maternal allele demethylation.

The fact that the overall risk of BWS in children conceived using IVF remains low and that BWS is, in most cases, associated with a good long-term outcome makes it unlikely that this finding will deter couples from using IVF. Nor does it seem necessary to offer prenatal diagnosis for BWS to couples undergoing IVF. Questions remain, however, about potential effects of IVF on other regions of the genome that are subject to epigenetic regulation. In this context, the observation of a possible association between IVF and Angelman syndrome, another disorder resulting from hypomethylation of the maternal genome, is of some concern (Cox et al. 2002; Orstavik et al. 2003). Although long-term follow-up data of children conceived by IVF are generally reassuring, it remains possible that alterations in genomic imprinting might have other unrecognized health implications for children and adults who were conceived by IVF. Our data reinforce the need for long-term follow-up studies of children conceived by IVF.

JANE HALLIDAY,[1,3,4] KAY OKE,[5] SUE BREHENY,[6] ELIZABETH ALGAR,[1,3] AND DAVID J. AMOR[1,2,3]

[1]*Murdoch Childrens Research Institute and* [2]*Genetic Health Services Victoria, Royal Children's Hospital, and* [3]*Department of Pediatrics, University of Melbourne, Parkville, Australia;* [4]*Victorian Birth Defects Register, Perinatal Data Collection Unit, Department of Human Services, Melbourne;* [5]*Melbourne IVF, East Melbourne, Australia; and* [6]*Monash IVF, Clayton, Australia*

Electronic-Database Information

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for BWS and Angelman syndrome)

References

Cox GF, Bürger J, Lip V, Mau UA, Sperling K, Wu B-L, Horsthemke B (2002) Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet 71: 162–164

DeBaun MR, Niemitz EL, Feinberg AP (2003) Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of *LIT1* and *H19*. Am J Hum Genet 72:156–160

DeBaun MR, Niemitz EL, McNeil DE, Brandenburg SA, Lee MP, Feinberg AP (2002) Epigenetic alterations of *H19* and *LIT1* distinguish patients with Beckwith-Wiedemann syndrome with cancer and birth defects. Am J Hum Genet 70: 604–611

DeBaun MR, Tucker MA (1998) Risk of cancer during the first four years of life in children from The Beckwith-Wiedemann Syndrome Registry. J Pediatr 132:398–400

Gicquel C, Gaston V, Mandelbaum J, Siffroi J-P, Flahault A, Le Bouc Y (2003) In vitro fertilization may increase the risk of Beckwith-Wiedemann syndrome related to the abnormal imprinting of the *KCNQ1OT* gene. Am J Hum Genet 72: 1338–1340

Maher ER, Brueton LA, Bowdin SC, Luharia A, Cooper W, Cole TR, Macdonald F, Sampson JR, Barratt CL, Reik W, Hawkins MM (2003) Beckwith-Wiedemann syndrome and assisted reproduction technology (ART). J Med Genet 40: 62–64

Niemitz EL, Feinberg AP (2004) Epigenetics and assisted reproductive technology: a call for investigation. Am J Hum Genet 74:599–609

Ørstavik KH, Eiklid K, van der Hagen CB, Spetalen S, Kierulf K, Skjeldal O, Buiting K (2003) Another case of imprinting defect in a girl with Angelman syndrome who was conceived by intracytoplasmic sperm injection. Am J Hum Genet 72: 218–219

Address for correspondence and reprints: Associate Professor Jane Halliday, Public Health Genetics, Murdoch Childrens Research Institute, Parkville, 3052, Australia. E-mail: jane.halliday@mcri.edu.au

0002-9297/2004/7503-0023$15.00