Diversity of Two Short Tandem Repeat Loci (CD4 and F13A1) in Three Brazilian Ethnic Groups

Author(s): HEIQUE M. BOGDAWA, MARA H. HUTZ, FRANCISCO M. SALZANO and TANIA A. WEIMER

Source: *Human Biology*, Vol. 72, No. 6 (December 2000), pp. 1045-1053
Published by: Wayne State University Press
Stable URL: http://www.jstor.org/stable/41465902

REFERENCES

*Diversity of Two Short Tandem Repeat Loci (*CD4 *and* F13A1*) in Three Brazilian Ethnic Groups*

HEIQUE M. BOGDAWA,[1] MARA H. HUTZ,[1] FRANCISCO M. SALZANO,[1] AND TANIA A. WEIMER[1]

Abstract Two microsatellites (*CD4* and *F13A1*) were investigated in seven Brazilian populations: one group each of European- and African-derived subjects from Porto Alegre, southern Brazil, and five Amerindian tribes (three Tupi-Mondé speaking [Gavião, Surui, and Zoró], one Macro-Gê [Xavante], and one Carib [Wai-Wai]). For both markers, neo-Brazilians presented with a high diversity, but Amerindians showed a low level of variability. Genotype frequency distributions were heterogeneous among populations, the only exception being similar *CD4* frequencies in Afro- and Euro-Brazilians. Gene diversity analysis revealed that most of the total variation is due to intrapopulational diversity in all populations. Because of the high information content of these markers in Afro- and Euro-Brazilians, these systems are most appropriate for forensic analyses. The comparison among Brazilian and other world populations revealed high similarity among populations of the same ethnic group, indicating a high discriminative power for these markers.

Brazilian groups are interesting and challenging for population genetics investigations due to their ethnic origins and interethnic admixture. Although numerous studies based on protein polymorphisms (for instance, Salzano et al. 1998) and DNA markers (Hutz et al. 1999) have already been performed, only the progressive accumulation of data can help provide a better understanding of the evolutionary history of these groups.

Short tandem repeats (STRs; microsatellites) have been identified in all eukaryote species so far investigated. Because of their relative abundance in the genome and their high degree of polymorphism, STRs have been widely used in genetic diversity studies of human populations (e.g., Bowcock et al. 1994; Gusmão et al. 1997).

[1] Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal, 15053, 91501–970 Porto Alegre, RS, Brazil.

Human Biology, December 2000, v. 72, no. 6, pp. 1045–1053.

KEY WORDS: SOUTH AMERICAN POPULATIONS, AMERINDIANS, SHORT TANDEM REPEATS, MICROSATELLITES

Materials and Methods

The present study provides information about two STR loci (*CD4* and *F13A1*) for three Brazilian ethnic groups: Amerindians from the Central and Amazonian regions, as well as European- and African-derived individuals living in Porto Alegre, in southern Brazil. Gene diversity and the relationships among Amerindian populations as well as among Brazilians and other world populations were considered, and the potential application of these two markers for forensic use was also evaluated. Specific questions asked were: (*a*) Do the results of the European- and African-derived groups agree with those obtained in their putative ancestors? (*b*) How effective are these markers in characterizing populations? and (*c*) Do the Amerindian gene diversity and population relationships obtained with these systems agree with those observed using other polymorphisms?

Porto Alegre (51°10′W, 30°5′S) is the capital of Brazil's southernmost state, Rio Grande do Sul, and has about 1,300,000 inhabitants. Persons of European ancestry in Porto Alegre are mainly descendants of Portuguese settlers, but other Europeans (Italians, Germans, and Spaniards) have also contributed to their gene pool. People of African ancestry constitute approximately 14% of the population and descend from individuals brought to Brazil between the 15th and 18th centuries, mainly from Africa's West Coast, but also from Angola and Mozambique.

The European-derived sample consisted of 61 individuals who came to the Genetics Department of the Federal University of Rio Grande do Sul (UFRGS) for paternity testing. Blood samples from 61 African-derived individuals were obtained at the central laboratory of a general public hospital (Santa Casa de Misericórdia), to which they went for routine blood examinations.

The Amerindian sample consisted of 143 individuals from five Brazilian tribes: three groups speaking a language classified by Rodrigues (1986) as belonging to the Mondé family of the Tupi stock, namely Surui (61°10′W; 10°50′S; $n = 24$), Zoró (60°20′W, 10°20′S; $n = 30$), and Gavião (61°8′W, 10°10′S; $n = 30$); a tribe speaking a Macro-Gê language (Rodrigues 1986; Greenberg 1987), Xavante (51°40′W; 13°20′S; $n = 32$); and a group speaking a language classified as belonging to the Parukoto-Charumã family of the Carib stock (Rodrigues 1986), Wai-Wai (57°55′W, 0°40′S; $n = 27$). For more details about these Amerindian populations, see Callegari-Jacques et al. (1996) and Salzano et al. (1997, 1998).

The Porto Alegre blood samples were refrigerated shortly after collection and immediately stored under appropriate conditions. The Amerindian material was transported by air to Porto Alegre and subsequently processed. DNA was obtained from leukocytes by salting-out procedures (Miller et al. 1988; Lahiri and Nurnberger 1991).

The *CD4* locus consists of a pentanucleotide repeat within a noncoding

region of the T lymphocyte marker *CD4* gene on chromosome 12pter-12p12, while the *F13A1* locus is a tetranucleotide repeat within an intron of the coagulation factor XIII gene at chromosome 6p24–6p25. Polymerase chain reaction (PCR) amplification of both markers was carried out using primers and protocols described by Wall et al. (1993). PCR products were separated by means of electrophoresis through a 10% nondenaturing vertical polyacrylamide gel in a continuous system and stained with ethidium bromide, as suggested by Lahiri et al. (1997). Allele sizes were estimated by comparing PCR products with either an allelic ladder (123 base pairs for *F13A1*) or with size markers (pBR322 DNA/ *Msp*I for both STRs, and pBR322 DNA/ *Hae*III for *CD4*).

Allele frequencies were estimated by gene counting. Heterogeneity among populations was tested according to Roff and Bentzen (1989). Average heterozygosities (*H*) and gene diversity analysis were performed according to Nei (1987) using the NJBAFD software (Takezaki 1999). The gene differentiation coefficient considering the number of subpopulations (G'_{ST}) was calculated as described by Livshits and Nei (1990), while standard errors of this estimate were determined according to Chakraborty (1974). Partitioning of the variance was also performed using the analysis of molecular variance (AMOVA) approach (Michalakis and Excoffier 1996) with the Arlequin software program, version 1.1 (Schneider et al. 1997). D_{SW} genetic distances among the Amerindians as well as among several world populations were obtained (Shriver et al. 1995), and the relationships among these groups were depicted by neighbor-joining (NJ) trees (Saitou and Nei 1987). The reliability of the trees was examined by a bootstrap test (Felsenstein 1985), with 3000 replications as suggested by Hedges (1992). These calculations were also made using the NJBAFD software. The probability of genotypic identity of two individuals chosen at random in a population (GI) was calculated according to Van Zeveren et al. (1990), and the probability of exclusion of a falsely accused man (PE) was estimated using the method of Chakravarti and Li (1983).

Results and Discussion

Allele frequencies and expected average heterozygosities are presented in Table 1. High variability was observed for both markers in the urban Euro- and African-Brazilians. For *CD4* a total of nine alleles were detected; allele **10* occurred exclusively in Afro-Brazilians; and two alleles only were found in almost all Amerindians. For *F13A1A* a total of 11 alleles were observed, three of them being present solely in Afro-Brazilians. Amerindians showed reduced variability, with a greater effect in the Zoró.

Heterogeneity in genotype distributions among populations was observed for both markers, the only exception being the Afro- and Euro-Brazilians, who presented similar *CD4* frequencies (data not shown).

Table 1. Allele Frequencies and Average Heterozygosities (*H*) at the *CD4* and *F13A1* Loci

				Amerindians				
Loci	*Alleles*	*Euro-Brazilians* $n=61$	*Afro-Brazilians* $n=61$	*Suruí* $n=24$	*Zoró* $n=30$	*Gavião* $n=30$	*Wai-Wai* $n=27$	*Xavante* $n=32$
CD4	*8	0.37	0.25	0.56	0.77	0.62	0.59	0.81
	*9	0.28	0.21			0.05		
	*10		0.01					
	*11	0.02	0.12					
	*12	0.02	0.02					
	*13	0.26	0.17	0.44	0.23	0.33	0.41	0.19
	*14	0.03	0.15					
	*15	0.01	0.05					
	*16	0.01	0.02					
	H	0.72	0.83	0.50	0.36	0.51	0.49	0.31
F13A1	*4	0.09	0.01	0.11		0.25	0.02	0.05
	*5	0.20	0.13	0.55	0.57	0.46	0.20	0.25
	*6	0.21	0.42	0.34	0.43	0.10	0.50	0.45
	*7	0.21	0.12			0.17		0.16
	*8	0.25	0.19			0.02	0.14	0.09
	*9	0.01	0.05				0.14	
	*10		0.01					
	*11		0.02					
	*15		0.02					
	*16	0.01	0.01					
	*19	0.02	0.02					
	H	0.80	0.75	0.58	0.50	0.70	0.68	0.72

n = sample sizes.

Table 2. Gene Diversity Analysis for Seven Brazilian Populations

Levels of Analysis	*No. of Subpopulations*	*No. of Loci*	H_T	H_S	D_{ST}	*Dm*	$G'_{ST} \pm SE$
Level 1							
All populations	7	2	0.661	0.595	0.066	0.077	0.115 ± 0.005
Level 2							
All Amerindians	5	2	0.567	0.526	0.041	0.051	0.088 ± 0.027
Level 3							
Tupi-Mondé–speaking tribes	3	2	0.543	0.516	0.027	0.040	0.067 ± 0.019

H_T = total genetic variation; H_S = intrapopulational heterozygosity; D_{ST} = gene differentiation among populations; *Dm* = average minimum genetic distance among subpopulations; G'_{ST} = coefficient of gene differentiation considering the number of populations examined (Nei 1987; Livshits and Nei 1990); SE = standard error. Essentially the same results were obtained using the AMOVA methodology (data not shown).

The degree of heterozygosity found in Afro- and Euro-Brazilians is of the same order of magnitude as that found in other populations of the same ethnic group (Edwards et al. 1991; Wall et al. 1993; Hammond et al. 1994; Pastore et al. 1996; Gusmão et al. 1997). The low diversity observed in Amerindians as compared to Europeans and Africans was already observed for other markers in studies of the same tribes (Bortolini et al. 1998; Hutz et al. 1999), but in relation to the systems studied here it is similar to those found in Asians (Wall et al. 1993; Hammond et al. 1994). Therefore, the distributions of these markers suggest the absence of an important bottleneck in the early colonization of South America, in agreement with the findings of Bevilaqua et al. (1995). The lower level of diversity found in the Zoró was also observed for this population in relation to *D1S80* (Heidrich et al. 1995) and *HB* haplotypes (Bevilaqua et al. 1995), but not for mtDNA or proteins (Bortolini et al. 1998) APO B, DRD2, and LPL (Hutz et al. 1999). Therefore, genetic drift cannot, in isolation, explain the present findings.

Table 2 presents three levels of gene diversity analysis for the seven Brazilian populations. The total gene diversity (H_T) value for the Tupi-Mondé–speaking tribes (54%) was lower than the values for the other population groupings (57% and 66%, respectively). By comparing H_S and H_T it can be seen that most of the total diversity can be attributed to the intrapopulational variability at all population levels. The gene differentiation coefficient was larger for level 1 (G'_{ST} = 11%), but standard errors of these estimates are high. Similar gene diversity coefficients for these Amerindians were obtained by Bevilaqua et al. (1995). The small G'_{ST} value observed for the Tupi-Mondé Indians was expected, since they live nearby, and G'_{ST} is generally correlated with geographic distances (Livshits and Nei 1990).

Not much information is available for populations in which these two systems have been studied simultaneously. We have been able to locate five

Table 3. Genotypic Identity and Probability of Exclusion Based on *CD4* and *F13A1* (× 100)

	Genotypic Identity			*Probability of Exclusion*		
Population	*CD4*	*F13A1*	*Combined*	*CD4*	*F13A1*	*Combined*
Euro-Brazilians	14.1	7.1	1.0	46.8	60.3	78.9
Afro-Brazilians	5.7	9.4	0.5	64.6	54.8	83.9
Suruí	37.8	26.9	10.2	18.0	28.2	41.1
Zoró	47.7	38.1	18.2	15.0	18.9	31.1
Gavião	33.1	14.4	4.8	23.1	43.5	56.6
Wai-Wai	38.4	15.4	5.9	18.2	43.0	53.8
Xavante	53.6	13.4	7.2	13.0	46.5	53.5

other samples, and they were considered together with ours in the NJ tree shown in Figure 1. Clustering was as expected, with populations of the same ethnic derivation grouping together. Amerindians clustered together with Asians, the Wai-Wai showing an intermediate position, near the Whites. This could be due to interethnic admixture, although no signs of admixture were detected based on a study of 27 protein systems (Callegari-Jacques et al. 1996). These Indians, however, have been in contact with non-Indians for at least 100 years (Hutz et al. 1999).

Similar analyses using the two markers in isolation could be made, including a larger number of populations; in these cases *CD4* furnished a similar pattern to that presented in Figure 1, but *F13A1* did not show a discriminative power, no clear cluster being observed (data not shown).

To verify the usefulness of the STRs investigated here for forensic analyses in Brazilian populations, two estimates were made and are presented in Table 3. Although some differences were observed considering each locus separately, Afro-Brazilians showed less genotypic identity (GI = 0.5%) and higher values of exclusion probability (PE = 84%). Due to lower variability, Amerindians as a whole presented a higher individual identity and lower probability of exclusion, the Zoró having the most extreme values (GI = 18%, PE = 31%). Of the two loci, *CD4* was more informative for Porto Alegre's African-derived population, but *F13A1* was more informative for the other populations. These results indicate the importance of these markers for paternity determinations. Using a set of 18 protein markers Robinson et al. (1995) estimated an exclusion probability of 89% in the Euro-Brazilians of this same population. Therefore, using just two STRs it is possible to obtain a similar result with a more effective cost-benefit ratio. The usefulness of these markers in this context among the Indians is, however, lower.

Acknowledgments Thanks are due to the Fundação Nacional do Índio (FUNAI) for permission to study the Indians and for logistic assistance. The Indian leaders and

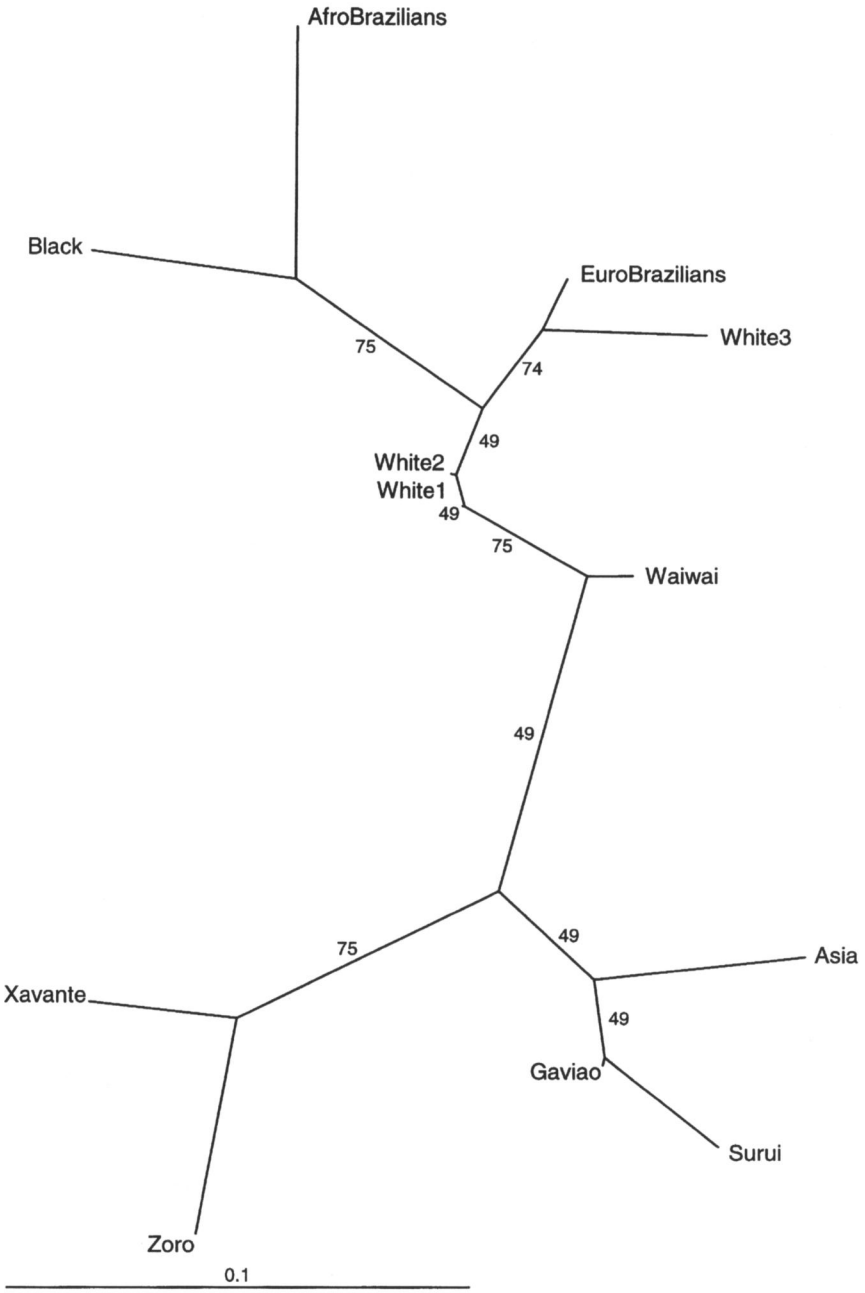

Figure 1. Neighbor-joining unrooted tree obtained from D_{SW} distances for 12 populations for *CD4* and *F13A1*. Asia: US Asians; Black: US blacks; and White 1: US whites (Hammond et al. 1994). White 2: whites from Italy (Pastore et al. 1996). White 3: northern European whites (Wall et al. 1993). The other samples are those reported here. Numbers indicate bootstrap values.

the subjects of the investigation were adequately informed about the aims of the study and gave their approval, which is also acknowledged. We are very grateful to Ricardo V. Santos and Carlos E. A. Coimbra Jr. for the collection of four of the Amerindian samples. Financial support was provided by Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (FINEP), and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

Received 7 July 1999; revision received 27 March 2000.

Literature Cited

Bevilaqua, L.R.M., V.S. Mattevi, G.M. Ewald et al. 1995. Beta-globin gene cluster haplotype distribution in five Brazilian Indian tribes. *Am. J. Phys. Anthropol.* 98:395–401.

Bortolini, M.C., C. Baptista, S.M. Callegari-Jacques et al. 1998. Diversity in protein, nuclear DNA, and mtDNA in South Amerinds—agreement or discrepancy? *Ann. Hum. Genet.* 62:133–145.

Bowcock, A.M., A. Ruiz-Linares, J. Tomfohrde et al. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457.

Callegari-Jacques, S.M., F.M. Salzano, T.A. Weimer et al. 1996. The Wai Wai Indians of South America: History and genetics. *Ann. Hum. Biol.* 23:189–201.

Chakraborty, R. 1974. A note on Nei's measure of gene diversity in a substructured population. *Humangenetik* 21:85–88.

Chakravarti, A., and C.C. Li. 1983. The effect of linkage on paternity testing. In *Inclusion Probabilities in Parentage Testing*, R. H. Walker, ed. Arlington, VA: Amer. Assoc. Blood Banks, 411–422.

Edwards, M.C., P.R. Clemens, M. Tristan et al. 1991. Pentanucleotide repeat length polymorphism at the human CD4 locus. *Nucl. Acids Res.* 19:4791.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.

Greenberg, J.H. 1987. *Language in the Americas.* Stanford, CA: Stanford University Press.

Gusmão, L., M. J. Prata, A. Amorim et al. 1997. Characterization of four short tandem repeat loci (TH01, VWA31/A, CD4, and TP53) in Northern Portugal. *Hum. Biol.* 69:31–40.

Hammond, H.A., L. Jin, Y. Zhong et al. 1994. Evaluation of 13 short tandem repeats for use in personal identification applications. *Am. J. Hum. Genet.* 55:175–189.

Hedges, S.B. 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol. Biol. Evol.* 9:366–369.

Heidrich, E.M., M.H. Hutz, F.M. Salzano et al. 1995. D1S80 locus variability in three Brazilian ethnic groups. *Hum. Biol.* 67:311–319.

Hutz, M.H., S.M. Callegari-Jacques, M.C. Bortolini et al. 1999. Variability in nDNA, mtDNA, and proteins—a test case. In: *Genomic Diversity: Applications in Human Population Genetics*, S.S. Papiha, and R. Deka, eds. New York, NY: Plenum (in press).

Lahiri, D.K., and J.I. Nurnberger Jr. 1991. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucl. Acids Res.* 19:5444.

Lahiri, D.K., A. Zhang, and J.I. Nurnberger Jr. 1997. High-resolution of PCR products from a microsatellite marker using a nonradioisotopic technique. *Biochem. Mol. Med.* 60:70–75.

Livshits, G., and M. Nei. 1990. Relationships between intrapopulational and interpopulational genetic diversity in man. *Ann. Hum. Biol.* 17:501–513.

Michalakis, Y., and L. Excoffier. 1996. A genetic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142:1061–1064.

Miller, S.A., D.D. Dykes, and H.F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl. Acids Res.* 16:1215.

Nei, M. 1987. *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.

Pastore, L., E. Vuttariello, C. Sarrantonio et al. 1996. Allele frequency distributions at several variable number of tandem repeat (VNTR) and short tandem repeat (STR) loci in a restricted Caucasian population from South Italy and their evaluation for paternity and forensic use. *Molec. Cell. Probes* 10:299–308.

Robinson, W.M., A.C. Hickmann, T.A. Weimer et al. 1995. Protein genetics and paternity determinations in southern Brazil. *Ind. J. Hum. Genet.* 1:105–110.

Rodrigues, A.D. 1986. *Linguas Brasileiras. Para o Conhecimento das Linguas Indigenas.* São Paulo, Brazil: Edições Loyola.

Roff, D.A., and P. Bentzen. 1989. The statistical analysis of mitochondrial DNA polymorphism: (2 and the problem of small samples. *Mol. Biol. Evol.* 6:539–545.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.

Salzano, F.M., M.H.L.P. Franco, T.A. Weimer et al. 1997. The Brazilian Xavante Indians revisited: New protein genetic studies. *Am. J. Phys. Anthropol.* 104:23–34.

Salzano, F.M., T.A. Weimer, M.H.L.P. Franco et al. 1998. Protein genetic studies among the Tupi-Mondé Indians of the Brazilian Amazon. *Am. J. Hum. Biol.* 10:711–722.

Schneider, S., J.M. Kueffer, D. Roessli et al. *ARLEQUIN* (*version 1.1*) Geneva, Switzerland: University of Geneva.

Shriver, M.D., L. Jin, E. Boerwinkle et al. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* 12:914–920.

Takezaki, N. 1999. NJBAFD: Neighbor-joining tree construction from allele frequency data. [On-line.] Available: ftp://ftp.nig.ac.jp/pub/Bio/njbafd/dos/

Van Zeveren, A., Y. Bouquet, A. van de Weghe et al. 1990. A genetic blood marker study on 4 pig breeds. *J. Anim. Breed. Genet.* 107:104–112.

Wall, W.J., R. Willianson, M. Petron et al. 1993. Variation of short tandem repeats within and between populations. *Hum. Molec. Genet.* 2:1023–1029.