Brief Communication: Patterns of Linkage Disequilibrium and Haplotype Diversity at Xq13 in Six Native American Populations

Sijia Wang,[1]* Gabriel Bedoya,[2] Damian Labuda,[3] and Andres Ruiz-Linares[1]*

[1]*Department of Genetics, Evolution and Environment, University College London, 4 Stephenson Way, London NW1 2HE, UK*
[2]*Laboratorio de Genética Molecular, Universidad de Antioquia, Medellín, Colombia*
[3]*CHU Sainte-Justine, Département de Pédiatrie, Université de Montréal, Montréal, PQ, Canada*

KEY WORDS linkage disequilibrium; Xq13; Native Americans; haplotype

ABSTRACT Comparative studies of linkage disequilibrium (LD) can provide insights into human demographic history. Here, we characterize LD in six Native American populations using seven microsatellite markers in Xq13, a region of the genome extensively studied in populations around the world. Native Americans show relatively low diversity and high LD, in agreement with recent genome-wide survey and a scenario of sequential founder effects accompanying human population dispersal around the globe. LD in Native Americans is similar to that observed in some recently described small population isolates and higher than in large European isolates (e.g., Finns), which have been extensively analyzed in medical genetics studies. Haplotype analyses are consistent with a colonization of the New World by a differentiated East Asian population, followed by extensive genetic drift in the Americas. Am J Phys Anthropol 142:476–480, 2010.

Patterns of linkage disequilibrium (LD) across the genome are influenced by a range of factors, including variable mutation and recombination rates, natural selection, and population demography (Ardlie et al., 2002). Genome-wide comparisons of LD in different human populations have been carried out in the CEPH-HGDP panel and the HapMap reference set. Other extensive population surveys have been performed for a few regions of the genome, including a ~13 Mb segment on Xq13, which has been examined in a range of populations across the world (Laan and Paabo, 1997; Zavattari et al., 2000; Angius et al., 2001; Kaessmann et al., 2002; Katoh et al., 2002; Latini et al., 2004; Laan et al., 2005; Marroni et al., 2006; Branco et al., 2008; Bellis et al., 2008, Leite et al., 2009). So far, comparative studies of LD including Native Americans are fairly scant, often limited to the five populations of the CEPH-HGDP panel (Sawyer et al., 2005; Conrad et al., 2006; Jakobsson et al., 2008; Li et al., 2008; Bosch et al., 2009). To further the analysis of LD in Native Americans here we examine the Xq13 region, previously studied around the world, in six Native American populations.

MATERIALS AND METHODS

Samples

DNA samples (isolated from peripheral blood) were obtained from consenting individuals representing six Native American populations: Wayuu (n = 66 chromosomes), Ingano (n = 38), Kogi (n = 44), Zenu (n = 46), and Ticuna (n = 30), from Colombia, and Cree (n = 25) from Saskatchewan, Canada. Following the linguistic classification of Ruhlen (1991), Wayuu and Ticuna both belong to the Equatorial-Tucanoan linguistic stock. Wayuu is one of the largest Native American groups in Colombia, with an estimated population size of 135,000, whereas the Ticuna have a population size of ~8,000. The Ingano is an Andean population, with a population size of ~18,000. Kogi and Zenu are Chibchan-Paezan populations, with estimated population sizes of 3,000 and 34,000, respectively. See Mesa et al. (2000) for more information of the Native Americans in Colombia. Cree belongs to a large population (~200,000 in Canada) organized into many smaller groups. The Cree in Saskatchewan have a census of roughly 73,500.

Our genotyping data were combined with published datasets using the same markers on seven East Asian (Katoh et al., 2002; Laan et al., 2005) (Buriat, n = 78; Evenki, n = 71; Japanese, n = 100, Khalkha, n = 83; Khoton, n = 40; Uriankhai, n = 55; and Zahkchin, n = 59), five Volga-Ural (Laan et al., 2005) (Chuvashi, n = 40; Komi, n = 46; Mari, n = 44; Mordva, n = 48; and Udmurt, n = 49), and eight Western European populations (Laan and Paabo, 1997; Zavattari et al., 2000; Laan et al., 2005) (Dutch, n = 70; Estonian, n = 45; Finnish, n = 80; German, n = 41; Italian, n = 92; Russian, n = 66; Saami, n = 54; and Swedish, n = 41). See Supporting Information Table 1 for census size for all 26 populations.

Additional supporting information may be found in the online version of this article.

*Correspondence to: Sijia Wang, FAS Center for Systems Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138. E-mail: swang@oeb.harvard.edu; Andres Ruiz-Linares, Department of Genetics, Evolution and Environment, University College London, 4 Stephenson Way, London NW1 2HE, UK. E-mail: a.ruizlin@ucl.ac.uk

Received 22 July 2009; accepted 23 October 2009

DOI 10.1002/ajpa.21234
Published online 23 December 2009 in Wiley InterScience (www.interscience.wiley.com).

Genotyping

We studied seven microsatellite markers at Xq13: DXS983, DXS8037, DXS8092, DXS1225, DXS8082, DXS986, and DXS995, spanning more than 13 Mb, from physical map position 69.36–82.64 Mb (GenBank Build 36.2) and about 3.4 cM, from genetic map position 83.93–87.29 cM (Kong et al., 2002). Microsatellites were typed on ABI PRISM 377 DNA analyzer using PCR products obtained as described by Laan and Paabo (1997) and data processing by GENESCAN version 3.1 and GENOTYPER version 2.5. The missing data rate is 2%.

Statistical analyses

Gene diversities were computed using Arlequin 2.0 (Schneider et al., 2000). Extracting full information of the phased male samples and unphased female samples, GENECOUNTING (Zhao, 2004) was used to obtain maximum-likelihood estimate of haplotype frequencies. Pair-wise LD was assessed using a Monte Carlo approximation to Fisher's exact test with the POWERMARKER 3.0 program (Liu and Muse, 2005). A randomized sampling correction was used to avoid a bias due to differences in sample size. Multilocus LD was estimated with the r_d statistic using the MULTILOCUS program (Agapow and Burt, 2001). A matrix of Nei's D_A distances (Nei et al., 1983) between populations was obtained from two-locus (DXS1225-DXS8082) haplotype frequencies using PowerMarker 3.0, and the results displayed by multidimensional scaling (MDS) using the SPSS package 12.0.1.

RESULTS

Gene diversity and LD

Native Americans show Xq13 microsatellite gene diversities that are mostly lower than in Eurasian populations, ranging between 0.325–0.620 and 0.594–0.755, respectively (Supporting Information Table 2). Considering each region as a single group, gene diversity is lower in Native Americans (0.638) than in East Asians (0.682), Volga-Ural populations (0.754), and Europeans (0.729).

On average, 57.3% marker pairs (12 of 21) show significant LD across Native American populations (Fig. 1 and Supporting Information Table 3), a considerably higher proportion than observed in non-isolated Eurasian populations, where 14.3% marker pairs (3 of 21) are in significant LD. The increased pair-wise LD in Native Americans is comparable to that reported for some small isolated Eurasian populations, such as the Saami and Khoton. A similar pattern is observed for multilocus LD (see Fig. 1), Native Americans averaging an r_d of 0.172 compared to an average of 0.025 in European and Asian populations. Again, only the Saami and Khoton have values of r_d comparable to those observed in Native Americans (0.14 and 0.12, respectively). There is a significant negative correlation between the logarithm transformation of population size and LD, measured by multilocus r_d ($r = -0.616$, $P < 0.01$), or by proportion of significant LD pairs ($r = -0.664$, $P < 0.01$). There is no significant difference in the proportion of LD pairs between Cree from Canada and the other five populations from Colombia (two-tailed t-test: $P = 0.205$).

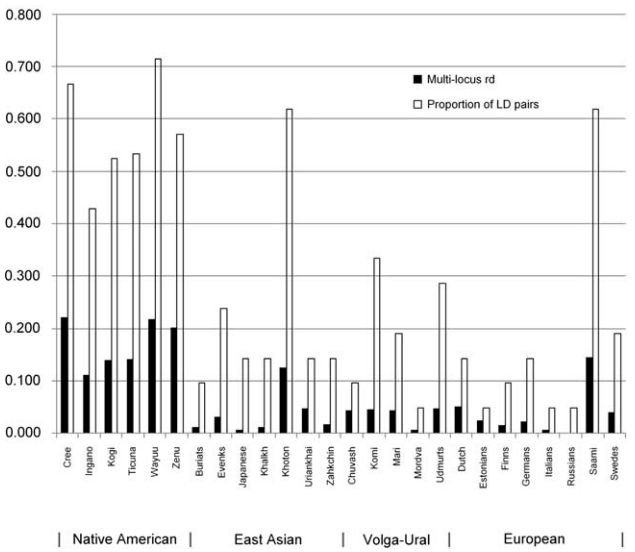

Fig. 1. LD evaluated by the proportion of marker pairs in significant LD ($P < 0.05$ using Fisher's exact test) and multilocus r_d in 26 populations, ordered from left to right in geographic groups: Native American, East Asian, Volga-Ural, European.

Haplotype diversity at DXS1225-DXS8082

Very strong LD has been observed between markers DXS1225 and DXS8082 (located 162 kb apart), in populations from around the world (Supporting Information Table 3). The haplotype frequency distribution for these two markers in Native Americans and other continental groups is shown in Table 1. The three most common haplotypes in Native Americans (defined using allele sizes) are 198–225, 198–227, and 202–221. These three haplotypes are found at elevated frequencies in the Kogi. Two of them predominate in the Wayuu (198–225 at 33% and 202–221 at 17%), Ingano (198–227 at 37% and 202–221 at 32%), and Zenu (198–225 at 39% and 198–227 at 22%). One haplotype is markedly prevalent in the Cree (198–227 at 60%) and the Ticuna (202–221 at 80%). An important differentiation in haplotype frequency is seen between continental groups. Haplotype 198–225 is relatively common in East Asians (~12%) but, of the other two common Native American haplotypes, 198–227 is rare (<6%) outside of the Americas, and 202–221 has very low frequency in East Asians and is absent from Volga-Urals and Europeans. Conversely, the most common haplotype in East Asia (202–217 with a frequency of 25%) and two prominent Volga-Ural and European haplotypes (202–211 and 210–219 with frequencies of 10–28%) are rare or absent in Amerindian populations. These two most common European haplotypes are present at low frequencies in the Wayuu, Ingano, and Zenu. This likely reflects a low level of non-native admixture in these populations, as observed in a larger dataset (Wang et al., 2007).

MDS of a distance matrix calculated from the DXS1225-DXS8082 haplotype frequencies (see Fig. 2) shows three main clusters—Europeans, East Asians, and Native Americans—corresponding to continental populations examined, with Volga-Urals occupying an intermediate position between Europeans and East Asians. Europeans cluster together, separately from Volga-Ural populations, with the exception of the Mari. Russians and Saami are closer to the remaining Volga-Urals than

TABLE 1. Common haplotypes of DXS1225-DXS8082 marker pair

DXS1225-DXS8082	Cree	Ticuna	Wayuu	Kogi	Ingano	Zenu	NA	EA	VU	EU
192–227	*							*	0.06	*
192–229								*	0.06	0.07
198–219	*		*				*			*
198–221			**0.13**		*	**0.11**	0.06	*	*	*
198–223						**0.13**	*		*	*
198–225	*	**0.1**	**0.33**	**0.41**		**0.39**	**0.25**	**0.12**	*	*
198–227	**0.6**		0.06	**0.2**	**0.37**	**0.22**	**0.21**	0.05	0.06	*
198–229	*			**0.16**			*	*	*	0.08
200–221					*					*
200–225								0.08		
200–229								*		
202–209			*				*			
202–211			*		*		*	*	**0.11**	**0.1**
202–217								**0.25**	0.08	*
202–219	*	0.07	*				*	0.06		
202–221	0.08	**0.8**	**0.17**	**0.18**	**0.32**	0.07	**0.24**	*		
202–223			0.08				*			
202–225		*	*			*	*	*		
202–227					0.05	*	*	*		
202–229				*				*		
206–217					**0.11**		*			
206–219								*	0.08	0.05
210–219			*		0.05	*	*	0.05	**0.22**	**0.28**
212–219	0.08		0.09	*			*	*	*	*
214–219			*				*			
216–219	*								*	*

Haplotypes with frequency >0.1 are in bold. The most common haplotype in each population or group is underlined. Haplotype frequency between 0.005 and 0.05 is indicated as *. NA, Native American; EA, East Asian; VU, Volga-Ural; EU, European.

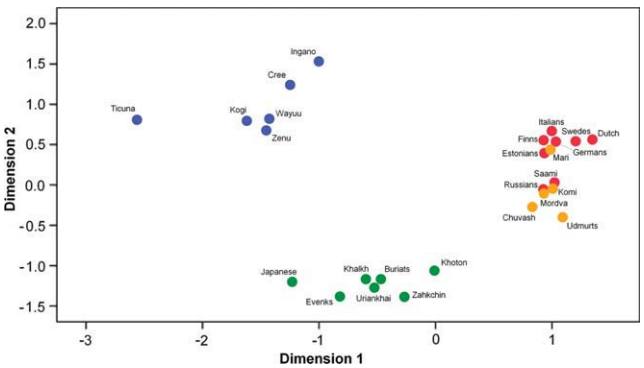

Fig. 2. Multidimensional scaling on Nei's D_A distance matrix derived from frequency of haplotypes at markers DXS1225 and DXS8082. Native American populations are shown in blue, East Asian in green, Volga-Ural in yellow, and European in red. RSQ = 0.930.

to other European populations. Native American populations display the highest within-group distances, whereas Europeans and Volga-Ural populations form tighter clusters. This reflects the considerable variation in haplotype frequencies across native populations, resulting in substantially a higher F_{ST} amongst Native Americans than amongst populations from other regions (0.17 vs. 0.02–0.04, respectively).

DISCUSSION

A low-genetic diversity and high LD at Xq13 was observed in all the Native American populations examined here, with increased LD being apparent both in two-locus and mutilocus analyses. It is worth noting though that results from using different genetic markers could lead to different conclusions (Sawyer et al., 2005). Studies with genome-wide coverage are therefore needed to verify the findings. Our observations are consistent with previous genome-wide surveys, indicating that Native Americans have lower diversity and higher LD relative to other continental regions (Conrad et al., 2006; Jakobsson et al., 2008; Li et al., 2008). These patterns have been interpreted as resulting from sequential bottleneck effects during the dispersal of human populations around the world with entry into the Americas representing the last of these founder events (Prugnolle et al., 2005; Ramachandran et al., 2005; Wang et al., 2007). The population contraction at the colonization of the American continent appears to have been quite substantial, with recent estimates, suggesting that as few as ~100 individuals could have been the initial colonizers (Ray et al., 2009). Our results suggests that the increased LD in Native American populations is comparable to that seen in some small population isolates described in other parts of the world, such as the Saami, and considerably higher than in larger isolates, such as the Finns, which have been extensively examined in medical genetics studies. Interestingly, gene diversity in Native Americans is often considerably lower than in those isolates, suggesting that Native American populations could provide further advantages for trait gene identification (Terwilliger et al., 1998; Peltonen et al., 2000).

Our analysis of haplotypes at markers DXS1225-DXS8082 demonstrates the considerable informativeness of this region for exploring the relatedness of human populations. It is well established that the Americas were colonized by individuals migrating from Asia across Beringia [reviewed by Goebel et al. (2008)], and this is reflected in the relatively close-genetic relatedness of these populations (Wang et al., 2007). Furthermore, the population that colonized the New World seems to have undergone some differentiation from other Asian populations, before its dispersal throughout the Americas, as

evidenced by the occurrence of genetic variants shared by populations across the Americas that are not observed in Asia (Neel, 1978; Wang et al., 2007; Bourgeious et al., 2009; Schroeder et al., 2009). This overall picture is consistent with the haplotype analysis at markers DXS1225-DXS8082. There is evidence of shared ancestry with Asia (haplotype 198–225), loss of diversity in the Americas (including the loss of East Asian haplotype 202–217), and the presence of American-specific haplotypes shared by native populations from Canada to South America (haplotypes 198–227 and 202–221). MDS further illustrates this overall picture with Native Americans appearing closer to East Asians, Volga-Ural populations occupying an intermediate position between East Asians and Europeans; consistent with their geographic location and possibly reflecting genetic influences from both neighboring regions (see Fig. 2). The greater spread of Native Americans on this plot, in comparison with the other three population clusters, reflects the relatively important variation in haplotype frequency between Native American populations. This is consistent with genome-wide surveys documenting the relatively large differentiation in allele frequencies between populations across the Americas, possibly as a result of extensive genetic drift during the process of human dispersal in the continent, which was probably followed by substantial population isolation.

ACKNOWLEDGMENTS

This work was partly supported by grants from Colciencias (1115-04-16471) and Universidad de Antioquia (Sostenibilidad 2009–2010). SW acknowledges support of a K.C. Wong Scholarship and a UK Overseas Research Studentship. DL acknowledges support of the Canadian Institute of Health Research. We thank Maris Laan for sharing published genotype data.

LITERATURE CITED

Agapow PM, Burt A. 2001. Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1:101–102.

Angius A, Melis PM, Morelli L, Petretto E, Casu G, Maestrale GB, Fraumene C, Bebbere D, Forabosco P, Pirastu M. 2001. Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. Hum Genet 109:198–209.

Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299–309.

Bellis C, Cox HC, Ovcaric M, Begley KN, Lea RA, Quinlan S, Burgner D, Heath SC, Blangero J, Griffiths LR. 2008. Linkage disequilibrium analysis in the genetically isolated Norfolk Island population. Heredity 100:366–373.

Bosch E, Laayouni H, Morcillo-Suarez C, Casals F, Moreno-Estrada A, Ferrer-Admetlla A, Gardner M, Rosa A, Navarro A, Comas D, Graffelman J, Calafell F, Bertranpetit J. 2009. Decay of linkage disequilibrium within genes across HGDP-CEPH human samples: most population isolates do not show increased LD. BMC Genom 10:338.

Bourgeois S, Yotova V, Wang S, Bourtoumieu S, Moreau C, Michalski R, Moisan JP, Hill K, Hurtado AM, Ruiz-Linares A, Labuda D. 2009. X-chromosome lineages and the settlement of the Americas. Am J Phys Anthropol 140:417–428.

Branco CC, Cabrol E, Sao BM, Gomes CT, Cabral R, Vicente AM, Pacheco PR, Mota-Vieira L. 2008. Evaluation of linkage disequilibrium on the Xq13.3 region: comparison between the Azores islands and mainland Portugal. Am J Hum Biol 20:364–366.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK. 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat Genet 38:1251–1260.

Goebel T, Waters MR, O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319: 1497–1502.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de LJ, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451:998–1003.

Kaessmann H, Zollner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, Uhlen M, Paabo S. 2002. Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet 70:673–685.

Katoh T, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanishi T, Inoko H, Tamiya G. 2002a. Genetic isolates in East Asia: a study of linkage disequilibrium in the X chromosome. Am J Hum Genet 71:395–400.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. 2002. A high-resolution recombination map of the human genome. Nat Genet 31:241–247.

Laan M, Paabo S. 1997a. Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435–438.

Laan M, Wiebe V, Khusnutdinova E, Remm M, Paabo S. 2005. X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. Eur J Hum Genet 13:452–462.

Latini V, Sole G, Doratiotto S, Poddie D, Memmi M, Varesi L, Vona G, Cao A, Ristaldi MS. 2004. Genetic isolates in Corsica (France): linkage disequilibrium extension analysis on the Xq13 region. Eur J Hum Genet 12:613–619.

Leite FP, Santos SE, Rodriguez EM, Callegari-Jacques SM, Demarchi DA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH. 2009. Linkage disequilibrium patterns and genetic structure of Amerindian and non-Amerindian Brazilian populations revealed by long-range X-STR markers. Am J Phys Anthropol 139:404–412.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.

Liu K, Muse SV. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128–2129.

Marroni F, Pichler I, De Grandi A, Beu VC, Vogl FD, Pinggera GK, Bailey-Wilson JE, Pramstaller PP. 2006. Population isolates in South Tyrol and their value for genetic dissection of complex diseases. Ann Hum Genet 70:812–821.

Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A. 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. Am J Hum Genet 67:1277–1286.

Neel JV. 1978. Rare variants, private polymorphisms, and locus heterozygosity in Amerindian populations. Am J Hum Genet 30:465–490.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153–170.

Peltonen L, Palotie A, Lange K. 2000. Use of population isolates for mapping complex traits. Nat Rev Genet 1:182–190.

Prugnolle F, Manica A, Balloux F. 2005. Geography predicts neutral genetic diversity of human populations. Curr Biol 15:R159–R160.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci USA 102:15942–15947.

Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A, Excoffier L. 2009. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. Mol Biol Evol msp238.

Ruhlen M. 1991. A guide to the World's languages. Stanford, CA: Stanford University Press.

Sawyer SL, Mukherjee N, Pakstis AJ, Feuk L, Kidd JR, Brookes AJ, Kidd KK. 2005. Linkage disequilibrium patterns vary substantially among populations. Eur J Hum Genet 13:677–686.

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

Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI, Wall JD, Pritchard JK, Malhi RS, Smith DG, Rosenberg NA. 2009. Haplotypic background of a private allele at high frequency in the Americas. Mol Biol Evol 26:995–1016.

Terwilliger JD, Zollner S, Laan M, Paabo S. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. Hum Hered 48:138–154.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in native Americans. PLoS Genet 3:e185.

Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, Eaves I, Mastio G, Todd JA, Cucca F. 2000a. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. Hum Mol Genet 9:2947–2957.

Zhao JH. 2004. 2LD. GENECOUNTING and HAP: computer programs for linkage disequilibrium analysis. Bioinformatics 20:1325–1326.