GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Short Communication

Significance of linkage disequilibrium heterogeneous patterns in the 21q22.3 region for mapping 21 trisomy individuals

Y. Valle[1], J.R. Padilla-Gutiérrez[1], A. Quintero-Ramos[2], I.J. García-González[3] and F. Rivas[4]

[1]Departamento de Clínicas Médicas,
Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Mexico
[2]Departamento de Fisiología, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Mexico
[3]Doctorado en Genética Humana, CUCS, Universidad de Guadalajara, Guadalajara, Mexico
Hospital General de Occidente, Secretaría de Salud Jalisco, Guadalajara, Mexico

Corresponding author: J.R. Padilla-Gutiérrez
E-mail: imey_99@yahoo.com

Genet. Mol. Res. 12 (3): 2821-2828 (2013)
Received November 14, 2012
Accepted March 15, 2013
Published August 8, 2013
DOI http://dx.doi.org/10.4238/2013.August.8.2

ABSTRACT. Recombination patterns can be indirectly inferred by means of linkage disequilibrium (LD) estimates, since LD is negatively correlated with genetic distance. However, LD does not necessarily have absolute correspondence with genetic distance. We estimated LD at 5 loci located in the 21q22.3 region. These STRs (D21S1440, D21S168, D21S1260, D21S1446, and D21S1411) covered 8.81 Mb of the 21q22.3 region. They were genotyped by conventional PCR. Similar size samples previously validated by sequencing were used as a genotyping control. Three hundred and sixty-nine individuals (62 families) living in Guadalajara, Mexico, were included. As an inclusion criterion,

each family had a positive paternity test by autosomal markers for the CODIS core loci. Two hundred and thirty phase known haplotypes were identified by familial segregation. Only those haplotypes whose frequency was higher than 4% were taken into account for LD estimation, expressed as Lewontin's D' coefficient and Bonferroni's correction P values. For all 5 loci, the genetic distributions were in agreement with Hardy-Weinberg expectations. Heterozygosity and haplotype diversity were ≥0.69 and 99.58%, respectively. D21S1440-D21S168 (4.51 cM) and D21S1446-D21S1411 (4.58 cM) marker haplotype frequencies were significantly different from those expected by random distribution. The remaining haplotypes, including those with minimal inter-distance (D21S1260-D21S1446, 1.44 Mb), did not show LD. The 5 STRs at the 21q22.3 region in this Mexican population showed a non-homogeneous LD pattern, which demonstrates that recombination or linkage should not be assumed solely on the basis of genetic distance.

Key words: Genetic distance; Linkage disequilibrium; Recombination; Short tandem repeats; 21q22.3 region

INTRODUCTION

It is well known that recombination pattern varies according to gender, genomic regions and populations, among other factors. This is true for the chromosome 21q region, where diverse recombination rates are observed and which contains the Down syndrome critical region (DSCR), located from 21q22.2 to 21q22.3 (Antonarakis et al., 2004). In addition to be related to Down syndrome, DSCR is of interest as some genes within this region have been associated with cancer (Hwang et al., 2008), diabetes (Concannon et al., 2008), and other disorders.

Recombination patterns may be inferred by the estimation of linkage disequilibrium (LD), since LD is negatively correlated with genetic distance. Because of the existence of hot spots that decrease the probability of association between 2 flanking sequences that include alleles of 2 genetic loci (Morton, 2005), LD does not necessarily have an absolute correspondence with genetic distance. LD estimation may also be influenced by factors such as polymorphism types and statistical methods for LD calculation. In this regard, informativity and other population parameters are presented in this study, as well as the LD pattern of 5 short tandem repeat (STR) markers within the DSCR, which covers 8.81 Mb (16.96 cM) in the 21q22.3 region in a Mexican population. The knowledge of LD patterns could be useful in future association studies, particularly involving the DSCR, where it has been proposed that some haplotype combinations could explain the clinical variability seen in Down syndrome (Kerstann et al., 2004).

MATERIAL AND METHODS

The present study included 369 healthy individuals living in Guadalajara, Jalisco, Mexico. They included 115 parents and 254 children from 62 families. Each family had a positive paternity test by autosomal markers from the CODIS core loci (paternity index >10,000), non-consanguineous parents, and at least 2 children.

Ethical considerations

Informed written consent was obtained from all individuals before enrollment in the study, according to the ethical guidelines of the Declaration of Helsinki amended in 2008.

Genotyping

Five informative STRs [each marker with expected heterozygosity (H_E) >0.74] were genotyped in 369 individuals by conventional PCR: D21S1440, D21S168, D21S1260, D21S1446, and D21S1411 (namely S1, S2, S3, S4, and S5, respectively) located at 21q22.3. Marker order was based on genetic distance (ALFRED, 2012; NCBI, 2013). Primer sequence, melting temperature, and other technical characteristics of the markers are presented in Table 1. For genotyping, 10 ng genomic DNA was used in a total volume of 10 µL containing 1X PCR buffer, 1.5 mM $MgCl_2$, 100 µM dNTPs, 0.03 U Taq polymerase (Invitrogen, Carlsbad, CA, USA), and variable primer concentration (0.8 µM S1, S3 and S4; 0.20 µM S2 and S5). Electrophoresis was carried out using 7 (tri- and tetranucleotide markers) or 10% (dinucleotide markers) polyacrylamide gels and silver staining. Allele identification was done by using a 10-bp (Invitrogen) and homemade allelic ladders containing pooled samples (Pacek et al., 1993). In addition, as genotyping control, we used some similar size samples previously validated by sequencing (ABI PRISM 377 Genetic Analyzer, Applied Biosystems, Foster City, CA, USA).

Table 1. Structure and thermal PCR characteristics of 5 short tandem repeat on chromosome 21.

Marker[a]	Sequence	cM	Repeat motif	Size (bp) (allele)	Tm (°C)
D21S1440	GAGTTTGAAAATAAAGTGTTCTGCG CCCCACCCCTTTTAGTTTTA	45.39	TAA	148-178 (5-15)	57
D21S168	ATGCAATGTTATGTAGGCTG CGGCATCACAGTCTGATAAA	49.90	GT	106-120 (16-23)	61
D21S1260	TCCAAGGGGTTCATCC CCCAAGGCACTGTTCC	56.33	CA	196-21 (10-21)	60
D21S1446	ATGTACGATACGTAATACTTGACAA GTCCCAAAGGACCTGCTC	57.77	CTAT	205-229 (7-13)	60
D21S1411	ATGATGAATGCATAGATGGATG AATGTGTGTCCTTCCAGGC	62.35	GATA	265-329 (16-32)	64

[a]Data according to Allele Frequency Database and the National Center for Biotechnology Information. Tm = annealing temperature employed in the PCR program.

Statistical analysis

Allele frequency, derived from observed genotype counts, and parental genotype proportions were compared with those of Hardy-Weinberg expectations (HWE) by a chi-square test. Qualitative data were analyzed with the chi-square test or the Fisher exact test, where P = 0.05 was the statistically significant threshold. Parental haplotypes were determined by family segregation (looking at transmission of markers, two by two loci, in the offspring). LD was expressed as Lewontin's D' corrected coefficient. Bergholdt et al. (2005) and Ziegler and Koenig (2006) recommend not including either observed or expected haplotype frequencies lower than 3% in the analysis to exclude the low haplotype frequency bias. Considering the sample size, we set up a cutoff of 4% to make the analysis more stringent. In addition, the P values were adjusted with Bonferroni's correction for multiple testing (Lander and Kruglyak, 1995). H_E was calculated as $1 - \sum p_i^2$, where p_i is the observed frequency of each allele, and

haplotype diversity as 1 - $\sum H_i^2$, where H_i is the frequency of each haplotype in the population tested. Haplotype counts obtained by family segregation were compared with those calculated from parental genotypes using the Arlequin® software.

RESULTS

Allele frequencies in parents, H_E, and P values for HWE are presented in Table 2. Genotype distribution for all 5 loci did not deviate from HWE ($P > 0.05$). The most informative STR was S5 (14 alleles, $H_E = 0.89$) and the least informative was S1 (5 alleles, 0.69).

Table 2. Allele frequencies of 5 short tandem repeats of the chromosome 21 in healthy subjects from Guadalajara, Jalisco, Mexico.

Allele	Observed (2n = 230)	Frequency ± SE (%)	HWE	H_E
D21S1440 (S1)				
8	77	33.5 ± 0.0968		
9	35	15.2 ± 0.0561		
10	96	41.7 ± 0.1057		
11	21	9.1 ± 0.0361	P = 0.06	0.69
12	1	0.5 ± 0.0018		
D21S168 (S2)				
16	13	5.7 ± 0.0231		
17	3	1.3 ± 0.0056		
18	19	8.3 ± 0.0329		
19	38	16.5 ± 0.0599		
20	96	41.7 ± 0.1057		
21	44	19.7 ± 0.0672		
22	11	4.8 ± 0.0198		
23	6	2.6 ± 0.0111	P = 0.24	0.75
D21S1260 (S3)				
10	13	5.7 ± 0.0232		
11	10	4.3 ± 0.0181		
12	11	4.8 ± 0.0198		
13	64	27.8 ± 0.0873		
14	21	9.1 ± 0.0361		
15	42	18.3 ± 0.0649		
16	22	9.6 ± 0.0376		
17	15	6.5 ± 0.0265		
18	14	6.1 ± 0.0249		
19	7	3.0 ± 0.0129		
20	9	3.9 ± 0.0163		
21	2	0.9 ± 0.0037	P = 0.06	0.86
D21S1446 (S4)				
7	74	32.1 ± 0.0522		
8	29	12.6 ± 0.0225		
9	33	14.3 ± 0.0253		
10	77	33.5 ± 0.0541		
11	16	7.0 ± 0.0128		
12	1	0.5 ± 0.0009	P = 0.35	0.74
D21S1411 (S5)				
16	3	1.3 ± 0.0056		
17	3	1.3 ± 0.0056		
18	5	2.1 ± 0.0092		
19	17	7.4 ± 0.0298		
20	32	13.9 ± 0.0521		
21	22	9.6 ± 0.0376		
22	23	10.0 ± 0.0391		
23	26	11.3 ± 0.0436		
24	29	12.6 ± 0.0479		
25	43	18.7 ± 0.0661		
26	13	5.7 ± 0.0231		
27	9	3.9 ± 0.0163		
28	4	1.7 ± 0.0074		
29	1	0.5 ± 0.0019	P = 0.45	0.89

HWE = Hardy-Weinberg expectations; SE = standard error; H_E = expected heterozygosity.

Eight haplotypes were observed twice (8-19-15-7-21, 8-20-13-7-26, 10-20-13-7-23, 10-20-13-7-25, 10-20-13-8-25, 10-20-15-10-20, 10-20-16-10-24, and 10-21-13-7-25), while the other haplotypes were observed only once (data available upon request).

Haplotype distribution comparisons (those obtained by family segregation *vs* inferred) showed minor differences, because some haplotypes were only seen in family pedigrees. From 230 haplotypes, ≤158 had a frequency higher than 4% and were considered for LD analysis.

Pairwise inter-distance and LD data are presented in Figure 1 and Table 3, respectively. Significant two-site LD values were found for S1-S2 (LD' = 0.1954, P = 0.002) and for S4-S5 (LD' = 0.2800, P = 0.0065). Other marker combinations, even S3-S4 with the smallest inter-distance, did not show LD, and thus, the independent segregation hypothesis could not be rejected. Haplotype data by gender are presented in Tables 4 and 5. These distribution comparisons did not yield significant differences.

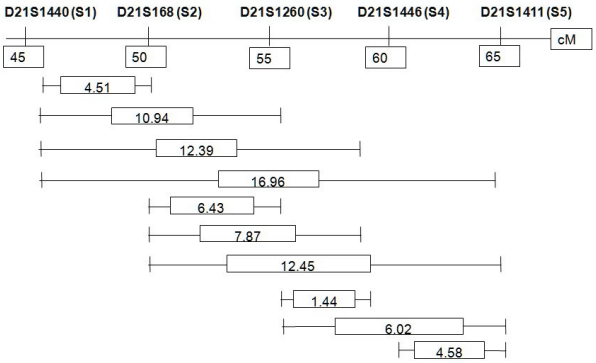

Figure 1. Genetic inter-distance of 5 short tandem repeats on chromosome 21. Genetic distance is reported as cM, between pair-site combinations. The minimal distance was between S3-S4 (1.44 cM) and the maximal was between S1-S5 (16.96 cM).

Table 3. Haplotype counts and linkage disequilibrium (LD) values.

(S1-S2)[a]	n (230)	Frequency		D'	P_c
		Observed	Expected		
8-18	10	0.04	0.03	0.29	0.04
8-19	17	0.07	0.06	0.16	0.16
8-20	22	0.10	0.14	0.31	0.01
8-21	12	0.05	0.06	0.18	0.32
9-20	13	0.06	0.06	0.11	0.69
10-19	12	0.05	0.07	0.24	0.07
10-20	**52**	**0.22**	**0.17**	**0.19**	**0.0021**
10-21	20	0.09	0.08	0.06	0.95
Pooled haplotypes[b]	72	0.31	0.33	NA	NA
S4-S5[c]					
20-7	12	0.05	0.04	0.08	0.79
20-10	13	0.06	0.05	0.10	0.33
22-10	11	0.05	0.03	0.21	0.06
24-7	7	0.03	0.04	0.25	0.17
24-10	14	0.06	0.04	0.22	0.07
25-7	**22**	**0.10**	**0.06**	**0.28**	**0.0065**
25-10	9	0.04	0.06	0.37	0.09
Pooled haplotypes[b]	142	0.62	0.72	NA	NA

[a]S1-S2, $P_c < 0.0071$. [b]Haplotypes with a frequency <4% were pooled; however, they were not considered for LD interpretation. [c]S4-S5, $P_c < 0.0083$. Significant P values are in bold. NA = not applicable. P_c = P value adjusted with Bonferroni's correction.

Table 4. S1-S2 haplotype comparison according to gender.

Haplotype	Female [n (frequency)]	Male [n (frequency)]	P
8-18	5 (0.04)	5 (0.04)	1.00
8-19	8 (0.07)	9 (0.08)	0.73
8-20	11 (0.09)	11 (0.09)	1.00
8-21	7 (0.06)	5 (0.04)	0.76
9-20	5 (0.04)	8 (0.07)	0.53
10-19	3 (0.03)	9 (0.08)	0.12
10-20[a]	27 (0.23)	25 (0.22)	-
10-21	15 (0.13)	5 (0.04)	0.11
Pooled haplotypes[b]	34 (0.31)	38 (0.34)	0.60

[a]Reference haplotype. [b]Haplotypes with a frequency <4% were pooled.

Table 5. S4-S5 haplotype comparison according to gender.

Haplotype	Female [n (frequency)]	Male [n (frequency)]	P
20-7	3 (0.03)	9 (0.08)	0.29
20-10	6 (0.05)	7 (0.06)	0.96
22-10	6 (0.05)	5 (0.04)	0.62
24-7	3 (0.03)	4 (0.03)	1.00
24-10	4 (0.03)	10 (0.09)	0.48
25-7[a]	10 (0.09)	12 (0.10)	-
25-10	5 (0.04)	4 (0.03)	0.70
Pooled haplotypes[b]	78 (0.71)	64 (0.57)	0.40

[a]Reference haplotype. [b]Haplotypes with a frequency <4% were pooled.

DISCUSSION

There are limitations for direct analysis of recombination rates. For instance, large human pedigrees, desirable for obtaining a large number of informative meiosis events, are scant; spermatozoa analysis provides data only in males, while oocyte manipulation demands high and cutting-edge technology (Jeffreys et al., 2000; Tiemann-Boege et al., 2006). Therefore, indirect LD estimation methods are commonly used to infer recombination. LD analysis has also provided valuable information in studies on human population divergence, disease association, and gene candidate identification (Sved et al., 2008; Zhang et al., 2009; Beuten et al., 2011; Li et al., 2012).

In this study, marker informativity was high (H_E >0.69), and 93% of haplotypes were identified by segregation in families. There was no evidence of LD in haplotypes at 5 loci, as can be foreseen on the basis of high haplotype diversity (99.58%). Haplotype comparisons (segregation in families *vs* inferring) showed minimal differences with some rare haplotypes only observed in families. This makes sense since the Arlequin® software is based on expectation-maximization, a robust algorithm against deviations from HWE, but it does not handle a large number of loci or detect rare haplotypes (Ziegler and Koenig, 2006). The present results suggest that both haplotype identification methods are reliable enough for haplotype and frequency estimates for the multiallelic loci used here.

LD, as it happens with recombination rates, is irregularly distributed across the genome, and it is influenced by factors such as population type (O'Connor et al., 2011), DNA sequence (Fullerton et al., 2001), number of transcriptional factors (Dawson et al., 2002), gene

density, gene function and chromosome length (Smith et al., 2005), and epigenetic effects (Jeffreys et al., 2000).

It has been suggested that LD is likely to be found at distances <1 cM between two loci, (Ziegler and Koenig et al., 2006); longer inter-spaces, as in this study, make it improbable to detect LD. Notwithstanding, the present results suggest linkage between S1-S2 and S4-S5, with inter-distances >4 cM. Analogous findings have been seen at more than 3 map units (Gordon et al., 2000; Service et al., 2001). Otherwise, neither the highest D' values nor the least marker inter-distance was significant. For instance, haplotype 12-27 at S1-S5 disclosed 100% D' ($P < 0.00001$), but had a low frequency (data not shown); S3-S4, the two closest STRs, showed no significant LD. No significant differences were found in haplotype distribution by gender (Tables 4 and 5).

To summarize, only 2 of the 4 paired STRs showed significant LD, where they showed the shortest distances within the region studied. The lack of significant D' between internal pairs can be indirectly explained also by crossover interference (Zhao et al., 1995). The non-homogeneous LD estimates in the 21q22.3 region in this report stress the importance of not determining LD solely on the basis of genetic distance.

As stated above, LD estimates can be used for gene localization and association mapping (Morton, 2005). In the context of complex diseases, haplotype analyses could be more useful (and less expensive) than single marker studies, since explanations such as 2 or more cis-acting mutations can be better derived from the former (Ziegler and Koenig, 2006). An analogous approach could be applied in trisomy association studies. For instance, considering the possibility of Down syndrome clinical and complication heterogeneity due to allele variation and combination (Kerstann et al., 2004), this variation is expected to correlate with haplotype diversity and distribution at DSCR, hence, the relevance of haplotype studies, such as the present one in DSCR. Knowledge of the distribution of LD in particular genomic regions, concentrating first on small fractions of the genome, will provide better designs in large-scale studies.

ACKNOWLEDGMENTS

We thank all the individuals who participated in this study.

Conflict of interests

The authors declare that they have no conflict of interests.

REFERENCES

ALFRED (2012). Allele Frequency Database. Available at [http://alfred.med.yale.edu]. Accessed November 5, 2012.

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, et al. (2004). Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat. Rev. Genet.* 5: 725-738.

Bergholdt R, Nerup J and Pociot F (2005). Fine mapping of a region on chromosome 21q21.11-q22.3 showing linkage to type 1 diabetes. *J. Med. Genet.* 42: 17-25.

Beuten J, Gelfond JA, Piwkham D, Pollock BH, et al. (2011). Candidate gene association analysis of acute lymphoblastic leukemia identifies new susceptibility locus at 11p15 (LMO1). *Carcinogenesis* 32: 1349-1353.

Concannon P, Onengut-Gumuscu S, Todd JA, Smyth DJ, et al. (2008). A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3. *Diabetes* 57: 2858-2861.

Dawson E, Abecasis GR, Bumpstead S, Chen Y, et al. (2002). A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 418: 544-548.

Fullerton SM, Bernardo CA and Clark AG (2001). Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* 18: 1139-1142.

Gordon D, Simonic I and Ott J (2000). Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. *Genomics* 66: 87-92.

Hwang KT, Han W, Cho J, Lee JW, et al. (2008). Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. *Int. J. Cancer* 123: 1807-1815.

Jeffreys AJ, Ritchie A and Neumann R (2000). High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum. Mol. Genet.* 9: 725-733.

Kerstann KF, Feingold E, Freeman SB, Bean LJ, et al. (2004). Linkage disequilibrium mapping in trisomic populations: analytical approaches and an application to congenital heart defects in Down syndrome. *Genet. Epidemiol.* 27: 240-251.

Lander E and Kruglyak L (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11: 241-247.

Li Y, Li Z, Zhang X, Yan C, et al. (2012). Association of ALOX5AP haplotypes with susceptibility to coronary artery disease in a Chinese Han population. *Eur. J. Intern. Med.* 23: e119-e123.

Morton NE (2005). Linkage disequilibrium maps and association mapping. *J. Clin. Invest.* 115: 1425-1430.

NCBI (2013). National Center for Biotechnology Information. Available at [http://www.ncbi.nlm.nih.gov/genome/sts/]. Accessed November 5, 2012.

O'Connor KL, Hill CR, Vallone PM and Butler JM (2011). Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples. *Forensic Sci. Int. Genet.* 5: 538-540.

Pacek P, Sajantila A and Syvänen AC (1993). Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. *PCR Methods Appl.* 2: 313-317.

Service SK, Ophoff RA and Freimer NB (2001). The genome-wide distribution of background linkage disequilibrium in a population isolate. *Hum. Mol. Genet.* 10: 545-551.

Smith AV, Thomas DJ, Munro HM and Abecasis GR (2005). Sequence features in regions of weak and strong linkage disequilibrium. *Genome Res.* 15: 1519-1534.

Sved JA, McRae AF and Visscher PM (2008). Divergence between human populations estimated from linkage disequilibrium. *Am. J. Hum. Genet.* 83: 737-743.

Tiemann-Boege I, Calabrese P, Cochran DM, Sokol R, et al. (2006). High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. *PLoS Genet.* 2: e70.

Zhang H, Sol-Church K, Rydbeck H, Stabley D, et al. (2009). High resolution linkage and linkage disequilibrium analyses of chromosome 1p36 SNPs identify new positional candidate genes for low bone mineral density. *Osteoporos. Int.* 20: 341-346.

Zhao H, Speed TP and McPeek MS (1995). Statistical analysis of crossover interference using the chi-square model. *Genetics* 139: 1045-1056.

Ziegler A and Koenig IR (2006). A Statistical Approach to Genetic Epidemiology. 1st edn. Wiley-VCH, Weinheim.