Genomics & Informatics Vol. 6(3) 130-135, September 2008

Haplotype Phylogeny of a 200kb Region in the Human Chromosome X Terminal Band (q28)

Sangsoo Kim*

Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Korea

Abstract

The haplotypes of a 200 kb region in the human chromosome X terminal band (q28) were analyzed using the International HapMap Project PhaseII data, which had been collected for three analysis panels (YRI, CEU, and CHB+JPT). When multiple linkage disequilibrium blocks were encountered for a panel, the neighboring haplotypes that had crossover rate of 5% or more in the panel were combined to generate 'haploid' configurations. This resulted in 8, 7, and 5 'haploid' configurations for the panels of YRI, CEU, and CHB+JPT, respectively. The multiple sequence alignment of these 'haploids' was used for the calculation of allele-sharing distances and the subsequent principal coordinate analysis. Two 'haploids' in CEU and CHB+JPT were hypothesized as 'parental' in light of the observations that the successive recombinants of these haploids can model two other haploids in CEU and CHB+JPT, and that their configurations were consistent with those in YRI. This study demonstrates the utility of haplotype phylogeny in understanding population evolution.

Keywords: haplotype, SNP, phylogeny, haploview, HapMap

Introduction

Linkage disequilibrium (LD), the nonrandom association of alleles at different loci, is the result of evolutionary history of the population involving mutations, selection, recombination, population bottleneck and random genetic drift (Xiong & Jin, 2007). A haplotype is a set of co-occurring polymorphic alleles on the same chromosome. As haplotypes partition the genome into blocks that show strong LD, while the blocks are separated by weaker association or recombination hotspots. The haplotype block model has particular implications in the study of dense markers such as single nucleotide polymorphism (SNP) as it implies that a smaller number of markers (tagging SNPs) are necessary to uniquely distinguish different haplotypes.

The International HapMap Project aims to produce such valuable haplotype information of the human genome (The International HapMap Consortium, 2003). A total of 270 samples, 90 samples from each of three major ethnic groups, African, Asian, and European, were genotyped, phased and release to the public freely (Thorisson *et al.,* 2005). Taking advantage of such rich and yet free resource, various research hypotheses can be tested. As an example, the readily downloadable haplotype configurations of the panels can be compared and phylogenetic relationship among them can be pursued to glean how the haplotypes might have been evolved. Phylogenetic studies based on haplotypes are not new; genetic association studies have been done with haplotypes inferred from restriction endonuclease mapping in fruitfly (Templeton *et al.,* 1987). Recently, haplotype phylogeny has been advocated for detecting disease susceptibility loci (Bardel *et al.,* 2005). A haplotype phylogenetic tree was constructed based on six tagging SNPs of DTNBP1 (Mutsuddi *et al.,* 2006). Genealogical relationships among haplotypes in a chromosome 2 8.4 kb region without obligate recombination events were demonstrated using the CEU samples only (The International HapMap Consortium, 2005). If other population samples such as YRI, CHB, and JPT had been included, the haplotype blocks would have been fragmented due to a number of historical recombination events and phylogenetic studies with such a small block would have not been informative.

In this study, instead of conventional tree-based phylogeny, principal coordinate analysis (PCoA) (Higgins, 1992) was employed using the haplotype data on a region encompassing multiple blocks. As PCoA, albeit distance-based, is useful to grasp the major trend among the sequences, it would be worth to try how PCoA performs with such a dataset. As an illustrative purpose, a region of 200 kb in chromosome Xq28, which is about 1 Mb away from the pseudoautosomal region (PAR2) at the tip of X chromosome long arm, was chosen and the haplotype structures of three ethnic groups that showed apparent recombination events were compared. This region of the human genome harbors several important disease genes such as glucose-6-phosphate dehydrogenase (G6PD), cancer/testis antigens (CTAG1B, CTAG2), and Gab3 protein (GAB3).

*Corresponding author: E-mail sskimb@ssu.ac.kr
Tel +82-2-820-0457, Fax +82-2-824-4383
Accepted 10 September 2008

Methods

Genotype and haplotype data

The genotype data used in this study were of the Release 21/phaseII of the International HapMap Project (The International HapMap Consortium, 2005) and were downloaded from the Project web site (Thorisson *et al.,* 2005) using Haploview 4.1 (Barrett *et al.,* 2005). The HapMap Project intended to cover all three major ethnic groups and collected DNA samples from four populations: 90 each from 30 trios of African Yoruba people (YRI) and 30 trios of western European ancestry in the US (CEU), and 45 each from unrelated Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) (The International HapMap Consortium, 2003). As the genotypes of Chinese and Japanese were very similar to each other, the HapMap Consortium pooled these groups as a single analysis panel (CHB+JPT) (The International HapMap Consortium, 2005). Consequently, genotype and phased haplotype data were available in three separate analysis panels, namely, YRI, CEU, and CHB+JPT. Upon downloading the genotype data separately for each analysis panel, various pairwise LD measures were calculated for markers having minor allele frequency of 5% or more. LD blocks were defined usinge the default parameters settings and haplotypes for each block were estimated by the EM algorithm implemented in Haploview (Barrett *et al.,* 2005). Included in the Haploview output for each haplotype were population frequency and transition probability (crossover percentage) to the haplotypes in the neighboring blocks. The International HapMap Project also provides phased haploid structures for each individual in the analysis panels. They were downloaded from the ftp site.

Generation of 'haploids'

Haplotypes from neighboring blocks were connected to generate a long 'haploid' configuration. All possible combinations having crossover rate of 5% or more were used in enumeration of the putative 'haploid' configurations. A 'haploid' was denoted as *ceu.i.j.k*, indicating that it was generated by combining the *i, j,* and *k*-th haplotype from the first, second, and third block of the CEU panel, respectively. Those 'haploids' having identical configuration were grouped together and one of them was arbitrarily chosen to represent the group.

Calculation of allele-sharing distance and PCoA

Dissimilarity between a pair of 'haploids' was calculated as the ratio of the number of different alleles over the number of all the markers. No correction for multiple substitutions was made, as only the polymorphic sites were included in the multiple alignments. The resulting dissimilarity matrix was then used to lay out the 'haploids' onto a multidimensional space using a classical multidimensional scaling routine implemented in R statistical system (*http://www.r-project.org*). The first two principal coordinates were taken and plotted for visual inspection of the relationship between 'haploids'.

Inference of ancestral alleles

The alleles of chimpanzee, orangutan, and macaque orthologous to the human Xq28 region were downloaded using the Table Browser of UCSC Genome Browser. The human ancestral allele was inferred parsimoniously, *i.e.*, as the chimpanzee allele if and only if the chimpanzee allele was the same with one of the human bi-alleles. If the chimpanzee allele was missing, the orangutan and macaque alleles were looked up in turn.

Results

The genotype data of the chromosome Xq28 200kb region were downloaded from International HapMap Project web site (*http://www.hapmap.org*) using Haploview (Barrett *et al.,* 2005). See Table 1 for the summary information of the markers. For each analysis panel, LD blocks and recombination hot spots were identified using Haploview (data not shown). Haplotypes reconstructed by Haploview for each block are shown in Fig. 1. For YRI (Fig. 1A), seven haplotypes having frequency greater than 1% were found for the first block, while three haplotypes were found respectively for the other two short blocks. Similarly for CEU (Fig. 1B) and CHB+JPT (Fig. 1C) panels, haplotypes were reconstructed separately for three blocks. It should be emphasized that for YRI, unlike the other non-African panels, about the first half of the SNPs listed in Table 1 had high recombination rate and consequently not included in the haplotypes. Also shown in Fig. 1 are the connections between neighboring haplotypes. Thick and thin lines were drawn for the crossover percentage greater than 10% and 1%, respectively. As the panels were different one another in haplotype boundaries, it was not possible to directly compare the haplotypes from different panels. Instead we regenerated the 'haploid' sequences that covered the 200 kb region by recombining the neighboring haplotypes having crossover rate of 5% or more. There were eight, seven, and five such 'haploids' for YRI, CEU, and CHB+JPT, respectively.

In order to find the phylogenetic relationship among

Table 1. SNP markers used in the analysis

Common ID	rs #	Position	Allele†		Genotype*			Minor Allele Frequency			Panel ID†		
			Major	Ancest.	YRI	CEU	C+J	YRI	CEU	C+J	YRI	CEU	C+J
1	rs2071429	153324212	A	G	G:A	A:G	A:G	0.022	0.156	0.133	01	01	01
	rs1050828	153327921	C	C	C:T			0.111			02		
	rs762516	153328367	C	C	C:T			0.156			03		
	rs743544	153328870	G	G		G:A	G:A		0.022	0.496		02	02
2	rs2472393	153335000	C	T	T:C	C:T	C:T	0.022	0.156	0.126	04	03	03
3	rs2472394	153335055	C	C	C:A	C:A	C:A	0.133	0.122	0.067	05	04	04
	rs5986997	153391253	C	C	C:T			0.011			06		
4	rs4898389	153391341	A	G	G:A	A:G	A:G	0.033	0.156	0.126	07	05	05
	rs4898390	153391916	C	T			C:T			0.007			06
5	rs7879049	153393397	A	G	A:G	A:G	A:G	0.344	0.122	0.067	08	06	07
	rs7053878	153397804	T	A	T:A		T:A	0.022		0.119	09		08
	rs5987002	153441074			C:T			0.222			10		
6	rs4526543	153441705	T	C	C:T	T:C	T:C	0.1	0.389	0.185	11	07	09
	rs5987003	153444584	G	G	G:C		G:C	0.1		0.007	12		10
7	rs10127172	153449042	C	C	T:C	C:T	C:T	0.022	0.389	0.185	13	08	11
8	rs4898348	153450098	C	C	C:T	C:T	C:T	0.433	0.267	0.119	14	09	12
9	rs5945222	153450908	C	T	T:C	C:T	C:T	0.022	0.389	0.185	15	10	13
10	rs5945223	153451374	G	G	T:G	G:T	G:T	0.1	0.389	0.185	16	11	14
11	rs5945224	153451839	G	G	A:G	G:A	G:A	0.022	0.389	0.185	17	12	15
12	rs5987005	153456571	T	C	C:T	T:C	T:C	0.022	0.389	0.185	18	13	16
	rs4425917	153467117	T	T	T:C			0.022			19		
	rs3813455	153467839	C	C		C:G	C:G		0.067	0.059		14	17
13	rs5945108	153468101	A	A	A:G	A:G	A:G	0.022	0.267	0.119	20	15	18
14	rs4431759	153468177	C	T	T:C	C:T	C:T	0.022	0.389	0.185	21	16	19
	rs12395061	153469224	A	A	A:G			0.222			22		
15	rs5945109	153471453	A	G	G:A	A:G	A:G	0.022	0.389	0.185	23	17	20
	rs17281342	153471843	A	A		A:T			0.067			18	
16	rs5987011	153472536	G	G	A:G	G:A	G:A	0.044	0.389	0.185	24	19	21
17	rs6571262	153483962			C:A	A:C	A:C	0.022	0.389	0.185	25	20	22
18	rs884840	153491735	C	C	C:T	T:C	T:C	0.022	0.389	0.185	26	21	23
19	rs6643615	153499102	C	C	C:T	T:C	T:C	0.022	0.389	0.185	27	22	24
20	rs5945232	153501932	G	G	G:C	G:C	G:C	0.133	0.3	0.148	28	23	25
21	rs5945233	153503029	T	T	T:A	T:A	T:A	0.244	0.311	0.141	29	24	26
	rs5987014	153504114	C	C	C:T			0.011			30		
	rs17281349	153504565	G	G		G:A			0.067			25	
	rs7066018	153504638	T	T	T:C			0.278			31		
	rs6643696	153506325	G	A	G:A			0.278			32		
	rs6571263	153507934	C	C	C:T			0.278			33		
22	rs2664170	153509306	G	G	G:A	A:G	A:G	0.444	0.322	0.163	34	26	27
23	rs5987015	153511685	G	A	G:A	A:G	A:G	0.467	0.322	0.163	35	27	28

the haplotypes covering the same genomic region, a multiple sequence alignment of these 'haploid' sequences has been created. As shown in Table 1, a total of 47 SNPs had been used at least once in the haplotype reconstruction. However, some of them were monomorphic in at least one other panel and removed from the multiple alignment. A total of 25 markers had been polymorphic in all three panels and used in the subsequent analysis. Consequently some redundancies were noticed among the 'haploid' sequences as shown in Table 2. There were four and five non-redundant 'haploids' for YRI and non-YRI panels, respectively. As noted above, a large portion of loci had high recombination rate in YRI panel and haplotypes were not constructed in that portion ('un-haplotyped'). For the YRI individuals, the phased individual genome-wide haploid structures constructed by the International HapMap Project were downloaded and grouped according to the 'haploid' structures that were generated from this study. The 'un-haplotyped' portion was then filled in by the consensus among the grouped haploid structures, while the polymorphic sites were left with 'N'. In order to validate

Table 1. Continued

Common ID	rs #	Position	Allele†		Genotype*			Minor Allele Frequency			Panel ID†		
			Major	Ancest.	YRI	CEU	C+J	YRI	CEU	C+J	YRI	CEU	C+J
24	rs2728526	153512921	C	T	C:T	T:C	T:C	0.467	0.322	0.163	36	28	29
	rs1848764	153514246	A	A	A:T			0.222			37		
	rs1509789	153517046	G	G	G:A			0.222			38		
	rs6643697	153518993	A	A	A:G			0.011			39		
	rs17051988	153521732	G	G	G:A			0.233			40		
	rs7885619	153524421	T	T	T:G			0.133			41		
25	rs2728725	153527460	T	T	C:T	C:T	C:T	0.411	0.322	0.163	42	29	30

* Monomorphic SNPs in a given analysis panel are omitted.
† Marker IDs in a given analysis panel and used in Fig. 1.
‡ Major alleles are for the forward strand, while the ancestral alleles were putatively assigned parsimoniously by referring to those of chimpanzee, orangutan, and macaque.

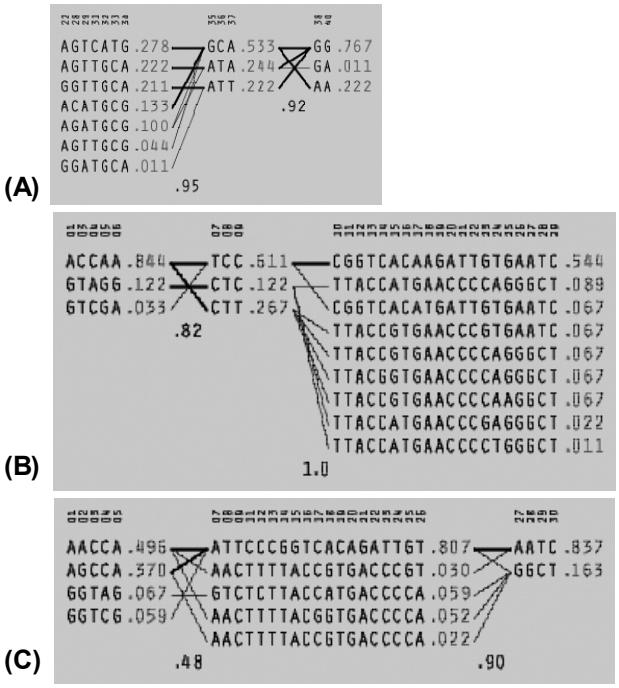

Fig. 1. Haploview outputs showing the haplotypes of the 200kb Xq28 region for (A) YRI, (B) CEU, and (C) CHB+JPT.

the presence of other non-YRI 'haploids' as well, they were also compared with the phased individual haploid structures constructed by the International HapMap Project. Their frequencies observed in that individual data were shown in Table 2. Although the frequencies estimated from the multi-haplotype crossover rate were qualitatively similar to those observed values, all the 'haploids' constructed in this study had the observed frequency higher that 5%, the same cutoff used for the crossover percentage. In fact, all the individual haploids with frequency higher than 5% were represented in the non-redundant 'haploids' constructed in this study.

For each pair of the non-redundant 'haploids', the allele-sharing distance was calculated to form a dissimilarity matrix. PCoA was then performed to lay out those nine 'haploids' in a multidimensional space, while preserving the distance relationship as much as possible. The first two dimensions were plotted (Fig. 2). The ancestral and YRI 'haploids' were separated into two groups along the first principal coordinate (PC1) axis, showing a progressive substitution pattern in the last six loci for *yri.3.0.0, yri.4.0.0, yri.0.0.0,* and ancestral 'haploid', while *yri.1.0.0* was the most distinct with a highly polymorphic pattern. Interestingly there were nearby non-YRI 'haploids' that showed congruent patterns: for example, *c+j.2.1.1* was located on the lower PC1 side as *yri.3.0.0*, while *c+j.0.0.0* was on the higher PC1 side as *yri.0.0.0*. Among those five non-YRI 'haploids', these two were the most distant, sharing just two alleles out of 25, while the other three were intermediate. In fact, the first five alleles of *c+j.0.2.1* were that of *c+j.0.0.0*, while the rest were that of *c+j.0.2.1*. It is cautiously hypothesized that *c+j.0.2.1* might have been a recombinant of *c+j.2.1.1* and *c+j.0.0.0*. Even more interesting is that of ceu.0.1.3, which showed the pattern of *c+j.0.2.1* in the first 19 loci, while the last six were that of *c+j.0.0.0*. It implies that another round of recombination between and *c+j.0.0.0* might have produced *ceu.0.1.3*. This is the most parsimonious phylogenetic model among these non-YRI 'haploids'. Taken together, it can be hypothesized that *c+j.2.1.1* and might have been the 'parental' haploids, while the putative ancestral 'haploid' had a structure interlaced with both 'parental' haploids, implying they might have been diverged from the ancestral 'haploid'. The haploid *c+j.3.0.0* had the first four loci that were different from *c+j.0.0.0* and yet

Table 2. Haploid configurations* constructed by combining haplotypes

Haploid ID†	rs2071429	rs2472393	rs2472394	rs4898389	rs7879049	rs4526543	rs1012717	rs4898348	rs5945222	rs5945223	rs5945224	rs5987005	rs5945108	rs4431759	rs5945109	rs5987011	rs6571262	rs884840	rs6643615	rs5945232	rs5945233	rs2664170	rs5987015	rs2728526	rs2728725	Estimated frequency in each panel‡	Observed frequency in each panel‡
ceu.1.2.1	G	T	A	G	G	C	T	C	T	T	A	C	A	T	G	A	C	C	C	C	A	G	G	C	T	1.09%	10.67
c+j.2.1.1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	0.35%	5.93
ceu.0.1.4	A	C	C	A	A	C	T	T	T	T	A	C	G	T	G	A	C	C	C	C	A	G	G	C	T	1.79%	20.00
ceu.0.1.5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1.79%	—
ceu.0.1.6	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1.79%	—
c+j.0.2.1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	0.27%	7.41
ceu.0.1.3	A	C	C	A	A	C	T	T	T	T	A	C	G	T	G	A	C	C	C	G	T	A	A	T	C	1.79%	6.67
ceu.0.0.0	A	C	C	A	A	T	C	C	C	G	G	T	A	C	A	G	A	T	T	G	T	A	A	T	C	31.44%	55.33
ceu.0.0.2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	3.87%	—
c+j.0.0.0	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	31.16%	72.59
c+j.1.0.0	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	27.60%	—
c+j.3.0.0	G	T	C	G	A	T	C	C	C	G	G	T	A	C	A	G	A	T	T	G	T	A	A	T	C	4.68%	5.93
yri.3.0.0	G	T	N	G	N	N	N	C	N	N	N	N	A	N	N	N	N	N	N	C	A	G	G	C	T	4.14%	13.33
yri.3.0.1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2.95%	—
yri.4.0.0	G	T	C	G	N	C	T	N	T	T	A	C	A	T	G	A	C	C	C	G	A	G	G	C	T	2.43%	8.00
yri.4.0.1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1.73%	—
yri.0.0.0	G	T	C	N	N	C	T	N	T	N	A	C	A	T	G	N	C	C	C	G	T	G	G	C	N	8.65%	31.33
yri.0.0.1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	6.17%	—
yri.1.1.0	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	G	T	A	A	T	C	5.17%	43.33
yri.2.2.0	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	4.68%	—

*'·' Denotes identical nucleotide as the one above.

†Representative IDs are given in bold letters.

‡Estimated frequency was calculated by multiplying haplotype frequencies and crossover rates, while the observed frequency was calculated by counting the number of individual haploids constructed by the HapMap Project that matched the shown 'haploid' structure constructed by this study.

congruent with those of *yri.3.0.0, yri.4.0.0, yri.0.0.0,* and the putative ancestral 'haploid', implying that it might be 'parental' to *c+j.0.0.0*.

Discussion

This study demonstrated the usefulness of haplotype phylogeny in reconstructing evolutionary history of the haplotypes. When applied to various ethnic groups or populations, it has the potential to suggest putative migration pathways. Although similar study can be performed with the genotype or 'haploid' configuration of the individuals, such strategies typically rely on some sort of genetic distance measure and the detailed configuration would be hidden by the distances. Consequently if a 'haploid' is found in-between two others, it cannot be definitely inferred whether it is a mixture of the latter or not. Character-based phylogenetic methods might be suitable in resolving these issues. However, these methods assume independent variation of the alleles and cannot incorporate the evolutionary concepts of linkage disequilibrium (LD) and nonrandom recombination. The method presented here, haplotype phylogeny, takes advantage of the simplification presented by LD and haplotypes. For cases with historical recombination events, the conventional tree-based methods such as UPGMA or neighbor-joining (Saitou & Nei, 1987) may not be appropriate, while PCoA still works by showing the major trends in sequences (Higgins, 1992).

The analysis of an admixture population in genomic scale may be best handled by a model-based clustering method (Xu *et al.,* 2008) such as STRUCTURE (Falush *et al.,* 2003). However, it is for testing model hypothesis and not for the discovery of admixture. It also requires lengthy computational time. On the other hand, the method presented here is appropriate for locating the potential loci of crossover events at the fine scale utilizing all the common markers. The results, then, may be tested with STRUCTURE.

However it should be noted that the work presented here has some limitations. For example, the spacing between neighboring SNPs were rather coarse (~10 kb)

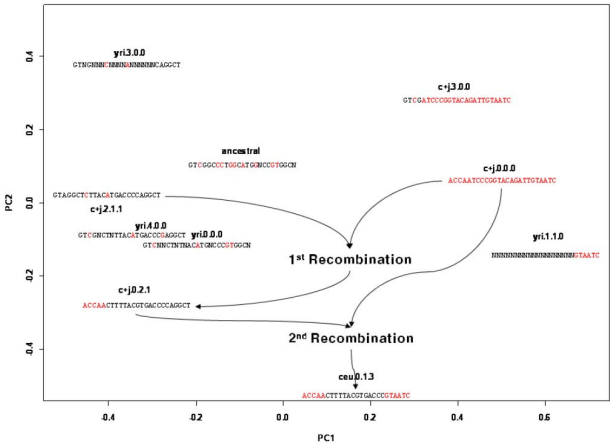

Fig. 2. Principal coordinate plot of the non-redundant haploids based on their allele-sharing distances. Alleles that were observed in one of the ancestral haploids, *c+j.0.0.0,* were shown in red. Note the separation of YRI haploids along the 1st principal coordinate axis: one formed by *yri.3.0.0, yri.4.0.0,* and *yri.0.0.0,* and the other by *yri.1.0.0.* The two putative 'parental' haploids in CEU and CHB+JPT, *c+j.2.1.1* and *c+j.0.0.0* had alleles congruent to the two most distant haplotypes *yri.3.0.0* and *yri.1.1.0,* respectively. The haploid *c+j.0.2.1* might be a recombinant of the two putative 'parental' haploids, while another round of recombination between *c+j.0.2.1* and *c+j.0.0.0* might have produced *ceu.0.1.3.*

and the number of samples is only 270 involving three distinct ethnic groups. Although the detailed result presented here may be subject to change as higher resolution data coupled with more samples become available, the basic idea and procedure proposed here should be applicable. With small number of unique haploids, it was possible to work out the phylogenetic relationships manually. With bigger data, it would be necessary that the analysis result be annotated with some sort of statistical significance.

Acknowledgements

This work has been supported by the Soongsil University Research Fund.

References

Bardel, C., Danjean, V., Hugot, J.P., Darlu, P., and Génin, E. (2005). On the use of haplotype phylogeny to detect disease susceptibility loci. *BMC Genetics* 6, 24.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.

Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567-1587.

Higgins, D.G. (1992). Sequence ordinations: a multivariate analysis approach to analysing large sequence data sets. *Bioinformatics* 8, 15-22.

Mutsuddi, M., Morris, D.W., Waggoner, S.G., Daly, M.J., Scolnick, E.M., and Sklar, P. (2006). Analysis of High-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. *The American Journal of Human Genetics* 79, 903-909.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4, 406-425.

Templeton, A.R., Boerwinkle, E., and Sing, C.F. (1987). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in drosophila. *Genetics* 117, 343-351.

The International HapMap Consortium. (2003). The International HapMap Project. *Nature* 426, 789-796.

The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature* 437, 1299-1320.

Thorisson, G.A., Smith, A.V., Krishnan, L., and Stein, L.D. (2005). The international hapmap project web site. *Genome Research* 15, 1591-1593.

Xiong, M., and Jin, L. (2007). Association studies of complex diseases. In *bioinformatics - from genomes to therapies Vol. 3, T.* Lengauer, ed. (Wiley-VCH, Germany), pp.1375-1426.

Xu, S., Huang, W., Qian, J., and Jin, L. (2008). Analysis of genomic admixture in uyghur and its implication in mapping strategy. *The American Journal of Human Genetics* 82, 883-894.