Journal of Biomedical Science (2006) 13:489–498
DOI 10.1007/s11373-006-9077-7

A comparison of major histocompatibility complex SNPs in Han Chinese residing in Taiwan and Caucasians

Hsin-Chou Yang[1], Chien-Hsin Lin[2], Chia-Ling Hsu[1], Shuen-Iu Hung[1], Jer-Yuan Wu[1], Wen-Harn Pan[1], Yuan-Tsong Chen[1] & Cathy S. J. Fann[1,2,*]
[1]*Academia Sinica, Institute of Biomedical Sciences, 128, Academia Road, Section 2 Nankang, Taipei, 115, Taiwan, ROC;* [2]*Institute of Genetics, Yang-Ming University, Taipei, 112, Taiwan, ROC*

Received 1 December 2005; accepted 30 January 2006

Key words: disease association mapping, ethnic heterogeneity, haplotype block, phylogenetic analysis, population admixture

Summary

Genetic dissection of complex diseases is both important and challenging. The human major histocompatibility complex is involved in many human diseases and genetic mechanisms. This highly polymorphic chromosome region has been extensively studied in Caucasians but not as well in Asians. Thus, we compared genotypic distributions, linkage disequilibria and haplotype blocks between Caucasian and Taiwan's Han Chinese populations. Moreover, we investigated the population admixture and phylogenetic system in Han Chinese residing in Taiwan. The results show that Taiwan's Han Chinese differ drastically in genotypic information compared with Caucasians but are relatively homogeneous among the three major ethnic subgroups, Minnan, Hakka and Mainlanders. Differences in allele frequency (AF) between Taiwanese and Caucasians in some disease-associated loci may reveal clues to differences in disease prevalence. The results of ethnic heterogeneity imply that public databases should be used with caution in cases where the study population(s) differs from the population characterized in the database. The high homogeneity we observed among the Taiwanese subpopulations mitigates the possibility of spurious association caused by ignoring population stratification in Taiwanese disease gene association studies. These results are useful for understanding our genetic background and designing future disease gene mapping studies.

Introduction

In this post-genome era, utilizing genomic/genetic analyses to study mechanisms of complex diseases is important but challenging. An understanding of the genetic background of study populations contributes to association mapping when identifying disease susceptibility genes underlying complex disorders [1–3]. Careful investigation of ethnic heterogeneity helps to avoid misuse of publicly available databases derived from other populations, therefore reduce a potential inference bias when applied to disease gene mapping studies [4–6].

Human major histocompatibility complex (MHC) on chromosomal region 6p21.3 is a highly polymorphic region that has been analyzed to study historical recombination, population evolution, and the development of autoimmune and infectious disorders [7–11]. This region has high within-population specificity and between-population heterogeneity, and has been well studied in Caucasians but not in Asians.

The Han Chinese are a major population in Asia. In this study, we consider Han Chinese residing in Taiwan, which is a small island of 36,000 km^2 with 23 million residents. The

*To whom correspondence should be addressed. Tel.: +886-2-27899144; Fax: +886-2-27823047; E-mail: csjfann@ibms.sinica.edu.tw

Taiwanese population consists of three subgroups, each with its own dialect and culture: the Minnan (70% of the population) immigrated to Taiwan about 300–400 years ago, the Hakka (13%) immigrated about 200 years ago, and the Mainlanders (14%) about 50 years ago. The origins of the Minnan and Hakka groups remain uncertain, although some studies suggest that the Minnan and Hakka might have originated from the indigenous population of southern China, whose ancestry was from the Hundred Yueh, Burmese, or Thais [12–14]. Other theories, however, support the assertion that the three ethnic groups have a common origin in the northern China Hwa-Shia [15, 16]. The Mainlander group is relatively heterogeneous, with members having originated from almost all geographic regions of China. Thus, they likely are a representative sample of people now living in China. Here we investigated the population admixture and phylogenetic system of Han Chinese residing in Taiwan.

Materials and methods

Study populations and samples

We studied data from Caucasian and three Taiwanese subgroup (Minnan, Hakka, and Mainlanders) samples. The Caucasian samples comprised cell lines of 100 founders of Caucasian descendents from the Coriell Cell Repositories as well as AF data obtained from 136 independent chromosomes of Centre d'Etude du Polymorphism Humain (CEPH) families published in an MHC region study (http://www.broad.mit.edu/mpg/idrg/projects/hla.html). From blood samples previously collected from 3385 healthy Taiwanese residents through the end of the year 2003 [17], we randomly selected 229 "pure resident" (i.e., three generations from the same group) samples to reduce possible dilution of the ethnic effect from each of the three main Taiwanese subgroups. This sample size provided about 80% power for pairwise comparisons with a significance level of 0.001.

Genotype data

We analyzed 185 SNPs after excluding those with a genotypic call rate smaller than 0.85. The region under consideration encompassed 3.96 Mb. The mean, median, and standard deviation of intermarker spacing were 21.5, 15.7, and 21.3 kb, respectively. SNP information, including rs number, position, the corresponding gene, functionality, and AF, are summarized in Supplementary Table 1 (online). The Puregene DNA purification kit (Gentra Systems, Minneapolis, MN, USA) was used to extract genomic DNA from blood. DNA concentration was quantitated by measuring UV absorbance at 260 nm. We used SpectroDESIGNER (Sequenom, San Diego, CA, USA) to design PCR primers and genotyping probes, and PCR-ABI 9700 thermocyclers (Applied Biosystems, Foster City, CA, USA) were used for all PCR amplifications and primer extension reactions. We carried out genotyping using a high-throughput genotyping platform based on MALDI-TOF mass spectrometry (Sequenom). The details of SNP genotyping were discussed in our previous study [18].

Allele frequency and comparative analysis

We estimated AFs of all SNPs in each population using an allele-counting approach. The Fisher exact test and adjustment for false discovery rate (FDR) [19] were conducted to identify SNPs that showed significant differences in AF distribution among different populations. Additionally, we examined the polymorphism status of each marker under the criterion, minor AF (MAF) > 0.01. The kappa coefficient of any two subgroups and the corresponding 95% confidence interval were used to assess the magnitude of polymorphism concordance.

Analysis of haplotype blocks and htSNPs

Based on the software HAPLOVIEW [20], we used the Lewontin D' measure [21] to estimate the intermarker coefficient of linkage disequilibrium (LD). The confidence interval of LD was estimated using a resampling procedure and was used to construct the haplotype blocks [22]. HtSNPs were selected on a block-by-block basis [20].

Analyses of population admixture and phylogeny

To investigate population admixture for the Taiwanese residents, we inferred the admixture proportion for the different populations for each individual based on a Bayesian Markov Chain

Monte Carlo algorithm using the software STRUCTURE [23]. The length of the burn-in period was 10,000, and the number of MCMC replications after the burn-in period was 10,000. We used the software PHYLIP [24] to calculate the Nei's distance [25] using AF data and then constructed a phylogenetic tree using the neighbor joining approach. A bootstrapping procedure with 10,000 replications was used to confirm the reproducibility of the resultant phylogenetic tree.

Results

We analyzed 185 SNPs within the MHC region using samples from Caucasians and Taiwanese non-aboriginal residents. The data sources in this paper included: (1) Caucasian samples obtained from the Coriell Cell Repositories (http://www.locus.umdnj.edu/ccr/) and genotyped by our laboratory, (2) Taiwanese samples obtained from the three subgroups collected in Taiwan Han Chinese Cell and Genome Bank [17], also genotyped by our laboratory, and (3) Caucasian AF data from an online database (http://www.broad.mit.edu/mpg/idrg/projects/hla.html). Information on the SNPs, AFs, and gene annotations is available (Supplementary Table 1 online).

Figure 1 shows histograms for the AFs of the two Caucasian sample groups and three Taiwanese subgroups. The two Caucasian histograms were highly consistent, and the three Taiwanese histograms showed similar AF distributions with only minor differences. However, there were large discrepancies between the Caucasian and Taiwanese samples.

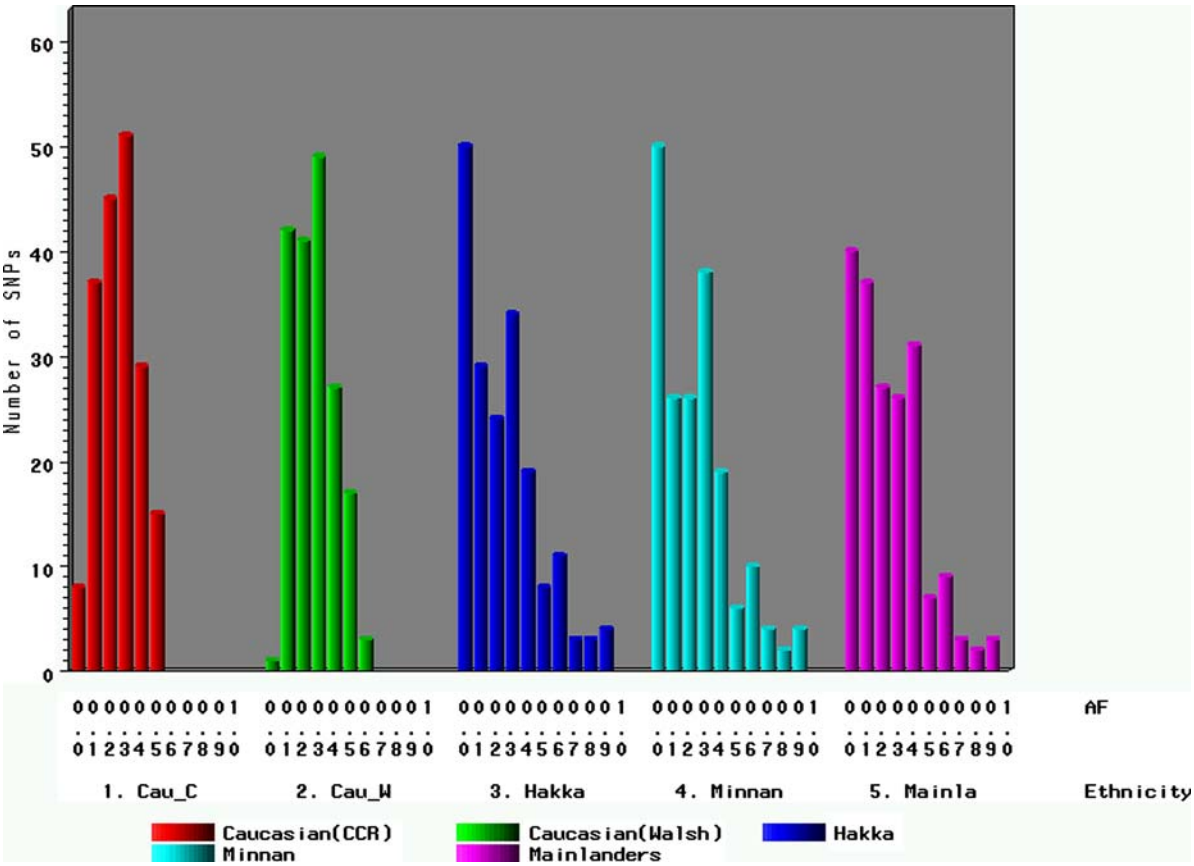


Figure 1. Allele frequency histograms of Caucasian samples from the Coriell Cell Repositories that we genotyped, a Caucasian population using the AF data in available database (http://www.broad.mit.edu/mpg/idrg/projects/hla.html), and the three Taiwanese ethnic subgroups, Hakka, Minnan, and Mainlanders. Because of the duality of the alleles of each SNP, we used the minor alleles of the Caucasian population in our genotyping as reference alleles for all of the histograms. The histograms show a smaller difference in AF distribution among the three Taiwanese subgroups relative to a large discrepancy between the Caucasians and Taiwanese.

In a comparison of genotypic distributions between the Taiwanese and Caucasian samples, Fisher exact tests identified significant differences in 101 of the 185 SNPs, after adjusting for the FDR (Supplementary Figure 1 online), demonstrating a high degree of ethnic heterogeneity between the Taiwanese and Caucasian samples. Among the 101 SNPs, 48 were located within 39 distinct gene regions, where 12 genes have been found to be associated with disease etiology

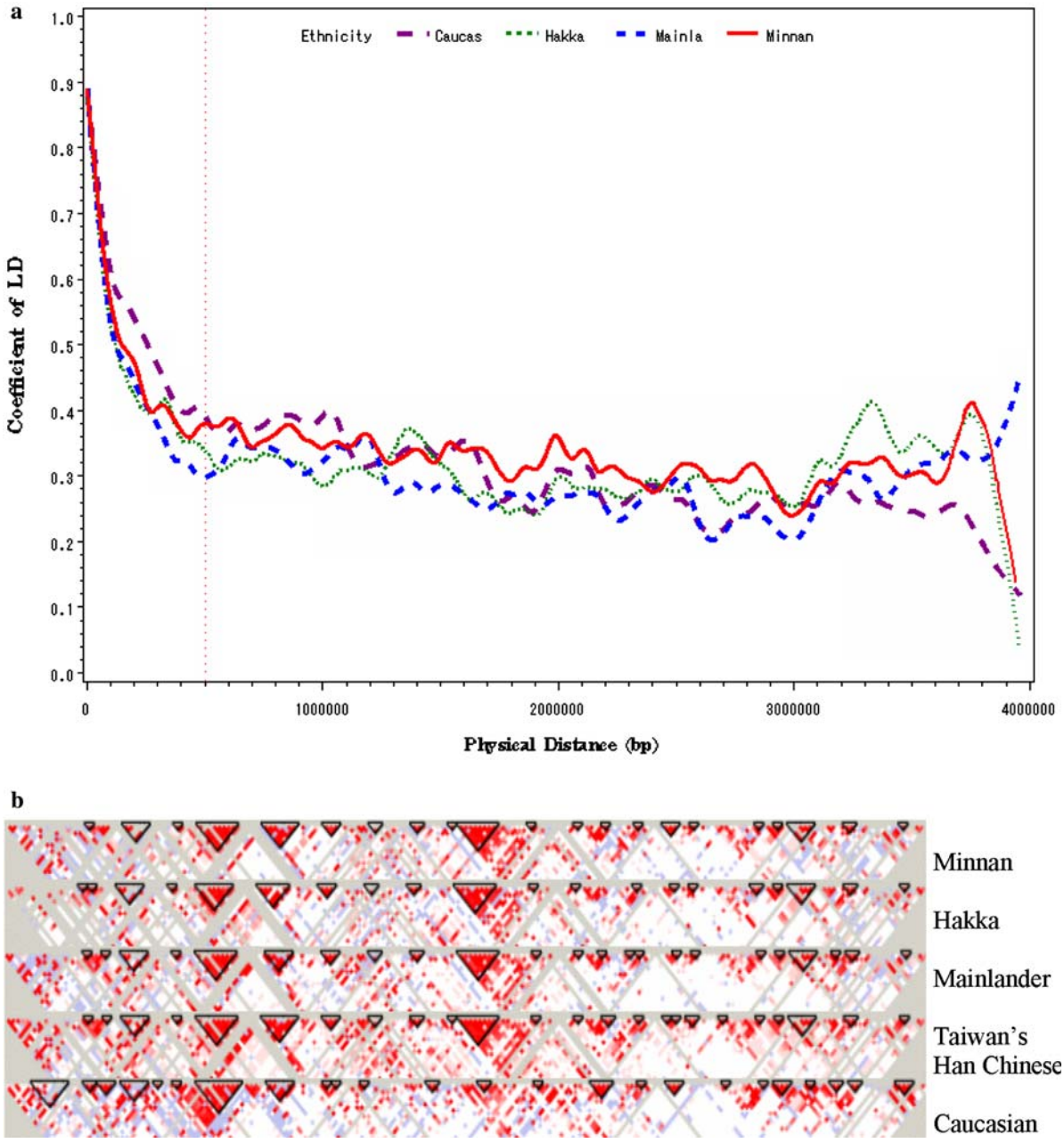


Figure 2. (a) Relationship between LD and physical distances of the intermarkers. The obvious decay of LD with physical distance was apparent in all samples within an intermarker distance of 500 kb (as shown by the dotted reference line). The Caucasian group had the smallest slope, followed by the Minnan, Hakka, and Mainlander groups. (b) LD maps and haplotype blocks for different populations. The blocks were constructed based on the confidence interval approach using the software HAPLOVIEW [20]. The red triangles denote the haplotype blocks and white areas represent evidence of strong recombination.

Table 1. Summary statistics of haplotype blocks in different populations.

Ethnicity	No. of blocks	Mean of block size (bp)	Median of block size (bp)	Std of block size (bp)	Maximum block size (bp)	Minimum block size (bp)	Coverage rate	No. of htSNPs	Tagging efficiency (%)
Minnan	20	34,551	16,494	39,782	135,800	634	0.174	48	71
Hakka	18	34,499	16,037	39,730	135,800	634	0.157	41	67
Mainlander	21	29,555	11,874	37,098	135,800	634	0.157	49	74
Taiwan's Han Chinese	24	29,710	11,817	37,606	135,800	634	0.180	54	73
Caucasian	22	34,940	25,875	32,254	121,583	4031	0.194	60	81

(Supplementary Table 2 online). The gene-disease association was summarized from Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nih.gov/entrez/query.fcgi?db = OMIM); Genetic Association Database (GAD) (http://www.geneticassociationdb.nih.gov/cgi-bin/index.cgi), and MHC studies [9, 11, 18]. Comparing the prevalence of inherited diseases in Han Chinese and Caucasian populations, atherosclerosis related to *AIF1* and psoriasis related to *C6orf18* are more prevalent in Caucasians [26, 27]; allopurinol-induced severe cutaneous adverse reaction related to *BAT3* [18] and hepatitis B virus related to *HLA-DRA* are more prevalent in Han Chinese [28, 29]. Furthermore, we also investigated genomic regions highly linked to the 101 SNPs and found that some are related to the etiology of diseases such as nasopharyngeal carcinoma, which is highly prevalent in Asian populations. Its prevalence in the southern Han is 25 times higher than that in Caucasians [30]. Using Taiwanese patients, the susceptibility locus has been previously mapped to an *HLA-A*-containing segment between microsatellite markers D6S211 and D6S510 within a 132 kb segment [31]. Of the 185 total SNPs in our study, we found three successive SNPs (rs2517862, rs1611750, and rs1655930) having significant AF differences between the Taiwanese and Caucasian samples within this gene-rich segment. These results may be related to the observed differences in prevalence and linkage signal of candidate genes after properly adjusting for environmental impacts.

Fisher exact tests identified 13 SNPs with unequal genotypic distributions among the three Taiwanese subgroups after adjusting for the FDR. Among these SNPs, follow-up pair-wise comparisons between the Minnan, Hakka, and Mainlanders further identified three, seven, and nine SNPs significantly different in genotypic distributions (Supplementary Table 3 online). SNPs rs2021722, rs2763979 and rs2075799 had $-\log_{10}(p\text{-value}) > 5$. SNP rs2021722 is located within *TRIM26*, which has been found to directly interact with hepatocyte nuclear factor 4- α, whose misregulation contributes to type 2 diabetes [32]. SNP rs2763979 is located within *HSPA1B*, which also has susceptibility implications in type 2 diabetes [33] and is prevalent in southern Asians [34]. SNP rs2075799 is located within *HSPA1L*, which is related to type 1 diabetes [35].

We also compared the polymorphic status of the different groups (Supplementary Table 4 online). Thirteen percent of the polymorphic SNPs in the Caucasian samples could not be validated in the Taiwanese samples (Supplementary Table 4 online), in which eight SNPs (rs3132129, rs3094628, rs3094054, rs3129820, rs3129973, rs558702, rs433061, and rs1269852) had only one type of allele and where three SNPs (rs3132129, rs558702, and rs433061) were within the functional genes *C6orf12*, *C6orf46*, and *TNXB* (Supplementary Table 5 online). The three Taiwanese subgroups had less than 7.5% inconsistency rates of polymorphism status (MAF > 0.01), and all pair-wise concordance indices, kappa (κ), were greater than 0.66. In the case of 17 distinct SNPs whose polymorphism status differed among the Taiwanese subgroups, 9 of the 17 were located in functional regions (Supplementary Table 6 online). Among the markers having inconsistent polymorphic status among the different groups, the maximum difference in AFs was 0.06, which was smaller than that between the Taiwanese and Caucasians (0.21). These analyses further reflected the differences in genetic background between the Taiwanese and Caucasian populations.

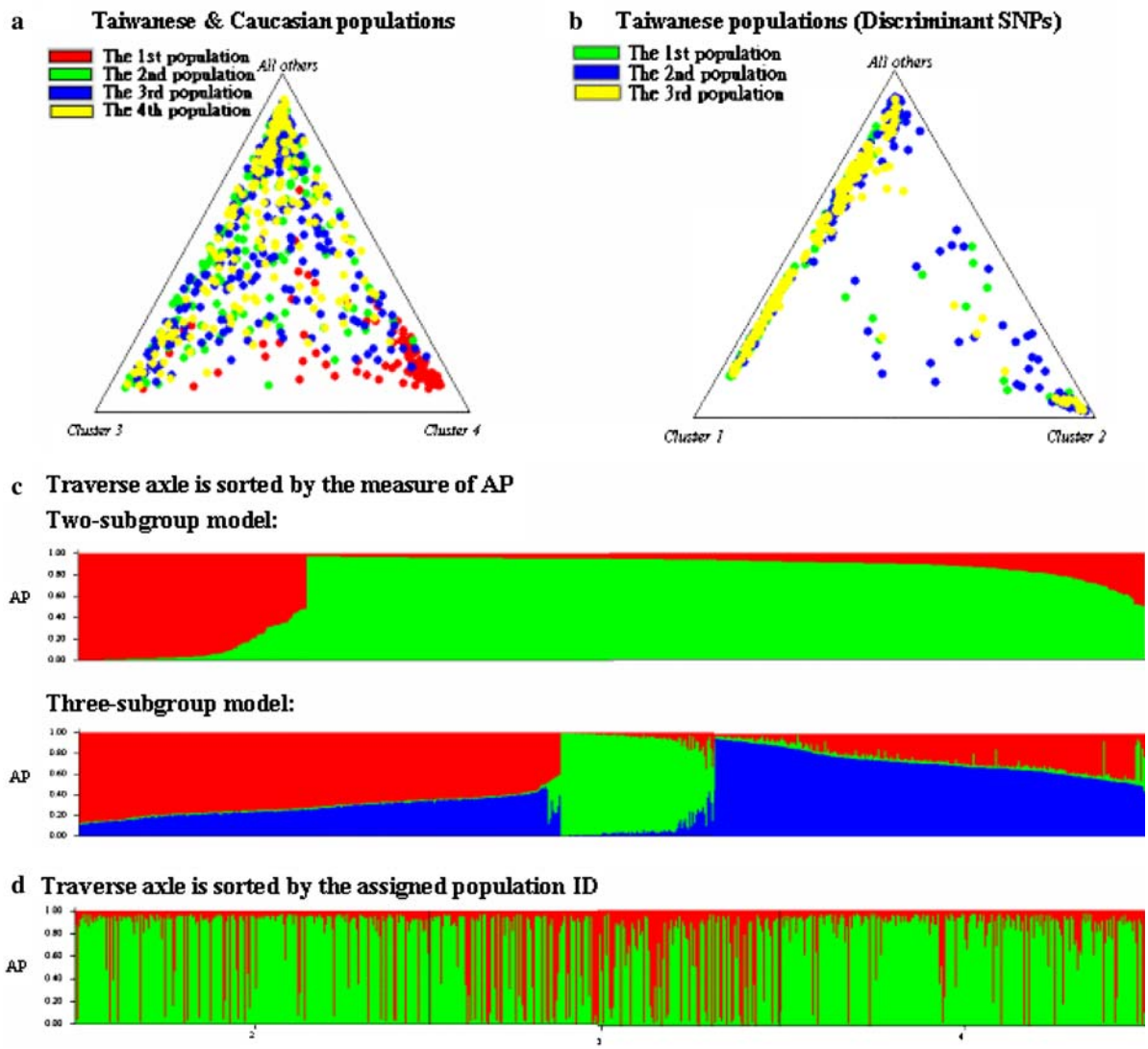


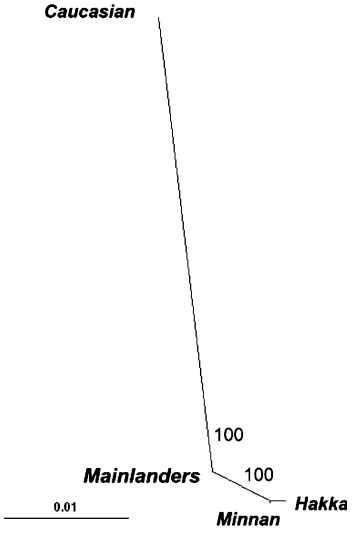
Figure 3. The estimated individual admixture proportion (AP) of different populations. (a) Admixture proportions for each individual in the Caucasian and three Taiwanese subgroups were estimated based on the algorithm STRUCTURE [23]. The Caucasian individuals (red) were grouped together and were assigned to a cluster. The Hakka (green), Mainlander (blue), and Minnan (yellow) residents were separated from the Caucasians but formed a mixture. (b) The 13 discriminant SNPs identified in Supplementary Table 3 were used to estimate individual admixture proportions. The Mainlander residents (blue) were slightly different from the other two Taiwanese subgroups. (c) Individuals were sorted according to the individual admixture proportions, for which two-subgroup model and three-subgroup model were assumed in the graphics in the top and bottom panels, respectively. The individual admixture proportions under the two-subgroup model can assign an individual to a group with less ambiguity relative to the three-subgroup model. (d) Under the two-subgroup model, i.e., non-Mainlander and Mainlander, many Mainlander residents can be separated from the other two Taiwanese subgroups.

To understand the relationships of multiple loci among different groups across the human MHC region, we performed structure analyses using LD based on markers having an MAF greater than 0.01 that also followed the Hardy–Weinberg equilibrium. In general, LD decayed with increasing intermarker distance, but the patterns were different between the study samples (Figure 2a). Within the region where the intermarker distance was less than 500 kb, the Caucasian samples decayed slower than the Taiwanese subgroup samples, among which the Minnan had the lowest slope.

a Phylogenetic distances among the populations of interest.

Populations	Hakka	Mainlanders	Southern Mainlanders	Northern Mainlanders	Caucasian
Minnan	0.001	0.003	0.002	0.006	0.043
Hakka		0.003	0.002	0.007	0.044
Mainlanders				---	0.034
Southern Mainlanders				0.005	0.038
Northern Mainlanders					0.031

b Phylogenetic tree among the Taiwanese and Caucasian populations.



c Phylogenetic tree among the Caucasians, Minnan, Hakka, southern Mainlanders and northern Mainlanders.

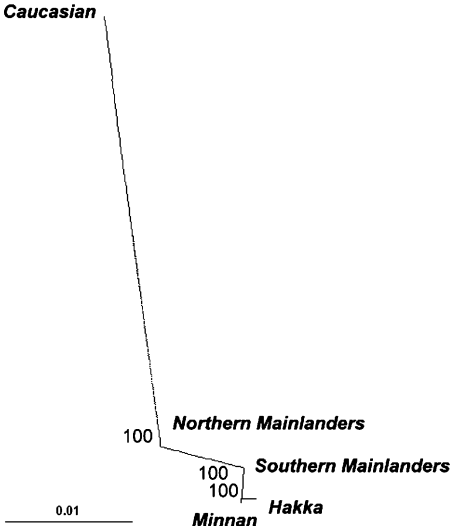


Figure 4. The phylogenetic analysis. (a) The distance between any two populations was calculated based on the Nei's distance [25], which was used to construct the phylogenetic tree based on the neighbor joining method using the software, PHYLIP [24], with 10,000 bootstrap replications. (b) The analysis shows the large difference between the Taiwanese and Caucasian populations. (c) The Minnan and Hakka are closer to each other relative to the Mainlanders, whereas the southern Mainlanders is closer to the cluster of the Minnan and Hakka relative to the northern Mainlanders.

The decay in each group was retarded after the threshold, and the LD fluctuated. Figure 2b presents the profiles of haplotype blocks for the Caucasian and Taiwanese populations, and Table 1 shows the corresponding descriptive statistics. There were apparent differences in LD structure. The maximum haplotype block in the Taiwanese population was located between rs2844476 and rs1150793, with a width of 135.8 kb; the genes *AIF1*, *BAT3*, *CSNK2B*, *LY6G5C*, *BAT5*, *LY6G6C*, and *MSH5* lie within this block. However, in this region the Caucasians samples showed only a small haplotype block between rs805256 and rs805281, with a width of 25.8 kb. The summary results (Table 1) suggest that the Caucasian population had a larger median block size and a smaller standard deviation of the block relative to the Taiwanese groups. The coverage rate of the haplotype block in the Taiwanese and Caucasian samples was 18% and 19%, respectively. We identified the representative SNPs for haplotype blocks (htSNPs) in the different populations. The ratio of the number of htSNPs to the haplotype block size was larger for the Caucasian group, implying a lower SNP tag efficiency relative to the Taiwanese population. To examine the stability of

the LD map construction and selection of htSNPs, we performed the same analysis using data for MAF > 0.05 and obtained very similar results.

To examine the magnitude of population admixture within the Taiwanese populations, we calculated individual admixture proportions. With only a few exceptions, the Caucasians were located close to the lower-right corner of the triangle plot and were separated from the other clusters (Figure 3a). This result demonstrates a genetic demarcation between the Caucasian and Taiwanese samples. The three major Taiwanese subgroups comprised an indistinguishable mixture. Based on the 13 discriminating SNPs (Supplementary Table 3 online), we found a slight difference between the Mainlander subgroup and the other two subgroups (Figure 3b). Moreover, the comparison of two versus three Taiwanese subpopulations suggested that the Mainlander residents might belong to a lineage relatively remote from the mixed lineages of the Hakka and Minnan (Figure 3c and d). This observation was further confirmed using phylogenetic analysis. The resulting bootstrapped phylogenetic tree showed that the Minnan and Hakka were more closed related relative to the Mainlanders, and that the Caucasians were genetically distant from the Taiwanese population (Figure 4). The high bootstrap confidence values indicated that the same tree topology was obtained among all bootstrapped samples and demonstrated that the resultant phylogenetic tree was very reliable.

Discussion

A key component of association studies that are successful in identifying disease susceptibility genes is a good understanding of the genetic background of the study populations. Our present study investigated the ethnic heterogeneity between Han Chinese and Caucasians and provides an important genetic dissection of the Han Chinese residing in Taiwan, an epitome of the Asian population. The results have some important implications: (1) The large discrepancy of genetic distributions between the Taiwanese and Caucasian populations was found. This suggests that public databases should be used with caution in cases where the study population(s) differs. For the need of future Taiwan's genetic studies, establishment of Taiwan-specific genetic databases is critically important. (2) A few studied SNPs in the human MHC region with high between-population heterogeneity presented different genotypic distributions among Taiwanese subgroups. The results suggest that impact of population stratification in Taiwan's association mapping cannot be precluded completely. However, high homogeneity we observed among the Taiwanese subpopulations alleviates the possibility of false positive conclusions due to ignoring population stratification in Taiwanese disease gene association studies. (3) The useful information of dense SNPs and the associated disorders in the human MHC region are provided. Characterization of haplotype blocks and htSNPs in the study populations helps design future gene mapping experiments and improve statistical power in genetic association studies. (4) Differences in AF between the Taiwanese and Caucasian populations were found in some disease-associated loci. These loci are related to the etiology of the diseases whose disease prevalence is also different between the two populations. The relationship between these loci and disease prevalence is warranted for further investigation.

Acknowledgments

This project was funded by grants from the Taiwan National Science Council (NSC 93-2320-B-001-0.26 and NSC 94-311-B001-007-Y) and the Genomics and Proteomics Program of Academia Sinica (94F003-2).

Electronic supplementary material

Supplementary material is available for this article at http://dx.doi.org/10.1007/s11373-006-9077-7 and is accessible for authorized users.

References

1. Wang W.Y.S., Barratt B.J., Clayton D.G. and Todd J.A., Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6: 109–118, 2005.
2. Hirschhorn J.N. and Daly M.J., Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6: 95–108, 2005.
3. Maniatis N., Collins A. and Gibson J., Positional cloning by linkage disequilibrium. Am. J. Hum. Genet. 74: 846–855, 2004.

4. Freedman M.L., Reich D., Penney K.L., McDonald G.J., Mignault A.A., Patterson N., Gabriel S.B., Topol E.J., Smoller J.W., Pato C.N., Pato M.T., Petryshen T.L., Kolonel L.N., Lander E.S., Sklar P., Henderson B., Hirschhorn J.N. and Altshuler D., Assessing the impact of population stratification on genetic association studies. Nat. Genet. 36: 388–393, 2004.
5. Marchini J., Cardon L.R., Phillips M.S. and Donnelly P., The effects of human population structure on large genetic association studies. Nat. Genet. 36: 512–517, 2004.
6. Helgason A., Yngvadóttir B., Hrafnkelsson B., Gulcher J. and Stefánsson K., An Icelandic example of the impact of population structure on association studies. Nat. Genet. 37: 90–95, 2005.
7. Walsh E.C., Mather K.A., Schaffner S.F., Farwell L., Daly M.J., Patterson N., Cullen M., Carrington M., Bugawan T.L., Erlich H., Campbell J., Barrett J., Miller K., Thomson G., Lander E.S. and Rioux J.D., An integrated haplotype map of the human major histocompatibility complex. Am. J. Hum. Genet. 73: 580–590, 2003.
8. Stenzel A., Lu T., Koch W.A., Hampe J., Guenther S.M., De La Vega F.M., Krawczak M. and Schreiber S., Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. Hum. Genet. 114: 377–385, 2004.
9. Horton R., Wilming L., Rand V., Lovering R.C., Bruford E.A., Khodiyar V.K., Lush M.J., Povey S., Talbot C.C., Wright M.W., Wain H.M., Trowsdale J., Ziegler A. and Beck S., Gene map of the extended human MHC. Nat. Rev. Genet. 5: 889–899, 2004.
10. Miretti M.M., Walsh E.C., Ke X., Delgado M., Griffiths M., Hunt S., Morrison J., Whittaker P., Lander E.S., Cardon L.R., Bentley D.R., Rioux J.D., Beck S. and Deloukas P., A high-resolution linkage disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. Am. J. Hum. Genet. 76: 634–646, 2005.
11. Shiina T., Inoko H. and Kulski J.K., An update of the HLA genomic region, locus information and disease associations. Tissue Antigens 64: 631–649, 2004.
12. Meacham W., Origins and development of the Yueh coastal Neolithic: a microcosm of culture change of the mainland of East Asia, In: Keightly D.K. (Ed), The Origins of Chinese Civilization. University of California Press, California, 1981.
13. Fang H.K. (ed.), Deep Investigation of the Origin of Hakka. Woolin Publishing, Taipei, Taiwan, 1996.
14. Lin M., Chu C.C., Chang S.L., Lee H.L., Loo J.H., Akaza T., Juji T., Ohashi J. and Tokunaga K., The origin of Minnan and Hakka, the so-called ''Taiwanese'', inferred by HLA study. Tissue Antigens 57: 192–199, 2001.
15. Zhao T.M. and Lee T.D., Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese Nation. Hum. Genet. 83: 101–110, 1998.
16. Lo H.L. (ed.), An Introduction to the Study of the Hakkas in its Ethnic, Historical, and Cultural Aspects. SMC Publishing, Taipei, Taiwan, 1933.
17. Pan W.H., Fann C.S.J., Wu J.Y., Hung S.I., Hung Y.T., Chen Y.J., Hsu C.L., Liao C.J. and Chen, Y.T., Establishment of Taiwan Han Chinese cell and gene bank: comparing SNP profiles in MHC region with Caucasians. The American Society of Human Genetics 54th Annual Meeting, Toronto, Ontario, Canada, October 26–30, 2004.
18. Hung S.I., Chung W.H., Liou L.B., Chu C.C., Lin M., Huang H.P., Lin Y.L., Lan J.L., Yang L.C., Hong H.S., Chen M.J., Lai P.C., Wu M.S., Chu C.Y., Wang K.H., Chen C.H., Fann C.S.J., Wu J.Y. and Chen Y.T., HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. Proc. Natl Acad. Sci. USA 102: 4134–4139, 2005.
19. Benjamini Y. and Hochberg Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B Met. 57: 289–300, 1995.
20. Barrett J.C., Fry B., Maller J. and Daly M.J., Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265, 2005.
21. Lewontin R.C., The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49: 49–67, 1964.
22. Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E.S., Daly M.J. and Altshuler D., The structure of haplotype blocks in the human genome. Science 296: 2225–2229, 2002.
23. Pritchard J.K., Stephens M. and Donnelly P., Inference of population structure using multilocus genotype data. Genetics 155: 945–959, 2000.
24. Felsenstein J., Phylip: Phylogeny inference package, ver. 3.6, Software User Manual. University of Washington, Washington, 1995.
25. Nei M., Genetic distance between populations. Am. Nat. 106: 283–292, 1972.
26. Woo K.S., Chook P., Chan W.B., So W.Y., Cockram C.S. and Celermajer D.S., Type 1 diabetes and arterial dysfunction in asymptomatic Chinese adults. Diabetes Care 24: 173, 2001.
27. Lench N., Iles M.M., Mackay I., Patel R., Sagoo G.S., Ward S.J., Dechairo B., Olavesen M., Carey A., Duff G.W., Cork M.J. and Tazi-Ahnini R., Single-point haplotype scores telomeric to human leukocyte antigen-C give a high susceptibility major histocompatability complex haplotype for psoriasis in a Caucasian population. J. Invest. Dermatol. 124: 545–552, 2005.
28. Hung S.I., Chung W.H. and Chen Y.T., HLA-B genotyping to detect carbamazipine-induced Stevens–Johnson syndrome: implications for personalizing medicine. Pers. Med. 2: 225–237, 2005.
29. Bhimma R. and Coovadia H.M., Hepatitis B virus associated nephropathy. Am. J. Nephrol. 24: 198–211, 2004.
30. Lin J.C., Cherng J.M., Lin H.J., Tsang C.W., Liu Y.X. and Lee S.P., Amino acid changes in functional domains of latent membrane protein 1 of Epstein-Barr virus in nasopharyngeal carcinoma of southern China and Taiwan: prevalence of an HLA A2-restricted ‚epitope-loss variant'. J. Gen. Virol. 85: 2023–2034, 2004.
31. Lu C.C., Chen J.C., Tsai S.T., Jin Y.T., Tsai J.C., Chan S.H. and Su I.J., Nasopharyngeal carcinoma–susceptibility locus is localized to a 132 kb segment containing HLA-A using high-resolution microsatellite mapping. Int. J. Cancer 115: 742–746, 2005.
32. Odom D.T., Zizlsperger N., Gordon D.B., Bell G.W., Rinaldi N.J., Murray H.L., Volkert T.L., Schreiber J., Rolfe P.A., Gifford D.K., Fraenkel E., Bell G.I. and

Young R.A., Control of pancreas and liver gene expression by HNF transcription factors. Science 303: 1378–1381, 2004.
33. Zouari B.K., Chouchane L., Jellouli K., Cherif S., Haddad S., Gabbouj S. and Danguir J., Polymorphism of stress protein HSP70–2 gene in Tunisians: susceptibility implications in type 2 diabetes and obesity. Diabetes Metab. 30: 175–180, 2004.
34. Zimmet P., Alberti K.G. and Shaw J., Global and societal implications of the diabetes epidemic. Nature 414: 782–787, 2001.
35. Pociot F., Ronningen K.S. and Nerup J., Polymorphic analysis of the human MHC-linked heat shock protein 70 (HSP70–2) and HSP70-Hom genes in insulin-dependent diabetes mellitus (IDDM). Scand. J. Immunol. 38: 491–495, 1993.