

Similarity of the allele frequency and linkage disequilibrium pattern of single nucleotide polymorphisms in drug-related gene loci between Thai and northern East Asian populations: implications for tagging SNP selection in Thais

Surakameth Mahasirimongkol · Wasun Chantratita · Somying Promso · Ekawat Pasomsab · Natini Jinawath · Wallaya Jongjaroenprasert · Viraphong Lulitanond · Phanida Krittayapoositpot · Sissades Tongsima · Pathom Sawanpanyalert · Naoyuki Kamatani · Yusuke Nakamura · Thanyachai Sura

Received: 6 July 2006 / Accepted: 20 July 2006 / Published online: 7 September 2006
© The Japan Society of Human Genetics and Springer-Verlag 2006

Abstract The similarity of the marker allele frequency and linkage disequilibrium structure between two populations are major factors for the determination of the transferability and efficiency of haplotype tagging SNP derived from one population to use for an indirect association study in another population. To prove the similarity between northern East Asian populations in Hapmap and Thais, 861 SNP in 166 drug-related genes shared between Thais, Han Chinese and Japanese were analyzed for their correlation statistics. Allele frequency, *F_{st}* statistics and linkage

disequilibrium statistics (r^2) showed a high correlation between these populations. TagSNP sets derived by an aggressive tagging algorithm from these 861 SNP in Japanese and Chinese were used to test the coverage of East Asia-derived tagSNP in Thais. TagSNP derived from Japanese and Chinese are comparable in the percentage of coverage of the alleles captured with tagSNP at $r^2 \geq 0.8$ (93% vs. 93%) in these drug-related gene loci. Additional tagSNP sets derived from the combination of Japanese- and Chinese-derived tagging SNP sets were used to test the coverage in Thais. The

S. Mahasirimongkol (✉) · P. Sawanpanyalert
Department of Medical Sciences, Ministry of Public Health,
Center for International Cooperation,
Foreign Affairs Section, 3rd Floor, Building 1,
Nonthaburi 11000, Thailand
e-mail: surakameth@dmsc.moph.go.th

W. Chantratita · S. Promso · E. Pasomsab
Department of Pathology, Faculty of Medicine,
Ramathibodhi Hospital, Mahidol University,
Bangkok, Thailand

W. Jongjaroenprasert
Endocrine unit, Department of Medicine,
Faculty of Medicine, Ramathibodhi Hospital,
Mahidol University, Bangkok, Thailand

V. Lulitanond
Department of Microbiology, Khon Kaen University,
Khon Kaen, Thailand

P. Krittayapoositpot
Department of Internal Medicine, Suratthani hospital,
Suratthani, Thailand

S. Tongsima
National Center for Genetic Engineering
and Biotechnology, Pathumthani, Thailand

N. Kamatani
Institute of Rheumatology,
Tokyo Women's Medical University,
Tokyo, Japan

Y. Nakamura
Research Group for Personalized Medicine,
SRC, RIKEN, Yokohama, Japan

N. Jinawath · Y. Nakamura
Laboratory of Molecular Medicine,
Human Genome Center,
Institute of Medical Science, University of Tokyo,
Tokyo, Japan

T. Sura
Medical Genetics and Molecular Medicine Unit,
Department of Medicine, Faculty of Medicine,
Ramathibodhi Hospital, Mahidol University,
Bangkok, Thailand

later set improved the percentage of coverage of alleles captured with tagSNP at $r^2 \geq 0.8$ –98% for these sites. High similarity between Thais and northern East Asian allele frequency and linkage disequilibrium statistics supported that tagSNPs derived from the northern East Asian population should be useful for an indirect association study in Thais. The combination of non-overlapping Japanese derived tagSNP and Chinese-derived tagSNP improved the percentage of genomic coverage in Thais, at least in these drug-related gene loci.

Keywords Thai · Thailand · Allele frequency · Linkage disequilibrium · Hapmap · TagSNP · Transferability

Introduction

The additive effect of individual genetic variations in loci encoding metabolic enzymes, drug transporters, cell surface markers, and cellular growth and differentiation factors may play a significant role in the variability of the response and toxicity of a number of drugs. It is expected that pharmacogenomics will contribute to a better health care system through decreasing unavoidable side effects and increasing the efficacy of drugs (Mushiroda et al. 2005, 2006; Saeki et al. 2005). Research conducted in pharmacogenomics usually involves retrospective case-control association studies to determine the difference in allele frequencies of variations at markers of interest between the case and control populations.

Genetic markers selected for genotyping in pharmacogenomic studies are single nucleotide polymorphisms (SNP) because of their abundance in genomes and the maturity of high-throughput genotyping technology. For indirect association studies, several SNP selection methods have been developed based on linkage disequilibrium information to reduce the redundancy in genotyping information. These methods also aim at increased marker coverage of the genome. Current methods for marker selection with the highest level of coverage are the marker selection method based on linkage disequilibrium statistics derived from genotyping of common SNP in the target population; markers selected by this method are called TagSNP (Altshuler et al. 2005). Through international collaboration efforts like the International HapMap Project, linkage disequilibrium information has been made available for four major populations, the Han Chinese, Japanese, Caucasians and Yorubans (Deloukas and Bentley

2004). This information is deposited in web databases and is made accessible to all researchers. Applying tagSNP based on linkage disequilibrium information from Hapmap populations to other populations will result in a reduction of the power of the study unless the allele frequency and linkage disequilibrium structure between Hapmap populations and target populations are highly similar (Lim et al. 2006; Montpetit et al. 2006; Mueller et al. 2005; Ribas et al. 2006). Among five major populations, it is comprehensible that tagSNP set coverage will not be good enough to use tagSNP derived from one population for an indirect association study in other populations. The quantifying correlation of the allele frequency and the linkage disequilibrium pattern between the Thai and Hapmap northeast Asian population, and the reduction of tagging efficiency when using Japanese Hapmap tagSNP for an indirect association study in Thais, should be useful for the development of a better SNP selection strategy using Hapmap information.

In Thailand, the Pharmacogenomics Project was established in 2003 with the aim of promoting the research, development and application of pharmacogenomics in Thailand. Study of allele frequency and linkage disequilibrium of markers in drug-related gene loci is relevant to the objective of this project. Recently, information from extensive analysis of these genes in the Japanese population was made publicly available (Kamatani et al. 2004). In this study, we genotyped 1,536 SNP derived from the Japanese population by haplotype-based selection in 280 Thais from four different geographic locations in Thailand. These SNPs were derived by the genotyping of 4,104 markers discovered by the resequencing of 48 Japanese individuals in 752 Japanese individuals to capture common SNP (minimal allele frequency >0.1), linkage disequilibrium structure and haplotype inference (Iida et al. 2001). Haplotype tagging SNPs were selected from individual genotyping data with additional rare functional SNP chosen to increase the coverage for an association analysis in these genes (Iida et al. 2006). To try to understand the relationship between northern East Asian populations in the Hapmap database and the utilization of Hapmap database in Thais, allele frequency and haplotype frequency were determined for each population and comparisons between these populations were made. An empirical assessment of the tagging performance of Japanese tagSNP and Chinese tagSNP in Thais was shown. Results of this comparison study will benefit researchers who plan to conduct indirect association studies in Thais.

Materials and methods

Population samples

This study was reviewed by the Ethical Committee of the Faculty of Medicine, Ramathibodi Hospital, Mahidol University. After receiving informed consent, 280 individual blood samples were obtained for the genotyping study. Each of 70 individuals was ascertained to be from one of four regions representing four different parts of Thailand as follows: ChiangRai (a province in the northern region of Thailand), Supanburi and Kanjanaburi (provinces in the central region of Thailand), KhonKhaen (a province in the northeastern region of Thailand) and Surajthani (a province in the southern region of Thailand). Family histories were reviewed for consanguinity and relatedness of the individuals. Only individuals with similar self-reported parental ethnicity were selected for genotyping. The people of these four regional populations speak their local Thai dialects specific to their regions. Therefore, they were considered to represent four major different regions in Thailand and were believed to be good estimators of the distribution of the population allele frequency from each region.

Marker selection

Single nucleotide polymorphisms (SNP) were selected based on resequencing information of 166 drug-related genes in 48 Japanese individuals. Every novel SNP discovered was confirmed by sequencing both strands of each PCR product (Iida et al. 2001). SNPs confirmed by resequencing were genotyped in 752 samples of unrelated individual Japanese. Haplotype tagging SNPs were computed base on individual genotyping information of common SNPs (minor allele frequency >0.1). Additional rare, functional, and out-of-block SNPs were also included in this drug-related SNP set. Overall, 1,536-plex microarray SNP genotyping sets were established for the pharmacogenomics study in the Japanese population. We applied this drug-related 1,536-plex SNP microarray set to Thai populations.

SNP genotyping

DNA extraction was performed by a standard phenol-chloroform protocol at the Medical Genetics and Molecular Medicine Unit, the Faculty of Medicine, Ramathibodi Medical School in Thailand. DNA concentrations were measured and diluted to the final concentration of 50 ng/μl using the Picogreen DNA quantification kit. Genotyping of SNP was done using

the Illumina Goldengate genotyping platform. This platform utilized 1,536-plex of genotyping in a 96-well plate platform. After DNA samples from each individual were prepared on 96-well plates, SNP genotyping following the standard protocol (Gunderson et al. 2005) was performed.

In brief, DNA samples were activated for binding to paramagnetic particles. In the hybridization step, activated DNA samples were combined with assay oligonucleotides, hybridization buffer and paramagnetic particles. There are three oligonucleotides designed for each SNP locus. Two oligonucleotides are specific to each allele of the SNP site, called the allele-specific oligonucleotides (ASOs). A third oligonucleotide that hybridizes several bases downstream from the SNP site is the locus-specific oligonucleotide (LSO). All three oligonucleotide sequences contain regions of genomic complementary and universal PCR primer sites; the LSO also contains a unique address sequence that targets a particular bead type. The set of 1,536-plex SNP was chosen based on a pharmacogenomics discovery by the Japanese millennium project as previously described (Iida et al. 2006). During the primer hybridization process, the assay oligonucleotides hybridize to the genomic DNA samples bound to paramagnetic particles. Extension of the appropriate ASO and ligation of the extended product to the LSO linked information about the genotype present at the SNP site to the address sequence on the LSO. These linked, full-length products provide a template for PCRs using the universal PCR primers P1, P2, and P3. The universal PCR primers P1 and P2 are Cy3- and Cy5-labeled. After downstream-processing, the single-stranded, dye-labeled DNAs were hybridized to their complement bead type through their unique address sequences. Hybridization of the GoldenGate assay products onto the array matrix allows for the separation of the assay products in solution onto a solid surface for individual SNP genotype readouts. After hybridization, a microarray scanner was used to analyze the fluorescence signal using software for automated genotype clustering and calling (Gunderson et al. 2004).

Statistical analyses

Japanese Hapmap (Japanese) and Han Chinese Hapmap (Chinese) genotyping data of these SNP were retrieved from the Hapmap website (<http://www.hapmap.org>) by query of the Hapmap release 20. Overall, 1,312 markers were successfully genotyped in the Thai population. Complete failure to obtain the genotype in Thais occurred only for 84 markers. In February 2006, there were only 870 SNPs with individual genotyping

information available in the Han Chinese population and Japanese population SNP database from the Hapmap database that were successfully genotyped in at least one of the regional Thai populations. Nine SNPs were eliminated in our analysis because of possible errors in genotyping reported in the Hapmap database, as remarkable differences in allele frequencies between the Hapmap database and Japanese Pharmacogenomic SNP Consortium database were detected. Comparison of population genetic marker characteristics between the Japanese and Chinese with Thais was done using 861 SNPs.

The Hardy-Weinberg equilibrium (HWE) test for each SNP within each population was calculated by standard χ^2 statistics, and SNPs with HWE test *P* values <0.001 were excluded from further analysis. Pair-wise linkage disequilibrium statistics (r^2 , D') and the χ^2 test for differences of allele frequency and haplotype frequency were calculated by HAPLOVIEW (Barrett et al. 2005). Pair-wise and hierarchical *Fst* statistics were calculated with the hierfstat package in *R* (Goudet 2005). Pearson's product moment correlation, scatter plots of pair-wise population allele frequency and scatter plots of pair-wise population linkage disequilibrium statistics were done using the *R* statistics program (Team 2004).

TagSNPs were selected using an aggressive tagging strategy by TAGGER panel in HAPLOVIEW. LOD score cutoff at three was used in all tagging selection. Fold saving of tagSNP is defined as n/n_h . In this study, n_h is the number of SNP selected to cover the region and n is the total number of SNPs.

Results

Comparison of allele frequency distribution

Considering MAF comparison, most of the MAF distribution was similar between Thais and other popu-

Table 1 Minor allele frequency (MAF) distribution

	Japanese ^a	Chinese ^a	Thais
MAF = 0	20 (2.3%)	17 (2.0%)	15 (1.7%)
>0–0.1	63 (7.3%)	83 (9.6%)	61 (7.1%)
>0.1–0.2	190 (22.1%)	205 (23.8%)	186 (21.6%)
>0.2–0.3	234 (27.2%)	219 (25.4%)	265 (30.8%)
>0.3–0.4	176 (20.4%)	171 (19.9%)	169 (19.6%)
>0.4–0.5	178 (20.7%)	166 (19.3%)	165 (19.2%)
Total	861	861	861

^a SNP genotyping information retrieved from the Hapmap database

lations. As shown in Table 1, there were only 15 SNPs (1.74%) in the Japanese found to be homozygous in the Thais. Twenty SNPs were homozygous from the Japanese Hapmap database. These polymorphisms were selected based on non-homozygosity in allele frequency determination in 752 Japanese samples; therefore, these SNPs were supposed to be heterozygous in Japanese. These homozygosities reported in the Hapmap database can be explained by the difference in sample sizes of the samples used for the determination of allele frequency (752 individual samples in the Pharmacogenomic SNP Consortium versus 44 Japanese samples used in the Hapmap project). In other words, if the Japanese Hapmap data were used, 20 markers would be excluded from the genotyping because of their reported homozygosity due to the small size of the samples used to obtain allele frequency.

Differences between Japanese and Chinese population genetic characteristics were considered insignificant and were analyzed as one population in the Hapmap data analysis (Altshuler et al. 2005). Correlation statistics calculated by a comparison of Japanese and Chinese SNP genotyping information were used in our analysis as the draft threshold to quantify differences between two populations, especially within northern East Asian populations. While the correlation statistics for Japanese and Chinese allele frequency is 0.9514, the correlation statistics are 0.9493 when comparing between Japanese and Thais, and 0.9298 when comparing Chinese and Thais. There were trends of higher correlations between the Chinese and all regional Thai populations than of those between the Japanese and all regional Thai populations (Fig. 1, lower diagonal matrix).

Genetic distance between regional Thai populations and northern East Asian populations

Pair-wise *Fst* statistics (Cockerham and Weir 1986) values for 861 markers show the same patterns of population distance with lower *Fst* values for comparison within regional Thai population pair-wise *Fst* statistics than pair-wise *Fst* statistics of Japanese–Chinese populations (Table 2). Interestingly, southern and northeastern Thais were more distant from northern East Asian populations than northern and central Thais. Lower pair-wise *Fst* values between Japanese and Thais compared to Chinese and Thais could possibly be explained by the nature of these SNP sets where they originated from Japanese tagging selection. Hierarchical *Fst* statistics for these markers in Thais are 0.00086.

Fig. 1 Scatter plots of pair-wise comparison of allele frequency distribution between Japanese, Chinese, and regional Thai populations: northern Thai, central Thai, southern Thai and northeastern Thai populations are shown. Pearson's product moment correlation of allele frequency between two populations presented in *lower diagonal matrix*

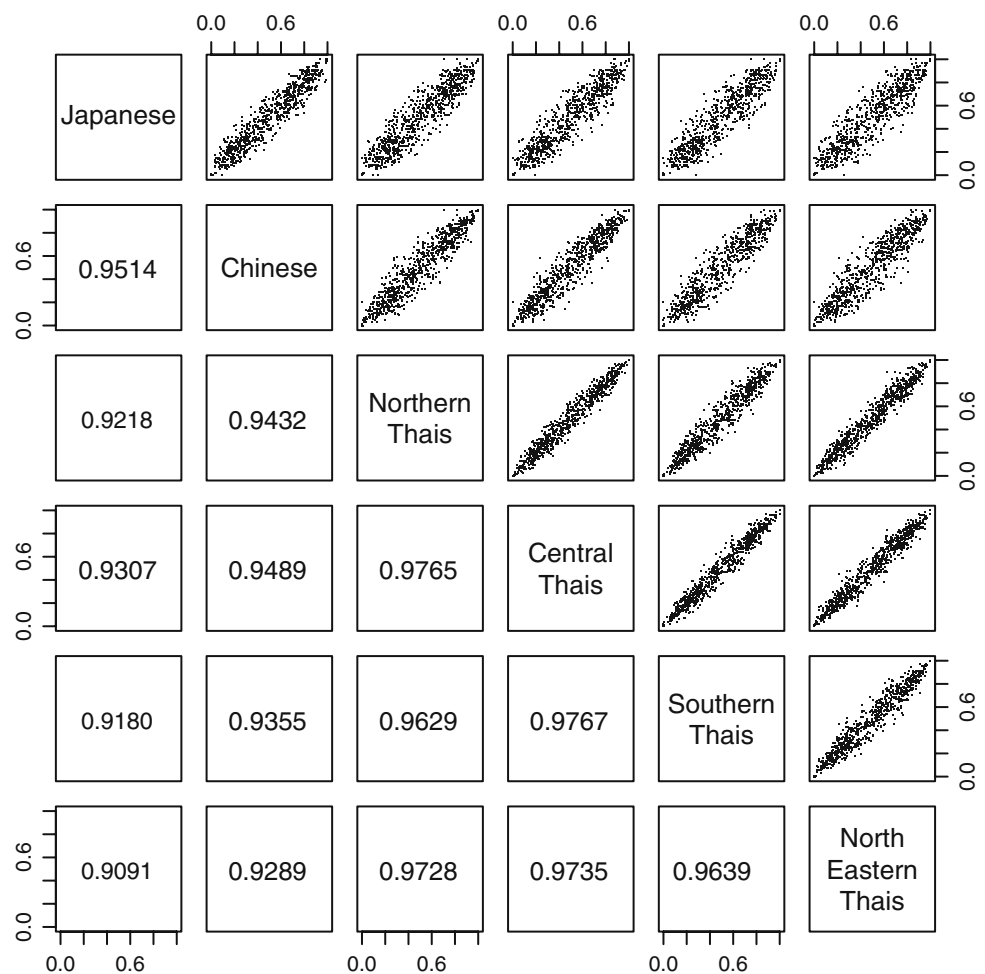


Table 2 Pair-wise *Fst* statistics between populations

	Chinese	Northern Thai	Central Thai	Southern Thai	Northeastern Thai
Japanese	0.00752	0.01234	0.01187	0.01502	0.01384
Chinese		0.01781	0.01863	0.01983	0.02063
Northern Thai			0.00107	0.00122	0.00013
Central Thai				0.00165	0.00005
Southern Thai					0.00315

HWE testing

Hardy-Weinberg equilibrium tests were tested against genotyping results. Genotyping results with the HWE test with $P < 0.001$ were excluded from further analysis. All 861 SNPs passed the HWE test, being significant at $P < 0.001$.

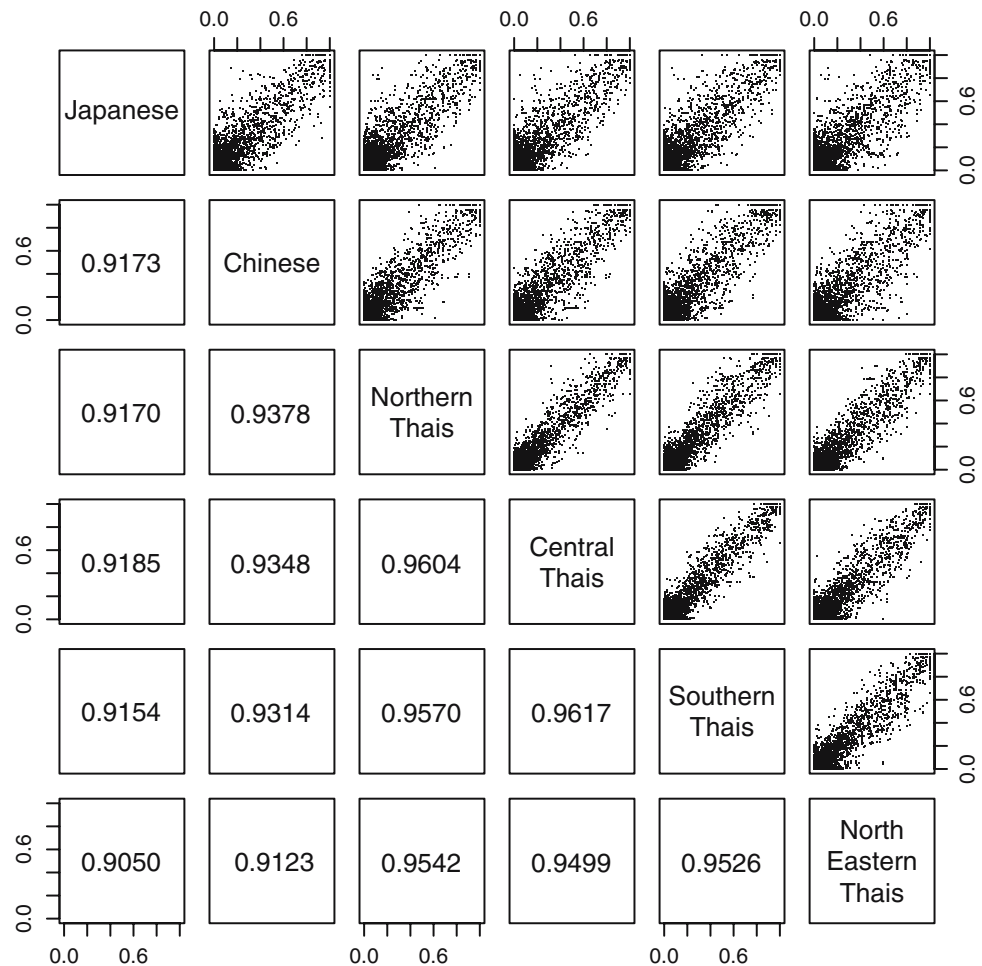
Linkage disequilibrium measure and comparison

The r^2 statistics were calculated for pairs of SNPs with a distance less than 500 kb; pair-wise r^2 were calculated and compared among regional Thai populations, Chinese and Japanese populations using only SNPs that were shared by all populations. In total, 691 SNPs contributed to the 8,419 pair-wise linkage disequilibrium statistics in these comparisons. Scatter plots comparing r^2 statistics between populations are presented in Fig. 2. The correlation of r^2 statistics between Chinese and regional Thai populations is higher than the correlation between Japanese and regional Thai populations. Within regional Thai populations, r^2 statistical correlations are higher than correlation statistics between Japanese and Chinese correlations.

D' at the difference distance

D' statistics calculated from 691 SNPs revealed the same pattern with rapid decay of D' statistics when

Fig. 2 Scatter plots of pairwise comparison of r^2 statistics among Japanese, Chinese and regional Thai populations: northern Thai, central Thai, southern Thai and northeastern Thai populations are shown in this figure with Pearson's product moment correlation of r^2 statistics between two populations presented in lower diagonal matrix



physical distance between pairs of SNPs increased. At close physical distance with less than 100 base pair intervals, there is more variation in D' statistics resulting in lower mean D' statistics for these markers. These findings are similar to those observed in Malays (Cha et al. 2004).

Haplotype inference

Because allele frequency and linkage disequilibrium characters between different regional Thai populations are highly correlated (Figs. 1, 2), further analysis was done by combining regional Thai populations together. Haplotype inference using Gabriel et al.'s haplotype definition (Gabriel et al. 2002) inferred the total number of blocks at 111, 114, and 179 blocks in Japanese, Chinese, and Thais, respectively. The number of haplotype blocks inferred by this method is higher in Thais compared to northern East Asian populations. The greater number of haplotype blocks in our population may reflect the older age of the Thai population than the northern East Asian population. (Fig. 3)

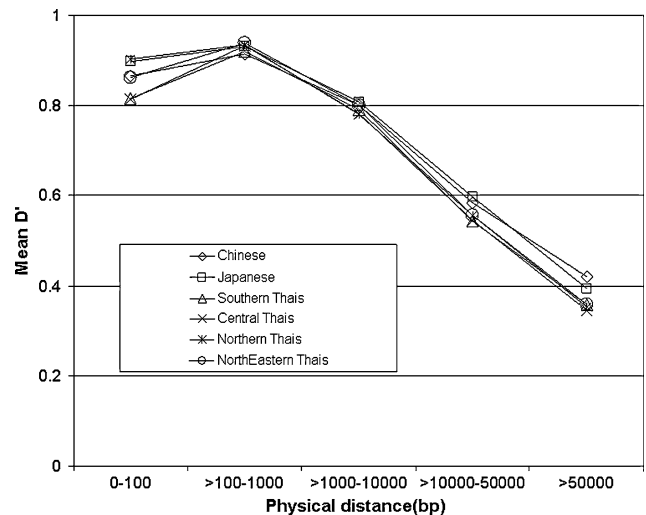


Fig. 3 Relationship between D' statistics and physical distance separating two single nucleotide polymorphisms in a pair

TagSNP selection and tagging efficiency in Thais

An aggressive tagging algorithm was used to select the tagSNP set within 861 SNPs from Japanese, Chinese,

Table 3 Number of TagSNPs computed from northern East Asian populations and percentage of alleles captured with $r^2 \geq 0.8$ in Thais

	Japanese	Chinese	Japanese + Chinese
Number of tagSNPs	618	624	724
Mean r^2	0.965	0.964	0.985
Percentage of alleles captured with $r^2 \geq 0.8$	93%	93%	98%
Mean r^2 for alleles with $r^2 < 0.8$	0.644±0.020	0.608±0.031	0.629±0.013

and Thai genotyping data (de Bakker et al. 2005). Because these are mostly haplotype tagging SNPs, the fold savings in the numbers of SNPs are 618/861 (1.39 folds), 624/861 (1.38 folds) and 605/861 (1.42 folds). These fold reduction numbers are different from data generated from tagging SNP selection using pre-defined interval genotyping data like in Hapmap (usually between three- and four-fold savings in Asian populations). However, the nature of transferability of these SNPs should not be different from interval genotyping data. We tested the coverage of the tagSNP derived from Japanese, Chinese, and a combination of Japanese and Chinese (Japanese + Chinese) tagSNPs with the Thai data. The results are shown in Table 3.

Difference in allele frequency and haplotype frequency between Thais and northern East Asian populations

Permutation chi-square statistics were used to determine the northern East Asian and Thai differences in allele frequency and haplotype frequency with 100,000 permutations. Only 5% of these sites have significant allele frequency differences between the two populations at $P < 0.05$. For haplotype frequency, 12% of the haplotype blocks inferred have a significant difference at $P < 0.05$. A full list of allele frequency, haplotype frequency, and their chi square statistics is available at (<http://www.pharmagtc.net/perm861.htm>).

Discussion

For the genetic association study in the population not included in the Hapmap database, a demonstration of the similarity and transferability of tagSNPs from the populations in the Hapmap database to their target populations will be useful for the researcher to choose the tagging SNP from the Hapmap database for their indirect association study. A transferability study in

Caucasians and Asians was encouraging, with evidence supporting the transferability for the Hapmap population-derived tagSNPs to their populations (Cha et al. 2004; Lim et al. 2006; Mueller et al. 2005; Ribas et al. 2006). We demonstrated here the similarity of Thais with East Asian Hapmap populations using 280 individuals. These individuals were a sampling from the four major regions of Thailand, and these genotyping data could also be used to explain the within-population variation of Thais.

Genotyping data were generated by microarray genotyping of 1,536-plex drug-related gene SNP sets from the Japanese population. With an admirable extensive effort for the genotyping of 752 unrelated Japanese individuals, Katamani and Nakamura et al. showed that haplotype-based SNP selection from these samples could be used to capture an association between traits and SNPs with allele frequency > 0.03 (Kamatani et al. 2004). The 1,536-plex SNP set was established using information from this analysis. Using these genotyping results to determine the population genetics data will not be straightforward because the majority of these SNPs are derived from haplotype tagging SNP selection in the Japanese population. But the inter-population genetics parameters analyzed here should not be affected by the bias introduced by the marker selection strategy.

Evidence from the Y-chromosome haplotype and mitochondrial markers supported that mainland southern East Asia was probably the first settlement of modern humans from Africa about 18,000 to 60,000 years ago (Chu et al. 1998; Horai 1995; Jin and Su 2000; Ke et al. 2001). After the settlement, mainland southern East Asians migrated northward to China and Siberia (Su et al. 1999). There was also a contribution from central Asia to the present-day northern East Asian population (Hammer 1995; Qian et al. 2001), especially the Japanese. There was also evidence of a high similarity of the Malaysian population with the Japanese for polymorphisms within drug-related genes (Cha et al. 2004).

A similarity between populations was assessed by pair-wise allele frequency correlation statistics, *Fst* statistics, and pair-wise linkage disequilibrium statistics between Chinese, Japanese, and regional Thai populations. Within these sites, four regional Thai populations are highly similar and were analyzed as one population in the later analysis. Within northern East Asian populations (Japanese and Chinese), the comparisons were more similar with their correlation statistics and *Fst* statistics than the comparisons between northern East Asians and the regional Thai population. The Chinese were closer to the Thais than the

Japanese, and these results are in concordance with the population history and previous mitochondrial marker analysis (Fucharoen et al. 2001). Tagging transferabilities were demonstrated in a European population with a Caucasian Hapmap database with a comparative level of correlations (Ribas et al. 2006). The difference between northern East Asians and Thais was also determined by chi-square statistics; 5% of these allele frequencies and 12% of haplotype frequencies are different at $P < 0.05$.

Tagging SNP performance in Thais was assessed by the percentage of coverage of the markers with $r^2 \geq 0.8$ when using tagging SNPs from other populations in Thais. The tagging SNP was selected using an aggressive pair-wise tagging approach in the TAGGER program. The percentage of coverage of markers with $r^2 \geq 0.8$ was at 93% for Japanese-derived tagging SNP and 93% for Chinese-derived tagging SNPs. The percentage of coverage was increased to 98% when using tagging sets selected by a combination of tagSNP sets from Japanese and Chinese. While the percentage of coverage may not hold true if Thais have common SNPs that are not genotyped in the Hapmap database, it is unlikely that the Thai population will have a large number of common SNPs that are not shared with the Japanese and Chinese populations. The increase of coverage may be explained by the fact that the Japanese and Chinese were considered in separate clades of Asians (Qian et al. 2001). At least, using a combination of tagging SNPs from the Chinese and Japanese Hapmap database should be useful to capture the linkage disequilibrium structure in an association study by the genotyping of a small number of samples to determine the linkage disequilibrium structure for the selection of markers to use in an indirect association study. Since this linkage disequilibrium-based tagSNP selection algorithm will pick different markers from a bin each time it runs, further study for the evaluation of a combination of tagSNP from Hapmap populations for use in an indirect association study for other Asian populations should be explored.

Acknowledgments We would like to acknowledge the Thailand Center for Excellence in Life Sciences for funding support to S.M. and this research. We would also like to thank Matsuhiro Shinoda, Kazuna Morohoshi, and Susumu Saito for their discussion and support on genotyping.

References

- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. *Nature* 437:1299–320
- de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. *Nat Genet* 37:1217–1223
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Cha PC, Yamada R, Sekine A, Nakamura Y, Koh CL (2004) Inference from the relationships between linkage disequilibrium and allele frequency distributions of 240 candidate SNP in 109 drug-related genes in four Asian populations. *J Hum Genet* 49:558–572
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95:11763–11768
- Cockerham CC, Weir BS (1986) Estimation of inbreeding parameters in stratified populations. *Ann Hum Genet* 50 (Pt 3):271–281
- Deloukas P, Bentley D (2004) The HapMap project and its application to genetic studies of drug response. *Pharmacogenomics J* 4:88–90
- Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet* 46:115–125
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
- Goudet J (2005) hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* 5:184–186
- Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E, Bierle J, Doucet D, Milewski M, Yang R, Siegmund C, Haas J, Zhou L, Oliphant A, Fan JB, Barnard S, Chee MS (2004) Decoding randomly ordered DNA arrays. *Genome Res* 14:870–877
- Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 37:549–54
- Hammer MF (1995) A recent common ancestry for human Y chromosomes. *Nature* 378:376–378
- Horai S (1995) Evolution and the origins of man: clues from complete sequences of hominoid mitochondrial DNA. *Southeast Asian J Trop Med Public Health* 26 (Suppl 1):146–154
- Iida A, Saito S, Sekine A, Kitamoto T, Kitamura Y, Mishima C, Osawa S, Kondo K, Harigae S, Nakamura Y (2001) Catalog of 434 single-nucleotide polymorphisms (SNP) in genes of the alcohol dehydrogenase, glutathione S-transferase, and nicotinamide adenine dinucleotide, reduced (NADH) ubiquinone oxidoreductase families. *J Hum Genet* 46:385–407
- Iida A, Saito S, Sekine A, Takahashi A, Kamatani N, Nakamura Y (2006) Japanese single nucleotide polymorphism database for 267 possible drug-related genes. *Cancer Sci* 97:16–24
- Jin L, Su B (2000) Natives or immigrants: modern human origin in east Asia. *Nat Rev Genet* 1:126–133
- Kamatani N, Sekine A, Kitamoto T, Iida A, Saito S, Kogame A, Inoue E, Kawamoto M, Harigai M, Nakamura Y (2004) Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP Maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNP within haplotype blocks and the haplotypes constructed with haplotype-tagging SNP. *Am J Hum Genet* 75:190–203

- Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, Xiao C, Shriver M, Lell J, Wallace D, Wells RS, Seielstad M, Oefner P, Zhu D, Jin J, Huang W, Chakraborty R, Chen Z, Jin L (2001) African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 292:1151–1153
- Lim J, Kim YJ, Yoon Y, Kim SO, Kang H, Park J, Han AR, Han B, Oh B, Kimm K, Yoon B, Song K (2006) Comparative study of the linkage disequilibrium of an ENCODE region, chromosome 7p15, in Korean, Japanese, and Han Chinese samples. *Genomics* 87:392–398
- Montpetit A, Nelis M, Laflamme P, Magi R, Ke X, Remm M, Cardon L, Hudson TJ, Metspalu A (2006) An evaluation of the performance of Tag SNP derived from HapMap in a Caucasian Population. *PLoS Genet* 2:e27
- Mueller JC, Lohmusaar E, Magi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, Schreiber S, Pramstaller P, Pichler I, Romeo G, Gaddi A, Testa A, Wichmann HE, Metspalu A, Meitinger T (2005) Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am J Hum Genet* 76:387–398
- Mushiroda T, Saito S, Tanaka Y, Takasaki J, Kamatani N, Beck Y, Tahara H, Nakamura Y, Ohnishi Y (2005) A model of prediction system for adverse cardiovascular reactions by calcineurin inhibitors among patients with renal transplants using gene-based single-nucleotide polymorphisms. *J Hum Genet* 50:442–447
- Mushiroda T, Ohnishi Y, Saito S, Takahashi A, Kikuchi Y, Shimomura H, Wanibuchi Y, Suzuki T, Kamatani N, Nakamura Y (2006) Association of VKORC1 and CYP2C9 polymorphisms with warfarin dose requirements in Japanese patients. *J Hum Genet* 51(3):249–253
- Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S (2001) Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J Hum Genet* 46:211–220
- Ribas G, Gonzalez-Neira A, Salas A, Milne RL, Vega A, Carracedo B, Gonzalez E, Barroso E, Fernandez LP, Yankilevich P, Robledo M, Carracedo A, Benitez J (2006) Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum Genet* 118:669–679
- Saeki M, Saito Y, Jinno H, Sai K, Hachisuka A, Kaniwa N, Ozawa S, Kawamoto M, Kamatani N, Shirao K, Minami H, Ohtsu A, Yoshida T, Saijo N, Komamura K, Kotake T, Morishita H, Kamakura S, Kitakaze M, Tomoike H, Sawada J (2005) Genetic variations and haplotypes of UGT1A4 in a Japanese population. *Drug Metab Pharmacokinet* 20:144–151
- Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet* 65:1718–1724
- Team, RDC (2004) R: A language and environment for statistical computing. Foundation for Statistical Computing