Population genetic structure of variable drug response

James F. Wilson[1,2], Michael E. Weale[3,4], Alice C. Smith[1], Fiona Gratrix[1], Benjamin Fletcher[3], Mark G. Thomas[3], Neil Bradman[3] & David B. Goldstein[1]

Published online: 29 October 2001, DOI: 10.1038/ng761

© 2001 Nature Publishing Group http://genetics.nature.com

Geographic patterns of genetic variation, including variation at drug metabolizing enzyme (DME) loci and drug targets, indicate that geographic structuring of inter-individual variation in drug response may occur frequently. This raises two questions: how to represent human population genetic structure in the evaluation of drug safety and efficacy, and how to relate this structure to drug response. We address these by (i) inferring the genetic structure present in a heterogeneous sample and (ii) comparing the distribution of DME variants across the inferred genetic clusters of individuals. We find that commonly used ethnic labels are both insufficient and inaccurate representations of the inferred genetic clusters, and that drug-metabolizing profiles, defined by the distribution of DME variants, differ significantly among the clusters. We note, however, that the complexity of human demographic history means that there is no obvious natural clustering scheme, nor an obvious appropriate degree of resolution. Our comparison of drug-metabolizing profiles across the inferred clusters establishes a framework for assessing the appropriate level of resolution in relating genetic structure to drug response.

Introduction

Many drugs that show therapeutic potential never reach the market because of adverse reactions in some individuals, whereas other drugs in common use are effective for only a fraction of the population in which they are prescribed. This variation in drug response depends on many factors, such as sex, age and the environment, as well as genetic determinants. Since the 1950s, pharmacogenetic studies have systematically identified allelic variants at genes that influence drug response, including those of both drug-metabolizing enzymes (DMEs)[1] and drug targets[2], such as the cytochrome P450 monooxygenase CYP2D6 (refs. 3,4) and the N-acetyl transferase NAT2 (ref. 5) genes. Detailed functional analysis of variants at genes such as these has clearly shown the importance of genetic variation in drug responses. For example, analysis of *NAT2* alleles has identified amino acid–replacement mutations that reduce activity and a noncoding mutation that reduces translation, lowering the concentration of the enzyme[5]. In the case of CYP2D6, common variants include a frameshift leading to a truncated, nonfunctional protein and a splice-site mutation resulting in the absence of the protein[3,4]. These and other examples indicate that genetic tests might predict an individual's response to specific drugs, allowing medicines to be tailored to specific genetic makeups. Because of the potential commercial and clinical significance of such personalized medicines, an understanding of the genetic role of variable drug response is an important goal of biomedical research.

In addition to concerns surrounding individual variation in drug response, the geographic distribution of certain variants has highlighted the possible importance of average differences in drug response across populations. Genetic polymorphisms in DMEs, which probably contribute significantly to phenotypic variation in drug response, all vary in frequency among populations[2], some by as much as twelvefold[1]. For example, the well-known poor-metabolizer phenotype of debrisoquine oxidation is due to variant alleles of *CYP2D6*. Between 5% and 10% of Europeans, but only ~1% of Japanese, have loss-of-function variants at this locus that affect the metabolism of more than 40 drugs, including such commonly used agents as β-blockers, codeine and tricyclic antidepressants. The *CYP2D6* ultra-rapid metabolizer alleles also vary in frequency, even within Europe, from ~10% in Northern Spain to 1–2% in Sweden[6]. Polymorphisms in DMEs can lead to acute toxic responses and unwanted drug–drug interactions or to therapeutic failure from augmented drug metabolism (as in the case of *CYP2D6* duplications)[1,7].

These observations show that for some drugs, the tradeoffs between efficacy and adverse drug reaction not only will differ between individuals but also will show differences in average effects across different populations[8]. Genetically structured populations may be composed of two or more subpopulations with distinct drug-reaction profiles and thus may be better considered separately in some contexts. This raises the questions of the appropriate way to infer human population genetic structure in the context of the evaluation of drug safety and efficacy, and of how to relate this inferred genetic structure to drug response. To address this problem, we have used presumably neutral microsatellite markers to infer genetic clusters for a heterogeneous population, such as may be used in drug trials large enough to allow detection of both genetic and environmental

[1]*Galton Laboratory, Department of Biology, University College London, London, UK.* [2]*Department of Zoology, University of Oxford, Oxford, UK.* [3]*The Centre for Genetic Anthropology, Departments of Biology, University College London, London, UK.* [4]*Genostics Ltd, 28/30 Little Russell Street, London WC1A 2HN, UK. Correspondence should be addressed to D.B.G. (e-mail: d.goldstein@ucl.ac.uk).*

© 2001 Nature Publishing Group http://genetics.nature.com

© 2001 Nature Publishing Group http://genetics.nature.com

Table 1 • Inferring the number of clusters

K	ln Pr(*X*|*K*)	Pr(*K*|*X*)
1	–33680.97	~0
2	–32650.80	~0
3	–32046.80	~0
4	–31943.23	1.000
5	–31972.33	~0
6	–31987.10	~0

effects (for instance, Phase III trials). We compared the frequencies of functionally significant alleles at DME loci across the inferred clusters as an easily defined surrogate for drug response. Using this approach, we (i) show that there is considerable scope for population-genetic structuring in drug response in diverse metropolitan populations, because of the variation they harbor in DME allele frequency differences among identifiable genetic clusters (ii) establish a framework for determining the appropriate level of resolution (that is, the number of inferred clusters that should be used) in relating this population-genetic structuring to drug response and (iii) show that commonly used ethnic labels (such as Black, Caucasian and Asian) are insufficient and inaccurate descriptions of human genetic structure.

Results

We genotyped 16 chromosome 1 microsatellites from the ABI prism panel 1 (an average of 17 cM apart) and 23 X-linked microsatellites (≥2 cM apart)[9] in each of eight populations: South African Bantu speakers (46), Amharic- and Oromo-speaking Ethiopians from Shewa and Wollo provinces collected in Addis Ababa (48), Ashkenazi Jews (48), Armenians (48), Norwegian speakers from Oslo (47), Chinese from Sichuan in southwestern China (39), Papua New Guineans from Madang (48) and Afro-Caribbeans collected in London (30).

Genetic structure

We used a model-based clustering method implemented by the program STRUCTURE[10] to assign individuals to subclusters on the basis of these genetic data, ignoring their actual population affiliations. This mimics a scenario in which there is cryptic population structure, or no information as to the ethnic origin of the individuals. Briefly, the model implemented in STRUCTURE assumes *K* clusters, each characterized by a set of allele frequencies at each locus; the admixture model then estimates the proportion of each individual's genome having ancestry in each cluster. We estimated Pr(*X*|*K*), where *X* represents the data, using a model allowing admixture, for *K* between 1 and 6. From this and a uniform prior on *K* between 1 and 6, we estimated Pr(*K*|*X*) using Bayes's theorem (Table 1)[10]. Virtually all of the posterior probability density is on *K*=4.

The apportionment of individuals (the average per-individual proportion of ancestry) from each of the eight populations into the four STRUCTURE-defined clusters (Table 2) broadly corresponds to four geographical areas: Western Eurasia, Sub-Saharan Africa, China and New Guinea. Notably, 62% of the Ethiopians fall in the first cluster, which encompasses the majority of the Jews, Norwegians and Armenians, indicating that placement of these individuals in a 'Black' cluster would be an inaccurate reflection of the genetic structure. Only 24% of the Ethiopians are placed in the cluster with the Bantu and most of the Afro-Caribbeans; however,

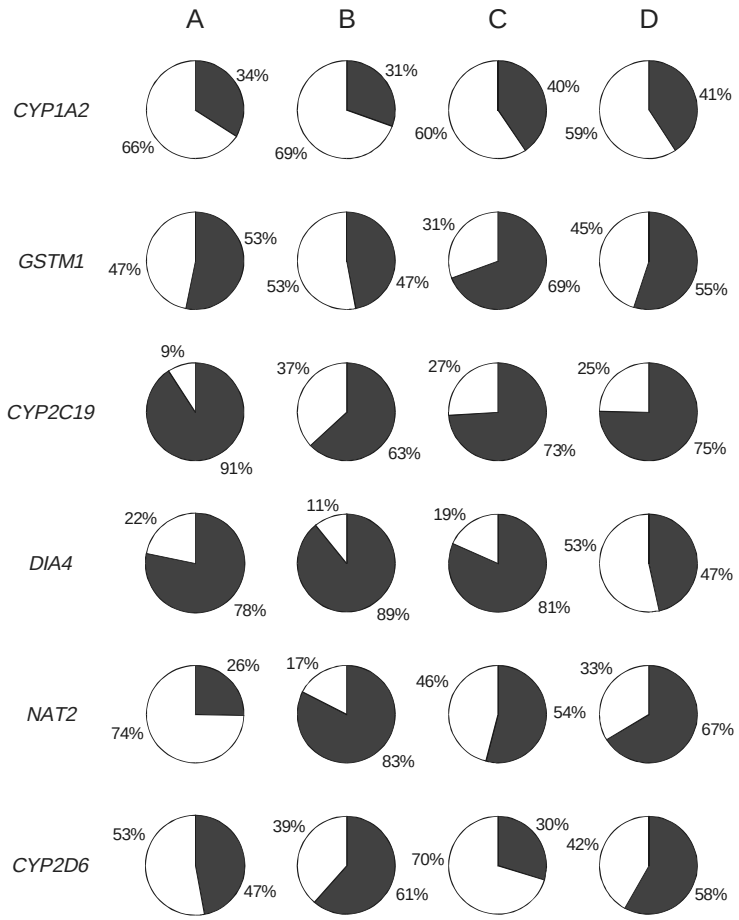

Fig. 1 Allele frequencies at each DME gene in the STRUCTURE-defined clusters. In all but the last two, black indicates wildtype and white, mutant; for *CYP2D6*, all mutant alleles are pooled as white, and for *NAT2* both tested mutant alleles (*5 and *6) are pooled as white. Cytochrome P450 1A2 (CYP1A2) metabolizes several drugs and carcinogens, including the analgesic acetaminophen (Tylenol) and probably antipsychotic drugs[18]. CYP2C19 metabolizes diazepam, barbiturates and antidepressants, and a polymorphic variant is responsible for the classical mephenytoin poor-metabolizer phenotype[19]. The classical debrisoquine poor-metabolizer phenotype is due to a variant of CYP2D6[7], and NAT2 is responsible for the classical isoniazid polymorphism[5]. NAD(P):quinone oxidoreductase (DIA4) converts quinones to stable hydroxyquinones and bioactivates antitumor quinones and nitrobenzenes[15]. Glutathione-*S*-transferase M1 (GSTM1) conjugates various electrophilic compounds, including potent environmental carcinogens such as aflatoxin B_1 epoxides[1]. The two *NAT2* polymorphisms we genotyped both result in slow acetylator alleles which lead to increased risks of drug toxicity and of certain cancers[1,5]. Of the *CYP2D6* alleles we assayed, *CYP2D6**1 is wildtype, *3 and *4 have no activity (which can lead to an acute toxic response to some drugs) and *2, *9 and *10 have reduced activity[17,20]. The *CYP1A2* variant genotyped leads to increased enzyme inducibility in smokers[21]. We genotyped the major polymorphism in *CYP2C19* responsible for the mephenytoin poor-metabolizer trait. After the administration of various drugs, this variant can lead to bone marrow toxicity, fatal blood dyscrasias and other adverse responses[1]. Increased susceptibilities to various cancers are associated with the deletion polymorphism in *GSTM1* genotyped here, dramatically so for smokers[1,14]. The mutation in *DIA4* leads to a complete absence of the protein and thus loss of protection against the toxic and carcinogenic effects of quinones[15]. Frequencies are shown for groupings corresponding to those shown in Table 2.

© 2001 Nature Publishing Group http://genetics.nature.com

21% of the Afro-Caribbeans are placed in a cluster with the West Eurasians (presumably reflecting genetic exchange with Europeans). Finally, China and New Guinea are placed almost entirely in separate clusters, indicating that the ethnic label 'Asian' is also an inaccurate description of population structure.

Consideration of only the X-linked microsatellites for the purposes of clustering supports K=3 with a clustering very similar to that for the entire dataset, except that the Chinese and New Guinean clusters are combined into one. When only the chromosome 1 microsatellites are used, the clustering is essentially the same as for the whole dataset. This discrepancy may be explained by one of two factors: (i) a lack of resolution in the X chromosome microsatellites or (ii) a biological factor such as the different number of X chromosomes and autosomes carried by males and females. To test these hypotheses, we carried out structure runs on the chromosome 1 data using an amount of information equal to that available from the X chromosome (22 alleles). The chromosome 1 microsatellites continued to support K=4, indicating that a lack of resolution in the X chromosome microsatellites may not have been the explanation. Perhaps, because the X chromosome spends more time in the female germline than does chromosome 1 and because females have a higher migration rate than males[11], the X-linked loci have less genetic structure. Smaller random subsets of the loci support a variety of values for K and do not agree on the clustering scheme (data not shown). This is probably because there are no natural clusters, as there has not been a history of bifurcation in human populations. Our results indicate that a reasonably high number of loci should be used to obtain consistency in clustering; one approach would be to use one marker from each chromosome arm. All of the analyses we present use the full dataset, resulting in four clusters (Table 2).

Table 2 • Proportion of membership of each sampled population in STRUCTURE-defined subclusters

Population	A	B	C	D
Bantu	0.04	0.02	0.93	0.02
Ashkenazi	0.96	0.01	0.01	0.02
Ethiopia	0.62	0.08	0.24	0.06
Norway	0.96	0.02	0.01	0.01
Armenia	0.90	0.04	0.02	0.05
China	0.09	0.05	0.01	0.84
Papua New Guinea	0.02	0.95	0.01	0.02
Afro-Caribbean	0.21	0.03	0.73	0.03

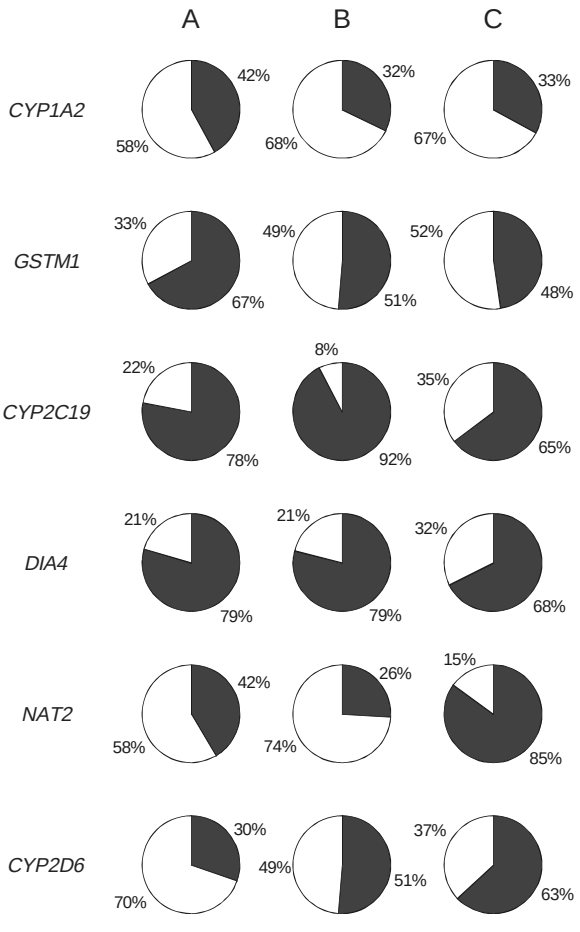

Fig. 2 Allele frequencies at each of the DME variants in the ethnically labeled groups. See Fig. 1 legend for details. A, Bantu, Ethiopian and Afro-Caribbean frequencies; B, those for Norwegians, Ashkenazi Jews and Armenians; C, those for Chinese and New Guineans.

Drug-metabolizing enzymes

Our selection of DMEs includes representatives of both phase I (oxidation or reduction) and phase II (conjugation) drug metabolism. We included three enzymes of the phase I cytochrome P450 family: CYP1A2, CYP2C19 and CYP2D6. We also included three conjugating or phase II metabolism enzymes: NAT2, NAD(P):quinone oxidoreductase (DIA4) and glutathione-*S*-transferase M1 (GSTM1). We determined allele frequencies at 11 variants in the genes encoding these six DMEs, all of which are known to be functionally significant (Fig. 1).

There are notable differences in the allele frequencies of DME-encoding genes between the genetically identified clusters (Fig. 1) for five of six reported loci. To assess differentiation across clusters, we counted allele frequencies in each of the clusters and calculated χ^2; we also tested for differences in allele frequencies using logistic regression. Using both methods, and correcting for multiple comparisons, the allele frequency distributions are significantly different for four of the six loci (significant for *NAT2, CYP2C19, DIA4* and *CYP2D6*). The pattern is particularly striking at *CYP2C19*, where the frequency of the mutant allele (the mephenytoin polymorphism) in cluster B is more than four times that of cluster A (P<0.0001). We also observed extreme differentiation between clusters B and D for *DIA4*, for which the frequency of the mutant allele (which provides no protection against the toxic effects of quinones) differs by almost five-fold (P<0.0001). This is a notable difference, as clusters B and D would be combined as 'Asian' in current drug evaluation using ethnic labels. *NAT2* also shows significant differentiation between these two clusters, as well as among the others. We observed strong to modest differences in allele frequencies for the other DME genes between at least two pairs of the clusters in each case. To further explore cluster differentiation we counted the number of loci for which there are significant allele frequency differences (using χ^2) for each of the pairs of clusters. Without correcting for multiple comparisons, this number varied from 2 (of 6 loci) for B versus D, to 5 (of 6) for B versus C. Given the important differences in drug response determined by these variants, the scope for genetic structuring in drug response clearly is high. For some drugs, therefore, the trade-off between therapeutic response and adverse drug reactions will differ between the clusters identified here, making this kind of genetic analysis important in checking for such effects in any phase III clinical trial.

We compared the predictive value of the genetic clusters to that of commonly used ethnic labels by counting the DME allele frequencies in the grouping resulting from those labels: Caucasian

© 2001 Nature Publishing Group http://genetics.nature.com

(Norwegian, Ashkenazi Jew, Armenian), Black (Bantu, Ethiopian, Afro-Caribbean) and Asian (Chinese, New Guinean; Fig. 2). Notably, for *DIA4*, the large frequency difference between clusters B and D (driven by the differentiation between China and New Guinea) is averaged when both populations are lumped; the mutant allele frequency is thus only one and a half times as high as that in the other two groups. Indeed, the overall differentiation for the ethnic groups is not significant after correction for multiple comparisons. Note that in no case did we observe the reverse in our data: that is, the ethnic labels never show sharp differentiation that is not observed in the clusters. In addition, only in the case of *CYP2D6* are the allele frequency differentials as high as they are for genetically defined clusters. Although there is some DME allele frequency differentiation between ethnically labeled groups, in most cases it is less than that seen for the genetic clusters. To confirm this, we fitted logistic regression models to the allele data using membership in the genetic clusters as the explanatory variables, and tested for the increase in goodness of fit obtained by adding the ethnic labels as explanatory variables. We then compared this to the increase in goodness of fit obtained by adding the genetic cluster information to the ethnic group information. Of those DME loci (*NAT2, CYP2C19*, *DIA4* and *CYP2D6)* that showed significant differentiation in either the clusters or the ethnic groups, in three of four cases, adding genetic cluster information to ethnic labels was more significant than adding ethnic labels to genetic clusters. For *CYP2D6*, the opposite was true.

Multilocus interactions

Undesirable drug reactions or interactions, as well as environmental sensitivities, may also be due to the existence of variants at two (or more) loci. An example of this may be the case of the increased susceptibility to colorectal cancer in individuals with a rapid/rapid metabolizer phenotype at *CYP1A2* and *NAT2*, especially for those who prefer well-cooked meat[12]. It is important to consider not only differences in allele frequency between the inferred clusters but also differences in frequency for multilocus genotypes. There are large frequency differentials between the clusters we have identified for multilocus genotypes, which may give rise to phenotypic combinations like this; in fact, the frequency of the combination *CYP1A2*-A/A, *NAT2**4/– observed in cluster B (47%) is more than twice that seen in clusters A (19%) or C (22%; $P<0.01$ for overall differentiation). When such interactions are important, they may be apparent in the genetic analysis described here, from the distribution of drug response across inferred clusters.

Discussion

By carrying out the clustering analysis with the number of clusters set to different values, we can compare the extent of differentiation among the clusters to assess the appropriate level of resolution. In the context of a Phase III trial, the appropriate benchmark would reflect the amount of the total variation in drug response explained by the genetic clusters. A surrogate test would be to carry out exact tests of differentiation[13] on relevant functional polymorphisms, stopping when an increase in the number of clusters does not appreciably increase the degree of differentiation. The clustering properties of STRUCTURE, however, can be unstable across different values of *K*, which complicates the implementation of such an analysis.

It is well known that there are inter-ethnic differences in DME allele frequencies and thus in drug response. Our focus here, however, has been to assess the scope for average difference in drug response across genetically inferred clusters. Not only can these clusters be derived in the absence of knowledge about ethnicity (or geographic origin), but they are also more informative than commonly used ethnic labels. Because of the potential clinical significance of average differences in drug response, we conclude that it is not only feasible but a clinical priority to assess genetic structure as a routine part of drug evaluation.

When the most important genes influencing response to a particular drug or group of drugs have been identified, it should be possible to personalize medicine on the basis of an individual's genotype, assuming that routine individual genotyping is commercially and technically feasible. Short of such detailed knowledge, however, it is important to assess whether drugs work similarly in different genetic subgroups. The appropriate level of clustering may be evaluated empirically by assessing the amount of variation in response explained by the inferred clusters. In addition, we have shown that the common ethnic labels currently available to regulatory authorities show a poor correspondence with genetically inferred clusters.

Analysis of population structure in biomedical research

Our implementation of STRUCTURE is primarily meant to show that familiar ethnic labels are not accurate guides to genetic structure. We have not attempted to provide a definitive description of human population structure. The results of STRUCTURE can, in fact, be quite difficult to interpret. Notably, statistical difficulties may arise when assessing convergence, and the assessment of the appropriate value of *K* is currently not rigorous[10]. These and other issues can lead to anomalous outcomes; for example, an implausible value of *K* may be supported where one of the clusters is more or less empty. In addition, results may vary for biological reasons, such as when markers are affected differentially by forces acting on the genome, such as gene flow. Detailed analysis of STRUCTURE output and other clustering schemes, using a standard battery of markers in a global sample of human populations, will be needed to arrive at a canonical clustering scheme for use in biomedical research. Such an evaluation would need to be geographically exhaustive and to include a sufficient number of markers throughout the genome to ensure that the resulting clustering scheme is robust; consistent results should be obtained with different marker and sample sets.

Methods

Microsatellite markers and structure inference. All subjects were unrelated males. We genotyped[9] the following X-linked microsatellites: *DXS984*, *996*, *1036*, *1053*, *1062*, *1203*, *1204*, *1205*, *1206*, *1211*, *1212*, *1220*, *1223*, *7103*, *8014*, *8061*, *8068*, *8073*, *8085*, *8086*, *8087* and *8099*. We genotyped the following chromosome 1 microsatellites: *D1S196*, *206*, *213*, *249*, *255*, *450*, *484*, *2667*, *2726*, *2785*, *2797*, *2800*, *2836*, *2842*, *2878* and *2890*. The chromosome 1 markers form part of the ABI Prism linkage mapping panel 1 and were amplified according to the manufacturer's instructions. We assigned individuals into clusters using the admixture model in the program STRUCTURE[10], with no correlation in allele frequencies among populations and a burn-in time of at least 1 million steps, followed by another 1 million steps of the Markov Chain for data collection. We carried out multiple runs for each set of conditions to be sure that the chain had converged; in total, we carried out more than 500 runs.

DME genotyping. We sequenced the intronic C734A transversion in *CYP1A2* and two SNPs in *NAT2*: C481T, defining allele *5 (in complete allelic association with Ile113Thr) and G590A (giving Arg197Gln), defining allele *6. We classified all other alleles as *4, and combined the two mutant allele frequencies for the purpose of binary analysis. We genotyped the deletion allele of glutathione-*S*-transferase M1 (*GSTM1*) using *GSTM4* amplification as an internal control[14]. We genotyped the C191T transition (giving Pro187Ser) in *DIA4* (ref. 15) and the G117A transition (leading to a truncated protein) in *CYP2C19* (ref. 16) using polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP). We labeled *GSTM1* and RFLP amplicons fluorescently and determined sizes on an ABI 3100 automated sequencer

© 2001 Nature Publishing Group http://genetics.nature.com

(Applied Biosystems). We typed *CYP2D6* SNPs by gene-specific PCR, followed by nested multiplex reamplification–RFLP detection of the following 'key' mutations[17]: C100T (Pro34Ser; alleles *10 and *4), G1846A (splicing defect; allele *4), A2549del (frameshift; allele *3), 2613–2615AGAdel (Lys281del; allele *9) and C2850T (Arg296Cys; allele *2). All other chromosomes were denoted *1 (thus, this category includes some non-wildtype alleles). For the binary analyses, we considered *CYP2D6**1 as having normal activity and all other alleles as having reduced activity. We labeled *CYP2D6* amplicons using fluorescent primers and sized them on an ABI 377 automated sequencer (Applied Biosystems; genotyping details available from B.F.). In the case of *GSTM1*, the assay does not allow differentiation between homozygous and heterozygous presence of the nondeletion allele. For this case, we carried out calculations on genotype frequencies and homozygous deletion versus homozygous or heterozygous for the nondeletion allele. We estimated the accuracy of our genotyping by retesting a number of samples from each population. Error rates varied from 0 to 7% for the DME SNPs and from 0 to 5% for the microsatellites.

DME differentiation across clusters. We calculated DME allele frequencies in the clusters by distributing an individual's genotype among the clusters, according to the proportion of ancestry that the individual had in each cluster, as determined by STRUCTURE output. When individuals were placed in the cluster in which they had the most ancestry, the results changed very little (data not shown). To meet the assumption of a multinomial distribution, we evaluated χ^2 tables after placing individuals in the clusters in which they had most ancestry.

Acknowledgments

D.B.G. is a Royal Society/Wolfson Research Merit Award holder.

Received 30 July; accepted 4 October 2001.

1. Weber, W.W. *Pharmacogenetics* (Oxford University Press, Oxford, 1997).
2. Evans, W.E. & Relling, M.V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487–491 (1999).
3. Gough, A.C. *et al.* Identification of the primary gene defect at the cytochrome P450 CYP2D locus. *Nature* **347**, 773–776 (1990).
4. Kagimoto, M., Heim, M., Kagimoto, K., Zeugin, T. & Meyer, U.A. Multiple mutations of the human cytochrome P450IID6 gene (CYP2D6) in poor metabolizers of debrisoquine. Study of the functional significance of individual mutations by expression of chimeric genes. *J. Biol. Chem.* **265**, 17209–17214 (1990).
5. Blum, M., Demierre, A., Grant, D.M., Heim, M. & Meyer, U.A. Molecular mechanism of slow acetylation of drugs and carcinogens in humans. *Proc. Natl Acad. Sci. USA* **88**, 5237–5241 (1991).
6. Bernal, M.L. *et al.* Ten percent of North Spanish individuals carry duplicated or triplicated CYP2D6 genes associated with ultrarapid metabolism of debrisoquine. *Pharmacogenetics* **9**, 657–660 (1999).
7. Meyer, U.A. & Zanger, U.M. Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* **37**, 269–296 (1997).
8. International Conference on Harmonisation. Ethnic Factors in the Acceptability of Foreign Clinical Data. (International Conference on Harmonisation, 1998).
9. Wilson, J.F. & Goldstein, D.B. Consistent long-range linkage disequilibrium generated by admixture in a Bantu–Semitic hybrid population. *Am. J. Hum. Genet.* **67**, 926–935 (2000).
10. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
11. Seielstad, M.T., Minch, E. & Cavalli-Sforza, L.L. Genetic evidence for a higher female migration rate in humans. *Nature Genet.* **20**, 278–280 (1998).
12. Kohlmeier, L., DeMarini, D. & Piegorsch, W. Gene-nutrient interactions in nutritional epidemiology. in *Design Concepts in Nutritional Epidemiology* (eds. Margetts, B. & Nelson, M.) 312–337 (Oxford University Press, Oxford, 1997).
13. Raymond, M. & Rousset, F. An exact test for population differentiation. *Evolution* **49**, 1280–1283 (1995).
14. Krajinovic, M., Labuda, D., Richer, C., Karimi, S. & Sinnett, D. Susceptibility to childhood acute lymphoblastic leukemia: influence of CYP1A1, CYP2D6, GSTM1, and GSTT1 genetic polymorphisms. *Blood* **93**, 1496–1501 (1999).
15. Gaedigk, A. *et al.* NAD(P)H:quinone oxidoreductase: polymorphisms and allele frequencies in Caucasian, Chinese and Canadian Native Indian and Inuit populations. *Pharmacogenetics* **8**, 305–313 (1998).
16. Goldstein, J.A. & Blaisdell, J. Genetic tests which identify the principal defects in CYP2C19 responsible for the polymorphism in mephenytoin metabolism. *Methods Enzymol.* **272**, 210–218 (1996).
17. Gaedigk, A. *et al.* Optimization of cytochrome P4502D6 (CYP2D6) phenotype assignment using a genotyping algorithm based on allele frequency data. *Pharmacogenetics* **9**, 669–682 (1999).
18. Basile, V.S. *et al.* A functional polymorphism of the cytochrome P450 1A2 (CYP1A2) gene: association with tardive dyskinesia in schizophrenia. *Mol. Psychiatry* **5**, 410–417 (2000).
19. Ferguson, R.J. *et al.* A new genetic defect in human CYP2C19: mutation of the initiation codon is responsible for poor metabolism of S-mephenytoin. *J. Pharmacol. Exp. Ther.* **284**, 356–361 (1998).
20. Daly, A.K. *et al.* Nomenclature for human CYP2D6 alleles. *Pharmacogenetics* **6**, 193–201 (1996).
21. Sachse, C., Brockmoller, J., Bauer, S. & Roots, I. Functional significance of a C→A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br. J. Clin. Pharmacol.* **47**, 445–449 (1999).

© 2001 Nature Publishing Group http://genetics.nature.com