

Heather E. Collins-Schramm · Rick A. Kittles
Darwin J. Operario · James L. Weber
Lindsey A. Criswell · Richard S. Cooper
Michael F. Seldin

Markers that discriminate between European and African ancestry show limited variation within Africa

Received: 7 June 2002 / Accepted: 29 July 2002 / Published online: 14 September 2002

© Springer-Verlag 2002

Abstract Markers informative for ancestry are necessary for admixture mapping and improving case-control association analyses. In particular, African Americans are an admixed population for which genetic studies require accurately evaluating admixture. This will require markers that can be used in African Americans to determine if a given genomic region is of European or African ancestry. This report shows that, despite studies indicating high intra-African sequence variation, markers with large inter-ethnic differences have only small variations in allele distribution among divergent African populations and should be valuable for evaluating admixture in complex disease genetic studies.

The use of natural population admixture for uncovering disease susceptibility genes is receiving considerable attention (McKeigue et al. 2000; Smith et al. 2001). Proposed methods examine linkage disequilibrium in the admixed population or are based on defining the ancestry of

chromosomal regions in an admixed population. Each approach utilizes chromosomal markers to look for differences in the population likelihood of inheritance from one of the founding populations. Differences between controls and probands can be examined for each region, or the inheritance pattern of one genomic region in probands can be compared to the pattern in the rest of their genome. The presence of these differences would presumably reflect the different risk of disease conferred by alleles of susceptibility genes contributed by the founding populations. Several diseases – including type 2 diabetes, propensity for nephropathy and systemic lupus erythematosus – are thought to have major differences in the genetic risk of disease between African and European populations, and therefore may be appropriate for admixture mapping studies in African Americans.

The application of admixture mapping methods requires markers that can distinguish between the founding populations. In addition, such markers are useful in the development and application of statistical methods to adjust for population stratification that otherwise confounds accurately assessing linkage disequilibrium in association-based genetic studies (Pritchard et al. 2000). Recently, a large number of markers, with large differences in allele frequencies between African American and European American or African and European derived populations, have been reported (Smith et al. 2001; Collins-Schramm et al. 2002). However, some investigators have questioned the usefulness of such markers given previous data indicating very large variation in allele frequencies within populations, particularly those of African derivation (Lewontin 1972; Nei and Roychoudhury 1974).

Sequence variation within the African population is thought to exceed the variation in sequence between populations (e.g. between the African and European population). For example, based on comparisons of F_{ST} values from examination of 36 dystrophin polymorphisms, Zietkiewicz et al. (1997) have suggested that most of the divergence within the world is caused by differences among particular populations and not by differences between Africans and non-Africans. Previous data from our labo-

H.E. Collins-Schramm · D.J. Operario · M.F. Seldin (✉)
Rowe Program in Human Genetics,
Departments of Biological Chemistry and Medicine,
University of California at Davis,
One Shields Avenue, Davis, CA 95616-8669, USA
e-mail: mfseldin@ucdavis.edu,
Tel.: +1-530-7546017, Fax: +1-530-7546015

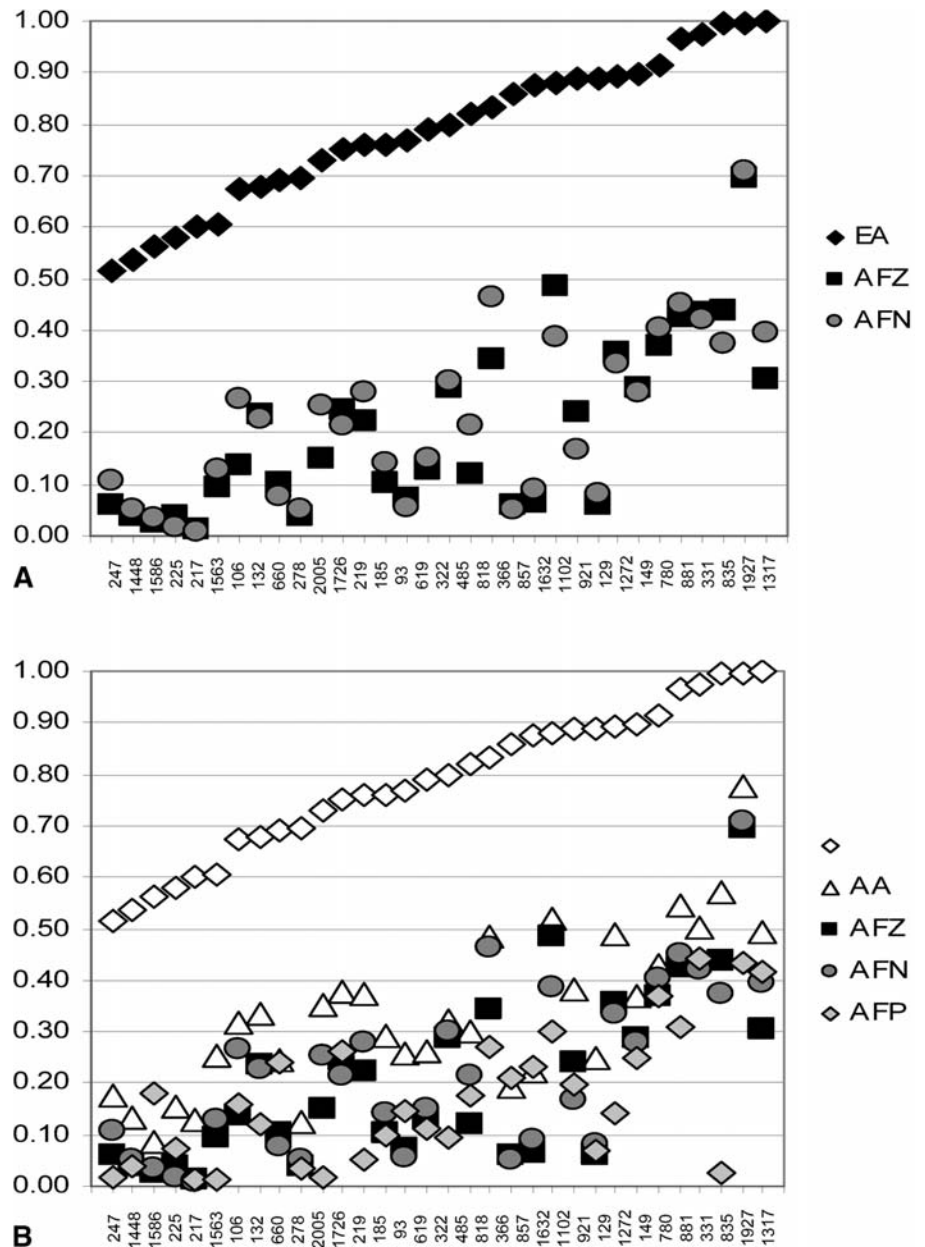
R.A. Kittles
National Human Genome Center at Howard University,
2041 Georgia Ave., Rm. 615, Washington, DC 20060, USA

J.L. Weber
Center for Medical Genetics,
Marshfield Medical Research Foundation,
Marshfield, WI 54449, USA

L.A. Criswell
Rosalind Russell Medical Research Center for Arthritis,
University of California, San Francisco,
San Francisco, CA 94143, USA

R.S. Cooper
Preventive Medicine and Epidemiology, Loyola University,
Maywood, IL 60153, USA

Fig. 1a, b Allele frequency comparisons between European American, African American and African populations for 32 diallelic insertion/deletion polymorphisms. The specific indel marker numbers (<http://research.marshfield-clinic.org/genetics>) are shown along the abscissa. The ordinate shows the frequency of the more common allele of each marker in the European American (*EA*) population for each population. **a** A comparison between *EA* (individual typing; mean subjects/marker = 153) and African populations from Zimbabwe (*AFZ*) (individual typing; mean subjects/marker = 152) and Nigeria (*AFN*) (individual typing; mean subjects/marker = 159). **b** Additional comparisons with the inclusion of African American (*AA*) (individual typing; mean subjects/marker = 264) and Pygmy (*AFP*) (typing results from a pooled sample containing DNA from 12 Mbuti and 9 Biaka individuals determined for 30 of 32 markers)



laboratories, as well as other investigators, have suggested that markers selected for large inter-ethnic differences in allele frequencies might behave very differently. For example, modeling suggested that for markers selected for large differences between European Americans and Africans, the allele frequencies in the Shona tribal group from Zimbabwe could approximate the African contribution to Californian African Americans (Collins-Schramm et al. 2002). This might be considered surprising since the Shona are located in Southeast Africa, far from the regions contributing to the slave trade (Iliffe 1995). However, these studies neither directly compared different African populations, nor examined admixture in individual subjects within any population.

The current study utilized 32 markers determined – as part of a large-scale screen for markers useful for admix-

ture mapping – to have large allele frequency differences between Africans from Zimbabwe and European Americans (Collins-Schramm et al. 2002). All markers were diallelic insertion/deletion polymorphisms (indels) originally identified by the Marshfield Center for Medical Genetics (<http://research.marshfieldclinic.org/genetics>). Markers were chosen such that linkage disequilibrium between any marker pair in any African or European population would be very unlikely; they were distributed on 16 chromosomes, and the minimum separation between a marker pair was 13 cM. These markers were individually genotyped in DNA samples obtained from Africans in Nigeria and Zimbabwe, European Americans, and African Americans. Nigerians representing the Edo (Bini) ethnic group were recruited in the Udo community near Benin City, Nigeria. The Zimbabwe samples were recruited from in-

dividuals of the Shona (Southeast Bantu) ethnic group. The European American and African American individuals were classified by self-described ethnicity and derived primarily from Northern California. In addition, 30 of these indels were genotyped in a pooled sample of DNA from Pygmy subjects (12 Mbuti and 9 Biaka individuals). Previously constructed genetic trees have placed Southeast Bantu, West African, and Mbuti and Biaka populations in different clusters throughout a sub-Saharan major cluster, thus genotyping these populations should give a representation of variation in sub-Saharan Africa (Cavalli-Sforza et al. 1994). All samples were obtained under protocols and informed-consent procedures approved by institutional review boards and were labeled with anonymous code numbers.

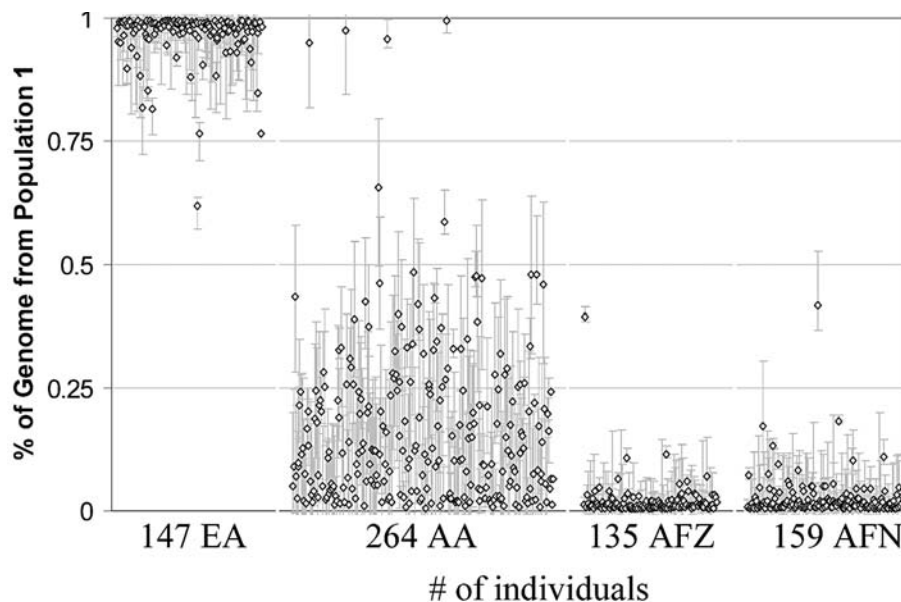
For each of these 32 indels, the allele frequencies in Nigerian subjects were nearly identical to the allele frequencies in subjects from Zimbabwe and dramatically contrasted with those in European American subjects (Fig. 1a). A table of the actual allele frequencies in each population is available at http://roweprogram.ucdavis.edu/Allele_Frequencies. Typing studies also indicated that for each of these markers the allele frequencies in African Americans were between the African and European American allele frequencies, consistent with previous estimations (Fig. 1b) (McKeigue et al. 2000; Collins-Schramm et al. 2002). Perhaps more surprising, there was a strong correlation between the allele distributions in the above noted African groups and those observed in pooled DNA samples from 21 Pygmy subjects (Fig. 1b). The variation for the intra-African comparisons, Nigerian/Zimbabwe (mean Wahlund variance, $f=0.005$), Nigerian/Pygmy ($f=0.02$) and Zimbabwe/Pygmy ($f=0.02$), was very small compared with inter-ethnic African/European American comparisons: Nigerian/European American ($f=0.38$); Zimbabwe/European American ($f=0.375$); and Pygmy/European American ($f=0.40$).

In addition, individual admixture was examined using the above markers and the Bayesian algorithms developed by Pritchard (Structure Program; Pritchard et al. 2000). This analysis (Fig. 2) separated the African individuals from the European American individuals without prior knowledge of population affiliation. The African subjects, whether from Zimbabwe or Nigeria, were clearly separated from the European American subjects, with most very tightly grouped, and no distinction was evident between the Zimbabwe and Nigerian populations. As expected, the admixed population, African Americans, showed considerable variation in the proportion of the putative contribution of the two populations. This finding emphasizes the importance of controlling for individual admixture ratios in admixture mapping studies. It is also interesting to note that there are outliers in all populations; the biological explanation for these is uncertain but they may represent unreported admixture history.

The current results can also be compared with those utilizing classical protein polymorphisms (Cavalli-Sforza et al. 1994). F_{ST} values were calculated from our data and compared with those of similar ethnic groupings reported in this previous work. As expected, based on our selection method, our markers had substantially higher inter-ethnic F_{ST} values (e.g. European American/Zimbabwe, $F_{ST}=0.375$) than those from classical polymorphisms (e.g. English/Bantu, $F_{ST}=0.229$). In contrast, for intra-African comparisons smaller F_{ST} values were observed for our markers selected for ancestry information (e.g. Zimbabwe/Nigerians, $F_{ST}=0.005$) than those observed with classical polymorphisms (e.g. Bantu/West Africans, $F_{ST}=0.019$). While these comparisons use different sample sets and marker classes, the results suggest that intra-ethnic variation is smaller for markers chosen for large inter-ethnic differences than for randomly selected markers.

We believe that these studies provide compelling evidence that the African contribution to individual African

Fig. 2 Individual admixture in European American (EA), African (AFZ and AFN) and African American (AA) subjects estimated using the Structure program (Pritchard et al. 2000). Each point represents an individual subject that was analyzed with the indels described in Fig. 1 (mean 28 markers/individual examined; minimum markers/individual = 23 markers). The 80% confidence interval for each individual is shown. The Structure program was run without prior assignment of population affinity using 50,000 replicates in both the burn-in and simulation phases.



American subjects can be accurately analyzed with a set of obtainable markers that distinguish between European and African ancestry. In addition, it may be noteworthy that for the majority of markers, the African populations are more nearly fixed for one allele than the European American population (Fig.1). Thus, the lack of divergence within Africa for these markers cannot be explained solely by higher levels of sequence diversity.

Although many additional markers will need to be used for admixture mapping of medically important traits, the current study indicates that differences in allele frequencies of such ancestry informative/ethnic-difference markers within the various sub-populations that were subjected to the slave trade will not be a major confounding problem. These markers can also be readily applied to detect and control for unwanted admixture in association studies, as suggested by others (Pritchard et al. 2000), and will be applicable to specific forensic applications. It is unclear what limitations may be present for sets of such ancestry informative markers selected for use in other admixed populations or in populations that have major contributions from several different large ethnic/ancestral populations. Finally, the explanation for the lack of African variation observed with these ancestry informative markers, but not with previously examined unselected markers, which likely includes some combination of selection, bottlenecks, origin of polymorphisms and stability of polymorphisms, will require additional studies for precise inferences.

Acknowledgements Support for this research was provided by National Institutes of Health grants U01-DK57249, N01-HV-48141, AR44804, R01-HV-62681 and HL45508.

References

- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes, Princeton University Press, Princeton, New Jersey
- Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL, Knowler WC, Cooper R, Li H, Seldin MF (2002) Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 70:737–750
- Iliffe J (1995) Africans: the history of a continent. Cambridge University Press, Cambridge
- Lewontin RC (1972) The apportionment of human diversity. *Evol Biol* 6:381–398
- McKeigue PM, Carpenter JR, Parra EJ, Shriver MD (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet* 64:171–186
- Nei M, Roychoudhury AK (1974) Genetic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. *Am J Hum Genet* 26:421–443
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* 67:170–181
- Smith MW, Lautenberger JA, Doo Shin H, Chretien J, Shrestha S, Gilbert DA, O'Brien SJ (2001) Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. *Am J Hum Genet* 69:1080–1094
- Zietkiewicz E, Yotova V, Jarnik M, Korab-Lakowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D (1997) Nuclear DNA diversity in worldwide distributed human populations. *Gene* 205:161–171