Evidence for positive selection and population structure at the human *MAO-A* gene

Yoav Gilad*†, Shai Rosenberg‡, Molly Przeworski§, Doron Lancet*, and Karl Skorecki†‡

*Department of Molecular Genetics and the Crown Human Genome Center, The Weizmann Institute of Science, Rehovot 76100, Israel; ‡Rappaport Faculty of Medicine and Research Institute, Technion–Israel Institute of Technology, and Rambam Medical Center, Haifa 31096, Israel; and §Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, United Kingdom

Communicated by Eviatar Nevo, University of Haifa, Haifa, Israel, November 19, 2001 (received for review May 15, 2001)

We report the analysis of human nucleotide diversity at a genetic locus known to be involved in a behavioral phenotype, the monoamine oxidase A gene. Sequencing of five regions totaling 18.8 kb and spanning 90 kb of the monoamine oxidase A gene was carried out in 56 male individuals from seven different ethnogeographic groups. We uncovered 41 segregating sites, which formed 46 distinct haplotypes. A permutation test detected substantial population structure in these samples. Consistent with differentiation between populations, linkage disequilibrium is higher than expected under panmixia, with no evidence of a decay with distance. The extent of linkage disequilibrium is not typical of nuclear loci and suggests that the underlying population structure may have been accentuated by a selective sweep that fixed different haplotypes in different populations, or by local adaptation. In support of this suggestion, we find both a reduction in levels of diversity (as measured by a Hudson–Kreitman–Aguade test with the DMD44 locus) and an excess of high frequency-derived variants, as expected after a recent episode of positive selection.

Association studies at the monoamine oxidase A (*MAO-A*) locus have been motivated by the striking finding that mutations in the gene result in borderline mental retardation and abnormal behavior, including increased impulsive behavior (1). It has been proposed that a broad range of interindividual human variability in related behavioral phenotypes may be associated with nucleotide variation at this locus, in particular with the well-documented range of interindividual variability in the activity level of the gene product (2).

Numerous studies have been carried out on the association of this genetic locus to behavioral phenotypes (3–5), but positive and negative associations have been accepted with reservation and have been difficult to replicate in subsequent studies. Some of these difficulties may result from a failure to take into account the evolutionary history of the region, including the effects of population stratification and/or of natural selection. These factors shape the distribution of linkage disequilibrium (LD) and hence the likelihood of an association.

As a step in our understanding of genetic variability at this locus, we examined polymorphism patterns in normal, unrelated individuals. We used direct resequencing of the *MAO-A* gene. A number of indirect approaches have been used for large-scale single nucleotide polymorphism (SNP) discovery and analysis (6, 7). However, direct resequencing is the most reliable approach to SNP discovery, affording a complete picture of the sequence variation for a given genomic region. To establish the phase of segregating sites across long genomic segments of autosomal loci, previous studies have often inferred haplotypes by means of a variety of algorithms (e.g., ref. 8). These have difficulty in reconstructing the phase of SNPs at low frequency. Here, we are able to determine haplotypes directly in males, because *MAO-A* is sex-linked. The region reported in this study is one of the longest stretches of DNA in a recombining part of the genome for which haplotypes have been obtained directly.

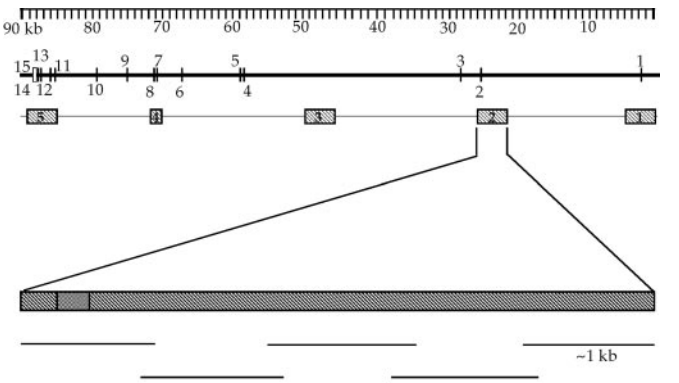


Fig. 1. Overall genomic structure and sequencing strategy for the MAO-A gene. The arrangement of exons is shown relative to the scale provided at the top. We indicate the position of each of the five resequenced regions. The sequencing strategy is illustrated for region 2, where the PCR products are shown as overlapping segments.

Methods

DNA Samples. Human genomic DNA was derived from two sources. (*i*) Thirty three DNA samples were provided by Coriel Cell Repositories, Camden, NJ. These consisted of: seven Pygmy samples, nine Aboriginal Taiwanese, three Chinese, two Japanese, five Mexicans, and seven Russians. (*ii*) Samples from 23 unrelated individuals were provided by the National Laboratory for the Genetics of Israeli Populations at Tel Aviv University; these came from two ethnic groups: Ashkenazi Jews (13 individuals) and Bedouins (10 individuals). We isolated genomic DNA from two common chimpanzees (*Pan troglodytes*) from blood kindly provided by Yigal Horvitz of the Israeli Safari Zoo (Ramat-Gan, Israel), using the Genomix DNA preparation kit (Talent SRL, Trieste, Italy).

Sequencing Strategy. The *MAO-A* gene spans more than 90 kb. We chose five segments that varied from 2 to 5 kb in length and totaled 18.8 kb (Fig. 1). We tried to include as much exon sequence as possible while keeping the segments equally distributed across the entire gene. Overlapping ≈1-kb PCR products were sequenced across each segment. The sequence we screened consisted of 95.7% introns and 4.3% exons.

PCR Procedures. We designed specific PCR primers for the amplification of the ≈1-kb segments of the *MAO-A* gene, based

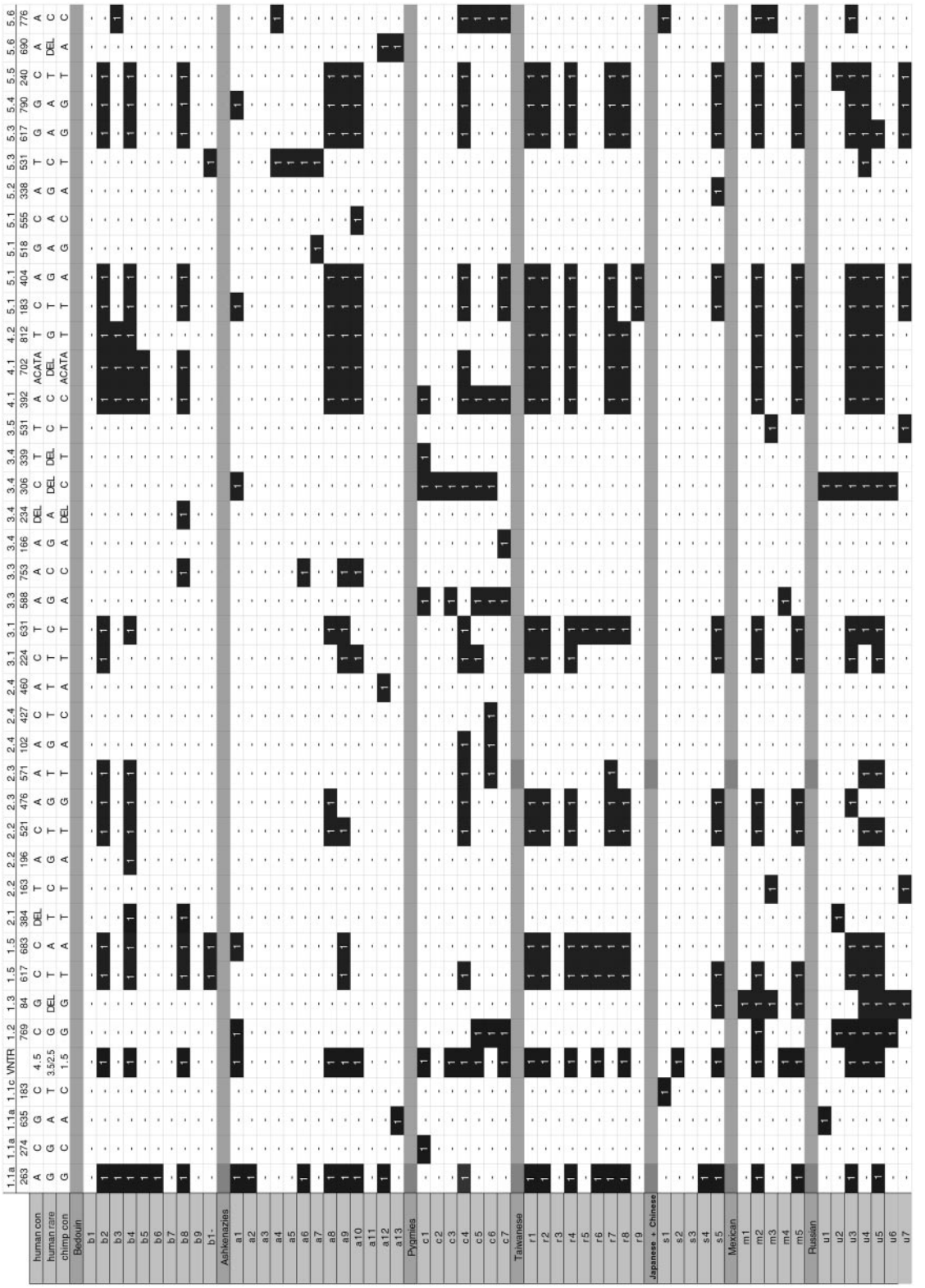
Abbreviations: MAO-A, monoamine oxidase A; SNP, single nucleotide polymorphism; LD, linkage disequilibrium.

†To whom reprint requests may be addressed. E-mail: yoav.gilad@weizmann.ac.il or skorecki@techunix.technion.ac.il.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

www.pnas.org/cgi/doi/10.1073/pnas.022614799

Table 1. Individual haplotypes



Haplotypes of 56 males, spanning the 90-kb MAO-A gene. The 41 polymorphic sites are indicated by their segment name (first row) and the position within the segment (second row). The human consensus (frequent allele), human rare allele, and the chimpanzee consensus are indicated in rows 3–5. The rare variant is represented by a dark rectangle with the digit 1. We also indicate the human and chimpanzee consensus sequences.

on the available sequences. We performed PCR in a total volume of 25 μl, containing 0.2 mM of each deoxynucleotide (Promega), 50 pMol of each primer, PCR buffer containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris (pH 8.3), 1 unit of *Taq* DNA polymerase (Roche Molecular Biochemicals), and 50 ng of genomic DNA. PCR conditions were as follows: 35 cycles of denaturation at

94°C, annealing at either 55°C or 52°C, and extension at 72°C, each step for 1 min. The first step of denaturation and the last step of extension were 3 min and 10 min, respectively. PCR products were separated on a 1% agarose gel to view their size, and they were purified by using the High Pure PCR Product Purification Kit (Roche Molecular Biochemicals).

DNA Sequencing. Sequencing reactions were performed on PCR products or clones in both directions with dye terminators (dye terminator cycle sequencing kit; Perkin–Elmer) on an Applied Biosystems 3700 automated sequencer.

After base calling with Applied Biosystems ANALYSIS software (version 3.0), the analyzed data were edited by using the SEQUENCHER program, version 3.0 (Gene Codes, Ann Arbor, MI).

Determination of Polymorphism and Divergence. We sequenced each ≈1-kb genomic segment from both ends for each individual. The SEQUENCHER software was used to assemble the sequences and identify DNA polymorphisms. We repeated the sequencing reaction of any segment originally identified as containing a singleton. The human sequences were aligned with the chimpanzee sequence to identify fixed differences.

Data Analysis. We calculated three summaries of diversity levels: Watterson's θ_W (9), based on the number of segregating sites in the sample; π (10), the average number of pairwise differences in the sample; and θ_H, a summary that gives more weight to high frequency-derived variants (11). Under the standard neutral model of a random-mating population of constant size, all three summaries estimate the population mutation parameter $\theta = 3N\mu$ (for X-linked loci), where N is the diploid long-term inbreeding effective population size, and μ is the mutation rate per generation. To test whether the frequency spectrum of mutations conformed to the expectations of this standard neutral model, we calculated the value of three test statistics: Tajima's D (12), which considers the difference between π and θ_W, Fay and Wu's H test (11), which considers the difference between π and θ_H, and the HKA (Hudson–Kreitman–Aguade) test (13), which tests whether levels of polymorphism are consistent with levels of divergence, as expected under the neutral model, by comparison with one or more reference loci. The P values for D and H were estimated from 10^4 coalescent simulations of an infinite site locus that condition on the sample size; these simulations are implemented for a fixed number of segregating sites rather than with a population mutation rate (cf. ref. 14). All but one of the P values reported were for no recombination. The assumption of no recombination is a conservative one as determined by using a modification of the program of R. Hudson, University of Chicago (see http://home.uchicago.edu/~rhudson1/), which implements the coalescent with recombination. For a given population recombination rate, simulations were run conditional on the actual number of base pairs and sample size. The resulting analysis shows that with recombination the variance of H decreases. Thus, there is a reduction in the proportion of runs in which $P < 0.05$, indicating that the test becomes more conservative when critical values for no recombination are used. This is demonstrated below by the H test P values reported for different rates of recombination.

To test for differentiation between populations, we used the S_{nn} test (15). This test is based on the idea that, in the presence of population structure, the nearest neighbors (in sequence space) of a haplotype will tend to be found in the same population as that haplotype more often than they would be under panmixia. This test has been shown to be more powerful than χ^2 tests of homogeneity for small samples, especially in the presence of recombination (16).

To summarize pairwise LD, we used the common measure, D' (17), a summary of pairwise LD normalized so that the range would be between −1 and 1. We used Fisher's Exact Test to determine whether the pairs of sites were in significant LD.

The recombination rate per generation was estimated by using the approach of Payseur and Nachman (18) (see also http://eebweb.arizona.edu/nachman/publications/data/microsats.html), which is based on a comparison of a physical map (the GB4 radiation hybrid map) and the Genethon genetic map (see ref. 18 for details). To estimate for the rate for this locus, we used two microsatellites, DXS1201 and DXS1043, in close proximity to the *MAO-A* gene (according to the National Center for Biotechnology Information map viewer). Estimates obtained by this method are effectively estimates of the rate of crossing-over alone, because gene conversion contributes little to the rate of gamete exchange for markers far apart (cf. ref. 19).

We estimated the population recombination rate, $C = 2Nr$ (r is the recombination rate in females) from this estimate of r and an estimate of N. An estimate of N can be obtained from diversity levels, assuming a mutation rate per generation. Here, N was estimated by dividing the summary of diversity, π, by the mutation rate, and a factor of 3 that takes into account the sex linkage. The mutation rate was estimated based on the divergence values (see below).

We used the estimate of C obtained from N and r estimates to gauge our power to detect a decay of LD under the standard neutral model. Specifically, we ran 10^4 coalescent simulations of the standard neutral model, with the population mutation rate estimated as π. These simulations conditioned on the sample size and our estimate of C. Singletons were excluded from each run. The power to detect a decay was estimated as the proportion of runs with a significantly negative correlation of D' values and distance.

Table 2. Population variability parameters for 18.8 kb of the human MAO-A gene

	Sample size	Length	S	bp/SNPs	Singletons	K	θ_W	π	θ_H	%D	Tajima's D	Fu and Li's D	H test	H test P value
Total group	56	18,820	41	459.02	11	46	0.047	0.050	0.105	1.01	0.34	−1.38	**−10.31**	**0.041**
Askenazies	13	18,820	26	723.85	6	12	0.045	0.042	0.111	1.03	−0.66	0.12	**−13.04**	**0.013**
Beduins	10	18,820	23	818.26	8	8	0.043	0.048	0.097	1.03	0.92	−0.89	**−9.26**	**0.030**
Africans	7	18,820	24	784.17	13	7	0.052	0.048	0.094	1.04	−0.45	**−3.35**	**−8.66**	**0.042**
Taiwanese	9	18,820	17	1107.06	1	7	0.033	0.046	0.057	1.06	**1.86**	**1.49**	−2.09	0.167
Asians	5	18,820	16	1176.25	14	4	0.041	0.036	0.073	1.06	−0.93	**−17.13**	**−6.92**	**0.040**
Mexicans	5	18,820	21	896.19	5	5	0.054	0.063	0.077	1.04	1.26	0.55	−2.54	0.198
Russians	7	18,820	26	723.85	7	7	0.056	0.064	0.112	1.02	0.74	0.18	−9.04	0.052

S is the number of polymorphic sites. D is the number of fixed differences per base pairs in percentage. π and θ values are given per bp in percentage. Significant values are in bold. P values are given for the H test. The designation of total group refers to all of the populations taken together such that some alleles that appear as singletons within an individual population may not be singletons in the total group.

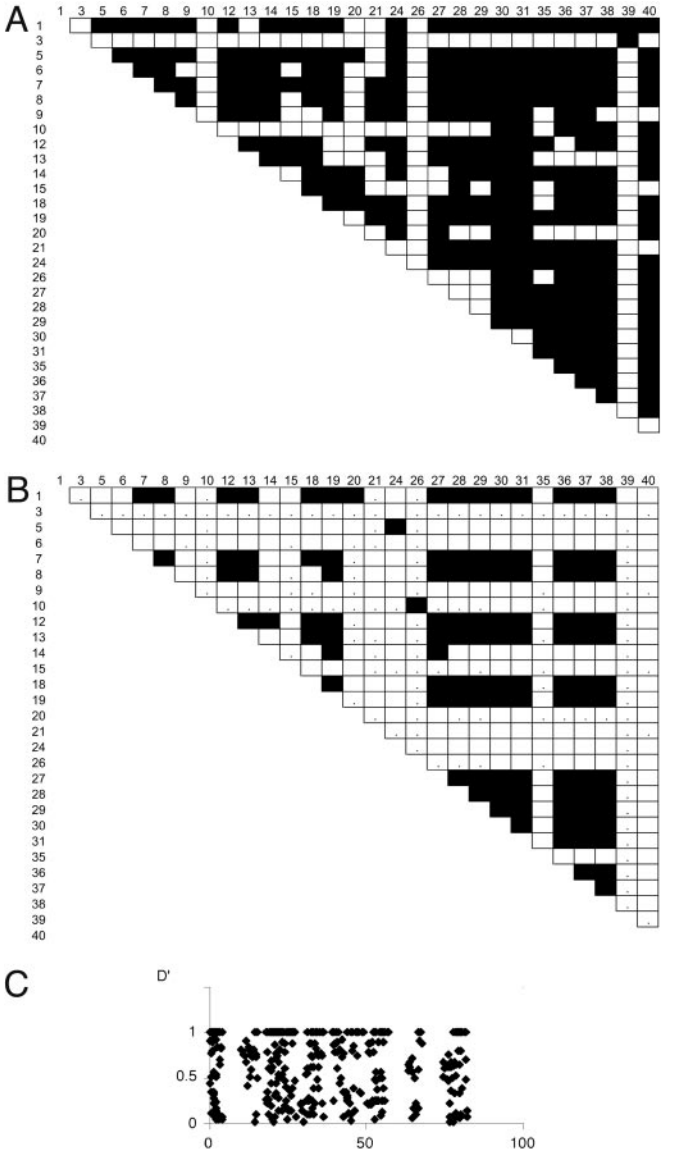


Fig. 2. Recombination and the pattern of LD at the MAO-A gene. (*A*) Number of pairs with four gametes. Singletons were excluded from this analysis, because they cannot have four gametes by definition. The variable number tandem repeat (VNTR) was also excluded from the analysis because it is not binary. A dark square indicates that four gametes were observed at a pair of sites. An open square indicates that fewer then four gametes were observed at a pair of sites. (*B*) Pairs in significant LD. Singletons were excluded from this analysis because they cannot be significant on mathematical grounds. The VNTR was also excluded because it is not binary and hence not suitable for LD analysis. Dark squares indicate pairs in significant LD at the 5% level, assessed by a Fisher's Exact Test. Dots indicate complete or absolute LD, which is not significant. (*C*) Scatterplot of the decay of LD with physical distance. Each point is the absolute D' value for a pair of sites separated by a given physical distance. Singletons were excluded from this analysis.

Results and Discussion

We sequenced a total of 1,053 kb of human genomic DNA. For each individual, we sequenced five regions totaling 18.8 kb and approximately evenly spaced across 90 kb of the MAO-A gene (Fig. 1). We surveyed 56 males from seven ethnogeographic groups. Two orthologous chimpanzee sequences were obtained to infer the ancestral state at each polymorphic site and to estimate the number of fixed differences between humans and chimpanzees. There were no shared polymorphisms between humans and chimpanzees. Human–chimpanzee divergence was 1% (191 sites) and did not differ significantly between exons and introns (0.98% and 1.02%, respectively) in our sample. Using the number of fixed differences between humans and chimpanzees and a time to the common ancestor of 250,000 generations (we used 500,000 because both branches of the species tree have to be considered), we obtained an estimate of the mutation rate, μ, of 2.1×10^{-8} per base pair per generation. This result is very similar to the estimate obtained from the study of six X-linked pseudogenes (20), suggesting a fairly relaxed level of selective constraint in this genomic region.

Within humans, we identified 41 segregating sites, 39 of which were single nucleotide substitutions and 11 of which were seen only once in the sample. The polymorphic sites were found on 46 distinct haplotypes (Table 1). We found 37 SNPs in introns and two synonymous mutations in exons. Thus, levels of diversity in exons and introns in our sample were quite similar, at roughly $\pi = 0.05\%$ per bp (Table 2). The nucleotide diversity levels observed for our data set are comparable to the average values reported for seven X-linked introns (21).

This genomic region experiences high levels of recombination, as estimated from a comparison of physical and genetic maps ($r = 2.1$–4.58 cM/Mb, depending on the choice of physical map and microsatellite, DXS1201 or DXS1043, respectively; ref. 18). The high number of recombination events is also apparent in the polymorphism data, with many pairs of sites (65%) showing all four gametes (Fig. 2*A*). The minimum number of recombination events needed to explain these data is $R_m = 20$ (22). A subset of these four gametes may be caused by multiple hits (a violation of the infinite-sites model) rather than by genetic exchange. In particular, transitions from CpG dinucleotide sites are known to occur at roughly 10 times the rate of other base substitutions (20). In this data set, however, only six of the 41 segregating sites occurred at CpG sites (two of which are singletons), and their exclusion still left $R_m = 19$. Thus, most of the four gametes probably were formed by recombination. Furthermore, exclusion of these sites from subsequent analyses did not alter our conclusions. Excluding the only other three segregating sites that were in other sequence motifs known to have higher mutation rates [mononucleotide tracts of length >5, and also potential DNA polymerase alpha-arrest motifs (TG(A/G)(A/G)GA and anything ≤3 bp of this motif; ref. 23] also did not further alter our results. One should generally be concerned about multiple mutation events when applying the infinite sites model to simulate the likelihood of a given data set. Nevertheless, for analysis of segregating sites in the current data set, application of the infinite sites model appears to be valid.

Despite this evidence for extensive recombination, there are a large number of pairs in significant LD (238 of 406 possible informative pairs, Fig. 2*B*). Although the pairwise comparisons are strongly suggestive of high levels of LD, because the comparisons are not independent, it is not possible to assign a meaningful statistic for the overall multiple comparisons. Accordingly, we examined the pattern of LD decay and found a clear-cut pattern in which no decay of LD was observed across the region of 90 kb (using only informative sites and D' as a measure of pairwise LD, $P = 0.138$ by a permutation test; Fig. 2*C*). Comparisons with other studies suggest that this pattern is highly atypical (cf. refs. 24 and 25); it should be noted, however, that LD is expected to extend further on the X chromosome because C is halved relative to autosomes.

To test whether our inability to detect a decay is unusual, we estimated the power we would have under the standard neutral assumptions, given an estimate of the population recombination rate, $C = 2Nr$ (see *Methods*). For this data set, we estimated N to be 8,412, consistent with estimates of effective population sizes for many loci across the genome (5,000 to 20,000 range) (21). Our estimates of N and r yield a range of C values from 32

to 73 for 90 kb. For C values within this range, simulations suggested that the power to detect a decay of LD by using D' is at least 95% under the standard neutral model. (Note that although multiple hits might contribute to the number of apparent recombination events, they make the excess LD less likely.) This power analysis suggests that a lack of decay over this scale is unexpected under the standard neutral model, given our independent estimate of r and observed levels of diversity.

A possible explanation for the observation of numerous sites with four gametes yet no decay of LD with distance is rampant gene conversion (but not short tract length conversion events, because if these were abundant we would expect clustering of four-gamete states along the diagonal in Fig. 2*A*, which is not the case), and very little crossing-over. Indeed, gene conversion will increase the rate of recombination on small scales (on the order of the mean tract length) but will have little effect on LD at larger scales (26). However, in the current case, there is independent evidence of a high rate of crossing-over (at least at the megabase scale).

Instead, the lack of decay over a distance of 90 kb may result from population structure. We applied the nearest-neighbor statistic (S_{nn}) permutation test (15) to assess whether there was evidence for differentiation among the populations in our sample. The S_{nn} is highly significant for the overall sample ($P < 10^{-7}$). Applications of the S_{nn} permutation test to all possible pairs of subpopulations revealed substructure for 11/21 pairs (at the 5% level). In particular, almost all comparisons with the Pygmy population detected significant pairwise differentiation.

To compare these results with other human data sets, we applied the S_{nn} to five other human data sets (27–30). Interestingly, the results were highly significant in all of them (results not shown). Although this fine-scale population differentiation appears to be frequent, a lack of LD decay over a comparable distance has not been reported. Because demographic forces should affect all loci similarly, it seems likely that for the *MAO-A* gene, the underlying population structure has been accentuated by a locus-specific influence.

Positive selection, such as through local adaptation, is one example of a locus-specific influence that might explain the lack of LD decay. Indeed, there are an accumulating number of examples where distinct selective pressures appear to apply in different environments (e.g., refs. 31–33). Alternatively, the differentiation may result from the occurrence of a global selective sweep (34) in this region. For simple models of structure with restricted migration, it is known that sweeps can increase differentiation between populations at linked neutral sites, as distinct haplotypes are fixed in different subpopulations (35). Accordingly, we performed tests for deviation from neutrality in this region. Because levels of diversity appear to be positively correlated with recombination rates in humans (21, 36), we applied the HKA (Hudson–Kreitman–Aguade) test (13) to the *MAO-A* and a locus with a similarly high levels of recombination, *DMD44* (28). *DMD44* is an intron of the x-linked *DMD* gene, for which recombination rates are estimated to be roughly $r = 4.3$ cM/Mb (28), and where the pattern of polymorphism appears to conform to standard neutral expectations (28). The HKA test is significant ($P < 0.01$), suggesting that levels of diversity at the *MAO-A* are below expected values, given recombination environment of the locus, supporting the hypothesis of a recent selective sweep.

We also used two tests of the frequency spectrum to examine the fit of the observed frequency spectrum of the polymorphic sites to that expected under neutrality. The first was Tajima's D, which was not significant for our data set ($D = 0.34$, $P = 0.70$ one-tailed, Table 2). The second test was the recent Fay and Wu H test (11). When applied to the entire data set, the H test is significant ($P = 0.04$) even under the assumption of no recombination. In light of the high rate of recombination at the locus, this P value is conservative. If we use our estimate of r and N to estimate C for the 18.8 kb and (conservatively) ignore the fact that the regions are not contiguous, we would estimate C to be 7–15 for this region. For these values of the population recombination rate, P is 0.017 and 0.009, respectively. A significant H test is thought to be the unique signature of a very recent sweep (11, 37).

Although population growth leads to significant values of Tajima's D, the same is not true for Fay and Wu's H statistic. Under population growth (fixing the diversity level), high frequency-derived alleles are less abundant than under a constant population size model. One way to think of this is that the internal branches of the genealogical tree relating the individuals in the sample tend to be smaller in the presence of population growth. There is therefore less opportunity for a mutation to be at high frequency in the sample. In fact, the H test is highly conservative in the presence of growth (M.P., unpublished results). In addition, population growth tends to decrease levels of LD (cf. refs. 19 and 38). Thus, population growth would make both our observation of a significant H test and high levels of LD less likely.

However, some caution has to be exercised in interpreting the results, as in the presence of structure, highly unequal sampling from the different populations can also lead to a significant H test (M.P., unpublished results). Yet, the H values are low for five of the seven subpopulations in our study and negative in every one (see Table 2), suggesting that no particular ethnogeographic group is responsible for the result.

Conclusions

The pattern of polymorphism at MAO-A reveals high levels of LD and substantial differentiation between populations. The H test and the low diversity levels suggest that the underlying population structure may have been accentuated by positive selection, potentially acting on *MAO-A*-related phenotypes. This finding should motivate further studies of this region as a candidate in genetic association studies. In particular, the next step might be to genotype unlinked markers in the same populations and to try and untangle the effects of demography and selection.

Note Added in Proof. One of the segregating sites reported in this paper (SNP 1.5-684) was later found to be an error. The recomputed Tajima's D is now 0.33 (instead of 0.34) and the H test is now -9.70 (instead of -10.31), with no change in the P values. There is no change of the decay of LD with distance. Thus this error has no effect on any of our conclusions.

We thank Sara Selig both for lab work assistance and useful discussions. We are grateful to Dick Hudson and Michael Nachman for comments on an earlier version of the manuscript. Gil McVean kindly performed the power analysis and the permutation test for D'. D.L. holds the Ralph and Lois Silver Chair in Human Genomics. This work was supported by the Crown Human Genome Center at the Weizmann Institute of Science, the Alfried Krupp Foundation, and a Israel Ministry of Science grant to the National Laboratory for Genome Infrastructure. K.S. is supported by the Israel Academy of Sciences-FIRST award.

1. Brunner, H. G., Nelen, M. R., Zandvoort, P., Abeling, N., Gennip, A. H., Wolter, E. C., Kuiper, M. A., Roper, H. H. & Oost, B. A. (1993) *Am. J. Hum. Genet.* **52,** 578–580.
2. Hotamisligil, G. S. & Breakefield, X. O. (1991) *Am. J. Hum. Genet.* **49,** 383–392.
3. Cases, O., Seif, I., Grimsby, J., Gaspar, P., Chen, K., Pournin, S., Muller, U., Aguet, M., Babinet, C. & Shih, J. C. (1995) *Science* **268,** 1763–1766.
4. Tivol, E. A., Shalish, C., Schuback, D. E., Hsu, Y. P. & Breakefield, X. O. (1996) *Am. J. Med. Genet.* **67,** 92–97.
5. Shin, J. C., Chen, K. & Ridd, M. J. (1999) *Annu. Rev. Neurosci.* **22,** 197–217.

6. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., *et al*. (1999) *Nat. Genet.* **22,** 231–238.
7. Halushka, M. K., Fan, J. B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. & Chakravarti, A. (1999) *Nat. Genet.* **22,** 239–247.
8. Clark, A. G. (1990) *Mol. Biol. Evol.* **7,** 111–122.
9. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 256–276.
10. Nei, M. & Li, W. H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5269–5273.
11. Fay, J. C. & Wu, C. I. (2000) *Genetics* **155,** 1405–1413.
12. Tajima, F. (1989) *Genetics* **123,** 585–595.
13. Hudson, R. R., Kreitman, M. & Aguade, M. (1987) *Genetics* **116,** 153–159.
14. Wall, J. D. & Hudson, R. R. (2001) *Mol. Biol. Evol.* **18,** 1134–1135.
15. Hudson, R. R. (2000) *Genetics* **155,** 2011–2014.
16. Hudson, R. R. (2001) *in Handbook of Statistical Genetics*, eds. Balding, D., Bishop, N. & Cannings, C. (Wiley, New York), pp. 309–324.
17. Lewontin, R. C. (1995) *Genetics* **140,** 377–388.
18. Payseur, B. A. & Nachman, M. W. (2000) *Genetics* **156,** 1285–1298.
19. Pritchard, J. K. & Przeworski, M. (2001) *Am. J. Hum. Genet.* **68,** 1–14.
20. Nachman, M. W. & Crowell, S. L. (2000) *Genetics* **156,** 297–304.
21. Nachman, M. W., Bauer, V. L., Crowell, S. L. & Aquadro, C. F. (1998) *Genetics* **150,** 1133–1141.
22. Hudson, R. R., Kaplan, N. L., Rieder, M. J., Taylor, S. L., Clark, A. G. & Nickerson, D. A. (1985) *Genetics* **111,** 147–164.
23. Templeton, A. R., Clark, A. G., Weiss, K. M., Nickerson, D. A., Boerwinkle, E. & Sing, C. F. (2000) *Am. J. Hum. Genet.* **66,** 69–83.
24. Clark, A. G., Weiss, K. M., Nickerson, D. A., Taylor, S. L., Buchanan, A., Stengard, J., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E. & Sing, C. F. (1998) *Am. J. Hum. Genet.* **63,** 595–612.
25. Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S. (2001) *Nature (London)* **411,** 199–204.
26. Andolfatto, P. & Nordborg, M. (1998) *Genetics* **148,** 1397–1399.
27. Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. & Sing, C. F. (1998) *Nat. Genet.* **19,** 233–240.
28. Nachman, M. W. & Crowell, S. L. (2000) *Genetics* **155,** 1855–1864.
29. Rieder, M. J., Taylor, S. L., Clark, A. G. & Nickerson, D. A. (1999) *Nat. Genet.* **22,** 59–62.
30. Harding, R. M., Fullerton, S. M., Griffiths, R. C., Bond, J., Cox, M. J., Schneider, J. A., Moulin, D. S. & Clegg, J. B. (1997) *Am. J. Hum. Genet.* **60,** 772–789.
31. Harris, E. E. & Hey, J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3320–3324.
32. Hamblin, M. T. & Di Rienzo, A. (2000) *Am. J. Hum. Genet.* **66,** 1669–1679.
33. Rana, B. K., Hewett-Emmett, D., Jin, L., Chang, B. H., Sambuughin, N., Lin, M., Watkins, S., Bamshad, M., Jorde, L. B., Ramsay, M., *et al*. (1999) *Genetics* **151,** 1547–1557.
34. Maynard-Smith, J. M. & Haigh, J. (1974) *Genet. Res.* **23,** 23–35.
35. Slatkin, M. & Wiehe, T. (1998) *Genet. Res.* **71,** 155–160.
36. Przeworski, M., Hudson, R. R. & Di Rienzo, A. (2000) *Trends Genet.* **16,** 296–302.
37. Otto, S. P. (2000) *Trends Genet.* **16,** 526–529.
38. Slatkin, M. (1994) *Genetics* **137,** 331–336.