cases), renal failure (two cases), coronary thrombosis, acute lymphocytic leukemia, amyotrophic lateral sclerosis, pancreatic carcinoma, pulmonary embolism, and aspiration pneumonia. For the five female subjects who did not die of AIDS, the causes of death were systemic lupus erythematosus, pancreatic carcinoma, liver failure (two cases), and abdominal sepsis secondary to renal transplantation. All six of the heterosexual male AIDS patients and three of the homosexual men had histories of intravenous drug abuse. Three of the women, two heterosexual men who did not have AIDS, and one homosexual man had histories of chronic alcohol abuse.

9. Criteria for inclusion of subjects in the study were as follows: (i) age 18 to 60, (ii) availability of medical records, (iii) in AIDS patients, statement in the records of at least one AIDS risk group to which the patient belonged (homosexual, intravenous drug abuser, or recipient of blood transfusions), (iv) no evidence of pathological changes in the hypothalamus, and (v) no damage to the INAH nuclei during removal of the brain or transection of these nuclei in the initial slicing of the brain. Fourteen specimens (over and above the 41 used in the study) were rejected for one of these reasons; in all cases the decision to reject was made before decoding.
10. INAH 1 is the same as the nucleus named the "sexually dimorphic nucleus" and reported to be larger in men than women [D. F. Swaab and E. Fliers, *Science* **228**, 1112 (1985)]. My results support the contention by Allen *et al.* (*6*) that this nucleus is not dimorphic.
11. The ratio of the mean INAH 3 volumes for the heterosexual and homosexual male groups was calculated. The INAH 3 volume values were then randomly reassigned to the subjects, and the ratio of means was recalculated. The procedure was repeated 1000 times, and the ordinal position of the actual ratio in the set of shuffled ratios was used as a measure of the probability that the actual difference between groups arose by chance. Only one of the shuffled ratios was larger than the actual ratio, giving a probability of 0.001.
12. Application of ANOVA or correlation measures failed to identify any confounding effects of age, race, brain weight, hospital of origin, length of time between death and autopsy, nature of fixative (10 or 20% formalin), duration of fixation, or, in the AIDS patients, duration of survival after diagnosis, occurrence of particular complications, or the nature of the complication or complications that caused death. There were no significant positive or negative correlations between the volumes of the four individual nuclei across the entire sample, suggesting that there were no unidentified common-mode effects such as might be caused by variations in tissue shrinkage. The mean brain weight for the women (1256 ± 41 g) was smaller than that for either the heterosexual (1364 ± 46 g) or the homosexual (1392 ± 32 g) men, but normalizing the data for brain weight had no effect on the results. There was no correlation between subject age and the volume of any of the four nuclei, whether for the whole sample or for any subject group; this finding does not necessarily conflict with the report in (*6*) of age effects in INAH 1, and possibly INAH 2, because in (*6*) a much wider range of ages was examined than was used in the present study.
13. J. S. Chmiel *et al.*, *Am. J. Epidemiol.* **126**, 568 (1987); W. Winkenstein, Jr., *et al.*, *J. Am. Med. Assoc.* **257**, 321 (1987).
14. In the largest relevant study [A. P. Bell and M. S. Weinberg, *Homosexualities: A Study of Diversity among Men and Women* (Simon and Schuster, New York, 1978)], nearly half the homosexual male respondents reported having had over 500 sexual partners.
15. R. Detels *et al.*, *J. AIDS* **2**, 77 (1989).
16. R. A. Gorski, J. H. Gordon, J. E. Shryne, A. M. Southam, *Brain Res.* **148**, 333 (1978).
17. K. D. Döhler *et al.*, *ibid.* **302**, 291 (1984); R. E. Dodson, J. E. Shryne, R. A. Gorski, *J. Comp. Neurol.* **275**, 623 (1988); G. J. Bloch and R. A. Gorski, *ibid.*, p. 613; R. W. Rhees, J. E. Shryne, R. A. Gorski, *Dev. Brain Res.* **52**, 17 (1990).
18. R. H. Anderson, D. E. Fleming, R. W. Rhees, E. Kinghorn, *Brain Res.* **370**, 1 (1986).
19. I thank the pathologists who made this study possible by providing access to autopsy tissue; P. Sawchenko, C. Rivier, S. Rivest, G. Torres, G. Carman, D. MacLeod, S. Lockery, and J. Rice for comments and suggestions; and B. Wamsley for assistance with preparation of the manuscript. Supported by a PHS Biomedical Research Support Grant to the Salk Institute.

29 January 1991; accepted 24 June 1991

Technical Comments

Forensic DNA Tests and Hardy-Weinberg Equilibrium

DNA tests based on biochemical procedures are being widely used for the identification of accused individuals (*1*). When the DNA pattern obtained from a specimen at the scene of a crime matches that obtained from a suspect, the prosecution seeks to prove that the suspect is the only possible source of the specimen. That inference depends on knowing something about the distribution of genotypes of the entire population of other people, any one of whom might be the actual criminal. In forensic applications of DNA testing so far, that inference has been based on an assumption of Hardy-Weinberg equilibrium (H-W). H-W justifies the assumption of statistical independence implicit in formulas used to calculate the probability that the DNA patterns of a specimen and of a suspect would match by chance alone. H-W can (*2, 3*), and sometimes does (*4, 5*), fail under realistic conditions.

To evaluate H-W, Devlin *et al.* (*6*) developed methods "to test for an overall excess or dearth of heterozygotes" in a sample of humans and applied these methods to a database provided by Lifecodes, Inc., one of the major vendors of services for forensic DNA testing. Devlin *et al.* have provided a useful service in drawing further attention to the problem of coalescence, that is, the appearance of a single, blurred band in autoradiographic films resulting from DNA fragments of different but similar size. However, their assertion that "the arguments so far presented against [H-W] are incorrect" is unconvincing for several reasons.

1) Devlin *et al.* reject the finding by Lander (*2*) of an excess of homozygosity in a Hispanic population. They use a data set drawn from a Caucasian population [reference 18 of (*4*)] and report no direct test of the logistic model for Hispanics, but instead use the model from the Caucasian data to interpret the Hispanic data. Their model is untested on the population from which Lander drew his data.

2) Devlin *et al.* have not used the data on apparent homozygotes. These are the data most likely to reveal an excess of homozygosity. They eliminate a subset of data that deviates from the expectations under H-W, and then test the remaining data for agreement with H-W. This predisposes them toward finding no deviation from H-W.

3) Devlin *et al.* note correctly that population subdivision must affect the overall number of heterozygotes, but they do not acknowledge that not all allelic classes need have too few heterozygotes relative to H-W. Some heterozygote classes may be in H-W, others in excess, and still others deficient: it is only the total of all heterozygotes that is necessarily deficient when the population is subdivided (*7*). Because the method of Devlin *et al.* tests only a subset of the heterozygote data, they might observe no deviation from H-W in that subset and incorrectly conclude that there is no departure from H-W overall, when, in fact, there is.

4) No information is given by Devlin *et al.* about how the populations of Caucasians, blacks, and Hispanics were sampled. There is no reason to believe that these samples are random or representative samples of the corresponding self-identified cultural groups in the United States. Hence inference from the given samples to the population at large, or to the entire self-identified cultural groups, is perilous. For example, the Hispanic population around New York is primarily of Puerto Rican origin, that around Miami of Cuban origin, and that in the southwestern states of Mexican origin; there are varying mixtures of other Hispanic origins in all three regions. If the Hispanic data studied by Devlin *et al.* were drawn primarily from the New York region, the conclusions could well be invalid for the other major Hispanic subpopulations separately or for all Hispanics as a group.

5) Devlin *et al.* say that it is not appropriate to pool data from different races, yet they treat "black" and "Hispanic" as if these were biologically meaningful races. The population identified as "black" in the United States is a continuum of individuals ranging from people of primarily African origin to people of primarily European origin (and

some with a mixture of Amerindian origin), all of whom have in common only that they identify themselves culturally as black. The term "black" has more cultural than biological meaning. The same is true for the term "Hispanic." Since the methods of Devlin *et al.* do not detect deviations from H-W for groups as internally diverse as blacks and Hispanics, one wonders whether the same methods would detect deviations even if all U.S. citizens were combined into a single group that is known to be heterogeneous at these loci (*8*).

6) Devlin *et al.* give no analysis of their methods' statistical power to detect deviations from H-W. They should have performed a simulation study of artificially generated data, with sample sizes corresponding to the actual sample sizes, with varying amounts of deviation from H-W. Applying their methods to these data sets would show how small a deviation from H-W can be detected.

7) For the Caucasian data set, Devlin *et al.* observe an increase in the ratio of the observed to the expected number of heterozygotes as a function of the difference τ between fragment pair lengths (in kilobases) for three different probes (D17S79, D14S13, and D2S44) and approximate the pattern of increase by a logistic curve (their figure 3). They interpret the observed patterns as a result only of coalescence. While coalescence contributes to the patterns observed, it would not seem to explain why the patterns are so different for the three loci. Their estimated thresholds for coalescence range from 0.099 kb for the D17S79 locus to 0.434 kb for the D2S44 locus. If the patterns were a simple function of the physical properties of DNA separation on a gel and of human visual discrimination, the thresholds for coalescence should be independent of the probed locus. Factors other than coalescence appear to be important; population-level processes may be among them. For example, pairs of fragments of similar length (small τ) could have more recent common ancestry and be more geographically concentrated than pairs of fragments of greatly differing length (large τ).

8) The complete set of data available to Devlin *et al.* could have been used in a simple way to detect excessive or deficient homozygosity relative to H-W. According to the points plotted in their figure 3, as τ increases, the observed heterozygosity comes to match and often to exceed the expected heterozygosity under H-W. If only coalescence were responsible, then for large τ, the observed heterozygosity should randomly fluctuate above and below the expected heterozygosity; if other factors besides coalescence, such as population substructure, were at work, the observed heterozygosity should consistently deviate from the expected heterozygosity for large τ. Inclusion of pairs of fragments with length differences greater than those shown in figure 3 of Devlin *et al.* would resolve this issue.

Devlin *et al.* write that their "results do not prove multiplicability across loci." We concur. Their results also do not prove multiplicability within loci. In the future, it may become possible to avoid using the product rule. As the DNA patterns of large numbers of individuals are now being assembled, one could simply determine the fraction of known individuals that match the given multilocus genotype. Realistic statistical methods of this sort are required for forensic DNA pattern matching to be as useful and as reliable as it has the potential to be.

JOEL E. COHEN
Rockefeller University, New York, NY 10021–6399
MICHAEL LYNCH
University of Oregon, Eugene, OR 97403–1210
CHARLES E. TAYLOR
University of California, Los Angeles, CA 90024–1606

REFERENCES

1. Office of Technology Assessment, *Genetic Witness: Forensic Uses of DNA Tests* (Government Printing Office, Washington, DC, 1990).
2. E. Lander, *Nature* **339**, 501 (1989).
3. J. E. Cohen, *Am. J. Hum. Genet.* **46**, 358 (1990).
4. L. D. Mueller, in *Forensic DNA Technology*, M. Farley and J. Harrington, Eds. (Lewis, Ann Arbor, MI, 1990), pp. 51–62.
5. S. J. Odelberg *et al.*, *Genomics* **5**, 915 (1989).
6. B. Devlin, N. Risch, K. Roeder, *Science* **249**, 1416 (1990).
7. M. Nei, *Evolution* **19**, 256 (1965).
8. I. Balazs *et al.*, *Am. J. Hum. Genet.* **44**, 182 (1989).

26 December 1990; accepted 25 June 1991

In the landmark DNA fingerprinting case *New York* v. *Castro* (*1*), it was found that the Hispanic population database compiled by Lifecodes, Inc., showed marked deviation from random mating for several genetic markers. This conclusion was based on a significant excess of homozygotes observed compared with the number expected. The issue was important because forensic labs calculate genotype frequencies by multiplying individual allele frequencies, a calculation that is correct only if the population does not contain genetically differentiated subgroups.

Devlin *et al.* (*2*) suggest that excess homozygosity was found because these studies ignored the fact that observed homozygotes actually consist of two classes: true homozygotes and "pseudohomozygotes" (heterozygotes with two alleles so close that they coalesce into one band on an autoradiogram) were counted. Using statistical inference, Devlin *et al.* conclude that the observed excess homozygosity is explained by the presence of pseudohomozygotes. In fact, the data do not support these conclusions.

Previous genetic analyses *have* directly addressed the issue of pseudohomozygotes. The expert witnesses in the *Castro* case used a mathematical correction virtually identical to that used by Devlin *et al.* The published summary (*3*) of the *Castro* case clearly describes the issue of pseudohomozygotes, stating that the analysis of expected frequency of homozygotes was based on "the empirical probability of randomly drawing two alleles from the population sample that are either identical or so close together as to be scored as a single band; the minimum size difference needed to discriminate between one versus two bands in Lifecodes' experiments was stated explicitly in testimony and in a paper."

The analyses in the *Castro* case and by Devlin *et al.* differ not in basic methodology but in a quantitative assumption: the minimum size difference (b) needed to ensure that two bands do not coalesce. The expert witnesses in the Castro case used estimates of $b \approx 0.6\%$ of molecular weight, based on Lifecodes' information about its control experiments (*1*, transcript, pp. 3608–3609 and Court Exhibit 13-D). In contrast, Devlin *et al.* used a statistical procedure to obtain much larger values of b, ranging from 2.8 to 4.0% of molecular weight (*4*). Thus, a key issue is Lifecodes' actual experimental resolution—a question that will be hard to resolve without independent re-examination of both the DNA samples and the autoradiograms, given the conflict between experimental report and statistical estimation.

Even if we assume the larger value for b, however, the conclusions reached by Devlin *et al.* are not justified for three reasons.

1) In general, tests of excess homozygosity have low statistical power to detect deviations. With a large value of b, the statistical power of the test is so greatly weakened (evidence of excess homozygosity is swamped when true homozygotes are lumped with a large number of heterozygotes) that failure to detect excess homozygosity does not mean it is not present. Without explicitly calculating the statistical power to detect deviations (which Devlin *et al.* do not discuss), it is difficult to reach meaningful conclusions.

2) Even with the use of a large value of b, Devlin *et al.* report a substantial deviation from expectation ($z = 2.25$), but conclude that it is not statistically significant. There are two problems with this conclusion: (i) Although Devlin *et al.* explicitly stated (in the paper's title) that they are testing for *excess* homozygosity, they use a two-sided test rather than the correct one-sided test for

detecting an excess. (ii) Excess homozygosity was previously asserted only for Hispanics, but Devlin *et al.* study the three loci in Caucasians, blacks, and Hispanics. They then use a multiple-testing correction based on nine tests (three loci in three populations), rather than a correction based on three tests. If one applies a three-test correction for the one-sided test, deviations are significant if $-z > 2.12$ (rather than, as Devlin *et al.* state, $|z| > 2.8$). If one applies this threshold to table 2 of Devlin *et al.*, there is significant excess homozygosity at locus D17S79 in Hispanics even after appropriate correction for multiple testing.

3) Most important, pseudohomozygosity caused by closely spaced alleles is not a sufficient explanation for the observed excess homozygosity. If pseudohomozygosity were the only problem, then evidence of excess homozygosity would be present for small b, but would disappear for a sufficiently large b. In fact, the opposite occurs. Excess homozygosity for D17S79 in Hispanics increases to a maximum of $-z = 3.85$ at $b = 7.2\%$ (Fig. 1) (*5*). (Similar results may also hold for Caucasians and for blacks at D17S79, although we do not have access to the raw data.) For such large values of b, excess homozygosity cannot be accounted for by bands clearly coalescing. The excess is consistent with the hypothesis of subpopulations with different allele frequencies at this locus. One cannot rule out possible experimental artifacts, but such explanations would be speculative in the absence of experimental demonstration.

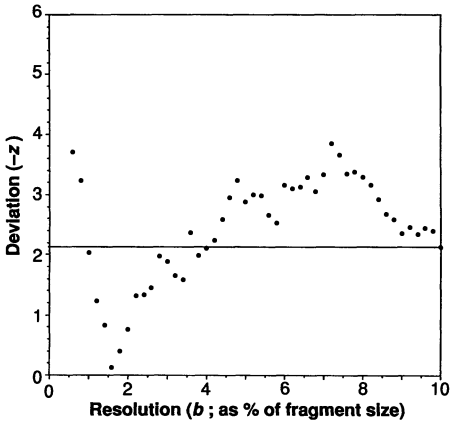

Fig. 1. Difference between observed and expected numbers of heterozygotes for D17S79 having alleles differing by more than b% (expressed in standard deviation units, z) plotted against b. Large deviations correspond to a deficit of observed heterozygotes, or an excess of apparent homozygotes (true homozygotes plus heterozygotes with alleles closer than b%). Horizontal line at $-z = 2.125$ represents 5% critical level for the Bonferroni three-test. Devlin *et al.* assume $b \approx 2.8\%$ for D17S79, which falls in the small range of values for which excess homozygosity is *not* found. Genotypes are taken from Lifecodes' database of 187 Hispanics used in (*1*, Court Exhibit II).

Because tests of excess homozygosity are statistically weak, the best way to examine population differentiation is the most straightforward way: compare actual samples from well-chosen ethnic subpopulations. Fortunately, such studies are now under way in several laboratories.

PHILIP GREEN
Washington University School of Medicine, St. Louis, MO 63110
ERIC S. LANDER
Massachusetts Institute of Technology, Cambridge, MA 02142

REFERENCES AND NOTES

1. *New York* v. *Castro*, 545 N.Y.S. 2d 985 (Sup. Ct. 1989).
2. B. Devlin, N. Risch, K. Roeder, *Science* **249**, 1416 (1990).
3. E. Lander, *Nature* **339**, 501 (1989).
4. The minimum size difference b needed to ensure that two bands do not coalesce can be stated in proportional terms (as a percentage of molecular weight) or in absolute terms (as a fixed number of base pairs)—with the former being slightly preferable because resolution tends to vary in proportion to molecular weight across the gel. Because Devlin *et al.* report b in absolute terms, we convert their figures for simplicity into proportional terms by dividing by median allele size. This yields $b = 4.0\%$ (= 434 bp per 11 kb) for D2S44, 3.1% (= 172 bp per 5.6 kb) for D14S13, and 2.8% (= 99 bp per 3.6 kb) for D17S79. Whether one measures b in absolute or proportional terms has no significant effect on the conclusions.
5. This finding is significant even at the threshold used by Devlin *et al.* and at the threshold appropriate for testing with a one-parameter family of b values. It contradicts reference 19 of Devlin *et al.*, which states that the results do not vary significantly for larger values of b.

11 October 1990; revised 14 January 1991; accepted 25 June 1991

Response: Human populations, especially in the United States, tend not to be homogeneous, but are composed of ethnically and racially diverse groups, such as Caucasians, blacks, and Hispanics. There are further subgroups within each of these major groupings. These subgroups are not entirely distinct, however, as matings among members of different subgroups often occur (*1*).

Statistical deviation from Hardy-Weinberg (H-W) predictions resulting from subpopulation admixture can occur only when two criteria are met: (i) there is limited mating among subgroups and (ii) there are differences in allele frequencies across the subgroups. Because of the nature of human populations, it is *never* assumed that such populations absolutely conform to H-W. Rather, what is assumed is that H-W gives an excellent approximation to the actual genotype frequencies because of gene flow (intermarriage) among subpopulations and only modest differences, at best, in allele frequencies among subgroups. In fact, these assumptions have been verified repeatedly in human genetics—a vast array of conventional genetic markers show no deviation from H-W.

Variable number tandem repeat (VNTR) loci pose difficulty in the assessment of H-W because of measurement error and coalescence, as we indicated in our original paper (*2*). Improper treatment of such loci can lead to the incorrect conclusion of homozygote excess. For example, the two references cited by Cohen *et al.* for deviation from H-W are unconvincing. In these reports, VNTR measurements were classified into arbitrary bins, which were subsequently treated as alleles for the test of H-W. In the cited study by Odelberg *et al.*, three loci appeared to have excess homozygosity. All three have large measurement error and coalescence. However, for some of the loci studied, measurement error was small, so that alleles could be unambiguously identified; for these loci, there was no significant excess of homozygotes.

Originally (*3*), Lander said there were "spectacular" deviations from H-W (homozygote excess) for two loci (D17S79 and D2S44) in Hispanics. We were surprised by this statement because of the extraordinary population dynamics required for such an excess. What we showed (*2*) is that the reported excess could be explained by coalescence. We did not say there are no "subgroups" within Hispanics. What we said was that such subgrouping was unlikely to contribute much to the stated homozygote excess and that admixture is unlikely to cause substantial deviations from H-W for the VNTR loci and populations studied.

Cohen *et al.* criticize our use of the Caucasian data set to estimate the probability of coalescence as a function of fragment size difference, and then our application of this function to Hispanics. The problem of coalescence is purely a physical process, independent of population characteristics. The probability of two bands coalescing on the gel (say of size 10.0 and 10.2 kb) could not possibly be different if the bands come from a Hispanic or if they come from a Caucasian. We estimated coalescence probabilities from the Caucasian data set because it was large. Furthermore, Cohen *et al.* question why our threshold values, b, for the three loci D17S79, D14S13, and D2S44 are so different and conclude that this difference is likely the result of population admixture. However, as we stated (*2*, reference 19), the probability of coalescence increases with fragment size for a given fragment size difference (see also the comment by Green and Lander). This explains the different values of b for the three loci, which have very different mean fragment sizes.

Cohen *et al.* criticize the exclusion of single band phenotypes ("apparent homozygotes") from our test. The set of apparent homozygotes is an indistinguishable mix-

ture of true homozygotes and close heterozygotes. The expected frequency of each of these is difficult to calculate without substantial assumptions and very precise estimates for measurement error and coalescence probabilities (*4*). By contrast, our test has no such requirements. Our test looks for an excess or dearth of heterozygotes that could not coalesce (that is, those with size difference $\tau > b$). By the symmetry of the test, it is equivalent to comparing the expected versus the observed number of phenotypes with $\tau \leq b$ (that is, homozygotes and close heterozygotes). If we, in fact, "eliminate a subset of the data that deviates from the expectations under H-W," its complement, the portion not eliminated, must also deviate from its expectation by the same amount in the opposite direction. We agree with Cohen *et al.* that population subdivision results only in an overall dearth of heterozygotes. However, we can imagine only trivial and unlikely examples where the deficiency is restricted to heterozygotes with allele size differences less than $b \approx 2\%$.

Cohen *et al.* and Green and Lander question the power of our test to detect admixture. Since our test is simply a sophisticated binomial test, it is easy to examine its power. Let z equal the critical value, n the sample size, t the proportional dearth of heterozygotes, and p the probability of a heterozygote being outside of the bound under H-W (unlikely to be misclassified as a homozygote). Its power is

$$\Phi(-z_\alpha\{np(1-p) - 2v(p)\}^{1/2}/ \{n(p-t)(1-p+t) - 2v(p)\}^{1/2})$$

where Φ indicates the cumulative normal distribution function evaluated from $-\infty$ to the argument and v is a small covariance defined in reference 17 of (*2*). Taking $p = 0.8$, $v = 0.1$ (*2*), and $z = -1.65$, we calculated the power for various levels of t and n (*5*). The results are given in Fig. 1. Lander (*3*) argued that the excess of homozygotes in the Hispanic sample for the loci D2S44 and D17S79 was 13 and 9%, respectively. As can be seen from Fig. 1, it would be relatively easy to detect such huge excesses of homozygotes (or equivalently huge dearths of heterozygotes).

Cohen *et al.* are surprised that we find no deviation from H-W for groups as "internally diverse" as blacks and Hispanics. However, it is internal diversity of *allele frequencies* that matters, and such diversity is unlikely to be large enough to create significant deviations from H-W. They also wonder if combining racial groups is sufficient to create significant violations of H-W. Of course, the answer depends on the magnitude of heterogeneity between groups. To examine this question, we used the Lifecodes database to

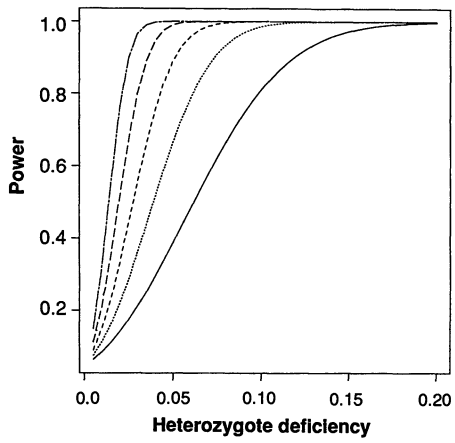

Fig. 1. Power curves for various values of t and n, where n = 100, 250, 500, 1000, and 2000, respectively, for the rightmost to leftmost line.

artificially create a mixed population of Caucasians and blacks, with 1000 genotypes from each. We then analyzed these samples for an excess of homozygotes using the critical value for a one-sided test of $z = -1.65$. We repeated this procedure 200 times. We rejected the null hypothesis of no admixture 21, 55.5, and 100% of the time for loci D17S79, D2S44, and D14S13, respectively. Our results correspond well with the histograms for these data for Caucasians and blacks (*7*): the histograms look similar for locus D17S79, somewhat different for locus D2S44, and different for locus D14S13.

To provide a further, heuristic example of admixture, we estimated the allele frequencies for locus D17S79 from the Lifecodes Caucasian population sample (*7*). It is reproduced as the top portion of Fig. 2. We then reversed the allele frequencies, giving the left-most allele the frequency of the right-most allele, and vice versa for all pairs of alleles. This allele distribution is displayed on the bottom portion of Fig. 2. Mixing these populations would clearly result in violent admixture. The proportion of homozygotes, if one ignores admixture, is 6%, whereas the actual proportion expected is 11%. Hence, extreme admixture results in only a 5% increase (that is, less than Lander's 9%). We created 500 genotypes from each of these allele distributions using Lifecodes' measurement error and rules for coalescence derived in (*2*). Simulating 30 populations from this mixture, we obtained 100% power to reject the null hypothesis of no admixture. No test statistic, $-z$, was less than 5.8.

We agree with Green and Lander that the interpretation of single-band phenotypes being due to homozygote excess or pseudohomozygosity depends critically on the assumed value of b, the coalescence threshold. However, their analysis is not "mathematically identical" to ours, nor do we agree on the plausibility of the low value of b (0.4%, 8) used in their analyses. Lander's analysis (*3*) involved the comparison of the observed number of "homozygotes" (actually single-band phenotypes) with the expected number of phenotypes with bands *within* a distance b apart. Our test compares the observed and expected number of heterozygotes *outside* a distance b apart. The former test is invalidated by the fact that, unless an unrealistically small value of b is used, there will be heterozygotes with an allele size difference of less than b which presumably are not included in the observed count if only single band phenotypes are considered. Our test does not have this limitation. The difference is Lander's assumption (*8*) that bands within a distance b apart will *always* coalesce, while bands greater than a distance b apart will *never* coalesce. This is obviously not the case; a logistic model gives a good fit to the probability of coalescence (*2*). Therefore, even if we used the same value of b as Green and Lander, our methods would not be the same. Furthermore, we are unable to find a derivation of a variance formula for their test statistic (observed minus expected) which takes into account the correlation in the observed and expected, as given in reference 17 of (*2*).

Although Lifecodes did not publish resolvability distances, one might expect their resolution values to be approximately 2% from their published reports (*6, 9*), in which it was stated that the resolvability of bands across lanes in a gel was 2%. Simple calculations show how unrealistic the 0.4% value is. A pair of alleles, both in the 10- to 12-kb range, which differ in size by 0.4%, would migrate in the gel to locations separated by about 0.3 mm. Similarly, bands in the 3- to 5-kb range would be separated by about 0.5 mm. Because the band widths themselves

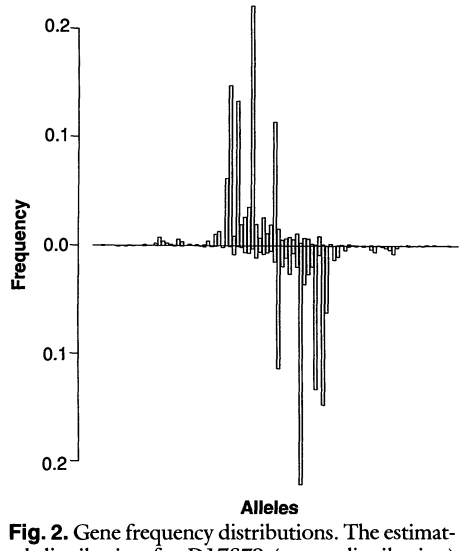

Fig. 2. Gene frequency distributions. The estimated distribution for D17S79 (upper distribution) and its mirror image (hanging distribution).

are 2 to 3 mm when visualized (*6, 9*), it would be impossible to resolve two bands of such similar size. Furthermore, for locus D17S79, 0.4% resolvability corresponds to 14 base pairs for a fragment of average size (3.5 kb), or only about one-third of the size of the repeat unit (38 base pairs). In fact, the values of b we derived through statistical inference correspond well with what would be expected from the experimental data (*6, 9*).

Green and Lander correctly point out the interesting behavior of our z statistic with an increasing value of b for locus D17S79; namely, z increases to near 0 at $b = 1.6\%$, then decreases to a negative peak of -3.85 at $b = 7.2\%$, and increases again. We replicated this graph with the Lifecodes Hispanic and Caucasian data sets (which, we understand, are generally available for scientific inquiry), but used absolute difference in band size (τ) instead of b (Fig. 3). Also, we extended the plot beyond 10% ($\tau = 0.35$) to about 20% (or $\tau = 0.70$). As can be seen in Fig. 3, for Hispanics, z returns to 0 at $\tau = 0.7$, indicating no difference between observed and expected at this point. It is noteworthy that the Caucasian data display a similar behavior, with a negative peak around 230 base pairs ($b \approx 7\%$).

Are the negative values for z, especially around $\tau = 0.26$, the result of deviation from H-W? Examination of the allele frequency distribution for this locus (Fig. 2) suggests a much more likely explanation. There are several common alleles (spikes) in the distribution that are separated by 2, 4, 6, 10, and 12 repeat units. Hence, a large number of heterozygotes separated by these numbers of repeats is expected. In our calculation of z, we assumed measurement errors of the two bands in a genotype to be independent. In fact, we now know they are positively correlated (*10*). Hence, when τ is approximately equal to the difference in common allele sizes, a larger number of heterozygotes should be observed than expected, while neighboring values of τ should have a deficiency of observed heterozygotes.

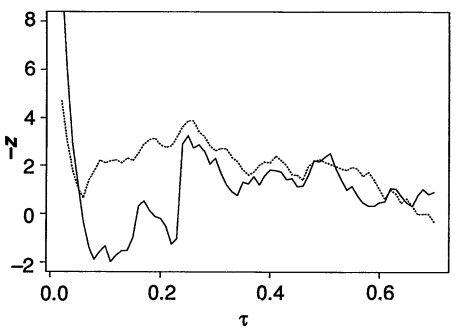

Fig. 3. The relationship between the negative of the z statistic and the bound τ. The z statistic is calculated by using the observed and expected number of heterozygotes whose difference in fragment size exceeds τ. The solid line is the relationship for the Caucasian data set, and the broken line is the relationship for the Hispanic data set.

The highest expected frequency of heterozygotes are those separated by six repeat units (228 base pairs). Correlated measurement errors increase substantially the observed heterozygotes around $\tau = 228$ and then decrease the observed heterozygotes afterward below what would be expected with no correlation. For example, in the extreme case of a correlation of 1, all such heterozygotes would differ by $\tau = 228$, and none by larger or smaller values.

The peaks and valleys in our Fig. 3 can be largely explained by the heterozygotes for common alleles and correlated measurement error. Because there is no statistical deficit of heterozygotes overall for $b > 2.8\%$ (beyond which coalescence is unlikely), the phenomena revealed in these plots must represent excess and deficiency of certain types of *heterozygotes* only (that is, those separated by various distances τ) and cannot be explained by homozygote excess, as "speculated" by Green and Lander.

Although we do not agree with Green and Lander's discussion of the correct significance level for our z test, examination of their figure 1 for D17S79 makes the controversy essentially moot. As can be seen in the figure, $-z$ is less than 2.12 (the 5% significance level for three *one-sided* tests) for all values of b between 1.0 and 3.3% (as also noted by Green and Lander), a region least likely to be affected by coalescence and the correlated measurement effect described above. Hence, we are still obliged to conclude that there is no statistical excess of homozygotes at this locus in Hispanics.

Furthermore, Lander (*3*) originally argued that there was an even larger homozygote excess in Hispanics (17% versus 4% expected) for locus D2S44 than for D17S79, yet this locus is absent from Green and Lander's discussion. In fact, for locus D2S44, our analysis (*2*) gave a slightly positive value for z (heterozygote excess). We have also examined the behavior of this locus for values of $b > 3.9\%$; it does not show the same behavior as D17S79, which is consistent with the fact that its frequency distribution lacks spikes (7). It appears that Green and Lander no longer believe there is a homozygote excess for D2S44 in Hispanics.

We agree with Green and Lander that tests of H-W, such as tests of excess homozygosity, are weak in detecting population differentiation. But that is the value of the H-W law—it is extremely robust with respect to population admixture—and why it is so generally applied, even when admixture is suspected. If a population shows no deviation from H-W, then comparing allele frequencies in subpopulations, as suggested by Green and Lander, may be of theoretical interest to population geneticists, but will have little practical consequence on the applicability of the H-W law.

We do not agree with the concern of Cohen *et al.* regarding the racial classification of individuals. In forensic application, it is usually possible only to classify an individual into one of the major racial groups, at best. Obtaining a genotype probability in this case can be conceived of as multiplying allele frequencies that were obtained by averaging across all subpopulations for that race. The correct value is the genotype frequency averaged across subpopulations. However, the former provides an excellent approximation of the latter unless there is *extreme* variation among subpopulations in allele frequencies. Such extreme heterogeneity is unlikely, even for blacks and Hispanics. In fact, given the vast empirical evidence supporting H-W in human populations, it is our belief that multiplication provides an appropriate approximation to genotype frequencies until proven otherwise. Furthermore, the suggestion of determining the proportion of "matching" genotypes from a given database is unrealistic. Given the enormous number of possible multilocus genotypes and their population infrequency, the observed sample will rarely, if ever, match any in the database, as has been the experience to date. Using a value of 0 seems far less satisfactory than using allele frequencies and multiplication.

B. DEVLIN
School of Medicine, Yale University, New Haven, CT 06510
NEIL RISCH
School of Medicine, Yale University
KATHRYN ROEDER
Department of Statistics, Yale University

REFERENCES AND NOTES

1. Hence, each major racial group should be conceived of as neither a "stew" with distinct components, nor as a completely blended "soup," but perhaps as a "ragout," where there is some blending together of the different constituents.
2. B. Devlin, N. Risch, K. Roeder, *Science* **249**, 1416 (1990).
3. E. Lander, *Nature* **339**, 501 (1989).
4. Reasonable approximations are not sufficient because lack of fit of the model could result from either inaccurate estimates of parameters or violation of H-W. These causes would be indistinguishable.
5. The z value of -1.65 is the 5% significance level for a single, one-sided test. This is the correct value for testing H-W against a specific alternative hypothesis (homozygote excess) for a single locus and one population. In our original analysis, we used a value of -2.8 to correct for the fact that nine tests were performed and to test for significant excess or deficiency of heterozygotes, as either alternative would be of interest.
6. B. Devlin *et al.*, *Am. J. Hum. Genet.* **48**, 662 (1991).
7. I. Balazs *et al.*, *ibid.* **44**, 182 (1989).
8. *New York* v. *Castro*, 545 N.Y.S. 2d 985 (Sup. Ct. 1989).
9. M. Baird *et al.*, *Am. J. Hum. Genet.* **39**, 489 (1986).
10. B. Devlin, N. Risch, K. Roeder, in preparation.

10 December 1990; revised 1 August 1991; accepted 7 August 1991