*Am. J. Hum. Genet.* 43:870–893, 1988

# Protein Variants in Hiroshima and Nagasaki: Tales of Two Cities

James V. Neel,* Chiyoko Satoh,† Peter Smouse,* Jun-ichi Asakawa,† Norio Takahashi,† Kazuaki Goriki,† Mikio Fujita,† Takeshi Kageoka,† and Ryuji Hazama†

*Department of Human Genetics, University of Michigan Medical School, Ann Arbor; and †Radiation Effects Research Foundation, Hiroshima

## Summary

The results of 1,465,423 allele product determinations based on blood samples from Hiroshima and Nagasaki, involving 30 different proteins representing 32 different gene products, are analyzed in a variety of ways, with the following conclusions: (1) Sibships and their parents are included in the sample. Our analysis reveals that statistical procedures designed to reduce the sample to equivalent independent genomes do not in population comparisons compensate for the familial cluster effect of rare variants. Accordingly, the data set was reduced to one representative of each sibship (937,427 allele products). (2) Both $\chi^2$-type contrasts and a genetic distance measure ($\Delta$) reveal that rare variants ($P < .01$) are collectively as effective as polymorphisms in establishing genetic differences between the two cities. (3) We suggest that rare variants that individually exhibit significant intercity differences are probably the legacy of tribal private polymorphisms that occurred during prehistoric times. (4) Despite the great differences in the known histories of the two cities, both the overall frequency of rare variants and the number of different rare variants are essentially identical in the two cities. (5) The well-known differences in locus variability are confirmed, now after adjustment for sample size differences for the various locus products; in this large series we failed to detect variants at only three of 29 loci for which sample size exceeded 23,000. (6) The number of alleles identified per locus correlates positively with subunit molecular weight. (7) Loci supporting genetic polymorphisms are characterized by more rare variants than are loci at which polymorphisms were not encountered. (8) Loci whose products do not appear to be essential for health support more variants than do loci the absence of whose product is detrimental to health. (9) There is a striking excess of rare variants over the expectation under the neutral mutation/drift/equilibrium theory. We suggest that this finding is primarily due to the relatively recent (in genetic time) agglomeration of previously separated tribal populations; efforts to test for agreement with the expectations of this theory by using data from modern cosmopolitan populations are exercises in futility. (10) All of these findings should characterize DNA variants in exons as more data become available, since the findings are the protein expression of such variants.

## I. Introduction

Between 1972 and 1984, the effort to use one-dimensional electrophoresis to search for an increase in mutations due to the atomic bombings of Hiroshima and Nagasaki resulted in 1,465,423 allele product determinations in the two cities, distributed over 30 different proteins representing 32 different gene products. An analysis of these data from the standpoint of radiation effects has just been published (Neel et al. 1988*b*). The purpose of the present paper is to present these data in a detail not previously available and then to develop their implications for a variety of questions, such as (1) differences in the origins of the inhabitants of the two cities, (2) the extent of interlocus variability in numbers of alleles finding expression as electrophoretic variants, (3) the distribution of allele frequencies in populations and its interpretation, and (4) the relationship between protein size, structure, and function, on the one hand, and the frequency of genetic variation, on the other. We will particularly direct attention toward both the implication of "recent" changes in human

Received June 7, 1988; revision received August 10, 1988.

Address for correspondence and reprints: James V. Neel, M.D., Department of Human Genetics, 4708 Medical Sciences II, Box 015, University of Michigan Medical School, Ann Arbor, MI 48109.



0002-9297/88/4306-0009$02.00

population structure for the interpretation of these findings and also to the implications of some of these findings for genetic variation at the DNA level. With reference to this latter topic, we point out that all of the 145 different variants included in the final data set correspond to nucleotide substitutions in coding (exon) regions. Because of the unusual circumstances that led to the collection of the large body of data we will analyze, comparable data on DNA polymorphisms will probably not become available during the next decade. Thus, the present findings will for the immediate future provide (indirect) insights not otherwise available into certain aspects of the population genetics of DNA markers.

## II. The Populations of Hiroshima and Nagasaki

The extensive genetic data available from Hiroshima and Nagasaki permit a more fine-grained comparison of allele frequencies in two populations than has ever before been possible. Such a comparison is of unusual interest both with reference to what is known of the earliest inhabitants of the Japanese Islands and with reference to historical developments subsequent to the discovery of Japan by the Western world in 1542.

During the periods of extensive glaciation, there were two principal land bridges between the present Japanese archipelago and the Asian mainland. The one was a broad wedge to the south, for which the Korean Peninsula may be regarded as the axis. The other, more tenuous, was to the north, connecting the present archipelago with Sakhalin and Siberia. The first ancestors of the modern Japanese are generally presumed to have reached the archipelago through one or both of these bridges, some 30,000–40,000 years ago, during the Pleistocene epoch. The possibility of even earlier human arrivals, perhaps utilizing a land bridge to the south, today represented by the Ryukyu Islands, cannot be excluded. The exact nature and timing of these and subsequent human migrations into Japan are still subject to much conjecture (see discussions in Japanese National Commission for UNESCO 1958; Smith and Beardsley 1962; Komatsu 1963, pp. xi, 64; Egami et al. 1981; Ogata 1981; Suzuki 1981; Aikens and Higuchi 1982, pp. xv, 354). By 8,000 B.C. a people with a well-defined culture complex and distinctive morphology, referred to as Jomon, were widely distributed throughout the Japanese islands. Whether they were the descendants of the earliest Japanese or of more recent immigrants is debatable. Beginning about 300 B.C. this Jomon culture was replaced in central Japan over a period of 500 years by the Yayoi culture complex, characterized, inter alia, by different styles in pottery, the introduction of rice cultivation, and an apparently rapid population increase. At the same time, there was a change in the mean skeletal characteristics of this region, suggesting an influx of people as well as culture (although the possible effect of a changing diet on skeletal traits should not be overlooked).

Genetic studies have now clearly demonstrated that the rapidly disappearing Ainu of the extreme north of Japan, who on the basis of morphology have in the past been labeled australoid or europoid, are, by virtue of the possession of the transferrin TF*DCHI allele and the Diego DI*A allele and by the absences of the red cell acid phosphatase allele ACP*C, the adenylate kinase allele AK*2, and the serum cholinesterase allele EI*A, definitely mongoloid (see Omoto 1975). Furthermore, detailed anthropometric studies suggest that the Ainu and the inhabitants both of the Ryukyu Islands (which include Okinawa) and of southwestern Kyushi (including Nagasaki) resemble each other somewhat more than they do the Japanese of northern Kyushu and Honshu (including Hiroshima) and that these resemblances are shared, especially by the Ainu, with the Jomon skeletal remains. Thus, it is tempting to postulate that a rapid expansion of the Yoyoi population (based on immigrants intermarrying with a locally differentiated population) effectively split the earlier Jomon peoples, of whom the Ainu are now an isolated relic, pushing them north and south. The people to the south may have had additional exchanges with the Chinese mainland and Taiwan. Biological evidence supportive of this hypothesis may be found in the similarity of the antigenic properties of the human hepatitis B virus in the extreme north and southwest of Japan (Nishioka 1984), as well as in the virtual restriction to these two regions of carriers of HTLV-I, the human T-cell leukemia virus (Ishida and Hinuma 1986). The Japanese are notable among mongoloids for the hirsutism of some individuals, especially evident in the Ainu but to be seen throughout Japan. Under the assumption of the rather direct relationship of the Ainu to Jomon man, this hirsutism of the central Japanese is perhaps their most obvious heritage from the Jomon people.

Early Chinese, Korean, and Japanese writings refer to the existence in Japan of numerous clan-tribes (*ujizoku*) which were presumably of considerable antiquity. With the passage of time, these gradually coalesced or the smaller and weaker were forcibly subjugated, resulting in the nuclei of later, feudal states. We would presume that these early tribes shared many

aspects of the isolation and breeding structure of some of the Amerindian tribes that we have studied in recent years, a point to which we return later.

An interesting recent development regarding Japanese origins is a phylogenetic tree, based on polymorphisms of mitochondrial DNA restriction sites, for residents of Shizuoka Prefecture, which is in the central of the Japanese islands (Honshu). This indicates two principal lineages thought to have diverged as early as about 125,000 years ago (Horai et al. 1984; Horai and Matsunaga 1986). (No Ainu were included in the present study, although the possibility of some Ainu admixture in the individuals sampled cannot be excluded.) A more recent study (Horai et al. 1987) on Okinawans revealed many new mitochondrial morphs, clearly differentiating the Okinawans from Japanese living on Honshu, but the evidence for two principal lineages was confirmed. This suggests a duality of Japanese origins antedating the separation of Jomon man from the mainland mongoloid stock.

As to more recent times, Nagasaki differs from Hiroshima in its rich contacts with Western and Chinese culture, beginning with its expansion in 1567 from a small fishing village to one of the principal points of Japanese contact with the outside world. We have briefly summarized these contacts elsewhere (Neel and Schull 1956). Suffice it to note here that it seems unlikely that either the contacts during the early period (≈1567–1639) or the limited contacts during Japan's self-imposed period of isolation from outside influences (1640–1853), when the Dutch maintained a very limited trade with Japan through Nagasaki, resulted in any significant genetic influence on the inhabitants of modern Nagasaki. The situation is less clear as regards the contacts after 1853, when Japan began to emerge from the period of seclusion and Nagasaki became one of the early Treaty ports and, in proportion to its size, probably the most cosmopolitan city in Japan, drawing entrepeneurs from all over the country. In addition, there are abundant data on the presence of Caucasians in Nagasaki, documented by a well-preserved Foreign Quarter which has become a major tourist attraction. Anecdotal evidence for mixed marriages abounds, some of which can be verified in the Foreign Cemetery in Nagasaki; but, despite considerable recent effort, we have been unable to generate anything approaching quantitative data on the degree of gene flow from Europeans and Americans. There has also for several hundred years been a relatively large Chinese colony in Nagasaki, but one apparently resulting in little admixture. People of Chinese ancestry, because of lack of a family (*honseki*) record, did not enter into the F1 Mortality Study from which our subjects were chosen. With the approach of World War II, there was a general exodus of non-Japanese citizens from Japan; we assume that in the event of an ethnically mixed marriage, the (Japanese) wife (and any small children) usually accompanied her husband. Those of mixed ancestry who remained behind assumed Japanese names and merged into the general population. As with so many historical developments of genetic interest, although the affairs of the chief players are well recorded, data to document overall admixture are scant (and even scantier concerning the ultimate fate of the admixed).

Throughout this period, Hiroshima remained a relatively quiet backwater in the stream of Japanese history with, so far as we can determine, virtually no foreign representation—and certainly none approaching that in Nagasaki. With the reopening of Japan in the nineteenth century, while there was no influx of foreigners to Hiroshima, the city did become the headquarters of the Second Japanese Army and an important military seaport. Thus, opportunities for the perturbation of the "original" gene pools of Hiroshima and Nagasaki are obvious. Even so, given the strictness with which the commoners were bound to their lord's fiefdom during the long period of feudalism in Japan (a system terminated by the Meiji Restoration of 1868), it is not unreasonable, despite the present-day mobility of the Japanese, to hope that the present data might supply some clues to both the remote and the historic past of these two cities. Rare variants, many of which must be derived from the private tribal polymorphisms of earlier days (see Neel 1978), can be expected to be especially useful in dissecting questions of origin such as those presented by modern Japanese (see Kirk 1975, 1982; Hill et. al. 1985; Wainscoat et al. 1986). Furthermore, any substantial recent admixture with a distinctly different ethnic group, as may have occurred in Nagasaki, should result in an increased number of different rare variants (but not in an increase in the total representation of such variants). We return to these matters in section V below.

## III. Technical Details

The design of the data collection and the electrophoretic techniques employed in generating these data have been referenced in a series of earlier contributions (Ferrell et al. 1977; Satoh et al. 1977; Ueda et al. 1977; Tanis et al. 1978; Neel et al. 1980, 1986, 1988*b*). The proteins examined were selected, on the basis of clarity

of resolution, from among approximately twice as many for which analytic techniques based on electrophoresis have been described. Throughout the study the standard supporting medium for electrophoresis was starch gel. Although as data collections progressed more convenient techniques became available, we held all procedures as constant as possible throughout the study, in order, in a study of radiation effects, to avoid the potential confounding of radiation effects by a changing mix of radiation-exposure histories and/or laboratory techniques. The findings on certain of the proteins which form the basis for the present study have been the object of separate communications, in which further biochemical details, especially regarding the nature of the variants, will be found. These latter communications are referenced in table 2, which summarizes the data. It is anticipated that in due course almost all of the findings on specific proteins will be the subject of separate communications. Thus the bare-bones data presentation of the present paper either has been or will be adequately fleshed out elsewhere.

## IV. The Data

Because all the variants encountered in the present study exhibit codominant inheritance, the findings can be presented in terms of allele counts. For all variants except the very well-known polymorphisms, a system of nonmenclature has been adopted in which a variant, if similar (not necessarily identical) in mobility to an already described variant in the literature, has been given the designation of that variant, followed by the city of discovery (HR or NG) and a numerical suffix indicating the order of discovery of variants of this general type. For the proteins for which other nomenclatures had already been established (G6PD and HB), the established conventions have been followed. No systematic attempt has been made to compare the variants encountered in the present study with apparently similar variants reported elsewhere. On the other hand, the variants of a particular isozyme or serum protein discovered during the present study have all been rigorously compared inter se, and each designated variant represents a unique electrophoretic class. The basic unit in the tabulation is a polypeptide (gene product) rather than a functional protein. Thirty proteins were examined. They represent the products of 32 genes. (For present purposes we score the [duplicated] α-hemoglobin loci as one.)

The data were collected in two phases. In phase I, a pilot study of the feasibility of evaluating mutation rates by electrophoretic studies of proteins, blood samples were collected from 4,648 adults undergoing periodic health appraisals at the Radiation Effects Research Foundation. When, in phase II, the formal genetic study was initiated, samples were collected from two age/sex/city-matched cohorts, one based on the children of parents one or both of whom had been proximally exposed at the time of the bombings (ATB), the other based on children of distally exposed parents. (A proximally exposed person was within 2,000 m of the hypocenter ATB, the zone within which there was a significant release of radiation, whereas a distally exposed person was 2,500 m or more from the hypocenter, at which distance kerma-in-air exposures only very rarely exceed 0.01 gray [Gy].) No attempt was made to obtain the necessary venous blood samples until the children were at least 13 years of age. The total number of children sampled from the two cohorts was 23,661. Since various of the determinations were introduced in the study at different times, and because there were some unsatisfactory determinations for all systems for which repeats could not be performed, the numbers of determinations vary from system to system. With three mutations in one cohort and four in the other, there is no evidence of a radiation effect (Neel et al. 1988*b*), and we feel justified both in combining the two series of children and in including the demonstrable mutations in the analysis.

Some of the adults examined during phase I were parents of the children examined in phase II. Furthermore, since we studied as many of the children in the two cohorts as feasible, each child being an independent sample for purposes of directly estimating the mutation rate, some of the children were related to each other as siblings. The precise matrix of genetic relationship between the individuals studied varies from genetic system to genetic system, as the number of electrophoretic determinations varies. For instance, with respect to PGM1 (see table 1A), for which the allele structure is especially complex, in addition to the 1,439 adults contacted in phase I who were subsequently represented by one or more children in phase II, there were 1,179 additional adults in phase I who had no children in phase II. Thus the total adult sample from the pilot study examined for this enzyme was 2,618. The number of children examined for this enzyme was 23,095, the sibship representation varying from 1 to 8. The total sample consists of 25,713 determinations (51,426 allele products). Applying a sampling adjustment suggested by Chakraborty (1978) for familial data of this type, we find that there are an estimated 11,851 independent genomes in the Hiroshima sample and 9,454 in the Nagasaki

**Table 1**

**Genetic Relationship Matrix of Individuals Studied, as Illustrated by Two of the 32 Systems**

A. PGM1, Hiroshima

| No. of Children Sampled/Sibship | No. of Parents Sampled: None | | One (579) | | Both (50) | | No Children in Series (732) |
|---|---|---|---|---|---|---|---|
| | Sibships | Children | Sibships | Children | Sibships | Children | |
| 1 | 5,269 | 5,269 | 354 | 354 | 17 | 17 | |
| 2 | 2,360 | 4,720 | 177 | 354 | 7 | 14 | |
| 3 | 394 | 1,182 | 39 | 117 | 1 | 3 | |
| 4 | 50 | 200 | 8 | 32 | | | |
| 5 | 6 | 30 | | | | | |
| 6 | | | 1 | 6 | | | |
| Total | 8,079 | 11,401 | 579 | 863 | 25 | 34 | |

B. PGM1, Nagasaki

| No. of Children Sampled/Sibship | No. of Parents Sampled: None | | One (766) | | Both (44) | | No Children in Series (446) |
|---|---|---|---|---|---|---|---|
| | Sibships | Children | Sibships | Children | Sibships | Children | |
| 1 | 3,259 | 3,259 | 363 | 363 | 12 | 12 | |
| 2 | 1,719 | 3,438 | 243 | 486 | 8 | 16 | |
| 3 | 614 | 1,842 | 104 | 312 | 2 | 6 | |
| 4 | 137 | 548 | 35 | 140 | | | |
| 5 | 27 | 135 | 14 | 70 | | | |
| 6 | 14 | 84 | 7 | 42 | | | |
| 7 | 4 | 28 | | | | | |
| 8 | 2 | 16 | | | | | |
| Total | 5,776 | 9,350 | 766 | 1,413 | 22 | 34 | |

*(continued)*

sample. Otherwise stated, the average sampled genome represents .868 independent genomes in Hiroshima and .784 independent genomes in Nagasaki. The corresponding fraction for the cities combined is .814. We have conducted a comparable analysis for TF, another locus for which variants were common (table 1B). For this protein, in addition to the 2,516 adults sampled in phase I who were parents of one or more of the children in phase II, there were 1,568 additional adults in Hiroshima and 556 in Nagasaki. Sample size adjustment reduces the numbers shown in table 1B to the equivalent of 13,279 independent genomes in Hiroshima and 9,804 in Nagasaki, i.e., an average equivalence per individual of .839 independent genomes in Hiroshima and .787 in Nagasaki. Note that for both of the proteins, the independent-genome equivalent is lower in Nagasaki than in Hiroshima, owing to the larger sibship sizes in Nagasaki.

The Chakraborty correction was not designed for a set of data in which rare variants play such a prominent role. Nonbinomial variation created by the aggregation of rare variants in sibships should be particularly prominent for such data. We have conducted two analyses of the findings with respect to PGM1 and TF (data not shown), analyses designed to explore the role of nonbinomial variation in biasing contrasts of variant frequencies in Hiroshima and Nagasaki. These two systems were selected for analysis because of a relatively high frequency of variants with an allele frequency $<.01$. For the first analysis of allele frequencies, all alleles represented in frequencies $<.001$ were pooled into an "other alleles" class. Two cross-city contrasts of allele frequencies were then undertaken, one based on correcting the observed allele frequencies by the Chakraborty factor, the other based on samples derived from selecting the first child of each sibship contributing to the phase II

**Table 1 (continued)**

C. TF, Hiroshima

| No. of Children Sampled/Sibship | No. of Parents Sampled: None | | One (1310) | | Both (180) | | No Children in Series (1,568) |
|---|---|---|---|---|---|---|---|
| | Sibships | Children | Sibships | Children | Sibships | Children | |
| 1 | 4,941 | 4,941 | 773 | 773 | 59 | 59 | |
| 2 | 2,229 | 4,458 | 412 | 824 | 28 | 56 | |
| 3 | 354 | 1,062 | 101 | 303 | 3 | 9 | |
| 4 | 40 | 160 | 22 | 88 | | | |
| 5 | 6 | 30 | 1 | 5 | | | |
| 6 | | | 1 | 6 | | | |
| Total | 7,570 | 10,651 | 1,310 | 1,999 | 90 | 124 | |

D. TF, Nagasaki

| No. of Children Sampled/Sibship | No. of Parents Sampled: None | | One (776) | | Both (44) | | No Children in Series (556) |
|---|---|---|---|---|---|---|---|
| | Sibships | Children | Sibships | Children | Sibships | Children | |
| 1 | 3,189 | 3,189 | 451 | 451 | 17 | 17 | |
| 2 | 1,666 | 3,332 | 307 | 614 | 13 | 26 | |
| 3 | 590 | 1,770 | 126 | 378 | 5 | 15 | |
| 4 | 134 | 536 | 43 | 172 | | | |
| 5 | 27 | 135 | 14 | 70 | 1 | 5 | |
| 6 | 9 | 54 | 12 | 72 | | | |
| 7 | 3 | 21 | 1 | 7 | | | |
| 8 | 2 | 16 | | | | | |
| Total | 5,620 | 9,053 | 954 | 1,764 | 36 | 63 | |

NOTE.—Numbers in parentheses are the numbers of adults sampled in phase I, categorized by their relationship to the children in phase II.

sample (i.e., by single selection). The second analysis was a cross-city comparison of the frequencies of the individual alleles in the "other alleles" category, where only those alleles with frequencies of <.001 were utilized. Two contrasts were again undertaken, as described above. Many of the entries in the cells of these latter tabulations are too small to justify use of standard $\chi^2$ approximations, and the analysis has been based on more exact methods described in the next section.

For the first analysis, based on the Chakraborty correction, a contrast of allele frequencies in Hiroshima and Nagasaki yields a $\chi^2$ value of 54.28 (df = 4, $P < .00001$) for the PGM1 locus. The corresponding analysis for the TF locus results in a $\chi^2$ of 14.43 (df = 3, $P < .0027$). A similar analysis based on single selection yielded a $\chi^2$ value of 49.51 (df = 4, $P < .00001$) for PGM1 and a value of 12.68 (df = 3, $P < .0054$) for TF. For the second analysis, the $\chi^2$ value on data corrected by the Chakraborty method were 30.39 (df = 16, $P < .0004$) for PGM1 and 26.13 (df = 14, $P < .0005$) for TF. A similar analysis based on the single-selection method results in a $\chi^2$ of 20.92 (df = 12, $P < .0126$) for PGM1 and of 20.15 (df = 12, $P < .028$) for TF. The differences between the two sets of $\chi^2$ values were much greater for the analysis of the variants with frequencies <.001. We suggest that the higher $\chi^2$ values and lower $P$ values yielded by those contrasts employing the "adjusted" data are primarily due to the influence of nonbinomial variation, whose role is more prominent at very low allele frequencies than at higher frequencies. The use of parental data in conjunction with offspring data would also introduce nonbinomial variation. Under the circumstances, it is more conservative to base the analyses on "single" selection from the offspring data, despite the sacrifice in numbers this entails. Table 2 presents the data on all the systems on

## Table 2

**Allele Numbers and Frequencies at the 32 Loci Examined in the Present Study**

| Enzyme and Allele | Hiroshima | | Nagasaki | | Total | |
|---|---|---|---|---|---|---|
| | *N* | Frequency | *N* | Frequency | *N* | Frequency |
| Haptoglobin (HP): | | | | | | |
| 1 | 4,556 | .255898 | 3,210 | .244851 | 7,766 | .251213 |
| 2 | 13,233 | .743260 | 9,887 | .754157 | 23,120 | .747881 |
| 1ANG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| 1BHR1 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| 1CHR1 | 3 | .000169 | 0 | .000000 | 3 | .000097 |
| 2AHR1 | 2 | .000112 | 0 | .000000 | 2 | .000065 |
| 2BHR1 | 6 | .000337 | 2 | .000153 | 8 | .000259 |
| 2BNG1 | 0 | .000000 | 2 | .000153 | 2 | .000065 |
| 2SNG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| 3AHR1 | 2 | .000112 | 5 | .000381 | 7 | .000226 |
| 3ANG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| 3ANG2 | 1 | .000056 | 1 | .000076 | 2 | .000065 |
| Total | 17,804 | 1.000000 | 13,110 | 1.000000 | 30,914 | 1.000000 |
| Transferrin (TF):[a] | | | | | | |
| C | 17,723 | .987904 | 13,108 | .991528 | 30,831 | .989442 |
| BHR1 | 7 | .000390 | 0 | .000000 | 7 | .000225 |
| BHR2 | 56 | .003122 | 19 | .001437 | 75 | .002407 |
| BHR3 | 14 | .000780 | 9 | .000681 | 23 | .000738 |
| BHR4 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| BHR5 | 9 | .000502 | 7 | .000530 | 16 | .000513 |
| BHR6 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| BHR7 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| BNG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| DCH1 | 113 | .006299 | 67 | .005068 | 180 | .005777 |
| DHR1 | 3 | .000167 | 6 | .000454 | 9 | .000289 |
| DHR3 | 3 | .000167 | 0 | .000000 | 3 | .000096 |
| DHR4 | 3 | .000167 | 0 | .000000 | 3 | .000096 |
| DHR5 | 6 | .000334 | 1 | .000076 | 7 | .000225 |
| DHR6 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| DNG2 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,940 | 1.000000 | 13,220 | 1.000000 | 31,160 | 1.000000 |
| Albumin (ALB): | | | | | | |
| A | 17,923 | .998941 | 13,191 | .997806 | 31,114 | .998460 |
| HR1 | 5 | .000279 | 0 | .000000 | 5 | .000160 |
| HR2 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| NG1 | 12 | .000669 | 23 | .001740 | 35 | .001123 |
| NG2 | 0 | .000000 | 5 | .000378 | 5 | .000160 |
| NG3 | 1 | .000056 | 1 | .000076 | 2 | .000064 |
| Total | 17,942 | 1.00000 | 13,220 | 1.000000 | 31,162 | 1.000000 |
| Ceruloplasmin (CP):[b] | | | | | | |
| B | 17,764 | .990300 | 13,074 | .990905 | 30,838 | .990556 |
| CNG1 | 166 | .009254 | 112 | .008489 | 278 | .008930 |
| CHR1 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| CHR2 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| AHR1 | 3 | .000167 | 0 | .000000 | 3 | .000096 |
| ANG1 | 3 | .000167 | 7 | .000531 | 10 | .000321 |
| ANG2 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,938 | 1.000000 | 13,194 | 1.000000 | 31,132 | 1.000000 |

*(continued)*

**Table 2 (continued)**

| Enzyme and Allele | Hiroshima | | Nagasaki | | Total | |
|---|---|---|---|---|---|---|
| | *N* | Frequency | *N* | Frequency | *N* | Frequency |
| Adenosine deaminase (ADA), E.C.3.5.4.5: | | | | | | |
| 1 | 17,481 | .974306 | 12,803 | .968750 | 30,284 | .971949 |
| 2 | 454 | .025304 | 410 | .031023 | 864 | .027730 |
| 4NG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| 5HR1 | 7 | .000390 | 1 | .000076 | 8 | .000257 |
| 6NG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,942 | 1.000000 | 13,216 | 1.000000 | 31,158 | 1.000000 |
| 6-Phosphogluconate dehydrogenase (PGD), E.C.1.1.1.44: | | | | | | |
| A | 16,197 | .903145 | 11,889 | .899864 | 28,086 | .901753 |
| C | 1,734 | .096688 | 1,312 | .099304 | 3,046 | .097797 |
| HR3 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| HR4 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| HG1 | 0 | .000000 | 2 | .000151 | 2 | .000064 |
| NG2 | 0 | .000000 | 2 | .000151 | 2 | .000064 |
| NG3 | 1 | .000056 | 3 | .000227 | 4 | .000128 |
| NG4 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| NG5 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| NG6 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| NG7 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,934 | 1.000000 | 13,212 | 1.000000 | 31,146 | 1.000000 |
| Adenylate kinase-1 (AK1), E.C.2.7.4.3: | | | | | | |
| 1 | 17,490 | 1.000000 | 13,076 | 1.000000 | 30,566 | 1.000000 |
| Total | 17,490 | 1.00000 | 13,076 | 1.000000 | 30,566 | 1.000000 |
| Phosphoglucomutase-1 (PGM1) E.C.2.7.5.1:[c] | | | | | | |
| 1 | 13,014 | .749395 | 10,230 | .779310 | 23,244 | .762273 |
| 2 | 4,057 | .233617 | 2,713 | .206673 | 6,770 | .222018 |
| 7 | 253 | .014569 | 135 | .010284 | 388 | .012724 |
| 3NG1 | 19 | .001094 | 27 | .002057 | 46 | .001509 |
| 4HR2 | 3 | .000173 | 0 | .000000 | 3 | .000098 |
| 4HR3 | 2 | .000115 | 0 | .000000 | 2 | .000066 |
| 4NG1 | 0 | .000000 | 1 | .000076 | 1 | .000033 |
| 5HR1 | 1 | .000058 | 0 | .000000 | 1 | .000033 |
| 6HR1 | 3 | .000173 | 0 | .000000 | 3 | .000098 |
| 6HR2 | 3 | .000173 | 2 | .000152 | 5 | .000164 |
| 6HR3 | 3 | .000173 | 1 | .000076 | 4 | .000131 |
| 6NG2 | 5 | .000288 | 5 | .000381 | 10 | .000328 |
| 6NG3 | 0 | .000000 | 1 | .000076 | 1 | .000033 |
| 7HR1 | 1 | .000058 | 2 | .000152 | 3 | .000098 |
| 8NG1 | 1 | .000058 | 9 | .000686 | 10 | .000328 |
| 9HR1 | 1 | .000058 | 0 | .000000 | 1 | .000033 |
| 9NG2 | 0 | .000000 | 1 | .000076 | 1 | .000033 |
| Total | 17,366 | 1.000000 | 13,127 | 1.000000 | 30,493 | 1.000000 |

*(continued)*

**Table 2 (continued)**

| Enzyme and Allele | Hiroshima | | Nagasaki | | Total | |
|---|---|---|---|---|---|---|
| | *N* | Frequency | *N* | Frequency | *N* | Frequency |
| Phosphoglucomutase-2 (PGM2) E.C.2.7.5.1:[d] | | | | | | |
| 1 | 17,373 | .999482 | 13,126 | .999695 | 30,499 | .999574 |
| 2HR1 | 1 | .000058 | 0 | .000000 | 1 | .000033 |
| 2HR2 | 1 | .000058 | 0 | .000000 | 1 | .000033 |
| 5NG2 | 0 | .000000 | 1 | .000076 | 1 | .000033 |
| 9HR1 | 1 | .000058 | 0 | .000000 | 1 | .000033 |
| 9NG1 | 6 | .000345 | 3 | .000228 | 9 | .000295 |
| Total | 17,382 | 1.000000 | 13,130 | 1.000000 | 30,512 | 1.000000 |
| Phosphoglucomutase-3 (PGM3) E.C.2.7.5.1: | | | | | | |
| 1 | 4,102 | .692671 | 2,747 | .692637 | 6,849 | .692658 |
| 2 | 1,820 | .307329 | 1,219 | .307363 | 3,039 | .307342 |
| Total | 5,922 | 1.000000 | 3,966 | 1.000000 | 9,888 | 1.000000 |
| Acid phosphatase (ACP1), E.C.3.1.3.2: | | | | | | |
| A | 3,642 | .209214 | 2,616 | .207685 | 6,258 | .208572 |
| B | 13,764 | .790671 | 9,980 | .792315 | 23,744 | .791361 |
| AHR1 | 1 | .000057 | 0 | .000000 | 1 | .000033 |
| BHR1 | 1 | .000057 | 0 | .000000 | 1 | .000033 |
| Total | 17,408 | 1.000000 | 12,596 | 1.000000 | 30,004 | 1.000000 |
| Triosephosphate isomerase (TPI) E.C.5.3.1.1:[e] | | | | | | |
| 1 | 15,620 | .999872 | 11,724 | .999829 | 27,344 | .999854 |
| 2HR1 | 1 | .000064 | 0 | .000000 | 1 | .000037 |
| 2NG1 | 0 | .000000 | 1 | .000085 | 1 | .000037 |
| 3HR1 | 1 | .000064 | 0 | .000000 | 1 | .000037 |
| 4NG1 | 0 | .000000 | 1 | .000085 | 1 | .000037 |
| Total | 15,622 | 1.000000 | 11,726 | 1.000000 | 27,348 | 1.000000 |
| Nucleoside phosphorylase (NP), E.C.2.4.2.1: | | | | | | |
| 1 | 17,084 | .998947 | 12,411 | .999436 | 29,495 | .999153 |
| 2NG1 | 1 | .000000 | 7 | .000564 | 7 | .000237 |
| 2NG2 | 17 | .000994 | 0 | .000000 | 17 | .000576 |
| 3HR1 | 1 | .000058 | 0 | .000000 | 1 | .000034 |
| Total | 17,102 | 1.000000 | 12,418 | 1.000000 | 29,520 | 1.000000 |
| Esterase B (ESB) E.C.3.1.1.1: | | | | | | |
| 1 | 16,560 | 1.000000 | 11,620 | 1.000000 | 28,180 | 1.000000 |
| Total | 16,560 | 1.000000 | 11,620 | 1.000000 | 28,180 | 1.000000 |
| Esterase D (ESD) E.C.3.1.1.1: | | | | | | |
| 1 | 10,775 | .638556 | 7,769 | .625322 | 18,544 | .632944 |
| 2 | 6,099 | .036144 | 4,653 | .374517 | 10,752 | .366988 |
| 6NG1 | 0 | .000000 | 1 | .000080 | 1 | .000034 |
| 6NG2 | 0 | .000000 | 1 | .000080 | 1 | .000034 |
| Total | 16,874 | 1.000000 | 12,424 | 1.000000 | 29,298 | 1.000000 |

*(continued)*

**Table 2 (continued)**

| Enzyme and Allele | Hiroshima | | Nagasaki | | Total | |
|---|---|---|---|---|---|---|
| | *N* | Frequency | *N* | Frequency | *N* | Frequency |
| Esterase A1 (ESA1), E.C.1.1.1: | | | | | | |
| A | 16,692 | .999162 | 12,205 | .999263 | 28,897 | .999205 |
| BNG1 | 0 | .000000 | 6 | .000491 | 6 | .000207 |
| BHR1 | 9 | .000539 | 1 | .000082 | 10 | .000346 |
| CHR1 | 5 | .000299 | 2 | .000164 | 7 | .000242 |
| Total | 16,706 | 1.000000 | 12,214 | 1.000000 | 28,920 | 1.000000 |
| PeptidaseA (PEPA), E.C.3.4.11.*: | | | | | | |
| 1 | 17,922 | .999554 | 13,143 | .999012 | 31,065 | .999324 |
| 4HR1 | 8 | .000446 | 13 | .000988 | 21 | .000676 |
| Total | 17,930 | 1.000000 | 13,156 | 1.000000 | 31,086 | 1.000000 |
| Peptidase B (PEPB) E.C.3.4.11.*: | | | | | | |
| 1 | 17,934 | .999443 | 13,215 | .999622 | 31,149 | .999519 |
| 2NG1 | 1 | .000056 | 1 | .000076 | 2 | .000064 |
| 3HR1 | 8 | .000446 | 4 | .000303 | 12 | .000385 |
| 8HR1 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| Total | 17,944 | 1.000000 | 13,220 | 1.000000 | 31,164 | 1.000000 |
| Glucose isomerase (GPI) E.C.5.3.1.9: | | | | | | |
| 1 | 17,888 | .996990 | 13,143 | .994326 | 31,031 | .995860 |
| 2NG1 | 5 | .000279 | 9 | .000681 | 14 | .000449 |
| 3HR1 | 10 | .000557 | 2 | .000151 | 12 | .000385 |
| 4HR1 | 21 | .001170 | 48 | .003631 | 69 | .002214 |
| 5HR1 | 6 | .000334 | 7 | .000530 | 13 | .000417 |
| 5NG1 | 9 | .000502 | 9 | .000681 | 18 | .000578 |
| 7HR1 | 2 | .000111 | 0 | .000000 | 2 | .000064 |
| 9HR1 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| Total | 17,942 | 1.000000 | 13,218 | 1.000000 | 31,160 | 1.000000 |
| Lactate dehydrogenase (LDHB) E.C.1.1.1.27: | | | | | | |
| B | 17,932 | 1.000000 | 13,203 | .999773 | 31,135 | .999904 |
| BNG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| BNG2 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| BNG3 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,932 | 1.000000 | 13,206 | 1.000000 | 31,138 | 1.000000 |
| Lactate dehydrogenase (LDHA), E.C.1.1.1.27: | | | | | | |
| A | 17,932 | 1.000000 | 13,205 | .999924 | 31,137 | .999968 |
| ANG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,932 | 1.000000 | 13,206 | 1.000000 | 31,138 | 1.000000 |
| Malate dehydrogenase-1 (MDH1), E.C.1.1.1.37: | | | | | | |
| 1 | 17,943 | .999944 | 13,215 | .999924 | 31,158 | .999936 |
| 3NG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| 7HR2 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| Total | 17,944 | 1.000000 | 13,216 | 1.000000 | 31,160 | 1.000000 |

*(continued)*

**Table 2 (continued)**

| Enzyme and Allele | Hiroshima | | Nagasaki | | Total | |
|---|---|---|---|---|---|---|
| | *N* | Frequency | *N* | Frequency | *N* | Frequency |
| Carbonic anhydrase-1 (CA1), E.C.4.2.1.1: | | | | | | |
| 1 | 17,900 | .999442 | 13,153 | .999772 | 31,053 | .999582 |
| HR1 | 8 | .000447 | 0 | .000000 | 8 | .000258 |
| HR2 | 2 | .000112 | 2 | .000152 | 4 | .000129 |
| NG1 | 0 | .000000 | 1 | .000076 | 1 | .000032 |
| Total | 17,910 | 1.000000 | 13,156 | 1.000000 | 31,066 | 1.000000 |
| Carbonic anhydrase 2 (CA2) E.C.4.2.1.1: | | | | | | |
| 1 | 17,917 | .999944 | 13,206 | 1.000000 | 31,123 | .999968 |
| HR1 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| Total | 17,918 | 1.000000 | 13,206 | 1.000000 | 31,124 | 1.000000 |
| Glucose-6-phosphate dehydrogenase (G6PD) E.C.1.1.1.49:[f] | | | | | | |
| N | 6,170 | .998220 | 4,488 | .998443 | 10,658 | .998314 |
| HIROSH | 1 | .000162 | 0 | .000000 | 1 | .000094 |
| USHIDA | 2 | .000324 | 0 | .000000 | 2 | .000187 |
| KANNON | 1 | .000162 | 0 | .000000 | 1 | .000094 |
| UBE-LK | 1 | .000162 | 1 | .000222 | 2 | .000187 |
| NAGASA | 0 | .000000 | 1 | .000222 | 1 | .000094 |
| NISHID | 0 | .000000 | 1 | .000222 | 1 | .000094 |
| OTONAS | 0 | .000000 | 2 | .000445 | 2 | .000187 |
| NAKAJI | 1 | .000162 | 0 | .000000 | 1 | .000094 |
| Unclassified | 5 | .000809 | 2 | .000445 | 7 | .000656 |
| Total | 6,181 | 1.000000 | 4,495 | 1.000000 | 10,676 | 1.000000 |
| Glutamate-oxaloacetate transaminase-1 (GOT1) E.C.2.6.1.1:[g] | | | | | | |
| 1 | 13,150 | .981197 | 9,927 | .985604 | 23,077 | .983088 |
| 2HR1 | 207 | .015445 | 117 | .011616 | 324 | .013803 |
| 3NG1 | 20 | .001492 | 10 | .000993 | 30 | .001278 |
| 4NG1 | 22 | .001642 | 14 | .001390 | 36 | .001534 |
| 5NG1 | 1 | .000075 | 2 | .000199 | 3 | .000128 |
| 6HR1 | 2 | .000149 | 1 | .000099 | 3 | .000128 |
| 7NG1 | 0 | .000000 | 1 | .000099 | 1 | .000043 |
| Total | 13,402 | 1.000000 | 10,072 | 1.000000 | 23,474 | 1.000000 |
| Glutamate-pyruvate transaminase-1 (GPT1) E.C.2.6.1.1: | | | | | | |
| 1 | 7,943 | .593071 | 5,905 | .586453 | 13,848 | .590231 |
| 2 | 5,417 | .404465 | 4,126 | .409773 | 9,543 | .406743 |
| 4NG1 | 16 | .001195 | 16 | .001589 | 32 | .001364 |
| 6NG1 | 3 | .000224 | 3 | .000298 | 6 | .000256 |
| 6HR1 | 1 | .000075 | 0 | .000000 | 1 | .000043 |
| 7HR1 | 1 | .000075 | 0 | .000000 | 1 | .000043 |
| 8HR1 | 2 | .000149 | 0 | .000000 | 2 | .000085 |
| 8HR2 | 2 | .000149 | 0 | .000000 | 2 | .000085 |
| 8NG1 | 7 | .000523 | 19 | .001887 | 26 | .001108 |
| 9HR1 | 1 | .000075 | 0 | .000000 | 1 | .000043 |
| Total | 13,393 | 1.000000 | 10,069 | 1.000000 | 23,462 | 1.000000 |

*(continued)*

**Table 2 (continued)**

| Enzyme and Allele | Hiroshima | | Nagasaki | | Total | |
|---|---|---|---|---|---|---|
| | *N* | Frequency | *N* | Frequency | *N* | Frequency |
| Hemoglobin β (HBB): | | | | | | |
| N | 17,931 | .999833 | 13,218 | 1.000000 | 31,149 | .999904 |
| HIKARI | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| HIJIYA | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| PROVID | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| Total | 17,934 | 1.000000 | 13,218 | 1.000000 | 31,152 | 1.000000 |
| Hemoglobin α HBA1 + HBA2): | | | | | | |
| N | 35,864 | .999888 | 26,435 | .999962 | 62,299 | .999920 |
| UBE-2 | 3 | .000084 | 0 | .000000 | 3 | .000048 |
| SHIMON | 1 | .000028 | 0 | .000000 | 1 | .000016 |
| Unclassified | 0 | .000000 | 1 | .000038 | 1 | .000016 |
| Total | 35,868 | 1.000000 | 26,436 | 1.000000 | 62,304 | 1.000000 |
| Hemoglobin δ (HBD): | | | | | | |
| N | 17,928 | 1.000000 | 13,218 | 1.000000 | 31,146 | 1.000000 |
| Total | 17,928 | 1.000000 | 13,218 | 1.000000 | 31,146 | 1.000000 |
| Isocitrate dehydrogenase-1 (IDH1) E.C.1.1.1.42: | | | | | | |
| 1 | 17,882 | .999329 | 13,202 | .998941 | 31,084 | .999164 |
| 2HR1 | 10 | .000559 | 4 | .000303 | 14 | .000450 |
| 2NG1 | 0 | .000000 | 2 | .000151 | 2 | .000064 |
| 3NG1 | 1 | .000056 | 6 | .000454 | 7 | .000225 |
| 4HR1 | 1 | .000056 | 0 | .000000 | 1 | .000032 |
| 4NG1 | 0 | .000000 | 2 | .000151 | 2 | .000064 |
| Total | 17,894 | 1.000000 | 13,216 | 1.000000 | 31,110 | 1.000000 |
| Phosphoglycerate kinase-1 (PGK1) E.C.2.7.2.3: | | | | | | |
| 1 | 8,010 | 1.000000 | 5,656 | 1.000000 | 13,666 | 1.000000 |
| Total | 8,010 | 1.000000 | 5,656 | 1.000000 | 13,666 | 1.000000 |

[a] Detailed reference: Fujita et al. 1985*b*.
[b] Detailed reference: Fujita et al. 1985*a*.
[c] Detailed reference: Satoh et al. 1984*b*.
[d] Detailed reference: Satoh et al. 1984*a*.
[e] Detailed reference: Asakowa et al. 1984.
[f] Detailed reference: Kageoka et a. 1985.
[g] Detailed reference: Satoh et al. 1986.

the basis of single ascertainment. The number of allele product determinations in this reduced data set is 539,994 for Hiroshima and 397,433 for Nagasaki. (The few null alleles of PGM1 and GPT whose presence was recognized are excluded from the tabulation.) The locus abbreviations and the Enzyme Commission (E.C.) numbers (for enzymes) are also indicated in table 2.

There is an additional argument for basing this analysis on the single-ascertainment data. In a subsequent paper we will undertake detailed contrasts of our findings for Japanese with our findings for Caucasians and American blacks in other studies of the same proteins (Mohrenweiser et al. 1987; Neel et al. 1988*a*). Both of these latter series are based on ascertainment of a single member of each nuclear family, all born within the time span of a single generation; our use in the present analysis of the single-ascertainment data from siblings studied over a 12-year period will maximize the comparability of the three series.

## V. Analysis

Any manipulation of data of this type must at the outset acknowledge the possibility of heterogeneity within electrophoretic classes. In the present series, such heterogeneity has already been demonstrated by an examination of the thermostability of the $4_{HIR\ 1}$ variant of GPI encountered in the present study; three distinct (and genetically transmitted) subclasses have been defined (Satoh and Mohrenweiser 1979). Thus, the allele frequencies presented in table 2 are, for some loci, overestimates, and the numbers of allele classes are underestimates, a point to which we return later.

### A. Differences between Hiroshima and Nagasaki

We begin with a bipartite analysis of the data of table 2. On the one hand, we contrast the two cities with respect to the frequency of polymorphisms at the nine loci where genetic polymorphisms have been encountered. Variants with a frequency <.01 do not enter into this analysis. On the other hand, we present a separate analysis for the 20 loci at which variants with frequencies <.01 were observed. The occurrence of cells with few or no entries in these latter contrasts necessitates more exact statistical tests than the usual approximate $\chi^2$. The problem is that the tail probabilities for $\chi^2$ are not reliably represented by the usual $\chi^2$ approximation. Our first task is to construct an exact test of the null hypothesis that the allele frequency spectra of the two cities are identical. To do that, we need to specify what we mean by an event more extreme than that we have observed. For the classic 2-×-2 table, this is straightforward enough and leads to Fisher's exact test (Leslie 1955). The same philosophy has been extended to the 2-×-$K$ table by Freeman and Halton (1951).

Following Freeman and Halton (1951) we note that, under the null hypothesis, the exact probability of observing ($r_1$) $A_1$'s in Hiroshima and ($N_1 - r_1$) $A_1$'s in Nagasaki, ($r_2$) $A_2$'s in Hiroshima and ($N_2 - r_2$) $A_2$'s in Nagasaki, and so on, subject to the constraint that $\Sigma_k r_k = N_R$ and that $\Sigma_k (N_k + r_k) = (N - N_R) = N_G$, is

$$\Pr(r_1, r_2, \ldots, r_K | N_1, \ldots, N_K, N_R, N_G) = \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}\binom{N_K}{r_K}}{\binom{N_R + N_G}{N_R}}. \quad (1)$$

To implement equation (1) we need to permute cell counts, subject them to various sample size constraints, evaluate $\chi^2$ for each permutation, and tally the resulting distribution. The most laborious aspect of the calculation is working out the permutational probabilities. For any small (even modest) table, these permutations and their probabilities can be enumerated exhaustively, and we have devised a computer program that will do so.

For contingency tables that are too large to handle by the exact procedure, approximate methods must be developed. For some of the tables we have encountered, the sheer numbers of permutations run into the millions, and the computational cost of the exact test is onerous. We have resorted to simulation, having drawn 100,000 random (2-×-$K$) contingency tables, with probabilities imposed by equation (1). For each of the 100,000 tables, we compute the $\chi^2$ test criterion (which can be converted into the standard Euclidean distance metric we need [Smouse and Williams 1982]), and we have tallied an empiric distribution. Careful comparison of a small number of very large tables shows that the resulting $P$ values are accurate to at least the fourth decimal place.

Returning now to table 2, we note that the question at issue is whether the common polymorphisms are as useful as the rare variants in distinguishing between the populations. The results of the two sets of analyses mentioned in the introduction to this section are presented in table 3. Because the numbers involved in the first contrast are so much greater than those in the second contrast, a much smaller difference between the two cities can emerge as significant in the first contrast than can emerge in the second contrast. To provide a perspective independent of sample size, we also present in table 3 a computation of the between-cities genetic distances, as revealed by the various loci contributing to the two sets of data. We use a Euclidean distance measure (Smouse 1982), defined as

$$\Delta_{RG} = [\,(P_R - P_G)^T V^{-1} (P_R - P_G) / (K - 1)\,]^{1/2}, \quad (2)$$

where $P_R$ is the vector of allele frequencies for Hiroshima, where $P_G$ is the corresponding vector for Nagasaki, both of length ($K - 1$) for $K$ independent alleles, and where $V^{-1}$ is the inverse of the covariance matrix for different alleles. It has been shown (Smouse and Williams 1982) that $\Delta$ can be extracted from the heterogeneity $\chi^2$ we have been employing as

$$\Delta_{RG} = \left[\frac{(N_R + N_G)}{N_R \cdot N_G} \cdot \frac{\chi^2}{(K - 1)}\right]^{1/2} \quad (3)$$

(where the $N$ values are sample sizes).

The allele frequencies in the two cities differ significantly at four of the nine loci supporting polymorphisms. (For these purposes, "significance" is defined as $P < .05$. Since multiple $\chi^2$ determinations were performed, the statistical assessment should involve some allowance for this fact. In the present context, however, we are more concerned with the *relative* values yielded by the two approaches than with their significance per se.) However, for three of these loci, the actual allele frequency difference between the cities is approximately .01 or less. For the fourth locus, PGM1, the large $\chi^2$ value is based primarily on a difference of .03 in the frequencies of the *PGM1*1* and *PGM1*2* alleles. With respect to the rare variants, nine of the 20 city differences are significant. For the six loci at which both polymorphisms and rare variants were encountered, the $\chi^2$ resulting from polymorphisms is larger for three of the contrasts.

A comparison of the relative value of the two data sets in revealing city differences can be achieved by summing $\chi^2$ across independent loci. For the polymorphisms, the sum is 71.16 with 10 df; for the rare variants, it is 211.01 with 103 df; $P$ is $<10^{-8}$ for both data sets. Even orders of magnitude at this level of improbability have little meaning, and we simply conclude that both sets of data are very powerful in differentiating between the cities.

With respect to the loci at which rare variants exhibit significant intercity differences, there are several rather striking findings. The previously noted greater frequency of *GPI*4HR1* in Nagasaki (Neel et al. 1978) is sustained. In addition, the *TF*BHR2* variant is twice as common in Hiroshima as in Nagasaki, whereas the reverse is true for *ALB*NG1* and PGM1*3NG1. These differences are the primary determinants of the relatively high $\chi^2$ values obtained for the corresponding locus contrasts. When each of these alleles is made the object of a simple 2-×-2 $\chi^2$ comparison, all other alleles being pooled, the values are as follows: *GPI*4HR1*, $\chi^2 = 19.76$ and $P < .000005$; *TF*BHR2*, $\chi^2 = 8.31$ and $P < .0016$; *ALB*NG1*, $\chi^2 = 6.86$ and $P < .0046$; and *PGM1*3NG1*, $\chi^2 = 3.98$ and $P < .024$. The fact that these comparisons were singled out after inspection of the data necessitates conservative interpretation of their exact significance level, but that there are biologically meaningful differences seems certain.

The $\Delta$ values reveal a somewhat different picture (table 3). The average distance based on the polymorphisms (.022) is much smaller than that based on the rare variants (.763). For the six loci where distance measurements can be calculated for both polymorphisms and rare variants, the average value for the former is .027 and that for the latter is .527. Thus, the rare variants are much more discriminating in revealing genetic differences between these populations than are the polymorphisms, but because of the smaller numbers involved, the $\Delta$ values for rare variants must be interpreted more cautiously than the $\Delta$ values for the common polymorphisms.

We turn now to a comparison of the cumulative frequency of all rare variants encountered in the two cities—and then to a comparison of the number of different kinds in the two cities. With reference to the first type of comparison, the cumulative frequency of all variant alleles whose individual frequencies are <.01 is nearly identical in the two cities, in Hiroshima 714/539,994, (1.32/1,000) and in Nagasaki, 550/397, 433, (1.38/1,000). This particular contrast of the two cities with respect to "variant structure" appears to be valid even if sample sizes differ. With respect to the second type of comparison, however, since the curve of discovery of *different* rare variants in a population is an increasing but nonlinear function of sample size (Ewens 1972; Rothman and Adams 1978), we must, for any comparison of the number of *different* rare variants in the two cities, adjust the number of alleles tested to comparable sample sizes for the two cities. We therefore adjusted the number of variants encountered at each locus in the two cities to the number we would expect to see in a standard sample of 10,000 alleles, by using the approach of Chakraborty et al. (1988). The adjusted number of different variants in a sample of size 10,000 is

$$E(k) = K - \sum_{i=1}^{K} \exp\{-10{,}000\, p_i\}, \qquad (4)$$

where $P_i$ is the observed frequency of the $i$th allele in the population and where $K$ is the total number of alleles observed in that population. We do not compute adjusted values for PGM3, PGK1, and G6PD because these loci do not have sample sizes of at least 10,000. As shown in table 4, the observed number of different variants for the 29 loci remaining after these exclusions was 127 in Hiroshima and 117 in Nagasaki. Adjusted to 10,000 determinations/locus/city, however, the number of alleles is 99.9 in Hiroshima and 96.8 in Nagasaki. The variances for these two numbers are 13.4 and 11.6 respectively; the difference in adjusted allele numbers between the two cities is statistically negligible.

### *B. Examination of Some Parameters Relevant to Variation in the Number of Rare Alleles at a Locus*

We move now to some comparisons across loci. These

**Table 3**

**Comparison by Locus of the Differences between Hiroshima and Nagasaki, as Revealed by Genetic Polymorphisms and Rare Variants**

| | Polymorphisms | | | | Rare Variants | | | |
|---|---|---|---|---|---|---|---|---|
| Locus | Δ | $\chi^2$ | df | *P* | Δ | $\chi^2$ | df | *P* |
| HP | .025 | 4.87 | 1 | .027 | .476 | 14.22 | 9 | .051 |
| TF | ... | ... | ... | ... | .152 | 23.83 | 14 | .019 |
| ALB | ... | ... | ... | ... | .531 | 12.94 | 4 | .003 |
| CP | ... | ... | ... | ... | .154 | 8.46 | 5 | .062 |
| ADA | .035 | 9.22 | 1 | .002 | 1.178 | 5.83 | 2 | .067 |
| 6PGD | .009 | .62 | 1 | .431 | .711 | 9.55 | 8 | .462 |
| AK1 | ... | ... | ... | ... | ... | ... | ... | ... |
| PGM1 | .054 | 44.00 | 2 | .000+ | .273 | 21.92 | 13 | .015 |
| PGM2 | ... | ... | ... | ... | .571 | 3.61 | 4 | 8.24 |
| PGM3 | .000 | .00+ | 1 | .998 | ... | ... | ... | ... |
| ACP1 | .004 | .11 | 1 | .744 | ... | ... | ... | ... |
| TPI | ... | ... | ... | ... | 1.155 | 4.00 | 3 | 1.000 |
| NP | ... | ... | ... | ... | 1.575 | 25.00 | 2 | .000+ |
| ESB | ... | ... | ... | ... | ... | ... | ... | ... |
| ESD | .027 | 5.31 | 1 | .021 | ... | ... | ... | ... |
| ESA1 | ... | ... | ... | ... | 1.099 | 13.22 | 2 | .001 |
| PEPA | ... | ... | ... | ... | ... | ... | ... | ... |
| PEPB | ... | ... | ... | ... | .335 | .75 | 2 | 1.000 |
| GPI | ... | ... | ... | ... | .302 | 17.15 | 6 | .004 |
| LDHB | ... | ... | ... | ... | ... | ... | ... | ... |
| LDHA | ... | ... | ... | ... | ... | ... | ... | ... |
| MDH1 | ... | ... | ... | ... | 2.000 | 2.00 | 1 | 1.000 |
| CA1 | ... | ... | ... | ... | 1.263 | 7.37 | 2 | .035 |
| CA2 | ... | ... | ... | ... | ... | ... | ... | ... |
| G6PD | ... | ... | ... | ... | .537 | 9.89 | 8 | .328 |
| GOT1 | .033 | 6.22 | 1 | .013 | .208 | 2.98 | 4 | .662 |
| GPT1 | .012 | .81 | 1 | .633 | .315 | 12.25 | 7 | .037 |
| HBB | ... | ... | ... | ... | ... | ... | ... | ... |
| HBA | ... | ... | ... | ... | 1.768 | 5.00 | 2 | .400 |
| HBD | ... | ... | ... | ... | ... | ... | ... | ... |
| IDH1 | ... | ... | ... | ... | .654 | 11.05 | 4 | .009 |
| PGK1 | ... | ... | ... | ... | ... | ... | ... | ... |

Note.—See text for explanation.

must always be approached with caution. One of the criteria for the inclusion of a protein system in the present study was that the normal phenotype, whether identified by a protein- or enzyme-activity stain, yielded a sharp band on starch gel electrophoresis. While this should facilitate the detection of variants, it does not follow automatically that charge-change variants are equally well detected in all these systems. Molecular size and configuration undoubtedly influence the ease of detection of charge changes. Consequently, while within-population comparisons of variants of apparently duplicated loci (e.g., PGM1 and PGM2, HBA and HBB)—and comparisons of corresponding loci across populations—should be reliable, other cross-locus comparisons are inevitably softer.

Having emphasized the differences between Nagasaki and Hiroshima, we will nevertheless at this point initiate a number of comparisons based on data pooled for the two cities, on the premise that the resulting population is no more heterogeneous than the Caucasian populations sampled in Ann Arbor, MI and London, on whom comparable studies have been performed. For such comparisons to be as relevant as possible, the sample sizes should be similar. We have used the same adjustment procedure employed earlier to compare the number of different variants in Hiroshima and Naga-

**Table 4**

**Comparison of Estimated Numbers of Different Alleles in Hiroshima and Nagasaki on the Basis of a Sample of 10,000 Determinations from Each City**

| | Observed Alleles | | | Hiroshima | | Nagasaki | | Cities Combined | |
|---|---|---|---|---|---|---|---|---|---|
| Locus | Hiroshima | Nagasaki | Combined | Expectation | Variance | Expectation | Variance | Expectation | Variance |
| HP | 8 | 9 | 12 | 5.99 | 1.113 | 6.68 | 1.357 | 7.04 | 1.980 |
| TF | 13 | 10 | 16 | 10.66 | 1.253 | 8.11 | 1.013 | 10.49 | 1.974 |
| ALB | 5 | 4 | 6 | 3.79 | .549 | 3.51 | .271 | 4.33 | .749 |
| ADA | 3 | 5 | 5 | 2.98 | .020 | 3.59 | .747 | 3.53 | .519 |
| 6PGD | 5 | 9 | 11 | 3.28 | .734 | 6.58 | 1.432 | 5.57 | 1.914 |
| AK1 | 1 | 1 | 1 | 1.00 | .000 | 1.00 | .000 | 1.00 | .000 |
| PGM1 | 14 | 12 | 17 | 10.67 | 1.838 | 9.67 | 1.359 | 11.15 | 2.407 |
| PGM2 | 5 | 3 | 6 | 3.28 | .769 | 2.43 | .340 | 3.01 | .832 |
| ACP1 | 4 | 2 | 4 | 2.87 | .492 | 2.00 | .000 | 2.50 | .375 |
| TPI | 3 | 3 | 5 | 1.95 | .499 | 2.15 | .489 | 2.24 | .851 |
| NP | 3 | 2 | 4 | 2.44 | .247 | 1.10 | .004 | 3.19 | .252 |
| ESB | 1 | 1 | 1 | 1.00 | .000 | 1.00 | .000 | 1.00 | .000 |
| ESD | 2 | 4 | 4 | 2.00 | .000 | 3.11 | .494 | 2.66 | .443 |
| ESA1 | 3 | 4 | 4 | 2.95 | .052 | 3.36 | .411 | 3.77 | .210 |
| PEPA | 2 | 2 | 2 | 1.99 | .011 | 1.10 | .000 | 1.10 | .001 |
| PEPB | 4 | 3 | 4 | 2.84 | .501 | 2.48 | .295 | 2.70 | .457 |
| GPI | 8 | 6 | 8 | 6.99 | .567 | 5.77 | .179 | 6.62 | .481 |
| LDHB | 1 | 4 | 4 | 1.00 | .000 | 2.59 | .747 | 1.95 | .648 |
| LDHA | 1 | 2 | 2 | 1.00 | .000 | 1.53 | .249 | 1.32 | .216 |
| MDH1 | 2 | 2 | 3 | 1.43 | .245 | 1.53 | .249 | 1.56 | .400 |
| CA1 | 3 | 3 | 4 | 2.66 | .232 | 2.31 | .420 | 2.94 | .508 |
| CA2 | 2 | 1 | 2 | 1.43 | .245 | 1.00 | .000 | 1.24 | .184 |
| GOT1 | 6 | 7 | 7 | 5.30 | .424 | 6.12 | .585 | 5.85 | .634 |
| GPT1 | 10 | 5 | 10 | 8.02 | 1.197 | 4.95 | .048 | 6.91 | 1.210 |
| CP | 6 | 4 | 7 | 4.48 | .795 | 3.53 | .254 | 4.34 | .859 |
| HBB | 4 | 1 | 4 | 2.28 | .734 | 1.00 | .000 | 1.73 | .552 |
| HBA | 3 | 2 | 4 | 1.81 | .430 | 1.32 | .216 | 1.64 | .481 |
| HBD | 1 | 1 | 1 | 1.00 | .000 | 1.00 | .000 | 1.00 | .000 |
| IDH1 | 4 | 5 | 6 | 2.85 | .493 | 4.50 | .400 | 4.21 | .768 |
| Total | 127 | 117 | 164 | 99.94 | 13.438 | 96.81 | 11.560 | 107.48 | 19.904 |

Note.—See text for explanation of procedure.

saki. Three loci, G6PD, PGM3, and PGK1, have again been excluded from consideration because of the relatively small number of determinations and the sex linkage of two of the loci, a fact that may subject alleles at these two loci to probabilities of loss that are different than those to which alleles at autosomal loci are subjected. The total number of allele determinations for each of the remaining loci has been reduced to 23,000 by the adjustment described earlier, a number determined by the observations on GPT1 (23,462) and GOT1 (23,474).

Table 5 and figure 1 present the numbers of alleles encountered at each of the 29 loci, numbers that permitted a sample size reduction to 23,000. The adjusted number of alleles per locus varies from one to 14. The variances for allele number are small enough that even after some intuitive allowance for locus differences in the ease of variant detection, many of these differences in allele numbers must be regarded as statistically significant. The distribution curve for the expected numbers of alleles per locus is right skewed, with no suggestion of bimodality. Only three of the 29 loci qualifying for table 5 did not yield genetic variants. Harris et al. (1974) failed to detect variants at 15 of the 43 enzyme loci they studied in a Caucasian sample; the difference is attributable to our much larger sample sizes.

*1. A comparison of duplicated loci.*—The most accurate comparisons regarding genetic variability involve apparently duplicated loci, of which there are four sets in these data: CA1 and CA2; HBA, HBB, and HBD;

**Table 5**

**Comparison of Across-Locus Variability on the Basis of a Standardized Sample Size of 23,000 Alleles (Cities Combined)**

| Locus | No. of Alleles | | Variance |
|---|---|---|---|
| | Pooled Sample | Adjusted Sample | |
| HP | 12 | 9.31 | 1.626 |
| TF | 16 | 12.90 | 1.704 |
| ALB | 6 | 5.24 | .475 |
| CP | 7 | 5.46 | .846 |
| ADA | 5 | 4.04 | .502 |
| 6PGD | 11 | 7.62 | 1.899 |
| AK1 | 1 | 1.00 | .000 |
| PGM1 | 17 | 14.04 | 1.768 |
| PGM2 | 6 | 4.12 | .998 |
| ACP1 | 4 | 3.07 | .498 |
| TPI | 5 | 3.28 | .981 |
| NP | 4 | 3.54 | .253 |
| ESB | 1 | 1.00 | .000 |
| ESD | 4 | 3.09 | .496 |
| ESA1 | 4 | 3.99 | .013 |
| PEPA | 2 | 2.00 | .000 |
| PEPB | 4 | 3.29 | .426 |
| GPI | 8 | 7.29 | .426 |
| LDHB | 4 | 2.57 | .749 |
| LDHA | 2 | 1.52 | .250 |
| MDH1 | 3 | 2.04 | .499 |
| CA1 | 4 | 3.47 | .301 |
| CA2 | 2 | 1.52 | .250 |
| GOT1 | 7 | 6.52 | .335 |
| GPT1 | 10 | 8.59 | .948 |
| HBB | 4 | 2.57 | .749 |
| HBA | 4 | 2.29 | .648 |
| HBD | 1 | 1.00 | .000 |
| IDH1 | 6 | 5.06 | .607 |
| Total | 164 | 131.43 | 18.247 |

Note.—See text for explanation.

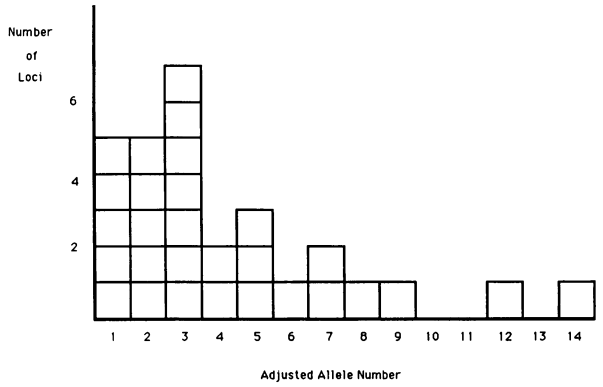


**Figure 1** A histogram of the expected number of alleles per locus, after adjustment of sample size to a standardized 23,000 determinations for those 29 loci with more than 23,000 locus product determinations (see table 5).

LDHA and LDHB; and PGM1, PGM2, and PGM3. The components of the first three sets of duplicated loci are quite similar with respect to number of variants, especially in view of the fact that the relatively small amount of hemoglobin A2 ($\alpha_2\delta_2$) in erythrocytes renders the recognition of δ-chain variants more difficult than the recognition of α- and β-chain variants, these latter two chains being the basis for the major hemoglobin fraction A ($\alpha_2\beta_2$). The situation is more complex as regards the PGM series. PGM3 is excluded from consideration because of the relatively small number of determinations. In phase I (Neel et al. 1978), we encountered no rare variants of PGM2 in 1,892 examination (3,784 locus tests) but a total of 15 variants of PGM1, of seven different phenotypes, in the same number of examinations. When the additional data from phase II are considered, the situation becomes more complicated. We observed 16 different variant alleles, two in polymorphic proportions, among 30,492 allele tests at the PGM1 locus and five variant alleles, none polymorphic, among 30,512 allele tests at the PGM2 locus. The ratio of the difference in expected allele numbers to its variance is 5.968. This difference, considered as a normal deviate, is highly significant ($P < 10^{-5}$). Thus, loci that appear to have a common evolutionary origin and whose enzyme products retain many substrate similarities may support very different numbers of variants.

2. *Variant frequency in relation to molecular size and number of molecular subunits.*—Koehn and Eanes (1978) and Mohrenweiser et al. (1987) have found a positive correlation between the numbers of variants at a locus and polypeptide size in human populations. Mohrenweiser et al. (1987) also found a negative correlation between variant numbers and protein complexity, this latter being defined by the number of polypeptide subunits in the protein. From the numerically adjusted data of table 5, we find a correlation of .61 ± .12 between polypeptide molecular weight and total number of alleles at the locus. In the few instances in which the molecular weight of the human polypeptide was not available, we have followed the example of Mohrenweiser et al. (1987) and used data from the species most closely related to humans. For those polypeptides that function as monomers, the mean adjusted number of alleles was 4.93 ± 1.23; for those that function in mul-

timers of various types (LDH, HB, NP, and HP), the mean number was 4.25 ± 0.64. We thus confirm (although the difference is not significant) the earlier findings mentioned above. It should be noted that these two observations are not independent, since the average molecular weight of the 11 polypeptides of table 5 that function as monomers (54,000) is greater than the average molecular weight of those 17 that function as multimers (41,000). When this fact is taken into consideration, there is in our series no relationship between variant number and protein complexity.

*3. Do polymorphic loci support more rare variants than do nonpolymorphic loci?*—We have recently reported that, following treatment of the human lymphocytoid cell line TK-6 with ethylnitrosourea, loci known to support protein polymorphisms yielded significantly more mutants detected by two-dimensional electrophoresis did than loci not known to support such polymorphisms (Hanash et al. 1988). In another study (Neel et al. 1988*b*) we also observed, as a statistically nonsignificant finding, that among seven newly arisen protein charge–altering or protein activity–altering mutations encountered in the course of studies on the genetic effects of the atomic bombs, five were encountered among the 15 proteins (of 30) exhibiting the greater number of electrophoretic variants. With respect to the present data, we can ask the question, Are rare variants more common at loci supporting genetic polymorphisms? Eight loci are polymorphic; when the sample is adjusted to 23,000, the average adjusted number of rare variants at these loci is 5.3. Twenty-one loci are not polymorphic; the average adjusted number of rare variants is 2.6. The normal deviate with respect to this difference is 1.761; $.01 < P < .05$.

Transferrin occupies a swing position in this tally. The *TF*DCHI* allele attains polymorphic proportions in many mongoloid and Amerindian populations (reviewed in Mourant et al. 1976; Tills et al. 1983) but attains an allele frequency of only .006 in our combined data. This is an exceptionally variable locus. If we yield to the temptation to treat *TF*DCHI* as a polymorphism, the average number of rare variants becomes 5.8/polymorphic locus and 2.0/nonpolymorphic locus. Now the ratio of difference to error is 2.626; $.005 < P < .01$. The present data may thus be viewed as consistent with the above-quoted data on greater mutability at loci supporting polymorphisms, the rare variants being seen as a reflection of locus mutation rates. Conversely, this association suggests that the mutation rate at a locus contributes to the probability that a locus supports polymorphisms. Harris et al. (1974) reported a similar correlation, but in their data the locus for placental alkaline phosphatase (not included in the present series) was entirely responsible for the apparent greater frequency of rare variants at polymorphic loci.

*4. The "dispensability" of a gene product and the number of variants at the locus encoding for that product.*—From a review of the literature, we conclude that genetic deficiency states and their functional impact have been satisfactorily documented for 14 of the polypeptides included in the present study. (We exclude the LDHB deficiency allele identified by Tanis et al. [1977] because in homozygotes the LDHA tetramer supplies enzyme activity. We also exclude hemoglobin $A_2$ ($\alpha_2\delta_2$) from consideration because of its minor representation in the erythrocyte.) It is a reasonable expectation that variants of those polypeptides whose absence is attended by no obvious ill effects should be less subject to selective forces than are variants of polypeptides whose absence results in disease. Accordingly, on the assumption of no average differences in mutability between "dispensable" and "indispensable" loci, more variants might be expected to accumulate at the former. The two proteins (G6PD and PGK) exhibiting sex-linked inheritance will again be excluded from further consideration because variants of these proteins are subject to a different type of selection than are those of autosomally inherited proteins. Of the remaining 12, the absence of four (HP, ALB, CA1, and PGM1) seems well tolerated, whereas absence of any of the remaining eight (TP1, ADA, AK, NP, GPI, CA2, HBA, and HBB) leads to severe disease. The average number of variants associated with the first four proteins, after sample size adjustment, is 7.0, whereas with the latter eight it is 2.2. The difference expressed as a normal deviate is 1.972; $.01 < P < .025$. The expectation is thus sustained.

### C. Variation in Allele Frequencies in the Total Sample

The frequencies with which various alleles are represented in a population is a topic of longstanding interest, bearing as it does on various theories pertaining to the maintenance of genetic variation in populations. A difficulty with data of this type is that, for the less common variants, the definition of "rare" is so sample-size dependent. For example, in a situation in which allelic sample size per locus is only 100 (the minimum size accepted by Chakraborty et al. [1980] in their thorough analysis of this problem), the minimum allele frequency for analytic purposes is .01. By contrast, the present data set can exhibit observed allele frequencies two orders of magnitude lower. Despite the limitations

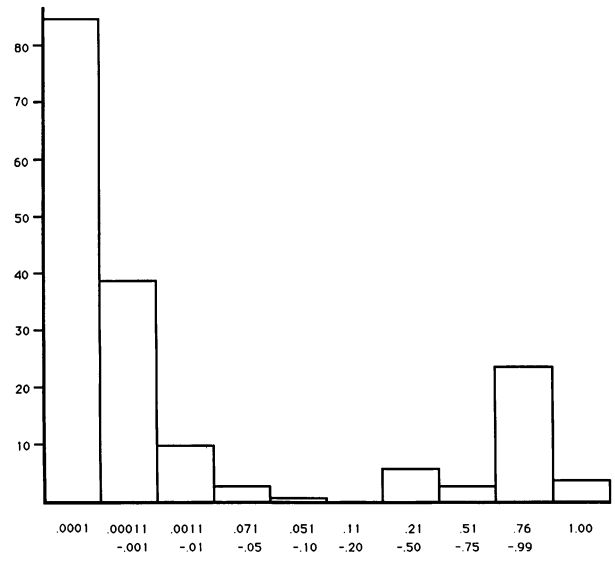


**Figure 2** Histogram of the frequencies presented by the 177 different allele products recognized in the course of the present study.

of sample size, however, the earlier data on the distribution of variant frequencies in diverse species—especially the human data of Harris et al. (1974) and those from our phase I study (Neel et al. 1978)—have, as pointed out by Chakraborty et al. (1980), shown an excess of rare variants, relative to mutation-drift expectations.

Figure 2 presents the data on allele frequencies for the combined Japanese material, as derived from table 1. The distribution is shaped like a reverse J, with a great preponderance of alleles falling into the rare category. Thus, 135 (76.3%) of the 177 alleles identified have a frequency < .01. (Again we score the α-hemoglobin loci as one.) As noted earlier, there is undoubtedly heterogeneity within the electrophoretic classes of some of the proteins studied (see Satoh and Mohrenweiser 1979; Wurzinger and Mohrenweiser 1982). Accordingly, not only is the real number of different alleles greater than 177, but, since subdivision of allele classes can only shift the curve to the left, the proportion of all variants that are rare should be even greater than 76.3%. Using a formula developed by Chakraborty et al. (1988), we expect (on the basis of neutral mutation/drift/equilibrium theory) to have encountered 20.4 alleles with frequencies <.01 in Hiroshima and 19.8 such alleles in Nagasaki. These expectations should be contrasted with the 99.9 and 96.8 totals of table 4. This calculation provides a frame of reference against which to view the numbers of the rare alleles recovered in the present study. The ways in which this "excess" of rare variants can be interpreted are considered below.

## Discussion

### *The Differences between Hiroshima and Nagasaki*

We presume a tribal structure in prehistoric Japan, one not unlike (but not identical with) that encountered in the Amerindians or Australian aborigines at time of first contact (but unlike the high-density agricultural populations of New Guinea). These tribes had arisen through successive bifurcations (and episodic fusions) of preexisting tribes. By analogy with the findings among tribal Amerindians (Neel 1978), we suggest that the electrophoretic variants of proteins in Japanese tribes in prehistoric times can conveniently be divided into the following three classes: (1) polymorphisms of widespread occurrence, present not only in most or all of the tribes, but also in other mongoloid and, usually, in caucasoid and negroid populations as well; (2) polymorphisms of highly restricted occurrence, present in a single tribe—or in several tribes that had only relatively recently diverged (the "private" polymorphisms of Amerindians); and (3) rare variants, with allele frequencies that by convention we set at < .01, sometimes apparently confined to single extended families. We assume that the majority of the (surviving) variants we have encountered are selectively neutral or nearly so. In a numerically stable population, the numbers in which a (surviving) variant occurs are a function of its age, with a very wide variance; and the age of the variant can be estimated if the breeding structure and life tables of the population have been established (Thompson 1976; Thompson and Neel 1978).

Japan is still so poorly mapped with respect to rare variants of the battery of proteins under consideration that it is hazardous to attempt precise calculations. To a first approximation, an allele frequency of .001 in a country of 120,000,000 implies the presence of 120,000 allele copies; a frequency of .0001 implies the presence of 12,000 allele copies, etc. A calculation of the age of such alleles is rendered difficult by the great expansion of the Japanese population since the beginning of the historical period (~A.D. 300)—but especially in recent years—and by lack of data on precise breeding structure in prehistoric times. By analogy with Amerindian populations (Thompson 1976), we suggest that any allele present in a frequency >.001 in either of the two cities probably dates back to the tribal

era, and even one with a frequency of .0001 may be that old. At that time it would probably have been a polymorphism restricted to a single tribe—or to several closely related tribes. Significant city differences in the frequency of these rare alleles may, then, reflect significant differences in the contribution of various tribes to the present-day population. In Amerindians, the empirical expectation for a "private" polymorphism such as we are postulating was once in every 40 genetic systems studied per tribe (Neel 1980). With regard to specific rare variant frequencies among the 32 loci listed in table 2, there were at least four significant differences between the two cities. A rough estimate of the number of tribes contributing quite differentially to the present-day populations of the two cities can be calculated as follows: (systems showing intercity differences in rare variant frequencies) ÷ (probability of a private polymorphism in a system); that is, (4/32) ÷ (1/40) = five tribes. In addition, there would be more or less equal contributions to the populations of the two cities by an unknown number of additional tribes (plus later, nontribal contributions).

Earlier, we failed to demonstrate a significantly greater number of different rare variants in Nagasaki than in Hiroshima despite the high probability of the more complex recent origins (both Japanese and non-Japanese) of the population of Nagasaki. We have also noted, however, (*a*) that the Hiroshima area may have participated to a greater extent than did the Nagasaki area in whatever influx of new genetic material accompanied the Yayoi expansion in central Japan and (*b*) that it has hosted a major military center. The implication would seem to be that, by whatever process, the present-day populaces of these two cities have reached about the same degree of genetic complexity. There is no excess of rare variants in Nagasaki as compared with Hiroshima that would be suggestive of a substantial infusion of non-Japanese genes into Nagasaki. Even so, as we have just seen, in the frequencies of the rare alleles there are still city differences that may well reflect tribal allele distributions in prehistoric times.

### *Some Locus Contrasts*

The data have provided a number of strong hints concerning the factors associated with genetic variation at a locus, such as the size and associations of the gene product, the occurrence of a polymorphism (a possible index of locus mutability), and the "dispensability" of the product associated with the locus. Unfortunately, even with a collection of data of this magnitude, none of these associations has a really impressive statistical base. Ultimately these associations must be fitted into a mutation-selection framework, but the pathway will be complex. For instance, given the evidence developed earlier (sec. V, B3) for the association between a polymorphism and the occurrence of rare variants at the same locus, we suggest that the driving force for the correlation between heterozygosity and the evolutionary rate of proteins noted by Skibinski and Ward (1982) might equally well be mutation, rather than the selective substitutions favored by these authors (see also Chakraborty and Hedrick 1983).

The term *mutation* must, however, be broadly interpreted. For example, in these data, the PGM1 locus contributes heavily to the association between the occurrence of genetic polymorphism at a locus and the occurrence of rare variants, as well as to the association between product dispensability and the accumulation of variants. On the basis of our data on the increased mutability, in somatic cells, of loci supporting polymorphisms, as compared with loci not supporting polymorphisms (Hanash et al. 1988), it is tempting to attribute these findings to a higher rate of point mutation resulting in nucleotide variants at this locus. However, when the products of the four most common, electrophoretically defined alleles at the PGM1 locus (*PGM1*1*, **2*, **3*, and **7*) were subjected to isoelectric focusing, each could be subdivided into a "+" and "–" type; these types were nonrandomly associated with the four electrophoretic classes. Thus, the four alleles became eight. On the basis of the allele frequencies plus the additive nature of the pI differences between allele products, as well as on the basis of the geographical distribution of the alleles, an allele phylogeny was constructed. This postulated three mutations resulting in nucleotide substitutions in a stem allele, plus four subsequent intragenic recombinations between these substitutions (Takahashi et al. 1982). The validity of this phylogeny will be tested as soon as suitable probes are available for this locus.

In the present context, it is important to emphasize that the probability of generating additional alleles from preexisting alleles is surely to some degree a function of the degree of polymorphism at the locus—and the PGM1 locus supports three electrophoretic polymorphisms in Asian populations. The generation of alleles from preexisting (polymorphic) alleles may provide a partial explanation of the skewness of figure 1. On the other hand, TF, with an equal number of alleles, supports only one electrophoretically defined, low-frequency polymorphism in Asian populations. Intragenic recombination seems a much less probable route to the

generation of new genetic variation at this locus. The growing evidence for intragenic recombination, such as has just been described for the PGM1 locus—recombination that is possibly complicated by hot spots or preferred nodes for recombination, which could be intragenic—as well as the evidence both for gene conversion of various types and for locus differences in mutability brings a new complexity to the interpretation of the forces responsible for allele frequencies in populations.

Not only is the generation of allele complexity in human populations assuming new dimensions, but so is the definition of an "equilibrium" or even "quasi equilibrium" population. Earlier we noted the striking excess ([$P < .01$] relative to expectation based on a neutral mutation/drift/equilibrium hypothesis) of rare variants in these data, a finding confirming an earlier observation of Ohta (1976) and Chakraborty et al. (1980) that was based on the data of Harris et al. (1974) and Neel et al. (1978). The three principal formal explanations to be considered are (1) selection against electromorphs (presumably primarily as heterozygotes), (2) a recent increase of mutation rates, or (3) a model inappropriate to the data set. We suggest that whatever the merits of the first two explanations, explanation (3) is so patently cogent for these populations that the data cannot in the present state of genetic theory be used to test any mutation-selection dichotomy. Elsewhere we have commented at length on the different profile of genetic variation encountered in conglomerate national as contrasted with tribal populations, a relatively high frequency of private polymorphisms and monomorphic loci occurring in the latter (Neel 1978; Chakraborty et al. 1988). It must be assumed that until some 2,000–3,000 years ago the ancestors of modern Japanese populations were organized as tribes. We have demonstrated empirically how private polymorphisms become rare variants as tribes amalgamate (Chakraborty et al. 1988). While, then, we do not feel it profitable to attempt to utilize the present data (or comparable data on other cosmopolitan populations) in any formulation that assumed equilibrium, we do suggest that data of this type can be used not only to shed light on the diverse origins of local populations, as discussed in the preceding sections, but also to evaluate the complexity of the origins of the various contemporary ethnic groups, a topic to be pursued elsewhere. All of these considerations are of course completely applicable to DNA polymorphisms and rare variants.

To the extent that the excessive frequency of rare variants, compared with expectation under a neutral mutation/drift/equilibrium hypothesis, is an index of population amalgamations, our observation of the extent of the excess has very fundamental implications for how one should view linkage disequilibrium either (*a*) at complex loci, such as the histocompatibility region in man, or (*b*) between RFLPs over DNA segments of several hundred kilobases. Prior to the tribal fusions of several thousand years ago which ultimately led to complex civilized populations, both the closely linked components of complex loci and the linked RFLPs occurred in populations with allele numbers and frequencies very different from those of the present populations. The histocompatibility haplotype frequencies in Amerindian tribes provide an example of this (see references in Dausset and Colombani 1973). Until sufficient time has elapsed for recombination to have brought the components of complex loci or the RFLPs into linkage equilibrium—that time of course depending on the closeness of the linkage—inferences about the selective forces maintaining certain disequilibria are almost certain to be wide of the mark.

We will delay detailed comparison of these findings and those resulting from studies of other ethnic groups until a later publication. For the data summarized in table 2, rare variant alleles have a frequency of 1.35/1,000. This frequency appears higher than that reported for Caucasians by Harris et al. (1974) (.88/1,000) and ourselves (Mohrenweiser et al. 1987) (.67/1,000) but quite similar to the frequency we have observed in American blacks (Mohrenweiser et al. 1987) (1.45/1,000). These various frequencies are based almost exclusively on the use of starch gel electrophoresis. However, the loci whose products were investigated—and the relative contributions of the various loci—differ from series to series; it is unwise at this stage of the analysis to overinterpret such comparisons.

## Acknowledgments

We thank Dr. Akira Hayashi of the Osaka Medical Center and Research Institute for Maternal and Child Health for the identification of the hemoglobin variants, and J.A. Tworek and T. McElrath for aid in the extensive tabulations and computations. Dr. Ranajit Chakraborty has been of great assistance in meeting many of the statistical issues raised by these data. These studies were supported by the Radiation Effects Research Foundation, a U.S.-Japan binational foundation, and by U.S. Department of Energy grant FGO2–87ER60533.

## References

Aikens, C. M., and T. Higuchi. 1982. Prehistory of Japan. Academic Press, New York.

Asakawa, J., C. Satoh, N. Takahashi, M. Fujita, J. Kaneko,

K. Goriki, R. Hazama, and T. Kageoka. 1984. Electrophoretic variants of blood proteins in Japanese. III. Triosephosphate isomerase. Hum. Genet. **68**:185–188.

Chakraborty, R. 1978. Number of independent genes examined in family surveys and its effects on gene frequency estimation. Am. J. Hum. Genet. **30**:550–552.

Chakraborty, R., P. A. Fuerst, and M. Nei. 1980. Statistical studies on protein polymorphisms in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. Genetics **94**:1039–1063.

Chakraborty, R., and P. W. Hedrick. 1983. Heterozygosity and genetic distance of proteins. Nature **304**:755.

Chakraborty, R., P. E. Smouse, and J. V. Neel. 1988. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am. J. Hum. Genet. **43**:709–726.

Dausset, J., and J. Colombani, eds. 1973. Histocompatibility testing 1972. Munksgaard, Copenhagen.

Egami, N., T. Umehara, S. Kamiyama, and C. Nakane. 1981. Nihonjin towa naninka: characteristics of the Japanese people. Kodansha, Tokyo.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3**:87–112.

Ferrell, R. E., N. Ueda, C. Satoh, R. J. Tanis, J. V. Neel, H. B. Hamilton, T. Inamizu, and K. Baba. 1977. The frequency in Japanese of genetic variants of 22 proteins. I. Albumin, ceruloplasmin, haptoglobin, and transferrin. Ann. Hum. Genet. **40**:407–418.

Freeman, G. H., and J. H. Halton. 1951. Note on an exact treatment of contigency, goodness of fit and other problems of significance. Biometrika **38**:141–149.

Fujita, M., C. Satoh, J. Asakawa, Y. Nagahata, Y. Tanaka, R. Hazama, and K. Goriki. 1985*a*. Electrophoretic variants of blood proteins in Japanese. V. Ceruloplasmin. Jpn. J. Hum. Genet. **30**:43–50.

Fujita, M., C. Satoh, J. Asakawa, Y. Nagahata, Y. Tanaka, R. Hazama, and T. Krasteff. 1985*b*. Electrophoretic variants of blood proteins in Japanese. VI. Transferrin. Jpn. J. Hum. Genet. **30**:191–200.

Hanash, S. M., M. Boehnke, E. H. Y. Chu, J. V. Neel, and R. D. Kuick. 1988. Non-random distribution of structural mutants following ethylnitrosourea treatment of cultured human lymphoblastoid cells. Proc. Natl. Acad. Sci. USA **85**:165–169.

Harris, H., D. A. Hopkinson, and E. B. Robson. 1974. The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. Ann. Hum. Genet. **37**:237–253.

Hill, A. V. S., D. K. Bowden, R. J. Trent, D. R. Higgs, S. J. Oppenheimer, S. L. Thein, K. N. P. Mickleson, D. J. Weatherall, and J. B. Clegg. 1985. Melanesians and Polynesians share a unique α-thalassemia mutation. Am. J. Hum. Genet. **37**:571–580.

Horai, S., T. Gojobori, and E. Matsunaga. 1984. Mitochondrial DNA polymorphism in Japanese. I. Analysis with restriction enzymes of six base pair recognition. Hum. Genet. **68**:324–332.

———. 1987. Evolutionary implications of mitochondrial DNA polymorphisms in human populations. Pp. 177–181 *in* F. Vogel and K. Sperling, eds. Human genetics. Springer, Berlin.

Horai, S., and E. Matsunaga. 1986. Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum. Genet. **72**: 105–117.

Ishida, T., and E. Hinuma. 1986. The origin of Japanese HTLV-I. Nature **322**:504.

Japanese National Commission for UNESCO. 1958. Japan: its land, people, and culture. Ministry of Education, Tokyo.

Kageoka, T., C. Satoh, K. Goriki, M. Fujita, S. Neriishi, K. Yamamura, J. Kaneko, and N. Masunari. 1985. Electrophoretic variants of blood proteins in Japanese. IV. Prevalence and enzymologic characteristics of glucose- 6-phosphate dehydrogenase variants in Hiroshima and Nagasaki. Hum. Genet. **70**:101–108.

Kirk, R. L. 1975. Isozyme variants as markers of population movement in man. Pp. 169–180 *in* C. L. Markert, ed. Isozymes. Vol. **4**. Academic Press, New York.

———. 1982. Microevolution and migration in the Pacific. Pp. 195–214 *in* B. Bonne-Tamir, T. Cohen, and R. M. Goodman, eds. Human genetics. Part A. The unfolding genome: proceedings of the Sixth International Congress of Human Genetics 1981, Jerusalem. Alan R. Liss, New York.

Koehn, R. K., and W. F. Eanes. 1978. Molecular structure and protein variation within and among populations. Evol. Biol. **11**:39–100.

Komatsu, I. 1963. The Japanese people: origins of the people and the language. East-West Center Press, Honolulu.

Leslie, P. H. 1955. A simple method of calculating the exact probability in 2×2 contingency tables with small marginal totals. Biometrics **42**:522–523.

Mohrenweiser, H. W., K. H. Wurzinger, and J. V. Neel. 1987. Frequency and distribution of rare electrophoretic mobility variants in a population of newborns in Ann Arbor, Michigan. Ann. Hum. Genet. **51**:303–316.

Mourant, A. E., A. C. Kopec, and K. Domaniewska-Soczak. 1976. The distribution of the human blood groups and other polymorphisms. 2d. ed. Oxford University Press, London.

Neel, J. V. 1978. Rare variants, private polymorphisms, and locus heterozygosity in Amerindian populations. Am. J. Hum. Genet. **30**:465–490.

———. 1980. Isolates and private polymorphisms. Pp. 173–193 *in* A. Eriksson, ed. Population structure and genetic disorders. Academic Press, London.

Neel, J. V., H. W. Mohrenweiser, and H. Gershowitz. 1988*a*. A pilot study of the use of placental cord blood samples in monitoring for mutational events. Mutat. Res. **204**: 365–377.

Neel, J. V., M. Otake, M. Fujita, C. Satoh, K. Goriki, S. Neri-

ishi, H. B. Hamilton, T. Kageoka, and J. Asakawa. 1980. A search for mutations affecting protein structure in children of proximally and distally exposed atomic bomb survivors: preliminary report. Proc. Natl. Acad. Sci. USA 77:4221–4225.

Neel, J. V., C. Satoh, K. Goriki, J.-I. Asakawa, M. Fujita, N. Takahashi, T. Kageoka, and R. Hazama. 1988*b*. Search for mutations affecting protein charge and/or function in children of atomic bomb survivors: final report. Am. J. Hum. Genet. 42:663–676.

Neel, J. V., C. Satoh, K. Goriki, M. Fujita, N. Takahashi, J. Asakawa, and R. Hazama. 1986. The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides. Proc. Natl. Acad. Sci. USA 83:389–393.

Neel, J. V., and W. J. Schull. 1956. The effect of exposure to the atomic bombs on pregnancy termination in Hiroshima and Nagasaki. NAS-NRC Publ. 461. Washington, DC.

Neel, J. V., N. Ueda, C. Satoh, R. E. Ferrell, R. J. Tanis, and H. B. Hamilton. 1978. The frequency in Japanese of genetic variants of 22 proteins. V. Summary and comparison with data on Caucasians from the British Isles. Ann. Hum. Genet. 41:429–441.

Nishioka, K. 1984. Predominant mode of transmission in Asia. Pp. 423–432 *in* G. N. Vias, J. L. Dienstag, and J. H. Hootnagel, eds. Viral hepatitis and liver disease. Grune & Stratton, Orlando.

Ogata, T., ed. 1981. The Japanese I (in Japanese). Anthropology. Vol. 5. Pp. 264. (No publication information available.)

Ohta, T. 1976. Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Popul. Biol. 10:254–275.

Omoto, K. 1975. Genetic affinities of the Ainu as assessed from data on polymorphic traits. Pp. 269–303 *in* S. Watanabe, A. Kondo, and E. Matsunaga, eds. Anthropological and genetic studies on the Japanese University of Tokyo Press, Tokyo.

Rothman, E. D., and J. Adams. 1978. Estimation of expected number of rare alleles of a locus and calculation of mutation rate. Proc. Natl. Acad. Sci. USA 75:5094–5098.

Satoh, C., R. E. Ferrell, R. J. Tanis, N. Ueda, S. Kishimoto, J. V. Neel, H. B. Hamilton, and K. Baba. 1977. The frequency in Japanese of genetic variants of 22 proteins. III. Phosphoglucomutase-1, phosphoglucomutase-2, 6-phosphogluconate dehydrogenase, adenylate kinase, and adenosine deaminase. Ann. Hum. Genet. 41:169–183.

Satoh, C., and H. W. Mohrenweiser. 1979. Genetic heterogeneity within an electrophoretic phenotype of phosphoglucose isomerase in a Japanese population. Ann. Hum. Genet. 42:283–292.

Satoh, C., N. Takahashi, J. Asakawa, N. Masunari, M. Fujita, K. Goriki, R. Hazama, and K. Iwamoto. 1984*a*. Electrophoretic variants of blood proteins in Japanese. I. Phosphoglucomutase-2 (PGM-2). Jpn. J. Hum. Genet. 29:89–104.

Satoh, C., N. Takahashi, J. Kaneko, Y. Kimura, M. Fujita, J. Asakawa, T. Kageoka, K. Goriki, and R. Hazama. 1984*b*. Electrophoretic variants of blood proteins in Japanese. II. Phosphoglucomutase-1 (PGM-1). Jpn. J. Hum. Genet. 29:287–310.

Satoh, C., N. Takahashi, Y. Kimura, A. Miura, J. Kaneko, M. Fujita, K. Toyama. 1986. Electrophoretic variants of blood proteins in Japanese. VII. Cytoplasmic glutamate-oxaloacetate transaminase (GOT1). Jpn. J. Hum. Genet. 31:1–14.

Skibinski, D. O. F., and R. D. Ward. 1982. Correlations between heterozygosity and evolutionary rate of proteins. Nature 298:490–492.

Smith, R. J., and R. K. Beardsley, eds. 1962. Japanese culture: its development and characteristics. Viking Fund Publications in Anthropology no. 34. Aldine, Chicago.

Smouse, P. E. 1982. Genetic architecture of swidden agricultural tribes from the South American rainforests. Pp. 139–178 *in* M. Crawford and J. M. Mielke, eds. Current developments in anthropolotical genetics; vol. 2. Ecology and population structure. Plenum, New York.

Smouse, P. E., and R. Williams. 1982. Multivariate analysis of HLA disease associations. Biometrics 38:757–768.

Suzuki, H. 1981. Racial history of the Japanese. Pp. 7–69 *in* Rassengeschichte der Menschheit. 8 Lieferung, Asien I: Japan, Indonesien, Ozeanien, R. Oldenberg, Munich.

Tanis, R. J., J. V. Neel, and R. J. de Arauz. 1977. Two more "private" polymorphisms of Amerindian tribes: $LDH_B$ GUA-1 and $ACP_1$ B GUA-1 in the Guaymi in Panama. Am. J. Hum. Genet. 29:419–430.

Tanis, R. J., N. Ueda, C. Satoh, R. E. Ferrell, S. Kishimoto, J. V. Neel, H. B. Hamilton, and N. Ohno. 1978. The frequency in Japanese of genetic variants of 22 proteins. IV. Acid phosphatase, NADP-isocitrate dehydrogenase, peptidase A, peptidase B and phosphohexose isomerase. Ann. Hum. Genet. 41:419–428.

Takahashi, N., J. V. Neel, C. Satoh, J. Nishizaki, and N. Masumari. 1982. A phylogeny for the principal alleles of the human phosphoglucomutase-1 locus. Proc. Natl. Acad. Sci. USA 79:6636–6640.

Thompson, E. A. 1976. Estimation of age and rate of increase of rare variants. Am. J. Hum. Genet. 28:442–452.

Thompson, E. A., and J. V. Neel. 1978. The probability of founder effect in a tribal population. Proc. Natl. Acad. Sci. USA 75:1442–1445.

Tills, D., A. C. Kopec, and R. E. Tills. 1983. The distribution of the human blood groups and other polymorphisms. Suppl. 1. Oxford University Press, Oxford.

Ueda, N., C. Satoh, R. J. Tanis, R. E. Ferrell, S. Kishimoto, J. V. Neel, H. B. Hamilton, and K. Baba. 1977. The frequency in Japanese of genetic variants of 22 proteins. II. Carbonic anhydrase I and II, lactate dehydrogenase, malate dehydrogenase, nucleoside phosphorylase, triose phosphate isomerase, hemoglobin A and hemoglobin $A_2$. Ann. Hum. Genet. 41:43–52.

Wainscoat, J. S., A. V. S. Hill, A. L. Boyce, J. Flint, M. Hernandex, S. L. Thein, J. M. Old, J. R. Lynch, A. G. Folusi, D. J. Weatherall, and J. B. Clegg. 1986. Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature **319**:491–493.

Wurzinger, K. H., and H. W. Mohrenweiser. 1982. Studies on genetic and nongenetic (physiological) variation of human erythrocyte glutamic oxaloacetic transaminase. Ann. Hum. Genet. **46**:191–201.