Am. J. Hum. Genet. 53:1107–1113, 1993

Independence Tests for VNTR Alleles Defined as Quantile Bins

B. S. Weir

Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh

Summary

VNTR fragment lengths in three databases maintained by the FBI for forensic purposes were partitioned into quantile bins, and tests for independence of the two bins at each of six loci were conducted. Whether independence was declared depended on the number of quantiles used. For a large number of quantile bins, equal to the number of fixed bins used by the FBI, 10 of 18 likelihood-ratio tests showed significant departures from independence when all genotypes were considered, and this changed to 7 of 18 when only heterozygotes were tested. This is in contrast to likelihood-ratio tests on fixed bins, when there were five significant departures over all genotypes and two departures for heterozygotes.

Introduction

The VNTR loci being used by forensic scientists have such high numbers of variants that it is virtually impossible to distinguish them all by gel electrophoresis. Instead, they are often assigned to a relatively small number of bins (e.g., see Budowle et al. 1991). In the system used by several forensic laboratories in the United States, the boundaries for these bins are defined by the known lengths of fragments resulting from viral digests. The advantage of these "fixed bins" is that the same boundaries apply to all sets of data, which simplifies the comparison of allelic frequencies in different data sets. A disadvantage of the fixed-bin approach is that different bins have different frequencies, and some bins may even be empty. In practice, empty bins are amalgamated with neighboring bins until some minimum frequency is obtained, but there can still be low numbers in bins at the ends of the length distributions. This complicates the task of testing for independence of bin frequencies and was the reason why Weir (1992*a*) chose to use likelihood-ratio tests for consistency of genotypic frequencies to Hardy-Weinberg expectations.

An alternative binning strategy was proposed by Geisser and Johnson (1992) and was applied by them to forensic databases (Geisser and Johnson 1993). Their "quantile bins" follow from dividing the ordered set of VNTR fragment frequencies into quantiles, i.e., sets of equally frequent bins. A set of n individuals will provide $2n$ fragments, and these can be placed into q sets of $2n/q$ fragments. The advantages of this binning strategy are that, provided that q is not too large, no bins have very low frequencies, and the usual χ^2 goodness-of-fit test for Hardy-Weinberg equilibrium can be used. Not only is this easy to calculate, but also χ^2 tables can be used to assess the significance of any test. In the forensic setting, another advantage is that all heterozygotes have the same frequency ($2n/q^2$) in a database, as do all homozygotes (n/q^2). The disadvantage is that bin boundaries depend on the size of the database, which makes comparisons between databases difficult. There can also be ambiguity over whether an individual is homozygous or heterozygous at the bin level.

Although the use of quantile bins simplifies the testing for Hardy-Weinberg equilibrium in forensic databases, it is shown here that similar results for databases maintained by the Federal Bureau of Investigation (FBI) are generally obtained from those using fixed bins, although there are differences for loci D10S28, D14S13, and D17S79. This paper differs from that of Geisser and Johnson (1993) in considering large numbers of quantile bins, as would be used in forensic applications.

The Data

Data for six VNTR loci—D1S7, D2S44, D4S139, D10S28, D14S139, and D17S79—were provided by the

Received June 21, 1993.

Address for correspondence and reprints: Dr. B. S. Weir, Program in Statistical Genetics, Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695.

0002-9297/93/5305-0015$02.00

Table 1

Significance Levels for Tests on FBI Databases

	Black		Caucasian		Hispanic	
Locus	Quantile	Fixed	Quantile	Fixed	Quantile	Fixed
D1S7:						
$n(Q)$	359	(26)	595	(26)	521	(24)
X^2_{Ho}	.00	.00	.17	.29	.93	.10
X^2_{GJ}	.04		.59		.07	
X^2_{LR}	.13	.39	.65	.37	.02	.00
X^2_{He}	.60	.86	.76	.55	.03	.00
D2S44:						
$n(Q)$	475	(24)	792	(21)	515	(19)
X^2_{Ho}	.00	.00	.00	.19	.01	.10
X^2_{GJ}	.00		.51		.14	
X^2_{LR}	.00	.02	.90	.66	.20	.72
X^2_{Hc}	.06	.47	.97	.80	.37	.70
D4S139:						
$n(Q)$	448	(18)	594	(17)	522	(16)
X^2_{Ho}	.28	.72	.04	.58	.04	.59
X^2_{GJ}	.42		.36		.12	
X^2_{LR}	.30	.51	.38	.62	.23	.42
X^2_{He}	.29	.42	.27	.71	.35	.39
D10S28:						
$n(Q)$	288	(24)	429	(23)	440	(21)
X^2_{Ho}	.01	.01	.00	.00	.00	.03
X^2_{GJ}	.05		.00		.00	
X^2_{LR}	.04	.64	.00	.24	.00	.16
X^2_{He}	.16	.81	.00	.83	.00	.31
D14S13:						
$n(Q)$	524	(25)	751	(24)	494	(23)
X^2_{Ho}	.00	.03	.00	.00	.00	.00
X^2_{GJ}	.33		.00		.00	
X^2_{LR}	.54	.65	.00	.09	.00	.31
X^2_{He}	.65	.82	.01	.28	.10	.48
D17S79:						
$n(Q)$	550	(15)	776	(13)	521	(9)
X^2_{Ho}	.00	.00	.00	.00	.00	.01
X^2_{GJ}	.00		.00		.00	
X^2_{LR}	.00	.00	.00	.00	.00	.00
X^2_{He}	.00	.12	.00	.67	.00	.01

FBI. These data are in three databases, identified as black, Caucasian, and Hispanic. Database sizes n, are shown in table 1, and these differ slightly from those described by Weir (1992*a*), although they correspond exactly to the data described by Budowle et al. (1991) and to the data currently used by the FBI in forensic calculations. The number, Q, of fixed bins used by the FBI is also shown in the table.

Quantiles were constructed by ordering all fragment lengths at a locus and assigning them to q equally frequent bins, as far as possible. It is often the case that one or more fragments had lengths equal to these bin boundaries. For simplicity, all such fragments were assigned to the same bin (e.g., see table 2), but this is not expected to have a major effect on the results. Values of q were allowed to vary between two and the FBI's number of fixed bins.

The Tests

Four separate tests were conducted at each locus in each database. First, a test for consistency of overall homozygosity with the value expected under Hardy-Weinberg equilibrium was performed. When allelic fre-

Table 2

Bin Boundaries and Counts for Caucasian Locus D10S28

	Fixed Bins		Quantile Bins	
Bin	Lower Boundary	Count	Lower Boundary	Count
1	1	13	1	37
2	964	44	1014	37
3	1078	38	1122	37
4	1197	15	1368	38
5	1353	34	1530	37
6	1508	67	1574	37
7	1638	75	1668	38
8	1789	71	1751	37
9	1925	40	1804	38
10	2089	51	1868	37
11	2352	16	1973	37
12	2523	14	2132	37
13	2693	36	2326	37
14	2863	42	2718	38
15	3034	41	2891	37
16	3330	56	3064	37
17	3675	39	3318	38
18	3980	62	3498	37
19	4324	58	3863	38
20	4822	12	4076	37
21	5220	6	4312	37
22	5686	23	4551	38
23	6369	5	4942	37

quencies are equal, as they are for quantile bins, the simple comparison of observed and expected values with a χ^2 test, or its normal equivalent (Geisser and Johnson 1992), is the same as the statistic used by Weir (1992*a*), following Levene (1949). To allow for small inequalities in sample quantile-bin frequencies (e.g., see table 2), the second form was used here:

$$X^2_{Ho} = \frac{n(H-h)^2}{\sum_i p_i^2 - 2\sum_i p_i^3 + (\sum_i p_i^2)^2}, \quad (1)$$

where n is the number of individuals in the database, p_i is the frequency of the ith bin (generally $1/q$), the sums range over $i = 1, \ldots q$, $h = \Sigma_i p_i^2$, and $H = \Sigma_i (n_{ii}/n)$ is the sum of homozygotes at the bin level. This statistic is distributed as χ^2 with 1 df when Hardy-Weinberg proportions hold.

The second test was the χ^2 goodness-of-fit test described by Geisser and Johnson (1992). If n_{ii} individuals are homozygotes for quantile i and n_{ij} individuals are heterozygotes for quantiles i and j, and quantile i has actual database frequency p_i, then the test statistic is

$$X^2_{GJ} = \sum_i \frac{(n_{ii} - np_i^2)^2}{np_i^2} + \sum_{i>j} \frac{(n_{ij} - 2np_ip_j)^2}{2np_ip_j}. \quad (2)$$

Under the Hardy-Weinberg hypothesis, this has an asymptotic χ^2 distribution with $q(q-1)/2$ df. Simulations showed that the χ^2 distribution holds well for samples of the size considered here. For $q = 2$, $X^2_{Ho} = X^2_{GJ}$, but since Geisser and Johnson (1993) performed the homozygosity test against a one-sided alternative, their p-values were twice as big for X^2_{GJ} as for X^2_{Ho}.

The third test was a likelihood-ratio test and was performed as by Weir (1992*a*), but for quantile bins. If n_{ii} and n_{ij} are defined as above, and n_i is the observed number of fragments in the ith quantile, the unconstrained likelihood is

$$L_1 = C \prod_{i,j} \left(\frac{n_{ij}}{n}\right)^{n_{ij}},$$

where C is a ratio of combinatorial coefficients. The likelihood constrained by the Hardy-Weinberg hypothesis is

$$L_0 = C2^{He} \prod_i \left(\frac{n_i}{2n}\right)^{n_i}$$

where He is the number of heterozygotes in the database. The test statistic is

$$X^2_{LR} = -2(\ln L_1 - \ln L_0) \quad (3)$$

to correct the expression given by Weir (1992*a*).

Simulations show that X^2_{LR} is *not* well approximated by a χ^2 distribution with $q(q-1)/2$ df in the present case. Empirical significance values were therefore found from sets of 1,000 samples simulated under the Hardy-Weinberg hypothesis. Each simulated sample was the same size as the database being considered but was drawn from a population with Hardy-Weinberg genotypic frequencies constructed from the database allelic frequencies.

As has been mentioned repeatedly, forensic VNTR databases almost always have an excess of homozygotes (see references in Weir 1992*b*). Several reasons for this phenomenon have been discussed, and there is agreement that the two main factors are coalescence of neighboring bands on a gel and loss of short bands from a gel. As a consequence, forensic scientists routinely estimate the frequency of bin homozygotes as twice the

frequency of that bin instead of the square of the bin frequency. The Hardy-Weinberg genotypic frequency is invoked only for heterozygotes, suggesting that it is appropriate to confine Hardy-Weinberg testing to bin heterozygotes. A test statistic was given by Weir (1992*a*):

$$X^2_{He} = -2(\ln L_1^* - \ln L_0^*), \quad (4)$$

where

$$L_1^* = C^* \prod_{i>j} \left(\frac{n_{ij}}{He}\right)^{n_{ij}}$$

$$L_0^* = C^* \prod_{i>j} \left(\frac{2p_i^* p_j^*}{1-h^*}\right)^{n_{ij}}.$$

The bin frequencies p_i^*, and their sum of squares h^*, are calculated from only the heterozygotes. These estimates are not strictly maximum likelihood and, as with the test statistic X^2_{LR}, significance levels for test statistic X^2_{He} are found empirically from the sets of 1,000 data sets simulated under the Hardy-Weinberg hypothesis. Finally, tests 1, 3, and 4 were repeated on the FBI fixed bins (Budowle et al. 1991).

Results

For locus D10S28 in the Caucasian database, the significance levels for several q values are shown in table 3. It is immediately evident that significance levels vary unpredictably with the number of bins. A similar phenomenon has often been noticed in population genetic data and was discussed in the context of linkage disequilibrium by Weir and Cockerham (1978). If the alleles are independent at the finest possible partitioning, then it is expected that independence in a population will persist as the partitions become more coarse (as q decreases). The converse is not true, since departures from Hardy-Weinberg equilibrium at a fine level may be opposite in sign and of such magnitude that they cancel out at coarser levels.

For samples, any pattern is possible. The simplest pair of examples relating $q = 2$ and $q = 3$ are those illustrated in tables 4 and 5. In table 4, the sample size of 9 shows a perfect fit to Hardy-Weinberg equilibrium for $q = 3$ but not for $q = 2$, whereas in table 5 the sample size of 12 shows a perfect fit to Hardy-Weinberg equilibrium for $q = 2$ but not for $q = 3$. It is doubtful that general statements about significance level for one q value can be based on tests conducted at another level.

Table 3

Significance Levels for Quantile Bin Tests Applied to D10S28 in the FBI Caucasian Database

q and Test	p	q and Test	p
2:		8:	
X^2_{Ho}	.58	X^2_{Ho}	.25
X^2_{GJ}	.58	X^2_{GJ}	.35
X^2_{LR}	.58	X^2_{LR}	.35
X^2_{He}	—	X^2_{He}	.63
3:		9:	
X^2_{Ho}	.38	X^2_{Ho}	.40
X^2_{GJ}	.06	X^2_{GJ}	.05
X^2_{LR}	.07	X^2_{LR}	.02
X^2_{He}	—	X^2_{He}	.41
4:		10:	
X^2_{Ho}	.19	X^2_{Ho}	.04
X^2_{GJ}	.03	X^2_{GJ}	.01
X^2_{LR}	.03	X^2_{LR}	.02
X^2_{He}	.02	X^2_{He}	.14
5:		15:	
X^2_{Ho}	.68	X^2_{Ho}	.00
X^2_{GJ}	.54	X^2_{GJ}	.00
X^2_{LR}	.55	X^2_{LR}	.00
X^2_{He}	.69	X^2_{He}	.03
6:		20:	
X^2_{Ho}	.20	X^2_{Ho}	.00
X^2_{GJ}	.28	X^2_{GJ}	.00
X^2_{LR}	.34	X^2_{LR}	.00
X^2_{He}	.62	X^2_{He}	.00
7:		23:	
X^2_{Ho}	.04	X^2_{Ho}	.00
X^2_{GJ}	.24	X^2_{GJ}	.00
X^2_{LR}	.26	X^2_{LR}	.00
X^2_{He}	.60	X^2_{He}	.00

Results for all loci and databases are summarized in table 1, for only the FBI's number of bins. The significance level of X^2_{LR} was generally, but not always, greater than that of X^2_{GJ}. The likelihood-ratio test for heterozygotes only is more likely to give the same conclusion as the likelihood-ratio test over all genotypes when the homozygosity test has a high significance level. When there are Hardy-Weinberg proportions in a population, there is no additional information among the homozygotes over that in the heterozygotes (Robertson and Hill 1984). There are several situations, however, when the homozygosity is substantially different from the Hardy-Weinberg value, but the collection of heterozygotes does not show much departure from Hardy-Weinberg values. This point was stressed by Weir

Table 4

Two Quantile Binnings for a Sample of Size $n = 9$

	Genotype	
Original Scale[a]	$q = 2$	$q = 3$
A_1A_2	A_1A_1	A_1A_1
A_3A_7	A_1A_1	A_1A_2
A_4A_8	A_1A_1	A_1A_2
A_5A_{13}	A_1A_2	A_1A_3
A_6A_{14}	A_1A_2	A_1A_3
A_9A_{10}	A_1A_2	A_2A_2
$A_{11}A_{15}$	A_2A_2	A_2A_3
$A_{12}A_{16}$	A_2A_2	A_2A_3
$A_{17}A_{18}$	A_2A_2	A_3A_3

[a] Alleles are ordered and numbered according to fragment length.

(1992*a*) and has considerable relevance for the forensic situation.

There is no general agreement between significance levels for tests performed on fixed and quantile bins, even if the numbers of bins are equal. The lack of agreement must be a consequence of the different binning strategies.

Discussion

Binning of VNTR alleles into quantiles is no more arbitrary than binning according to boundaries defined by viral digest fragments, and quantile binning simplifies Hardy-Weinberg testing if all genotypes are being considered. Such tests have been performed for three databases maintained by the FBI.

The forensic utility of quantile binning is a matter for debate. The bin boundaries change with sample size, so that comparisons across databases are made more difficult. The forensic practice of not invoking Hardy-Weinberg theory for homozygotes means that there is still interest in confining Hardy-Weinberg testing to heterozygotes, and there does not seem to be a simple modification of the χ^2 goodness-of-fit test for this situation. Another issue is that quantile bins can be quite narrow, as illustrated for D10S28 in the Caucasian database by figure 1 and table 2. Quantile bin 5, with boundaries 1530 and 1574, is more narrow than the "matching window" used by the FBI. Empirical studies have suggested that the true length of a VNTR fragment is included in an interval centered on the estimated length and of width 5% of that estimated length. The difference 1574 − 1530 = 44 is only 3% of the midpoint. This will hinder placing evidence fragment lengths into bins. The fixed bins, on the other hand, tend to have widths closer to 10% of their midpoints. The exact quantile-bin assignment of one fragment is of little concern, since all bins have the same frequency, but the ambiguity for very narrow quantile bins means that there could be uncertainty, when two fragments are being binned, as to whether the individual is homozygous or heterozygous at the bin level.

This discussion has concentrated on the high numbers of quantile bins desirable in forensic applications. The great power of VNTRs for human identification stems from their variability, and there would be little interest in limiting this variability by considering small numbers of bins. Why do fixed- and quantile-bin test values differ for the large q values? Of the 18 locus-database combinations, there are five instances where the quantile-bin test would declare significant departures for Hardy-Weinberg at the 5% significance level for quantile bins but not for fixed bins, whether the test was over all genotypes or only over heterozygotes. The explanation seems to be that of sampling effects of the type illustrated in tables 4 and 5. That there is underlying independence of the continuous fragment lengths for the two fragments within an individual is expected from a rich body of population genetic experience and is supported by the finding (Weir 1992*a*) of intraclass correlation coefficients not significantly different from zero. It is, of course, the fixed-bin approach that is used

Table 5

Two Quantile Binnings for a Sample of Size $n = 12$

	Genotype	
Original Scale[a]	$q = 2$	$q = 3$
A_1A_2	A_1A_1	A_1A_1
A_3A_4	A_1A_1	A_1A_1
A_5A_6	A_1A_1	A_1A_1
A_7A_{13}	A_1A_2	A_1A_2
A_8A_{14}	A_1A_2	A_1A_2
A_9A_{15}	A_1A_2	A_2A_2
$A_{10}A_{16}$	A_1A_2	A_2A_2
$A_{11}A_{17}$	A_1A_2	A_2A_3
$A_{12}A_{18}$	A_1A_2	A_2A_3
$A_{19}A_{20}$	A_2A_2	A_3A_3
$A_{21}A_{22}$	A_2A_2	A_3A_3
$A_{23}A_{24}$	A_2A_2	A_3A_3

[a] Alleles are ordered and numbered according to fragment length.

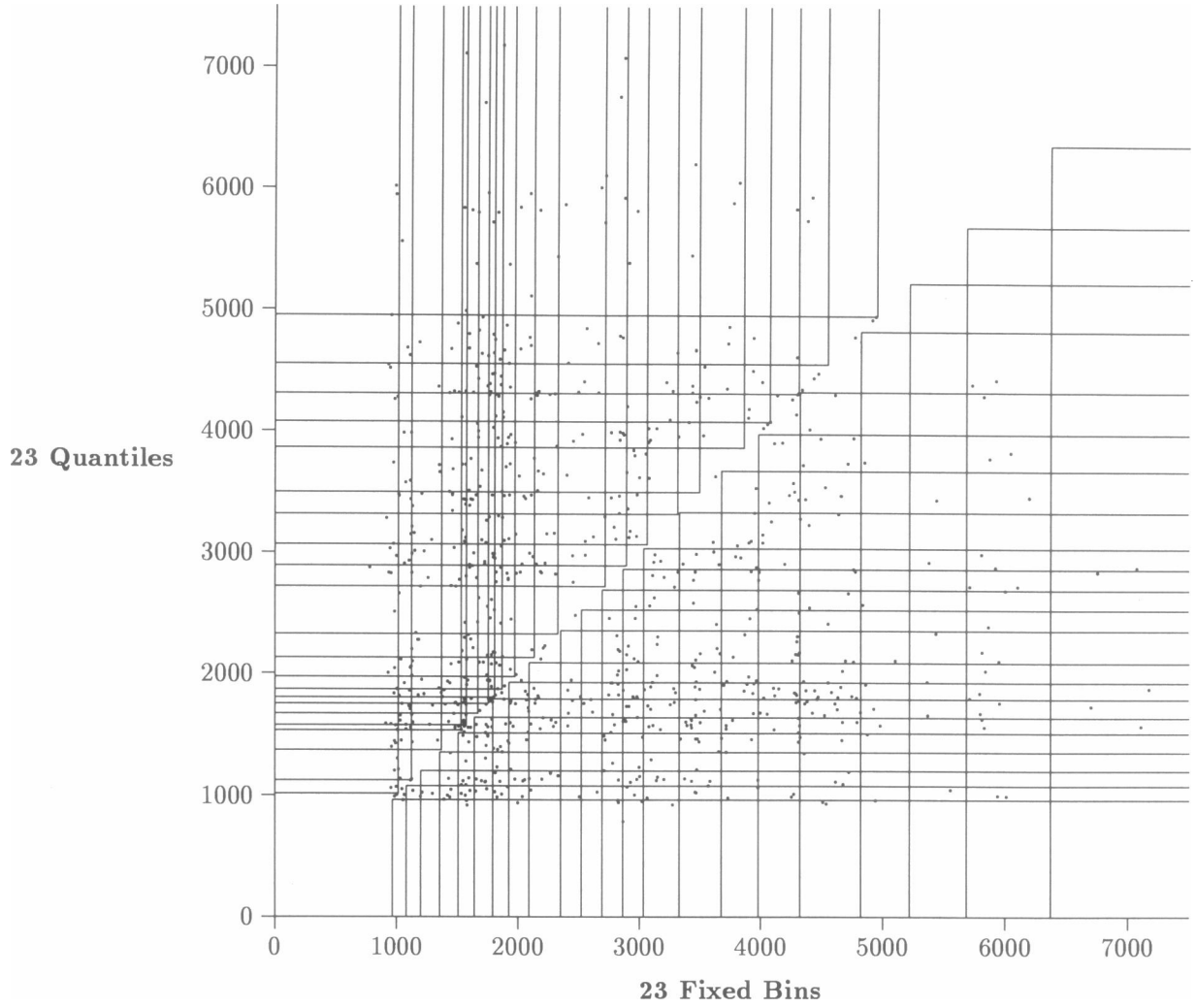

Figure 1 The Caucasian D10S28 database, plotted twice, with axes marked in base pairs. Each individual in the database is represented by a dot. Above the diagonal the fragments have been placed in 23 quantile bins, and below the diagonal fragments have been placed in the 23 fixed bins used by the FBI.

by the FBI in forensic calculations and for which testing of independence of VNTR fragments is needed. For this approach—with two exceptions, D1S7 and D17S79, in the Hispanic database—the heterozygote frequency arrays appear to be consistent with those predicted by the Hardy-Weinberg law.

Acknowledgments

This work was supported in part by National Institutes of Health grant GM43544. The continuing cooperation of Dr. Bruce Budowle of the FBI Forensic Science Research and Training Center is appreciated. Communication with Dr. Seymour Geisser was very helpful.

References

Budowle B, Monson KL, Anoe KS, Baechtel FS, Bergman DL, Buel E, Campbell PA, et al (1991) A preliminary report on binned general population data on six VNTR loci in Caucasians, Blacks and Hispanics from the United States. Crime Lab Digest 18:9–26

Geisser S, Johnson W (1992) Testing Hardy-Weinberg equilibrium on allelic data from VNTR loci. Am J Hum Genet 51:1084–1088

——— (1993) Testing independence of fragment lengths within VNTR loci. Am J Hum Genet 53:1103–1106

Levene H (1949) On a matching problem arising in genetics. Ann Math Stat 20:91–94

Robertson A, Hill WG (1984) Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. Genetics 107:703–718

Weir BS (1992*a*) Independence of VNTR alleles defined as fixed bins. Genetics 130:873–887

——— (1992*b*) Population genetics in the forensic DNA debate. Proc Natl Acad Sci USA 89:11654–11659

Weir BS, Cockerham CC (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. Genetics 88:633–642