J Mol Evol (1983) 19:449–454

Journal of
Molecular Evolution
© Springer-Verlag 1983

The Analysis of Hidden Electrophoretic Variation: Interspecific Electrophoretic Differentiation and Amino Acid Divergence

Paul A. Fuerst[1] and Robert E. Ferrell[2]

1 Department of Genetics, The Ohio State University, 1735 Neil Avenue, Columbus, OH 43210, USA

2 Center for Demographic and Population Genetics, University of Texas Health Science Center at Houston, TX 77025, USA

Summary. In a study of 25 human variants and 23 "evolutionary alleles" of hemoglobin we show that intra-specific and interspecific patterns of electrophoretic variability are not comparable. Significant deviation from the predicted electrophoretic differentiation between evolutionary alleles is normally found only when amino acid sequence divergence exceeds 10%. When two sequences had diverged at less than 30 out of 287 amino acid residues sites, only 7% of comparisons showed significant deviations from the expected difference of electrophoretic mobility, while significant deviation was shown by 57% of comparisons involving 30–40 residue differences, by 79% in the case of 51–60 differences and by all of the comparisons involving more than 60 differences. In contrast, human variants, which differ by only one or two amino acid residues (less than 1% difference), had significant deviations in 58% of comparisons. Those mutations that appear as fixed differences in the evolutionary material probably represent only a subset of the mutations which can appear within the species. The results suggest that statistical comparisons such as genetic distance may not measure the same process within a species as between species. This is due not to inherent problems with the statistic, but rather to inherent differences in the nature of molecular changes that are detectable by electrophoresis at different stages of population divergence.

Key words: Electrophoretic detectability – Hidden genetic variation – Polymorphism – Genetic distance – Hemoglobin

Offprint requests to: P.A. Fuerst

Introduction

Electrophoretic data have been extensively used for intra- and interspecific taxonomic comparisons (Ayala 1975), and to calculate statistics such as divergence time between taxa based on genetic distance (Nei 1972). Such studies presuppose that electrophoretic variation observed within populations and that between taxa such as species is based on amino acid substitutions that are similar in rate and kind. Standard survey methods that use support media such as starch, polyacrylamide or cellulose acetate, detect only a portion of the underlying biochemical variability present in a population. Two recent studies (Ramshaw et al. 1979; Fuerst and Ferrell 1980) examined the sensitivity of electrophoresis for detecting known alleles. Variants of hemoglobin which carried known molecular alterations were utilized in both investigations. Each study examined a reasonably large sample of human hemoglobin variants to predict variability patterns expected from population surveys. Fuerst and Ferrell (1980) also included a number of non-human hemoglobins. These "evolutionary variants" seemed to have equivalent variation to that found for the human variants. In this paper, however, we present a reexamination of the non-human variants and show that, contrary to earlier conclusions, the patterns of "hidden" variability are actually quite different at different taxonomic levels, and that these differences may have important implications for the interpretation of electrophoretic data collected over different evolutionary levels.

Methods and Materials

The data were originally collected as part of the work presented in Fuerst and Ferrell (1980). The hemoglobin variants considered in the present work are listed in Table 1; our earlier

Table 1. List of hemoglobin variants used in this study. Additional information on variants was given in Fuerst and Ferrell (1980). Human Hb-A is listed in both categories. Note that the net charge given for the rabbit sequence (+2) differs from that listed in Fuerst and Ferrell (1980), which showed a charge of (+4). Examination of the nucleotide sequence reported by Heindell et al. (1978), suggests a potential error in the amino acid sequence reported in Dayhoff (1972), which alters the calculated charge

Charge Class	Average Mobility	Human Variants	Evolutionary Sequences
+6	0.098	–	Chicken
+4	0.473	C	Chimpanzee A2
		E	Human A2
		O-Arab	Gorilla A2
			Spider monkey A2
+2	0.787	Russ	Slow loris
		G-Philadelphia	Racoon
		Matsu-Oki	Dog-1
		Tarrant	Dog-2
		S	Rabbit
		G-San Jose	Coyote
		G-Galveston	
		Korle Bu	
		Mobile	
		Baylor	
		Kempsey	
		D-Punjab	
0	1.007	Athens-Ga	Human A
		(Human A)	Chimpanzee
			Gorilla
			Japanese macaque
			Savannah monkey
			Mormoset
			Spider monkey
			Rhesus macaque
			Chapuchin monkey
			Mouse
-2	1.168	Camden	Horse-S
		Andrew-Minneapolis	
		Hiroshima	
		Cubujuqui	
		J-Baltimore	
-4	1.368	I	Horse-F
		N-Seattle	
		N-Baltimore	
		Hijiyama	

paper gave a complete description of the variants including specific details on mobilities and the error involved in their determination. Mobility was measured relative to human Hb-A, which is arbitrarily assigned a mobility of 1.00. Each human variant differs from human Hb-A by a characterized single amino acid substitution (a difference of two in the intact hemoglobin tetramer which is studied using electrophoresis). The 25 human variants fall into five electrophoretic charge classes, determined from the primary amino acid sequences considering only the net balance between basic and acidic amino acids. Twenty-three "evolutionary" variants (including human Hb-A) were also studied. These sequences fall into six charge classes (Table 1). Sequence differentiation among the evolutionary variants ranges from 0 differences (between human and chimpanzee, and between the dog-2 sequence and coyote), up to 91 amino acid differences (between the chicken and rabbit hemoglobin). These differences would be doubled in an intact hemoglobin molecule.

Five hemoglobin variants included in the earlier study are not included in the present analysis. These variants are Hb-Austin, Hb-Hope, cat HB-f, cat Hb-s, and human Hb-F. They were classified as statistical outliers by Fuerst and Ferrell (1980), based on the extreme deviations from expected electrophoretic mobility which were observed for the variants. The mobilities of the outlying sequences appear to be affected by additional biochemical factors not applicable to the remaining sequences. For instance, Hb-Austin migrates primarily as a dimer, rather than a tetramer, the cat hemoglobins are affected by a terminal amino acid modification, unlike that seen in any of the other samples, and Hb-F is the only fetal hemoglobin studied; the different physiological environment in which it must function may have contributed to the evolution of different residue interactions.

The average relative mobilities of each of the charge classes was estimated during the earlier study and these are given in Table 1. The difference in electrophoretic mobility between a pair of charge classes can be calculated directly. The expected distance between any two variants is taken to be the difference between the average mobilities of the net charge classes to which each of the variants belong. This assumes that the mobility differences are caused only by net charge differences attributable to the primary amino acid sequences; deviations from the expected values indicate cumulative effects of other biochemical factors which affect electrophoretic mobility. Following our earlier results, we will consider observed between-variant mobility differences which deviate from the expected difference by more than 0.03 (on a mobility scale relative to 1.00 mobility for human Hb-A) to be statistically significant deviations. This value (0.03) was estimated from the statistical sampling error of a large number of replicate electrophoretic gels containing the hemoglobin variants Hb-A, Hb-S, Hb-C, and human Hb-A2. Further details on the electrophoretic conditions used in the study can be found by referring to Fuerst and Ferrell (1980).

Results

To compare the within-species to the between-species electrophoretic behavior, we first reconsidered patterns of expected mobility, expanding upon earlier studies which attempted to predict mobility soleley on the basis of primary amino acid sequence. The observed mobility of each variant (given by Fuerst and Ferrell 1980) was compared with the mobility of its charge class. Table 2 lists the average observed deviations for the five charge classes of human variants. In Table 2, HbA has been included as a human variant, with an observed deviation of 0.007 from the expected class mobility of the 0 charge class (1.007). The average deviation of all 26 variants (0.028) is close to the value (0.030) chosen to represent a statistically significant difference. Three of the five charge classes exceed this value on average. An analysis of variance did not detect heterogeneity among charge classes for average deviation. The overall distribution of mobility deviations from expectations for the human variants is shown in Fig. 1.

Five variants reach polymorphic proportions in some human populations. These variants are Hb-S, Hb-C, Hb-E, Hb-D Punjab, and the common sequence Hb-A.

Table 2. Pattern of deviation from expected electrophoretic mobility for the human hemoglobin variants. Expected electrophoretic mobility is determined from the average mobility of a charge class as described in text

Charge Class	N	Average Deviation	S.D. among Variants
-4	4	0.027	0.016
-2	5	0.031	0.029
0	2	0.37	0.042
+2	12	0.031	0.020
+4	3	0.007	0.007
Total	26	0.028	0.022

Table 3. Pattern of deviation from expected electrophoretic mobility for the evolutionary sequences

Charge Class	N	Average Deviation	S.D. among Variants
-4	1	0.017	–
-2	1	0.047	–
0	10	0.014	0.008
+2	6	0.031	0.024
+4	4	0.013	0.012
+6	1	0.088	–
Total	23	0.023	0.022

Comparison of the "polymorphic" variants and the remaining "rare" variants showed that polymorphic variants have a smaller average deviation 0.021, compared to an average deviation of 0.030 for the rare variants. This difference was not statistically significant, however. Hb-S, which has the most severe phenotypic effects among the five "polymorphic" variants, substantially exceed the average deviation of the remaining four members of the group (0.046 versus 0.014).

Mobility deviations for the evolutionary sequences are listed in Table 3; when compared with Table 2 this shows that the evolutionary variants have mobilities which are slightly more predictable on average than human variants. The average deviation for the evolutionary sequences is 0.023. The distribution of mobility differences for the evolutionary variants is shown in Fig. 1 in comparison with the human variants. An analysis of variance fails to show any difference between the means for the two groups, however. It was just such a finding, and the assumption that the human variants represented a random sample of possible mutations, that lead us (Fuerst and Ferrell 1980) to conclude that results obtained from human variants could be extrapolated beyond the species. Further examination of the data, however, reveals this to be incorrect.

To show how and why our earlier conclusions were not justified by the data, we must examine the relationship between changes in electrophoretic mobility and amino acid replacement. Rather than further examine

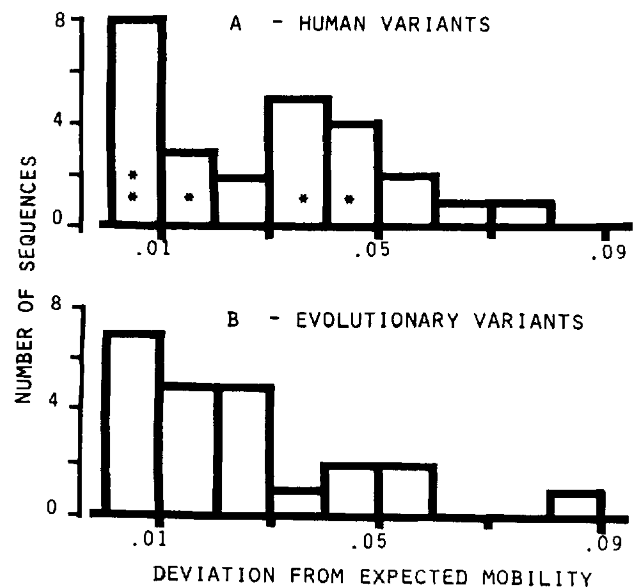

Fig. 1. Histograms showing the distribution of deviations from expected mobility for the two classes of hemoglobin variants. * – location of the five "polymorphic" human variants. A deviation of 0.03 is considered statistically significant

"expected mobility," consider the difference in mobility between variants.

Figure 2 presents a plot of observed deviations from expected mobility differences when evolutionary sequences are each compared with human Hb-A. There is a positive relationship between amino acid divergence and mobility deviation, significant below the 0.001 level, which predicts that protein sequence divergence of more than 10% (28 out of 287 residues) should result in significant deviation from the difference in electrophoretic mobility expected between a pair of variants. Of 13 comparisons with sequence divergence less than 10% from human Hb-A, two show deviations greater than 0.03, while six of nine comparisons with greater than 10% sequence divergence show significant deviations.

The same analysis can be applied to the human variants, which contain a single amino acid difference from the human Hb-A sequence. The deviations from expected mobility difference are plotted to the left of the ordinate in Fig. 2. Ten of the 25 variants show more than a 0.03 deviation from their expected difference in mobility when compared to human Hb-A.

Since there is no a priori reason for sequence divergence to be quantified only in reference to human Hb-A, a similar analysis was performed in which all pairwise distances within the evolutionary group and within the group of human variants were determined. Applying a linear regression as a first model of analysis for the pairwise comparisons (keeping in mind that we are now working with non-independent points), we obtained results similar to those obtained by standardizing against human Hb-A. For the extended pair-

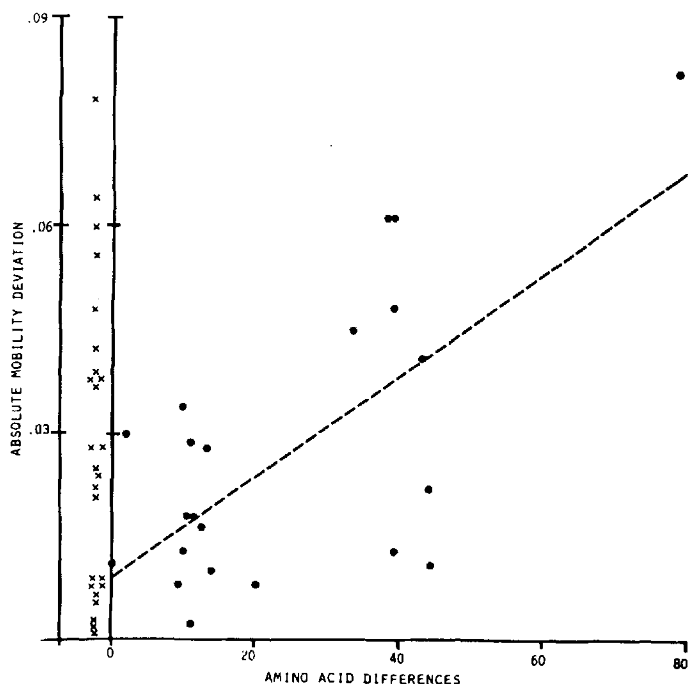

Fig. 2. Relationship between amino acid differentiation from human Hb-A and deviation from the expected difference in electrophoretic mobility between the variant and Hb-A. 0 – evolutionary variants, + – human variants (each of which differs from human Hb-A by one amino acid substitution). —— linear regression line obtained to represent the evolutionary data. The regression equation is: $Y = 0.0104 + 0.0007 X$. This predicts that significant mobility deviation (>0.03) will occur with more than 28 amino acid differences

wise analysis, significant deviation is predicted to occur at 11% sequence divergence (31 differences out of 254 residues). Because of the non-independent nature of the data in this analysis, we hesitate to place a significance statement on these results. Nevertheless, we feel the regression model is useful for predictive purposes.

Statistically, there is some question about the estimation of the expected mobility for the +6 charge class, since this class is represented by only a single variant, the chicken hemoglobin. However, when we excluded comparisons involving the chicken hemoglobin, the results were unchanged. Significant deviation is still estimated to occur with 31 amino acid changes, indicating that the results are not being distorted by comparisons involving only the most differentiated sequences.

The distribution of mobility deviation for various amounts of sequence divergence was examined. Only seven of 98 comparisons (7%) with sequence differences less than 30 residues showed significant divergence from expected distances between mobilities, while 28 of 47 (57%) comparisons with 30–40 residue differences, 46 of 72 (64%) comparisons with 41–50 differences, 11 of 14 (79%) comparisons with 51–60 differences and all of the 22 comparisons with more than 60 differences showed significant deviation. There was no taxonomic heterogeneity in this data not accounted for by differences in levels of amino acid sequence divergence. This is an important point, since it suggests that groups such as primates, which make up a major portion of the sequence data studied here, are not distorting the conclusions.

Mobility differences among the human variants are considerably less predictable than would be expected from their sequence divergence (which equal one amino acid difference for the comparisons involving Hb-A, and two amino acid differences for all other comparisons within the group). Over 58% of the pairwise comparisons have significant deviations in the distance between human variants. Twenty-four percent of human variants show mobility differences which deviate by more than 0.06 from their expected values, which is twice the deviation considered significant. This compares to the 15% of all evolutionary comparisons which have comparable deviations. However, 21 of the 38 evolutionary comparisons showing extreme deviations involve the chicken hemoglobin, which differs from various mammalian hemoglobins by 78 to 91 amino acid differences.

Deviations from expected distance between mobilities for the polymorphic variants were compared to other pairwise comparisons of the human variants. The average deviation for the polymorphic variants is 0.022, compared to 0.041 for the remaining pairwise comparisons. An analysis of variance shows that this difference is significant at the 0.05 level. The polymorphic variants, however, still have a significantly larger deviation among themselves than those observed in evolutionary comparisons among variants differing by 30 or fewer amino acids (average deviation of 0.015).

Our results suggest that electrophoretic differentiation of alleles for interspecific comparisons is not determined by the same set of biochemical factors that operate to differentiate alleles within a population. One aspect of this concerns the location of the intraspecific variants in the hemoglobin molecule, and the relationship of these locations to those of amino

Table 4. Distribution of the number of different sites in either the alpha or the beta chains of hemoglobin that are occupied by various numbers of different amino acids in the various evolutionary sequences studied here. Also included is the distribution at these sites of the different human variants. The human variants are divided into two groups, those which showed significant deviations (>0.03) in expected mobility difference from human Hb-A (see Fig. 1) and those whose deviations from Hb-A were not significant

Different Amino Acids	No. Sites	Significant Deviation	Nonsignificant Deviation
1	149	7	4
2	79	1	9
3	32	2	1
4	12	–	1
5	11	–	–
6	3	1	–
7	1	–	–

acid substitutions in the evolutionary material. Among the evolutionary variants studied here, amino acid differences occur at 138 of 287 potential sites. The number of sites occupied by different numbers of amino acid in the evolutionary sequences is given in Table 4, along with the distribution of the location of sites for the human variants. The human variants are separated into two groups, those variants which show significant deviation from their expected mobility distance from human Hb-A, and those whose deviations are not significant. These deviations were plotted in Fig. 2. Table 4 shows that variants with significant deviations occur more often at sites with only a single amino acid in the evolutionary material than would be predicted by a random sample of all sites. Nonsignificant variants often occur at sites with two or more different amino acids in the evolutionary material. The difference in occurrence at single versus multiple amino acid sites for the two groups of human variants was tested by using Fisher's exact test for a 2 x 2 table. The hypothesis being tested is that significant deviations are more likely to accompany amino acid substitutions at sites represented by only one amino acid in the evolutionary material. Such sites have only a single amino acid because they are "vital" to the function of the hemoglobin molecule, are occupied by an optimally functioning amino acid, and alternative amino acids have a great probability of causing molecular disruptions. It is these same disruptions that lead to both the physiological disturbances which have brought many of the human variants to the attention of investigators, and to the disturbance in electrophoretic behavior which causes the variant to be detectable. The Fisher's exact test indicated that the probability of observing at least the difference between deviant and nondeviant human variants was only 0.042, and therefore statistically significant. These results indicate that the evolutionary material corresponds to, at best, only a sub-set of the intraspecific sample, and that intraspecifically occurring rare alleles are more likely to be differentiable by electrophoresis than common alleles which have diverged during the process of speciation.

Discussion

Our results have important implications for the interpretation of electrophoretic data used to study interspecific differentiation. Electrophoretic data have been used to calculate divergence times, especially by means of Nei's genetic distance (Nei 1972). In the calculation of genetic distance, electrophoretic differences between "alleles" are assumed to represent single amino acid changes (as a first approximation). We can correct for electrophoretically "silent" substitutions by using the assumptions of the step-mutation model (Nei and Chakraboraty 1973). Attempts to use genetic distance as a time estimator have been moderately successful, depending on the clock calibration which is used (Gorman et al 1976). However, attempts to estimate times of racial separation, especially for humans, have sometimes met with highly suspect results (Kirk 1976). These attempts often yield estimates which suggest unreasonably long racial separation. The results presented above indicate clearly that interspecific differentiation of electromorphs is not equivalent to intraspecific change. Since many intraspecific divergence estimates yield apparent overestimation of separation time, the results presented here are relevant. Intraspecific differences among alleles are more likely to be ascertained than interspecific differences. Genetic distance estimates between populations within a species are affected principally by allele frequency differences at polypmorphic loci. Existing allelic differences would usually be detectable. In contrast, genetic distance between species is most affected by fixed allelic differences (gene substitutions). Our results indicate that these differences are likely to remain hidden when closely related species are compared electrophoretically. The genetic distances which are estimated between populations and between species would not be strictly comparable, since they would contain different levels of cryptic variation. An investigator estimating racial divergence times using a clock calibration based upon interspecific differentiation should obtain overestimates of racial time since separation.

Acknowledgements. We wish to thank R. Shaw, M. Nicholas and J. E. Brown for their constructive comments on early drafts of this paper. This work was partially supported by a grant from the National Science Foundation (DEB-8110220).

References

Ayala FJ (1975) Genetic differentiation during the speciation process. Evol Biol 8:1–78

Bunn HF, Forget BG, Ranney HM (1977) Human Hemoglobins. W.B. Saunders, Philadelphia

Dayhoff MO (1972) Atlas of Protein Sequence and Structure vol 5, National Biomedical Research Foundation, Washington, DC

Fitch WM (1974) A comparison between evolutionary substitutions and variants in human hemoglobins. Ann NY Acad Sci 241:439–448

Fuerst PA, Ferrell RE (1980) The stepwise mutation model: An experimental evaluation utilizing hemoglobin variants. Genetics 94:185–201

Gorman GC, Kim YJ, Rubinoff R (1976) Genetic relationships of three species of *Bathyogobius* from the Atlantic and Pacific sides of Panama. Copeia 1976:361–364

Heindell HC, Liu A, Paddock GV, Studnicka GM, Salser WA (1978) The primary sequence of the rabbit α-globin mRNA. Cell 15:43–54

Kirk RL (1976) Genetic distance – A pragmatic approach. Proceedings V. International Congress Human Genetics, Mexico City

Langley CH, Fitch WM (1974) An examination of the constancy of the rate of molecular evolution. J Mol Evol 3:161–177

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M, Chakraborty R (1973) Genetic distance and the electrophoretic identity of protein between taxa. J Mol Evol 2:323–328

Ohta T, Kimura M (1971) On the constancy of the evolutionary rate of cistrons. J Mol Evol 1:18–25

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204

Ramshaw JAM, Coyne JA, Lewontin RC (1979) The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics 93:1019–1037

Received September 1, 1982/ Revised June 6, 1983