Animal Blood Groups and Biochemical Genetics 16 (1985) 93-108

Blood group and protein polymorphism gene frequencies for seven breeds of horses in the United States

Ann Trommershausen Bowling and Robert S. Clark

Serology Laboratory, Department of Reproduction, School of Veterinary Medicine, University of California, Davis, CA 95616, USA

Received 12 February 1985; accepted 5 March 1985

Keywords: horse, blood groups, protein polymorphisms, gene frequencies

Summary

Gene frequencies at 20 blood group and protein polymorphism loci (*A, C, D, K, P, Q, U, Al, Tf, Pi, Xk, Es, Gc, PGD, CA, Cat, PGM, AP, Hb* and *PHI*) are given for seven horse breeds in the United States (Thoroughbred, Arabian, Standardbred, Morgan, Quarter Horse, Paso Fino and Peruvian Paso). These data are used to calculate that the battery of tests is at least 96% effective for recognizing incorrect paternity in these breeds. In addition to paternity testing, these tests can be applied to studies of breed relationships.

Introduction

The horse has a fossil record of at least 75 million years (Simpson, 1951). Domestication and concommitant human intervention in mating selection have influenced the horse for only the last few thousand years, but have resulted in breeds which serve a variety of purposes. In the last 100 years, several hundred breed societies throughout the world have defined salient characteristics and established studbooks in an attempt to preserve unique collections of traits.

It is of interest to science, breeders and breed societies to know the types and extent of variation present in these breeding groups. Among traits which vary, external form, color and disposition are relatively difficult to measure; blood groups and protein polymorphisms can be revealed by laboratory methods which allow precise definition and discrimination of variants. The biological significance of the wealth of horse blood polymorphisms is presently unknown. What factors in any species maintain polymorphism is a controversial topic (Kimura, 1982; Nei & Koehn, 1983) and will not be considered here.

This paper will address variation at 20 loci for seven horse breeds and will relate

this variation to its application in parentage determination. The relationship of these data to historical accounts of breed origins and breed structure will also be considered.

The breeds

Each breed of this study is defined by a stud-book maintained in the United States. Descriptive breed information was obtained from breed registries and from Edwards (1980). Two of the breeds (Thoroughbred and Standardbred) are bred primarily for sport (racing) and the other five primarily for pleasure riding and show ring competition. A few Arabians and some Quarter Horses are raced. Arabians and Thoroughbreds are worldwide in distribution with international exchange a practice recognized by national stud-book keepers. The Standardbred, Quarter Horse and Morgan were developed in the US from horses brought in by early settlers and have now been exported to many countries. The two gaited breeds (Paso Fino and Peruvian Paso), so called because their footfall cadence is distinctive from that of the majority of modern horses, are relatively recent arrivals to the US from Caribbean and South American countries. All seven breeds are undergoing active growth in numbers.

Thoroughbred (TB)

The recorded origins of this breed trace to the use in the early 1700s of imported 'eastern' ('Arabian', 'Barb' or 'Turk') stallions with English mares to beget horses for racing. Weatherby's 'General Stud Book' (GSB), Volume One, appeared in 1808. Registered TBs must trace in all lines to horses registered in the GSB or in similar TB stud-books maintained outside Great Britain. Thus this breed, originally a cross-bred, has had a closed stud-book for nearly 200 years. Individual TBs may excel at jumping, cross-country racing and dressage, but for the most part these horses have been selectively bred only for galloping speed. Since speed has a reasonably high heritability (More O'Farrell & Cunningham, 1974), such selection could be expected to restrict the gene pool despite the cross-bred origins of the breed. Currently, the US Jockey Club's annual registration is approaching 50 000 horses. The Jockey Club requires that all breeding stallions and most mares be blood-typed before their foals can be registered.

Arabian (AR)

In comparison to the TB, the Arabian as a breed has no origin in recorded history. Desert nomads in the Arabian peninsula developed a horse noted for such beauty, stamina and conformational strength that for centuries horse breeders went to the Arabian desert to import horses for improvement of local stock. The TB is a noteworthy result of this process. Historically, a few breeders in Europe and Egypt imported ARs not only for cross-breeding, but also to establish AR breeding pro-

grams. Most modern AR pedigrees trace to desert imports from about the mid-1800s to early 1900s. AR stud-books are maintained throughout the world. The international gene pool of ARs is essentially closed, but it is not homogeneous. Stud-books could still be acquiring new genetic material from other countries. ARs have been particularly successful in events requiring endurance, but are primarily bred for competitive show ring events in which excellence is subjectively judged. The Arabian Horse Registry of America requires all breeding stallions to be blood-typed and has set up a program to select foals at random for parentage verification. Annual foal registration is approaching 27 000.

Standardbred (ST)

This breed is primarily used for harness racing and has been developed in the US from stock derived by crossing Thoroughbred stallions (particularly, one named Messenger) with mares of sundry breeds brought by European settlers in the late 1700s and early 1800s. The name of the breed dates from John H. Wallace's 'American Trotting Register', first published in 1882. Horses listed met a standard racing speed over a mile of 2 min 30 s for trotters, or 2 min 25 s for pacers. Prominent in Register pedigrees was a stallion named Hambletonian and today few STs fail to trace to this stallion. The racing gait of the ST can be either the trot or the pace. The ability to perform one gait or the other is considered to be highly heritable and the usual selective breeding practice is to breed trotters to trotters or pacers to pacers. Thus this breed has been selected from cross-bred origins for highly heritable traits. Selection for racing performance would be expected to restrict the gene pool, despite the lack of rigorous pedigree definition in the early formation of the breed. The United States Trotting Association requires breeding stallions be blood-typed. Annual registration is about 21 000 foals.

Morgan Horse (MH)

All horses of this breed trace to one stallion, a horse owned in the late 1700s by a Vermont school teacher named Justin Morgan. The breeding origins of this stallion are not reliably documented. Thoroughbred, Arabian, Dutch Draft and Welsh Cob breeding have all been suggested, but nothing can be proved. He was bred to the variety of mares available locally and his descendents demonstrated tractability and versatility, from weight-pulling to saddle and harness racing, and use as a utility saddle horse – all by the same horse. A registry was established in 1909 to keep track of these horses and for a time cross-breeding with such breeds as Arabian, Standardbred and American Saddlebred, followed by back-crosses to Morgans, was allowed. At the present time, the registry is closed. The gene pool of this breed is expected to be quite diverse, despite descent from a single sire. Reasons for this include the variety of mare lines used in the establishment of the breed, the more recent use of cross-breeding, and selection practices based on somewhat subjective characters such as type which have a low heritability. The American Morgan Horse

Registry requires breeding stallions be blood-typed and has set up a program to select foals at random for parentage verification. Annual registration is approaching 6000 foals.

Quarter Horse (QH)
This breed had its origin in the colonial states on the East Coast of the US in the 17th and 18th centuries, although today it is more often associated with the ranching tradition of the West. Thoroughbreds and other imports were combined with horses descending from those brought to the New World by the Spanish explorers to create a versatile breed which gradually evolved distinction for two special traits: the ability to gallop very fast over short distances and the instincts to work cattle. The breed today is primarily raised for pleasure and the show ring, but QH racing and competition involving cattle work are still very important uses for which these horses are selectively bred. The stud-book was started in 1941 and remained open until recently to horses of unknown pedigree but with appropriate stock horse type and abilities. Cross-breeding with TBs is still allowed. This breed has registered over 2 000 000 animals since its inception, and is the largest horse breed registry in the world. The relative newness of the registry, the open stud-book, the broad genetic base of the stud-book entries, the large number of animals and the different directions in which the animals may be selectively bred predict that the amount of genetic diversity is quite large. The American Quarter Horse Association requires stallion blood-typing for those siring over a specified number of foals per year.

Paso Fino (PF)
This naturally-gaited riding breed is being developed in the US from Caribbean and South American sources. The horses descend from Barbs, Andalusians and Jennets brought to the New World by Spanish explorers and colonists as long as 500 years ago. The PFs have a natural lateral gait, a characteristic of most of the world's riding horses prior to the 17th century, that contrasts sharply with the square trotting gait seen in most horses today. This gait is extremely comfortable for the rider and is relatively non-tiring for horses to perform. In the US, these horses are primarily used for pleasure riding and showing. To date about 7000 horses have been registered. The Paso Fino Owners and Breeders Association requires breeding stallions be blood-typed.

Peruvian Paso (PP)
This is a naturally-gaited riding breed, developed in Peru, that has similar basic origins as the PFs, namely, from Barb, Jennet and Andalusian horses brought by early Spanish explorers and colonists. In Peru, horses of the Spanish importations were selectively bred for a natural lateral gait (probably like that of the Spanish Jennet) as well as for the 'termino' trait, a characteristic foreleg movement which is described as being like the arm motions of a swimmer. According to breed history, no

outside sources of genetic material have been introduced into the breed for several centuries. Only a few thousand of these horses are found in the US, but they are becoming popular for pleasure riding and showing. Breeding records in the US are presently being kept by several registries, including The Peruvian Paso Horse Registry of North America and the American Association of Owners and Breeders of Peruvian Paso Horses. Both registries require blood-typing of breeding stallions.

Materials and methods

Blood samples were sent to the laboratory in the period 1973 to 1983 in conjunction with breed registry regulations requiring blood-typing. For this study, the breeds involved and the number of samples tested were Thoroughbred (TB) 78 105, Arabian (AR) 25 196, Standardbred (ST) 5631, Morgan (MH) 4757, Quarter Horse (QH) 2336, Paso Fino (PF 251 and Peruvian Paso (PP) 320. The total number of samples considered was 116 596.

Two 10-ml blood samples were submitted for each horse, one in ACD anti-coagulant to use as a source of red cells and one in a dry tube to use as a serum source. Blood samples were sent to the laboratory at ambient temperature, then refrigerated until tests were completed.

Standard immunological procedures involving hemagglutination and complement mediated hemolysis (Stormont & Suzuki, 1964; Stormont, Suzuki & Rhode, 1964) were used to detect red cell alloantigens at seven internationally recognized blood group loci: *A, C, D, K, P, Q* and *U*. The reagents used to detect the antigenic specificities were produced by alloimmunization, followed by extensive testing and absorption of the immune sera until each behaved as monospecific. Assignment of alleles was based on reagent reaction patterns and was in accord with internationally agreed terminology (for an excellent recent review see Bell, 1983), with additions to the *A* and *D* loci.

In the A system, internationally defined alleles A^{adf}, A^{adg}, A^{a}, A^{b}, A^{c}, A^{e}, A^{bc}, A^{ce}, A^{bce} and A^{-} are recognized with seven reagents. An eleventh allele, A^{be}, which has not been previously noted, was also detected by our A system reagents. Addition of the alleles D^{dek} and D^{cf} to the internationally recognized D system alleles D^{ad}, D^{bc}, D^{cg}, D^{cegi}, D^{dk}, D^{d}, D^{cefg}, D^{def}, D^{de}, D^{dgh} and D^{-}, brought to 13 the number of alleles recognized by reagents for factors Da, Db, Dc, Dd, De, Df, Dg, Di and Dk. Since D^{-} has not been encountered by this laboratory, for purposes of gene frequency calculations, *D* was regarded as a 12-allele locus, functionally closed with respect to specificities Dc and Dd. Six new factor specificities, NFs 12, 23, 57, 71, 72 and D2, were extremely valuable not so much in defining new alleles as in reducing ambiguities in genotype assignment. Only Ddgh (which could be D^{dgh}/D^{dgh} or D^{dgh}/D^{d}) and Ddek (which could be D^{dek}/D^{dek}, D^{dek}/D^{dk}, D^{dek}/D^{de} or D^{dek}/D^{d}) remain ambiguous. The locus can be extended still further by reagents Dn and Do as well as reagents designated by us as NFs 19, 69 and 75, which define an additional 7 alleles

(D^{deo}, D^{dn}, D^{cg19}, D^{dgh19}, D^{ad69}, D^{adn} and D^{cg75}), but those data were not considered here.

Factor X and NFs 13 and 29 provide further definition of P system alleles (Trommershausen Bowling & Williams, 1985) but were not considered here.

Standard methods of starch (reviewed by Braend, 1973; Sandberg, 1974) and polyacrylamide (Juneja, Gahne & Sandberg, 1978) gel electrophoresis were used to identify inherited variants at the following enzyme and other protein loci: albumin (*Al*), transferrin (*Tf*), esterase (*Es*), *Xk*, *Gc*, protease inhibitor (*Pi*), 6-phosphogluconate dehydrogenase (*PGD*), phosphoglucomutase (*PGM*), phosphohexose isomerase (*PHI*), catalase (*Cat*), carbonic anhydrase (*CA*) and acid phosphatase (*AP*). Polyacrylamide gel isoelectric focusing was used for the detection of haemoglobin (*Hb*) variants (Braend & Johansen, 1983). With the exception of *Es*, these loci were considered as closed systems, with each phenotype defining a unique genotype. In the Es and Tf systems, silent alleles which have little or no detectable protein product have been reported by using standard electrophoretic techniques. Silent alleles behave as recessives to those for which protein products can be detected. From parentage information we assigned an esterase genotype heterozygous for a null allele to one MH and one PF. Only two horses have been typed which were negative for esterase; both were QHs. No conclusive evidence of a silent Tf allele was encountered among the randomly selected samples of this study, but the allele may not be confined to the single family in which it was reported (Yokohama et al., 1980).

Nomenclature has not been standardized for some electrophoretic variants, identified in the tables with brackets. A uniform nomenclature for these loci is currently under consideration by an ISABR committee.

These 20 loci are autosomal but not independent. For a review of horse linkage groups see Sandberg & Andersson (1984). *K* and *PGD* are in linkage group (LG) I; *Al*, *Gc* and *Es* are in LG II; *PHI* and *Xk* are in LG IV. *A* is closely linked to *ELA* (Equine Lymphocyte Antigen) in LG III. *C*, *D*, *P*, *Q*, *Tf*, *PGM*, *CA*, *AP* and *Hb* have been shown to be independent of the defined LGs and of each other. Of the 20 loci only three (*U*, *Pi* and *Cat*) have undefined linkage relationships.

Not every sample was tested for all the specificities of this report, particularly those tested prior to 1980, so estimates of gene frequencies were made on the basis of a random sample of 100 horses of each breed by using current records. For ARs, 100 consecutive samples of foals from the Arabian Horse Registry's random sampling program were used. For MHs data from 82 foals of the registry's random sampling program were used (reported at the 1982 ISABR Conference in Ottawa), augmented by a current sampling of every other incoming sample to provide updated estimates (based on 100 samples) for *Xk* and *A* loci. Random samples for TBs, STs, QHs, PFs and PPs were selected from incoming samples. For TBs, every 40th record in February 1983 was used (mostly young mares); for QHs and STs, every other February to May 1983 record (mostly stallions); for PFs, every second re-

cord from March 1978 to May 1983 (mostly stallions); for PPs, every record from March 1982 to May 1983.

For *CA* and *Cat*, frequencies were estimated using samples of TBs, ARs, STs, MHs and QHs tested in 1972 to 1976 when these tests were run routinely. *PGM* frequencies were estimated using every sample tested from each breed in the period April to June 1983. *Hb* frequencies were estimated for TBs by using 200 consecutive samples in August 1983, and for ARs, MHs, STs and QHs by using all samples tested between September 1983 and January 1984.

All blood-typing data for the 116 596 horses of this study were entered into a laboratory minicomputer. Search programs were used to identify rare variants in each breed which might not have been encountered in the particular random sample selected. Rare variants were required to have occurred in three or more animals not known to be related to each other as parent and offspring to minimize the effect of a few undetected cross-breds should they occur. Such variants are indicated in Tables 1 and 2 by parentheses if they occurred at a frequency greater than or equal to 0.001 overall in the breed. The number of occurrences of a rare type for each breed would need to be: 90 (TB); 30 (AR); 6 (ST); 5 (MH) and 3 (QH). Too few samples of PFs and PPs were studied to meet these stringent criteria.

Blood group frequencies were calculated as follows: for *C, K* and *U* by the square-root method; for *P*, with Bernstein's equation developed for human *ABO;* for *A* and *Q*, with the square-root method to obtain frequencies of the null alleles, followed by calculation and adjustment of the other allele frequencies to obtain the best fit to the data under the assumption of Hardy-Weinberg equilibrium; for *D*, by direct counting assuming no ambiguous phenotypes. Computation of allelic frequencies for protein polymorphisms was done by direct counting from phenotypes. Probabilities of exclusion (PEs) for the *D* and protein polymorphism loci were estimated according to Jamieson (1965). Estimates of PEs for other blood group loci were obtained in a manner similiar to that used by Rendel & Gahne (1961) for PE estimates of cattle tests. For simplification of PE calculations, *A* and *Q* were assumed to be 3-allele loci, which resulted in an underestimate of PE. For the *D* locus, the assumption of no ambiguous phenotypes resulted in a slight overestimate of PE.

Results and discussion

Gene frequencies

Alloantigenic red cell marker frequencies are presented in Table 1 and serum and red cell protein marker frequencies in Tables 2 and 3. Observed and expected phenotypes (data not given) showed the loci to be in genetic equilibrium, according to Hardy-Weinburg law, with a few minor deviations ($P \leq 0.05$), e.g. *Al* in PP, *Pi* in ST, *PGD* in AR. Such exceptions may be due to the non-random breeding structure of these populations or possibly to faulty assignment of genotypes but were not con-

Table 1. Gene frequencies of alloantigenic red cell markers.

Locus	Allele	Breed						
		TB (n = 100)	AR (n = 100)	ST (n = 100)	MH (n = 84)	QH (n = 100)	PF (n = 108)	PP (n = 100)
A	*adf*	0.849	0.636	0.566	0.529	0.440	0.263	0.321
	adg	(0.001)*	0.182	(0.001)	0.040	0.050	0.233	0.214
	a	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	b	0.010	0.055	0.389	0.201	0.150	0.144	0.065
	c	0.007	0.007	0.000	0.037	0.010	0.042	0.004
	e	0.000	0.000	0.015	0.055	0.010	0.000	0.000
	bc	0.000	0.020	0.020	0.006	0.010	0.014	0.000
	be	0.000	0.000	(0.001)	0.000	0.000	0.000	0.000
	ce	0.000	0.000	0.000	0.010	0.000	0.000	0.000
	bce	0.000	0.000	0.000	(0.001)	0.000	0.000	0.000
	–	0.134	0.100	0.010	0.122	0.330	0.304	0.396
C	*a*	0.588	0.827	0.588	0.688	0.755	0.369	0.735
	–	0.412	0.173	0.412	0.312	0.245	0.631	0.265
D	*ad*	0.000	(0.003)	0.005	0.135	0.050	0.079	0.100
	dk	0.326	0.455	0.230	0.255	0.285	0.213	0.350
	d	0.022	0.000	0.030	0.005	0.045	0.027	0.004
	dgh	0.063	0.015	0.095	0.085	0.075	0.060	0.006
	de	0.031	0.125	0.105	0.100	0.155	0.153	0.145
	dek	0.000	0.000	0.000	0.005	0.005	0.000	0.000
	dfk	0.031	(0.002)	0.070	0.005	0.035	0.065	0.015
	bc	0.205	0.260	0.045	0.065	0.185	0.218	0.170
	cg	0.214	0.130	0.345	0.300	0.110	0.144	0.100
	cegi	0.107	0.015	(0.001)	0.005	0.050	0.005	0.010
	cefg	0.000	0.000	(0.002)	0.040	0.005	0.037	0.100
	cf	0.000	0.000	0.075	0.000	(0.001)	0.000	0.000
K	*a*	0.025	(0.001)	0.322	0.006	0.041	0.001	0.025
	–	0.975	1.000	0.678	0.994	0.959	0.999	0.975
P	*a*	0.192	0.405	0.292	0.220	0.260	0.307	0.200
	b	0.080	0.030	0.045	0.073	0.020	0.069	0.020
	–	0.728	0.565	0.663	0.707	0.721	0.624	0.780
Q	*abc*	0.613	0.205	(0.006)	0.006	0.175	0.076	0.020
	ac	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	b	0.011	0.080	0.000	0.037	0.099	0.082	0.219
	c	0.022	0.290	0.067	0.296	0.206	0.210	0.250
	–	0.355	0.424	0.933	0.661	0.520	0.632	0.510
U	*a*	0.094	0.213	0.340	0.267	0.252	0.347	0.329
	–	0.906	0.787	0.660	0.733	0.748	0.653	0.671

* Specificity is present in the breed at frequency greater than or equal to 0.001, but was not found in the random sample under study.

Table 2. Gene frequencies of serum and red cell protein markers.

Locus	Marker	Breed						
		TB (n = 100)	AR (n = 100)	ST (n = 100)	MH (n = 82)	QH (n = 100)	PF (n = 109)	PP (n = 100)
Al	A	0.195	0.435	0.615	0.445	0.355	0.402	0.275
	I	0.000	0.000	(0.003)*	(0.001)	0.000	0.000	0.000
	B	0.805	0.565	0.385	0.555	0.645	0.598	0.725
Tf	D	0.320	0.290	0.245	0.293	0.225	0.262	0.295
	[E]**	0.000	0.000	(0.003)	0.006	0.005	0.009	0.000
	F_1	0.310	0.000	0.000	0.000	0.185	0.000	0.000
	F_2	0.155	0.455	0.535	0.543	0.335	0.178	0.225
	F_3	0.000	0.000	0.000	(0.001)	0.005	0.064	0.020
	G	0.000	0.000	(0.001)	0.006	0.005	0.000	0.000
	[H_1]	0.025	0.135	0.005	0.067	0.050	0.145	0.130
	[H_2]	0.000	0.000	0.000	0.000	0.005	0.065	0.025
	J	0.000	0.000	0.000	0.000	0.000	0.011	0.000
	M	0.000	0.000	0.000	0.043	0.000	0.000	0.000
	O	0.055	0.120	0.100	0.018	0.070	0.154	0.195
	R	0.135	(0.002)	0.115	0.024	0.115	0.112	0.110
Pi	F	0.030	0.090	(0.002)	0.019	0.005	0.000	0.000
	G	0.015	0.090	0.005	0.037	0.040	0.066	0.000
	I	0.060	(0.002)	0.045	0.049	0.040	0.014	0.000
	L	0.450	0.220	0.335	0.303	0.379	0.448	0.455
	N	0.185	0.010	0.110	0.179	0.126	0.042	0.025
	[O]	0.000	0.070	0.000	0.000	0.000	0.000	0.000
	S	0.130	0.255	0.070	0.079	0.218	0.236	0.140
	U	0.130	0.250	0.435	0.291	0.177	0.099	0.145
	other	(0.001)	0.015	(0.001)	0.043	0.015	0.094	0.235
Xk	F	0.020	0.000	(0.011)	0.035	0.010	0.017	0.005
	K	0.980	0.940	1.000	0.910	0.940	0.896	0.780
	S	0.000	0.059	(0.006)	0.055	0.050	0.087	0.215
Es	F	0.060	0.020	0.055	0.177	0.045	0.056	0.125
	G	0.000	0.035	0.155	0.085	0.061	0.178	0.240
	H	0.000	0.000	(0.001)	0.006	0.020	0.023	0.000
	I	0.905	0.935	0.705	0.689	0.834	0.701	0.465
	S	0.035	0.010	0.085	0.037	0.040	0.037	0.170
	O	0.000	0.000	0.000	0.006	(0.001)	0.005	0.000
	R	0.000	0.000	0.000	(0.010)	(0.001)	0.000	0.000
Gc	F	0.939	0.970	0.760	0.877	0.878	0.982	0.983
	S	0.061	0.030	0.240	0.123	0.122	0.018	0.017
PGD	D	0.000	0.005	0.000	0.000	0.005	0.006	0.000
	F	0.648	0.652	0.845	0.835	0.677	0.867	0.725
	S	0.352	0.343	0.155	0.165	0.318	0.127	0.275

* Specificity is present in the breed at frequency indicated in parentheses, but was not found in the random sample under study.

** Brackets indicate that nomenclature has not been standardized.

Table 3. Gene frequencies for additional red cell proteins.

Locus	Variant	Breed						
		TB	AR	ST	MH	QH	PF	PP
		(239)*	(505)	(192)	(72)	(273)		
CA	F	0.029	0.002	0.055	0.035	0.041	–	–
	I	0.971	0.998	0.940	0.965	0.954	–	–
	L	0.000	0.000	0.005	0.000	0.005	–	–
		(205)	(481)	(186)	(72)	(253)		
Ca	F	0.002	0.032	0.223	0.326	0.190	–	–
	S	0.998	0.968	0.777	0.674	0.810	–	–
		(1568)	(987)	(484)	(155)	(131)	(75)	(107)
PGM	F	0.002	0.111	0.150	0.148	0.027	0.133	0.136
	S	0.998	0.889	0.850	0.852	0.973	0.867	0.864
	V	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		(825)	(265)	(72)	(107)	(120)		
AP	F	0.006	0.096	0.063	0.051	0.025	–	–
	S	0.994	0.904	0.937	0.949	0.975	–	–
		(200)	(682)	(242)	(215)	(124)		
Hb	A**	0.000	0.000	0.019	0.000	0.024	–	–
	BI	0.198	0.565	0.459	0.581	0.381	–	–
	BII	0.802	0.435	0.522	0.419	0.595	–	–
		(646)	(561)	(139)	(153)	(118)		
PHI	F	0.000	0.011	0.097	0.013	0.013	–	–
	I	1.000	0.989	0.903	0.967	0.974	–	–
	S	0.000	0.000	0.000	0.020	0.013	–	–

* Number of animals tested to provide frequency estimate is given in parentheses above the first entry for each locus.

** Includes both A and AII.

sidered of such magnitude to invalidate the frequencies calculated on an assumption of population equilibrium.

Each breed had a unique profile of gene frequency data. In general, these data appeared to be similar to gene frequencies previously reported for TBs, ARs, STs and QHs (e.g. Scott, 1978; Suzuki, 1978; Trommershausen Bowling et al., 1980). We are not aware of previous reports for MHs, PFs or PPs.

Efficacy of blood-typing tests

The calculated effectiveness of the tests for detecting incorrect paternity when blood samples are tested from sire, dam and offspring was at least 96 % using 20 internationally defined loci (Table 4). In actual practice, the effectiveness could be

Table 4. Probability of exclusion (PE) at 20 loci.

Locus	Breed						
	TB	AR	ST	MH	QH	PF	PP
A	0.01	0.04	0.08	0.06	0.14	0.13	0.09
C	0.02	0.00	0.02	0.01	0.00	0.06	0.00
D	0.59	0.46	0.61	0.62	0.67	0.69	0.61
K	0.02	0.00	0.07	0.01	0.03	0.00	0.02
P	0.16	0.09	0.12	0.15	0.08	0.14	0.11
Q	0.01	0.08	0.06	0.08	0.08	0.06	0.20
U	0.06	0.08	0.05	0.08	0.08	0.06	0.07
Al	0.13	0.18	0.18	0.19	0.18	0.18	0.16
Tf	0.53	0.42	0.38	0.36	0.58	0.67	0.59
Es	0.09	0.06	0.27	0.28	0.17	0.26	0.43
Xk	0.02	0.05	0.00	0.08	0.06	0.09	0.14
Pi	0.51	0.61	0.43	0.58	0.54	0.50	0.46
PGD	0.18	0.18	0.11	0.12	0.18	0.10	0.16
Gc	0.05	0.03	0.15	0.10	0.10	0.02	0.02
CA	0.03	0.00	0.06	0.03	0.04	–	–
Cat	0.00	0.03	0.14	0.17	0.13	–	–
PGM	0.00	0.09	0.11	0.11	0.03	0.10	0.10
AP	0.00	0.08	0.05	0.05	0.02	–	–
Hb	0.13	0.18	0.21	0.18	0.21	–	–
PHI	0.00	0.01	0.08	0.03	0.03	–	–
Total PE for 20 loci	0.96	0.97	0.98	0.98	0.99	>0.98	>0.98

slightly lower when a mare has been bred to two stallions since owners tend to breed mares to related stallions (e.g. a stallion and his son) rather than two stallions selected at random from the breed. It is expected that the tests would be slightly more effective than 96 % if both parents were incorrect.

The efficacy of a locus depends on the number of alleles, their frequencies and whether the genotypes can be directly determined from the phenotypes (Rendel & Gahne, 1961). The most effective loci in this study were those closed loci having five or more alleles with appreciable frequencies, namely *D, Tf,* and *Pi.* Each of these had an estimated PE of 0.36 or greater. Taken together, the three loci had a theoretical PE in each of the seven breeds of between 0.86 and 0.98. The additional 17 loci tested thus contributed only slightly to the cumulative total PE.

Other serum (plasma) and red cell loci have been described (e.g. Weitkamp, Guttormsen & Castello-Leary, 1982; Weitkamp, Castello-Leary & Guttormsen, 1983). Tests have been developed as well for genetic markers of lymphocytes, and undoubtedly the most useful single locus by which the test battery can be augmented is that for Equine Lymphocyte Antigen (*ELA*). ELA currently has ten specificities behaving as alleles at a single locus which have been recognized and accepted by two international ELA workshops (Bull, 1983; Bailey et al., 1984).

Table 5. Number of variants per locus with a frequency equal to or greater than 0.001.

Locus	Breed						
	TB	AR	ST	MH	QH	PF	PP
A	5	6	6	8	7	6	6
C	2	2	2	2	2	2	2
D	8	8	11	11	12	10	10
K	2	2	2	2	2	1	2
P	3	3	3	3	3	3	3
Q	4	4	3	4	4	4	4
U	2	2	2	2	2	2	2
RBC total	26	27	29	32	32	28	29
Al	2	2	3	3	2	2	2
Tf	6	5	7	9	10	9	7
Es	3	4	5	7	7	7	4
Xk	2	2	3	3	3	3	3
Pi	8	9	7	8	8	7	5
PGD	2	3	2	2	3	3	2
Gc	2	2	2	2	2	2	2
Protein total	25	27	29	34	35	33	25
Combined RBC and protein total	51	54	58	66	67	61	54

Breed relationships and breed structure

The number of blood type polymorphisms with allelic frequencies greater than or equal to 0.001 is shown for each breed in Table 5. Of the 75 variants assayed, the smallest number of polymorphisms was found in TBs (51 or 68 %), and the greatest number in QHs (67 or 89 %).

The computer program of Dowling & Moore (1984) was used to calculate Nei's measures of normalized genetic identity (*I*) and standard genetic distance (*D*) between two populations (Nei, 1972) for seven breeds over 14 loci: *A, C, D, K, P, Q, U, Al, Tf, Pi, Xk, Es, Gc* and *PGD*. Average heterozygosity (*Jx*) for each breed was also calculated with this program. The results are shown in the matrix of Table 6.

Average heterozygosity values ranged from a low of 0.378 ± 0.069 in TBs to a high of 0.481 ± 0.066 in PFs. In various wild and domestic animal species, the probability of heterozygosity per individual ranges from 5-15 % (Selander et al., 1970), except for populations having a small pool of breeding animals. Calculations for those values included monomorphic loci, unlike this study, which only considered polymorphic loci likely to be useful in parentage determinations.

Table 6. Normalized genetic identity, *I*, standard genetic distance, *D*, and average heterozygosity, *Jx**, for seven breeds at 14 polymorphic loci**.

Breed	TB	AR	ST	MH	QH	PF	PP
TB	0.378 ± 0.069	0.062 ± 0.026	0.141 ± 0.052	0.085 ± 0.041	0.043 ± 0.022	0.098 ± 0.040	0.104 ± 0.043
AR	0.940	0.416 ± 0.071	0.094 ± 0.028	0.038 ± 0.016	0.022 ± 0.013	0.073 ± 0.033	0.065 ± 0.026
ST	0.868	0.910	0.447 ± 0.054	0.036 ± 0.015	0.073 ± 0.022	0.088 ± 0.030	0.122 ± 0.034
MH	0.918	0.962	0.965	0.450 ± 0.061	0.025 ± 0.013	0.049 ± 0.024	0.051 ± 0.021
QH	0.958	0.978	0.929	0.975	0.464 ± 0.066	0.042 ± 0.023	0.035 ± 0.017
PF	0.907	0.930	0.916	0.952	0.958	0.470 ± 0.073	0.039 ± 0.021
PP	0.901	0.937	0.885	0.950	0.966	0.961	0.481 ± 0.066

* Nei's *I* (genetic identity) is below the diagonal. *D* (genetic distance) is above the diagonal. *Jx* (average heterozygosity) is on the diagonal. Standard error values are given for *D* and *Jx*.
** Loci and frequency values used were from Tables 1 and 2.

Thoroughbred. TBs had comparatively little variation in blood type markers, consistent with their history of nearly two hundred years of a closed stud-book combined with intense selection for a single trait of rather high heritability. They had the fewest variants (51 alleles at 14 loci had frequencies greater than or equal to 0.001) and the lowest heterozygosity value (0.378 ± 0.069) compared with the other breeds of this study. Variants did occur with frequencies less than 0.001 (e.g. D^{ad}, Al^{l}, Es^{G}). Blood-typing many animals of a breed may negate general statements, made on the basis of small sample size, that certain factors never occur. Clearly, blood-typing alone cannot be used to determine whether a certain horse is a TB (or any other breed). Such a determination is based on meeting requirements for stud-book entry which may include blood type qualification to registered parents.

The breed that was most closely related to the TB based on these blood type data was the QH. This corresponds to the historical and contemporary use of TB stock within the QH breed structure. The AR relationship was not as close and was consistent with the minimal amount of overlap of contemporary ARs with TBs, even though the AR was historically important in the foundation of the TB. The distance between TB and ST was greater than any other distance of this study. Despite the historical contribution of the TB to the ST, the dissimilar selective breeding goals of

the breeds – racing speed at the gallop versus racing speed at the trot or pace – appears to have been sufficient to create the considerable difference.

Arabian. The narrow pedigree (genetic) base of the AR breed, in which no one trait is currently highly selected, provided an interesting contrast to the TB, in which selection for one trait has narrowed a genetic base which at one time may have been much broader. The small set of AR foundation animals was reflected in the rather small number of blood type variants found in the breed (54 out of 71 assayed) and in the relatively low value for heterozygosity (0.416 ± 0.071). These values were only slightly different from those of TBs.

QH, MH, ST and TB stud-books all have horses which trace directly to stock in the stud-book of the Arabian Horse Registry of America. For the most part, these connections were not closer than three generations for any animal which might have been blood-typed in this study. The number of animals which can show such associations would be few in any of these breeds. None of these breeds appeared to have the Pi-O type which seems unique to ARs but perhaps this variant is in AR bloodlines which have not been crossed into those stud-books. The rather close relationship of the QH to the AR by Nei's genetic distance calculations did not conform to expectations. If this relationship were due to the QH's association with the TB, then the TB should also be relatively close to the AR by Nei's measure, but it was not.

Standardbred and Morgan. These two breeds were both derived from early colonial horses and have been considered as separate breeds only within less than the last hundred years. The distance data clearly showed their similarity, both in relationships to each other and to all the other breeds. The STs have been intensively selected for harness racing speed, whereas the MHs have been selected from a similar founder population on the basis of descent from a single sire line. Despite the divergent selection programs, the two populations were still rather similar. STs had fewer polymorphisms (58 versus 66) but they had approximately the same mid-range heterozygosity values.

Surprisingly, the rather high gene frequencies for D^{ad} and Tf^{M} in MHs compared to STs (and to all other breeds of this study) did not have much impact on the calculated distance relationships of these breeds. It is tempting to speculate that the founder stallion of the Morgan breed in fact may have possessed these factors. Tf-M was unique to MHs among the breeds of this study. Among breeds of the world, it is reported in significant frequency only among pony breeds, especially Shetlands (Kaminski & Urbanska-Nicolas, 1979).

Quarter Horses. The diversity of sources which contributed to the formation of this breed was evident from the distance figures, in the calculations of heterozygosity and in the polymorphism count. The contribution of the TB is most clearly seen in a

breedwise comparison of the Tf-F_1 variant, which was a frequently encountered variant in TBs and QHs, but not present in significant levels in the other breeds. Perhaps the single most conspicuous difference between TBs and QHs was in the gene frequencies of the A locus alleles. QHs had a rather high frequency of *A–*, similar to the PFs and PPs, and unlike TBs.

Paso Fino and Peruvian Paso. The PF and PP were recently imported to the US and are, of course, not expected to have made any contribution to the other breeds of this study. For both breeds, the closest relationships shown by the data were with each other. This is consistent with their derivation from Spanish horses and the high degree of selection practiced by breeders for a gait which is now nearly unique to these breeds. Their proximate relationship to the QH in the data may reflect the influence of the early Spanish imports to the US for the QH.

In light of the isolation of the PP for several centuries from other horses except those also derived from colonial Spanish imports, it was surprising that the breed showed higher heterozygosity than either TBs or ARs. In fact, the theoretical calculations indicated that PPs had the highest heterozygosity value of any breed of this study. It could be that the population was not homogeneous; the horses we typed were not in a single stud-book, although many animals were double registered.

The diversity of polymorphisms and the mid-range value of homozygosity for the PF probably reflect the variety of sources from which this breed is being created. The presence of Tf-J, which has heretofore been reported only in Andalusians, provided evidence of the relationship of PFs to that classic Spanish breed.

Acknowledgements

The authors gratefully acknowledge the efforts and diligence over the past ten years of the staff of the Serology Laboratory, the support of the various breed associations affiliated with the project, and the critical reading of the manuscript by Michael Bowling, Cecilia Penedo and Linda Nickel.

References

Bailey, E., D. F. Antczak, D. Bernoco, R. W. Bull, R. Fister, G. Guerin, S. Lazary, S. Matthews, J. McClure, J. Meyer, V. D. Mottironi & J. Templeton, 1984. Joint Report of the Second International Workshop on Lymphocyte Alloantigens of the Horse, held 3-8 October 1982. *Animal Blood Groups and Biochemical Genetics* 15: 123-132.

Bell, K., 1983. The blood groups of domestic mammals. In: N. S. Agar & P. G. Board (Eds), Red blood cells of domestic mammals. Elsevier Biomedical Press, Amsterdam, p. 133-164.

Braend, M., 1973. Genetic variation in equine blood proteins. In: J. T. Bryans & H. Geber (Eds), Proceedings of the 3rd International Conference on Equine Infectious Diseases (Paris, 1972). Karger, Basel, p. 394-406.

Braend, M. & K. E. Johansen, 1983. Haemoglobin types in Norwegian horses. *Animal Blood Groups and Biochemical Genetics* 14: 305-307.

Bull, R. W. (Ed), 1983. Joint Report of the First International Workshop on Lymphocyte Alloantigens of the Horse, held 24-29 October 1981. *Animal Blood Groups and Biochemical Genetics* 14: 119-137.

Dowling, T. E. & W. S. Moore, 1984. A program for estimating genetic variability within and between populations. *Journal of Heredity* 75: 416.

Edwards, E. H., 1980. A Standard Guide of Horse and Pony Breeds. McGraw-Hill, New York.

Jamieson, A., 1965. The genetics of transferrins in cattle. *Heredity* 20: 419-441.

Juneja, R. K., B. Gahne & K. Sandberg, 1978. Genetic polymorphism of the vitamin D binding protein and another post-albumin protein in horse serum. *Animal Blood Groups and Biochemical Genetics* 9: 29-36.

Kaminski, M. & H. Urbanski-Nicolas, 1979. Electrophoretic polymorphism of proteins in the blood of horses: studies of eleven pony breeds or populations. *Biochemical Systematics and Ecology* 7: 229-237.

Kimura, M. (Ed), 1982. Molecular Evolution, Protein Polymorphism and the Neutral Theory. Springer.

More O'Ferrall, G. J. & E. P. Cunningham, 1974. Heritability of racing performance in Thoroughbred horses. *Livestock Production Science* 1: 87-97.

Nei, M., 1972. Genetic distance between populations. *American Naturalist* 106: 283-292.

Nei, M. & R. K. Koehn (Eds), 1983. Evolution of Genes and Proteins. Sinauer.

Rendel, J. & B. Gahne, 1961. Parentage tests in cattle using erythrocyte antigens and serum transferrins. *Animal Production* 3: 307-314.

Sandberg, K., 1974. Blood typing of horses: current status and application to identification problems. In: Proceedings of the 1st World Congress of Genetics applied to Livestock Production (Madrid, 1974), p. 253-265.

Sandberg, K. & L. Andersson, 1984. Genetic linkage in the horse. I. Linkage relationships among 15 blood marker loci. *Hereditas* 100: 199-208.

Scott, A. M., 1978. Immunogenetic analysis as a means of identification in horses. In: J. T. Bryans & H. Gerber (Eds), Equine Infectious Diseases. Veterinary Publications, Princeton, p. 259-268.

Selander, R. K., S. Y. Yang, R. C. Lewontin & W. E. Johnson, 1970. Genetic variation in the horseshoe crab (*Limulus polyphemus*), a phylogenetic 'relic'. *Evolution* 24: 402-414.

Simpson, G. G., 1951. Horses. Oxford University Press, New York.

Stormont, C. & Y. Suzuki, 1964. Genetic systems of blood groups in horses. *Genetics* 50: 915-929.

Stormont, C., Y. Suzuki & E. A. Rhode, 1964. Serology of horse blood groups. *Cornell Veterinarian* 54: 439-452.

Suzuki, Y., 1978. Studies on blood groups of horses. *Memoirs of the Tokyo University of Agriculture* 20: 1-50.

Trommershausen Bowling, A., J. Dodd, Y. Suzuki & C. Stormont, 1980. Population and family data on Arabian horses in the United States (abstract). *Animal Blood Groups and Biochemical Genetics* 11: Supplement 1: 19-20.

Trommershausen Bowling, A. & M. J. Williams, 1985. Expansion of the P blood group system of the horse. *Animal Blood Groups and Biochemical Genetics* 16: 145-148.

Trommershausen-Smith, A., 1979. Positive horse identification. Part 3: coat color genetics. *Equine Practice* 1: 24-35.

Weitkamp, L. R., P. Costello-Leary & S. A. Guttormsen, 1983. Equine marker genes: polymorphism for plasminogen. *Animal Blood Groups and Biochemical Genetics* 14: 219-223.

Weitkamp, L. R., S. A. Guttormsen & P. Costello-Leary, 1982. Equine gene mapping: close linkage between the loci for soluble malic enzyme and Xk (Pa). *Animal Blood Groups and Biochemical Genetics* 13: 279-284.

Yokohama, M., M. Kuwazima, N. Miura, M. Inoue, T. Mogi, T. Hosoda & T. Nakajima, 1980. On existence of a silent gene in equine transferrin system. *Japanese Journal of Zootechnical Science* 51: 331-335.