Genet. Res., Camb. (1983), **41**, *pp.* 241–258
With 5 text-figures
Printed in Great Britain

Macrogeographic genetic variation in a human commensal: *Aedes aegypti*, the yellow fever mosquito

By GRAHAM P. WALLIS, WALTER J. TABACHNICK[1] and JEFFREY R. POWELL

Department of Biology, Yale University, P.O. Box 6666, New Haven CT, 06511

(*Received 27 August 1982 and in revised form 6 December 1982*)

SUMMARY

Genetic variation at 11 enzyme loci in *Aedes aegypti* populations collected from 63 localities around the world is presented. A UPGMA tree based on genetic distance values clusters populations of sylvan *A. aegypti formosus* from West Africa and East Africa together, along with Asian and south-eastern U.S. populations. Domestic *A. aegypti aegypti* from East Africa and all other New World populations form the other major cluster. Multivariate discriminant analysis allows recognition of seven major global 'genetic–geographic' groups, which are consistent with the genetic distance data. Populations from the south-eastern U.S. are clearly genetically distinct from other U.S. and Mexican populations. This distinct gene pool may be indicative of reinfestation of areas where *A. aegypti* appears to have been absent in the recent past. Other evolutionary and epidemiological implications of the genetic population structure of *Aedes aegypti* are discussed.

1. INTRODUCTION

Human commensals have an important contribution to make to all areas of biology by their ready availability and general compliance to laboratory manipulation. In particular, they have great potential in increasing our understanding of several evolutionary processes as a result of their historical association with man. As we have expanded the region of our domain to include new environments, closely associated species have had to adapt to new pressures. Areas where man exists free of a certain species may suddenly become available to that species, and one may see rapid evolution of a recently introduced form. As man's patterns of migration and trade change, so does gene flow between populations of commensals. Secondary contact between two specialized forms may be monitored in the wild or imitated in the laboratory as an aid to understanding the speciation process. Commensals suffer direct and indirect interference from man, and have to respond appropriately to survive. In the case of an eradication attempt by chemical means, for example, the necessary genetic change in the population will rely on rebuilding from a few selected individuals, and the resulting population may well

[1] Present Address: Department of Biology, Loyola University, Chicago, IL 60626.

differ from the original one at a number of loci due to founder effects. In this situation, neutral or even detrimental genetic change may arise as a result of hitch-hiking by linked genes.

The house-mouse, *Mus musculus*, and the fruit-fly, *Drosophila melanogaster*, are good examples of human commensals that have been extensively studied genetically. Using large numbers of enzyme loci, local genetic differentiation (Selander & Yang, 1969) and subspecific differentiation (Selander, Hunt & Yang, 1969; Hunt & Selander, 1973) have been studied in *M. musculus;* Berry (1977) reviews work done on some island populations. Singh, Hickey & David (1982) present data for *D. melanogaster*. However, knowledge of world-wide isozyme variation in these species is not as extensive as it is for the mosquito species *Aedes aegypti*.

From a public-health viewpoint, mosquito-borne disease is of primary concern. *Aedes aegypti* is unusual in that its domestic form exists in strict association with humans, breeding in water containers inside and close to dwelling places, making it a commensal in a full sense of the word. Recent dengue epidemics in the Caribbean, and the ever-present threat of jungle and urban yellow fever in South America and Africa highlight the importance of *A. aegypti* as a target for control programmes, as it is the primary vector of these diseases. *A. aegypti* is one of the most widespread insect species, being found in close association with man throughout the tropical and subtropical world. The species is particularly interesting as one can study the extant ancestral form in Africa, where domestic and sylvan subspecies exist sympatrically (Petersen, 1977). Pioneer work by Craig and co-workers (e.g. Craig & Hickey, 1966) revealed considerable genetic heterogeneity in wild populations of *A. aegypti*. The species has latterly been shown to exhibit considerable morphological (McClelland, 1974; Petersen, 1977; VandeHey, Leahy & Booth, 1978), physiological (Machado-Allison & Craig, 1972), and behavioural (Trpis & Hausermann, 1975, 1978; Petersen, 1977; Leahy, VandeHey & Booth, 1978) differentiation. In such a polytypic species, there is also likely to be heterogeneity in those genetically controlled features of the vector that affect its overall ability to transmit disease. As efficiency of disease transmission can depend upon the geographic origin of the vector population (Aitken, Downs & Shope, 1977; Beaty & Aitken, 1979; Gubler, Novak & Mitchell, 1982; Tabachnick *et al.*, 1982), genetic delineation of geographic regions may be helpful in assessing the potential threat of invading populations. Routes of gene flow may also be elucidated by this means. The assumption made is that a gene pool that is an entity in terms of isozyme variability will possess other features common to the group as a whole. It has therefore been our contention that knowledge of the population genetics and evolutionary history of *A. aegypti* is of importance in understanding its role as a carrier of epidemic disease.

Previous electrophoretic work on *A. aegypti* includes an investigation of the two sympatric East African subspecies (Tabachnick, Munstermann & Powell, 1979), examination of microgeographic and temporal variation (Tabachnick & Powell, 1978; Tabachnick, 1982) and world-wide surveys (Tabachnick & Powell, 1979; Powell, Tabachnick & Arnold, 1980; Powell, Tabachnick & Wallis, 1982). The formal genetics of *A. aegypti* is advanced enough to provide an extensive genetic

map of morphological, physiological and biochemical markers (Munstermann & Craig, 1979; Wallis & Tabachnick, 1982). In our analyses we use loci that are spread throughout the genome.

The aims of the present study are to characterise further the world-wide genetic structure of *A. aegypti* and illustrate the possible epidemiological implications. Previous studies (Tabachnick & Powell, 1979) have supported the description of the two subspecies, *A. aegypti aegypti* (domestic form) and *A. a. formosus* (sylvan form), made by Mattingly (1957), and the idea that *A. aegypti* has only recently been introduced into Asia (Smith, 1956). Here we give additional data supporting these observations, as well as focusing more closely on *A. aegypti* populations in the southern United States.

2. MATERIALS AND METHODS

A total of 96 collections representing 63 localities from 21 countries (Table 1) have been analysed electrophoretically for 11 enzyme coding loci (8 enzymes). Many populations have been analysed for up to 22 loci, but these additional loci will not be considered here as they are either monomorphic or otherwise uninformative. Where multiple collections have been received from the same locality, unweighted composite mean gene frequencies have been calculated, and are presented here. Table 2 lists collections that have been analysed since

Table 1. *The 63 localities from which* Aedes aegypti *has been sampled*

Country	Population	Country	Population
Kenya	Kwa Bendegwa	Indonesia	Grogol
	Mgandini		Semarang
	Majengo		Tandjung Priok
	Kombeni	India	Bangalore
	Shimba Hills	Fiji	Walu Bay
	Kwa Dzivo	Guyana	Georgetown
	Nyali	Suriname	Paramaribo
Uganda	Bwerenga	Venezuela	Caracus
Tanzania	Dar Es Salaam		Maracay
Nigeria	Ogui	Guatemala	Escuintla
	Ukana	Trinidad	Port of Spain
	Mamu		Trinidad City
	Egede		Felicity
	Abor		Sealots
	Enugu	Puerto Rico	San Juan
Upper Volta	Bobo		Mayaguez
	Kari		Arecibo
	Bwombi		Barcelonata
	Ouagadougou	Jamaica	Montego Bay
Senegal	Dakar	United States	Florida (2)
	Kedougou		Alabama
	N'goyé		Mississippi
Ghana	Akosombo		Louisiana (3)
Gambia	Diabugu		Texas (9)
Taiwan	Kaohsiung	Mexico	Piedras Negras
			Victoria
			Montemorelos

Tabachnick & Powell (1979) with their abbreviations and origins. Tables of gene frequencies presented here are for these new populations alone, with the exception of the three hexokinase loci (Table 3), the data for which were not presented in Tabachnick & Powell (1979). We have restricted our analysis as far as possible to

Table 2. *Collections of* Aedes aegypti *analysed electrophoretically since Tabachnick & Powell (1979)*

Sample name	No. of collections	Country	Location	Date collected	Subsp.	Habitat
BWE	1	Uganda	Bwerenga	8/80	f	Outdoors, urban
GAMBIA	1	Gambia	Diabugu	1/79	f	Outdoors, village
KED	4	Senegal	Kedougou	9/80	f	Outdoors, urban
DAKAR	4	Senegal	Dakar	9/80	f	Outdoors, urban
NGOYE	1	Senegal	N'goyé	9/81	f	Outdoors, village
AKO	1	Ghana	Akosombo	5/81	f	Outdoors, village
OUAGA	1	Upper Volta	Ouagadougou	7/81	f	Outdoors, village
TANP	1	Indonesia	Jakarta	1/80	a	Indoors, urban
FIJI	1	Fiji	Walu Bay	9/80	a	
VERO	1	U.S.	Vero Beach FL	7/79	a	
GULF	1	U.S.	Gulfport MS	9/81	a	
ABBE	1	U.S.	Abbeville LA	9/81	a	
DEQUIN	1	U.S.	De Quincy LA	9/81	a	
MOBILE	3	U.S.	Mobile AL	9/81	a	Outdoors, urban
BEAUTEX	2	U.S.	Beaumont TX	9/81	a	
GALTEX	1	U.S.	Galveston TX	11/80	a	
AUSTEX	1	U.S.	Austin TX	7/79	a	
SANTEX	1	U.S.	San Antonio TX	12/79	a	
CCTEX	4	U.S.	Corpus Christi TX	5/80	a	
LARTEX	3	U.S.	Laredo TX	1/80	a	
EAGTEX	1	U.S.	Eagle Pass TX	5/81	a	—
WESTEX	2	U.S.	Weslaco TX	9/80	a	—
HOUTEX	1	U.S.	Houston TX	9/81	a	—
MONMEX	1	Mexico	Montemorelos	5/81	a	—
VICMEX	2	Mexico	Victoria	8/80	a	—
AREC	1	Puerto Rico	Arecibo	10/79	a	Outdoors, urban
BAR	1	Puerto Rico	Barcelonata	10/79	a	Outdoors, urban
TRIN	1	Trinidad	Port of Spain	6/80	a	Outdoors, urban
SEALOTS	1	Trinidad	Port of Spain	6/81	a	Outdoors, urban
TRINCITY	1	Trinidad	Trincity	6/81	a	Outdoors, urban
FELICITY	1	Trinidad	Felicity	6/81	a	Outdoors, urban
PARA	2	Suriname	Paramaribo	5/81	a	Outdoors, urban
GTOWN	1	Guyana	Georgetown	10/81	a	—
GUAT	1	Guatemala	Escuintla	9/80	a	—

material collected directly from the field, but some colonized material of recent origin was included in the analysis. However, we have avoided making any generalizations or inferences from colony material alone.

Electrophoretic methods described elsewhere (Tabachnick & Powell, 1979) were used to score genetic variation at the following 11 enzyme loci in all collections: *Gpd*, *Mdh*, *Me*, *Idh-1*, *Idh-2*, *Pgd*, *Hk-2*, *Hk-3*, *Hk-4*, *Pgm* and *Pgi*, with the exception of seven earlier populations in which *Idh-1* was not successfully resolved.

D values were calculated by Nei (1972) and UPGMA tree construction was according to Sneath & Sokal (1973). Stepwise multivariate discriminant analysis was performed using a biomedical computer program package (BMDP7M: Dixon, 1981).

3. RESULTS

Tables 3–7 show allele frequencies at the eight most polymorphic loci (n = number of genes sampled). *Gpd* is largely monomorphic; only 12 of the 63 localities exhibit variation (5 East African *formosus*, 5 South-Eastern U.S., 1 West African, and

Table 3. *Allele frequencies at the hexokinase loci* (Hk-2, Hk-3, Hk-4)

		Hk-2				*Hk-3*			*Hk-4*			
Population	n	100	113	Null	Others	90	100	111	100	109	Null	Others
KBW	170	0·762	0·230	0·008	0	0	0·766	0·234	0·667	0·260	0·073	0
MAJ	226	0·905	0·031	0·064	0	0	0·981	0·019	0·357	0·064	0·579	0
MGN	108	0·855	0·075	0·070	0	0	0·921	0·079	0·626	0·124	0·250	0
KOM	122	0·951	0·008	0	0·041	0·041	0·951	0·008	0·909	0·008	0·042	0·041
SHH	148	0·865	0	0	0·135	0·135	0·865	0	0·865	0	0	0·135
KDZ	256	0·931	0·005	0·051	0·013	0·013	0·982	0·005	0·877	0·019	0·094	0·010
NYA	106	0·981	0	0·019	0	0	1·0	0	0·962	0	0·038	0
DAR	180	0·961	0·039	0	0	0	0·961	0·039	0·961	0·039	0	0
BWE	84	1·0	0	0	0	0	1·0	0	0·976	0·024	0	0
GAMBIA	118	0·966	0·017	0·017	0	0	1·0	0	0·898	0	0·102	0
KED	134	1·0	0	0	0	0	1·0	0	0·988	0	0·012	0
DAKAR	274	1·0	0	0	0	0	1·0	0	0·982	0	0·018	0
OUAGA	60	0·983	0·017	0	0	0	0·983	0·017	0·983	0·017	0	0
GROGOL	154	1·0	0	0	0	0	1·0	0	0·942	0	0·058	0
SEMAR	54	1·0	0	0	0	0	1·0	0	0·944	0	0·056	0
TANP	120	1·0	0	0	0	0	1·0	0	0·967	0	0·033	0
SANTEX	108	1·0	0	0	0	0	1·0	0	0·981	0	0·019	0
CCTEX	212	0·896	0·104	0	0	0	0·896	0·104	0·888	0·104	0·008	0
EAGTEX	116	1·0	0	0	0	0	1·0	0	0·974	0	0·026	0
MONMEX	114	1·0	0	0	0	0	0·868	0·132	0·965	0	0·035	0
CARAC	108	1·0	0	0	0	0	1·0	0	0·926	0	0·074	0
MOBAY	86	0·733	0	0·267	0	0	1·0	0	0·593	0	0·407	0
MAYA	150	1·0	0	0	0	0	1·0	0	0·987	0	0·013	0
SANJ	52	0·923	0·019	0·058	0	0	0·962	0·038	0·808	0·019	0·173	0

No variants were detected at any *Hk* locus in the other 39 localities.

Population abbreviations appearing in this table but absent from Table 2 are defined in Tabachnick & Powell (1979).

1 Mexican) and the frequency of Gpd^{100} is above 0·91 in all cases except ABBE. *Me* is variable in only six localities (4 East African, 1 West African and 1 Texan) with Me^{100} never below a frequency of 0·75. Data for *Idh-1* may be found in Wallis & Tabachnick (1982). This locus is largely monomorphic outside of Africa, with the exception of the north coast of South America.

Variation for hexokinase (Table 3) is fairly concordant across all three loci (*Hk-2*, *Hk-3*, *Hk-4*). Genic variability is mostly restricted to the East African populations, and is highest of all in the *aegypti* subspecies. Some variation, especially for *Hk-4*, is found in Asian and New World populations.

Tables 4–7 give gene frequencies at 5 loci for the 34 populations sampled since Tabachnick & Powell (1979); the descriptions given here refer to gene frequencies from all 63 localities.

Malate dehydrogenase is of some geographic diagnostic use. Mdh^{84} is essentially an *A. a. formosus* allele whereas Mdh^{120} is only found at low frequencies in this subspecies. West African populations exhibit lowest heterozygosities and United States populations very much higher ones.

Table 4. *Allele frequencies at the* Mdh *locus*

Population	*n*	84	100	120	Others
BWE	132	0·402	0·598	0	0
GAMBIA	122	0·090	0·910	0	0
KED	102	0·043	0·948	0·009	0
DAKAR	276	0·102	0·867	0·031	0
NGOYE	100	0·200	0·680	0·120	0
AKO	80	0	0·837	0·163	0
OUAGA	60	0·417	0·583	0	0
TANP	120	0·067	0·625	0·308	0
FIJI	202	0·059	0·292	0·599	0·050
VERO	120	0·159	0·683	0·158	0
GULF	42	0	0·167	0·809	0·024
ABBE	68	0	0·456	0·529	0·015
DEQUIN	46	0	0·609	0·391	0
MOBILE	164	0·118	0·646	0·236	0
BEAUTEX	130	0	0·484	0·516	0
GALTEX	240	0	0·313	0·687	0
AUSTEX	102	0	0·333	0·667	0
SANTEX	112	0	0·259	0·741	0
CCTEX	278	0	0·617	0·383	0
LARTEX	378	0·050	0·401	0·549	0
EAGTEX	116	0	0·422	0·578	0
WESTEX	158	0	0·677	0·323	0
HOUTEX	252	0	0·421	0·575	0·004
MONMEX	114	0	0·342	0·658	0
VICMEX	168	0	0·384	0·616	0
AREC	194	0·026	0·361	0·613	0
BAR	80	0·013	0·474	0·513	0
TRIN	60	0	0·233	0·767	0
SEALOTS	40	0	0·350	0·650	0
TRINCITY	124	0	0·653	0·347	0
FELICITY	22	0	0·773	0·227	0
PARA	196	0·011	0·580	0·409	0
GTOWN	88	0	0·989	0·011	0
GUAT	362	0	0·055	0·945	0

Isocitrate dehydrogenase 2 is essentially under diallelic control with extreme disparate allele frequencies between regions making it of great diagnostic use. $Idh\text{-}2^{116}$ is rare in all sylvan populations, but reaches high frequencies in domestic populations. *A. a. aegypti* from Texas, Central America and the Caribbean have highest $Idh\text{-}2^{116}$ frequencies, but populations in the south-eastern United States resemble Asian populations having lower frequencies of this allele.

Phosphoglucomutase again supports the Mattingly subspecies division; all

sylvan populations from West and East Africa are highly heterozygous whereas Pgm^{100} predominates in domestic collections.

Phosphogluconate dehydrogenase and phosphoglucose isomerase are generally weakly polymorphic. Exceptions are some East African *formosus* populations and

Table 5. *Allele frequencies at the* Idh-2 *locus*

Populations	*n*	87	100	116	125
BWE	126	0	0·937	0·063	0
GAMBIA	122	0	0·918	0·082	0
KED	116	0	0·966	0·034	0
DAKAR	274	0	0·962	0·038	0
NGOYE	100	0	0·930	0·070	0
AKO	80	0	0·975	0·025	0
OUAGA	60	0·033	0·967	0	0
TANP	120	0	0·875	0·125	0
FIJI	202	0	0·881	0·119	0
VERO	120	0	0·792	0·208	0
GULF	42	0	0·833	0·167	0
ABBE	70	0	0·786	0·214	0
DEQUIN	46	0	0·891	0·109	0
MOBILE	170	0	0·870	0·130	0
BEAUTEX	136	0	0·833	0·167	0
GALTEX	240	0	0·463	0·537	0
AUSTEX	102	0	0·686	0·314	0
SANTEX	112	0	0·500	0·500	0
CCTEX	278	0	0·528	0·462	0·010
LARTEX	308	0	0·396	0·604	0
EAGTEX	116	0	0·509	0·491	0
WESTEX	106	0	0·953	0·047	0
HOUTEX	254	0	0·343	0·657	0
MONMEX	114	0	0·491	0·509	0
VICMEX	168	0	0·714	0·286	0
AREC	194	0	0·485	0·515	0
BAR	80	0	0·587	0·413	0
TRIN	60	0	0·067	0·933	0
SEALOTS	40	0	0·325	0·675	0
TRINCITY	118	0	0·551	0·449	0
FELICITY	22	0	0·364	0·636	0
PARA	198	0	0·247	0·753	0
GTOWN	88	0	0·693	0·307	0
GUAT	342	0	0·591	0·409	0

the Caribbean where Pgd^{116} is common, and some New World populations where Pgi^{93} and Pgi^{105} sporadically reach higher frequencies.

Mean expected heterozygosity over 10 loci (excluding *Idh-1*) is highest in East African populations, and lowest in Asia (Table 8).

(i) *Genetic–geographic groupings*

From the gene frequency data for the 29 localities then available, Tabachnick & Powell (1979) described 4 geographic areas, and separated East African populations according to subspecies giving 5 groupings. With the addition of

collections from 5 more localities, Powell, Tabachnick & Arnold (1980) divided the New World group into 3: Caribbean, United States and South America. Using multivariate discriminant analysis it proved possible to delineate all 7 groups in two dimensions without overlap.

Since then, the number of localities from which data are available has approximately doubled, with most emphasis being placed on areas bordering the Gulf of Mexico. Nei's overall D values were calculated for every pairwise population comparison yielding a matrix of nearly 2000 data points. It became clear that the

Table 6. *Allele frequencies at the* Pgm *locus*

Population	n	80	100	120	Others
BWE	126	0·254	0·381	0·333	0·032
GAMBIA	118	0·195	0·763	0·025	0·017
KED	132	0·039	0·549	0·339	0·073
DAKAR	276	0·061	0·578	0·335	0·026
NGOYE	100	0·100	0·660	0·180	0·060
AKO	80	0·225	0·238	0·325	0·212
OUAGA	60	0·083	0·517	0·400	0
FIJI	202	0	0·886	0	0·114
VERO	120	0·158	0·809	0·033	0
GULF	42	0·024	0·952	0·024	0
ABBE	70	0·014	0·814	0·143	0·029
DEQUIN	46	0·109	0·804	0·022	0·065
MOBILE	52	0·115	0·807	0·039	0·039
GALTEX	240	0	0·846	0·154	0
AUSTEX	102	0	0·990	0·010	0
SANTEX	112	0	0·946	0·054	0
CCTEX	278	0·020	0·926	0·054	0
EAGTEX	116	0	0·819	0·181	0
HOUTEX	182	0	0·879	0·121	0
VICMEX	170	0·017	0·977	0·006	0
TRINCITY	124	0	0·871	0·129	0
FELCITY	22	0	0·864	0·136	0
PARA	196	0·011	0·989	0	0
GTOWN	88	0	0·989	0·011	0
GUAT	330	0·015	0·985	0	0

The following populations were fixed for Pgm^{100}: TANP, BEAUTEX, LARTEX, WESTEX, MONMEX, AREC, BAR, TRIN, and SEALOTS ($\bar{n} = 138{\cdot}2$)

United States may be divided into south-east and south-west groupings resulting in a total of 8 geographic groups: East Africa *aegypti* (EAA), East Africa *formosus* (EAF), West Africa (WA) Asia (ASIA), south-eastern United States (SEUS), south-western United States with Mexico (TEXMEX), south and central America with Trinidad (SCA) and the Caribbean (CAR). Table 9 presents a summarized version of the genetic distance matrix using these groupings, with intragroup distances on the diagonal. In doing this, three anomalous collections were omitted: WESTEX, TRIN, and GUAT. These have rather odd and extreme gene frequencies at one or more loci. In the case of GUAT, we have no neighbouring collections and it is therefore not known whether this is truly indicative of a separate genetic geographic region, sampling error, or a gene pool isolate whose aberrant frequencies

Table 7. *Allele frequencies at the* Pgd *and* Pgi *loci*

		Pgd					*Pgi*			
Population	*n*	86	100	116	130	Others	*n*	93	100	105
BWE	124	0·024	0·895	0·081	0	0	126	0	1·000	0
GAMBIA	122	0	0·984	0	0·008	0·008	122	0	1·000	0
KED	122	0	0·979	0	0	0·021	144	0	1·000	0
DAKAR	248	0	0·926	0	0	0·074	286	0	1·000	0
OUAGA	60	0	0·967	0·033	0	0	60	0	1·000	0
VERO	120	0·017	0·933	0·050	0	0	120	0·100	0·892	0·008
GULF	42	0·071	0·929	0	0	0	42	0	0·881	0·119
DEQUIN	46	0	1·000	0	0	0	46	0·217	0·783	0
MOBILE	170	0·029	0·803	0·168	0	0	170	0	0·962	0·038
BEAUTEX	132	0·063	0·937	0	0	0	136	0·010	0·990	0
GALTEX	240	0	0·996	0·004	0	0	240	0·013	0·987	0
AUSTEX	96	0·031	0·969	0	0	0	102	0·078	0·922	0
SANTEX	112	0	1·000	0	0	0	112	0·009	0·991	0
CCTEX	274	0	1·000	0	0	0	276	0·077	0·883	0·040
LARTEX	394	0	1·000	0	0	0	394	0·187	0·812	0·001
EAGTEX	116	0	1·000	0	0	0	116	0·052	0·948	0
WESTEX	94	0	1·000	0	0	0	100	0·057	0·943	0
HOUTEX	254	0	1·000	0	0	0	190	0·105	0·895	0
MONMEX	116	0	1·000	0	0	0	116	0·302	0·483	0·215
AREC	194	0	0·768	0·232	0	0	194	0·010	0·902	0·088
BAR	80	0	0·463	0·537	0	0	80	0·100	0·812	0·088
TRIN	60	0	1·000	0	0	0	60	0·250	0·750	0
TRINCITY	124	0	1·000	0	0	0	80	0·016	0·984	0
PARA	192	0	1·000	0	0	0	156	0·010	0·921	0·069
GUAT	184	0	1·000	0	0	0	188	0·293	0·707	0

The following populations are fixed for both Pgd^{100} and Pgi^{100}: NGOYE, AKO, TANP, FIJI, ABBE, VICMEX, SEALOTS, FELICITY and GTOWN ($\bar{n} = 90{\cdot}3$).

Table 8. *Mean expected heterozygosity values with standard errors for each genetic–geographic region based on 10* loci

(N = number of localities)

Group	N	$\bar{H}_e \pm$ S.E.
EAA	3	0·209 ± 0·014
EAF	6	0·185 ± 0·012
WA	15	0·107 ± 0·004
ASIA	6	0·081 ± 0·007
SEUS	8	0·137 ± 0·008
TEXMEX	11	0·130 ± 0·012
SCA	9	0·102 ± 0·010
CAR	5	0·179 ± 0·017
Overall	63	0·117 ± 0·005

result from founder effect and drift. WESTEX appears to be an example of such an isolate; two large collections directly from the field show the same high frequency of $Idh\text{-}2^{100}$ and low heterozygosity. In the case of TRIN, sampling appears to be at fault as several other collections from the island agree with the overall picture. We feel that omitting these three samples is valid as it represents

only a tiny fraction of the total data, and in such a large survey one must expect aberrant samples. Fig. 1 is an UPGMA dendrogram generated from the data in Table 9.

Intensive sampling of *A. aegypti* from the southern United States has revealed a distinct discontinuity between populations from the east (Florida, Alabama, Mississippi, Louisiana and Beaumont TX) and those to the west (Texas and

Table 9. *Nei's overall D values within (on diagonal) and between genetic-geographic groupings*

(n = number of pairwise population comparisons used.)

	WA	EAF	EAA	ASIA	SEUS	TEXMEX	SCA	CAR
WA	0·0095 ±0·0007 $n = 105$	—	—	—	—	—	—	—
EAF	0·0181 ±0·0010 $n = 90$	0·0185 ±0·0030 $n = 15$	—	—	—	—	—	—
EAA	0·0695 ±0·0030 $n = 45$	0·0624 ±0·0058 $n = 18$	0·0197 ±0·0082 $n = 3$	—	—	—	—	—
ASIA	0·0239 ±0·0012 $n = 90$	0·0193 ±0·0023 $n = 36$	0·0398 ±0·0033 $n = 18$	0·0113 ±0·0025 $n = 15$	—	—	—	—
SEUS	0·0357 ±0·0019 $n = 120$	0·0328 ±0·0033 $n = 48$	0·0487 ±0·0035 $n = 24$	0·0168 ±0·0020 $n = 48$	0·0157 ±0·0022 $n = 28$	—	—	—
TEXMEX	0·0759 ±0·0017 $n = 150$	0·0712 ±0·0033 $n = 60$	0·0458 ±0·0032 $n = 30$	0·0424 ±0·0022 $n = 60$	0·0284 ±0·0014 $n = 80$	0·0127 ±0·0012 $n = 45$	—	—
SCA	0·0559 ±0·0028 $n = 105$	0·0529 ±0·0048 $n = 42$	0·0371 ±0·0032 $n = 21$	0·0300 ±0·0032 $n = 42$	0·0301 ±0·0024 $n = 56$	0·0243 ±0·0020 $n = 70$	0·0208 ±0·0031 $n = 21$	—
CAR	0·0910 ±0·0020 $n = 75$	0·0762 ±0·0044 $n = 30$	0·0500 ±0·0052 $n = 15$	0·0536 ±0·0022 $n = 30$	0·0453 ±0·0023 $n = 40$	0·0332 ±0·0019 $n = 50$	0·0357 ±0·0026 $n = 35$	0·0370 ±0·0067 $n = 10$
	WA	EAF	EAA	ASIA	SEUS	TEXMEX	SCA	CAR

Mexico). This is most noticeable in terms of *Idh-2* allele frequencies (Fig. 2). Considering the 8 south-eastern US populations, the frequency of *Idh-2*^{116} reaches 0·33 in HAMF, but in all the other 7 is below 0·22. Contrarily, *Idh-2*^{116} exceeds 0·28 in all 10 populations to the west of Beaumont (excluding the aberrant WESTEX), and in 8 of these exceeds 0·46. Furthermore, despite the proximity of these populations, the TEXMEX group is more genetically similar with the SCA and CAR groups than it is to SEUS (Table 9). Thus, for example, GALTEX and HOUTEX are more genetically similar to VENEZ, PARA, SEALOTS and TRINCITY than they are to nearby BEAUTEX, DEQUIN and ABBE.

Whilst Nei's D value is perhaps the best objective measure of genetic distance

between populations, it is not necessarily quantitative distance *per se* that is most useful here. If we want to be able to define geographic–genetic groups of *Aedes aegypti*, then consistent qualitative differences between groups become important. Thus from a purely discriminatory viewpoint, one is interested in defining a group in the most productive manner, and a procedure is needed that takes account of

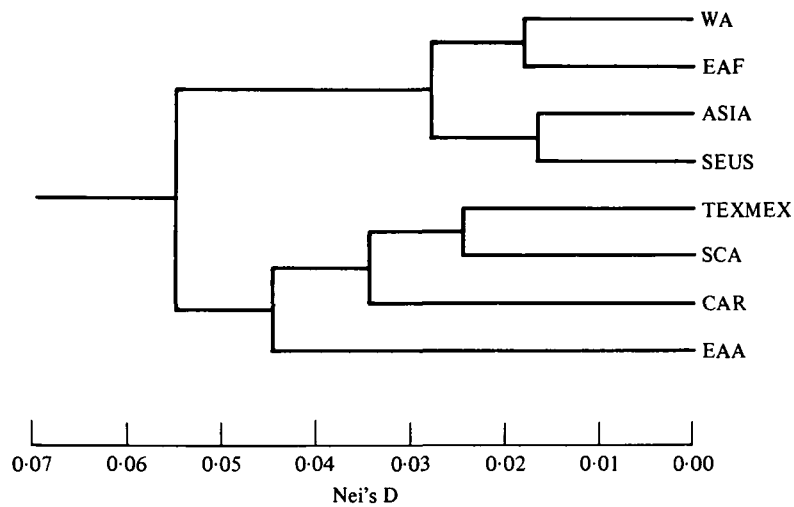

Fig. 1. UPGMA dendrogram generated from the genetic distance data given in Table 10.

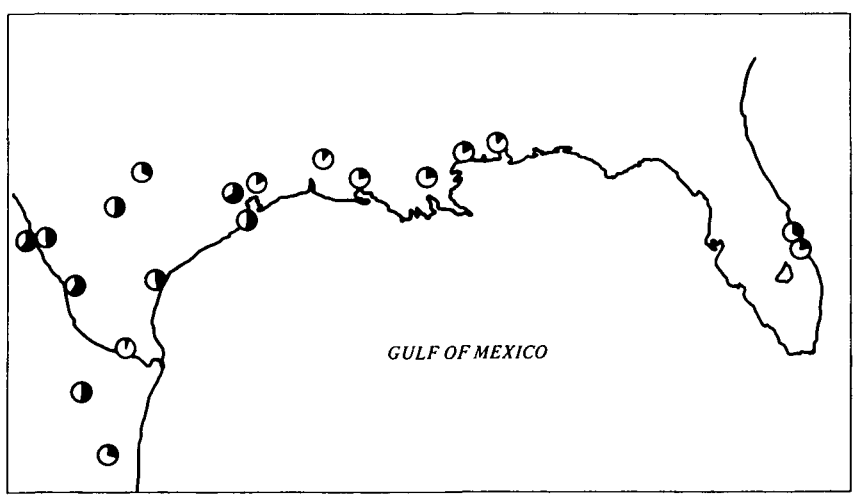

Fig. 2. *Idh-2*^{116} allele frequencies (shaded areas) in the southern U.S.

the useful consistent discriminatory differences in such a way that they are not swamped by larger less meaningful randomly distributed variation. Of course, it is best that discriminatory features are large as this allows assignment of a population of unknown origin to a group with a higher degree of certainty, but consistency has to be balanced against magnitude. We have already demonstrated

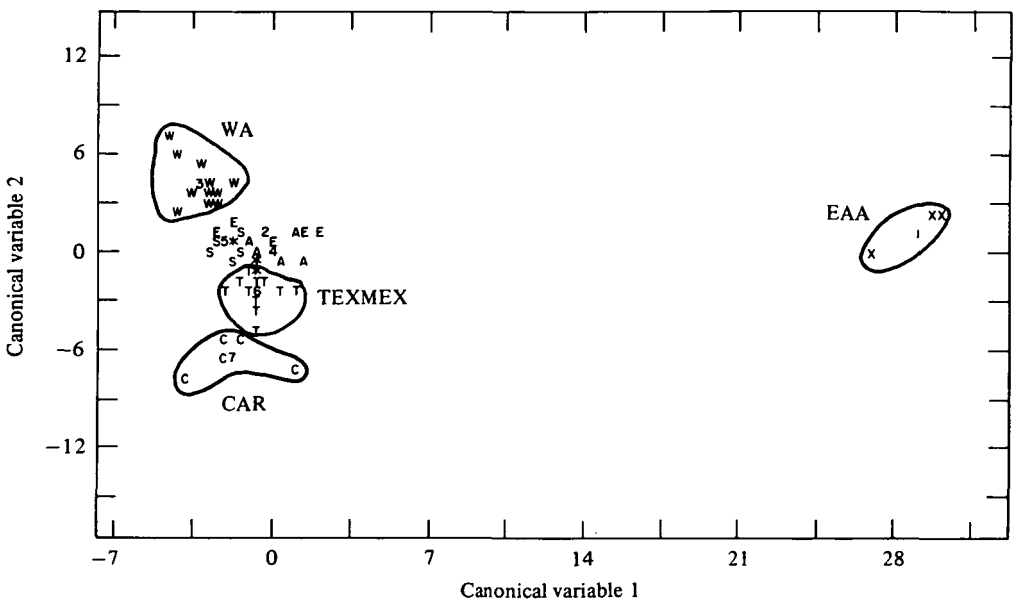

Fig. 3. Canonical plot of 62 *Aedes aegypti* populations based on 10 polymorphic loci. Each population is represented by the initial letter of its group (except *X* for EAA); overlap of populations from different groups is represented by *.

Table 10. *F-values and coefficients 1 and 2 (X and Y) for canonical variables in analysis of 59 localities (Fig. 4) and 52 localities (including* Idh-1) *(Fig. 5)*

Variable	Coefficient 1	Coefficient 2	F-value to enter or remove
	59 localities analysis		
Pgm^{100}	−8·860	−6·680	38·2005
Idh-2^{100}	5·398	−1·377	19·9160
Pgd^{100}	−16·200	−20·517	17·8895
Hk-2^{NULL}	−18·755	26·088	10·9124
Me^{100}	14·026	−0·852	8·1005
Pgd^{116}	−25·740	−4·343	6·9890
Gpd^{100}	−1·841	−5·993	5·2055
Constant	7·723	32·588	
	52 localities analysis		
Pgd^{116}	−23·915	5·498	42·008
Idh-2^{116}	−5·257	−0·099	38·104
Pgm^{100}	−9·061	−1·755	13·366
Idh-1^{160}	9·813	−0·595	8·308
Pgd^{100}	−15·971	−10·275	6·985
Me^{100}	8·695	−15·827	5·260
Idh-1^{80}	20·927	74·252	10·292
Constant	15·941	26·398	

https://doi.org/10.1017/S0016672300021315 Published online by Cambridge University Press

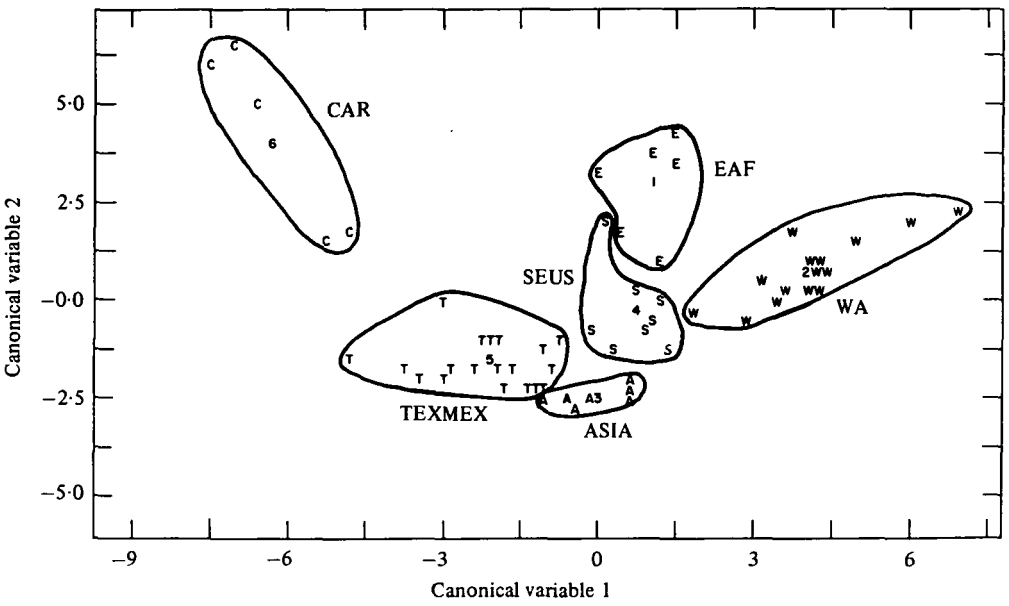

Fig. 4. Canonical plot of 59 *Aedes aegypti* populations based on 10 polymorphic loci.

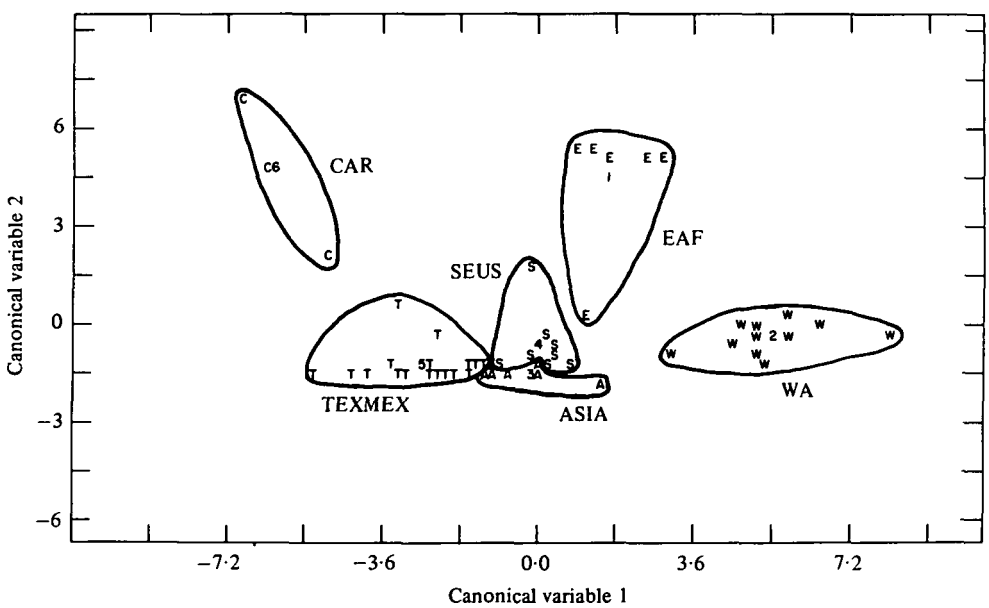

Fig. 5. Canonical plot of 52 *Aedes aegypti* populations based on 11 polymorphic loci.

https://doi.org/10.1017/S0016672300021315 Published online by Cambridge University Press

the use of stepwise multivariate discriminant analysis to this end (Powell *et al.* 1980; Powell *et al.* 1982). Using the most discriminatory characters, the technique provides several uncorrelated composite variables from which canonical coefficients are calculated. These are then applied to every locality's gene frequency set and a two-dimensional plot is made using the two canonical variables that account for the highest proportion of total dispersion.

When the gene frequency data from the 63 localities are analysed in this way, the situation demonstrated by Nei's D values and the dendrogram are essentially confirmed. This time only WESTEX is omitted; the very low frequency of *Idh-2*^{116} is an anomaly and disrupts the groupings.

Using the gene frequency data from these 62 localities, Fig. 3 shows the canonical plot. EAA is well separated from all the other groups, which form a mêlée of largely indistinguishable regions to the left with lower X values. The reason for the three East African *A. a. aegypti* samples forming such a distinct unit is their diagnostic *Hk* gene frequencies (Table 3). *Hk-4* is particularly variable and the high frequency of *Hk-4*^{109} is of great discriminatory value.

In order to discriminate between the other geographic regions, the analysis is re-run omitting the three EAA members. Fig. 4 is the resulting plot, showing fairly good separation of all groups. *Idh-1* was analysed in 52 of these 59 populations, and as it is of some discriminatory use, the analysis is run again for these populations including this locus, giving the plot in Fig. 5.

Table 10 presents the coefficients for canonical variables in the latter two analyses for 59 and 52 localities respectively. Alleles are listed in order of decreasing contributory importance. In these analyses, SCA was not upheld as a distinct group; this group is now merged with TEXMEX, and henceforth referred to as such.

4. DISCUSSION

The patterns of world-wide variation described in previous publications have been supported and enlarged upon here. Sylvan and domestic forms of *A. aegypti* from East Africa clearly represent distinct sympatric gene pools, with a mean genetic distance of $0{\cdot}0624 \pm 0{\cdot}0058$ between populations based upon 10 isozyme loci. There is little genetic differentiation between populations in West Africa (Table 9). There is three to four times as much differentiation between EAA and EAF, as well as EAA and WA, than there is between the sylvan groups EAF and WA. It is of interest to note that we have been unable to locate domestic type *A. a. aegypti* in West Africa. This presents a dilemma as most entomologists and epidemiologists believe that *A. a. aegypti* was introduced into the New World from West Africa via the slave trade commerce of the sixteenth–eighteenth centuries. Dispersal by human transport would seem more likely for the domestic subspecies than the sylvan form. It is possible that *A. a. aegypti* was once present in West Africa but has for some unknown reasons been replaced. We will discuss this and other alternatives in a forthcoming paper on the evolution of *A. aegypti*.

Populations from Asia and the south-eastern U.S.A. are most closely related to the sylvan forms (Fig. 1) with genetic distances of around 0·02–0·04. This similarity is to a large extent due to lower *Idh-2*^{116} frequencies. The high heterozygosity of

EAA populations and low heterozygosity in Asia is consistent with the likely history of the species which probably spread from East Africa to Asia. Apart from the SEUS group, New World populations have the greatest affinity with the domestic East African Group, forming the second cluster.

The canonical analysis confirms the existence of these groups as distinct genetic entities, and presents a means of 'classifying' new populations. Any single population can be removed and replotted using the resulting new coefficients to examine its effect on overall distortion of the group. For instance, if the Asian looking WESTEX is included, the increased within group variance disturbs the groupings. We know that WESTEX is a real genetic anomaly unlike any neighbouring localities and believe that this is an example of founder effect.

The most interesting new feature is the situation in the south-eastern U.S.A. This could be associated with the disappearance of and reinfestation by *A. aegypti* in areas around New Orleans over the last couple of decades (Trapido & Carmichael, 1974). An original founder effect coupled with the processes of drift and selection in the growing gene pool may be responsible for the genetic differences now observed. Aided by absence of competing populations, this SEUS type may have spread rapidly through Mississippi, Alabama and Florida to the east coast, and west through Louisiana. The zone of contact that now exists in eastern Texas may not be stable; if one population is spreading at the expense of the other, it is of use to know something of the epidemiological parameters of each. TEXMEX populations are genetically more similar to Caribbean populations than they are to ones from SEUS, and dengue fever has recently been epidemic in the Caribbean (PAHO, 1979). It would be of use to obtain some more inland U.S. populations to confirm and amplify our knowledge of the sudden discontinuity in eastern Texas.

Figure 4 demonstrates the distinctiveness of SEUS and TEXMEX, and how the former is rather similar to the EAF group. This sylvan element of SEUS is rather an enigma at present, but it is interesting to note that there are reports of New Orleans *A. aegypti* using tree holes as breeding sites, though this may merely be due to population overflow. The point on the plot that is closest to the EAF group is a composite of three collections from Mobile, Albama which have unusually high Pgd^{116} frequencies. When *Idh-1* is included, the picture changes little, but it becomes easier to differentiate between EAF and SEUS, as the former has *Idh-1* variants not commonly found in any US population. The EAF population with the low canonical variable 2 is NYALI, collected near Mombasa. This displacement is a result of low heterozygosity at *Pgd* and *Idh-1*, which may indicate migration between the port of Mombasa and Asia by sea.

So at present we recognise seven distinct genetic-geographic groupings of *A. aegypti*. The New World is divided into Caribbean (Puerto Rico and Jamaica), south-eastern U.S. (from Beaumont TX east to Vero Beach Florida) and the remaining regions bordering the Gulf of Mexico, the TEXMEX group (from Galveston TX west, Mexico, Guatemala, Venezuela, Guyana, Suriname, Trinidad). Some interesting epidemiological possibilities are suggested by these results. It has been known for some time that epidemic yellow fever has never existed on the Asian continent (Dudley, 1934), although dengue has been found there (Smith, 1956; Strode, 1951). Secondly, it is interesting that the southern United States has not

suffered dengue fever for 40 years or yellow fever since 1905, although the Caribbean continues to experience dengue, and in the case of Trinidad, yellow fever. Sporadic dengue cases have been reported in Mexico and south-west Texas in 1980. It is interesting that populations of mosquitoes in Asia and the south-eastern U.S. have a high genetic affinity with regard to isozymes. This is consistent with the hypothesis that isozyme-based genetic–geographic groups share other genetic features in common, vectorial efficiency perhaps being one.

Heterogeneity within some geographic regions may reflect the recent history of *Aedes aegypti* populations with regard to their association with humans. Compared to other well-studied commensals like the house mouse, *A. aegypti* has been more intensively subjected to human efforts for control and eradication. Insecticide control programmes impose intense selection for increased resistance to the compound(s), and will concomitantly alter the genetic constitution of the local population due to hitch-hiking and stochastic loss of variability. Eradication over large areas may ensue, leaving isolated 'island' populations. When control is relaxed, reinfestation by remaining isolates and invading migrants creates a new gene pool. The SEUS group may have arisen in such a manner.

An intuitive notion that has prevailed in population genetics for many years is the swamping effect of gene flow upon potentially genetically divergent populations (e.g. Kimura & Maruyama, 1971). From this viewpoint it may seem surprising that we are able to discern clear genetic–geographic regions, as levels of gene flow are generally thought to be high for prolific commensals. Our results would therefore seem to support circumstantially those who de-emphasize the role of gene flow as a mechanism for disrupting geographic differentiation (e.g. Endler, 1973). Alternatively, the notion that association with humans necessarily leads to high rates of gene flow among populations may be wrong.

We wish to acknowledge the help of the following entomologists who made this work possible by their generous provision of egg samples and/or valuable information: T. H. G. Aitken, R. Bartnette, R. Bentley, D. A. Boakye, R. G. Burton, M. K. Carroll, D. M. Chambers, M. Cornet, M. Dakin, J. B. Davies, R. L. Davis, J. D. Dickens, D. Espinosa, G. Faget, D. B. Francey, J. E. Hudson, E. A. Hughes, B. Hull, L. P. Lounibos, D. W. Micks, B. R. Miller, R. Novak, G. F. O'Meara, D. J. Pletsch, I. M. Rakai, C. Reed, S. S. Riché, D. Sykes, L. Terracina, E. S. Tikasingh, T. Villarreal, R. J. Wood, and M. Zuñiga.

We have also benefited greatly from continued communication with many of the workers acknowledged in Tabachnick & Powell (1979) and would like to thank them once again.

This work was funded by U.S. Army Contract DAMD-81-C-1026 and NIH grant no. AI 1132-10.

REFERENCES

Aitken, T. H. G., Downs, W. G. & Shope, R. E. (1977). *Aedes aegypti* strain fitness for yellow fever virus transmission. *American Journal of Tropical Medicine and Hygiene* **26**, 985–989.

Beaty, B. & Aitken, T. H. G. (1979). *In vitro* transmission of yellow fever virus by geographic strains of *Aedes aegypti*. *Mosquito News* **39**, 232–238.

Berry, R. J. (1977). *Inheritance and Natural History*. London: Collins.

Craig, G. B. & Hickey, W. A. (1966). Genetic variability in *Aedes aegypti* (Diptera: Culicidae). IV. Mutation load in some African populations. *Annals of the Entomological Society of America* **59**, 1228–1234.

Dixon, W. J. (ed.). (1981). *BMDP Statistical Software*. London: University of California Press.

Dudley, S. F. (1934). Can yellow fever spread into Asia? *Journal of Tropical Medicine and Hygiene* **37**, 273–278.

ENDLER, J. A. (1973). Gene flow and population differentiation. *Science* **179**, 243–250.

KIMURA, M. & MARUYAMA, T. (1971). Pattern of a neutral polymorphism in a geographically structured population. *Genetical Research* **18**, 125–131.

GUBLER, D. J., NOVAK, R. & MITCHELL, C. J. (1982). Arthropod vector competence – epidemiological, genetic and biological considerations. In *Recent Developments in the Genetics of Insect Disease Vectors*, A Symposium Proceedings (ed. W. W. M. Steiner, W. J. Tabachnick, K. S. Rai and S. Narang), pp. 343–378. Champaign, IL.: Stipes.

HUNT, W. G. & SELANDER, R. K. (1973). Biochemical genetics of hybridization in European house-mice. *Heredity* **31**, 11–33.

LEAHY, M. G., VANDEHEY, R. C. & BOOTH, K. S. (1978). Differential response to oviposition site by feral and domestic populations of *Aedes aegypti* (L.) (Diptera:Culicidae). *Bulletin of Entomological Research* **68**, 455–463.

MACHADO-ALLISON, C. E. & CRAIG, G. B. (1972). Geographic variation in resistance to desiccation in *Aedes aegypti* and *A. atropalpus* (Diptera, Culicidae). *Annals of the Entomological Society of America* **65**, 542–547.

MATTINGLY, P. F. (1957). Genetical aspects of the *Aedes aegypti* problem. I. Taxonomy and bionomics. *Annals of Tropical Medicine and Parasitology* **51**, 392–408.

MCCLELLAND, G. A. H. (1974). A worldwide survey of variation in scale pattern of the abdominal tergum of *Aedes aegypti* (L.) (Diptera:Culicidae). *Transactions of the Royal Entomological Society of London* **126**, 239–359.

MUNSTERMANN, L. E. & CRAIG, G. B. (1979). Genetics of *Aedes aegypti*. Updating the linkage map. *Journal of Heredity* **70**, 291–296.

NEI, M. (1972). Genetic distance between populations. *American Naturalist* **106**, 283–292.

PAN-AMERICAN HEALTH ORGANIZATION (1979). *Dengue in the Caribbean, 1977*. PAHO Scientific Publication, no. 375. Washington, D.C.

PETERSEN, J. L. (1977). Behavioral differences in two subspecies of *Aedes aegypti* (L.) (Diptera: Culicidae) in East Africa. Ph.D. Thesis, University of Notre Dame.

POWELL, J. R., TABACHNICK, W. J. & WALLIS, G. P. (1982). *Aedes aegypti* as a model of the usefulness of population genetics in studies of insect disease vectors. In *Recent Developments in the Genetics of Insect Disease Vectors*, A Symposium Procedings (ed. W. W. M. Steiner, W. J. Tabachnick, K. S. Rai and S. Narang), pp. 396–412. Champaign, IL: Stipes.

POWELL, J. R., TABACHNICK, W. J. & ARNOLD, J. (1980). Genetics and the origin of a vector population: *Aedes aegypti*, a case study. *Science* **208**, 1385–1387.

SELANDER, R. K., HUNT, W. G. & YANG, S. Y. (1969). Protein polymorphism and genetic heterozygosity in two European subspecies of the house-mouse. *Evolution* **23**, 379–390.

SELANDER, R. K. & YANG, S. Y. (1969). Protein polymorphism and genetic heterozygosity in a wild population of the house-mouse (*Mus musculus*). *Genetics* **63**, 653–667.

SINGH, R. S., HICKEY, D. A. & DAVID, J. (1982). Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. *Genetics* **101**, 235–256.

SMITH, C. E. G. (1956). The history of dengue in tropical Asia and its probable relationship to the mosquito *Aedes aegypti*. *Journal of Tropical Medicine and Hygiene* **59**, 3–11.

SNEATH, P. H. A. & SOKAL, R. R. (1973). *Numerical Taxonomy*. San Francisco: W. H. Freeman.

STRODE, G. K. (1951). *Yellow Fever*. New York: McGraw-Hill.

TABACHNICK, W. J. (1982). Geographic and temporal patterns of genetic variation of *Aedes aegypti* in New Orleans. *American Journal of Tropical Medicine and Hygiene* **31**, 849–853.

TABACHNICK, W. J., MUNSTERMANN, L. E. & POWELL, J. R. (1979). Genetic distinctness of sympatric forms of *Aedes aegypti* in East Africa. *Evolution* **33**, 287–295.

TABACHNICK, W. J. & POWELL, J. R. (1978). Genetic stucture of the East African domestic populations of *Aedes aegypti*. *Nature* **272**, 535–537.

TABACHNICK, W. J. & POWELL, J. R. (1979). A world-wide survey of genetic variation in the yellow fever mosquito, *Aedes aegypti*. *Genetical Research* **34**, 215–229.

TABACHNICK, W. J., AITKEN, T. H. G., BEATY, B. J., MILLER, B. R., POWELL, J. R. & WALLIS, G. P. (1982). Genetic approaches to the study of vector competency in *Aedes aegypti*. In *Recent Developments in the Genetics of Insect Disease Vectors*, A Symposium Proceedings (ed. W. W. M. Steiner, W. J. Tabachnick, K. S. Rai and S. Narang), pp. 413–432. Champaign, IL: Stipes.

TRAPIDO, H. & CARMICHAEL, G. T. (1974). Reappearance of *Aedes aegypti* in New Orleans, Louisiana. *Dengue Newsletter for the Americas* **3** (3), 7.

TRPIS, M. & HAUSERMANN, W. (1975). Demonstration of differential domesticity of *Aedes aegypti* (L.) (Diptera: Culicidae) in Africa by mark-release-recapture. *Bulletin of Entomological Research* **65**, 199–208.

TRPIS, M. & HAUSERMANN, W. (1978). Genetics of house entering behavior in East African populations of *Aedes aegypti* (L.) (Diptera: Culicidae) and its relevance to speciation. *Bulletin of Entomological Research* **68**, 521–532.

VANDEHEY, R. C., LEAHY, M. G. & BOOTH, K. S. (1978). Analysis of colour variations in feral, peridomestic and domestic populations of *Aedes aegypti* (L.) (Diptera: Culicidae). *Bulletin of Entomological Research* **68**, 443–453.

WALLIS, G. P. & TABACHNICK, W. J. (1982). Linkage of an isocitrate dehydrogenase locus with sex and two lethals in *Aedes aegypti*. *Journal of Heredity* **73**, 291–294.