Journal of Heredity 2005:96(5):502–512
doi:10.1093/jhered/esi093
Advance Access publication August 31, 2005

Geographic Differentiation in the House Fly Estimated by Microsatellite and Mitochondrial Variation

E. S. Krafsur, M. A. Cummings, M. A. Endsley, J. G. Marquez, and J. D. Nason

From the Departments of Entomology (Krafsur, Cummings, Endsley, and Marquez) and Ecology, Evolution, and Organismal Biology (Nason), Iowa State University, Ames, IA 50011.

Address correspondence to Dr. E. S. Krafsur, 403 Science 2, Iowa State University, Ames, IA 50011-3222, or e-mail: ekrafsur@iastate.edu.

Abstract

Gene flow over very large geographic scales has been investigated in few species. Examples include *Drosophila melanogaster*, *Drosophila subobscura*, *Drosophila simulans*, and the Mediterranean fruit fly (*Ceratitis capitata*). The cosmopolitan house fly, a highly vagile, fecund, colonizing species offers an additional exemplar. Genotypes at seven microsatellite loci were scored in 14 widely separated natural house fly populations from the Nearctic, neotropics, Afrotropics, Palearctic, and Asia. Allelic diversities and heterozygosities differed significantly among populations. Averaged over all populations, Weir and Cockerham's $\theta = 0.13$ and $R_{ST} = 0.20$. Pairwise genetic distance measures were uncorrelated with geographic distance. Microsatellite frequencies were compared with mitochondrial data from 13 of the same populations in which $\theta = 0.35$ and Nei's $G_{ST} = 0.72$. Mitochondrial variation indicated up to threefold greater indices of genetic differentiation than the microsatellites. We were unable to draw any biogeographical inferences from these results or from tree or network topologies constructed from the genetic data. It is likely that high microsatellite diversities, mutation rates, and homoplasy greatly compromised their usefulness in estimating gene flow. House fly colonization dynamics include a large number of primary and secondary colonizations coupled with substantial genetic drift, but no detectable bottlenecks.

Gene flow in colonizing species has not often been investigated over very large geographic scales. Perhaps the best examples are *Drosophila melanogaster* (Diptera: Drosophilidae), samples of which were derived from 114 cultured isofemale lines in which much variation was lost because of drift (Hale and Singh 1991); *Drosophila subobscura*, which only recently colonized the New World (e.g., Pascual et al. 2001); and *Ceratitis capitata* (Diptera: Tephritidae), the Mediterranean fruit fly (Gasperi et al. 2002; Malacrida et al. 1998), also known as the Med fly. Originating in Africa, *D. melanogaster* has become a commensal of mankind and virtually cosmopolitan. African populations were said to have greater diversities than New World populations, but recent work has shown that autosomal diversities are about the same and that bottlenecks in the colonization of the New World were unlikely; more complex explanations have been offered (Andolfatto 2001). Med flies are said to have originated in East Africa and their paths of primary and secondary colonization have been well documented because they are economically serious pests of commercial fruit. The genetic diversity of med fly colonies in the New World, the Mediterranean, and Asia tends to be much less than sylvan populations in Africa.

The common house fly, a highly vagile, fecund, colonizing species offers an additional exemplar. House flies (*Musca domestica* L. (Diptera: Muscidae)) are synanthropic and a public health threat because of feeding habits that include feces, vomitus, and decaying animal and vegetable matter, in addition to foodstuffs in the kitchen and on the dinner plate (Greenberg 1973; West 1951). House flies are interesting because they colonize habitats wherever mankind and domestic animals occur. These insects are highly fecund, have short life cycles, and demonstrate boom and bust dynamics (Black and Krafsur 1987; Krafsur 1985). In nature, house flies can readily establish colonies from the progeny of a single pair mating. *M. domestica* neither aestivates nor diapauses, and cannot tolerate subfreezing temperatures for long (Rosales et al. 1994), so must overwinter in restricted habitats by slow, continuous reproduction (Black and Krafsur 1986a; Krafsur 1985) which constitute annual population bottlenecks in which genetic variation is lost by drift (Black and Krafsur 1986b). Thus the longer the winter, the greater the expected effects of genetic drift, a hypothesis that predicts a progressive decline in mitochondrial diversity and genetic heterozygosity at higher latitudes. The hypothesis, however, was not confirmed by

data from samples in northern reaches of the Nearctic and Palearctic (Marquez and Krafsur 2003). On the other hand, house flies are highly vagile (Oldroyd 1964), as shown by mark, release, and recapture trials (Bishopp 1916; Quarterman et al. 1954), and they readily "hitchhike" in aircraft and shipping. Indeed, they are among the most common insects found in aircraft (Welch 1939; West 1951). Thus, over the long-term, forces of migration would seem likely to strongly oppose the forces of drift and local regimes of natural selection. The relative magnitudes of migration and drift can be estimated for selectively neutral loci that show negligible mutation rates.

House fly genetic diversities are high, no doubt because aggregate population densities are very large. Thirty-nine of 73 allozyme loci were polymorphic and mean diversities (Nei 1987) over all loci were 0.18, indicative of historically large populations (Krafsur et al. 1992). Mitochondrial diversities, estimated by single-strand conformational polymorphisms, were 0.26 ± 0.07, and a survey of 111 worldwide populations showed a substantial genetic differentiation (Marquez and Krafsur 2002). Because of their reduced copy number and uniparental transmission, mitochondrial genes are more sensitive than nuclear genes to demographic dynamics, and distribution patterns represent female dispersions (Avise 2004).

The development of microsatellite loci in house flies (Endsley et al. 2002) allows an independent view of genetic differentiation and gene flow among populations. In a preliminary examination of two widely separated populations, 14 of 19 dinucleotide loci and 1 of 6 trinucleotide loci were polymorphic. A total of 127 alleles were detected, 39 (31%) of which were private. The number of alleles per locus varied from 5 to 12, with a mean of 7.6 ± 0.6 in Cairo, Egypt, and 6.6 ± 0.6 in Palau, a small island in the western Pacific. Average heterozygosities were 0.76 ± 0.03 in the Cairo sample and 0.68 ± 0.05 in Palau. Yet Weir and Cockerham's (1984) θ was only 0.03 ± 0.01 averaged over 15 loci, an unexpectedly small level of differentiation that might suggest house flies consist of a weakly differentiated metapopulation.

Here we examine 14 house fly populations at seven microsatellite loci and compare the results with data on mitochondrial variation in 13 of the same populations. Nine Old World and five New World populations were examined. House flies are said to have originated in the southern Palearctic (Skidmore 1985) and were introduced to the New World in colonial times (Legner and McCoy 1966). But much earlier introductions, long before Europeans arrived, cannot be ruled out. Widely geographically diverse populations were chosen in an effort to uncover any underlying biogeographic patterns that may exist.

Materials and Methods

Sampling

House flies were sampled by colleagues as described (Marquez and Krafsur 2002). The flies were rapidly killed and preserved in ethanol. DNA was extracted by using a CTAB method (Shahjahan et al. 1995).

Microsatellites and Primers

The microsatellite loci were $(CA)_n$. The primers and GenBank accessions are set forth in Endsley et al. (2002).

Amplification of Loci and Demonstration of Genotypes and Mitochondrial Haplotypes

Amplifications of microsatellite loci were 10 μl reactions of 1× Biolase polymerase chain reaction (PCR) buffer, 1.5 mM $MgCl_2$, 0.4 mM dNTPs, 1 μM each primer, and 0.25 U Biolase polymerase (Bioline USA, Inc., Springfield NJ). The amplification profile was 94°C for 2 min and 35 cycles at 94°C for 30 s, 52°C for 30 s, 72°C for 30 s, and 72°C for 2 min. PCR products were diluted 1:5 with loading buffer (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol) and denatured at 95°C for 5 min before electrophoresis on 5% denaturing polyacrylamide gels at 50°C and 100 W. Fixation and silver staining followed the methods of Black and DuTeau (1997). Migration of each band was measured from the gel origin and each allele was assigned an index based on its migration relative to a standard band in a 20 bp ladder. Further details are offered in Krafsur and Endsley (2002). Genotypes in some populations were assigned using and ABI Genotyper and Genescan. These results were used to convert migration indices to base pair units.

Mitochondrial variation was assessed by examining single-strand conformational polymorphisms on silver-stained polyacrylamide gels (Black and DuTeau 1997) according to methods already set forth (Marquez and Krafsur 2002). The oligonucleotide primers used in PCR amplification of mitochondrial loci were N1-J-12585 (5′-GGT CCC TTA CGA ATT TGA ATA TAT CCT-3′) and LR-N-12866 (5′-ACA TGA TCT GAG TTC AAA CCG G-3′). They were used to amplify a 300 bp fragment of the 16S ribosomal RNA (*16S*) gene. Primers C2-J-3279 (5′-GGT CAA ACA ATT GAG TCT ATT TGA AC-3′) and C2-N-3494 (5′-GGT AAA ACT ACT CGA TTA TCA AC-3′) were used for amplification of a 214 bp fragment of the cytochrome oxidase II (*COII*) gene (Simon et al. 1994). The rationale for using these two loci was explained in Marquez and Krafsur (2002).

Genetic Analyses

Microsatellite genotypes were recorded in Microsatellite Toolkit version 3.0 (Park 2001) and output files were created for Arlequin (Schneider et al. 2000), FSTAT version 2.9.3 (Goudet 2001), and GENEPOP version 3.1c (Raymond and Rousset 1995). Linkage disequilibrium was examined by using GENEPOP. Agreement of genotypic frequencies with Hardy-Weinberg equilibrium in populations was tested by using Fisher's test in GENEPOP version 3.3 (Raymond and Rousset 1995). F_{ST} (Slatkin and Barton 1989; Wright 1978), G_{ST} (Nei 1987), and θ (Weir and Cockerham 1984)

measure departures from random mating among populations, but employ different assumptions in doing so. G_{ST}, like F_{ST}, is a ratio of heterozygosities, but it is based on identities, not expectations and probabilities. It can be considered as the mean F_{ST} weighted over all alleles. R_{ST} was estimated from the number of base pairs amplified at each allele by weighted analysis of variance (Michalakis and Excoffier 1996). The average relatedness of flies in samples vis-à-vis the relatedness of individuals in the combined samples was calculated by using FSTAT.

BOTTLENECK software (Piry et al. 1999) was used to see if heterozygosity (H_e) was greater than expected at mutation-drift equilibrium. Such an excess occurs after a bottleneck because loss of alleles is greater under the infinite allele mutation model than heterozygosity (Luikart et al. 1998). We used the two-phased mutation model with 95% single-step mutations and 5% multistep mutations, as recommended by Piry et al (1999). The Wilcoxon test was used to estimate the probability of heterozygosity excess.

Nei's standard genetic distances were used to show relationships among populations by using the unweighted pair group method using arithmetic averages (UPGMA) method. The standard genetic distance $D = -\log_e J_{XY}/(J_X J_Y)^{1/2}$, where J_X and J_Y are the gene identities over all loci and J_{XY} is the mean shared identity (Nei 1987:220). Isolation by distance was tested by regressing population pairwise estimates ($n = 91$) of G_{ST}, F_{ST}, and $F_{ST}/1 - F_{ST}$ (Rousset 1997) on the logarithm of geographic distances between them.

Relationships among populations were also quantified using population graphs, a multivariate graph theoretic approach (Dyer and Nason 2004). This approach does not assume a bifurcating tree (as does UPGMA), but instead represents population relationships in the form of a network topology. In the graph, the n populations are represented by n nodes, with node size reflecting the within component of microsatellite variation. The $n(n-1)/2$ edges between nodes are represented by edges with edge lengths reflecting the among component of microsatellite variation attributable to the connecting nodes. A final graph topology was obtained by determining the smallest edge set that sufficiently describes the among-population covariance structure. Procedures for edge deletion are described in Dyer and Nason (2004) and software for graph construction and visualization is available online (http://dyerlab.bio.vcu.edu/dloads/PopGraph.pdf). Restriction of gene flow between Old and New World populations should be reflected in the topology of the population graph, specifically by a deficit of edges connecting these two sets of nodes. This hypothesis was tested using the binomial approach described in Dyer and Nason (2004). We also tested for isolation by distance by regressing pairwise graph distances on untransformed and logarithm geographic distances between populations, with significance determined by Mantel test.

SAS version 8.2 (SAS Institute, Cary, NC) was used for analysis of variance (ANOVA), Wilcoxon ranked sum scores, Nei statistics, and Kruskal-Wallis tests. The results of the three tests closely agreed and only the ANOVAs are reported here. Mitochondrial variation was analyzed using the methods of Nei (1987) and SAS was used to perform the computations.

Results

Genotypic Disequilibrium

Seven loci afforded 21 pairwise comparisons of independence, 20 of which were consistent with hypotheses of independence ($P > .22$ in each comparison). Genotypes at *MdCA119* and *MdCA170*, however, were not independent ($P \ll .001$).

Microsatellite Diversities

Allele distributions, summed over all populations, were unimodal with long tails (Figure 1a–g). Differences between the shortest and longest alleles, averaged over loci, were 39.6 ± 7.3 bp. *MdCA104* and *MdCA121* showed high frequencies of alleles that differed by single base pairs. These are compound satellites that evidently consist of many insertions and deletions in the flanking regions (Colson and Goldstein 1999) in addition to CA repeats. A total of 135 alleles was detected among the seven loci (mean 19.35 ± 4.86 alleles per locus), 13 of which were singular (9.6%). There were 20 private alleles (16.4% of the total, not including the singulars). Mean private allele frequency was 0.065, ignoring singular alleles. About 24% of allelic diversity was unshared among two or more populations. Averaged over loci, population allelic diversities varied from 6.29 to 9.71, with an overall mean 7.94 (Table 1).

Expected and observed heterozygosities (Table 1) were 0.86 and 0.70, respectively, when averaged over loci and populations. The overall departure from random matings indicated by the data is $F_{IT} = 1 - H_o/H_e = 0.19$. Departures from random mating, indicated by F_{IS}, were significant in 7 of the 14 populations sampled. ANOVA indicated that heterozygosities ($F = 1.33$, df = 13, 84; $P \sim .21$) and allele diversities ($F = 1.26$, df = 13, 84; $P \sim .26$) did not differ significantly among the 14 populations.

Diversities in house flies sampled from the Palearctic and Ethiopian geographical regions were compared with those from presumptively colonized regions. Expected heterozygosities were 0.785 in the presumptively ancestral territories and 0.736 in the putatively colonized territories of the Nearctic, neotropical, and Australian regions ($F = 4.57$, df = 1, 13; $P \sim .054$). The corresponding allelic diversities were 8.52 and 7.38 alleles per locus, respectively, and differed significantly ($F = 4.78$, df =1, 13; $P \sim .049$).

Mitochondrial Diversities

Eighteen single-stranded conformational polymorphism (SSCP) haplotypes were recorded. Five of 13 populations were monomorphic (Canada, Kazakhstan, Korea, Palau, and Russia). Mean population diversity, H_S, was 0.232 ± 0.063 and varied in samples from 0 to 0.63. Diversity was greatest in the Afrotropical region (Table 2). The mean probability that two random flies in a population shared the same

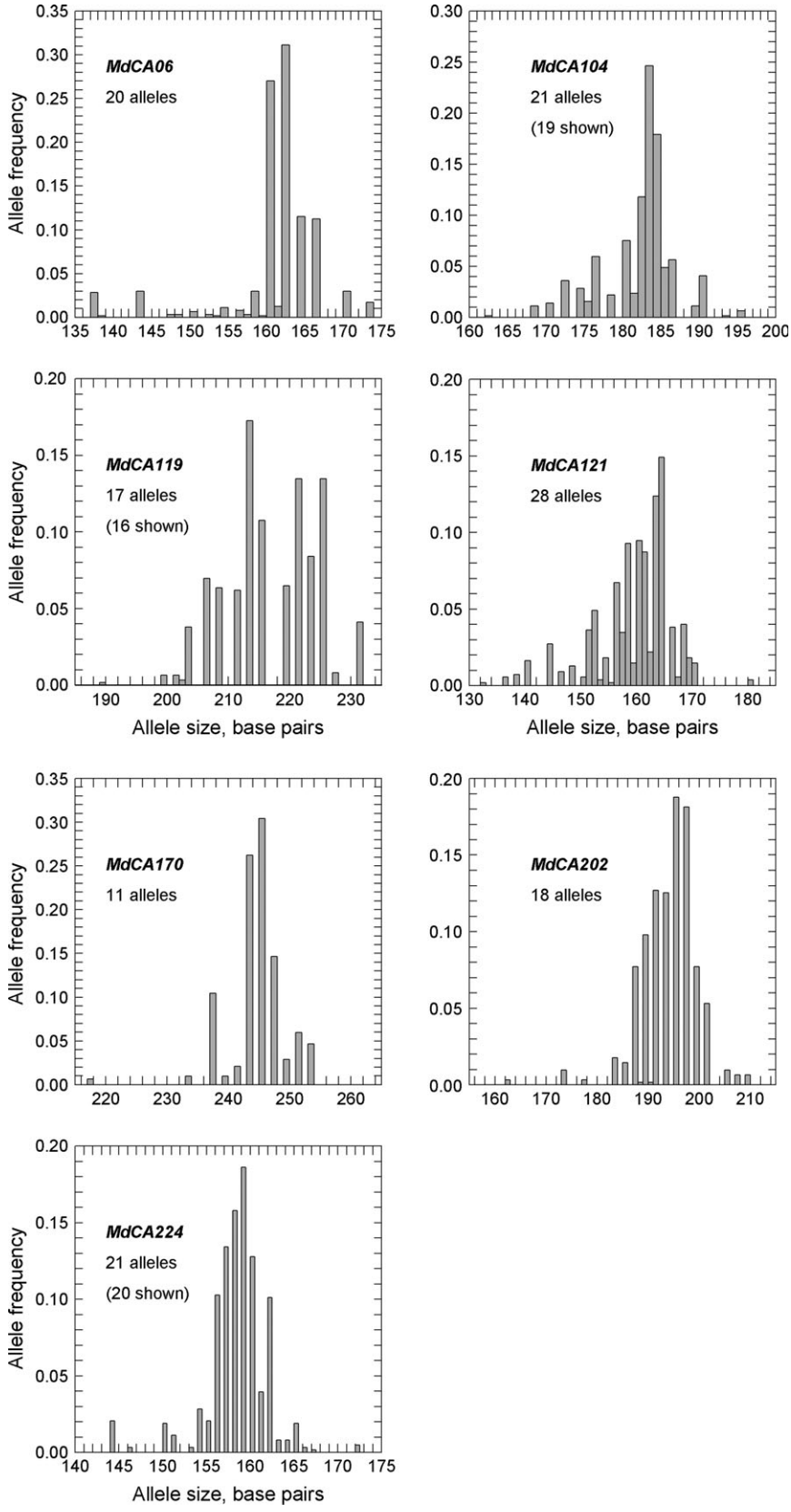

Figure 1. **(a–g)** Frequency distributions of allele sizes at CA loci in *M. domestica*.

Table 1. Genetic statistics for house fly populations averaged over seven microsatellite loci[a]

Populations	Heterozygosities				F_{IS}	Mean no.	
	Expected	SD	Observed	SD		Alleles	SD
Cairo, Egypt	0.816	0.026	0.739	0.036	0.097**	8.86	1.95
Dakar, Senegal	0.816	0.017	0.763	0.034	0.066*	9.71	3.15
Harare, Zimbabwe	0.730	0.029	0.713	0.036	0.025	7.29	1.11
Andong, S. Korea	0.803	0.024	0.776	0.034	0.034***	9.29	3.15
Chiang Mai, Thailand	0.800	0.044	0.733	0.035	0.085**	8.14	2.04
Almaty, Kazakhstan	0.783	0.026	0.770	0.034	0.018	8.14	2.41
Krapotkin, Russia	0.752	0.049	0.584	0.041	0.228***	9.43	2.99
Seibersdorf, Austria	0.746	0.048	0.743	0.036	0.005	6.71	2.14
Davis, California	0.647	0.063	0.617	0.038	0.047	6.86	1.46
Lethbridge, Canada	0.669	0.056	0.507	0.044	0.248**	7.71	3.20
Tippecanoe, Indiana	0.774	0.040	0.699	0.036	0.10	7.29	2.14
Honduras	0.798	0.041	0.695	0.039	0.131***	8.43	2.99
El Manantiol, Uruguay	0.722	0.057	0.709	0.035	0.019	6.29	2.69
Koror, Palau	0.770	0.030	0.742	0.035	0.037	7.00	2.58
Means	0.864	0.018	0.701	0.010	0.077	7.94	1.09

[a] Sample size = 24 flies.

Fisher's exact tests of deviation of observed and expected heterozygosities (14 df): $*P < .05$, $**P < .01$, $***P < .001$.

haplotype was 77%. The total diversity, $H_T = 0.85$, indicates that two randomly chosen flies from the worldwide pool are likely to have the same SSCP haplotypes with a probability of 15%.

The mean Australian, Neotropical, and Nearctic sample diversity was 0.188 and the putatively ancestral regional mean sample diversity was 0.26; these regional diversities were homogeneous ($F = 0.295$, df =1, 13; $P \sim .60$). A mean 2.0 haplotype was detected in putatively colonized regions and 2.9 in the ancestral regions ($F = 1.04$, df = 1, 13; $P \sim .33$).

Random Mating and Population Differentiation

Weir and Cockerham's (1984) *f*, analogous to F_{IS}, differed from zero (Table 3), indicative of null alleles and technical difficulties in identifying heterozygotes of alleles of molecular weights that did not sufficiently differ to distinguish them on gels. *MdCA22* was particularly troublesome. Weir and Cockerham's θ, estimated from microsatellite loci, seemed small (0.13) considering the geographic scale of sampling (Table 3). This translates to a mean 1.7 reproducing migrants per generation, according to the island model and its assumptions. The average relatedness of individuals in populations relative to the average relatedness over all populations (Queller and Goodnight 1989) was 0.22. The estimates in Table 3 include jackknife estimates of standard errors and 95% confidence intervals (CIs), statistics that show that indices of departures from random mating differed significantly from zero.

An analysis of molecular variance (AMOVA) indicated 87.5% of the microsatellite variance lay within populations (Table 4). Less than 1% of the variance was attributed to the New World–Old World dichotomy, and the fixation index for this was not significantly different from zero, as shown by permutation tests.

With the exception of R_{ST}, the various indices of genetic differentiation based on microsatellite loci were consistent, affording a mean estimate of among-population differentiation of approximately 0.13, equivalent to $N_em \sim 1.7$ (Table 5). R_{ST}, estimated from variances in microsatellite fragment lengths, suggested lesser rates of gene flow ($N_em \sim 1$).

All pairwise microsatellite G_{ST} (here termed G_{ij}) estimates (Table 6) differed significantly from zero and all suggested mean exchange rates of one or more flies per generation, if assumptions of the island model hold. Pairwise estimates of differentiation based on mitochondrial variation (Table 7) were very much greater than the estimates based on microsatellites. $G_{ij} = 1$ in 6 of 76 pairwise comparisons and only 7 pairwise comparisons suggested $N_em \geq 1$. Pairwise patterns of gene flow made little sense geographically, and no correlation of genetic with geographic distance was observed ($P \sim .5$ for the three measures of genetic distance tested; e.g., Figure 2). Therefore the foregoing data were consistent with a chief assumption of the island model, that populations exchange migrants at roughly the same rate. Mitochondrial and microsatellite UPGMA trees (Figures 3 and 4), based on Nei's standard genetic distance measure, were entirely incongruent and populations within continents did not cluster. For example, mitochondrial variation suggested that Thai,

Table 2. Mitochondrial diversities in house flies estimated by the SSCP method

Region	Samples	Haplotypes	Diversity (H_S)	SD
Australian	1	1	0	0
Oriental	1	3	0.24	0.11
Afrotropical	2	5	0.43	0.08
Palearctic	5	12	0.19	0.05
Nearctic	2	3	0.17	0.05
Neotropical	2	5	0.30	0.11
Mean	13	2.2	0.23	0.06
Total diversity (H_T)			0.85	0.01

Table 3. *F* statistics at microsatellite loci after Weir and Cockerham (1984)

Locus	F_{IS}	F_{ST}	F_{IT}	Relatedness[a]
MdCA06	0.085***	0.144	0.217	0.236
MdCA104	0.076***	0.153	0.217	0.251
MdCA119	0.052*	0.126	0.171	0.215
MdCA121	0.049***	0.136	0.179	0.231
MdCA170	0.028	0.215	0.237	0.347
MdCA202	0.079*	0.101	0.172	0.173
MdCA224	0.153**	0.041	0.187	0.069
Jackknife means over loci ± SE	0.077 ± 0.016	0.130 ± 0.019	0.196 ± 0.010	0.217 ± 0.031
95% confidence interval	0.051–0.108	0.093–0.165	0.180–0.215	0.157–0.272

[a] Relatedness = $(2F_{ST}/(1 + F_{IT}))$.

* $P < .05$; ** $P < .01$; *** $P < .005$.

Canadian, Korean, and Honduran samples clustered tightly together, as did the Austrian and Senegalese samples. Microsatellite loci suggested a close genetic distance between samples from Austria and Indiana. Microsatellites suggested the California sample was the most distantly related of all.

Although the topology of the population graph (Figure 5) reflected some of the same population relationships seen in the microsatellite UPGMA tree (e.g., Senegal, Kazakhstan, Palau, and Egypt), the deficit of edges between Old and New World populations was not significant (binomial test, $P = .38$), nor was the regression of graph distance on untransformed or logarithm geographic distance ($P = .46$ and .43, respectively).

Discussion

Genetic Diversities

Microsatellite heterozygosities and allelic diversities were substantial and homogeneous in all populations, as might be expected for a species with huge effective population sizes. Mitochondrial diversities, on the other hand, varied substantially among the six geographical regions. More extensive sampling than presented here confirms the foregoing result and demonstrates that most Australian populations were monomorphic (Marquez and Krafsur 2002).

Contrasts in diversity between presumptively ancestral, Old World, and colonized regions—the New World and Palau (Australian region)—gave the expected results. Lesser diversities might be expected in the colonized regions, unless immigration from the Old World has been heavy and continuous. Indeed, microsatellite diversities were greater in the presumptively ancestral regions. Mitochondrial diversities were also greater in the ancestral regions, but not significantly so. AMOVA on the microsatellite data showed that only 0.4% of the total variance was attributed to the ancestral-colonized regional dichotomy, and the topology of the population graph also revealed no significant differentiation between these regions.

It is unsurprising that no genetic bottlenecks were detected, because excess heterozygosity would exist in bottlenecked house flies for 1 to 3 years (Luikart et al. 1998), depending principally on latitude and dispersal rates. It is clear that New World house fly populations have been endemic for hundreds of years. Late spring and early summer populations in northern regions (say, north of 35°) well may show evidence of bottlenecks that accrue from overwintering refugia, but our sampling did not include any such flies. Earlier research in Iowa, however, showed significant drift among flies overwintering in livestock quarters on different farms in Ames; local populations became homogeneous, however, in June (Black and Krafsur 1986a,b).

Table 4. Analysis of molecular variance in house fly microsatellite allele frequencies

Source of variation	df	Sum of squares	Variance components	Percentage of variation
Among groups[a]	1	20.372	0.0094	0.36
Among populations within groups	12	207.415	0.3134	12.02
Within populations	656	1499.197	2.2854	87.62
Total	669	1726.984	2.6082	100

[a] New World: Palau and North, Central, and South America; Old World: Palearctic, Orient, Ethiopian region.

Fixation indices and permutation tests of $F = 0$: F_{ST}, 0.124 among all populations, $P \ll .001$; F_{SG}, 0.121 New World/Old World grouped populations, $P \ll .001$; F_{GT}, 0.004 between New World and Old World, $P \sim .24$.

Table 5. Indices of genetic differentiation at microsatellite and mitochondrial loci in house flies

Genomic class	Parameter	Estimate	Reference
Nuclear	R_{ST}	0.204	Michalakis and Excoffier (1996)
	θ	0.130	Weir and Cockerham (1984)
	G_{ST}	0.128	Nei (1987)
	F_{ST}	0.140	Private allele method Slatkin and Barton (1989)
	F_{ST}	0.124	Schneider et al. (2000)
	F_{DT}	0.121	Wright (1978)
Mitochondrial	θ	0.352	Weir (1996)
	G_{ST}	0.730	Nei (1987)
	F_{ST}	0.920	Slatkin and Barton (1989)

Table 6. Pairwise differentiation G_{ST} estimated from microsatellite loci (upper diagonal) and rates of gene flow (lower diagonal)

Sample	Egypt	Senegal	Zimbabwe	Korea	Thailand	Palau	Kazakhstan	Russia	Austria	California	Canada	Indiana	Honduras	Uruguay
Egypt		0.15	0.17	0.15	0.17	0.15	0.15	0.17	0.18	0.22	0.19	0.15	0.16	0.21
Senegal	1.45		0.18	0.15	0.17	0.18	0.15	0.16	0.17	0.25	0.19	0.15	0.16	0.21
Zimbabwe	1.19	1.12		0.19	0.20	0.20	0.20	0.22	0.21	0.27	0.24	0.19	0.21	0.24
Korea	1.41	1.46	1.05		0.17	0.17	0.16	0.16	0.17	0.22	0.20	0.16	0.17	0.22
Thailand	1.23	1.25	0.99	1.19		0.19	0.18	0.19	0.19	0.25	0.22	0.18	0.19	0.20
Palau	1.41	1.13	1.00	1.26	1.07		0.18	0.20	0.19	0.24	0.21	0.19	0.19	0.23
Kazakhstan	1.42	1.47	1.03	1.31	1.11	1.11		0.18	0.19	0.25	0.18	0.17	0.16	0.22
Russia	1.19	1.27	0.91	1.27	1.06	0.99	1.15		0.20	0.24	0.21	0.18	0.18	0.23
Austria	1.15	1.20	0.97	1.26	1.08	1.05	1.05	1.01		0.25	0.23	0.18	0.19	0.24
California	0.90	0.77	0.66	0.89	0.75	0.79	0.74	0.78	0.75		0.28	0.23	0.24	0.29
Canada	1.10	1.07	0.81	1.01	0.88	0.92	1.15	0.92	0.82	0.63		0.20	0.18	0.26
Indiana	1.38	1.37	1.09	1.30	1.12	1.09	1.26	1.15	1.17	0.81	0.98		0.17	0.23
Honduras	1.28	1.30	0.93	1.23	1.07	1.07	1.32	1.17	1.08	0.80	1.14	1.23		0.23
Uruguay	0.97	0.94	0.77	0.88	1.01	0.83	0.87	0.82	0.78	0.61	0.72	0.86	0.84	

F Statistics and Analogues

For microsatellites, F_{ST} and its analogues varied from 0.12 to 0.14, in remarkably close agreement. R_{ST} was greater at 0.20, but the estimate was inflated because of low-frequency outliers in allele size. Compensating for this is size homoplasy and allele size constraints. The mitochondrial equivalents varied enormously from Weir and Cockerham's method (θ = 0.35) to Slatkin and Barton's private allele method (F_{ST} = 0.92). Nei's (1987) method afforded a middling estimate of G_{ST} = 0.73. Thus mitochondrial indices of population differentiation were much greater than afforded by microsatellites. The mitochondrial genome is much more sensitive to demographic phenomena than the nuclear genome because of its essentially unrecombinant, matrilineal pattern of inheritance (Avise 2004). Even so, estimates of differentiation based on microsatellites seem highly conservative given the huge geographic distances involved and the absence of discernable biogeographic patterns. Moreover, a large proportion (24%) of microsatellite diversity was unshared among populations, testifying to strong genetic drift and suggesting that microsatellites greatly underestimated F_{ST} and related indices of differentiation. On the other hand, the frequencies of most alleles were 10% or less, and many were less than 5%, as can readily be seen in Figure 1. Larger sample sizes than ours would be required to consistently detect such "rare" alleles in samples. For example, a sample of 59 genomes (approximately 30 flies) is necessary to have a 95% chance of detecting an allele with an a priori frequency of 0.05.

Microsatellites have become the markers of choice for much ecological and population genetics research. Once a suite of loci have been developed, it is convenient to assemble genotypes from as many samples as can be made available. But microsatellites are not without problems (Balloux et al. 2000; Balloux and Goudet 2002; Balloux and Lugon-Moulin 2002; Colson and Goldstein 1999; Epperson 2005; Hedrick 1999; Nagylaki 1998; Nauta and Weissing 1996). Size homoplasy (alleles similar in state that differ by descent) leads to underestimation of differentiation. Neither the infinite allele nor stepping stone mutation models are appropriate. High variances and high and variable mutation rates can also limit the usefulness of microsatellites in estimating classical parameters of breeding structure. High gene diversities also introduce problems in assaying breeding structure. It is well known, for example, that Wright's F_{ST} cannot exceed the

Table 7. Pairwise differentiation G_{ij} (upper diagonal) and gene flow estimates (lower diagonal) from estimated by mitochondrial variation

Population	Egypt	Senegal	Zimbabwe	Korea	Thailand	Palau	Kazakhstan	Russia	Austria	California	Canada	Honduras	Uruguay
Egypt		0.70	0.51	0.81	0.70	0.81	0.81	0.81	0.52	0.64	0.81	0.70	0.63
Senegal	0.22		0.28	0.88	0.77	0.88	0.88	0.88	0.10	0.71	0.88	0.77	0.70
Zimbabwe	0.48	1.30		0.69	0.58	0.69	0.69	0.69	0.11	0.52	0.69	0.58	0.51
Korea	0.11	0.07	0.22		0.09	1.00	1.00	1.00	0.70	0.83	0.00	0.09	0.79
Thailand	0.22	0.15	0.37	4.84		0.88	0.88	0.88	0.59	0.71	0.09	0.04	0.66
Palau	0.11	0.07	0.22	0.00	0.07		1.00	1.00	0.70	0.83	1.00	0.88	0.82
Kazakhstan	0.11	0.07	0.22	0.00	0.07	0.00		0.00	0.70	0.83	1.00	0.88	0.82
Russia	0.11	0.07	0.22	0.00	0.07	0.00	0.00		0.70	0.83	1.00	0.88	0.82
Austria	0.46	4.30	4.01	0.21	0.35	0.21	0.21	0.21		0.53	0.70	0.59	0.53
CA	0.28	0.20	0.45	0.10	0.20	0.10	0.10	0.10	0.44		0.83	0.71	0.65
Canada	0.11	0.07	0.22	0.00	4.84	0.00	0.00	0.00	0.21	0.10		0.09	0.79
Honduras	0.22	0.15	0.37	4.84	13.57	0.07	0.07	0.07	0.35	0.20	4.84		0.63
Uruguay	0.30	0.21	0.48	0.14	0.26	0.11	0.11	0.11	0.44	0.26	0.14	0.30	

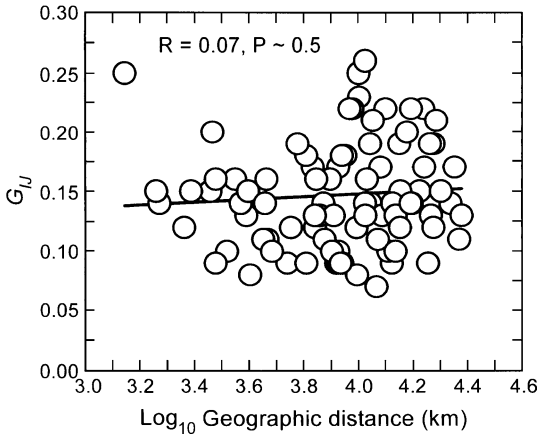

Figure 2. Pairwise genetic distance versus geographical distance.

level of homozygosity, and high allelic diversities make homozygotes less likely (Hedrick 1999; Nagylaki 1998; Wright 1978).

Gene Flow and Biogeography

Considering house fly vagility, we might expect multiple, independent colonizations from the Old World to the New, and the data, including UPGMA and population graph analyses (Figures 3–5), do not refute the expectation. Old World diversities were greater than New World diversities, as would be expected if colonizations and subsequent spread involved population bottlenecks. Diversities in presumptively colonized regions are now substantial, however. Pronounced drift, indicated by high frequencies of private nuclear and mitochondrial alleles in these regions suggest genetically diverse colonizing insects and longstanding endemicities. New World diversities include microsatellite and mitochondrial alleles not detected in the Old World, but more intensive and extensive sampling would be required to test this proposition adequately.

Comparison With Other Colonizing Species

Dispersal and colonization of new continents and islands in *Drosophila* species and the Mediterranean fruit fly have been incremental. The *Drosophila* picture does not afford unbiased estimates of geographic gene flow because most samples were derived from isofemale laboratory cultures that differed greatly in age. In Med flies, primary and secondary colonizations were accompanied by substantial losses of gene diversity when compared with populations in the home territories of East Africa (Gasperi et al. 2002). Med flies can utilize a wide variety of wild and commercial fruits for larval development and domestic populations can rapidly become feral. Moreover, good estimates of gene flow among Med flies have been documented because wild populations were sampled. In contrast to Med flies, all sampled house fly populations showed abundant nuclear gene diversities and little evidence of having undergone substantial bottlenecks in population size. Founding colonizers generally seem to have been substantial in number. Our data do not reject a hypothesis of multiple primary colonizations followed by abundant secondary, tertiary, and quaternary colonizations. The foregoing pattern, moreover, has been accompanied by substantial drift, as shown by mitochondrial variation. In an earlier survey, 38 of 111 (34%) population samples were monomorphic and the mean diversity averaged over six continents was 0.27 (Marquez and Krafsur 2002).

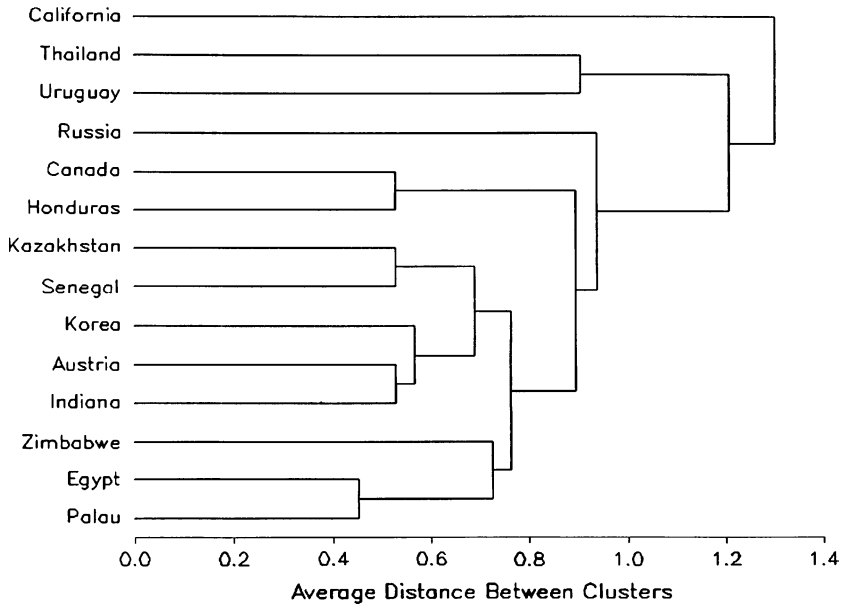

Figure 3. Nei's standard genetic distances in UPGMA format computed from seven microsatellite loci in *M. domestica*.

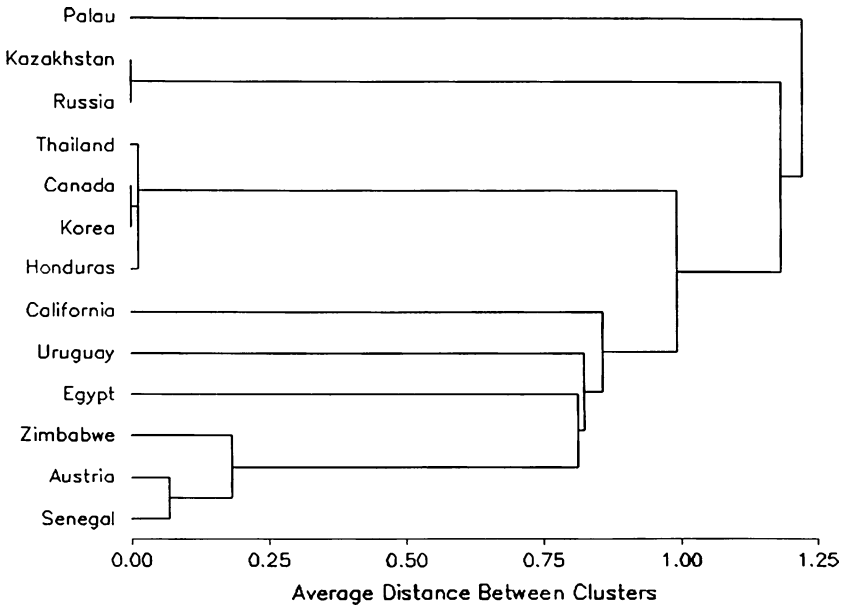

Figure 4. Nei's standard genetic distances in UPGMA format computed from mitochondrial loci in *M. domestica*.

No particular phylogeographic pattern to house fly colonization emerged and it seems that present day patterns of reproductive dispersal over large geographic distances is now constrained. Given the vagility of house flies, the nature of restraint is of interest. The distance between the sampled populations would seem the most likely constraint, but all tests of the hypothesis indicated no relationship between geographic and genetic distances. The pairwise data, moreover, do not suggest any discernable pattern of colonization. It seems clear, however, that house fly biogeography is different than that in the Med fly. Unlike Med flies, house flies can overwinter in temperate climates and exploit a wider variety of substrates for larval development; after all, fermenting organic matter is available just about everywhere, all year round.

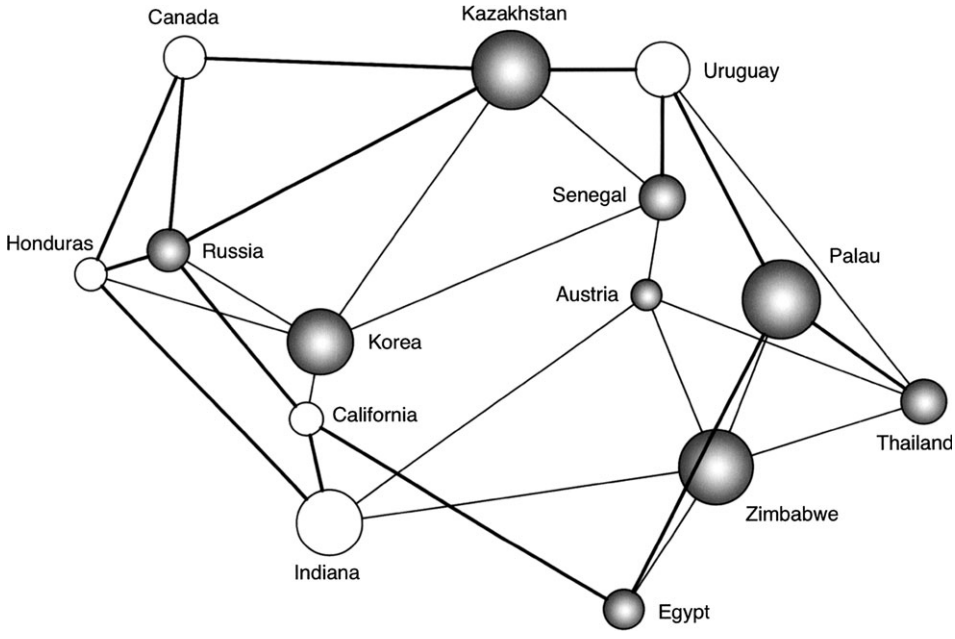

Figure 5. Population graph of interpopulation relationships computed from seven microsatellite loci in *M. domestica*. Dark spheres are Old World populations; white spheres are New World populations. Population diameters are proportional to their heterozygosities.

Therefore their geographic range is greater than that of the Med fly, their opportunities to colonize new territory are greater, and their effective population sizes are much greater.

Explicit comparison with *Drosophila* species is difficult because their levels of genetic resolution and sampling greatly differ. *Drosophila* and *Musca* are cosmopolitan and effectively commensals of mankind that can, however, prosper independently of mankind. House flies are large and robust compared with *Drosophila*. We might therefore expect house fly vagility to be greater than *Drosophila*, which would predict greater spatial homogeneity genetically. The expectation was not fully confirmed. We can compare mitochondrial indices of *Drosophila* with those of *Musca*, insofar as this is legitimate. Analysis of restriction fragment length polymorphisms (RFLPs) in 144 isofemale lines of *Drosophila* afforded a worldwide estimate of $G_{ST} = 0.66$ (Hale and Singh 1991), while that of *Musca* is rather close ($G_{ST} = 0.73$).

Acknowledgments

We thank David Bartholomew, Bruce Christiansen, Joel Coats, Hoda Farid, Nancy Hinkle, Yonggyun Kim, Tim Lysyk, Pat Matteson, Roger Moon, Marlin Rice, and Laura Weiser for collecting and sending us house flies. Jeff Scott kindly read and commented on the manuscript.

References

Andolfatto P, 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol 18:279–290.

Avise JC, 2004. Molecular markers, natural history and evolution, 2nd ed. Sunderland MA: Sinauer Associates.

Balloux F, Brunner H, Lugon-Moulin N, Hausser J, and Goudet, J, 2000. Microsatellites can be misleading: an empirical and simulation study. Evolution 54:1414–1422.

Balloux F and Goudet J, 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Mol Ecol 11:771–783.

Balloux F and Lugon-Moulin L, 2002. The estimation of population differentiation with microsatellite markers. Mol Ecol 11:155–165.

Bishopp FC, 1916. Dispersion of flies by flight. J Agric Res 21:729–766.

Black WC IV and DuTeau NM, 1997. RAPD-PCR and SSCP analysis for insect population genetic studies. In: Molecular biology of insect disease vectors: a methods manual (Crampton JM, Beard CB, and Louis C, eds). New York: Chapman & Hall; 361–373.

Black WC and Krafsur ES, 1986a. Population biology and genetics of winter house fly (Diptera: Muscidae) populations. Ann Entomol Soc Am 79: 636–644.

Black WC and Krafsur ES, 1986b. Seasonal breeding structure in house fly, *Musca domestica* L., populations. Heredity 56:289–298.

Black WC and Krafsur ES, 1987. Fecundity and size in the housefly: an investigation of environmental causes and genetic correlates of variation. Med Vet Entomol 1:369–382.

Colson I and Goldstein DB, 1999. Evidence for complex mutations at microsatellite loci in *Drosophila*. Genetics 152:617–627.

Dyer RJ and Nason JD, 2004. Population graphs: the graph theoretic shape of genetic structure. Mol Ecol 13:1713–1727.

Endsley MA, Baker MD, and Krafsur ES, 2002. Microsatellite loci in the house fly, *Musca domestica* L. (Diptera: Muscidae). Mol Ecol Notes 2: 72–74.

Epperson BK, 2005. Mutation at high rates reduces spatial structure within populations. Mol Ecol 14:703–710.

Gasperi G, Bonizzoni M, Gomulski LM, Murelli V, Torti, C, Malacrida AR, and Guglielmino, CR, 2002. Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*. Genetica 116:125–136.

Goudet J, 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at http://www.unil.ch/izea/softwares/fstat.html.

Greenberg JB, 1973. Flies and disease. Vol. 1, Ecology, classification, and biotic associations. Princeton, NJ: Princeton University Press.

Hale LR and Singh RS, 1991. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster* IV. Mitochondrial DNA variation and the role of history vs selection in the genetic structure of geographic populations. Genetics 129:103–117.

Hedrick PW, 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution 53:313–318.

Krafsur ES, 1985. Age composition and seasonal phenology of house fly (Diptera: Muscidae) populations. J Med Entomol 22:515–523.

Krafsur ES and Endsley MA, 2002. Microsatellite diversities and gene flow in the tsetse fly, *Glossina morsitans s.l.* Med Vet Entomol 16:292–300.

Krafsur ES, Helm JM, and Black WC IV, 1992. Genetic diversity at electrophoretic loci in the house fly, *Musca domestica* L. Biochem Genet 30:317–328.

Legner EF and McCoy CW, 1966. The house fly, *Musca domestica* Linnaeus, as an exotic species in the Western Hemisphere incites biological control studies. Can Entomol 98:243–248.

Luikart G, Allendorf FW, Cornuet J-W, and Sherwin WB, 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered 89:238–247.

Malacrida AR, Marinoni F, Torti C, Gomulski LM, Sebastiani F, Bonvicini C, Gasperi G, and Guglielmino CR, 1998. Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. J Hered 89:501–507.

Marquez JG and Krafsur ES, 2002. Gene flow among geographically diverse house fly populations (*Musca domestica* L.): a worldwide survey of mitochondrial diversity. J Hered 93:254–259.

Marquez JG and Krafsur ES, 2003. Mitochondrial diversity in New World house flies (Diptera: Muscidae). J Med Entomol 40:30–35.

Michalakis Y and Excoffier L, 1996. A generic estimation of population subdivision using distances between alleles with special interest to microsatellite loci. Genetics 142:1061–1064.

Nagylaki T, 1998. Fixation indices in subdivided populations. Genetics 148:1325–1332.

Nauta MJ and Weissing FJ, 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics 143:1021–1032.

Nei M, 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Oldroyd, H, 1964. The natural history of flies. New York: W.W. Norton .

Park SDE, 2001. Trypanotolerance in West African cattle and the population genetic effects of selection (PhD dissertation). Dublin, Ireland: University of Dublin.

Piry S, Luikart G, and Cornuet J-M, 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503.

Pascual M, Aquadro CF, Soto V, and Serra L, 2001. Microsatellite variation in colonizing and Palearctic populations of *Drosophila subobscura*. Mol Biol Evol 18:731–740.

Quarterman KD, Mathis W, and Kilpatrick JW, 1954. Urban fly dispersal in the area of Savannah, Georgia. Econ Entomol 47:405–412.

Queller DC and Goodnight KF, 1989. Estimating relatedness using genetic markers. Evolution 43:258–275.

Raymond M and Rousset F, 1995. GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

Rosales AL, Krafsur ES, and Kim Y, 1994. Cryobiology of the face fly and house fly (Diptera: Muscidae). J Med Entomol 31:671–680.

Rousset F, 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145:1219–1228.

Schneider S, Kueffer J-M, Roessli D, and Excoffier L, 2000. Arlequin ver. 2.0: a software for population genetic data analysis. Geneva, Switzerland: University of Geneva.

Shahjahan RM, Hughes KJ, Leopold RA, and DeVault JD, 1995. Lower incubation temperature increases yield of insect genomic DNA isolated by the CTAB method. Biotechniques 19:333–334.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, and Flook P, 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequence and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87:651–701.

Skidmore P. 1985. The biology of Muscidae of the world. Vol. 29, Series Entomologica. Dordrecht: Junk.

Slatkin M and Barton NH, 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349–1368.

Weir BS, 1996. Genetic data analysis II. Sunderland, MA: Sinauer Associates.

Weir BS and Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Welch EV, 1939. Insects found on aircraft at Miami, Fla., in 1938. Pub Hlth Rep 54:561–566.

West L, 1951. The housefly. Its natural history, medical importance, and control. Ithaca, NY: Comstock.

Wright S, 1978. Evolution and the genetics of populations. Vol. 4, Variability within and among natural populations. Chicago: University of Chicago Press.

Received August 30, 2004
Accepted May 26, 2005

Corresponding Editor: R. C. Woodruff