# Bioinvasions of the Medfly *Ceratitis capitata*: Source Estimation Using DNA Sequences at Multiple Intron Loci

**Neil Davies,*,† Francis X. Villablanca*,‡ and George K. Roderick*,†**

**Center for Conservation Research and Training, University of Hawaii, Honolulu, Hawaii 96822, †Department of Environmental Science, Policy and Management, University of California, Berkeley, California 94720-3112 and ‡Department of Biological Science, California Polytechnic State University, San Luis Obispo, California 93407*

Manuscript received February 17, 1999
Accepted for publication June 1, 1999

ABSTRACT

The Mediterranean fruit fly, *Ceratitis capitata*, is a devastating agricultural pest that threatens to become established in vulnerable areas such as California and Florida. Considerable controversy surrounds the status of Californian medfly infestations: Do they represent repeated introductions or the persistence of a resident population? Attempts to resolve this question using traditional population genetic markers and statistical methods are problematic because the most likely source populations in Latin America were themselves only recently colonized and are genetically very similar. Here, significant population structure among several New World medfly populations is demonstrated through the analysis of DNA sequence variation at four intron loci. Surprisingly, in these newly founded populations, estimates of population structure increase when measures of subdivision take into account the relatedness of alleles as well as their frequency. A nonequilibrium, likelihood-based statistical test that utilizes multilocus genotypes suggests that the sole medfly captured in California during 1996 was introduced from Latin America and was less likely to be a remnant of an ancestral Californian population. Many bioinvasions are hierarchical in nature, consisting of several sequential or overlapping invasion events, the totality of which can be termed a metainvasion. Phylogenetic data from multilocus DNA sequences will be vital to understanding the evolutionary and ecological processes that underlie metainvasions and to resolving their constituent levels.

BIOINVASIONS, especially those involving exotic insect pests, have been an undesirable element of U.S. and world agriculture for more than a century (Elton 1958; Frank and McCoy 1991; Metcalf 1995b). Despite the substantial economic impact of bioinvasions (U.S. Congress OTA 1993), management policy remains largely reactive because it is extremely difficult to predict which species will invade, when they will do so, and what the ecological consequences will be (Lawton and Brown 1986; Daehler and Strong 1993; Lodge 1993; New 1995; Kareiva 1996; Davies and Roderick 1999). Nearly all attempts at eradication or control (sterile insect release, biological control, the use of transgenic plants) require the determination of invasion sources and demographic parameters, such as population size, structure, and rate of growth. For example, very different management measures are required when invasions involve a pest species with panmictic population structure compared to those that involve a species with sharply defined and genetically isolated populations (Carey 1991). Here we address the fundamental issue of how to determine the source of a bioinvasion. Answering this apparently straightforward question has required the development of novel genetic markers and statistical analyses (reviewed by Davies *et al.* 1999). We apply these new techniques to the medfly *Ceratitis capitata*; in particular, can these approaches distinguish alternative hypotheses concerning the source of medfly infestations in California?

The medfly is arguably the world's most threatening agricultural pest, attacking over 200 different fruits, vegetables, and nuts (Carey 1991; Sheppard *et al.* 1992; McPheron *et al.* 1994). From their likely origin in Africa, medflies spread rapidly around the world (Figure 1). They were first reported from California in 1975 (for a review see Metcalf 1995a) and sporadic outbreaks have occurred ever since (Figure 2); it has been suggested that California may have had a resident population for up to 50 yr at a size below the level of detection (Carey 1991). The economic risk associated with the medfly is considerable: an endemic infestation in California might cost in excess of $1 billion per annum (Siebert 1994) and over $350 million has already been spent trying to prevent such an outcome (Metcalf 1995b).

The history of medfly invasions in California has been reviewed recently by several authors (McPheron *et al.* 1995; Roderick 1996b; Gasparich *et al.* 1997; Haymer *et al.* 1997) but the origin(s) of these infestations remain uncertain and controversial. Early work focused on mtDNA, but the many advantages of mtDNA over nu-

*Corresponding author:* Neil Davies, Division of Insect Biology, Department of Environmental Sciences, Policy and Management, 201 Wellman Hall, University of California, Berkeley, CA 94720-3112.
E-mail: ndavies@nature.berkeley.edu

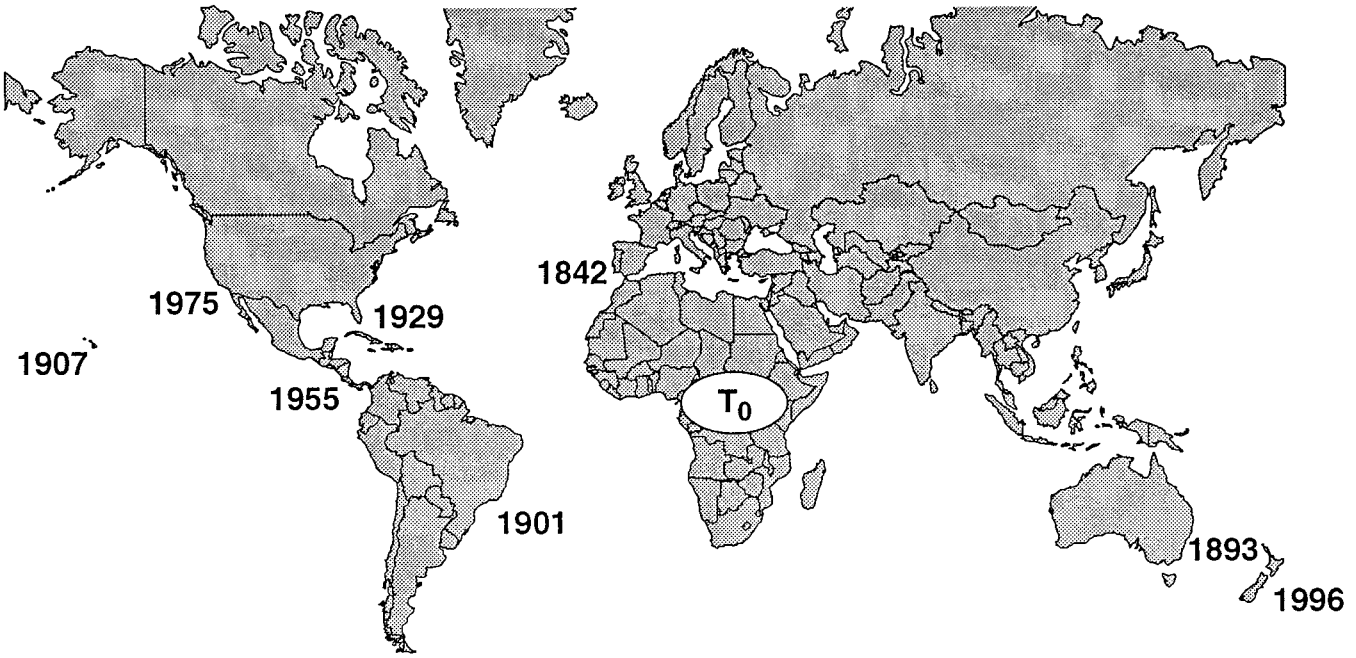


Figure 1.—The biogeography of the medfly invasion.

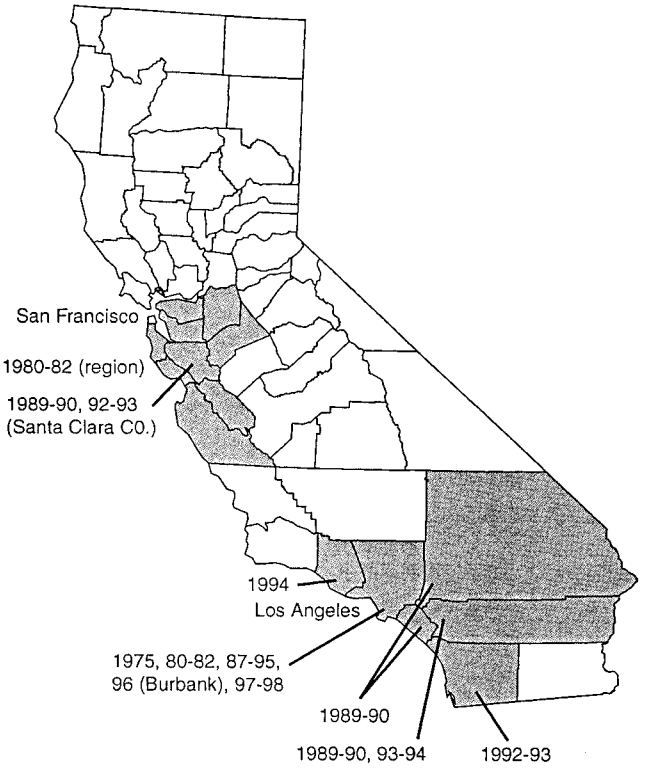


Figure 2.—Chronology of recent medfly infestations in California. Shaded regions are counties in which medflies have been trapped, prompting control programs. For more information see http://www.cdfa.ca.gov/pests/medfly/mediterranean_fly.html.

clear loci (Moore 1995; Roderick 1996a) may not apply to medfly invasions (Roderick *et al.* 1998). First, medflies exhibit lekking behavior, an extreme form of mating bias that reduces the effective population size of nuclear loci to levels similar to those of mtDNA (Hoelzer 1997). Second, restriction fragment length polymorphism (RFLP) analysis of mtDNA reveals low levels of genetic variation in New World populations, and medflies caught in Southern California prior to 1992 share a single mtDNA haplotype that is common throughout Central America and some other parts of the world (Sheppard *et al.* 1992; McPheron *et al.* 1994, 1995; Steck *et al.* 1996; Gasparich *et al.* 1997). Clearly, the precise identification of origins is impossible on the basis of these mtDNA data alone, and the single *vs.* multiple origin of Californian medflies remains unresolved (Gasparich *et al.* 1997).

Allozymes are variable in Mediterranean and African populations, but less so in populations outside of Africa (Gasperi *et al.* 1991; Malacrida *et al.* 1992; Baruffi *et al.* 1995). Fortunately, it is now possible to assess DNA variation directly, and high levels of diversity have been documented among invading medfly populations using several DNA-based nuclear markers, including microsatellites (B. McPheron and D. Prasher, personal communication), randomly amplified polymorphic DNA (RAPDs; Haymer *et al.* 1997), and introns (Gomulski *et al.* 1998; Villablanca *et al.* 1998). Villablanca *et al.* (1998) demonstrated that significantly more diversity is retained in invading populations at intron loci than is retained at mitochondrial or allozyme loci. The substantial amount of variation found with nuclear DNA

markers offers the possibility that potential sources might be resolved on a relatively fine geographic scale. Here we examine variation at the four intron loci sequenced by Villablanca *et al.* (1998) to investigate whether these markers can distinguish invading medfly populations in Central America and eastern South America, *i.e.*, those that mtDNA studies indicate as the most likely source of Californian medflies (Gasparich *et al.* 1997).

Three important questions are addressed by the current study: First, does sequence variation at intron loci reveal significant population structure? Second, how do frequency and distance-based approaches compare in their estimation of population subdivision? Finally, can we utilize genetic population structure to determine the origin of a recent medfly outbreak in California? For the last question, we focus on the single fly (*B-96*) captured at Burbank, California, in 1996. Using a statistical test that assigns multilocus genotypes to potential source populations based on the frequency of alleles in those sources (Rannala and Mountain 1997), we ask whether this fly was recently introduced to California, or whether it represents a remnant of a previous infestation. The results have direct consequences for medfly management strategy: either more effort must be directed toward eradication or interception policies should be focused on the most common source, if it can be identified.

## MATERIALS AND METHODS

**Collections and laboratory analysis:** Figure 1 details the recent spread of medflies from their ancestral range in sub-Saharan Africa (Hardy and Delfinado 1980; White and Elson-Harris 1992; McPheron *et al.* 1995; Metcalf 1995a; Foote 1997). Medfly samples were obtained from populations in California, Costa Rica, Guatemala, Mexico, Brazil, Peru, Greece, Hawaii, and Africa (Table 1). The Californian medflies were all collected in southern California between 1992 and 1996 (Figure 2). No medflies were found in 1995 and only a single individual (*B-96*) was captured during 1996. We treat all these Californian medflies as a single population for the purposes of statistical analysis. Under the null hypothesis that there is a resident medfly population in California, we assume that these flies, captured in the same geographic area, represent a single biological population (Carey 1991).

Extraction protocols and some of the sequences (from California, Brazil, Malawi, Hawaii, and Greece) have been published previously (see Villablanca *et al.* 1998). Briefly, DNA was isolated using a phenol-chloroform extraction and ethanol precipitation. Exon-primed, intron crossing (EPIC) primers (Palumbi and Baker 1994; Palumbi 1996) were constructed to amplify introns from within four different nuclear loci: the first intron of Cu/Zn superoxide dismutase (*Sod*), muscle-specific actin (*Actin*), vitellogenin 1 gamma intron 2 (*Vg*), and first chorion intron s36 (*Chorion*). Amplification and cloning were as in Villablanca *et al.* (1998) with the following modifications. PCR conditions, previously optimized for the Perkin-Elmer (PE; Norwalk, CT) 480 thermocycler, were adjusted for use with the PE 9600 and 9700. Reactions consisted of 30 μl water, 5 μl 8 mm dNTPs, 5 μl 10× PE *Taq* buffer, 2.4 μl of each 10 mm primer, 5 μl $MgCl_2$ (25 mm), 0.12 μl PE *Taq* polymerase (5 units/μl), and 1 μl of genomic DNA. Thermocycler parameters were 94° for 30 sec, 55° for 30 sec, and 72° for 40 sec, repeated for 40 cycles. Parameters for *Sod* were 94° for 30 sec, 60° for 1 min, and 72° for 1.5 min for 40 cycles.

As the PCR products in any reaction may include two alleles from the target locus, amplifications were cloned to isolate a single allele for sequencing. Following cloning, double-stranded DNA was prepared for sequencing using QuantumPrep (Bio-Rad Laboratories, Hercules, CA) following the manufacturer's guidelines. Sequences were obtained using an ABI 377 automated sequencer. Because of the large amount of sequencing that must be performed to recognize heterozygotes, we sequenced a single allele for most individuals. Cloned PCR products may contain errors incorporated by DNA polymerases (Pääbo and Wilson 1988), and thus different clones from the same PCR reaction are expected to differ by ~1:1000 bases due to *Taq* polymerase copying errors. We sought to reduce this source of error by ignoring singletons, which are nucleotide positions that vary in only one allele (*sensu* Villablanca *et al.* 1998).

**Phylogenetic analysis:** Phylogeography is a powerful means of investigating the history of populations (Avise 1994); however, phylogenetic analysis can be complicated for nuclear genes because of recombination (Begun and Aquadro 1992; Hudson 1994) and the persistence of ancestral alleles (Templeton and Sing 1993). Templeton *et al.* (1992) and Templeton and Sing (1993) provided a means for recognizing and dealing with alleles produced by recombination. In effect, the method focuses phylogenetic reconstruction on the most recent substitutions before attempting to resolve deeper nodes. Assuming that more recent changes are less likely to include homoplasy or recombination, sequences separated by one substitution are linked first, and the resulting groups form the basis for linking alleles separated by two substitutions, and so on. Details of the method are provided by Templeton (1998) and Templeton and Sing (1993).

**Population analysis and assignment test:** Population structure was determined through the analysis of molecular variation (AMOVA) method of Excoffier *et al.* (1992) as implemented by the computer program ARLEQUIN 1.1 (Schneider *et al.* 1997). Indices of population subdivision analogous to standard F-statistics (Wright 1951, 1965) are calculated for each locus and their significance assessed using permutation procedures that randomly redistribute individuals among populations. For each locus, this general approach was used to estimate two F-statistics, $\theta$ and $\Phi$. $\theta$ is based on allele frequencies only (Weir and Cockerham 1984; see Excoffier *et al.* 1992), while $\Phi$ also takes into account the genetic distance among alleles (Excoffier *et al.* 1992), in this case the average Kimura 2-parameter distance.

The computer program TFPGA (Miller 1998) was used to calculate the average $\theta$ (Weir and Cockerham 1984) across all loci on the basis of allele frequencies. The average $\Phi$ across loci, $\Phi_{AMOVA}$, was the mean of $\Phi$ values for the four loci, and its significance was assessed following Fisher's method for combining probabilities (see Sokal and Rohlf 1995). Population differentiation averaged over loci was also examined using the exact test (Raymond and Rousset 1995) as implemented by TFPGA (Miller 1998). Again, both frequency and distance-based approaches were used. Exact tests were performed on the basis of the distribution among populations of (i) alleles and (ii) phylogenetically derived allele groups, or clades. In the latter case, alleles in the same one-step clade of the Templeton networks were pooled. TFPGA was then used to calculate the average $\theta_{Templeton}$ (Weir and Cockerham 1984) across all loci on the basis of allele frequencies and then across all loci on the basis of the one-step allele groups. Significance of both statistics was assessed through 10,000 bootstrap replications over loci.

To determine the origin of the *B-96* medfly, Rannala and Mountain's (1997) assignment test (see Davies *et al.* 1999) was employed using their computer program IMMANC (http://allele.bio.sunysb.edu/). Assuming that loci are independent, the likelihoods of drawing a test multilocus genotype from a resident and from a potential source population are calculated. The difference in likelihood (ln λ) is negative if an individual is more likely to be an immigrant than a resident. Monte Carlo simulations are used to assess significance: random genotypes are generated from allele frequency distributions conditioned on the observed allele frequencies (the variance of these distributions decreases with increasing sample size). The probability of obtaining the observed value is assessed by comparison with the distribution of random values. Because the *B-96* genotype was included in the Californian population (and not the potential source) when estimating the "resident" allele frequencies, the test is conservative with respect to the null hypothesis that *B-96* is a resident—in this case of California (Davies *et al.* 1999). Currently, IMMANC accepts only diploid data and we therefore analyzed our data (which included only one of the two alleles possessed by each individual at each locus) by assuming that all (unobserved) genotypes were homozygotes. Although this modification does not alter the observed allele frequencies, it does artificially reduce their variance in the simulation analysis, making the interpretation of significance values problematic. The genotype probabilities calculated in this way are only approximate and the approximation will work best when the frequencies of the sampled alleles are high (*i.e.*, when genotypes have a high probability of being homozygous).

## RESULTS

A total of 237 medfly sequences were obtained for four intron loci in 74 individuals (Table 1). Not every individual was sequenced for all four loci due to problems in obtaining PCR amplification, which were probably due to the degraded state of some specimens. After a correction for errors associated with cloning PCR products (see above), 68 *Actin* sequences consisting of 293 bp revealed 19 alleles distinguished by 12 polymorphic sites. Fifty-seven *Chorion* sequences consisting of 394 bp revealed 18 alleles distinguished by 14 polymorphic sites. Sixty *Sod* sequences consisting of 812 bp revealed 38 alleles distinguished by 56 polymorphic sites including four insertion or deletion events. Fifty-two *Vg* sequences consisting of 409 bp revealed 32 alleles distinguished by 24 polymorphic sites including eight indels. The phylogenetic relationships of the alleles are shown for each locus in Figure 3. No striking phylogeographic structure was evident from visual inspection of the networks, suggesting that the distribution of alleles represents the random sorting of ancestral polymorphism.

Statistical analysis revealed substantial population structure among American medfly populations (Latin America plus California) and among Latin American populations only (excluding California). For American populations, both frequency and distance-based F-statistics were significant when combined across loci with the distance-based estimates, $\Phi_{AMOVA}$ and $\theta_{Templeton}$, being considerably larger than the frequency-based θ (Table 2). For Latin American populations, only θ and $\theta_{Templeton}$ were significant, and again $\theta_{Templeton}$ was twice as large as θ. AMOVA utilizes the same statistical framework and we present a comparison of locus by locus θ and $\Phi_{AMOVA}$ in Table 3. Consistent with the across-loci results, $\Phi_{AMOVA}$ is always $>\theta$ for all American populations but there is no clear pattern when only Latin America populations are considered. In all cases, *Chorion* and *Vg* revealed the most population differentiation.

The likelihood of drawing the *B-96* fly from the Californian population was compared to the likelihood that it was an immigrant from a Latin American population (Table 4). Whether *B-96* was more like an immigrant than one would expect by chance was determined through comparison of the observed ln λ values with the ln λ of 1000 multilocus genotypes randomly generated from the Californian population. Despite the conservative nature of the test, Table 4 shows that the *B-96* genotype was less likely ($\alpha < 0.05$) to be a resident of California than an immigrant from no less than four potential sources: Costa Rica, Guatemala, Mexico, or Peru. While the need to assume homozygosity in our data makes the interpretation of significance values difficult, the ancestral Californian population we sampled seems unlikely, according to this analysis, to be the source of the *B-96* medfly. To assess which of the four Latin American populations was the most likely source, we conducted further assignment tests. *B-96* was assumed to be a resident of Costa Rica, Guatemala, Mexico, and Peru, in turn. This null hypothesis was only rejected in one case: *B-96* is more likely to be an immigrant from Costa Rica than a resident of Peru ($\alpha < 0.05$). However, a Bonferroni correction for multiple comparisons would render this result insignificant.

## DISCUSSION

Contemporary bioinvasions, such as those of the medfly, are often characterized by a unique set of demographic and genetic features that result from a small number of colonizing individuals and the rapid growth and spread of new populations. Together these characteristics make it difficult to determine the source of bioinvasions, partly because it is hard to distinguish potential source populations that were only recently established themselves (Davies *et al.* 1999). In this study, we investigated sequence variation at nuclear intron loci to see if these markers could reveal population structure, and whether they could be used to determine the origin of a recent medfly "outbreak." Genetic analysis of DNA sequence variation at four intron loci, *Sod*, *Vg*, *Actin*, and *Chorion*, revealed significant population structure among previously indistinguishable medfly populations (Gasparich *et al.* 1997) in California, Central America, and eastern South America. Much of the subdivision was due to the Californian population, although significant structure was also evident when the Latin

**TABLE 1**

**Collection data**

| Region | Site | Actin | Chorion | Sod | VG |
|---|---|---|---|---|---|
| California | Pasadena, 1992 | A1 | C2 | S14 S15 | V4 |
| | Los Angeles, 1992 | A2 | C2 | S8 | V4 |
| | Westminster, 1993 | A9 | C2 | S3 | V4 |
| | Rosemaed, 1993 | A2 | C7 C6 | S3 | V4 |
| | Los Angeles, 1993 | A10 | | S1 | V4 |
| | Camarillo, 1994 | A1 A2 | C2 | S1 S1 | V4 |
| | Camarillo, 1994 | A2 | C2 | S1 | V4 |
| | Camarillo, 1994 | A1 | | S1 | V4 |
| | Camarillo, 1994 | A4 | C2 | S1 | V4 |
| | Camarillo, 1994 | A4 | C2 | S2 | V4 |
| | Camarillo, 1994 | A2 | C8 C2 | S7 S1 | V7 |
| | Burbank, 1996 | A1 | C2 | S3 | V5 |
| Brazil | Petrolina: Pernambuco | A2 | C3 | S5 | V10 V9 |
| | Petrolina: Pernambuco | A6 | C3 | S4 | |
| Costa Rica | Guanacaste: Caldera | A1 | C3 | | |
| | Guanacaste: Caldera | A2 | C1 | S3 | |
| | Guanacaste: Caldera | A8 A5 | C9 | | |
| | Guanacaste: Caldera | A12 | | S3 | |
| | Guanacaste: Caldera | A1 | C1 | S3 | |
| | Guanacaste: Caldera | A5 | C1 | S1 | |
| | Guanacaste: Caldera | | C2 | S16 | |
| Ecuador | Perucho | A13 A4 | C1 | S3 | V5 |
| | Perucho | A6 | C1 | S17 | V12 |
| | Perucho | A6 | C3 | | V13 |
| | Perucho | A6 | | S18 | V14 |
| Guatemala | Medfly repository: GU001A | A6 | C3 | | V17 V18 |
| | Medfly repository: GU001A | A5 | C2 | S23 | V4 |
| | Medfly repository: GU001A | A5 | C3 | S3 | V19 |
| | Medfly repository: GU001A | A1 | C2 | | V4 |
| | Medfly repository: GU001A | A6 | C2 | S3 | |
| | Medfly repository: GU001A | | | | |
| | Medfly repository: GU001A | A6 | C3 | | V20 |
| | Medfly repository: GU001A | A5 | C3 | | |
| | Medfly repository: GU001A | A6 | C3 | | V4 |
| Malawi | Monkey Bay | A6 | C3 | S30 S31 | V29 V30 |
| | Monkey Bay | A8 | C14 | S32 S33 | V6 |
| | Monkey Bay | A1 | | S36 S3 | |
| | Monkey Bay | A3 | C3 | | V32 |
| | Monkey Bay | A4 | C15 | S34 S35 | V31 |
| Mexico | Tapachula | | C2 | | V3 |
| | Tapachula | A6 | C3 | S37 | V3 |
| | Tapachula | A2 | C17 | S3 | V4 |
| | Tapachula | A1 | | S8 | V3 |
| | Tapachula | | C3 | S3 | V4 |
| Peru | Huaral | A1 | C18 | S12 | |
| | Huaral | A1 A5 | C3 | S12 | V4 |
| | Huaral | A1 | | | V4 |
| | Huaral | A6 | | S10 | V4 |
| | Huaral | A1 | | S38 | V7 |

(*continued*)

**TABLE 1**

**(Continued)**

| Region | Site | Actin | Chorion | Sod | VG |
|---|---|---|---|---|---|
| Kenya | Nairobi | A2 | C3 | S10 | V2 |
| | Nairobi | A3 | | S11 | V6 |
| | Nairobi | A14 A2 | C4 | S25 | V2 |
| | Nairobi | A7 | C6 | S26 | |
| | Nairobi | A7 | C12 | S27 | V23 |
| | Nairobi | A15 A16 | C1 | | V24 |
| | Nairobi | A7 | C13 | S11 | V25 |
| | Nairobi | A4 | C6 | S28 | V26 |
| | Nairobi | A17 | C2 | | V27 |
| | Nairobi | A18 | C1 | S29 | V28 |
| Hawaii | Maui | A19 A6 | C10 C16 | S6 | |
| | Hawaii | A4 A6 | C11 | S9 | V22 |
| | Hawaii | A2 | C3 C2 | S24 S6 | V21 |
| Greece | Chios | A2 | C3 | S20 S21 | V15 |
| | Southern Peloponnese | A2 | C3 | S22 S19 | V16 |

American populations were treated alone. The ability to discern differences among populations is a prerequisite for attempts to determine the origin of invading individuals, and intron variation appears to provide enough resolution to distinguish the source of very recent bioinvasions. An assignment test (Rannala and Mountain 1997) was applied to the single medfly captured in California (at Burbank, Los Angeles County) during 1996. The results suggested that the multilocus genotype of this fly was unlikely to have arisen from the ancestral Californian population that we sampled; it was more likely introduced from Latin America. There was less ability to distinguish sources within Latin America, although the demonstration of significant population structure indicates that additional intron data is likely to further reduce the number of potential sources.

Multilocus genotyping is clearly a powerful technique and many different markers, including allozymes, mtDNA, and microsatellites can be analyzed simultaneously following the statistical methodology outlined here (see also Davies *et al.* 1999). In most cases additional markers will increase the power of these tests, although some markers may be incompatible with this approach. RAPDs, for example, are very useful in providing high levels of genetic variation and their main advantage over introns and microsatellites is that they can be applied with very little prior genetic knowledge of a species (Williams *et al.* 1990). Unfortunately, it is not clear how RAPD data can be incorporated into the same statistical framework as introns and other markers, where genotypes can be identified at each locus.

Rannala and Mountain's (1997) assignment test applies statistical rigor to source estimation; however, further modifications of the test are desirable. First, laboratory scoring mistakes must be taken into account. For example, scoring errors occur in allozyme studies with a frequency of ~1% (Lathrop *et al.* 1983). One can attempt to correct for scoring mistakes prior to analysis, as we did for the errors associated with sequencing cloned PCR products. Alternatively, an error rate factor can be incorporated into the analysis; such an approach was used by Marshall *et al.* (1998) in their multilocus paternity test. A second source of error that should be considered is the implications of not sampling all the potential sources. Again with a focus on paternity testing, Marshall *et al.* (1998) presented a simulation method to assess the likelihood that a more probable source remains unsampled. Finally, assignment tests focus on the origin of single multilocus genotypes, and we have applied such a test to a bioinvasion consisting of a solitary fly. More often, however, bioinvasions consist of many invading individuals, each of which will have its own associated likelihood of being from one source or another. With such multiple assignments, one will be able to plot a distribution of likelihood statistics for the invading population as a whole, adding a new level of complexity to source estimation. For example, a bimodal distribution would imply that the invading population had two sources, but how can one assess the significance of such a conclusion, and how should one correct for multiple comparisons? We are currently examining these issues with respect to some large medfly infestations in California that have occurred since 1996.

The final purpose of our investigation was to explore and compare the utility of frequency- and distance-based approaches in the study of invading populations. Phylogeographic structure (Avise 1989, 1994; Roderick and Villablanca 1996; Roderick and Gillespie 1998) is not expected in recently founded populations because there has been little time for mutations to occur, and the relationships among alleles reflect evolutionary events in the ancestral range of the species

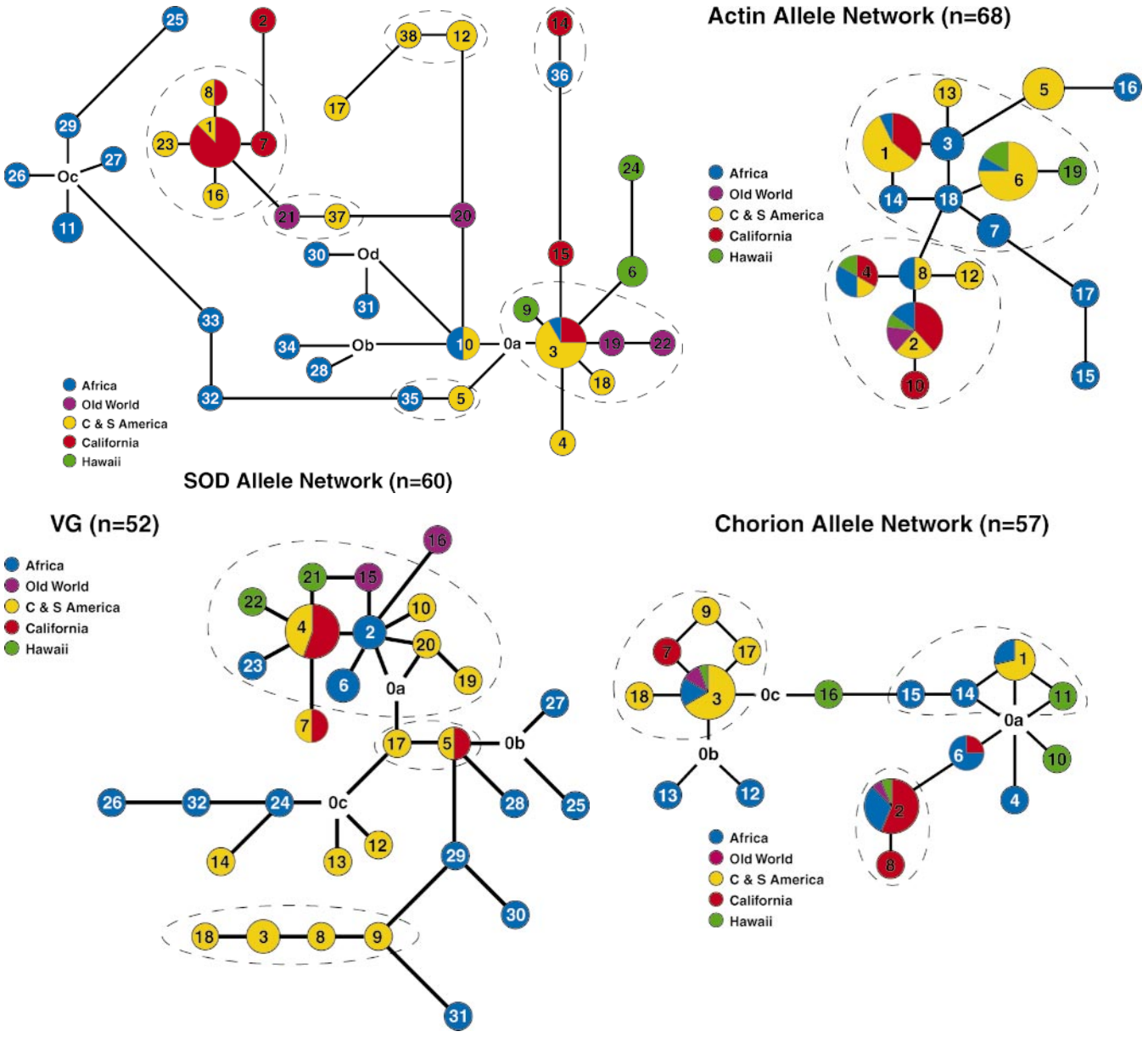


Figure 3.—The haplotype networks as estimated by the method described by Templeton (1998) for a, *Sod*; b, *Actin*; c, *Vg*; and d, *Chorion*. Haplotype identities are noted within circles by a number and distances between haplotypes are proportional to the number of base pair differences between haplotypes. The size of the circles is proportional to the number of individuals sharing a particular haplotype. Haplotypes marked O*x* are inferred haplotypes identified by parsimony. Dashed lines denote single-step clades (see text).

**TABLE 2**

**F-statistics and exact tests for population differentiation over all loci**

| | Frequency | | Distance | | |
|---|---|---|---|---|---|
| Region | θ | Exact | $\Phi_{AMOVA}$ | $\Phi_{Templeton}$ | $Exact_{Templeton}$ |
| Latin America and California | 0.16** | <0.0001 | 0.27* | 0.24** | <0.0001 |
| Latin only | 0.08** | <0.01 | 0.10NS | 0.16** | <0.01 |

Statistics are presented that are (i) solely based on the frequency of alleles, and (ii) that also consider the genetic distance among alleles. *P* values from 10,000 bootstrap replications over loci: * = <0.05, ** = <0.01.

**TABLE 3**

**Results of AMOVA for each locus**

| Region | Locus | θ | Φ |
|---|---|---|---|
| Latin America and California | Actin | 0.092 | 0.118 |
| | Chorion | 0.321 | 0.421 |
| | Sod | 0.083 | 0.174 |
| | Vitellogenin | 0.144 | 0.360 |
| Latin only | Actin | 0.067 | 0.067 |
| | Chorion | 0.086 | 0.005 |
| | Sod | 0.071 | 0.033 |
| | Vitellogenin | 0.103 | 0.285 |

A θ is an F-statistic based on allele frequencies while Φ is based on the genetic distance among alleles.

rather than their history in newly occupied areas (Villablanca *et al.* 1998; Davies *et al.* 1999). For example, McGuigan *et al.* (1998) reported significant differences in haplotype frequencies over a fine geographic scale among Australian frog populations; however, a smaller and insignificant F-statistic was obtained when the genetic distance among alleles was considered as well. McGuigan *et al.* (1998) suggested that the barriers to gene flow had not been in place long enough to influence the distribution of genetic relatedness. We expected American medflies to show a similar pattern; however, the opposite was true: distance estimates of population subdivision not only revealed significant population structure among American medflies, but yielded larger F-statistics than those based on allele frequencies alone. Various approaches have been developed that utilize phylogenetic data to detect gene flow (Slatkin 1994), yet current assignment tests do not consider the relatedness of alleles. Commonly used multilocus markers such as microsatellites (Glenn 1998) do not easily permit the phylogenetic analysis of alleles; however, when sequence data are available, our data suggest that frequency-based assignment tests might waste useful information. A new test should assess the multilocus likelihood of sampling a given set of alleles from a potential source population on the basis of the distance between alleles as well as their frequencies.

**TABLE 4**

**Results of assignment test (see text)**

| Assumed resident of: | Potential immigrant from: | ln λ | α | Power |
|---|---|---|---|---|
| California | Brazil | 13.31 | 0.182 | 1.000 |
| | Costa Rica | 1.26 | 0.025 | 0.995 |
| | Ecuador | 11.52 | 0.065 | 1.000 |
| | Guatemala | 3.84 | 0.017 | 1.000 |
| | Mexico | 2.96 | 0.028 | 0.999 |
| | Peru | 5.64 | 0.008 | 1.000 |
| Costa Rica | Guatemala | 7.75 | 0.210 | 0.999 |
| | Mexico | 6.84 | 0.177 | 0.999 |
| | Peru | 9.29 | 9.263 | 1.000 |
| Guatemala | Costa Rica | 1.58 | 0.068 | 0.978 |
| | Mexico | 3.87 | 0.224 | 0.978 |
| | Peru | 6.31 | 0.166 | 0.999 |
| Mexico | Costa Rica | 2.45 | 0.166 | 0.980 |
| | Guatemala | 5.92 | 0.314 | 0.988 |
| | Peru | 7.46 | 0.188 | 1.000 |
| Peru | Costa Rica | 0.95 | 0.037 | 0.990 |
| | Guatemala | 5.89 | 0.228 | 0.996 |
| | Mexico | 4.99 | 0.125 | 0.997 |

Interestingly, the inclusion of California increased distance- as well as frequency-based estimates of population subdivision. Californian medfly populations are the most recent, so founder events have not eroded phylogeographic signal; indeed they appear to have enhanced it. The average relatedness of alleles in derived populations might be increased simply by sampling from a single, phylogenetically heterogeneous population. Diversity within a species could be converted into divergence between populations when the species becomes fragmented because founder events sample only a small proportion of the ancestral variation and tend to include alleles of the dominant phylogenetic clade. Phylogeographic structure among recently founded populations might also occur if each derived population was founded from a different population of a phylogeographically structured source.

More work is clearly needed to explore the phylogenetic consequences of invasions and a better understanding of invasion genetic patterns will provide a deeper insight into the ecological and evolutionary processes that underlie bioinvasions. It is important to consider in such studies that invasions often involve a hierarchy of events, the totality of which might be termed a metainvasion. The metainvasion begins with a primary invasion, when a species first colonizes a new area from its ancestral source. Subsequently, secondary and tertiary invasions arise as the newly established populations themselves seed new areas. The genetic changes that result from these events are complex and phylogenetic analyses may be informative at some levels but not others. A primary invasion of the medfly occurred from Africa to the Mediterranean. The invasion of Latin America may be another primary invasion, direct from Africa, or a secondary invasion from the Mediterranean. Californian medfly infestations thus represent secondary or tertiary events in the global medfly metainvasion. Indeed, California may be subject to repeat invasions that could superimpose one another. The phylogeny of alleles as well as their frequencies in each population may need to be considered to unravel the history of the global medfly metainvasion and to distinguish its component levels. Phylogeographic structure in newly

colonized regions may arise as a second-hand signal resulting from multiple introductions or be due to the combination of different levels of the metainvasion—as here we combined California with Latin America. Whether the source (most probably in Africa) or sources of Latin American medflies is phylogeographically structured or simply a phylogenetically heterogeneous population will be determined by future genetic analysis of intron variation already under way in our laboratory. We are also addressing these questions theoretically through simulation studies.

To conclude, powerful genetic tools are now available to investigate the source of contemporary bioinvasions, and the statistical methods for analyzing these data continue to become more sophisticated. In the future, the challenge will be to integrate such genetic results with ecological and interception data to fully elucidate the process of range expansion exemplified by the medfly and other invasive species. It seems likely that the origin of Californian medfly infestations can be determined through the use of multilocus genotyping techniques such as microsatellites and methods such as single-strand conformation polymorphism (SSCP; Lessa 1992; Ortí *et al.* 1997) that rapidly screen intron variation. In the case of Californian medflies, source populations should be monitored for a suite of markers, and the database of allele frequencies regularly updated as these dynamic populations evolve due to drift or the introduction of new alleles through gene flow. (At the moment little is known about how the genetic make-up of source populations varies with time.) With adequate baseline data from source populations, new medfly infestations in California or elsewhere could be rapidly genotyped and assigned to a source. Due to its economic significance, the medfly infestation has become a model system for the study of contemporary bioinvasions and has several important lessons for other cases where limited funding is likely to restrict the amount of research effort. The genetic analysis of new bioinvasions should begin with mtDNA and allozymes; however, highly variable nuclear regions, such as introns, should also be considered. Multilocus genotyping provides a rapid method of determining the origin of invasions, whether using nonsequencing methods of screening intron variation and/or other types of markers. While the routine screening of new infestations may not require the expense of sequencing, for a full understanding of the bioinvasion process, DNA sequences at multiple loci are an invaluable and versatile tool.

We thank M. Soedarjo and B. Thorsby for their invaluable work in the lab. We are grateful to B. Rannala and J. Mountain for providing us with their computer program and for discussing its application to our study. We also thank S. R. Palumbi and A. Bohonak for comments on previous drafts of this article. Financial support was provided by the California Deparartment of Food and Agriculture, U.S. Department of Agriculture National Research Initiative and Agricultural Research Service, and National Science Foundation awards 9510822 (to F.X.V.), 9726573, 9813418, and 9816384.

## LITERATURE CITED

Avise, J. C., 1989 Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution **43:** 1192–1208.

Avise, J. C., 1994 *Molecular Markers, Natural History, and Evolution*. Chapman and Hall, London.

Baruffi, L., G. Damiani, C. R. Guglielmino, C. Mandi, A. R. Malacrida *et al.*, 1995 Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data. Heredity **74:** 425–437.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Carey, J. R., 1991 Establishment of the Mediterranean fruit fly in California. Science **253:** 1369–1373.

Daehler, C. C., and D. R. Strong, Jr., 1993 Prediction and biological invasions. Trends Ecol. Evol. **8:** 380–381.

Davies, N., and G. K. Roderick, 1999 Determining pathways of marine bioinvasion: genetical and statistical approaches, in *First Conference on Marine Bioinvasions*, edited by J. Pederson. Plenum Press, Boston (in press).

Davies, N., F. X. Villablanca and G. K. Roderick, 1999 Determining the source of newly founded populations: multilocus genotyping in nonequilibrium population genetics. Trends Ecol. Evol. **14:** 17–21.

Elton, C., 1958 *The Ecology of Invasions by Animals and Plants*. Chapman and Hall, London.

Excoffier, L., P. Smouse and J. Quattro, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479–491.

Foote, D., 1997 Selection in populations of the Mediterranean fruit fly (*Ceratitis capitata*) in Hawaii. Ph.D. Thesis, University of California, Davis.

Frank, J. H., and E. D. McCoy, 1991 The immigration of insects to Florida, with a tabulation of records published since 1970. Florida Entomologist **75:** 1–28.

Gasparich, G. E., J. G. Silva, H.-Y. Han, B. A. McPheron, G. J. Steck *et al.*, 1997 Population genetic structure of Mediterranean fruit fly (Diptera: Tephritidae) and implications for worldwide colonization patterns. Annal. Entomol. Soc. Amer. **90:** 790–797.

Gasperi, G., C. R. Guglielmino, A. R. Malacrida and R. Milani, 1991 Genetic variability and gene flow in geographical populations of *Ceratitis capitata* (Wied.) (medfly). Heredity **67:** 347–356.

Glenn, T., 1998 *Microsatellite Manual v. 6*. Smithsonian Institution <gopher:// nmnhgoph.si.edu:70 / 0ftp%3Aonyx.si.edu @ protocols/MsatManV6.rtf>, Washington, DC.

Gomulski, L. M., K. Bourtzis, S. Brogna, P. A. Morandi, C. Bonvicini *et al.*, 1998 Intron size polymorphism of the Adh1 gene parallels the worldwide colonization history of the Mediterranean fruit fly, *Ceratitis capitata*. Mol. Ecol. **7:** 1729–1742.

Hardy, D. E., and M. D. Delfinado, 1980 *Diptera: Cyclorrhapha III, Series Schizophora Section Acalypterae, exclusive of Family Drosophilidae*. University Press of Hawaii, Honolulu.

Haymer, D. S., M. He and D. O. McInnis, 1997 Genetic marker analysis of spatial and temporal relationships among existing populations of the Mediterranean fruit fly. Heredity **79:** 302–309.

Hoelzer, G. A., 1997 Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. Evolution **51:** 622–625.

Hudson, R. R., 1994 How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained. Proc. Natl. Acad. Sci. USA **91:** 6815–6818.

Kareiva, P., 1996 Developing a predictive ecology for non-indigenous species and ecological invasions. Ecology **77:** 1651–1652.

Lathrop, G. M., A. B. Hooper, J. W. Huntsman and R. H. Ward, 1983 Evaluating pedigree data. I. The estimation of pedigree error in the presence of marker mistyping. Am. J. Hum. Genet. **35:** 241–262.

Lawton, J. H., and K. C. Brown, 1986 The population and community ecology of invading insects. Phil. Trans. R. Soc. Lond. B **314:** 607–617.

Lessa, E. P., 1992 Rapid surveying of DNA sequence variation in natural populations. Mol. Biol. Evol. **9:** 323–330.

Lodge, D. M., 1993 Biological invasions: lessons for ecology. Trends Ecol. Evol. **8:** 133–137.

Malacrida, A. R., C. R. Guglielmino, G. Gasperi, L. Baruffi and R. Milani, 1992 Spatial and temporal differentiation in colonizing populations of *Ceratitis capitata*. Heredity **69:** 101–111.

Marshall, T. C., J. Slate, L. E. B. Kruuk and J. M. Pemberton, 1998 Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. **7:** 639–655.

McGuigan, K., K. McDonald, K. Parris and C. Moritz, 1998 Mitochondrial DNA diversity and historical biogeography of a wet forest-restricted frog (*Litoria pearsoniana*) from mid-east Australia. Mol. Ecol. **7:** 175–186.

McPheron, B. A., G. E. Gasparich, H. Han and G. J. Steck, 1994 Mitochondrial DNA restriction map for the Mediterranean fruit fly, *Ceratitis capitata*. Biochem. Genet. **32:** 25–33.

McPheron, B. A., W. S. Sheppard and G. J. Steck, 1995 Genetic research and the origin, establishment, and spread of the Mediterranean fruit fly, pp. 93–107 in *The Medfly in California: Defining Critical Research*, edited by J. G. Morse, R. L. Metcalf, J. R. Carey and R. V. Dowell. University of California, Center for Exotic Pest Research, Riverside.

Metcalf, R. L., 1995a Biography of the Medfly, pp. 43–48 in *The Medfly in California: Defining Critical Research*, edited by J. G. Morse, R. L. Metcalf, J. R. Carey and R. V. Dowell. University of California, Center for Exotic Pest Research, Riverside.

Metcalf, R. L., 1995b The need for research on exotic pests in California, pp. 5–39 in *The Medfly in California: Defining Critical Research*, edited by J. G. Morse, R. L. Metcalf, J. R. Carey and R. V. Dowell. University of California, Center for Exotic Pest Research, Riverside.

Miller, M. P., 1998 *Tools for Population Genetic Analysis*, Northern Arizona State University, Flagstaff.

Moore, W. S., 1995 Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution **49:** 718–726.

New, T. R., 1995 *Introduction to Invertebrate Conservation Biology*. Oxford University Press, Oxford.

Ortí, G., M. P. Hare and J. C. Avise, 1997 Detection and isolation of nuclear haplotypes by PCR-SSCP. Mol. Ecol. **6:** 575–580.

Pääbo, S., and A. C. Wilson, 1988 Polymerase chain reaction reveals cloning artifacts. Nature **334:** 387–388.

Palumbi, S. R., 1996 Nucleic acids II: the polymerase chain reaction, pp. 205–247 in *Molecular Systematics*, edited by D. Hillis and C. Moritz. Sinauer, Sunderland, MA.

Palumbi, S. R., and C. S. Baker, 1994 Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. **11:** 426–435.

Rannala, B., and J. L. Mountain, 1997 Detecting immigration using multilocus genotypes. Proc. Natl. Acad. Sci. USA **94:** 9197–9201.

Raymond, M., and F. Rousset, 1995 An exact test for population differentiation. Evolution **49:** 1280–1283.

Roderick, G. K., 1996a Geographic structure of insect populations: gene flow, phylogeography, and their uses. Annu. Rev. Entomol. **41:** 263–290.

Roderick, G. K., 1996b Population genetic studies of tephritid flies of economic importance, pp. 267–271 in *Fruit Fly Pests: A World Assessment of Their Biology and Management*, edited by B. A. McPheron and G. J. Steck. St. Lucie Press, Delray Beach, FL.

Roderick, G. K., and R. G. Gillespie, 1998 Speciation and phylogeography of Hawaiian terrestrial arthropods. Mol. Ecol. **7:** 519–531.

Roderick, G. K., and F. X. Villablanca, 1996 Genetic and statistical analysis of colonization, pp. 281–290 in *Fruit Fly Pests: A World Assessment of Their Biology and Management*, edited by B. A. McPheron and G. J. Steck. St. Lucie Press, Delray Beach, FL.

Roderick, G. K., N. Davies, A. J. Bohonak and F. X. Villablanca, 1998 The interface of population genetics and systematics: invasion genetics of the Mediterranean fruit fly (*Ceratitis capitata*), pp. 489–499 in *Pest Management—Future Challenges. Proceedings of the Sixth Australasian Applied Entomological Research Conference*, edited by M. P. Zalucki, R. A. I. Drew and G. G. White. University of Queensland, Brisbane, Australia.

Schneider, S., J.-M. Kueffer, D. Roessli and L. Excoffier, 1997 Arlequin. http://anthropologie.unige.ch/arlequin, University of Geneva, Geneva, Switzerland.

Sheppard, S. W., G. J. Steck and B. A. McPheron, 1992 Geographic populations of the medfly may be distinguished by mitochondrial DNA variation. Experientia **48:** 1010–1013.

Siebert, J. A., 1994 Economic impact of an embargo by Japan, Korea, Taiwan and Hong Kong of selected California exports due to a Mediterranean fruit fly infestation. Dept. Agric. and Res. Econom., University of California, Berkeley.

Slatkin, M., 1994 Cladistic analysis of DNA sequence data from subdivided populations, pp. 18–34 in *Ecological Genetics*, edited by L. A. Real. Princeton University Press, Princeton, NJ.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*. W. H. Freeman and Company, New York.

Steck, G., G. E. Gasparich, H.-Y. Han, B. A. McPheron and W. S. Sheppard, 1996 Distribution of mitochondrial DNA haplotypes among *Ceratitis capitata* populations worldwide, pp. 291–296 in *Fruit Fly Pests: A World Assessment of Their Biology and Management*, edited by B. A. McPheron and G. Steck. St. Lucie Press, Delray Beach, FL.

Templeton, A. R., 1998 Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol. **7:** 381–397.

Templeton, A. R., and C. F. Sing, 1993 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics **134:** 659–669.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

U.S. Congress OTA, 1993 Harmful Non-Indigenous Species in the United States. U.S. Government Printing Office, Washington, DC.

Villablanca, F. X., G. K. Roderick and S. R. Palumbi, 1998 Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. Mol. Ecol. **7:** 547–560.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

White, I. M., and M. M. Elson-Harris, 1992 *Fruit Flies of Economic Significance: Their Identification and Bionomics*. CAB International, London.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey, 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:** 6531–6535.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Wright, S., 1965 The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19:** 395–420.

Communicating editor: M. Slatkin