Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns

F. X. VILLABLANCA, G. K. RODERICK, and S. R. PALUMBI

Abstract

Biological invasions generally start from low initial population sizes, leading to reduced genetic variation in nuclear and especially mitochondrial DNA. Consequently, genetic approaches for the study of invasion history and population structure are difficult. An extreme example is the Mediterranean fruit fly, *Ceratitis capitata* (Medfly), for which successive invasions during this century have resulted in a loss of 60% of ancestral genetic variation in isozymes and 75% of variation in mitochondrial DNA. Using Medflies as an example, we present a new approach to invasion genetics that measures DNA sequence variation within introns from multiple nuclear loci. These loci are so variable that even relatively recently founded Medfly populations within California and Hawaii retain ample genetic diversity. Invading populations have only lost 35% of the ancestral genetic variation. Intron variation will allow high-resolution genetic character-ization of invading populations in both natural and managed systems, although non-equilibrium methods of analysis may be necessary if the genetic diversity represents sorting ancestral polymorphism.

Introduction

Invasions of exotic pests and pathogens are a growing threat to agriculture, human health, and the natural environment, mainly as a result of increasing global commerce and travel (Miller *et al*. 1992; US Congressional Office of Technology Assessment 1993; Metcalf 1995b; Schrag & Wiener 1995; Roderick & Howarth 1997). While many terrestrial invaders are insects, invasive species include other invertebrates, vertebrates, plants, and pathogens (Perring *et al*. 1993; Schiff & Sheppard 1993; Schrag & Wiener 1995). Colonizing populations of these invaders are usually founded by only a few individuals representing a minute sample of the source population (Elton 1958). Genetic consequences of such colonization events include the well-known phenomena of founder events and genetic bottlenecks (Lande & Barrowclough 1987; Nei *et al*. 1975).

Because of the low genetic variation following invasions, genetic tools have been of limited use in studying invasion biology. In particular, mitochondrial DNA (mtDNA) is poorly suited to studies of invasions unless the invading population is large or grows rapidly. This is because mtDNA is subject to strong genetic drift due to its maternal and haploid mode of inheritance. Nuclear genetic variation is retained for a longer period of time (Neigel & Avise 1986), but identification of variable DNA sequences in the nuclear genome can be difficult.

There are at least three major PCR-based techniques for understanding nuclear genetic variation in populations. The first, random amplified polymorphic DNA (RAPD) has the advantage of requiring little prior knowledge and short development times. However, RAPD data suffer from problems of interpretation and repeatability in ways that make their effective use cumbersome and time-consuming in many situations (Grosberg *et al*. 1996). Microsatellites have been an effective means of under-standing genetic variation in small populations (Queller *et al*. 1993; Strassman *et al*. 1996) but require a long phase

of technique development in new taxa (but see Moore *et al*. 1991; Glenn *et al*. 1996). Although many loci can be identified and analysed, the alleles observed in a population cannot be related to one another phylogenetically unless the underlying mechanism of mutation and the mutation rate are known. Therefore, geographical origins of individual microsatellite alleles are difficult to determine without a phylogeographic analysis. Population differentiation is examined by testing for shifts in allele frequency and variance (Slatkin 1995), yet this requires the samples from an invading population to be pooled, making it impossible to recognize when populations have been founded from multiple sources (Roderick & Villablanca 1996). In addition, DNA sequence data may be required to demonstrate homology among similar-sized alleles when comparing allele frequencies (Garza *et al*. 1995). The third approach to understanding nuclear genetic variation is the use of exon-priming intron-crossing (EPIC) PCR to amplify introns from targeted protein-coding genes (Lessa 1992; Lessa & Appelbaum 1993; Palumbi & Baker 1994; Slade *et al*. 1994). Although this technique is labour intensive (because PCR products must be cloned to separate alleles in diploids before sequencing) and requires troubleshooting (e.g. primer design and sequence analysis to identify homologous loci), the DNA data provide a great deal of information about particular alleles within populations and allow extensive phylogenetic and coalescence-based analyses of allelic variation. Techniques such as single-stranded conformational polymorphism (SSCP, Hare *et al*. 1996) can be used to identify homozygotes or unique alleles and thus avoiding cloning.

Introns are presumed to evolve more quickly than corresponding nuclear protein coding regions because they are subject to fewer functional constraints (Li 1997). Introns are also expected to retain variation due to their diploid and biparental inheritance. Unlike genetic variability revealed by RAPDs and microsatellites, intron sequence variation can be subjected to phylogenetic reconstruction, cladistic analysis of gene flow, as well as standard population genetic and coalescence analysis of alleles (Palumbi 1996a; Roderick 1996a). Even though variation in invading populations may be pre-existing variation (sorting ancestral polymorphism), a large amount of variation could allow for unique sorting events in different populations.

Here, we explore DNA sequence variation within multiple nuclear introns loci as a tool with which to study the genetics of invading populations. The Mediterranean fruit fly (*Ceratitis capitata*) is a classic example of this problem and represents a 'worse-case' scenario for the genetic study of invasions. While a genetic approach to understanding the invasion biology of the Medfly has been considered both essential and necessary (Carey 1991, 1996; Metcalf 1995b), little genetic variation has

been uncovered within invading populations with both allozymes (Gasparich *et al*. 1991; Malacrida *et al*. 1992; Baruffi *et al*. 1995) and mtDNA (Gasparich *et al*. 1997). For example, a sample of over 1200 flies collected worldwide revealed six mtDNA haplotypes in this species, with only four haplotypes found in invading populations (Gasparich *et al*. 1997). In contrast to these observations, the intron data presented here reveal a wealth of genetic variability within invading populations. Intron sequence diversity in invading and noninvading populations is substantially greater than the genetic diversity in allozymes or mtDNA.

Materials and methods

Specimens

Alleles were sampled from populations in Africa and California, Hawaii, Brazil and Greece (Table 1). Individual flies from Greece, Brazil, and California were obtained from the USDA APHIS Medfly Germplasm Repository coordinated by D. Prasher (Otis, MA). Flies from Hawaii were collected by T. Shelly (University of Hawaii) and D. Foote (Volcanoes National Park). DNA from these invading populations was isolated using a phenol–chloroform extraction and ethanol precipitation. DNA from African flies and the outgroup was obtained from the laboratory of B. McPheron (University of Pennsylvania State). The African population represents a noninvading population that exists on native hosts (Steck *et al*. 1995). The outgroup (for phylogenetic analysis) is a single specimen of *Ceratitis rosa*, a sister species to the Medfly.

Enzymatic amplification

Exon-primed, intron-crossing (EPIC) primers (Palumbi & Baker 1994; Palumbi 1996a,b) were constructed to amplify introns from within several different nuclear loci of the Medfly (Fig. 1). We selected single-copy nuclear genes in which the intron positions are conserved across species. For members of multigene families, primers were designed to preferentially amplify a single locus. Primer annealing is well within the exons such that locus homology can be established by comparing intron position and protein-coding sequences to known homologous and nonhomologous genes. This approach allows for the potential recognition of nonspecific amplification products (multiple gene copies or pseudogenes). Reference sequences for primer design were obtained from the literature (Konsolaki *et al*. 1990; Rina & Savakis 1991; He & Haymer 1992; Kwiatowski *et al*. 1992). The loci investigated were muscle-specific actin (Actin) intron 1, chorion s36 (s36) intron 1, vitellogenin 1 gamma (Vg1) intron 2, and Cu/Zn superoxide dismutase (SOD)

Table 1 Nuclear intron alleles and genotypes across multiple loci for Medfly (*Ceratitis capitata*) specimens from invading and noninvading populations. Intron allele names correspond to Fig. 3 and were arbitrarily chosen following alphabetical order. For individuals sampled to date, underlined alleles indicate alleles unique to that region worldwide. Known heterozygotes have two alleles at each locus. One possible actin allele and two possible chorion alleles were rejected based on an application of the singleton criteria to cloned PCR products

Individual	Region	Locality	Collecting date	ID No.	Independent loci			
					SOD	Chorion	Vitellogenin	Actin
	Invading							
1	California	LA Co: Duarte: 2222 Wesley Gr. A	92 Sept. 24	CA101.1	G,L	D	E	H
2	California	LA Co: Granada Hills, W Index St.	93 July 26	CA350	M	D	E	F
3	Hawaii	Maui: Hashimoto Farm	94 Jan. 31	Ma84	E	F,G	E	B,C
4	Hawaii	Big Island: Kipuka Ki	94 Jan. 25	Ha95	F	C	F	B,F
5	Brazil	Petrolina: Pernambuco	93?	BR004.04	H	A	E	B
6	Greece	Island of Chios	93 Aug. 15	GR001 A.010	H,I	A	C,I	F
7	Greece	Southern Peloponnese	83? Lab Col.		F	A	E	–
	African							
8	Africa	Malawi: Monkey Bay	95 Feb.	95042009	A,B	A	A,G	A
9	Africa	Malawi: Monkey Bay	95 Feb.	95042011	N,O	B	D	E
10	Africa	Malawi: Monkey Bay	95 Feb.	95042012	D,K	A	B	G,H
11	Africa	Malawi: Monkey Bay	95 Feb.	95042015	C,J	E	H	D

intron 1. Nomenclature and homology assessments follow Fyrberg *et al.* 1981; Sanchez *et al.* 1983; Konsolaki *et al.* 1990; Tolias *et al.* 1990; Rina & Savakis 1991; Kwiatowski *et al.* 1992.

PCR reactions were performed in a 25 µL volume including: 16 µL of water, 2.5 µL of 8 mM dNTPs, 2.5 µL of 10× *Taq* buffer, 1.2 µL of each 10 mM primer, 0.06 µL of Perkin Elmer *Taq* polymerase (5 U/µl) and 1 µL of genomic DNA (*Taq* buffer as in Palumbi 1996b). Thermocycler parameters were: 94° for 30 s, 55° for 30 s and 72° for 40 s, repeated for 40 cycles. Parameters for SOD were 94° for 30 s, 60° for 1 min and 72° for 1.5 min for 40 cycles.

Cloning and sequencing

The PCR products in any reaction may include alleles from the target locus as well as products from other loci. If both alleles from a heterozygote were sequenced simultaneously then ambiguous sequences would result. Because of diploidy, and the potential for heterozygosity, PCR products were cloned in order to isolate each allele before sequencing. PCR products were ligated into a TA cloning vector (pBluescript KS–). Colonies with inserts were screened by PCR to determine insert size. Single-stranded DNA (for sequencing) was generated by helper-phage infection. DNA for sequencing was recovered by a 20% PEG/2.5 M NaCl precipitation followed by phenol–chloroform extraction and ethanol precipitation. Complete details of the cloning procedure are presented by Palumbi & Baker (1994). Single-stranded DNA was sequenced manually using a radiolabelled nucleotide, the sequenase kit (Stratagene), 7 µL of template DNA, and either a cloning primer (modified from Stratagene's T3) or an internal sequencing primer (Fig. 1). Sequences were also obtained using the ABI 373 automated sequencer and 60 ng (Vg1, s36 and Actin) or 250 ng (SOD) of single-stranded template. Sequences were aligned by eye in SeqEdit (version 1.0.3). Alignments were entered into HEAPBIG (a program written by and available from S. R. Palumbi) in order to identify variable sites and make pairwise comparisons for percentage sequence difference.

Data analysis

It is well recognized that cloning reveals the errors incorporated into DNA by DNA polymerases (Pääbo & Wilson 1988): *Taq* polymerase is known to have a misincorporation (error) rate of about 1:1000 bases, so it can be expected that different clones from the same PCR reaction will differ by about this amount. One way to distinguish real alleles from *Taq* errors is to obtain and sequence multiple clones from each PCR reaction. Using this approach in heterozygotes normally requires sequencing at least three clones to verify one allele and at least three clones to verify the second. Thus at least six sequences would be required per individual. If both alleles are not equally represented in six clones it is possible that more clones would need to be sequenced. As an alternative approach we chose to sequence between one and three clones per individual and then identify and remove singletons.

A) Superoxide dismutase

```
 M  V  V  K  A  V  C  V  I  N  G  D  V  K  G  T  V  H  F  E  Q  Q
ATGGTGGTAAAAGCTGTATGCGTAATTAACGGAGACGTTAAGGGAACCGTTCACTTTGAACAGCAG>>>G-70
TAGGAAATAAATTTTCATGTTTTCTTGGTATTTTCGAGGTCCAAGGTCATGGCACCGGTTGACATAGTAG-140
AGCATTACATGTACTATCACAATTATGGAGTAAGAGACGTTATTTTATTTACGGTATTATACGCAAAAT-210
ACCGATCATAATATATCCAAATATTGTAATTTCAATCATCGTAAATAATATGCCTCCAAAAAATTTGGAT-280
AATTATCTTCTCAATTTAACAGCGTTAATTTGTGTGTGGGGTGTAGGTGAGGTAGATGAGATGTACTTC-350
GTCATTGCGGTTAAGTGGTATGATTTATTCAATGATAGCGAAAAATTAAGTATTTCGTTATCATTTGTAT-420
AAAAAAGATAATGGAAAATTGAAGAAGTTTATTGTATTTAAAAATTTATTATTTTTTCGAGGTTTTATAA-490
TTTAAACTTATTTCTTTATGCGGTTATTTGTATATTGTTAATGTGTTGCTTTTTGAGACGGTATCACTGA-560
TAATGAACCTATTTATTTCCAATTGAAATTGTATACGGTATGAGAGAGGAGTTCGTATGCATATATTAT-630
ATGTATGTATGTACTTTATTAGCATGAATTAGCACTTTTTTAATTTTTTATTTCATTTGTAACGTGAGTTA-700
             D  A  K  S  P  V  L  V  T  G  E  V
TTTGTATTGCTTATTTTTTATTTTCTCATTTAAG<<GATGCAAAGTCTCCCGTCCTAGTCACTGGTGAGGT-770
   N  G  L  A  K  G  L  R  G  F  H  V  H  E  F  G
TAACGGACTTGCTAAGGGTCTGCGTGGCTTCCATGTACACGAATTCGGG-819
```

B) Vitellogenin

```
 N  G  N  K  D  Y  D  Y  G  S  S  Q  G  N  Q  G  T  T  S  S  E  E  D
AATGGCAACAAGGACTACGACTACGGCAGCAGCCAAGGTAATCAAGGTACCACCTCAAGCGAAGAAGATT-70
 Y  S  E  S  W  K  N  Q  K  R  T  K  G  N  L  V
ACAGCGAATCCTGGAAGAACCAGAAGAGAACCAAGGGCAACCTTGTC>>>GTAAGTTTTATAACAAATAG-140
                                         I  I  N  L  G  A
CTTTTAATTTTCCACCTGCATTACAACTGTTTATTTTCTTTTGCTCAG<<<ATCATAAACTTGGGCGCTA-210
 T  L  S  N  M  K  R  F  A  L  L  D  V  E  Q  T  G  N  M  I  G  K  T  L
CGCTCTCAAACATGAAACGCTTCGCGCTCCTCGATGTGGAGCAAACCGGCAACATGATCGGCAAAACTCT-280
   V  Q  L  T  N  E  V  D  V  P  Q  E  I  I  H  I  V  A  Q  G  I  G  A
AGTTCAATTGACCAACGAAGTTGATGTCCCACAAGAGATCATCCACATTGTAGCCCAAGGTATTGGTGCT-350
 Q  V  A  G  A  A  G  R  Q  Y  K  R  L  T  G  H  Q  L  R  R  I  T  A
CAGGTTGCTGGCGCTGCTGGTCGTCAGTACAAACGTTTGACTGGACATCAACTCCGCCGTATCACCGCTT-420
 L  D  P  S  K  I  F  A  K
TGGATCCCTCCAAAATTTTCGGCCCAA-446
```

C) Chorion

```
                 M  K  C  F  L  F  T  L  F  F  G  G  A  P
GATCAGAGAGATCGTCATAAAATGAACCTGTTTCTATTCACACTTTTCTTCGGTGGCGCACCG>>>GTAC-70
GTGTACTCAATATTAATACAAACACATTCATTGAGAATCCGAAAAGAAAAATAAAGATTAGAAATTAAAA-140
CGCACACACTCACTTACACAGCGCATTGCATCAAAGAAGTTCCATACTAATTGTAAATGTTTCTTCTTT-210
            L  A  T  A  S  Y  G  S  S  S  G  G  G  G  G  ?  S
CGTCTTTATAAAG<<<<CTCGCAACAGCCTCTTACGGCTCTTCTTCCGGTGGCGGTGGTGGTGGNCAT-280
 S  Y  L  S  S  A  S  S  N  G  L  D  E  L  V  Q  A  A  A  G  G  A  Q  Q
CGTATTTATCCTCTGCCTCATCGAACGGTCTCGATGAATTAGTACAGGCTGCTGCTGGTGGTGCACAACA-350
   A  G  G  T  I  T  P  A  N  A  E  I  P  V  S  P  A  E  V  A  R  L  N
AGCTGGCGGTACCATTACGCCCGCTAATGCTGAAATACCCGTCTCACCCGCCGAGGTGGCTCGCCTCAAT-420
 Q  V  Q  A  Q  L  Q
CAAGTGCAAGCTCAATTGCAAGC-443
```

D) Actin

```
 I  H  E  T  V  Y  Q  S  I  M  K  C  D  V  D  I  R  K  D  L  Y  A  N
ATTCACGAGACTGTCTACCAGTCAATCATGAAGTGCGACGTGGACATCCGTAAGGACTTGTACGCCAACA-70
 N  V  L  S  G  G  T  T  M  Y  P
ACGTACTGTCCGGCGGCACCACCATGTACCCAG>>GTAAGAGGAGCCAATTTCATTTAATTGCAGCGGCT-140
          G  I  A  D  R  M  Q  K  E  I  T  A  L  A  P
TAATTCCTTTACTTACGTTCCTAG<<GTATTGCCGATCGCATGCAAAAGGAAATCACTGCCTTGGCTCCA-210
 S  T  I  K  I  K  I  I  A  P  P  E  R  K  Y  S  V  W  I  G  G  S  I
TCGACCATCAAGATTAAGATCATTGCACCACCTGAGCGTAAATACTCCGTATGGATTGGTGGTTCTATCT-280
 L  A  S  L  S  T  F  Q  Q  M  W  I  S  K  Q  E
TGGCTTCGCTCTCCACCTTCCAGCAGATGTGGATCTCTAAGCAGGAGTA-329
```

Fig. 1 Nucleotide sequences for multiple EPIC loci in an African Medfly (no. 95042009). The 5′ intron boundary is marked by > and the 3′ boundary by <. Translations for exon sequences are shown by the single-letter amino acid code. Primer sequences are underscored. A. Cu–Zn SOD including intron one. The first codon in this sequence is amino acid position 1. B. Vitellogenin 1, gamma including intron two. The first codon in this sequence is amino acid number 166. C. Chorion s36 with the first amino acid being the start codon. D. Muscle-specific actin including intron 1. The first amino acid is codon position 274.

Singletons are polymorphic character states that occur in only one sequence of an alignment. Singletons appear in two forms: (i) a particular nucleotide position that is variable in only one sequence; or (ii) a unique character state in a variable position. Of course, not all singletons are *Taq* errors, but considering them to be so results in a conservative measure of allelic diversity.

Singletons were removed as follows. In the alignment all polymorphic sites were verified against the raw data. Sequences in the completed alignment were then compared. A polymorphism that exists in only one sequence in the entire data set (singleton) is changed to match the consensus sequence. This change is marked by a different case in the alignment. If later sequences show that a singleton is actually a polymorphism, then the edited sequence would be changed back to the original base. This approach, similar to that for estimates of heterozygosity below, will produce a slight underestimate of the genetic diversity in introns, but does not change the conclusions drawn here.

Genetic diversity

To assess the level of variability in introns relative to mtDNA and allozymes we used a measure of diversity that allows comparisons between haploid and diploid genomes. Our present purposes only required that we determined levels of variation between populations and not necessarily within or between individuals. In particular, we were interested in the levels of variation shown in invading vs. ancestral populations of Medflies. Expected heterozygosity was thus estimated within populations ($1 - \sum p_i^2$, where p_i equals the frequency of the ith allele) rather than as an observed individual heterozygosity measure averaged across individuals. As only allele frequencies were required (p_is) it was not necessary to identify and sequence both alleles of an EPIC amplification. This estimate of intron genetic diversity might be biased downward if we sequenced the same allele in all clones from a single individual and erroneously estimated a heterozygote to be a homozygote. Only in chorion did we sequence multiple clones (< 3) from single individuals and find that all clones were identical. Therefore only for chorion might our estimate of heterozygosity be low for this reason. Our Medfly intron results were compared to estimates of heterozygosity for all allozymes by Baruffi *et al.* (1995) and mtDNA by Gasparich *et al.* (1997). In order to make these studies more comparable in sample size we have included, in the estimates of heterozygosity, recently collected EPIC sequences that allowed us to triple the number of individuals (F. X. Villablanca *et al.* unpublished).

Phylogenies of alleles

DNA sequences of alleles at homologous loci can be subjected to phylogenetic reconstruction resulting in an 'allele tree,' providing that recombination has not

contributed to the history of those alleles. However, recombination is a hallmark of eukaryotic genomes and may have contributed greatly to nuclear genetic diversity within populations (Begun & Aquadro 1992; Hudson 1994). As a result, phylogenetic analyses of nuclear genes must consider the effects of recombination. In addition, it is possible that 'ancestral' alleles may still be present in populations (Templeton & Sing 1993). In a phylogenetic sense these alleles are nodes or steps along the branch and not branch tips: they are ancestors to alleles in the tree. In phylogenetic analyses these ancestral alleles result in polytomies (multifurcations), not because of lack of resolution or because the tree is not dichotomously branching, but because phylogenetic analyses generally do not recognize extant sequence as nodes. Clearly, a method is needed that explicitly allows for extant ancestral sequences and recombination.

The network method of Templeton (Templeton *et al.* 1992; Templeton & Sing 1993) provides a means for recognizing and dealing with alleles produced by recombination. In addition, this method does not assume all alleles are branch tips, but allows alleles to be placed at nodes and along branches. For each locus, a test was performed to determine whether polymorphism rates per site (due to recombination and/or multiple mutations) were sufficiently low to justify a parsimony model (Hudson 1989; Templeton & Sing 1993; program kindly provided by A. Templeton). If a parsimony model is justified (back-mutations and recombination are considered rare), then a parsimony analysis was performed (Swofford 1993). Following this analysis the allele tree was redrawn as a network, thus allowing extant sequences (identified by zero length branches) to be nodes and internodes. If the parsimony model was rejected then the network was constructed following the linkage algorithm of Templeton *et al.* (1992). Briefly, this algorithm begins with the connection of alleles that show no parallel changes or reversals, then alleles of ever-increasing divergence are connected, until finally alleles might be found that can be parsimoniously connected to multiple alternative sites on the network. If for these most-divergent alleles it is more parsimonious to assume a single recombination event (rather than multiple parallelisms and reversals), then a recombination event is accepted. Recombination implies that an allele does not have a single history, but rather that parts of the allele have separate histories. A recombinant sequence therefore cannot be connected to only one branch of a network because portions of the allele belong to different branches. These multiple connections would look like triangles or boxes and represent reticulations.

It is possible to interpret the phylogeny of alleles at three hierarchical levels: alleles, individuals and populations. The mutational network indicates how alleles are hierarchically related within lineages. If alleles are related by recombination then the lineages reticulate, otherwise, mutation defines linear or bifurcating hierarchies. This hierarchical level is relevant to our analysis as it identifies sorting ancestral polymorphism (i.e. no monophyly of invading alleles, or invading alleles are ancestral to non-invading alleles). The hierarchical nesting of alleles within individuals would require multiple constraints. Recombination makes it such that the relatedness of alleles at a single locus does not imply particular degrees of relatedness among individuals, unless individuals are geographically subdivided and represent discrete 'recombination pools.' Alleles might be nested within individuals across loci in the presence of linkage. Unfortunately, our limited samples did not allow us to explore the possibility of recombination pools or linkage. The third hierarchical level is pertinent to the present case. Alleles or clades may be examined relative to their geographical locations (i.e. phylogeography). If this is the case, comparisons across loci allow independent tests of the observed pattern.

Results

PCR products

All the EPIC primer sets produced the expected PCR products including the targeted introns (Fig. 1). A single PCR product was produced from each amplification, except for vitellogenin (which occasionally produced a product 60 bp larger than the target; this product was not analysed). Only clones with the appropriate size inserts were sequenced. Sequencing produced only the expected results: sequences were easily aligned with Medfly references and contained introns in the expected positions. None of the Medfly sequences contained stop codons, frame-shifts, or an excess of replacement changes suggestive of pseudogenes or multigene families. Insertions/deletions were required for the alignment of SOD and Vitellogenin data sets, but only within intron portions of the sequence. An application of the singleton criterion caused us to recognize 1.9 errors per 1000 bases, or an average of one error every three sequences. Following this criterion we rejected a single actin clone and two chorion clones as putative alleles. The sequences obtained from four loci in one African individual are shown in Fig. 1. The sequences of all alleles are available from GenBank.

Homology

Identification of single-copy genes and confirmation that EPIC–PCR is assessing homologous loci in different individuals is an important part of the implementation of this

technique (Palumbi & Baker 1994; Slade *et al.* 1994). For our Medfly loci, homology was confirmed by comparing the exon sequence and intron position to known sequences (Fig. 1). Because some of these loci are duplicated (see below) we need to recognize homologous loci that are also orthologous (of one lineage to the exclusion of other lineages derived from gene duplications). Orthology can be inferred from a combination of amino acid similarity, intron position, and spatial, temporal and sex-limited tissue-specific expression. These inferences are especially robust when the existence of a locus has been determined by homologous probes and evaluation of all hybridized products as is the case for each of the loci considered here (Konsolaki *et al.* 1990; He & Haymer 1992; Kwiatowski *et al.* 1992 and references therein; Rina & Savakis 1991).

Muscle-specific actin

Three actin loci have been identified in Medflies (Haymer *et al.* 1990). Amino acid sequence, intron position and spatial and temporal expression patterns confirm that there is only one form of the muscle-specific actin in Medflies (He & Haymer 1994). The intron position differs between the muscle-specific form and the other two forms. There are also multiple fixed differences between these three genes (i.e. Vanderkerckhove & Weber 1978; Fyrberg *et al.* 1981), some of which are in the exon sequences amplified using EPIC primers. Our primers targeted the muscle-specific actin. The resultant sequences showed an intron position and exon sequence that allow us to conclude that all the sequences are orthologous muscle-specific actin.

Chorion s36

Orthology can be inferred for the chorion sequences based on a combination of characteristics: amino acid similarity, intron position, expression in choriogenic-stage follicular cells, and during early or late oogenesis (Konsolaki *et al.* 1990; Tolias *et al.* 1990). The chorion genes of Medflies have been compared to those of *Drosophila melanogaster* (Tolias *et al.* 1990) leading to the discovery of four chorion-like genes in Medflies. Two of these genes are expressed late in oogenesis and have low similarity to known *D. melanogaster* genes. The other two are expressed early in choriogenesis and are homologous with *D. melanogaster* loci (s36 and s38). Both s36 and s38 loci have a large 'central domain' that is distinctly different between the genes, but highly conserved between species (i.e. only four of 85 amino acids are different between *D. melanogaster* and Medflies s36 genes). Our EPIC sequences allows differentiation between the Medfly s36 and s38 loci. All sequences obtained are homologous to Medfly s36.

Gene conversion is a potential complicating factor in the analysis of chorion genes (Regier *et al.* 1994). It has been found that gene conversion occurs in all but one of the chorion gene families and produces homogenization among the sequences of different loci. The result is a linked array of nearly identical genes. Two of the Medfly chorion genes are localized on one chromosome (Zacharopoulou *et al.* 1992). Yet, it is likely these loci are not evolving by interlocus gene conversion because the sequences of both loci can be readily distinguished (Konsolaki *et al.* 1990; Tolias *et al.* 1990). It is most probable that concerted evolution, if it occurs in Medfly s36, would occur between alleles within a locus (intralocus gene conversion). Even with small sample sizes we can reject the suggestion that this process is common in Medfly s36: there are multiple alleles present within populations, and heterozygosity is comparable to that seen at other loci. One of the early choriogenesis chorion genes (ErB) does not evolve through concerted evolution (Regier *et al.* 1994). The locus for which we have designed our primers is also expressed during early choriogenesis (Konsolaki *et al.* 1990).

Cu/Zn superoxide dismutase

The three forms of SOD (Cu–Zn, Mn and Fe) appear to be unrelated based on amino acid differences, different metal cofactors, and distinct crystal structures (Parker & Blake 1988; Tainer *et al.* 1982). We designed primers for the Cu–Zn SOD locus. It has a conserved intron position and distinct amino acid sequence making it distinguishable from other forms. The Cu–Zn gene was duplicated early in eukaryote history, raising the possibility that the second (extracellular) form could exist in any lineage (Smith & Doolittle 1992). The Medfly is one of the few eukaryotes within which the second Cu–Zn gene is known (Banks *et al.* 1995) These two forms can be easily distinguished based on amino acid sequence (including the exon fragments studied here). All of our sequences are orthologous to the intracellular Cu–Zn SOD.

Vitellogenin 1 gamma

Four single-copy vitellogenins have been identified in Medflies (Rina & Savakis 1991; Terpstra 1988; Martinez & Bownes 1994). The Medfly vitellogenins have all been sequenced at least in part (Rina & Savakis 1991) and in the region under study in the present study all four Medfly genes can be distinguished. Homology can be inferred between Medfly vitellogenins Vg1-gamma and Vg2-delta and *D. melanogaster* YP3 and YP1, respectively, based on sequence similarity, expression, and intron position (Rina & Savakis 1991). The other Medfly vitellogenins, Vg1-alpha and Vg2-beta, are recent duplicates of Vg1-gamma

```
Vg1-gamma ATTCACTTTCAACCTGCATAACAGCAGTTTTTTTTATTTTCCTCagATCATCAAC
          |..................|...|...........|.....|............|.
Vg1-delta TTTCACTTTCAACCTGCATGACAACAGTTTTTTTTCATTTTGCTCagATCATCATC
```

Fig. 2 Sequence distinction between vitellogenin 1 gamma and Vg 1 delta (from Rina & Savakis 1991). The first position in this alignment is 21 bases 3′ of the 5′ intron boundary. The 3′ intron boundary is shown in lower case (ag). The last three bases are in amino acid position 207 and result in a replacement change from Asparagine (AAC in Vg1-gamma) to Isoleucine (ATC in Vg1-delta). All of the Medfly sequences in our analysis have the Asparagine codon as in Vg1-gamma.

and Vg2-delta, respectively. Both of the Vg2 vitellogenins (delta and its recent duplicate beta) lack the intron that we amplified. Vg1-gamma, and the recent duplicate Vg1-alpha, differ at multiple positions. Some of these differences can be easily discerned in our EPIC fragments, including five differences between the introns and one replacement change 3′ of the intron (Fig. 2). All of our sequences match that of Vg1-gamma at all distinguishing sites.

Pseudogenes and nonhomology

Our data show no evidence for the existence of pseudogenes for the proteins and enzymes we have investigated. It is always possible that very recent duplication events could produce undetected pseudogenes. Primer design and specificity seems to have been sufficiently accurate to prevent the production of mixed PCR products. In addition, all of the sequences generated are only minor variants of those previously published for the four Medfly loci. The singularity of PCR products and the resultant patterns of sequence variation strongly suggest that each of our four data sets contains only sequences homologous to known loci.

Allelic diversity

Significant genetic variation was found at each of the four intron loci, even for the loci with small (< 80 bp) introns

(Table 1). With all populations pooled, every locus had multiple alleles. The African population produced multiple alleles for every locus. In over half of the invading populations, allelic diversity was discovered in two or more loci, despite the small intrapopulation samples. On average, 33% of the clones we sequenced were unique alleles (Table 2). This is a remarkable result given that we controlled for *Taq* error by removing singletons. The locus with the longest intron (SOD) produced the greatest allelic diversity, as might be expected if intron sequence variation is neutral. Yet, even small introns show allelic diversity in this species.

Overall, 50% of our sequence data is from exons, which contain only about 16% of the variable sites. Alleles differed at an average of 3% of their nucleotides, with most of the variation occurring in introns (Table 2). For example, nucleotide diversity (the mean sequence difference among alleles) is three times higher for intron sequences alone compared to when exons are included. Among introns, average percentage of variable sites varies widely, with the vitellogenin intron showing 14% and chorion showing 2%.

Chorion is an exception: in this gene the intron has nearly the same percentage variable sites as does the coding region (Table 2). In other studies of nuclear DNA it has been observed that, as in Medfly chorion s36, levels of variation in introns do not always exceed those of exons (Kreitman 1983; Kliman & Hey 1993). This result may be due to differential selection on coding and noncoding regions, or it reflects greater than expected constraints on supposedly neutral sequences. There are no amino acid substitutions between coding regions in the Medfly s36 alleles, and the ratio of replacement to silent substitutions (compared to the outgroup *Ceratitis rosa*) is 3.2%/13.9%. The percentage sequence difference between congeners (*C. capitata* and *C. rosa*) is approximately the same for introns (11–12%) as for silent sites (13–17%), although the difference among silent sites is a little higher. Thus, there is probably no excess of replacement substitutions in exons, but rather an apparent reduction of mutations or substitutions in introns. If the greater variability in introns

Table 2 Variable introns examined in the Medfly. Identities of introns and loci were confirmed by comparison of exon sequences and intron positions to published sequences. A dash (–) indicates that the sequence is too short to estimate a value accurately

	Count		Length (bp)		% Variable sites		Nucleotide diversity		
Locus	Sequences	Allele	Total	Intron	Coding	Intron	Total	Exon	Intron
Cu, Zn SOD, Intron 1	17	15	778	673	0.95	4.90	0.01122	–	0.01287
Chorion s36	16	7	386	157	2.18	1.91	0.00727	0.00773	0.00667
Vitellogenin 1, Gamma, Intron 2	13	9	395	70	1.85	14.29	0.01477	0.00828	0.05284
Actin, Muscle Specific, Intron 1	13	8	277	59	0.36	10.17	0.01038	–	0.04567
Mean	15	10	459	240	1.34	7.82	0.01091	0.00800	0.02951

seen at other loci is in large part a function of recombination shuffling a larger proportion of neutral or nearly neutral mutations, then a decrease in the recombination rate may account for the decrease in intron diversity. Chromosome position is correlated with the level of polymorphism in *Drosophila* (Begun & Aquadro 1992) and may play a role in reducing chorion variability in Medflies. Alternatively, concerted evolution or purifying selection may be reducing the number of alleles present in a population and therefore the potential for recombination among alleles. Of these two, concerted evolution is the least likely given that the percentage sequence difference among exons is about 1.2 times than of introns both within and between species. If concerted evolution were a major factor, then it should act within but not between species.

Phylogenetic analysis

Templeton's network method allows for the reconstruction of phylogenies from potentially recombining DNA fragments (Templeton *et al.* 1992; Templeton & Sing 1993). This method requires a test (Hudson's H) to determine whether a phylogeny will require a substantial number of parallel or reverse mutations or recombination events. Multiple substitutions at any site are predicted if the per site probability of change is sufficiently large ($P > 0.05$ that

Table 3 Estimate of Hudson's H (probability that phylogenetic reconstruction of relationships among alleles will need to invoke multiple mutations or recombination)

Locus	Number of sequences	Hudson's H		
		Total	Exon	Intron
SOD	17	0.029	–	0.034*
s36	16	0.018	0.020	0.017*
Vg1	13	0.035	0.020	0.118
Actin	13	0.026	–	0.108

* $P < 0.05$.
Probabilities of < 0.05 justify phylogeny reconstruction using parsimony. Probabilities > 0.05 indicate that nonparsimonious reconstructions (multiple mutations at a site, or recombination) are needed to explain the variation in alleles and justify a network approach (Hudson 1989, Templeton *et al.* 1992; Templeton & Sing 1993). A dash (–) indicates the sequence fragment is too short for our calculation to be accurate.

a single nucleotide difference between two randomly chosen alleles is due to more than one mutation). Two of the Medfly loci had substitution rates low enough to justify parsimony reconstructions (Table 3, Hudson's H with $P < 0.05$ for SOD and Chorion s36). The networks for these two loci were produced by redrawing the parsimony tree,
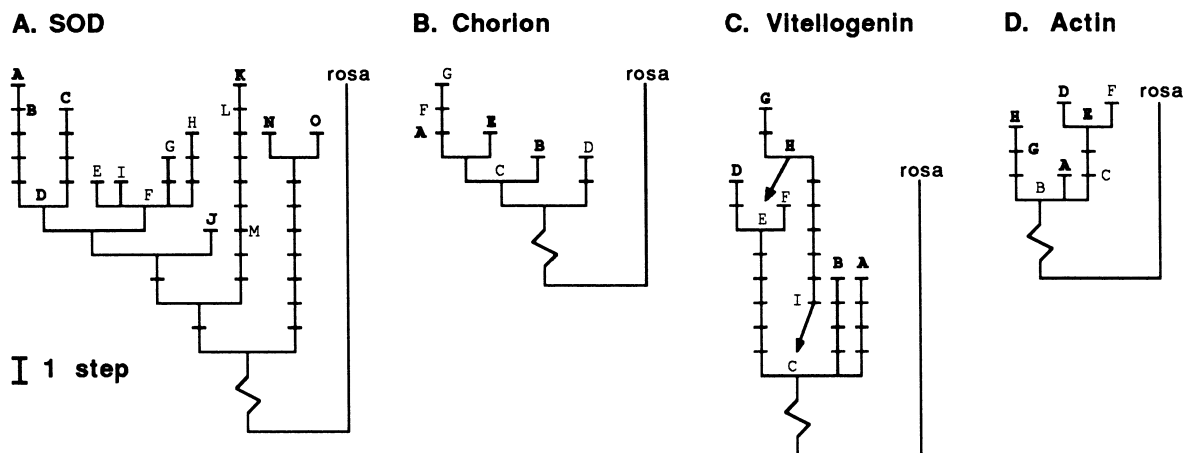
Fig. 3 Allele networks for multiple nuclear loci in the Medfly. Networks are constructed following the method of Templeton *et al.* (1993). The network is a redrawn (where sequences can fall along the branches) parsimony network (if parallelisms, reversals and recombination are estimated to be rare, Table 3), or it is produced following the linkage algorithm described by Templeton *et al.* (1993) (when a parsimony model is rejected). In all cases, except vitellogenin, Templeton's algorithm produced a tree identical to one of the equally parsimonious trees from parsimony analysis. This is because, generally, pairwise distances among sequences match the number of changes between sequences as inferred from the parsimony tree. The placements of roots were thus determined from parsimony analysis, using *Ceratitis rosa*, as the outgroup. Branch lengths are drawn to scale except along the branch leading to the outgroup. Two separate networks in vitellogenin indicates that some of the alleles could not be placed into a phylogenetic context because they are inferred to have arisen by recombination and thus are not derived from any single allele. For this locus, the number of pairwise differences between alleles do not equal the number of differences that would be inferred from the parsimony analysis. The arrows indicate possible relationships between alleles in the two vitellogenin networks. The geographical sources of all alleles are given in Table 1. African alleles are shown in bold.

(Swofford 1993) allowing extant sequences to be placed at nodes (Fig. 3), and then rooting with the outgroup sequence. In cases where the parsimony model is rejected all networks are built following Templeton's linkage algorithm and the unrooted network is then rooted with the outgroup sequence. Vitellogenin and actin require multiple mutations or recombination to explain observed polymorphisms (Table 3, Hudson's H with $P > 0.05$). In both of these cases the parsimony model was rejected only when analysing intron data and not when we considered the entire gene region. Most of the segregating sites and substitutions occur within introns and thus the test pertains mostly to these positions. Including exon sequence simply decreases the average substitution rate: it does not make intron positions less subject to multiple substitutions, it just obscures their higher substitution rate. The network for actin alleles (Fig. 3) required one parallelism and one reversal. Similarly, the vitellogenin network required one recombination event (Fig. 3). The alleles that are produced by inferred recombination are not connected to the remainder of the network because such connections would imply singular histories when these alleles are actually the product of multiple combined histories.

The phylogeny of alleles shows that there is no phylogeographic structuring at the population level. Few alleles are shared between African and invading populations. Yet all of the networks contain African alleles that are phylogenetically derived from alleles in invading populations (Table 1, Fig. 3). This result implies that invading alleles probably arose in Africa, as alleles derived from them can be found there. The phylogenies of alleles, similarly, do not provide evidence that any invading population is monophyletic. Indeed, some of the alleles (e.g. vitellogenin allele E and Actin allele F) are shared among invading populations and Africa. None of the loci follow concordant phylogeographic patterns. Given these patterns, and the recency of Medfly invasions, our result is clearly that of a phylogeny of alleles that do not correspond with the historical relationships among invading populations: invading Medfly populations are still in the process of sorting ancestral polymorphism. Although the phylogeny of alleles is not useful for phylogeographic analysis in this case, it is still essential in that it demonstrates that alleles might be shared among populations simply because all populations are ultimately derived from Africa and not because they share a common invasion history.

Discussion

Intron sequences show a exceptionally high level of allelic diversity. Despite the small numbers of individuals examined here, invading populations of Medflies demonstrate a retention of substantial allelic variation (Table 1). EPIC sequences from Medflies allow us to evaluate the levels of genetic diversity and population level heterozygosity in introns relative to allozymes and mitochondrial DNA. This comparison is particularly informative if made across invading and noninvading populations (Fig. 4). Noninvading (i.e. African) populations contain more variation in intron sequences than in mtDNA, while mtDNA shows more diversity than allozymes. In invading Medfly populations the high intron diversity is in contrast to the low diversity at allozyme loci and mtDNA (Fig. 4). Invading populations show an enormous loss of their original mtDNA diversity (up to 75%), while the proportional loss of variation in nuclear loci (both introns and allozymes) is less severe. Although the two types of nuclear loci (introns and allozymes) also show a loss (35% and 60%, respectively), for introns it is not as great as for other DNA data (mtDNA). It is probable that more of the nuclear diversity is retained in invading populations due to the larger effective population size and inheritance mode of nuclear genes relative to mtDNA (diploid vs. maternal and haploid, Tajima 1983; Birky *et al.* 1989). Introns are more variable than allozymes in source populations, and the absolute amount of variability remaining in introns following invasions is correspondingly greater (Fig. 4).
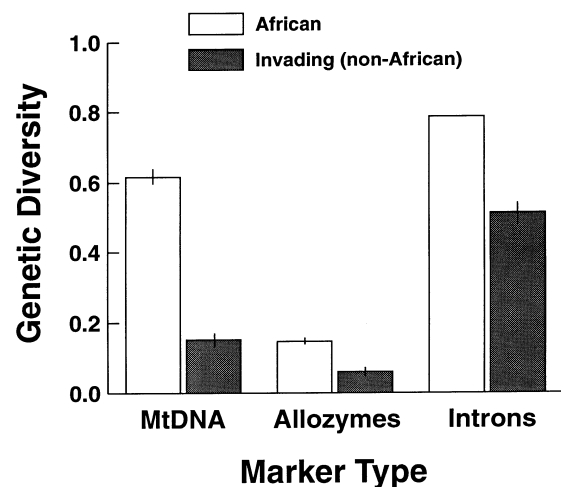

Fig. 4 Genetic variation within African and non-African (invading) populations of Medfly for mtDNA, allozymes and nuclear introns. For all types of data, mean (± SE) genetic variability within populations is calculated as 1 – (probability that two alleles are identical by descent), or $1 - \sum p_i^2$ which also estimates expected heterozygosity for nuclear loci. MtDNA data were obtained by RFLP's; allozyme data represent a minimum of 22 polymorphic loci (see text). For nuclear loci, expected heterozygosity was calculated across loci within populations first; therefore the error bars represent variation between populations.

In Medflies, introns exhibit more genetic variability throughout the invasion process compared to the other markers. This retained intron diversity in invading populations may allow for the reconstruction of invasion histories and is especially advantageous for studies requiring multiple loci to independently test demographic and phylogeographic models. The amount of genetic variation found within introns makes possible the future study of both origins and population structure. Clearly, the next step will be to sample populations more thoroughly and test for population subdivision.

These results have three general implications for the use of intron sequences in invasion biology. First, a large amount of nuclear genetic variability is retained in the Medfly. The genetic variability found at each of several loci implies that the analysis of DNA sequence variation within multiple nuclear introns will be a powerful tool for the study of invasion genetics, even if standard phylogeographic analyses may be inappropriate. This tool will be especially useful for tracing and understanding invasions where little genetic variation is otherwise expected. Importantly, this invading species was previously found to be genetically homogeneous and scientifically difficult to study if not intractable (Liedo & Carey 1996; Roderick 1996b; Roderick & Villablanca 1996). Although this is only one test, the previous intractability of Medfly invasion genetics suggests a useful new tool.

Second, on average, variation is higher in introns than in protein-coding regions, both in percentage variable sites and in nucleotide diversity (Table 2). Population surveys of *Drosophila* confirm this generalization. One example is the *adh* locus in *D. melanogaster*, which contains three introns. The two shorter introns (nos 2 and 3 combined to 135 nucleotides) have a nucleotide polymorphism frequency of 5.2%. Even the most slowly evolving intron (no. 1) has a polymorphism frequency (3.13%) greater than that across *adh* exons (1.8%). This is in spite of the fact that intron 1 has recognized secondary structure, regulatory elements expressed in the larval form, and is within a gene subject to strong purifying selection (Begun & Aquadro 1993; Kreitman 1983). This argues in favour of the general utility of introns for the development of multigenic population genetics and phylogeography. Introns are known to vary in their substitution rates (e.g. Hey & Kliman 1993; Begun & Aquadro 1994) and clearly further empirical work in this area would be helpful.

A third implication is that the history and genealogy of each nuclear locus provides unique information. Each locus tracks the sorting of alleles within and among invading populations. Unlinked loci thus represent independent histories (Pamilo & Nei 1988) in which each gene tree is a potential test of the hypothetical invasion history. Thus, phylogenetic concordance among gene trees from independent loci would provide clear support that gene trees correspond with the population history (Avise & Ball 1993). For example, each one of the Medfly loci could suggest a hypothesis of relationships among populations (Fig. 3). Yet, retained ancestral polymorphism and a lack of monophyly among populations suggests that these loci have not resolved the population (invasion) history. Similarly, shared alleles at a single locus might suggest particular relationships among populations (Table 1), with concordant patterns across multiple loci providing support. However, again, a lack of monophyly plus retained ancestral polymorphism among populations suggest that allele sharing in this case could be due to one common ultimate source for all populations. The disparate genealogies demonstrate that multiple loci will be required to obtain a complete picture of invasion histories. Indeed, the validity of a phylogeographical analysis itself can be tested with multiple phylogenies of alleles.

Abundant sequence variability in Medfly introns suggests that EPIC loci will be useful for other problems in population genetics. Allozymic and mtDNA variability in Medflies was found to be minimal and insufficient for population genetic study of invading populations (Roderick 1996b; Roderick & Villablanca 1996). However, these same populations demonstrated an enormous amount of genetic variability when evaluated for sequence diversity in introns. There has been a recent surge in research into nuclear genes (Palumbi 1996b; Roderick 1996a). The conceptual part of this surge is driven by tests for concordance among independently evolving genetic markers in order to support systematic and population genetic inferences (Avise & Ball 1993; Palumbi & Baker 1994). The empirical part of this surge has been driven by the need to identify and test potentially useful nuclear genes (Friedlander *et al.* 1992, 1994; Palumbi 1996b; Villablanca 1994). The next empirical step will be to determine if most plant and animal populations show the levels of intron sequence diversity identified here. There is already some evidence that this will be the case (Hey & Kliman 1993; Slade *et al.* 1994; Palumbi & Baker 1996). If so, EPIC loci, with more diversity and less propensity for random genetic drift, will be especially useful for studies of bottlenecks, founder events, multiple founders and hybridization.

Interpreting data from variable nuclear sequences in population genetics requires a consideration of issues beyond those normally considered with mtDNA. Chief among these issues is that the time required for the establishment of nuclear monophyletic lineages within populations is far greater than for mtDNA (Tajima 1983 Pamilo & Nei 1988; Moore 1995;). The average time to monophyly (or mean coalescence time) of allele lineages within

two populations is N_f generations for mtDNA and $4N_e$ generations for nuclear DNA (Tajima 1983; where N_f and N_e are the female and total effective population sizes, respectively). Thus, mitochondrial DNA is more likely than any single nuclear locus to track more closely the phylogeographic history of populations because nuclear alleles are retained for longer (Moore 1995). Birky *et al.* (1989) have shown that the approach to equilibrium of allele frequencies is much faster than that of nucleotide diversity or monophyly. Significant frequency differences in nuclear genes can be detected between populations when $N_e < t < 4N_e$, where t is the number of generations since population divergence. Thus, for nuclear sequences, population subdivision can be detected sooner using allele frequencies than it can be with phylogenetic or coalescence methods. Taking an allele-frequency approach allows one to recognize population subdivision, but may obfuscate the recovery of population histories. The inherent advantage to a phylogenetic approach is that individuals are analysed and they are not pooled into species or populations. The analysis of individuals allows the recognition of para- and polyphyletic populations and species if some exist (see Roderick & Villablanca 1996).

A second point to consider in interpreting data from variable nuclear sequences in population genetics is that much of the nucleotide diversity is being produced and maintained by recombination (Templeton *et al.* 1992; Hudson 1994). The analysis of phylogeographic patterns has proven invaluable to population biology (Avise & Hamrick 1996). Yet there are potential complications when using intron sequences in phylogeographic analyses. One of the most serious complications is the construction of phylogenies from DNA fragments that could be related by recombination. An additional problem is that if the resultant gene trees are not concordant among loci then one does not gain a phylogeographic perspective. But the lack of monophyly and concordance among nuclear gene trees within invading populations may itself provide further information about the invasion process by providing evidence that the genealogical pattern is one of segregating ancestral polymorphism. The continued segregation of ancestral polymorphism in the Medfly might have been predicted (Li 1997) given that most invading populations are less than 100 years old (Metcalf 1995a). One might also have predicted that some invading populations would be fixed for alternate ancestral alleles given the Medley's short generation time (as short as 29 days, Vargas *et al.* 1997).

Ancestral allelic diversity in invading populations suggests the hypothesis that effective population size in invading populations is large. This hypothesis is based on the assumption of neutrality, which seems valid given that all four loci considered show similar patterns of variation. It is improbable that selection has acted on all loci in a similar manner (relative to geography). Thus, patterns of diversity at these EPIC loci probably reflect the demographic history of Medflies. The large effective population size suggested here may result from two processes: (i) continued gene flow and a lack of population subdivision; or (ii) a large initial population size. These alternative hypotheses could be tested using EPIC sequences by looking for population subdivision given that genetic diversity has now been uncovered. Some localities have diagnostic mtDNA haplotypes (Gasparich *et al.* 1997) consistent with population subdivision (or a lack of neutrality). Intron sequence data could be used to test independently for population subdivision. Similarly, intron sequence data may ultimately allow one to disentangle whether abundant genetic diversity in invading populations is due to a large founder population or simply a founder population derived from an extremely genetically diverse source.

References

Avise JC, Ball RM Jr (1993) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology*, **10**, 45–67.

Avise JC, Hamrick JL (1996) *Conservation Genetics*. Chapman & Hall, New York.

Banks GK, Robinson AS, Kwiatowski J, Ayala FJ, Scott MJ, Kriticou D (1995) A second super oxide dismutase gene in the medfly, *Ceratitic capitata*. *Genetics*, **140**, 967–702.

Baruffi L, Damiani G, Guglielmino CR, Mandi C, Malacrida AR, Gasperi G (1995) Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multi-locus enzyme electrophoresis data. *Heredity*, **74**, 425–437.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, **356**, 519–520.

Begun DJ, Aquadro CF (1993) African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature*, **365**, 548–550.

Begun DJ, Aquadro CF (1994) Regional variation in fruitflies. *Nature*, **369**, 450.

Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of hetroplasmic cells, and comparison of nuclear genes. *Genetics*, **121**, 613–627.

Carey JR (1991) Establishment of the Mediterranean fruit fly in California. *Science*, **253**, 1369–1373.

Carey JR (1996) The incipient Mediterranean fruit fly population in California: Implications for invasion biology. *Ecology*, **77**, 1690–1697.

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

Friedlander TP, Regier JC, Mitter C (1992) Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. *Systematic Biology*, **41**, 483–490.

Friedlander TP, Regier JC, Mitter C (1994) Phylogenetic information content of five nuclear gene sequences in animals: initial assessment of character sets from concordance and divergence studies. *Systematic Biology*, **43**, 511–525.

Fyrberg EA, Bond BJ, Hershey ND, Mixter KS, Davidson N (1981) The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. *Cell*, **24**, 107–116.

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in human and chimpanzees, with implications for constraint on allele size. *Molecular Biology and Evolution*, **12**, 594–603.

Gasparich GE, Silva JG, Han H-Y, McPheron BA, Steck GJ, Sheppard WS (1997) Population genetic structure of Mediterranean fruit fly (Diptera: Tephritidae) and implications for world wide colonization patterns. *Annals of the Entomological Society of America*, **90**, 790–797.

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

Glenn TC, Stephan W, Dessauer HC, Braun MJ (1996) Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Molecular Biology and Evolution*, **13**, 1151–1154.

Grosberg RK, Levitan DR, Cameron BB (1996) Characterization of genetic structure and geneologies using RAPD–PCR markers: a random primer for the novice and nervous. In: *Molecular Zoology: Advances, Strategies, and Protocols* (eds Ferraris J, Palumbi SR), pp. 67–100. Wiley, New York.

Hare MP, Karl SA, Avise JC (1996) Anonymous nuclear DNA markers in the American Oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Molecular Biology and Evolution*, **13**, 334–345.

Haymer DS, Anleitner JE, He M, Thanaphum S, Saul S, Ivy J, Houtchens K, Arcangeli L (1990) Actin genes in the Mediterranean fruit fly, *Ceratitis capitata. Genetics*, **125**, 155–160.

He M, Haymer DS (1992) A muscle-specific actin gene from the Mediterranean fruit fly, *Ceratitis capitata. Insect Molecular Biology*, **1**, 15–24.

He M, Haymer DS (1994) The actin gene family in the oriental fruit fly *Bactrocera dorsalis*. Muscle specific actins. *Insect Biochemistry and Molecular Biology*, **24**, 891.

Hey J, Kliman M (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Molecular Biology and Evolution*, **10**, 804–822.

Hudson RR (1989) How often are polymorphic restriction sites due to a single mutation? *Theoretical Population Biology*, **36**, 23–33.

Hudson RR (1994) How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? *Proceedings of the National Academy of Sciences, USA*, **91**, 6815–6818.

Kliman RM, Hey J (1993) DNA sequence variation at the period locus within and among species of the *Drosophila melanogaster* complex. *Genetics*, **133**, 375–387.

Konsolaki M, Komitopoulou K, Tolias PP, King DL, Swimmer C, Kafatos FC (1990) The chorion genes of the medfly, *Ceratitis capitata*, I: structural and regulatory conservation of the s36 gene relative to two *Drosophila* sequences. *Nucleic Acids Research*, **18**, 1731–1737.

Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster. Nature*, **304**, 412–417.

Kwiatowski J, Skarecky D, Ayala FJ (1992) Structure and sequence of the *Cu,Zn Sod* gene in the Mediterranean fruit fly, *Ceratitis capitata*: intron insertion/deletion and evolution of the gene. *Molecular Phylogenetics and Evolution*, **1**, 72–82.

Lande R, Barrowclough GF (1987) Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (ed. Soule ME), pp. 87–123. Cambridge University Press.

Lessa EP (1992) Rapid surveying of DNA sequence variation in natural populations. *Molecular Biology and Evolution*, **9**, 323–330.

Lessa EP, Appelbaum G (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**, 119–129.

Li W-H (1997) *Molecular Evolution*. Sinauer Assoc, Massachusetts.

Liedo P, Carey JR (1996) Demography of fruit flies and implications to action programs. In: *Economic Fruit Flies: A World Assessment of Their Biology and Management* (eds McPheron BA, Steck GJ), pp. 299–308. St. Lucie Press, Delray Beach, Florida.

Malacrida AR, Guglielmino CR, Gasperi G, Baruffi L, Milani R (1992) Spatial and temporal differentiation in colonizing populations of *Ceratitis capitata. Heredity*, **69**, 101–111.

Martinez A, Bownes M (1994) The sequence and expression pattern of the *Calliphora erthrocephala* yolk protein A and B genes. *Journal of Molecular Evolution*, **38**, 336–351.

McPheron BA, Sheppard WS, Steck GJ (1995) Genetic research and the origin, establishment, and spread of the Mediterranean fruit fly. In: *The Medfly in California: Defining Critical Research* (eds Morse JG, Metcalf RL, Carey JR, Dowell RV), pp. 93–107. University of California, Center for Exotic Pest Research, Riverside, Riverside, California.

Metcalf RL (1995a) Biography of the Medfly. In: *The Medfly in California: Defining Critical Research* (eds Morse JG, Metcalf RL, Carey JR, Dowell RV), pp. 43–48. University of California, Center for Exotic Pest Research, Riverside, Riverside, California.

Metcalf RL (1995b) The need for research on exotic pests in California. In: *The Medfly in California: Defining Critical Research* (eds Morse JG, Metcalf RL, Carey JR, Dowell RV), pp. 5–39. University of California, Center for Exotic Pest Research, Riverside, Riverside, California.

Miller S, Holt A, Bailey J, Machida S (1992) *The Alien Pest Species Invasion in Hawaii: Background Study and Recommendations for Interagency Planning.* The Nature Conservancy of Hawaii & Natural Resources Defense Council, Honolulu, Hawaii.

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

Moore SS, Sargent LL, King TJ, Mattick JS, Georges M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primers in closely related species. *Genomics*, **10**, 654–660.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: *Evolutionary Process and Theory* (eds Nevo E, Karlin S), pp. 515–534. Academic, New York.

Pääbo S, Wilson AC (1988) Polymerase chain reaction reveals cloning artefacts. *Nature*, **334**, 387–388.

Palumbi SR (1996a) Macrospatial genetic structure and speciation in taxa with high dispersal abilities. In: *Molecular Zoology: Advances, Strategies, and Protocols* (eds Ferraris J, Palumbi SR), pp. 101–108. Wiley, New York.

Palumbi SR (1996b) Nucleic acids III: the polymerase chain reaction. In: *Molecular Systematics* (eds Hillis D, Moritz C). Sinauer, Sunderland, Mass.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

Palumbi SR, Baker CS (1996) Nuclear genetic analysis of population structure and genetic variation using intron primers. In: *Molecular Genetic Approaches in Conservation* (eds Smith TB, Wayne RK), pp. 25–37. Oxford University Press, Oxford.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.

Parker MW, Blake CCF (1988) Crystal structure of manganese superoxide dismutase from *Bacillus stearothermophilus* at 2.4 Å resolution. *Journal of Molecular Biology*, **199**, 649–661.

Perring TM, Cooper AD, Rodriquez RJ, Farrar CA, Bellows TS Jr (1993) Identification of a whitefly species by genomic and behavioral studies. *Science*, **259**, 74–77.

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. *Trends in Ecology and Evolution*, **8**, 285–288.

Regier JC, Wiegmann BM, Leclerc RF, Friedlander TP (1994) Loss of phylogenetic information in chorion gene families of *Bombyx mori* by gene conversion. *Molecular Biology and Evolution*, **11**, 72–87.

Rina M, Savakis C (1991) A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata*: sequence and structural conservation in dipteran yolk proteins and their genes. *Genetics*, **127**, 769–780.

Roderick GK (1996a) Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annual Review of Entomology*, **41**, 263–290.

Roderick GK (1996b) Population genetic studies of tephrited flies of economic importance. In: *Economic Fruit Flies: A World Assessment of Their Biology and Management* (eds McPheron BA, Steck GJ), pp. 267–271. St. Lucie Press, Delray Beach, Florida.

Roderick GK, Howarth FG (1997) Invasion genetics: natural colonizations, exotic pests, and classical biological control. In: *Biological Invasions of Ecosystems by Pests and Beneficial Organisms* (eds Miyazaki M, Andow D), pp. 349–357. National Institute of Agro-Environmental Sciences, Tsukuba, Japan.

Roderick GK, Villablanca FX (1996) Genetic and statistical analysis of colonization. In: *Economic Fruit Flies: A World Assessment of Their Biology and Management* (eds McPheron BA, Steck GJ), pp. 281–290. St. Lucie Press, Delray Beach, Florida.

Sanchez F, Tobin SL, Rdest U, Zulauf E, McCarthy BJ (1983) Two *Drosophila* actin genes in detail: gene structure, protein structure, and transcription during development. *Journal of Molecular Biology*, **163**, 533–551.

Schiff NM, Sheppard WS (1993) Mitochondrial DNA evidence for the 19th century introduction of African honey bees into the United States. *Experientia*, **49**, 530.

Schrag S, Wiener P (1995) Emerging infectious disease: what are the relative roles of ecology and evolution? *Trends in Ecology and Evolution*, **10**, 319–324.

Slade RW, Moritz C, Heideman A (1994) Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution*, **11**, 341–356.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Smith MW, Doolittle RF (1992) A comparison of evolutionary rates of the two major kinds of superoxide dismutase. *Journal of Molecular Evolution*, **34**, 175–184.

Steck G, Gasparich GE, Han HY, McPheron BA, Sheppard WS (1995) Distribution of mitochondrial DNA haplotypes among *Ceratitis capitata* populations worldwide. In: *4th International Congress of Fruit Flies of Economic Importance* (eds McPheron BA, Steck G), Tampa, Florida.

Strassman JE, Solis CR, Peters JM, Queller DC (1996)Strategies for finding and using highly polymorphic DNA microsatellite loci for studies of genetic relatedness and pedigrees. In: *Molecular Zoology: Advances, Strategies, and Protocols* (eds Ferraris J, Palumbi SR). Wiley, New York.

Swofford DL (1993) PAUP. Phylogenetic analysis using parsimony. Sinauer, Sunderland, MA.

Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC (1982) Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *Journal of Molecular Biology*, **160**, 181–217.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Templeton AR, Crandall KA, Sing CF (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.

Templeton AR, Sing CF (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.

Terpstra PABG (1988) Homology of *Drosophila* yolk proteins and the triacylglycerol lipase family. *Journal of Molecular Biology*, **202**, 663–665.

Tolias PP, Konsolaki M, Komitopoulou K, Kafatos FC (1990) The chorion genes of the medfly, *Ceratitis capitata*. II. Characterization of three novel cDNA clones obtained by dif-

ferential screening of an ovarian library. *Developmental Biology*, **140**, 105–112.

US Congress Office of Technology Assessment (1993) *Harmful Non-Indigenous Species in the United States*. US Government Printing Office, Washington, D.C.

Vanderkerckhove J, Weber K (1978) At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. *Journal of Molecular Biology*, **126**, 783–802.

Vargas RI, Walsh WA, Kanehisa D, Jang EB, Armstrong JW (1997) Demography of four Hawaiian fruit flies (Diptera: Tephritidae) reared at five constant temperatures. *Annals of the Entomological Society of America*, **90**, 162–168.

Villablanca FX (1994) Spatial and temporal aspects of populations revealed by mitochondrial DNA. In: *Ancient DNA* (eds Herrmann B, Hummel S), pp. 31–58. Springer-Verlag, New York.

Zacharopoulou A, Frisardi M, Savakis C, Robinson AS, Tolias P, Konsolaki M, Komitopoulou K, Kafatos FC (1992) The genome of the Mediterranean fruitfly *Ceratitis capitata*: localization of molecular markers by in situ hybridization to salivary gland polytene chromosomes. *Chromosoma*, **101**, 488–455.