Molecular Ecology (2003) **12**, 2941–2951 doi: 10.1046/j.1365-294X.2003.01978.x

Population structure of the pumpkin fruit fly *Bactrocera depressa* (Tephritidae) in Korea and Japan: Pliocene allopatry or recent invasion?

JEOMHEE MUN,* ANDREW J. BOHONAK† and GEORGE K. RODERICK*
University of California, Division of Insect Biology, Department of Environmental Science, Policy and Management, 201 Wellman Hall #3112, Berkeley, CA 94720–3112, USA, †San Diego State University, Department of Biology, 5500 Campanile Drive, San Diego, CA 92182–4614

Abstract

Because of their widespread agricultural impact and rapid range expansions, true fruit flies (Tephritidae) are the subject of quarantine and control efforts worldwide. Among these flies, the pumpkin fruit fly *Bactrocera depressa*, which infests squash and other cucurbitaceous plants in Korea, Japan and Taiwan, was recently isolated from produce shipments entering Japan and identified as a regulatory target. This species was described in 1933 from collections in Japan and discovered in 1974 in Korea, suggesting that it may have recently invaded mainland Asia. We analysed the genetic structure of Asian populations of *B. depressa* using sequence variation for mitochondrial gene *cytochrome-oxidase* I and three nuclear loci: *elongation factor* 1α, *tubulin* β1 and *tubulin* β3, using frequency-based approaches, nested clade analysis and assignment tests. Contrary to the hypothesis of recent invasion, high levels of genetic subdivision were found among five Korean and three Japanese populations. Nested clade analysis suggested a variety of processes operating over different time scales, including ancient isolation between Korea and Japan and more recent range expansions within each country. Contrary to a priori expectations, the results also suggested the recent introduction of a mitochondrial haplotype into Yokohama, Japan that is related closely to a widespread haplotype found throughout Korea. Assignment tests also supported these conclusions. The combination of a genealogical approach and probabilistic assignments of individuals to populations of origin was able to provide statistical support for the identification of cryptic introductions within an otherwise widespread indigenous species.

Keywords: assignment test, bioinvasion, cryptic invasion, EPIC loci, intron, mtDNA sequence, nDNA sequence, nested clade analysis, population genetic structure, statistical phylogeography

Received 14 May 2003; revision received 13 August 2003; accepted 13 August 2003

Introduction

True fruit flies (Diptera: Tephritidae) are found in nearly all habitats with suitable plant life. Their cosmopolitan distributions, broad larval host range and substantial economic impacts have placed tephritids among the world's most notorious agricultural pests (McPheron & Steck 1996). Within the United States, for example, six genera (*Anastrepha*, *Bactrocera*, *Ceratitis*, *Dacus*, *Rhagoletis* and *Toxotrypana*) have been identified as major threats to agricultural resources, and are consequently the subject of large-scale interdiction, quarantine and control efforts (USDA 2001). Elsewhere in the world, many of these species are indigenous or already established as exotic pests. As with other taxa, genetic studies of dipteran pests facilitate inferences regarding aspects of ecology and evolution not studied easily by other means. However, the biology of tephritids presents unique challenges for genetic data interpretation. First, tephritid populations rarely match the ideal *n*-population model upon which traditional analyses are based (e.g. Wright 1931). Local populations of tephritids as well as many other insect pests

Correspondence: Jeomhee Mun. E-mail: jhmun@nature.berkeley.edu

can be founded easily by a small number of colonists, go extinct due to environmental or anthropogenic causes, undergo extreme fluctuations in size and display source-sink dynamics (Eber & Brandl 1994; Stark *et al*. 1994; Roderick 1996a; Roderick *et al*. 1998; Sved *et al*. 2003). The resulting metapopulation structure can have greatly reduced levels of variation and depart significantly from drift–gene flow equilibrium (Davies *et al*. 1999b; Whitlock & McCauley 1990; Fonseca *et al*. 2000; Mun *et al*. 2000; Goodisman *et al*. 2001).

Despite these constraints, variable microsatellites, mtDNA restriction fragment length polymorphisms (RFLPs) and sequences and nDNA sequences have been identified for intraspecific studies of several tephritid pest species, including the medfly *Ceratitis capitata* (He & Haymer, 1997; Steck *et al*. 1996; Villablanca *et al*. 1998) and other *Bactrocera* species (Roderick *et al*. 1996; Kinnear *et al*. 1998; Mun *et al*. 1999; Smith *et al*. 2003). These markers have proved useful for identifying the source of individual invaders in areas where the species are not yet resident, and in characterizing the demography of invasions in recently established populations (Bohonak *et al*. 2001; Davies *et al*. 1999a; Bonizzoni *et al*. 2001; Meixner *et al*. 2002). However, laboratory and analytical techniques for analysing population structure in exotic pest insects are often species- and even region-specific. This is because levels of genetic variation tend to vary widely among taxa, even for the same gene. Further, frequency-based analyses used to study the movement of individuals among established populations are inappropriate for newly established populations (Bohonak & Roderick 2001; Davies *et al*. 1999b). The current paradigm for invasion genetics of tephritid species is based almost entirely on studies of the Mediterranean fruit fly *Ceratitis capitata*, whose cosmopolitan distribution and broad host range are probably exceptions in this family, rather than the norm (White & Elson-Harris 1992; Roderick 1996b).

Tephritids in the genus *Bactrocera* are of particular concern throughout much of Asia and Australia, where they constitute a significant threat to agricultural resources (Han *et al*. 1994; Takamatsu 1952; White 1996; Kinnear 1998; Kim *et al*. 1999). Of these, the pumpkin fruit fly *B. depressa* (Shiraki) is relatively unusual, with a distribution restricted to mountainous regions of Korea, Japan and Taiwan and primary association with the unculivated host *Trichosanthes kirilowii* (Chinese cucumber or 'snakegourd') (Carrol *et al*. 2002 onwards; Kim *et al*. 1999). *B. depressa* was described first from Japan in 1933 (Shiraki 1933) and was discovered in Korea in 1974 (Kim & Kim 1974). In areas of Japan and Korea where it occurs, *B. depressa* is currently a significant crop pest of cucurbitaceous plants, primarily *Cucurbita moschata* (winter squash, or pumpkin) and *C. maxima* (turban squash) (Takamatsu 1952; Shiraki 1968; Han *et al*. 1994; Kim *et al*. 1999). It is unknown whether *B. depressa* impacted personal-use pumpkin crops in Korea prior to 1974. Although pumpkins have been recognized only recently as economically valuable crops in Korea, *C. moschata* has been cultivated on a small scale there for over 1500 years (An 1999; Kim *et al*. 1999). The relatively recent discovery of *B. depressa* in Korea suggests that it may have invaded this region recently, or have only recently become a problem; perhaps as an expansion of range associated with new agricultural practices. Detection of *B. depressa* on crops exported from Korea to Japan supports a hypothesis that the species is now being transported regionally in shipments of infested produce. This led to its listing as a quarantine subject in 2002 by the United States (Kim 2002). However, if the pumpkin fruit fly were indigenous to both countries, its current distribution could have originated as early as the splitting of Japan from mainland Asia in the late Miocene or early Pliocene 5–10 mya (Smith *et al*. 1994).

Here, we used sequence-level variation from one mitochondrial gene and three nuclear loci to study the genetic structure of *B. depressa* in Korea and Japan and infer the recent history of these populations. We addressed two questions: (1) at what spatial scale, if any, are populations of the pumpkin fruit fly genetically subdivided? (2) What is the invasion history of populations in Japan and Korea? Elucidation of the history of *B. depressa* in these two countries may help to determine its potential for future invasions in Asia and elsewhere and provides a contrast for genetic studies of other invasive tephritid species. The work also demonstrates the combination of a genealogical approach and probabilistic assignments to detect cryptic, recently invading, genotypes within an already established species.

Materials and methods

Individual genotyping

From 1999 to 2000, 63 pumpkin fruit flies were collected as adults, pupae or larvae from eight regions of fruit infestation (up to 250 km apart) across much of South Korea and three populations in Japan (spanning approx. 450 km on the main island of Honshu, Table 1). Individuals were stored in 95% ethanol until DNA was extracted with either standard phenol/chloroform methods (Palumbi 1996) or commercial tissue extraction kits (DNEasy kits, Qiagen, Inc.). Protocols for 35 amplification cycles were adapted from previous studies of tephritid fruit flies (Villablanca *et al*. 1998; Mun *et al*. 2000). An 821-bp portion of the mitochondrial gene *cytochrome-oxidase I* (CO I) was amplified using the primer pair C1-J-2183 and TL2-N-3014 (Table 2). For the nuclear genes, universal exon-primed, intron-crossing (EPIC) primer pairs, EF1 and EF2 for *elongation factor* (EF) 1α, and Tub3 and Tub4 for beta *tubulin* (Tub) were used in initial screens (Palumbi 1996).

Table 1 Sample collection localities for the pumpkin fruit fly, *Bactrocera depressa*

Country	Date of collection	Latitude, longitude	No. of Ind.	Host plant population
Korea				
Chiak (Ch)	2 September 2000	37.4 N, 128.0 E	9	*Cucurbita moschata*
Jeongub (Je)	17 March 2000	35.5 N, 126.6 E	10	*Cucurbita moschata/Trichosanthes kirilowii*
Mooju (Mo)	17 March 2000	36.0 N, 127.6 E	7	*Cucurbita moschata*
Wanju (Wa)	12 March 2000	35.8 N, 127.2 E	7	*Cucurbita moschata*
Yangsan (Ya)	July 1999	35.3 N, 129.1 E	6	*Cucurbita moschata*
Total			39	
Japan				
Ofunato (Of)	October 1999	39.1 N 141.7 E	6	*Cucurbita moschata*
Nagano (Na)	December 1999	36.6 N 138.2 E	10	*Cucurbita moschata*
Yokohama (Yo)	8 September 2000	35.4 N 139.5 E	8	*Trichosanthes kirilowii*
Total			24	

Table 2 Primer pairs used for PCR amplifications (see text for references). Product size refers to the number of base pairs used in the final analysis, after removing primers and ambiguous end positions

Gene	Product size (bp)	Primer pairs (5′–3′)
Cytochrome oxidase I (mtDNA CO I)	821	CAA CAT TTA TTT TGA TTT TTT GG TCC ATT GCA CTA ATC TGC CAT ATT A
Tubulin β1 (Tub β1)	292	ATC AGT GCG CTC TGG CCC CTA T GTA CTC TTC ACG TAT TTT TGA GAT GAG
Tubulin β3 (Tub β3)	272	CTC GGT ACG TTC TGG CGC TTT C GTA CTC TTC TCG AAT TTT CGA TAT CAA
Elongation factor 1α (EF)	188	ACG TGA AGA ACG TTT CCG TTA AGG AG ATG TGA GCA GTG TGC AAT CCA ATA C

However, sequencing of cloned polymerase chain reaction (PCR) products revealed that two or more loci were being amplified in each case. Consequently, we developed novel, locus-specific primers for this study (Table 2).

For each individual fly and each locus, cycle sequencing of PCR products was carried out using the Big Dye Cycle Sequencing Kit (PE Applied Biosystems). In most cases (including all sequences with ambiguities), PCR products were sequenced in both directions. For each of the nuclear genes, either one or two sites were variable, and individuals were often heterozygous as described in earlier work on tephritids (e.g. Davies *et al*. 1999a). Heterozygosity was inferred based on the superimposition of two dye peaks in the corresponding chromatograms. Subsequent cloning and sequencing of approximately 1/4 of the heterozygous individuals for each locus verified the accuracy of this technique. Both alleles were sequenced for individuals with two heterozygous sites, to verify the *cis* or *trans* arrangement.

After excluding beginning or trailing sites that were ambiguous in some individuals, sequence length varied from 188 bp (EF 1α) to 821 bp (CO I). Because there were no insertions or deletions, sequences were inspected visually and aligned automatically using the program SEQUENCHER version 4.0 (Gene Codes Corporation). Genotypes were obtained for all 63 individuals at each locus, except for *tubulin* β3 ($n = 61$). For each locus, sequences for all haplotypes (i.e. novel alleles differing by at least one base pair from all others) were submitted to NCBI GenBank (AF477957–AF477976; AY349475–AY349482).

Analysis of mitochondrial and nuclear gene loci

Variation within populations was summarized using standard statistics: A (number of alleles = unique haplotypes), S (number of segregating sites = variable nucleotide positions) and Nei's (1987) gene diversity (equivalent to expected heterozygosity, with a sample size correction). Nucleotide diversity, π, was calculated as the average pairwise divergence between all gene combinations (Nei 1987). For each population-locus combination, we calculated Wright's inbreeding coefficient F_{IS} and tested for departures from Hardy–Weinberg equilibrium using exact tests. Statistics were calculated using the program ARLEQUIN version 2 (Schneider *et al*. 2001).

Population structure was analysed at different spatial scales using several approaches. Among all eight populations, subdivision was tested for statistical significance using exact tests and quantified using Wright's (1931) F_{ST} and Excoffier *et al*.'s (1992) analogous statistic Φ_{ST} (10 000

Table 3 Nucleotide variation within, and divergence among, all populations

					Differentiation among all populations		
Locus	*n*	*A*	*S*	π	*P* (exact test)	F_{ST}	Φ_{ST}
mtDNA *CO I*	63	20	34	1.33%	< 0.00001	0.316***	0.830***
Tubulin β3	61	3	2	0.31%	< 0.00001	0.517***	0.317***
Tubulin β1	63	2	1	0.23%	< 0.00001	0.318***	0.505***
Elongation factor 1α	63	3	2	0.15%	< 0.00001	0.312***	0.104***

n: sample size (individuals); *A*: number of alleles (unique haplotypes); *S*: number of segregating sites; π nucleotide diversity (see text). Divergence among populations was assessed statistically using exact tests, and calculated as Wright's (1931) F_{ST} and Excoffier *et al.*'s (1992) Φ_{ST}. ***Denotes values that are highly significantly different from 0 ($P < 0.00001$) for F_{ST} and Φ_{ST}.

randomizations or permutations in each case). In contrast to F_{ST}, Φ_{ST} considers the genealogical or evolutionary distance among alleles, as well as their frequencies in each population. Because the distances among alleles were small (Table 3) we used the matrix of observed pairwise distances for these calculations, rather than a corrected distance measure. Population subdivision was also analysed hierarchically at the levels of population, country and total by (1) estimating Φ_{ST} for Korea and Φ_{ST} for Japan with a separate analysis for each country, and (2) calculating Φ_{CT} (divergence between Korea and Japan) using a hierarchical model with populations nested within countries. (We chose to estimate differentiation within each country using separate analyses because the corresponding estimate Φ_{SC} from the hierarchical model assumes that differentiation within Korea is equal to differentiation within Japan. Preliminary analyses showed that this was an invalid assumption.) For the three nuclear genes, we calculated overall estimates of Φ_{ST} and Φ_{CT} by jackknifing over loci (Weir 1996).

Relationships among the alleles at each locus were estimated with the network reconstruction method developed by Templeton and colleagues (e.g. Templeton *et al*. 1992; Templeton 1998), which uses a parsimony criterion. Because this approach explicitly allows for multifurcations and the persistence of ancestral nodes, we judged it to be more appropriate in representing relationships among alleles for this intraspecific study than standard phylogenetic tree-building techniques (see review, Posada & Crandall 2001). Briefly (1) we constructed a network for each locus using the computer program TCS version 1.13 (Clement *et al*. 2000); (2) nested sets of relationships within the network were inferred using the algorithms of Templeton *et al*. (1987, 1992); (3) non-random geographical associations among haplotypes and clades were tested statistically with the computer program GEODIS (Posada *et al*. 2000); and (4) the results were interpreted using the inference key provided with the GEODIS version 2 application (see also Templeton 1998). This dichotomous key examines the geographical distribution of each clade member, assessing whether patterns are consistent with processes such as restricted gene flow (i.e. isolation by distance), allopatric fragmentation or long-distance colonizations. Identification of ancestral (node) and more recent 'tip' haplotypes is based primarily on the number of mutational connections, as well as haplotype frequencies. In accord with coalescent theory, the TCS program favours the most frequent sequences as ancestral (Clement *et al*. 2000). Templeton (1998, 2002) provides an overview of the entire process, known as 'nested clade analysis' (NCA).

The inference key portion of nested clade analysis has been criticized in that it does not assess error appropriately, and thus may not accurately infer, or distinguish among, alternative hypotheses (Knowles & Maddison 2002). To overcome this limitation we tested the hypotheses that particular genotypes had origins in the population in which they were collected with multilocus assignment tests (Davies *et al*. 1999b; Pritchard *et al*. 2000) using all four loci as implemented by the computer program STRUCTURE version 1 (Pritchard 2000). The analysis determines the probability ($Q > 0.95$) that migration can be rejected statistically with $\alpha < 0.05$. For individuals within each population we also tested for statistical outliers in the probability of origin. Statistical calculations for *t*-test were performed in DataDesk (Velleman 1997).

Results

For *tubulin*, universal primers used in the initial PCR produced two different products that were of similar size but divergent at approximately 30% of nucleotide positions. An NCBI BLAST search confirmed that the two different PCR products corresponded to *tubulin* β1 and *tubulin* β3 (characterized elsewhere, e.g. Guenette *et al*. 1991; Trivinos-Lagos *et al*. 1993). Two novel primer sets were subsequently developed to specifically target each locus (Table 2). Similarly, universal primers used in the initial PCR for *elongation factor* 1α produced two products: a 280-base pairs (bp) product, which contained a putative intron, and a 220-bp product, which did not. We designed

Table 4 Gene diversity (Nei 1987) for each gene in each population (equivalent to expected heterozygosity, with a sample size correction)

	mtDNA CO I	Tub β3	Tub β1	EF 1α	Mean
Korea					
Chiak	0.694	0.307	0.111	0.386	0.375
Jeongub	0.867	0.000	0.526	0.505	0.475
Mooju	0.857	0.363	0.495	0.275	0.497
Wanju	0.810	0.264	0.495	0.484	0.513
Yangsan	0.333	0.200	0.530	0.318	0.345
Mean for Korea	0.712	0.227	0.431	0.393	
Japan					
Ofunato	0.333	0.409	0.000	0.621	0.341
Nagano	0.378	0.442	0.000	0.337	0.289
Yokohama	0.714	0.500	0.000	0.525	0.435
Mean for Japan	0.475	0.450	0.000	0.494	
Mean	0.589	0.311	0.270	0.431	

a novel primer set to target the intron-containing product (Table 2).

Variation within populations

Variation within all four genes consisted entirely of single nucleotide polymorphisms; no insertions or deletions were found, even within the introns of *tubulin* β1, *tubulin* β3 and *elongation factor* 1α. Mitochondrial variation was considerably higher than variation for the nuclear genes (Table 3). In 63 individuals, we identified 20 unique haplotypes (four from Chiak, five from Jeongub, five from Mooju, four from Wanju, two from Yangsan, two from Ofunato, three from Nagano, three from Yokohama), and 34 segregating sites for CO I, in contrast with 1–2 segregating sites and 2–3 alleles for each of the nuclear genes. This led to a higher estimate of gene diversity for CO I ($\Pi = 1.33\%$) than for the other loci ($\Pi = 0.15\%–0.31\%$). Because the natural expectation for π (average nucleotide divergence between randomly selected pairs of sequences) is equal to θ ($= 4N_e\mu$), higher gene diversity in CO I may reflect a higher average mutation rate. Alternative interpretations would include a recent selective sweep at or near each nuclear gene, or balancing selection on the mitochondrial genome.

These results extended to the level of the population, where gene diversity was higher for CO I than for the three nuclear genes in six of eight populations. However, few other generalities were apparent when contrasting populations (Table 4). For example, the fixation of Tub β1 in Japan led to a significant decrease in gene diversity when contrasted with Korea ($P < 0.01$ from *t*-test, using each population as an independent sample), even though Japan possessed significantly higher levels of variation in Tub β3 ($P < 0.05$). Differences between the two countries for variation in CO I and EF 1α were not statistically significant.

Divergence among populations

The eight populations showed significant differentiation for all loci (Table 3). Across all populations exact tests for differentiation were highly significant, and estimates of F_{ST} ranged from 0.312 for EF 1α to 0.517 for Tub β3 (all $P < 0.00001$). Φ_{ST} estimates were also high but more variable among loci, ranging from $\Phi_{ST} = 0.104$ for EF 1α to $\Phi_{ST} = 0.830$ for CO I. Differentiation among Korean populations was less than differentiation among Japanese populations for three of the four loci (Table 5; contrasts were not possible for Tub β1 due to an absence of variation within Japan). In a hierarchical analysis focused specifically on divergence between Korea and Japan, differentiation between the countries was also quite high, with $\Phi_{CT} = 0.829$ for the mitochondrial gene CO I and $\Phi_{CT} = 0.551$ for a combined estimate across the three nuclear loci. With one exception, mitochondrial estimates of Φ_{ST} within each country and Φ_{CT} were always greater than the corresponding estimates for nuclear genes (Table 5).

Nested clade analysis (NCA)

We reconstructed intraspecific haplotype networks for all four loci. In each case, divergence among populations was statistically significant (as estimated by exact contingency tests), although a more detailed analysis of Tub β1 was not possible because only two alleles were found. (With only two alleles, the identity of the ancestor cannot be readily determined using NCA without an outgroup). The presence of three alleles for EF 1α and Tub β3 allowed us to analyse geographical structure at a single level for each gene, and the presence of 20 alleles for CO I was high enough to warrant reconstruction of a more complete set of higher-level clades (Fig. 2).

For EF 1α, Tub β3 and CO I, population subdivision was consistent with range expansions and/or allopatric

Gene	Korea: Φ_{ST}	Japan: Φ_{ST}	Korea vs. Japan: Φ_{CT}
MtDNA			
Cytochrome-oxidase I	0.130**	0.464***	0.829*
nDNA			
Overall (three nuclear loci) Jackknife estimate ± 1 SE	0.090 ± 0.046	0.407 ± 0.169	0.551 ± 0.282
Tubulin β3	−0.011	0.147	0.741*
Tubulin β1	0.151*	No variation	0.294
Elongation factor 1α	0.059	0.435**	0.084

For each single-locus estimate, significance was assessed by randomization: *$P < 0.05$, **$P < 0.005$, ***$P < 0.00001$. Statistical significance was not estimated for the combined estimate of nuclear divergence from jackknifing.

Table 5 Divergence among Korean and Japanese populations, and between Korea and Japan as estimated by AMOVA (Excoffier *et al*. 1992) and exact tests. Φ_{ST} estimates are derived from separate analyses for Korean and Japanese populations, while Φ_{CT} is estimated from a hierarchical island model (see text for details). Combined jackknife estimates for the three nuclear genes were calculated following Weir (1996)

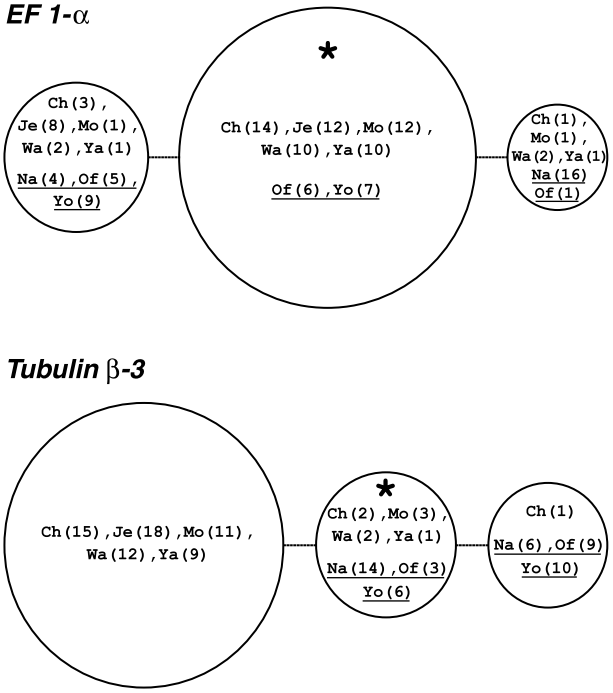

Fig. 1 Haplotype networks for Tub β3 and EF 1α. Localities are indicated for each unique sequence (Japanese localities are underlined), and each haplotype is connected by a single mutational step. The size of each circle is proportional to the sample size. Haplotypes marked with an asterisk are hypothesized to be ancestral based on the criteria of Castelloe & Templeton (1994). Numbers in brackets denote number of individuals for each haplotype from each population.

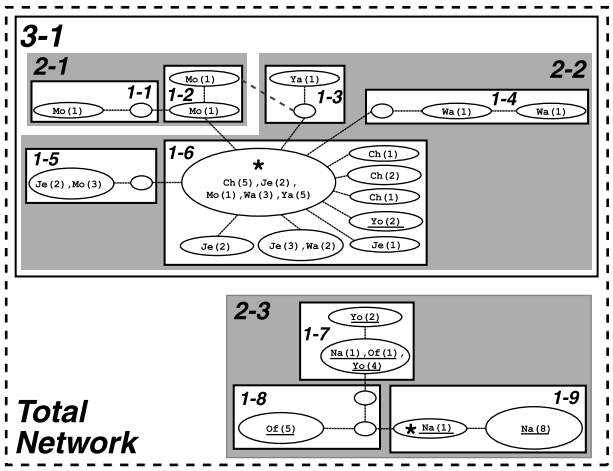

Fig. 2 Haplotype network for CO I, formatted as in Fig. 1. Empty ovals indicate haplotypes that were not sampled, and the dashed line indicates an alternative connection for one ambiguous 'loop' (see text). Boxes enclose nested clades at successively higher levels. Japanese localities are underlined. Numbers in brackets denote number of individuals for each haplotype from each population.

fragmentation events, with NCA interpretations differing somewhat due to varying levels of resolution for each gene. For EF 1α, all three alleles were found in Korea and Japan, with relatively low divergence between the two countries. However, one haplotype (on the left in Fig. 1) had a distribution that was significantly widespread, while the presumed ancestral haplotype was significantly restricted ($P < 0.05$) from randomization tests; (Posada *et al*. 2000). The average geographical displacement between the interior and 'tip' haplotypes was also statistically significant. The NCA inference key interpreted these patterns in EF 1α as a contiguous range expansion (see Templeton 1998 for justification; Templeton *et al*. 1995). For Tub β3, much more differentiation between Korea and Japan was apparent, and a somewhat contradictory pattern emerged (Fig. 1). In this case, the two 'tip' haplotypes distributions that were significantly restricted, as determined by analyses of each tip alone and a joint contrast with the ancestral haplotype ($P < 0.001$ for all). These patterns are consistent with allopatric fragmentation between Korea and Japan.

Extensive variation in CO I facilitated a more detailed geographical analysis of clade structure (Fig. 2). All evolutionary relationships among these haplotypes (i.e. mutational steps) were reconstructed using NCA, with two

exceptions. First, it was necessary to break a connection within an ambiguous 'loop' (i.e. an incompletely resolved portion of the network) using the criteria of Crandall, Templeton and colleagues (Crandall *et al*. 1994; Templeton *et al*. 1995). Second, the evolutionary relationship between clades 3–1 and 2–3 could not be inferred. These two groups diverged by 2–3%, in contrast to < 0.5% divergence within each clade. Because clade 3–1 contains Korean haplotypes with one exception and clade 2–3 contains only Japanese haplotypes, we concluded that *B. depressa* has evolved separately in these two regions for a long period of time.

Conclusions regarding population genetic processes were possible for CO I clades 1–6, 2–2 and 2–3. Significant geographical associations were present within two additional clades (3–1 and the total cladogram) although the interior vs. tip status could not be determined in these cases. Clade 1–6 contained a common, widespread haplotype, which was the presumed ancestor of the entire Korean network (Fig. 2). Seven haplotypes were connected to this ancestor by a single mutational step, and each was relatively rare and restricted in distribution. One of these seven was found in Yokohama, Japan (in two individuals), but nowhere in Korea. The NCA inference key interpreted these distributional patterns in terms of allopatric fragmentation, although the analysis does not easily facilitate separate interpretations for the Korean portions of the clade (which may be a simple result of geographically restricted gene flow) and the haplotype in Yokohama (which appears to be a recent, long-distance colonist). We repeated the NCA after removing the Yokohama haplotype, and found that the distribution of clade 1–6 in Korea was consistent with a range expansion or restricted gene flow, although not statistically significant ($P = 0.28$ for contingency test of the clade). Similar interpretations were evident in the more inclusive clade 2–2, which includes the common, widespread interior clade 1–6 and the restricted tip clades 1–4 and 1–5. NCA suggests that these patterns are consistent with restricted gene flow and isolation by distance between clades 1–5 and 1–6. This result was the same and statistically significant whether or not the Yokohama haplotype was included in the network.

Unlike clade 3–1, the Japanese clade 2–3 did not contain a single, common and widespread haplotype. The three most common haplotypes were several mutational steps apart from each other, and two were restricted to a single population. The NCA inference key suggested either a contiguous range expansion or long-distance colonization events, with further sampling in Japan necessary to distinguish between these alternatives.

Assignment tests

Multilocus assignment tests suggested that most individuals originated in the locality in which they were collected (Fig. 3). Over 90% of individuals sampled were assigned with a probability of origin of $Q > 0.95$ to their collection locality, thereby rejecting the hypothesis of recent immigration for most individuals. Two individuals (one from Jeongub, Korea and one from Yokohama, Japan) had a much lower probability of origin in their collection locality ($Q < 0.80$). The individual from Yokohama, Japan was one of those identified by NCA analysis of mitochondrial DNA as a potential recent immigrant from Korea to Japan. Populations varied somewhat in the degree to which individuals were likely to be assigned to their collection locality (Fig. 3). Some populations showed a greater variance in probability of local origins of individuals, with statistical outliers in four populations (Korea: Chiak, Jeongub; Japan: Nagano, Yokohama) suggesting that all individuals within populations may not share the same origin.

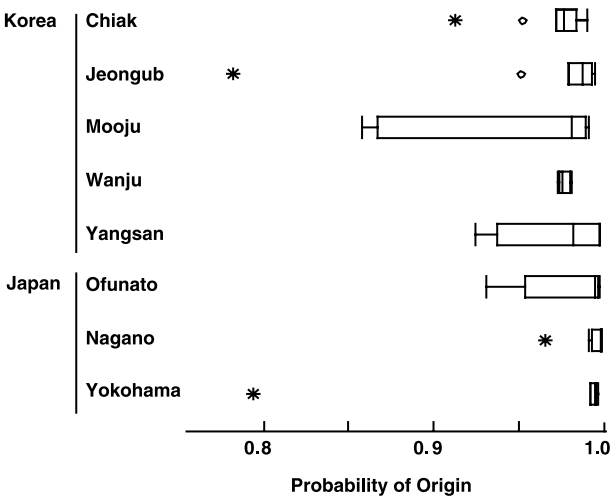

Fig. 3 Probabilities that each individual originated from the same gene pool from which it was sampled, summarized by population. Probabilities correspond to the *Q*-values of Pritchard *et al*. 2000). A probability > 0.95 indicates that migration can be statistically rejected with $\alpha < 0.05$. Boxplots for each population depict the median (centre line) and 25th and 75th percentiles (box hinges). Whiskers extend an additional 1.5 × (high hinge–low hinge), outliers are indicated with circles, and extreme outliers with stars (Velleman 1997).

Discussion

The economic impacts of exotic species have led to an increased interest in studies of their ability to disperse, colonize and become established (e.g. Nalepa & Schloesser 1992; Lodge 1993; Holway & Suarez 1999). One of the primary challenges in many cases is to simply determine which species are indigenous (or native) and which are introduced (Carlton 1996; Gillespie & Roderick 2002). For example, it may be very difficult to distinguish between a recent invasion and historically widespread distribution. In the case of agricultural pests, these determinations have

been the subject of considerable debate because of difficult decisions regarding the costs and benefits of alternative interdiction and eradication efforts (Carey 1992; Saul 1992; Voss 1992; Carey 1996). As a result, a more complete understanding of the historical distribution of the pumpkin fruit fly and contemporary patterns of movement are necessary for resource managers in Asia, and will provide baseline data for potential future invasions of temperate regions.

We were able to identify one mitochondrial and three nuclear markers that were variable in *B. depressa* at the nucleotide level. Levels of variation in the mitochondrial gene CO I exceeded those in EF 1α, Tub β1 and Tub β3, whether estimated as gene diversity, number of segregating sites, number of alleles or average pairwise divergence (Tables 3 and 4). These results contrast sharply with levels of variation in the well-studied Mediterranean fruit fly *Ceratitis capitata*, where nucleotide variation in intron-containing genes is extremely high (S = 12–56 and A = 18–38), and differentiation is low even on a global scale (Davies *et al*. 1999a and unpublished data). Beginning approximately 150 years ago, frequent large-scale invasions of *C. capitata* have homogenized populations worldwide and maintained high levels of genetic variation. In contrast, *B. depressa* has a relatively restricted distribution within Asia, and the data from these four loci suggest that dispersal or secondary invasions across the Sea of Japan are infrequent on relevant ecological time scales (Table 5). These results illustrate that the choice of markers, sampling strategies and analyses in molecular studies of tephritid pests needs to be chosen with regard to the distribution and ecology of each species.

We were able to separate several processes responsible for genetic structure in *B. depressa* using nested clade analysis. Geographic patterns in two of the nuclear genes were consistent with a range expansion (EF 1α) and allopatric fragmentation (Tub β3); each of these processes was embedded within the more detailed network available for CO I (see Fig. 2). Range expansions were evident in both Korea (clade 1–6) and Japan (clade 2–3), with the shorter branches in 1–6 indicating that the expansion within Korea occurred more recently. In Korea, this process is superimposed onto an older pattern of restricted gene flow and isolation by distance in clade 2–2. The NCA interprets the broad distribution of interior clade 1–6 and the restricted distributions of 1–2, 1–4 and 1–5 as being consistent with isolation by distance. The deep divergence between 3 and 1 and 2–3 supports the general conclusion from the Φ-statistics that Korea and Japan constitute very different gene pools.

Pliocene allopatry or recent invasion?

Ecological data available for *B. depressa*, the historical collections (Shiraki 1933; Kim & Kim 1974) and current taxonomic treatments (e.g. White *et al*. 1992) suggest recent introduction of the pumpkin fruit fly into Korea. However, our genetic data are inconsistent with this recent invasion hypothesis. Our data support the hypothesis that the pumpkin fruit fly is indigenous to both countries and the flies are not panmictic throughout this region. Because of the limitations of gene flow estimates (Bohonak, Roderick 2001), an accurate assessment of current gene flow between Korea and Japan may not be possible without knowing the time since divergence between the countries and the genetic composition of the colonizing populations. However, the conclusion that gene flow is low between the two countries is robust under many different types of analyses. For example, an equilibrium interpretation of Φ_{CT} (Wright 1931) translates to a single migrant across the Sea of Japan every five generations (combined nDNA) to 10 generations (mtDNA). Similarly, application of the cladistic gene flow estimate proposed by Slatkin & Maddison (1989) to CO I would suggest that less than one individual migrates every 10 generations between the two countries.

A qualitative interpretation of the CO I gene genealogy (Fig. 2) would suggest historical isolation across the Sea of Japan and one recent, long-distance migrating allele. However, it is surprising to note that the invading fly (or flies) carrying this allele moved from Korea to Japan, rather than the opposing direction. This result was also supported statistically for one of the Yokohama flies in an assignment test employing all four loci. Because the Yokohama allele in question is a single mutational step from the most common, widespread allele in Korea, it seems likely that the event occurred recently, and that the migrant originated in a Korean population that we did not sample or as a result of mutation since arrival in Japan. The nested clade analysis shows that this migration event took place at approximately the same time as a range expansion throughout Korea (see clade 1–6), and could be associated with changes in agricultural practices or in regional produce shipments.

The high levels of genetic divergence between the two countries indicate that the populations were on largely independent evolutionary trajectories prior to this event. We applied a simple molecular clock to estimate when CO I clades 3–1 and 2–3 may have diverged, acknowledging that there are many uncertainties associated with calculations of this type (Gillespie 1991; Ayala 1997). The maximum pairwise divergence between these clades is 3.05%. Assuming a divergence rate of 2.3% per million years for insect mitochondrial DNA (Brower 1994), this suggests that the Korean and Japanese gene pools have been separated for 1.3 million years. Although that date does not correspond precisely to the Pliocene separation between the two countries 5–10 million years ago (Smith *et al*. 1994), it is clear that the establishment of *B. depressa* in Korea and

Japan long predates the rise of agriculture and regionwide trade.

The recent invasion of undetected genotypes into areas where a species is already been established has been termed a 'cryptic invasion' (Geller 1999). Such invasions have been documented for similar species replacing each other (Geller 1999; Grosholz 2002), genotypes replacing others within species (Saltonstall 2002) and undetected hybridization of individuals within newly colonized areas (Gaskin & Schaal 2002). This study illustrates that probabilistic assignments of individuals to populations of origin can provide statistical support for the identification of cryptic introductions, which can then be used to confirm patterns identified by a genealogical approach such as nested clade analysis.

Conclusions

The global transport of produce and the homogenization of agriculture can lead to genetic panmixia in agricultural pest populations (Davies *et al*. 1999b; Pimentel *et al*. 2000; Bohonak & Roderick 2001). However, *B. depressa* shows strong regional subdivision in Asia that can be attributed to a variety of underlying processes. Genetic variation among pumpkin fruit fly populations indicates isolation between Korea and Japan for approximately 1 million years, as well as more recent range expansions in each country. Recent changes in population structure in Korea are coincident with the dispersal of a single mitochondrial allele into Japan via one or more individuals. The evidence for a cryptic invasion was supported by genealogical nested clade analysis and statistical assignment. More detailed inferences are likely to require further sampling throughout Asia and the isolation of nuclear genes that are more variable.

Acknowledgements

We would like to thank Dr Tomonari Watanabe and Dr Sujein Chang for help with sample collection. Our work on the population structure of insect pests is supported by grants from Korean Science Foundation, National Science Foundation, United States Department of Agriculture, California Department of Food and Agriculture and the Agricultural Experiment Station of the University of California.

References

An WS (1999) *Seed Plants of Korea*. Sakyejul Press, Seoul.

Ayala FJ (1997) Vagaries of the molecular clock. *Proceedings of the National Academy of Sciences of the USA*, **94**, 7776–7783.

Bohonak AJ, Davies N, Villablanca FX, Roderick GK (2001) Invasion genetics of New World medflies: testing alternative colonization scenarios. *Biological Invasions*, **3**, 103–111.

Bohonak AJ, Roderick GK (2001) Dispersal of invertebrates among temporary ponds: are genetic estimates accurate? *Israel Journal of Zoology*, **47**, 367–386.

Bonizzoni M, Zheng L, Guglielmino CR *et al.* (2001) Microsatellite analysis of medfly bioinfestations in California. *Molecular Ecology*, **10**, 2515–2524.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences USA*, **91**, 6491–6495.

Carey JR (1992) The medfly in California [Response]. *Science*, **255**, 515–516.

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

Carlton JT (1996) Biological invasions and cryptogenic species. *Ecology*, **77**, 1653–1655.

Carroll LE, White IM, Friedberg A *et al*. (2002 onwards) Pest *Fruit Flies of the World: Descriptions, Illustrations, Identification, and Information Retrieval*. Version: 8th August 2002. Available at: http://www.sel.barc.usda.gov/Diptera/tephriti/pests/adults/

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent. *Molecular Phylogenetics and Evolution*, **3**, 102–113.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Crandall KA, Templeton AR, Sing CF (1994) Intraspecific phylogenetics: problems and solutions. In: *Models in Phylogeny Reconstruction* (eds Scotland RW, Siebert DJ, Williams DM), pp. 273–297.Clarendon Press, Oxford.

Davies N, Villablanca FX, Roderick GK (1999a) Bioinvasions of the medfly, *Ceratitis capitata:* source estimation using DNA sequences at multiple intron loci. *Genetics*, **153**, 351–360.

Davies N, Villablanca FX, Roderick GK (1999b) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution*, **14**, 17–21.

Eber S, Brandl R (1994) Ecological and genetic spatial patterns of *Urophora cardui* (Diptera: Tephritidae) as evidence for population structure and biogeographical processes. *Journal of Animal Ecology*, **63**, 187–199.

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

Fonseca DM, LaPointe DA, Fleischer RC (2000) Bottlenecks and multiple introductions: population genetics of the vector of avian malaria in Hawaii. *Molecular Ecology*, **9**, 1803–1814.

Gaskin JF, Schaal BA (2002) Hybrid tamarix widespread in U.S. invasion and undetected in native Asian range. *Proceedings of the National Academy of Sciences USA*, **99**, 11256–11259.

Geller JB (1999) Decline of native species masked by sibling species invasion: the case of Californian marine mussels. *Conservation Biology*, **13**, 661–664.

Gillespie JH (1991) *The Causes of Molecular Evolution*. Oxford University Press, New York.

Gillespie RG, Roderick GK (2002) Arthropods on islands: colonization, speciation, and conservation. *Annual Review of Entomology*, **47**, 595–632.

Goodisman MAD, Matthews RW, Crozier RH (2001) Hierarchical genetic structure of the introduced wasp *Vespula germanica* in Australia. *Molecular Ecology*, **10**, 1423–1432.

Grosholz E (2002) Ecological and evolutionary consequences of coastal invasions. *Trends in Ecology and Evolution*, **17**, 22–27.

Guenette S, Prichard RK, Klein RD, Matlashewski G (1991) Characterization of a beta-tubulin gene and a beta-tubulin gene products of *Brugia pahangi*. *Molecular Biochemical Parasitology*, **44**, 153–164.

Han MJ, Lee SH, Ahn SB, Choi JY, Choi CM (1994) Distribution, damage and host plant of pumpkin fruit fly, *Paradacus depressa* (Shiraki). *RDA Journal of Agricultural Science*, **36**, 346–350.

He M, Haymer DS (1997) Polymorphic intron sequences detected within and between populations of the oriental fruit fly (Diptera: Tephritidae). *Annals of the EntoMolecular Society of America*, **90**, 825–831.

Holway DA, Suarez AV (1999) Animal behavior: an essential component of invasion biology. *Trends in Ecology and Evolution*, **14**, 328–330.

Kim J-K (2002) Exporting citrus to the USA. In: *The Farmer's Newspaper*, **3077**, 1.

Kim CW, Kim JI (1974) Insect fauna of national park, Mt. Naejangsan in summer season. *Report of the Korean Association for Conservation of Nature*, **8**, 95–126.

Kim TH, Kim JS, Mun JH (1999) Distribution and bionomics of *Bactrocera (Paradacus) depressa* (Shriaki) in Chonbuk province. *Korean Journal of Soil Zoology*, **4**, 26–32.

Kinnear MW, Bariana HS, Sved JA, Frommer M (1998) Polymorphic microsatellite markers for population analysis of a tephritid pest species, *Bactrocera tryoni*. *Molecular Ecology*, **7**, 1489–1495.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Lodge DM (1993) Biological invasions: lessons for ecology. *Trends in Ecology and Evolution*, **8**, 133–137.

McPheron BA, Steck GJ (1996) *Fruit Fly Pests. A World Assessment of Their Biology and Management*. St Lucie Press, Delray Beach, FL.

Meixner MD, McPheron BA, Silva JG, Gasparich GE, Sheppard WS (2002) The Mediterranean fruit fly in California: evidence for multiple introductions and persistent populations based on microsatellite and mitochondrial DNA variability. *Molecular Ecology*, **11**, 891–899.

Mun JH, Kim JS, Song YH, Kim TH, Roderick GK (2000) Molecular genetic diagnosis of four fruit fly species (Tephritidae). *Journal of Asian-Pacific Entomology*, **3**, 89–94.

Mun JH, Song YH, Heong KL, Roderick GK (1999) Genetic variation among Asian populations of rice planthoppers, *Nilaparvata lugens and Sogatella furcifera* (Hemiptera: Delphacidae): mitochondrial DNA sequences. *Bulletin of EntoMolecular Research*, **89**, 245–253.

Nalepa TF, Schloesser DW (1992) *Zebra Mussels: Biology, Impacts and Control*. Lewis Publishers, Chelsea, MI.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, NY.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 205–247. Sinauer Associates, Sunderland, MA.

Pimentel D, Lack L, Suniga R, Morrison D (2000) Environmental and economic costs of nonindigenous species in the United States. *Bioscience*, **50**, 53–65.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Pritchard JK (2000). *Structure*. University of Chicago, Chicago.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

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 tephritid flies of economic importance. In: *Fruit Fly Pests. A World Assessment of Their Biology and Management* (eds McPheron BA, Steck GJ), pp. 267–271. St Lucie Press, Delray Beach, FL.

Roderick GK, Davies N, Bohonak AJ, Villablanca FX (1998) The interface of population genetics and systematics. In: *Pest Management — Future Challenges. Proceedings of the Sixth Australasian Applied Entomological Research Conference* (eds Zalucki MP, Drew RAI, White GG), pp. 489–499. University of Queensland, Brisbane, Australia.

Saltonstall K (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences USA*, **99**, 2445–2449.

Saul SH (1992) The medfly in California [Letter]. *Science*, **255**, 514.

Schneider S, Roessli D, Excoffier L (2001) ARLEQUIN: a software for population genetics data analysis. Geneva, Switzerland.

Shiraki T (1933) A systematic study of Trypetidae in the Japanese empire. *Memoirs of the Faculty of Science and Agriculture, Taihoku Imperial University*, **8**, 90–91.

Shiraki T (1968) Fruit flies of the Ryukyu Islands. *United States National Museum Bulletin*, **263**, 1–104.

Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–614.

Smith PT, Kambhampati S, Armstrong KA (2003) Phylogenetic relaitonships among Bactrocera species (Diptera: Tephritidae) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **26**, 8–17.

Smith AG, Smith DG, Funnell BM (1994) *Atlas of Mesozoic and Cenozoic Coastlines*. Cambridge University Press, NY.

Stark JD, Vargas RI, Walsh WA (1994) Temporal synchrony and patterns in an exotic host–parasitoid community. *Oecologia*, **100**, 196–199.

Steck GJ, Gasparich GE, Han HY, McPheron BA, Sheppard WS (1996) Distribution of mitochondrial DNA haplotypes among *Ceratitis capitata* populations worldwide. In: *Fruit Fly Pests. A World Assessment of Their Biology and Management* (eds McPheron BA, Steck GJ), pp. 291–296. St Lucie Press, Delray Beach, FL.

Sved JA, Yu H, Dominiak B, Gilchrist AS (2003) Inferring modes of colonization for pest species using heterozygosity comparisons and a shared-allele test. *Genetics*, **163**, 823–831.

Takamatsu Y (1952) Biological notes on the squash fly, *Zeugodacus depressus* Shiraki. *Oyo-Kontyu (Journal of the Nippon Society of Applied Entomology)*, **8**, 14–18.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR (2002) Out of Africa again and again. *Nature*, **416**, 45–51.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–352.

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, Routman EJ, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

Trivinos-Lagos L, Ohmachi T, Albrightson C *et al.* (1993) The highly divergent alpha- and beta-tubulins from *Dictyostelium discoideum* are encoded by single genes. *Journal of Cell Science*, **105**, 903–911.

United States Department of Agriculture (USDA) (2001) Fruit fly cooperative control program: final environmental impact statement, 2001. USDA Marketing and Regulatory Programs, Animal and Plant Health Inspection Service, Riverdale, MD.

Velleman PF (1997) *DataDesk for Macintosh*. Data Description Inc., Ithaca, NY, USA.

Villablanca FX, Roderick GK, Palumbi SR (1998) Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Molecular Ecology*, **7**, 351–365.

Voss HJ (1992) The medfly in California [Letter]. *Science*, **255**, 514.

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Sunderland, MA.

White IM (1996) Fruit fly taxonomy: recent advances and new approaches. In: *Fruit Fly Pests. A World Assessment of Their Biology and Management* (eds McPheron BA, Steck GJ), pp. 253–258. St Lucie Press, Delray Beach, FL.

White IM, Elson-Harris MM (1992) *Fruit Flies of Economic Significance: Their Identification and Bionomics*. CAB International, Wallingford.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Jeomhee Mun (http://nature.berkeley.edu/~jhmun) is a postdoctoral fellow at the University of California, Berkeley studying the population genetics of insect pests including true fruit flies, rice pests and arthropod vectors of disease. Andrew Bohonak (http://www.bio.sdsu.edu/pub/andy) is an Assistant Professor at San Diego State University with interests in freshwater biology, invertebrate population biology and population genetics of invading species. George Roderick (http://nature.berkeley.edu/~roderick) is a Professor at UC Berkeley and conducts research on the population biology and genetics of plant-feeding insects and other arthropods.
