ORIGINAL ARTICLE

Analysis of olive fly invasion in California based on microsatellite markers

NE Zygouridis^{1,3}, AA Augustinos^{1,3}, FG Zalom² and KD Mathiopoulos¹¹Department of Biochemistry and Biotechnology, University of Thessaly, Larissa, Greece and ²Department of Entomology, University of California, Davis, CA, USA

The olive fruit fly, *Bactrocera oleae*, is the main pest of the olive fruit and its expansion is exclusively restricted to the cultivation zone of the olive tree. Even though olive production has a century-old history in California, the olive fly was first detected in the Los Angeles area in 1998. Within 5 years of the first observation, the insect was reported from all olive cultivation areas of the state. Field-collected flies from five locations in California and another from Israel were analyzed on the basis of microsatellite polymorphisms in 10 microsatellite loci. These results were integrated with those

of a previous study of olive fly populations around the European part of the Mediterranean basin. The analysis pointed to the eastern part of the Mediterranean as the putative source of the observed invasion. Moreover, samples from California were quite different from Mediterranean samples implying the participation of phenomena such as genetic drift during the invasion and expansion of the olive fly in California.

Heredity (2009) **102**, 402–412; doi:10.1038/hdy.2008.125; published online 24 December 2008

Keywords: *Bactrocera oleae*; microsatellites; population analysis; Tephritidae; invasion; olive fly

Introduction

The olive fly, *Bactrocera oleae* (Gmelin), is a member of the Diptera family Tephritidae. Being the most important pest of cultivated olives, its management is a matter of great importance. The olive fly damages the fruit by 'stings' resulting from oviposition. The hatched larva feeds on the mesocarp of the olive, creating channels inside the fruit. The oviposition site may also serve as a point of entry for microbial infection of the fruit, compromising the organoleptic qualities of the olive. This may result in reduced commercial value of table olives and unmarketable olive oil because of high levels of acidity. These factors can result in an annual production loss of between 5 and 30%, depending on environmental conditions (Mazomenos, 1989; Katsoyanos, 1992).

Being a monophagous insect, the range of the olive fly is limited to the olive cultivation zones, as well as areas where olive trees are indigenous. Traditionally, this was restricted to the Mediterranean basin where most olive trees are grown. However, olive cultivation is expanding in areas such as South Africa, Australia, China and the Americas.

The Spanish first carried olive tree cuttings to the Americas in the middle of the sixteenth century. Around 1700, Franciscan monks brought olive trees to Mexico and from there to California. Intensive cultivation of the olive tree started in central and southern California at the

end of the nineteenth century. Today the California olive industry produces table and oil olives valued at about US\$80 million (California Department of Food and Agriculture, 2006). The olive fly was first found in California near Los Angeles International Airport in October 1998 (Rice, 2000). A statewide monitoring program subsequently detected the olive fly in all olive cultivation areas of the state over the following 5 years (Zalom *et al.*, 2008). However, the geographical origin of the insect's invasion remains unknown.

Analyses of natural olive fly populations support the subdivision into three groups: Pakistan, Africa and Mediterranean plus America (Nardi *et al.*, 2005), with the Mediterranean population further differentiated into eastern (Cypriot), central (Greek–Italian) and western (Iberian) groups (Augustinos *et al.*, 2005). The fundamental aim of the present study was a precise determination of the geographical origin of the California invasion of the olive fly within the Mediterranean group. Field-collected samples from five different locations of California and one from Israel were analyzed on the basis of polymorphism of 10 microsatellite loci. These markers constitute a subset of the markers described and used in the analysis of endemic olive fly populations around the Mediterranean basin by Augustinos *et al.* (2002, 2005). Microsatellites are particularly informative in the study of recent population phenomena, such as biological invasions (Bruford and Wayne, 1993; Schlötterer and Pemberton, 1994; Tautz and Schlötterer, 1994). Moreover, they have been successfully used in the analysis of population structure of different Tephritidae species, such as *Ceratitis capitata* (Bonizzoni *et al.*, 2001; Gasperi *et al.*, 2002), *Bactrocera dorsalis* (Aketarawong *et al.*, 2007) and *B. tryoni* (Yu *et al.*, 2001; Gilchrist *et al.*, 2004). In the case of *C. capitata*, they have provided guidance in establishing the source of recent invasions in

Correspondence: Dr KD Mathiopoulos, Department of Biochemistry and Biotechnology, University of Thessaly, Ploutonos 26, Larissa 41221, Greece.

E-mail: kmathiop@bio.uth.gr

³These authors contributed equally to this study.

Received 2 September 2008; revised 19 November 2008; accepted 28 November 2008; published online 24 December 2008

different parts of the world (Meixner *et al.*, 2002; Bonizzoni *et al.*, 2004). Such analyses can contribute to better planning of control strategies to avoid future infestations, as a detailed knowledge of the biology, genetic structure and geographical variability of a given species is a prerequisite to establishing control measures such as quarantine, phytosanitary control and eradication (Roderick and Navajas, 2003).

Materials and methods

Collection of fly samples and DNA extraction

Collection sites and number of flies used in the study are shown in Figure 1 and Table 1. Individual flies were collected during the olive-harvesting season from infested olive fruit. Flies were stored either at -80°C or in 95% ethanol at -20°C . Genomic DNA was extracted using the Wizard Genomic DNA extraction kit (Promega, Madison, WI, USA).

Genotyping

We used 10 microsatellite markers in the present analysis, named *Boms2*, 10, 18, 21, 22, 25, 29, 30, 31 and

61 (Augustinos *et al.* 2002, 2005). PCRs were performed in a total volume of $10\mu\text{l}$, containing 10 ng of genomic DNA, $1\times$ complete reaction buffer (New England Biolabs, Inc., Ipswich, MA, USA), 0.2 mM of each dNTP, $0.5\mu\text{M}$ of each primer and 0.4 U of Taq polymerase (New England Biolabs, Inc.). Amplifications were performed with an initial denaturation step (5 min at 95°C), followed by 30 cycles of: 15 s at 94°C , 30 s at 50°C and 30 s at 72°C , with a final elongation step of 5 min at 72°C . PCR products were electrophoresed on 5% denaturing polyacrylamide gels and visualized according to the Silver Sequence DNA Sequencing System technical manual (Promega).

Data analysis

Genetic variability was measured as the mean number of alleles per locus (n_a), effective number of alleles (n_e), observed (H_o) and expected heterozygosity (H_e), using POPGENE version 1.31 (Yeh *et al.*, 1999) and allelic richness after correction for sample size, using FSTAT (Goudet, 2001). Deviations from Hardy–Weinberg equilibrium (HWE) were tested with the G^2 likelihood ratio test in POPGENE. G^2 criterion [defined as (Observed

Figure 1 Collection sites and geographical representation of the clustering outcome, assuming three hypothetical clusters. (a) Map showing the collection sites in the Mediterranean basin. 1, Gimmaraes-PO; 2, Lisbon-PO; 3, Murcia-SP; 4, Madrid-SP; 5, Arrhenys-SP; 6, Farfa-IT; 7, Vasto-IT; 8, Alexandroupolis-GR; 9, Lefkas-GR; 10, Patras-GR; 11, Mani-GR; 12, Ithaca-GR; 13, Kos-GR; 14, Kythira-GR; 15, Maladrino-GR; 16, Crete-GR; 17, Aidin-TU; 18, Nicosia-CY; 19, Limassol-CY; 20, Sde Boker-IS. TU, Turkey; IS, Israel; CY, Cyprus; GR, Greece; IT, Italy; SP, Spain; PO, Portugal. (b) Map showing the collection sites in California. 21, Calaveras; 22, Napa; 23, Solano; 24, Yolo Davis; 25, San Luis Obispo. Colored components in each pie show the co-ancestry distribution of individuals in each one of the three clusters. The colour reproduction of this figure is available on the html full text version of the manuscript.

Table 1 *B. oleae* field-collected samples

Country	Sample site	Sample name	Date	Coordinates		Sample size
				Latitude	Longitude	
Portugal	Gimmaraes Lisbon	Gim-PO	August 2002	41° 26' N	8° 20' W	30
		Lis-PO	August 2002	38° 45' N	9° 11' W	29
Spain	Murcia	Mur-SP	September 2001	37° 58' N	1° 8' W	50
	Madrid	Mad-SP	August 2002	40° 26' N	3° 43' W	29
	Arrhenys	Arr-SP	August 2002	41° 36' N	2° 33' E	30
Italy	Farfa	Far-IT	October 1999	42° 13' N	12° 43' E	50
	Vasto	Vas-IT	September 2002	42° 7' N	14° 43' E	30
Greece	Alexandroupolis	Ale-GR	September 2001	40° 52' N	25° 53' E	50
	Lefkada	Lef-GR	September 2001	38° 50' N	20° 42' E	48
	Patras	Pat-GR	September 2001	38° 15' N	21° 44' E	50
	Mani	Man-GR	September 2001	36° 33' N	22° 25' E	50
	Ithaca	Ith-GR	September 2001	38° 21' N	20° 43' E	23
	Kos	Kos-GR	September 2001	36° 54' N	27° 17' E	21
	Kythira	Kyt-GR	September 2001	36° 9' N	22° 59' E	25
	Maladrino	Mal-GR	September 2001	38° 27' N	22° 14' E	50
	Crete	Cre-GR	September 2001	35° 20' N	25° 8' E	43
Turkey	Aidin	Aid-TU	October 2002	37° 52' N	27° 50' E	9
Cyprus	Limassol	Lim-CY	September 2002	34° 41' N	33° 3' E	30
	Nicosia	Nic-CY	September 2002	35° 10' N	33° 22' E	24
Israel	Sde Boker	Sde-IS	October 2007	30° 52' N	34° 47' E	18
California	Calaveras	Cal-CA	April 2004	38° 27' N	120° 31' W	30
	Napa	Nap-CA	April 2004	38° 13' N	122° 17' W	30
	Solano	Sol-CA	April 2004	38° 35' N	121° 98' W	30
	Yolo Davis	Yol-CA	April 2004	38° 34' N	121° 51' W	30
	San Luis Obispo	SLO-CA	April 2004	35° 20' N	120° 43' W	30

value) $\times \ln(\text{Observed value}/\text{Expected value})$. Genotypic disequilibrium was tested with Genepop (Raymond and Rousset, 1995), with Fisher's exact test, for all pairs of loci in all samples and across samples.

Genetic distances were measured according to Nei (1972) using POPGENE (Yeh *et al.*, 1999). A UPGMA dendrogram was constructed in PHYLIP 3.6C (Felsenstein, 1994), using the allele frequencies. The robustness of each node was assessed by the bootstrap method. For this purpose, 100 pseudoreplicates were generated by random resampling of the original data, using SEQBOOT and GENDIST. The new distance matrices were then subjected to NEIGHBOR and CONSENSE to produce a consensus tree. Visualization of the dendrogram was performed by TreeView32 software (Page, 1996).

Pairwise F_{ST} values were estimated using FSTAT software. Inference on the degree of population subdivision based on hierarchical analysis of molecular variance (AMOVA) was performed by the ARLEQUIN 2.0 software (Schneider *et al.*, 1997), to compare the percentage of genetic variability attributed to the variance among the major geographical areas sampled to that observed among samples within each of them. The significance of the resultant F statistics and variance components were tested with 10 000 permutations.

STRUCTURE software was used to determine the number of possible genetic clusters in California (Pritchard *et al.*, 2000; Falush *et al.*, 2003). We used all four different models, with a burn-in period of 50 000 and 50 000 Markov chain Monte Carlo (MCMC) repetitions after the initial burn-in. We tested for $K=1$ to $K=5$ and

performed five runs of each model, to check the consistency of the results. In a previous study we observed at least three subpopulations of the olive fly in the Mediterranean, but these were not well differentiated due to gene flow. We used again the four different models, with a burn-in period of 50 000 and 50 000 MCMC repetitions after the initial burn-in, to analyze the structuring of all 25 samples. We ran five repetitions of all above models (to check the consistency of our results), assuming $K=1$ to $K=10$. Although the authors of the program recommend the use of the admixture model, they also suggest that the no-admixture model is more powerful at detecting subtle structure, such as in our case. Because the likelihood curve did not provide a clear K -value (Figure 3a), we estimated a modified K -value, taking also into account the consistency of the results. On the basis of modification described by Evanno *et al.* (2005) we divided the mean likelihood value of the five runs by their standard deviation and considered this as the modified new K -curve (Figure 3b).

GENALEX 6.1 software (Peakall and Smouse, 2006) was used to estimate pairwise population Φ_{PT} values and Φ_{PT} matrix was used to perform principal components analysis (PCA). Φ_{PT} measure suppresses within-population variance and simply calculates population differentiation based on the genotypic variance.

GeneClass 2.0 software (Piry *et al.*, 2004) was used to perform population assignment and exclusion test and to calculate the probability of origin for each individual and each sample. We also tried to assign California samples,

assuming four subpopulations in the Mediterranean basin (the three previously described along with the Israeli). We chose a Bayesian model using the Rannala and Mountain (1997) criterion, under a 0.05 threshold.

Finally, we used BOTTLENECK software (Cornuet and Luikart, 1997) to detect any recent bottleneck phenomena, especially in California. We chose the Wilcoxon sign-rank test because it performs better than the other two tests with few loci.

Results

The Mediterranean basin is presumably the source of the olive fly invasion in California (Nardi *et al.*, 2005). To further specify the origin of this invasion, microsatellite polymorphism data of California samples were compared with the preexisting data set of Mediterranean samples (Augustinos *et al.*, 2005). To expand the Mediterranean data set, a sample from Israel was also included in the analysis. For this purpose, 150 field-collected flies from five different locations in California (Calaveras, Napa, Yolo, Solano and San Luis Obispo) and 18 from Israel (Sde Boker), as shown in Figure 1 and Table 1, were genotyped at 10 microsatellite loci.

Hardy–Weinberg equilibrium and linkage disequilibrium Conformation to HWE was tested in California and Israeli samples according to the G^2 criterion at a significance level of 5%. Out of 60 tests performed (6 samples $\times$ 10 loci), none of them showed departure from HWE after the sequential Bonferroni correction (Rice, 1989). Interestingly, testing HWE under the assumption of a single California population, two loci deviated: *Boms25* due to homogeneity excess and *Boms29* due to an excess of one class of homozygotes. This contrasts the situation in the Mediterranean basin where no deviations were observed (Augustinos *et al.*, 2005) and may evidence the fact that there is not a single, homogeneous population of the olive fly in California. This could be the result of the very rapid expansion of the species in a large geographical area, where genetic drift and subsequent allele adaptation in particular ecological niches took place faster than the homogenizing effects of unrestricted gene flow. Linkage disequilibrium was observed only in few of the tests performed. However, after correcting for multiple tests, no linkage disequilibrium was statistically significant.

Microsatellite variability in California and Israel

Genetic variability was measured as the average number of actual (n_a) and effective alleles (n_e), allelic richness, and degree of observed (H_o) and expected (H_e) heterozygosity (Table 2). A westward decline in the levels of polymorphism (H_e values) in the olive fly Mediterranean populations has already been reported (Augustinos *et al.*, 2005). The trend holds for the Israeli sample ($H_e = 0.59$), whose value is greater than those of the Central Med samples, even though it is a small sample and estimation errors cannot be excluded. California n_e , H_o and H_e values are intermediate between those observed in the West Med and the Central Med groups. Allelic richness in California is the lowest among all regions, which is in accordance with the hypothesis of a Mediterranean source of olive fly invasion in California (Nardi *et al.*, 2005).

Table 2 Genetic variability estimates of olive fruit fly samples

Sample	N	n_a	AR	n_e	H_o	H_e
Gim-PO	30	4.00	3.27	2.26	0.48	0.49
Lisbon-PO	29	5.10	3.60	2.20	0.52	0.49
Mur-SP	50	5.20	3.47	2.30	0.49	0.49
Mad-SP	29	4.00	3.28	2.23	0.48	0.48
Arr-SP	30	5.00	3.59	2.28	0.51	0.50
West Med group	168	7.70	3.44	2.28	0.50	0.49
Far-IT	50	5.80	3.83	2.59	0.59	0.57
Vas-IT	30	5.70	3.94	2.44	0.55	0.55
Ale-GR	50	5.30	3.57	2.57	0.54	0.57
Lef-GR	48	5.70	3.92	2.64	0.60	0.58
Pat-GR	50	6.00	3.95	2.78	0.62	0.58
Man-GR	50	5.70	3.82	2.70	0.56	0.57
Ith-GR	23	4.70	3.63	2.65	0.57	0.58
Kos-GR	21	4.50	3.82	2.59	0.59	0.57
Kyt-GR	25	4.50	3.72	2.66	0.58	0.58
Mal-GR	50	6.10	3.66	2.56	0.57	0.56
Cre-GR	43	5.10	3.56	2.45	0.57	0.55
Aid-TU	9	3.90	3.90	2.64	0.60	0.58
Central Med group	449	9.60	3.78	2.66	0.58	0.57
Lim-CY	30	4.70	3.74	2.84	0.59	0.60
Nic-CY	24	4.70	3.91	2.84	0.62	0.61
Cypriot group	54	5.70	3.83	2.92	0.60	0.61
Sde-IS	18	4.50	3.79	2.58	0.56	0.59
Cal-CA	30	3.60	3.09	2.39	0.48	0.54
Nap-CA	30	3.70	3.13	2.38	0.48	0.53
Sol-CA	30	3.60	3.18	2.38	0.54	0.53
Yol-CA	30	3.60	3.15	2.41	0.53	0.52
SLO-CA	30	3.50	3.00	2.23	0.51	0.49
Californian group	150	3.60	3.11	2.36	0.51	0.52

Abbreviations: AR, allelic richness, corrected for sample size; H_e , heterozygosity expected; H_o , heterozygosity observed; N, sample size; n_a , actual allele number; n_e , effective allele number. Regional group values are indicated in bold.

Analysis of alleles gave one new allele in California, albeit at very low frequency. In addition, three of the California alleles were not sampled in the Iberian Peninsula. Given the large number of flies collected in Spain and Portugal (168), it is highly unlikely that these three alleles are present in this area. Moreover, no private alleles were found among West Med and California, West Med, East Med and California, and among West and East Med. This indicates that the West Med samples are quite discrete from the East Med and California samples.

Population structure in California

Estimated pairwise F_{ST} values in California were all not significant except for one (Sol-CA/SLO-CA), as shown in Table 3. Furthermore, all F_{ST} values among California samples and Mediterranean samples are significant (except those of Aid-TU), although at different probability levels. Genetic distances among California samples were measured according to Nei (1972; Table 3) and these varied between 0.0124 (Sol-CA/Yol-CA) and 0.0450 (SLO-CA/Sol-CA).

Further analysis of olive fly population structure in California was performed using the STRUCTURE software. Five independent runs of all four models were performed, assuming K ranging between 1 and 5. There seems to be no clustering of California samples. To further analyze the genetic structure, pairwise Φ_{PT} values were calculated and Φ_{PT} pairwise matrix was used to perform PCA. Still, no clustering of samples inside California was observed.

Table 3 Genetic distance values according to Nei (1972) and significance of F_{st} values

	Gim	Lis	Mur	Mad	Arr	Far	Vas	Ale	Lef	Pat	Mal	Man	Ith	Kos	Kyt	Cre	Aid	Lim	Nic	Sde	Cal	Nap	Sol	Yol	SLO
Gim-PO	-	NS 0.0027	NS 0.0055	NS 0.0149	NS -0.0033	NS 0.0373	NS 0.0372	NS 0.0288	NS 0.0399	NS 0.0330	NS 0.0316	NS 0.0373	NS 0.0304	NS 0.0532	NS 0.0314	NS 0.0351	NS 0.0403	NS 0.1122	NS 0.0982	NS 0.0907	NS 0.0719	NS 0.0951	NS 0.0906	NS 0.0832	NS 0.1145
Lis-PO	0.0185	-	NS 0.0084	NS 0.0118	NS 0.0106	NS 0.0495	NS 0.0469	NS 0.0322	NS 0.0431	NS 0.0408	NS 0.0368	NS 0.0428	NS 0.0419	NS 0.0620	NS 0.0373	NS 0.0424	NS 0.0437	NS 0.1015	NS 0.0943	NS 0.0813	NS 0.0706	NS 0.0848	NS 0.0861	NS 0.0793	NS 0.1114
Mur-SP	0.0184	0.0211	-	NS 0.0340	NS 0.0029	NS 0.0339	NS 0.0446	NS 0.0313	NS 0.0295	NS 0.0406	NS 0.0285	NS 0.0402	NS 0.0329	NS 0.0480	NS 0.0303	NS 0.0397	NS 0.0320	NS 0.1073	NS 0.1151	NS 0.0813	NS 0.0645	NS 0.0950	NS 0.0869	NS 0.0828	NS 0.1070
Mad-SP	0.0306	0.0270	0.0473	-	NS 0.0195	NS 0.0540	NS 0.0368	NS 0.0480	NS 0.0554	NS 0.0425	NS 0.0409	NS 0.0527	NS 0.0503	NS 0.0582	NS 0.0430	NS 0.0505	NS 0.0508	NS 0.1199	NS 0.0971	NS 0.0946	NS 0.0981	NS 0.1130	NS 0.1208	NS 0.1116	NS 0.1478
Arr-SP	0.0129	0.0261	0.0159	0.0353	-	NS 0.0398	NS 0.0334	NS 0.0341	NS 0.0348	NS 0.0357	NS 0.0291	NS 0.0391	NS 0.0356	NS 0.0456	NS 0.0326	NS 0.0433	NS 0.0432	NS 0.1096	NS 0.1121	NS 0.0845	NS 0.0763	NS 0.1052	NS 0.1042	NS 0.0969	NS 0.1231
Far-IT	0.0568	0.0723	0.0486	0.0777	0.0612	-	NS 0.0133	NS 0.0080	NS 0.0091	NS 0.0095	NS 0.0045	NS 0.0068	NS 0.0075	NS 0.0185	NS -0.0007	NS 0.0050	NS 0.0044	NS 0.0619	NS 0.0646	NS 0.0352	NS 0.0427	NS 0.0620	NS 0.0623	NS 0.0487	NS 0.0751
Vas-IT	0.0595	0.0707	0.0648	0.0579	0.0555	0.0341	-	NS 0.0195	NS 0.0206	NS 0.0133	NS 0.0115	NS 0.0130	NS 0.0227	NS 0.0072	NS 0.0119	NS 0.0217	NS -0.0057	NS 0.0640	NS 0.0610	NS 0.0326	NS 0.0428	NS 0.0577	NS 0.0653	NS 0.0507	NS 0.0798
Ale-GR	0.0469	0.0506	0.0458	0.0703	0.0545	0.0236	0.0430	-	NS 0.0083	NS 0.0002	NS 0.0012	NS -0.0020	NS -0.0041	NS 0.0078	NS -0.0002	NS 0.0016	NS 0.0001	NS 0.0374	NS 0.0274	NS 0.0253	NS 0.0142	NS 0.0250	NS 0.0325	NS 0.0199	NS 0.0385
Lef-GR	0.0612	0.0647	0.0434	0.0810	0.0555	0.0257	0.0453	0.0249	-	NS 0.0066	NS 0.0026	NS 0.0102	NS 0.0107	NS 0.0069	NS 0.0036	NS 0.0113	NS 0.0019	NS 0.0484	NS 0.0529	NS 0.0277	NS 0.0222	NS 0.0443	NS 0.0460	NS 0.0391	NS 0.0494
Pat-GR	0.0514	0.0611	0.0575	0.0626	0.0561	0.0257	0.0343	0.0133	0.0228	-	NS 0.0013	NS -0.0018	NS 0.0004	NS -0.0009	NS 0.0009	NS 0.0041	NS -0.0022	NS 0.0426	NS 0.0267	NS 0.0287	NS 0.0272	NS 0.0339	NS 0.0463	NS 0.0310	NS 0.0434
Mal-GR	0.0497	0.0556	0.0420	0.0602	0.0473	0.0185	0.0316	0.0146	0.0166	0.0143	-	NS 0.0031	NS -0.0013	NS -0.0001	NS -0.0024	NS 0.0041	NS -0.0045	NS 0.0468	NS 0.0459	NS 0.0280	NS 0.0321	NS 0.0492	NS 0.0482	NS 0.0441	NS 0.0663
Man-GR	0.0577	0.0644	0.0572	0.0770	0.0612	0.0223	0.0345	0.0109	0.0280	0.0109	0.0172	-	NS 0.0001	NS 0.0049	NS -0.0019	NS -0.0003	NS -0.0104	NS 0.0318	NS 0.0304	NS 0.0201	NS 0.0253	NS 0.0267	NS 0.0376	NS 0.0232	NS 0.0432
Ith-GR	0.0545	0.0680	0.0534	0.0786	0.0621	0.0311	0.0560	0.0163	0.0372	0.0217	0.0190	0.0221	-	NS 0.0063	NS -0.0002	NS -0.0024	NS 0.0021	NS 0.0443	NS 0.0365	NS 0.0395	NS 0.0315	NS 0.0411	NS 0.0540	NS 0.0354	NS 0.0639
Kos-GR	0.0836	0.0939	0.0731	0.0880	0.0749	0.0464	0.0348	0.0328	0.0319	0.0202	0.0212	0.0291	0.0389	-	NS 0.0062	NS 0.0189	NS -0.0124	NS 0.0388	NS 0.0342	NS 0.0250	NS 0.0206	NS 0.0367	NS 0.0457	NS 0.0381	NS 0.0499
Kyt-GR	0.0549	0.0613	0.0493	0.0682	0.0574	0.0189	0.0400	0.0204	0.0260	0.0215	0.0165	0.0184	0.0293	0.0377	-	NS -0.0033	NS -0.0069	NS 0.0249	NS 0.0347	NS 0.0145	NS 0.0279	NS 0.0387	NS 0.0503	NS 0.0375	NS 0.0602
Cre-GR	0.0541	0.0625	0.0561	0.0722	0.0652	0.0195	0.0450	0.0157	0.0285	0.0183	0.0184	0.0132	0.0177	0.0462	0.0156	-	NS 0.0052	NS 0.0481	NS 0.0440	NS 0.0337	NS 0.0378	NS 0.0374	NS 0.0517	NS 0.0315	NS 0.0561
Aid-TU	0.0839	0.0849	0.0700	0.0942	0.0880	0.0474	0.0384	0.0442	0.0460	0.0387	0.0358	0.0303	0.0556	0.0337	0.0421	0.0476	-	NS 0.0262	NS 0.0336	NS 0.0169	NS 0.0162	NS 0.0262	NS 0.0309	NS 0.0213	NS 0.0474
Lim-CY	0.1777	0.1572	0.1621	0.1877	0.1758	0.1141	0.1195	0.0737	0.0949	0.0842	0.0875	0.0660	0.0975	0.0858	0.0646	0.0879	0.0905	-	NS 0.0183	NS 0.0110	NS 0.0384	NS 0.0328	NS 0.0579	NS 0.0468	NS 0.0582
Nic-CY	0.1540	0.1454	0.1766	0.1486	0.1808	0.1214	0.1166	0.0609	0.1055	0.0610	0.0885	0.0667	0.0869	0.0805	0.0831	0.0838	0.1040	0.0561	-	NS 0.0272	NS 0.0346	NS 0.0174	NS 0.0406	NS 0.0331	NS 0.0446
Sde-IS	0.1428	0.1263	0.1232	0.1459	0.1352	0.0762	0.0753	0.0622	0.0675	0.0680	0.0647	0.0555	0.0961	0.0711	0.0560	0.0718	0.0836	0.0496	0.0778	-	NS 0.0391	NS 0.0278	NS 0.0383	NS 0.0387	NS 0.0523
Cal-CA	0.1042	0.1012	0.0902	0.1379	0.1114	0.0740	0.0770	0.0360	0.0469	0.0531	0.0588	0.0509	0.0678	0.0525	0.0617	0.0663	0.0687	0.0758	0.0724	0.0838	-	NS 0.0025	NS 0.0118	NS 0.0018	NS 0.0115
Nap-CA	0.1346	0.1187	0.1305	0.1572	0.1513	0.1010	0.0967	0.0493	0.0772	0.0614	0.0815	0.0519	0.0800	0.0728	0.0755	0.0648	0.0801	0.0658	0.0464	0.0661	0.0262	-	NS 0.0130	NS -0.0080	NS 0.0043
Sol-CA	0.1260	0.1184	0.1174	0.1675	0.1472	0.1001	0.1056	0.0579	0.0781	0.0779	0.0789	0.0654	0.0964	0.0829	0.0901	0.0828	0.0811	0.1018	0.0771	0.0778	0.0350	0.0359	-	NS 0.0085	NS 0.0249
Yol-CA	0.1162	0.1098	0.1124	0.1535	0.1373	0.0802	0.0853	0.0414	0.0683	0.0560	0.0732	0.0459	0.0698	0.0728	0.0720	0.0557	0.0699	0.0848	0.0665	0.0790	0.0237	0.0124	0.0292	-	NS 0.0074
SLO-CA	0.1510	0.1450	0.1384	0.1975	0.1654	0.1113	0.1181	0.0607	0.0760	0.0670	0.0977	0.0673	0.1010	0.0807	0.0956	0.0824	0.0933	0.0927	0.0740	0.0889	0.0320	0.0242	0.0450	0.0258	-

Abbreviations: CA, California; CY, Cyprus; GR, Greece; IS, Israel; IT, Italy; NS, not significant; PO, Portugal; SP, Spain; TU, Turkey.

*** $P < 0.001$; ** $P < 0.005$; * $P < 0.01$.

Below diagonal: Genetic distance values according to Nei (1972), based on allele frequency differences. ^[41] Above diagonal: F_{st} values and significance of population differentiation, estimated by F_{st} values.

West Med group	Central Med group
East Med group	California

Figure 2 UPGMA dendrogram constructed after 100 bootstrap resamples, based on Nei's (1972) genetic distances, showing the relationships among the 25 samples studied. TU, Turkey; IS, Israel; CY, Cyprus; GR, Greece; IT, Italy; SP, Spain; PO, Portugal; CA, California.

Worldwide structuring of olive fly samples

Analysis of significance of pairwise F_{ST} values among all samples showed the possibility of clustering in four subpopulations (Table 3). West Med samples are significantly different from all other samples. Similarly, California samples constitute a separate group, as they are significantly different from all other samples. Central Med samples form yet another group, whereas Cypriot and Israeli samples seem to comprise a further, East Med cluster. Aid-TU sample is very small (nine individuals) and this is probably the reason it presents no significant difference with almost any other sample. Genetic distances among California and all the other samples ranged between 0.0459 (Yol-CA/Man-GR) and 0.1975 (SLO-CA/Mad-SP). The Israeli sample (Sde-IS) showed a minimum genetic distance with Lim-CY (0.0496) and a maximum genetic distance with Mad-SP (0.1459). These values were used to construct an unrooted UPGMA dendrogram (Figure 2). The topology of the dendrogram was confirmed by relatively high bootstrap values at the main nodes (100 and 78%). The major clusters as well as further groupings within these clusters were identical with those of the Augustinos *et al.* (2005) analysis, albeit generated with 10 (of the original 12) microsatellite loci. There are the three previously described subpopulations of Western Mediterranean (Spain and Portugal), Central Mediterranean (Greece, Italy and Turkey) and Eastern Mediterranean (Cyprus). Interestingly, samples from Israel and California fall in this third cluster. These results point to the eastern part of the Mediterranean

basin as a putative source of the olive fly invasion in California.

STRUCTURE analysis was performed, assuming K ranging between 1 and 10. The no-admixture model provided a better clustering of our samples. Although we expect our samples to be admixed, the no-admixture model produces better results when genetic structure is not very clear due to high gene flow (as in our case). It seems that there are at least three subpopulations, but the likelihood curve is not clear as to whether K can be greater than 3 (Figure 3a). K -curve continues growing after the value of 3, presumably beyond the true K value. However, independent runs are not so consistent, especially when assumed K grows. To incorporate consistency to our results and based on the modification described by Evanno *et al.* (2005), we estimated a new K -curve by dividing the mean likelihood of five independent runs by their standard deviation (Figure 3b). This way it becomes apparent that the true number of clusters must be three and, surely, no more than four. Partition of color components in Figure 4 ($K=3$) substantially differentiates West Med (1–5) from East Med flies (18–20), whereas Central Med flies fall into a 'transition zone' that seems to share a great degree of ancestry either with Iberian or with Cypriot-Israeli populations. However, California flies (21–25) are clearly more alike those of the East Med. Table 4 presents the average number of ancestry probabilities of each one of the 25 samples in each of the three hypothetical clusters, whereas Figure 1 shows how the assumed three genetic

Figure 3 STRUCTURE analysis: mean variance of likelihood for all samples, assuming $K=1$ to $K=10$ under the no-admixture with correlated allele frequencies model. Each value represents the mean of five independent runs. (a) Estimation of the true number of clusters, based on the likelihood (K) curve. (b) Estimation of the true number of clusters, based on the likelihood (K) curve, taking into account consistency of the results. The colour reproduction of this figure is available on the html full text version of the manuscript.

Figure 4 STRUCTURE analysis: estimated population structure for all samples based on allele frequency variation. Each individual is represented by a vertical line which is partitioned into $K=3$ colored components. 1–2, Portugal; 3–5, Spain; 6–7, Italy; 8–16, Greece; 17–18, Cyprus; 19, Turkey; 20, Israel; 21–25, California. The colour reproduction of this figure is available on the html full text version of the manuscript.

Table 4 Average coefficients of ancestry obtained from a STRUCTURE run, with $K=3$, for the 839 wild flies collected from 25 sampling locations around Mediterranean and California

Area	Samples	Clusters		
		1	2	3
West Med group	Gim-PO	0.729	0.210	0.061
	Lisbon-PO	0.569	0.238	0.193
	Mur-SP	0.606	0.235	0.159
	Mad-SP	0.610	0.357	0.033
	Arr-SP	0.654	0.258	0.088
Central Med group	Far-IT	0.409	0.472	0.119
	Vas-IT	0.280	0.552	0.169
	Ale-GR	0.349	0.365	0.286
	Lef-GR	0.376	0.383	0.241
	Pat-GR	0.276	0.463	0.261
	Man-GR	0.375	0.452	0.173
	Ith-GR	0.359	0.423	0.218
	Kos-GR	0.350	0.520	0.130
	Kyt-GR	0.190	0.537	0.273
	Mal-GR	0.302	0.508	0.190
East Med group	Cre-GR	0.381	0.338	0.281
	Aid-TU	0.218	0.515	0.267
	Lim-CY	0.154	0.445	0.401
California	Nic-CY	0.126	0.570	0.304
	Sde-IS	0.227	0.341	0.432
	Cal-CA	0.157	0.135	0.708
California	Nap-CA	0.041	0.081	0.878
	Sol-CA	0.074	0.094	0.832
	Yol-CA	0.087	0.074	0.839
	SLO-CA	0.133	0.123	0.744

Abbreviations: CA, California; CY, Cyprus; GR, Greece; IS, Israel; IT, Italy; PO, Portugal; SP, Spain; TU, Turkey.

The highest value of co-ancestry of each population in a cluster is in bold.

groups are distributed among the 25 samples. It is obvious that individuals of each sample are distributed to all three classes, probably due to the high gene flow between olive fly natural populations in the Mediterranean basin.

To further analyze structuring of olive fly populations, PCA was performed with the GENALEX software. Initially, PhiPT values were estimated and the PhiPT pairwise matrix was used to perform PCA. This analysis clearly demonstrates that there are at least four clusters: the first is formed by West Med samples (Iberian Peninsula), the second by Central Med samples (Greece, Italy and Turkey), the third by East Med samples (Cyprus and Israel) and the fourth by California samples (Figure 5). However, it is clear that California samples are most closely related to the East Med group and most distantly to the West Med group.

Analysis of molecular variance

AMOVA was performed to test the degree of homogeneity among populations. Samples were grouped in four populations, based on geographical criteria (West Med, Central Med, East Med and California) and on PCA results. Most variation is attributed to within-sample variability (95.6% of the total variance, $F_{ST}=0.04397$). However, a substantial proportion is due to differences among groups (3.79%). Only a very small, nonsignificant proportion of the variance derives from within-group variation (0.6%).

Migration

GeneClass version 2.0 was used to assign all individuals to their respective samples and 373 of the 840 individuals were correctly identified (44.4%). When all Mediterranean samples were clustered into five subpopulations

Figure 5 Principal components analysis of molecular variance performed in GENALEX 6.1, using the PhiPT pairwise distance matrix. The first two axes explain 71.56% of the total variance. The colour reproduction of this figure is available on the html full text version of the manuscript.

(West Med, Central Med, Cypriot, Israeli and California), 651 of the 840 individuals (77.5%) were correctly assigned (Appendix 1). It should be noted that for this analysis we divided East Med cluster into Cypriot and Israeli, in an effort to identify more precisely the source of the California invasion. West Med individuals were properly assigned as such at 82.1%, whereas 12.5% were assigned as Central Med, very few as Israeli and Californian, and none as Cypriot. Individuals from the Central Med subpopulation presented the lowest degree of correct assignment (34.7% for respective samples—60.2% for the Central Med cluster), illustrating the extensive gene flow occurring between natural olive fly populations around in Central Mediterranean basin, where Greece and Italy have been trading centers for olive products for thousands of years. The same analysis for flies from California correctly assigned 69 of the 150 individuals (46%) in their respective samples and, more importantly, 131 of the 150 (87.3%) individuals were correctly identified as Californian. Therefore, California olive flies may form a discrete subpopulation that has differentiated from the ancestral Mediterranean populations. To define the possible origin of the invasion, the five California samples were assigned to one of the three described subpopulations around the Mediterranean plus Israel. Cal-CA and Nap-CA were identified as Cypriot, whereas the remaining three were identified as Israeli (Appendix 2).

Analysis of bottleneck

Olive fly presence in California is recent (first observed in 1998), but it has expanded in the whole cultivation zone in the state. We used BOTTLENECK software to test for any recent bottleneck phenomena in the area. Results suggest that the olive fly in California has not faced any bottlenecks.

Discussion

Invasions, such as those of the olive fly, are often characterized by a unique set of demographic and genetic features that result from a small number of colonizing individuals and the rapid growth and spread of new populations (Davies and Roderick, 1999). We searched for such features in the relatively recent olive fly invasion in California to determine its origin. For this purpose, 150 wild flies captured from five different sampling areas in California were genotyped for 10 microsatellite loci. These results were integrated to those of a previous study of olive fly populations around the European part of the Mediterranean basin (Augustinos *et al.*, 2005). In that study, three subpopulations were distinguished: Western Mediterranean (constituted of samples from the Iberian Peninsula), Central Mediterranean (samples from Greece, Italy and Turkey) and Eastern Mediterranean (samples from Cyprus).

Iberia (most likely) absolved

Normally, a colonization process goes along with a loss of genetic variability because the founder population carries only a fraction of the original population's variability. In the case of California samples, three alleles were encountered (in relatively high frequencies) that were not observed in the West Med subpopulation. Given the large number of flies analyzed from West Med, it is highly unlikely that these alleles are present in this area. In addition, there were no shared alleles between California and West Med (that is, alleles were present in these two areas but not anywhere else). Furthermore, genetic distance values are the greatest between California and West Med samples and these two areas are placed in two different nodes of a dendrogram with a bootstrap value of 100. STRUCTURE analysis also shows little possibility of California and West Med samples

being in the same genetic cluster. PCA also suggests that California samples are more distantly related to the West Med ones. Finally, Bayesian assignment of groups of individuals shows that it is highly unlikely that California genotypes derive from West Mediterranean genotypes.

Greece and Italy remotely linked

The first major discrimination derived from the UPGMA dendrogram (with high bootstrap value) is the one between Eastern Med plus California samples and the rest of the Mediterranean. All analyses presented in this study show that samples from Greece, Italy and possibly Turkey form a distinct subpopulation, which is also mixed with western and/or eastern Mediterranean subpopulations, likely due to extensive gene flow in the area (that could be explained by extensive trade throughout history). STRUCTURE analysis also suggests that the Central Med cluster might be different from the cluster of Eastern Mediterranean and California, although it seems to share a certain degree of ancestry with this group. Therefore, it does not seem very likely that areas of the Central Mediterranean are the source of the California invasion. Finally, assignment of groups of individuals performed by GeneClass software gives a low probability only to the Cal-CA sample to have a Central Med origin.

East Med culpa?

Instead, all analyses of our data point to the Eastern Mediterranean as the putative source of the observed invasion. Eastern Mediterranean populations are clearly more polymorphic than California samples. Although California samples tend to form a cluster of their own, it is obvious that they are more closely related to East Med samples, either based on genetic distance values and allele frequencies or under STRUCTURE analysis and PCA. An interesting finding comes from the GeneClass analysis. When asked to assign California samples to one of the four groups assumed, three were identified as Israeli (Sol-CA, Yol-CA and SLO-CA) and the remaining two as Cypriot (Cal-CA and Nap-CA). Therefore, all were identified as East Med samples.

Olive fly in California: single or multiple invasions?

Fly collection data from the California Department of Food and Agriculture and the statewide monitoring program (through which the flies used in this analysis were obtained) suggest a rapid and virtually complete colonization by the olive fly within the range of olive cultivation in California. Till now, it remains unclear whether the fly invaded once in 1998 when it was first detected, or additional times either before the first detection or since. The fact that only a small fraction of the alleles that are found in Mediterranean are also present in California (44 out of 106) can be evidence of a flow of a small number of alleles from the Mediterranean basin toward California. Analysis of olive flies collected in subsequent years throughout the invasion time and from more collection sites will show which of the two prementioned scenarios is more likely.

Acknowledgements

We thank Hannah Joy Burrack for all the California fly shipments, David Nestel for the Israeli samples and Zissis Mamuris for several useful discussions with the authors. This research was partially supported by grants to KDM from the Greek Ministry of National Education and Religious Affairs. NEZ was partially supported by a fellowship from the National Fellowship Foundation (Greece).

References

- Aketarawong N, Bonizzoni M, Thanaphum S, Gomulski LM, Gasperi G, Malacrida AR *et al.* (2007). Inferences on the population structure and colonization process of the invasive oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Mol Ecol* **16**: 3522–3532.
- Augustinos AA, Mamuris Z, Stratikopoulos E, D'Amelio S, Zacharopoulou A, Mathiopoulos KD (2005). Microsatellite analysis of olive fly populations in the Mediterranean indicates a westward expansion of the species. *Genetica* **125**: 231–241.
- Augustinos AA, Stratikopoulos EE, Zacharopoulou A, Mathiopoulos KD (2002). Polymorphic microsatellite markers in the olive fly, *Bactrocera oleae*. *Mol Ecol Notes* **2**: 278–280.
- Bonizzoni M, Guglielmino CR, Smallridge CJ, Gomulski M, Malacrida AR, Gasperi G (2004). On the origins of medfly invasion and expansion in Australia. *Mol Ecol* **13**: 3845–3855.
- Bonizzoni M, Zheng L, Guglielmino CR, Haymer DS, Gasperi G, Gomulski LM *et al.* (2001). Microsatellite analysis of medfly bioinfestations in California. *Mol Ecol* **10**: 2515–2524.
- Bruford MW, Wayne RK (1993). Microsatellites and their application to population genetic studies. *Cur Opin Genet Dev* **3**: 939–943.
- California Department of Food and Agriculture (2006). *California Agricultural Resource Directory*. CDFA Office of Public Affairs: Sacramento, CA.
- Cornuet JM, Luikart G (1997). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.
- Davies N, Roderick GK (1999). Determining pathways of marine bioinvasion: genetic and statistical approaches. In: Pederson J (ed). *First Conference of Marine Bioinvasions*. Plenum Press: Boston.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Felsenstein J (1994). *PHYLIP (Phylogeny Inference Package)* version 3.6. Distributed by the Author. Department of Genome Sciences, University of Washington: Seattle.
- Gasperi G, Bonizzoni M, Gomulski LM, Murelli V, Torti C, Malacrida AR *et al.* (2002). Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*. *Genetica* **116**: 125–135.
- Gilchrist AS, Sved JA, Meats A (2004). Genetic relations between outbreaks of the Queensland fruit fly, *Bactrocera tryoni* (Froggatt) (Diptera: Tephritidae), in Adelaide in 2000 and 2002. *Austr J Entomol* **43**: 157–163.
- Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available onhttp://www.unil.ch/izea/software/fstat.html. Updated from Goudet (1995).
- Katsoyannos P (1992). *Olive Pests and their Control in the Near East*. FAO Plant Production and Protection Paper No 115. FAO: Rome.

- Mazomenos BE (1989). *Dacus oleae*. In: Robinson AS, Hooper G (eds). *Word Crop Pests*, Vol. 3B, Elsevier: Amsterdam. pp 169–177.
- Meixner MD, McPherson 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. *Mol Ecol* **11**: 891–899.
- Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005). Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Mol Ecol* **14**: 2729–2738.
- Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283–292.
- Page RDM (1996). TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357–358.
- Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.
- Piry S, Alapetite A, Cornuet J, Paetkau D, Baudouin L, Estoup A (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered* **95**: 536–539.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* **94**: 9197–9201.
- Raymond M, Rousset F (1995). GENEPOP v.1.2: population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
- Roderick GK, Navajas M (2003). Genes in new environments: genetics and evolution in biological control. *Nat Rev Gen* **4**: 889–899.
- Rice RE (2000). Bionomics of the olive fruit fly *Bactrocera* (*Dacus*) *oleae*. *UC Plant Protection Quarterly* **10**: 1–5.
- Rice WR (1989). Analysis tables of statistical tests. *Evolution* **43**: 223–225.
- Schlötterer C, Pemberton J (1994). The use of microsatellites for genetic analysis of natural populations. *EXS* **69**: 203–214.
- Schneider S, Kueffer JM, Roessli D, Excoffier L (1997). ARLEQUIN, version 1.1. A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva: Switzerland.
- Tautz D, Schlötterer C (1994). Simple sequences. *Curr Opin Genet Dev* **4**: 832–837.
- Yeh FC, Boyle T, Rongcai Y, Ye Z, Xiyun JM (1999). POPGENE VERSION 1.32 Microsoft Window-based Freeware for Population Genetic Analysis. <http://www.ualberta.ca/~fyeh/>.
- Yu H, Frommer M, Robson MK, Meats AW, Shearman DCA, Sved JA (2001). Microsatellite analysis of the Queensland fruit fly *Bactrocera tryoni* (Diptera: Tephritidae) indicates spatial structuring: implications for population control. *Bull Entomol Res* **91**: 139–147.
- Zalom FG, Burrack HJ, Bingham R, Price R, Ferguson L (2008). Olive fruit fly (*Bactrocera oleae*) introduction and establishment in California. *Acta Horticulturae* **791**: 619–627.

Appendix 1

Assignment of individuals to five groups, using GeneClass 2.0 software

Samples to be assigned	Reference groups				
	West Med	Central Med	Cypriot	Israeli	Californian
Gim-PO	0.833	0.166	—	—	—
Lisbon-PO	0.793	0.172	—	0.034	—
Mur-SP	0.780	0.120	—	0.040	0.060
Mad-SP	0.827	0.137	—	0.034	0.034
Arr-SP	0.900	0.033	—	0.033	0.033
West Med	0.821	0.125	—	0.029	0.029
Far-IT	0.200	0.700	0.040	0.040	0.020
Vas-IT	0.166	0.600	—	0.166	0.066
Ale-GR	0.160	0.500	0.120	0.060	0.160
Lef-GR	0.125	0.562	0.104	0.125	0.083
Pat-GR	0.120	0.580	0.120	0.060	0.120
Man-GR	0.160	0.580	0.160	0.080	0.020
Ith-GR	0.130	0.695	0.130	0.043	—
Kos-GR	0.095	0.666	0.142	—	0.095
Kyt-GR	0.280	0.480	0.200	—	0.040
Mal-GR	0.120	0.600	0.120	0.080	0.080
Cre-GR	0.119	0.666	0.071	0.023	0.119
Aid-TU	0.111	0.777	0.111	—	—
Central Med	0.149	0.602	0.107	0.064	0.075
Lim-CY	0.066	0.066	0.600	0.133	0.133
Nic-CY	—	—	0.875	0.041	0.083
Cypriot	0.037	0.037	0.722	0.092	0.111
Sde-IS	—	—	0.055	0.944	—
Cal-CA	0.133	0.066	0.100	0.066	0.633
Nap-CA	—	0.033	0.033	0.033	0.900
Sol-CA	0.033	0.066	—	—	0.900
Yol-CA	—	—	0.033	0.033	0.933
SLO-CA	0.033	0.033	0.100	0.033	0.800
Californian	0.040	0.040	0.053	0.033	0.833

Regional group values are indicated in bold.

Appendix 2

Assignment of groups of individuals of California samples to the three described genetic groups plus Israel, according to GeneClass 2.0 software

<i>Samples to be assigned</i>	<i>Reference groups</i>							
	<i>1</i>	<i>%</i>	<i>2</i>	<i>%</i>	<i>3</i>	<i>%</i>	<i>4</i>	<i>%</i>
Cal-CA	Cypriot	99.828	Israeli	0.165	Central Med	0.007	Iberian	0.000
Nap-CA	Cypriot	77.477	Israeli	22.523	Central Med	0.000	Iberian	0.000
Sol-CA	Israeli	100.000	Cypriot	0.000	Central Med	0.000	Iberian	0.000
Yol-CA	Israeli	99.993	Cypriot	0.007	Central Med	0.000	Iberian	0.000
SLO-CA	Israeli	100.000	Cypriot	0.000	Central Med	0.000	Iberian	0.000