# Genetic Aspects of the Worldwide Colonization Process of *Ceratitis capitata*

**A. R. Malacrida, F. Marinoni, C. Torti, L. M. Gomulski, F. Sebastiani, C. Bonvicini, G. Gasperi, and C. R. Guglielmino**

**Multilocus enzyme electrophoresis data from 26 polymorphic loci (124 alleles) were used to analyze the genetic aspects of the worldwide colonization of *Ceratitis capitata* (medfly). Eighty-two samples of 17 populations were collected from six regions throughout the species range: Africa, extra-Mediterranean islands (Madeira and Gran Canaria), Mediterranean region, Latin America (Guatemala), Pacific (Hawaii), and Australia. The variability parameters (*H, P, A*) reveal that the geographical dispersal of medfly from its ancestral source area (East Africa) is associated with a great reduction in variability. The pattern of decreasing variability occurs at two regional levels: in the African–Mediterranean region where the differentiation is gradual, and in the Latin American–Pacific region where some ancestral variability is still present as a consequence of recent colonization. The UPGMA phylogenetic tree, derived from Nei's genetic distances, shows the presence of intraspecific differentiative processes affecting mainly the two island populations, Réunion and Hawaii. The population genetic changes observed in the species range are consistent with both the chronology and the historical circuitous course of the medfly colonization process.**

The Mediterranean fruit fly (medfly) (*Ceratitis capitata*) is a fast-colonizing species which in the last 100 years has spread from its supposed origin in tropical Africa to a number of countries including the Mediterranean basin, parts of South and Central America, and Australia (Fletcher 1989b). In these regions it exists in a wide variety of climates. History of infestation and geographical spread of this species are well documented (for a comprehensive review see Robinson and Hooper 1989). The oldest derived populations are found in the Mediterranean area. More recent colonization events characterize the New World populations. The question of why this species has become a major pest has been approached by studies on zoogeography (Maddison and Bartlett 1989), on population biology (Papadopoulos et al. 1996), and on the analysis of the life-history strategies that this species has evolved (Fletcher 1989a).

A description of medfly population structure is important both for a better understanding of the history and future evolutionary potential of this pest species and its populations as well as for risk assessment. The genetic aspects of the colonization process have been studied through the analysis of protein (Gasperi et al. 1991; Malacrida et al. 1992) and DNA variation (Baruffi et al. 1995; Gasparich et al. 1995; Gomulski et al. 1996; Haymer and McInnis 1994; McPheron et al. 1994). From these previous studies it has been possible to suggest an African origin for this species in the sub-Saharan East region (Kenya) and to evidence the presence of intraspecific differentiative processes of ancestral versus peripheral populations. Very high levels of genetic variability have been detected in the African native populations: this high level of genetic variability may reflect the genetic plasticity of this polyphagous species which, in a very short time, has reached a nearly cosmopolitan geographical distribution (Malacrida et al. 1996). With respect to the colonization pattern, we suggested that medfly populations, like those of *Drosophila melanogaster* (David and Capy 1988), can be divided into three main categories: ancestral, ancient, and new populations from sub-Saharan Africa, the Mediterranean basin, and the New World, respectively (Malacrida et al. 1992). Because historical processes seem to be important determinants of medfly geographic variation, we have undertaken more extensive medfly sampling from throughout the species range to study the genetic consequences of the

From the Dipartimento di Biologia Animale, Università di Pavia, Piazza Botta 9, I-27100, Pavia, Italy (Malacrida, Marinoni, Torti, Gomulski, Sebastiani, Bonvicini, and Gasperi), Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy (Guglielmino), and Istituto di Genetica Biochimica ed Evoluzionistica del CNR, Pavia, Italy (Guglielmino). We are very grateful to Jeffrey R. Powell for critical reading of the manuscript and for his valuable suggestions. We appreciate the help of two anonymous reviewers for their highly valuable comments. We are indebted to the following people for having provided the wild-collected specimens of medfly: T. Mukiama (University of Nairobi, Kenya); E. Osir (ICPE, Nairobi, Kenya); S. Quilici (IFRA, Saint-Pierre, La Réunion); A. Mazih (Institute de Agronomique et Vétérinaire-Hassan II, Agadir, Morocco); P. Ross (CIT-INIA, Madrid, Spain); A. Barbosa (Madeira); A. Economopoulos (IMBB, Heraklion, Greece); F. De Lima (Department of Agriculture, Perth, Western Australia); B. Katsoyannos (University of Thessaloniki, Thessaloniki, Greece); D. O. McInnis (USDA-ARS, Honolulu, Hawaii); P. Rendon (USDA-APHIS, San Miguel, Petapa, Guatemala); Y. Roessler (Citrus Marketing Board, Israel); L. Suess (University of Milan, Milan, Italy); G. Zervas (NCSR, Democritos, Athens, Greece). This research was supported by grants from the European Communities Commission (ECC) (Project FAIR PL96-1972), the National Ministry of the University and Scientific Research and Technology (M.U.R.S.T. "Funds 40%"), and the International Atomic Energy Agency (Vienna, Austria). Address correspondence to Dr. Malacrida at the address above or e-mail: malacrid@ipv36.unipv.it.

© 1998 The American Genetic Association 89:501–507

worldwide colonization process. In this article we examine the results obtained with a multilocus enzyme electrophoresis approach on the distribution of genetic variability in ancestral and derived populations from the Mediterranean basin, Australia, Latin America, and Pacific areas. The results are used to infer (1) a historical interpretation of the population genetic changes found in different geographic populations and (2) to assess the existence of population substructuring in the species range.

## Materials and Methods

### Populations of *C. capitata*

Eighty-two samples of 17 populations of *C. capitata* were collected from five regions throughout the species range: Africa, extra-Mediterranean islands, Mediterranean basin, New World, and Australia.

*African region.* In this region the following three populations were sampled: Kenya, Réunion, and Morocco. The Kenyan population was represented by 12 samples of pupae, collected on coffee berries, over 10 years (August 1984–August 1994) from two farms near Nairobi. The population on Réunion was sampled 14 times (October 1987–December 1994): 10 times near St. Denis and 4 times in other places on the island. Pupae were collected from several different host fruits. The Moroccan population was sampled nine times (February 1992–May 1993) from the *Argania spinosa* forest on the western coast of that country (Mazih and Debouzie 1996).

*Extra-Mediterranean islands.* One sample was collected in Gran Canaria (October 1994) from *Prunus persica* and one in Madeira (October 1995) from *Pirus communis.*

*Mediterranean region.* The following nine populations were sampled: Spain, Sardinia, Procida Island, north Italy, Libya, Israel, Crete, Chios Island, and Greece. The Spanish population was sampled twice, near Valencia (September 1994) and near Tarragona (October 1994). The population from Sardinia was sampled six times (October 1985–October 1990) in the western and eastern parts of central Sardinia. The Procida Island population was sampled seven times (June– August 1985). Three north Italy samples were collected around Milan: in October 1990, September 1992, and November 1995. The Libyan population was sampled once in August 1992. The Israel population was sampled twice near Tel Aviv in May 1991 and July 1993. The Crete population consisted of one sample from Sisses, collected in 1993, and two samples from Fodele collected in 1995. The Chios Island population consisted of three samples collected from August 1992 to September 1993. The Greek flies derive from two samples: one from Salonicco and the other from Attica, collected in March and July 1994, respectively.

*Latin American region.* This region is represented by the population from Guatemala. The Guatemalan flies derive from three samples collected near Antigua, from coffee berries, from December 1989 to March 1994.

*Pacific region.* The Hawaiian flies derive from two samples collected in Mauna Loa and Kauai in April 1992.

*Australian region.* The Australian flies derive from nine samples collected near Perth, from different hosts, from 1992 to 1994.

### Electrophoretic Studies

Preparation of samples and electrophoretic procedures are described in Gasperi et al. (1991). For each population sample at least 25 individuals were assayed at the following 26 enzyme loci: *Mpi, Mdh-1, Mdh-2, Hk-1, Hk-2, Est-1, Est-2, Pgi, Zw, Pgd, Fh, Had, Idh, Pgm, Got-1, Got-2, Ak-1, Ak-2, Adh-2, Gpt, Pgk, Me, α-Gpdh, Aox, Acon- 1, Acon-2.* Staining for enzyme activities, after electrophoresis, was based on the recipes of Meera Khan (1971) and Harris and Hopkinson (1976). For each population we calculated the genotypic absolute frequencies and the weighted mean allele frequency of each locus across the relative samples.

### Data Analysis

Every population was tested for conformance to the Hardy–Weinberg equilibrium at each polymorphic locus using the exact probability test provided by the BIOSYS-1 program (Swofford and Selander 1981). As recommended by Rice (1989), the sequential Bonferroni test was used on all the 156 tests to determine if any cases showed significant departure from Hardy–Weinberg expectations, with $\alpha = 0.05$. The single populations were tested for the degree of genetic variability. The average number of alleles per locus ($A$), percentage of polymorphic loci ($P$), and mean heterozygosity ($H$) were estimated for each population. We also estimated the expected number of alleles per locus: $n_e = 1/(1 - H)$.

The relationships between the populations were represented through a dendrogram obtained from Nei's (1978) unbiased genetic distances. In Nei's method it is assumed that all loci have the same rate of neutral mutation and that the genetic variability is at equilibrium between mutation and genetic drift. Nei's genetic distance is expected, for a sample of many loci, to rise linearly with time. The trees were constructed using the unweighted pair-group arithmetic average (UPGMA; Sneath and Sokal 1973) and neighbor-joining methods with the PHYLIP 3.5c package (Felsenstein 1993). The bootstrap method (Efron 1982) was applied to test the robustness of tree topology.

$F$ statistics were calculated in the medfly sampling area, grouping the 17 above mentioned populations into nine geographical groups. Groups and populations constitute the basis of a hierarchical design for studying the degree of the genetic isolation within and between groups. The standardized genetic variance ($F_{ST}$; Wright 1965) within and between each of the geographical groups was estimated. For each $F_{ST}$ value obtained for each locus, we tested the significance of deviation from zero with the heterogeneity chi-square test (Workman and Niswander 1970), $\chi^2 = 2NF_{ST}(k - 1)$ with $(k - 1)(s)$ degrees of freedom, where $N$ is the total sample size, $k$ is the number of alleles for the locus, and $s$ is the number of subpopulations. The total sample size ($N$) for each couple of groups is calculated on the basis of the average sample size per population.

To estimate the relationships between $F_{ST}$ values and geographical distances, log geographic distances were calculated between all pairs of groups and were regressed against $F_{ST}$ values. To test the significance of the relationships, we used the Mantel test (Mantel 1967) with 10,000 random permutations to correct for lack of independency between data points (GENEPOP, version 3, described by Raymond and Rousset 1995).

Gene flow estimates ($Nm$) were derived from the private allele method of Slatkin (1985) and calculated between the nine geographical groups. The mean frequency of private alleles [$p(1)$], indicates the average level of gene flow among subpopulations according to the formula

$$Nm = \frac{e^{\{-[\ln[p(1)]+2.44]/0.505\}}}{N/25},$$

where $N$ is the average number of individuals sampled per population.

## Results

### Estimates of Genetic Variability in the 17 Geographic Populations

We scored a total of 124 alleles for the 26 polymorphic loci which can be considered

**Table 1. Parameters of genetic variability in 17 populations of *C. capitata* from five geographic regions**

| Population | $N^a$ (±SE) | $A$ (±SE) | $n_e$ | $P$ | $H_O{}^b$ (±SE) | $H_E{}^c$ (±SE) |
|---|---|---|---|---|---|---|
| African area | | | | | | |
| Kenya | 309.2 (15.2) | 3.7 (0.4) | 1.16 | 50.0 | 0.137 (0.036) | 0.161 (0.042) |
| Réunion | 255.8 (7.5) | 2.4 (0.3) | 1.10 | 26.9 | 0.091 (0.031) | 0.103 (0.037) |
| Morocco | 224.2 (6.2) | 2.9 (0.2) | 1.10 | 23.1 | 0.090 (0.031) | 0.097 (0.034) |
| Extra-Mediterranean islands | | | | | | |
| Canary | 28.9 (0.6) | 1.4 (0.1) | 1.09 | 23.1 | 0.079 (0.029) | 0.075 (0.028) |
| Madeira | 28.0 (0.8) | 1.4 (0.2) | 1.09 | 26.9 | 0.083 (0.030) | 0.080 (0.028) |
| Mediterranean area | | | | | | |
| Spain | 55.6 (0.9) | 1.5 (0.1) | 1.07 | 23.1 | 0.068 (0.026) | 0.074 (0.030) |
| Sardinia | 103.6 (5.2) | 1.5 (0.2) | 1.06 | 23.1 | 0.057 (0.021) | 0.059 (0.022) |
| Procida | 99.8 (6.1) | 1.5 (0.1) | 1.06 | 19.2 | 0.057 (0.023) | 0.059 (0.023) |
| North Italy | 44.2 (1.6) | 1.3 (0.1) | 1.04 | 19.2 | 0.035 (0.017) | 0.037 (0.018) |
| Libya | 27.5 (1.5) | 1.2 (0.1) | 1.03 | 15.4 | 0.032 (0.016) | 0.044 (0.024) |
| Israel | 65.9 (2.0) | 1.3 (0.1) | 1.04 | 11.5 | 0.035 (0.019) | 0.040 (0.022) |
| Crete | 73.7 (1.9) | 1.2 (0.1) | 1.03 | 7.7 | 0.027 (0.022) | 0.032 (0.026) |
| Chios | 94.5 (3.9) | 1.3 (0.1) | 1.01 | 7.7 | 0.015 (0.008) | 0.017 (0.009) |
| Greece | 58.6 (0.6) | 1.4 (0.1) | 1.04 | 11.5 | 0.040 (0.020) | 0.044 (0.023) |
| Pacific | | | | | | |
| Hawaii | 51.6 (3.6) | 1.3 (0.1) | 1.5 | 26.9 | 0.052 (0.021) | 0.068 (0.026) |
| Latin American | | | | | | |
| Guatemala | 83.7 (3.2) | 1.3 (0.2) | 1.06 | 11.5 | 0.054 (0.033) | 0.060 (0.035) |
| Australia | 208.8 (4.8) | 1.3 (0.1) | 1.03 | 11.5 | 0.032 (0.018) | 0.034 (0.018) |

[a] Average number of individuals per locus.
[b] Observed heterozygosity based on direct count of heterozygotes.
[c] Unbiased heterozygosity (Nei 1978).

**Table 2. Chronological records of the worldwide colonization of *C. capitata***

| Diffusion area: Geographic region | Country | Date of the earliest record | |
|---|---|---|---|
| Africa | South East Africa[a] | putative source area | Fletcher 1989b |
| | South Africa | 1889 | Back and Pemberton 1918 |
| Extra-Mediterranean islands | Canary[a] | early 1800 | Fimiani 1989 |
| | Madeira[a] | 1829 | ″ |
| Mediterranean area | Spain[a] | 1842 | Fimiani 1989 |
| | Algeria | 1850 | ″ |
| | Tunisia | 1855 | ″ |
| | South Italy[a] | 1863 | ″ |
| | France | 1885 | ″ |
| | Portugal | 1898 | ″ |
| | Israel[a] | end 1800 | ″ |
| | Turkey | 1904 | ″ |
| | Greece[a] | 1915 | ″ |
| | ex-Yugoslavia | 1947 | ″ |
| Latin American | Brazil | 1905 | Enkerlin et al. 1989 |
| | Costa Rica | 1955 | ″ |
| | Nicaragua | 1960 | ″ |
| | Panama | 1963 | ″ |
| | Guatemala[a] | 1975 | ″ |
| | Mexico | 1977 | ″ |
| Pacific | Hawaii[a] | 1910 | Harris 1989 |
| Australia | Australia[a] | 1897 | Hooper and Drew 1989 |

[a] Countries for which population variability data were obtained in this study.

to be a random, although small, sample of the genome. These loci are widely distributed over the genetic map of *C. capitata* (Malacrida et al. 1988). Most of the polymorphic loci in each population were in good agreement with Hardy–Weinberg expectations. Only 5 (*Est-*$_1$ and *Got-*$_2$ in Kenya, *Hk-*$_2$ and *Mpi* in Réunion, *Zw* in Israel) out of 156 exact probability tests were found to show significant statistical deviations after Bonferroni correction.

The amount of allozyme variability within each of the 17 populations, collected in the six geographic regions, are estimated by the parameters $A$, $n_e$, $P$, and $H$ (Table 1). The African ancestral populations (Kenya, Réunion, Morocco) are more polymorphic with respect to the derived populations, that is, those from the Mediterranean basin, Guatemala, Hawaii, and Australia. Among African populations, Kenya has the highest average number of alleles per locus and it is polymorphic for 50% of its loci. Furthermore, the observed number of alleles ($A$) is three times higher than the expected number ($n_e$) on the basis of the observed heterozygosity; this high $A$ estimate is due to the presence of several low frequency endemic alleles. These results, which indicate Kenya as the most polymorphic population, confirm the general trend of decreasing genetic variation from this putative African source area toward the periphery of the species range (Baruffi et al. 1995; Gasperi et al. 1991). We tried to interpret these variability data, estimated in the species' range, on the basis of the history of medfly worldwide colonization, the stages of which are summarized in Table 2. The relationship between the date of the earliest record of each considered geographical population (as presented in Table 2) and its degree of variability, in terms of heterozygosity, are shown in Figure 1. The correlation coefficient between the time of colonization and the heterozygosity is statistically significant ($r = -0.69$, $P < .05$). This result suggests that the steps in the decrease of heterozygosity are related to the timing of the colonization, especially in the Mediterranean basin. The colonization of this area is an old event (Fimiani 1989), and we can suppose that the populations are derived from each other through subsequent bottlenecks. On the other hand, the medfly invasion into Latin America and the Pacific area has been caused by more recent and unrelated events (Harris 1989).

### Genetic Distances and Cluster Analysis

The Nei unbiased genetic distance ($D$) estimates reveal the presence of differentia-

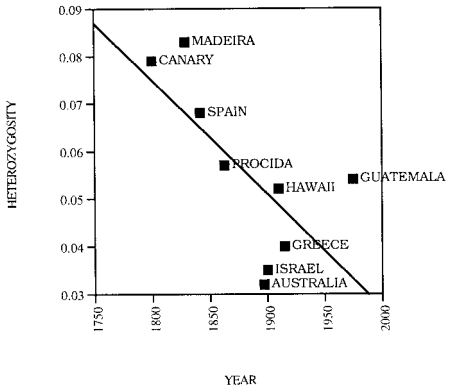


**Figure 1.** Correlation between the date of the earliest record of each considered geographic population (as presented in Table 2) and the average heterozygosity ($H$) at 26 biochemical loci ($r = -0.69$, $P < .05$).

tion processes among the 17 geographical medfly populations. The largest $D$ values occur with all the comparisons involving the populations of Réunion ($D$ = 0.110–0.145) and Hawaii ($D$ = 0.072–0.145). Therefore Réunion and Hawaii are the most differentiated populations. Low $D$ values are observed between Kenya and all the remaining populations, confirming Kenya's ancestral status. All the comparisons involving the other populations show very low $D$ values, indicating a general high genetic similarity among the other derived populations. Among the lowest $D$ values are those between the Australian population and the populations from the Mediterranean area ($D$ = 0.010–0.028).

A UPGMA tree obtained with Nei's genetic distances representing the 17 populations from the six regions is shown in Figure 2. The corresponding consensus tree obtained from 100 bootstrap resamplings of the original dataset shows a similar pattern. In Figure 2 we report, under the corresponding node, only those bootstrap values greater than 50%. Réunion and Hawaii are the most divergent populations; subsequently, Kenya and Guatemala separate from the populations of the Mediterranean area together with Australia and the extra-Mediterranean islands. It is clear from Figure 2 that the differentiation among the Mediterranean populations is minimal and it is characterized by multifurcations; this means that the Mediterranean clusters are composed of populations whose genetic distances are not statistically significant. The affinity of the Australian and the Mediterranean populations is evident: Australia is included within the Mediterranean cluster in 51% of the bootstrapped replications. A high consensus value (75%) supports the association between Morocco and Spain. The neighbor-joining tree obtained with the same dataset shows the same differentiative processes among populations.

### Hierarchical *F* Statistics Analysis

To determine the effect of spatial scale on gene frequencies we conducted a hierarchical analysis of population differentiation in the species range. For this purpose we partitioned the 17 populations into nine geographical groups (Table 3) representing (1) the ancestral Kenya populations, (2) the African-related population from Réunion, (3–6) four groups of old adventive populations from the Mediterranean area, (7) the Australian group, and (8 and 9) the two recently derived populations from Latin America and the Pacific—Guatemala and Hawaii. $F_{ST}$ estimates, calculated as the weighted average over all loci within and between the nine geographical groups, are shown in Table 3. Almost all $F_{ST}$ values differ significantly from zero according to the heterogeneity chi-square test, suggesting a high degree of heterogeneity in the species range.

The important outcome of Table 3 is that very low $F_{ST}$ values are observed between the source area (south East Africa) and all the other derived groups, excluding Réunion. The genetic heterogeneity of the two island populations, Hawaii and Réunion, with respect to the other geographical groups is evident; however, Hawaii doesn't appear differentiated from the ancestral population of Kenya ($F_{ST}$ = 0.066). Australia shows low levels of genetic diversity with respect to each of the two Mediterranean groups. The within-group estimates suggest that the eastern Mediterranean group is the most heterogeneous.

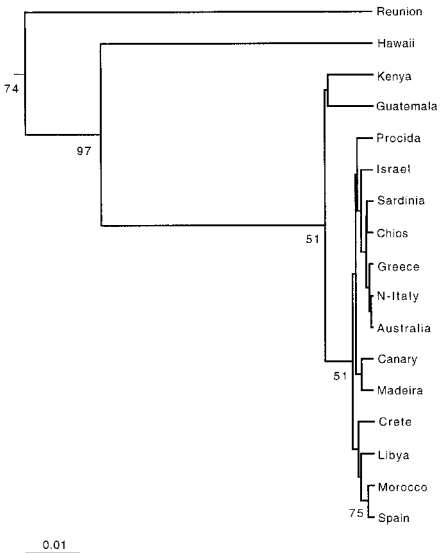


**Figure 2.** Tree derived from Nei's unbiased genetic distances representing the 17 populations of *C. capitata*. Under the corresponding node are reported only the bootstrap values greater than 50%, obtained from 100 bootstrap resamplings of the original dataset.

To test for isolation by distance in the species range, we regressed the $F_{ST}$ estimates with log of the geographic distance between the pairwise combinations of all the nine geographic groups of populations. This regression analysis showed that there was no significant relation between the $F_{ST}$ estimates and the geographical separation ($F_{ST} = -288 + 0.108$ log distance, $r^2 = 0.051$, Mantel $P = 0.149$). However, when we removed from this analysis the two recently derived populations from Latin America and the Pacific (Guatemala and Hawaii) and the Australian group, there was a clear pattern of isolation by distance (Figure 3) in the pairwise comparisons involving the African populations and the four groups of the old adventive populations from the Mediterranean area ($F_{ST} = -1.198 + 0.364$ log distance, $r^2 = 0.38$, Mantel $P = 0.033$). We can conclude that in the African–Mediterranean area, the genetic differentiation between populations increases with geographic distance.

### Gene Flow

From the above analysis it seems clear that the history and the time of colonization must have influenced the contemporary genetic structure of medfly populations. Taking into account the important distinction between contemporary and historical gene flow (Slatkin 1987), we used allozyme data to derive indirect estimates of migration using Slatkin's private allele approach. Table 4 reports the partition of the number of private alleles and their average frequencies detected in each pairwise comparison between the nine geographic groups of populations. In each comparison, low values of the total average frequency of private alleles correspond to high $Nm$ estimate (Table 5). High $Nm$ values are observed in all the comparisons involving the ancestral Kenya population: a maximal $Nm$ estimate of 56.9 has been ascertained in the pairwise comparison with the Iberian–African group, which in turn shows high levels of gene flow with the western and eastern Mediterranean areas ($Nm$ = 132.0 and 139.0, respectively) and Australia ($Nm$ = 42.8).

Genetic isolation, in terms of gene flow, has been ascertained for the two island populations—Réunion and Hawaii—which maintain gene flow levels greater than 1

**Table 3.** $F_{ST}$ estimates and $\chi^2$ analysis within[a] and between[b] the nine geographical groups of *C. capitata* populations

| Groups | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 South East Africa (Kenya) | — | 0.261<br>5917.8<br>(168)* | 0.014<br>500.1<br>(146)* | 0.023<br>1861.1<br>(178)* | 0.070<br>7619.8<br>(150)* | 0.063<br>1986.6<br>(152)* | 0.021<br>812.5<br>(148)* | 0.066<br>1411.5<br>(146)* | 0.067<br>3305.9<br>(150)* |
| 2 Réunion | | — | 0.182<br>3033.2<br>(84)* | 0.341<br>7418.1<br>(138)* | 0.429<br>7215.7<br>(88)* | 0.431<br>9845.0<br>(90)* | 0.248<br>2341.9<br>(78)* | 0.182<br>3220.3<br>(86)* | 0.442<br>6809.2<br>(82)* |
| 3 Extra-Mediterranean islands (Canary, Madeira) | | | 0.026<br>36.6<br>(24) | 0.006<br>135.4<br>(104) | −0.008<br>— | −0.014<br>— | 0.102<br>349.0<br>(34)* | 0.242<br>889.9<br>(34)* | 0.007<br>189.7<br>(32)* |
| 4 Iberian–African group (Morocco, Spain) | | | | −0.004<br>— | 0.023<br>564.32<br>(112)* | 0.017<br>486.0<br>(110)* | 0.028<br>394.0<br>(108)* | 0.133<br>1847.2<br>(108)* | 0.032<br>586.3<br>(106)* |
| 5 West Mediterranean group (Sardinia, Procida, North Italy) | | | | | 0.009<br>71.3<br>(57) | −0.010<br>— | 0.056<br>518.1<br>(40)* | 0.209<br>2067.5<br>(46)* | 0.006<br>108.4<br>(42)* |
| 6 East Mediterranean group (Libya, Israel, Crete, Chios, Greece) | | | | | | 0.119<br>184.4<br>(90)* | 0.055<br>334.4<br>(40)* | 0.303<br>1525.5<br>(44)* | −0.023<br>— |
| 7 Guatemala | | | | | | | — | 0.302<br>821.5<br>(26)* | 0.095<br>479.1<br>(26)* |
| 8 Hawaii | | | | | | | | — | 0.316<br>1886.3<br>(28)* |
| 9 Australia | | | | | | | | | — |

[a] Only if the group is composed of at least two populations. For groups 3 and 4, composed of only two populations, single samples of each population are at the low level of the hierarchy.

[b] When each of the compared groups is composed of only one population, single samples are considered at the low level of the hierarchy.

Chi-square estimates have not been computed for negative $F_{ST}$ values, which result from negative variance components in the hierarchical analysis.

The degrees of freedom are in parentheses. The asterisk indicates tablewide Bonferroni statistical significance at $\alpha = 0.05$ (Rice 1989).

with the ancestral Kenyan population. Réunion has a high number of private alleles, some of which are fixed (Baruffi et al. 1995), while Hawaii contains few private alleles, but at very high frequencies.

## Discussion

### Genetic Consequences of the Colonization Process

The analysis of variability parameters, such as *H, P,* $n_e$, and *A,* clearly reveals the dynamic aspects of the population genetic changes associated with the worldwide spread of the medfly. The East African area, represented here by the Kenyan population, is confirmed as the ancestral source area of the species (Baruffi et al. 1995; Gasperi et al. 1991). This native population is characterized by a high number of alleles per locus, most of which are at low frequency and found only in this population: we consider these alleles as "ancestral" ones. The geographical dispersal of medfly from the source area is associated with a gradual and great reduction of variability, which is indicative of a colonization process probably through subsequent bottlenecks. The loss in variability is quantifiable: from the ancestral level of $H = 0.137$ in Kenya, 34% is lost within the African area ($H = 0.090$ in Morocco and 0.091 in Réunion), 41% in the extra-Mediterranean islands ($H = 0.081$), 70% within the Mediterranean area ($H = 0.041$), and 77% in Australia ($H = 0.032$). A loss of variability of 61% is observed within the populations from Hawaii and Guatemala ($H = 0.053$).

We observe that the loss of variability in the derived populations occurs at two regional levels. The first is at the African–Mediterranean level: the decrease is gradual from Kenya to Greece and it is tightly correlated with the timing of the medfly expansion. In this range the genetic differentiation is also positively related with the geographic distance, with a clear pattern of isolation by distance. The second regional level involves the recent expansion into Latin America and the Pacific, where some ancestral variability is still present in the derived populations, since probably there has not been sufficient time for the two populations to reach a proper equilibrium variability.

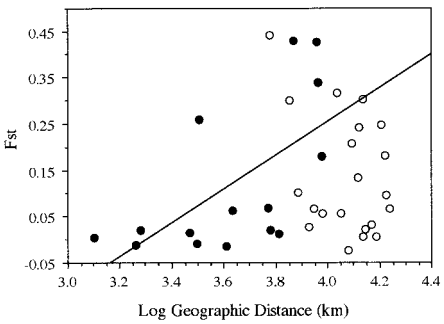


**Figure 3.** Plot of $F_{ST}$ against log geographic distance for all pairs of nine population groups. White circles indicate comparisons including New World and Australian populations; black circles indicate comparisons not including New World and Australian populations. The regression line is shown only for the latter comparisons ($F_{ST} = -1.198 + 0.364$ log distance; $r^2 = 0.38$).

The phylogenetic tree portrays well all the above described aspects of the medfly colonization process. In addition it reflects the presence of a great amount of differentiation affecting the two island populations, Réunion and Hawaii. The old population from Réunion and the new one from Hawaii may represent two examples of incipient geographic differentiation due to island isolation. All the other derived populations still maintain genetic affinity with the ancestral Kenyan population; among them, the Guatemalan population is the closest. The Mediterranean populations form a compact cluster with a very short range of genetic distances. However, incipient genetic substructuring within this area is supported by the hierarchical analysis data showing the eastern Mediterranean group as the most heterogeneous.

### Agreement of the Genetic Data With the Proposed Worldwide Colonization Pattern

From its supposed origin in southern East Africa (Hagen et al. 1981), the medfly

**Table 4. Partition of the number of private alleles (npa) and their average frequencies, *p*(1), in parentheses, in each pair group comparison**

| Pair of groups | | npa (*p*(1)) | | | | | |
|---|---|---|---|---|---|---|---|
| a | b | Group a | | Group b | | Total | |
| SE Africa | Réunion | 47 | (0.049) | 10 | (0.204) | 57 | (0.076) |
| SE Africa | Extra-Medit. islands | 60 | (0.021) | — | | 60 | (0.021) |
| SE Africa | Iberian-African group | 37 | (0.012) | 16 | (0.008) | 53 | (0.011) |
| SE Africa | West Medit. group | 57 | (0.019) | 2 | (0.018) | 59 | (0.019) |
| SE Africa | East Medit. group | 58 | (0.019) | 3 | (0.008) | 61 | (0.019) |
| SE Africa | Guatemala | 68 | (0.031) | 2 | (0.100) | 70 | (0.033) |
| SE Africa | Hawaii | 66 | (0.027) | 2 | (0.199) | 68 | (0.032) |
| SE Africa | Australia | 65 | (0.023) | 1 | (0.026) | 66 | (0.023) |
| Réunion | Extra-Medit. islands | 28 | (0.133) | 5 | (0.402) | 33 | (0.173) |
| Réunion | Iberian-African group | 17 | (0.162) | 33 | (0.063) | 50 | (0.097) |
| Réunion | West Medit. group | 25 | (0.137) | 7 | (0.311) | 32 | (0.175) |
| Réunion | East Medit. group | 28 | (0.103) | 10 | (0.201) | 38 | (0.129) |
| Réunion | Guatemala | 32 | (0.119) | 3 | (0.731) | 35 | (0.171) |
| Réunion | Hawaii | 33 | (0.109) | 6 | (0.400) | 39 | (0.154) |
| Réunion | Australia | 31 | (0.105) | 4 | (0.500) | 35 | (0.150) |
| Extra-Mediterranean islands | | — | | — | | 5[a] | (0.025)[a] |
| Extra-Medit. isl. | Iberian-African group | — | | 39 | (0.009) | 39 | (0.009) |
| Extra-Medit. isl. | West Medit. group | 5 | (0.039) | 10 | (0.026) | 15 | (0.030) |
| Extra-Medit. isl. | East Medit. group | 5 | (0.039) | 10 | (0.004) | 15 | (0.015) |
| Extra-Medit. isl. | Guatemala | 10 | (0.160) | 4 | (0.154) | 14 | (0.158) |
| Extra-Medit. isl. | Hawaii | 8 | (0.224) | 4 | (0.157) | 12 | (0.202) |
| Extra-Medit. isl. | Australia | 7 | (0.046) | 3 | (0.012) | 10 | (0.036) |
| Iberian-African group | | — | | — | | 43[a] | (0.009)[a] |
| Iber.-Afr. group | West Medit. group | 39 | (0.004) | 5 | (0.041) | 44 | (0.008) |
| Iber.-Afr. group | East Medit. group | 37 | (0.007) | 3 | (0.002) | 40 | (0.007) |
| Iber.-Afr. group | Guatemala | 47 | (0.025) | 2 | (0.121) | 49 | (0.028) |
| Iber.-Afr. group | Hawaii | 45 | (0.036) | 2 | (0.222) | 47 | (0.044) |
| Iber.-Afr. group | Australia | 44 | (0.014) | 1 | (0.002) | 45 | (0.014) |
| West Mediterranean group | | — | | — | | 8[a] | (0.055)[a] |
| West Medit. group | East Medit. group | 8 | (0.030) | 8 | (0.002) | 16 | (0.016) |
| West Medit. group | Guatemala | 12 | (0.102) | 1 | (0.192) | 13 | (0.109) |
| West Medit. group | Hawaii | 13 | (0.146) | 4 | (0.121) | 17 | (0.140) |
| West Medit. group | Australia | 11 | (0.039) | 2 | (0.005) | 13 | (0.034) |
| East Mediterranean group | | — | | — | | 12[a] | (0.023)[a] |
| East Medit. group | Guatemala | 13 | (0.057) | 2 | (0.121) | 15 | (0.065) |
| East Medit. group | Hawaii | 13 | (0.131) | 4 | (0.139) | 17 | (0.133) |
| East Medit. group | Australia | 10 | (0.013) | 1 | (0.002) | 11 | (0.012) |
| Guatemala | Hawaii | 4 | (0.331) | 6 | (0.468) | 10 | (0.413) |
| Guatemala | Australia | 4 | (0.200) | 6 | (0.141) | 10 | (0.165) |
| Hawaii | Australia | 5 | (0.134) | 5 | (0.331) | 10 | (0.233) |

[a] npa and *p*(1) estimates within each geographical group.

seems to have traveled with man to the Mediterranean coast (Maddison and Bartlett 1989). The fly may have established itself progressively from Africa to the Mediterranean regions of Spain (De Breme 1842) and then to other coastal, northern and eastern Mediterranean locations (Fimiani 1989). It was introduced into Australia from Europe around 1897 as a secondary colonization event (Hooper and Drew 1989). In the Americas it was first reported in central America in Costa Rica in 1955 (Gallo et al. 1970), spreading to other countries, being detected in 1976 in Guatemala, and moving to Mexico along the coffee belt (Harris 1989).

The relatively recent historical association of the medfly populations explains the observed genetic similarities between the African ancestral and the derived populations. It is not surprising that the Mediterranean populations exhibit a modest genetic differentiation with respect to the others, in view of the presumed longer history of occupation of the Mediterranean area contrasted with the more recent colonization of the Pacific and Latin American regions. In the plot of Figure 2, genetic distances between African, European, Latin American, and Pacific populations are the results of a sequence of historical events during the colonization.

Also the Slatkin's *Nm* estimates reflect the common ancestry of populations in the recent past, since the majority of the considered private alleles are "ancestral" African alleles. On this basis, the high *Nm* estimates detected at a macrogeographical level between Africa and the Mediterranean basin are consistent with both the chronology and with the above mentioned historical circuitous course of medfly spreading (Figure 4). The *Nm* estimates between south East Africa to the Iberian–African area and from this region to the western and eastern parts of the Mediterranean area are consistent with the proposed route of colonization from the source area to northwest Africa and from there to Spain. The Iberian–African area seems to have played the major role in the entry of medfly into the Mediterranean basin. Within this area our data are consistent with a colonization pattern from the west to the east. Secondary colonization events such as those from the Mediterranean basin to Australia (Fimiani 1989) explain our high *Nm* estimates between these groups. However, Slatkin gene flow estimates depend on relatively few rare alleles and can be strongly influenced by ongoing gene flow deriving from trading activities. This seems to be the case for gene flow estimates between Guatemala and the Iberian–African region: the *Nm* value (8.6) appears slightly greater compared to the *Nm* estimate between Guatemala and Kenya (6.2), in spite of the greater genetic affinity between these two last populations. In the same way, the high *Nm* estimate (87.0) between the Iberian–African group and the extra-Mediterranean islands

**Table 5. Gene flow estimates within and between the nine geographical groups of *C. capitata* populations according to Slatkin's method**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 South East Africa (Kenya) | — | 1.5 | 15.2 | 56.9 | 24.6 | 19.4 | 6.2 | 7.2 | 15.7 |
| 2 Réunion | | — | 0.3 | 0.9 | 0.4 | 0.5 | 0.3 | 0.4 | 0.5 |
| 3 Extra-Mediterranean islands (Canary, Madeira) | | | 10.2 | 87.0 | 9.3 | 26.9 | 0.3 | 0.2 | 6.0 |
| 4 Iberian–African group (Morocco, Spain) | | | | 94.4 | 132.0 | 139.0 | 8.6 | 3.8 | 42.8 |
| 5 West Mediterranean group (Sardinia, Procida, North Italy) | | | | | 4.1 | 31.8 | 0.7 | 0.5 | 9.3 |
| 6 East Mediterranean group (Libya, Israel, Crete, Chios, Greece) | | | | | | 11.6 | 1.5 | 0.4 | 54.6 |
| 7 Guatemala | | | | | | | — | 0.04 | 0.3 |
| 8 Hawaii | | | | | | | | — | 0.2 |
| 9 Australia | | | | | | | | | — |

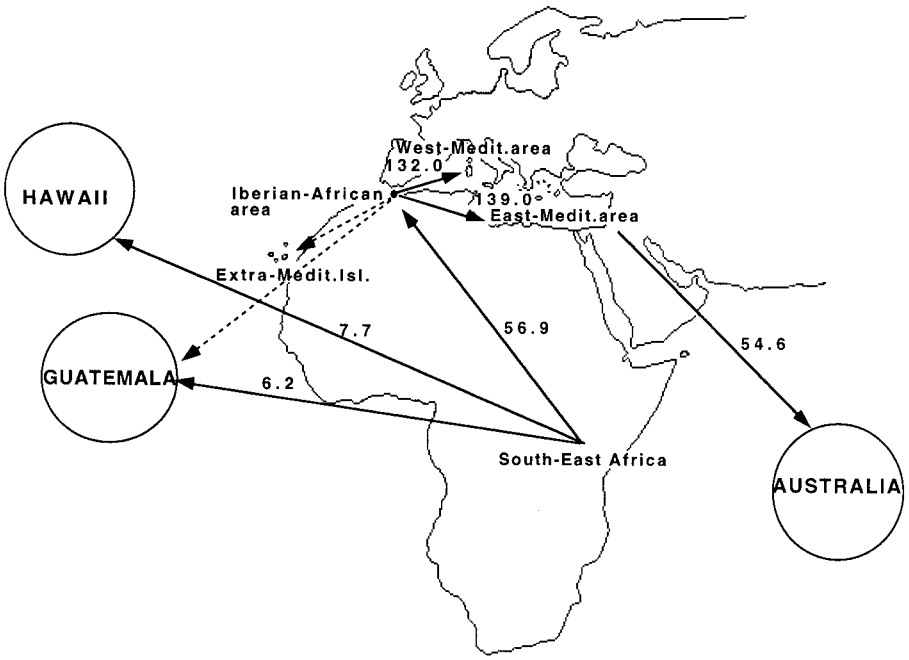


**Figure 4.** Slatkin's *Nm* estimates along the probable route of *C. capitata* colonization process from the south East Africa source area to the Mediterranean basin, to Australia, and to Latin America and Pacific areas. The dashes arrows indicate the probable gene flow influenced by ongoing trading activities.

could be influenced by trading activities subsequent to the establishment of the medfly in these islands.

In conclusion, from these genetic data it appears that medfly could be considered as a case in which single ancestral populations in eastern and southern Africa give rise to several populations in the recent past. The isolation by distance evidenced in the African–Mediterranean area is compatible with a hierarchical migration structure. On the other hand, the structure of Latin American and Pacific populations can be the product of few geographically separated colonization events followed by subsequent expansions. These two colonization processes, one ancient and one new, are correlated with human mobility capacities at different times and are therefore connected with the history of human trading activities.

**References**


Back EA and Pemberton CE, 1918. The Mediterranean fruit fly. USDA Bulletin 640. Washington, DC; U.S. Department of Agriculture; 1–44.

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

David JR and Capy P, 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trend Genet 4:106–111.

De Breme F, 1842. Note sur le genre *Ceratitis* de M. Mac heay (Diptera). Ann Soc Entomol France 11:183–190.

Efron B, 1982. The jackknife, the bootstrap, and other resampling plans. Conf Board Math Sci Soc Ind Appl Math 38:1–92.

Enkerlin D, Garcia LR, Fidel Lopez M, 1989. Mexico, Central and South America. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3A. Amsterdam: Elsevier Science; 83–90.

Felsenstein J, 1993. PHYLIP (phylogeny inference package) version 3.5c. Seattle: University of Washington.

Fimiani P, 1989. Mediterranean region. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3A. Amsterdam: Elsevier Science; 39–50.

Fletcher BS, 1989a. Life history strategies of tephritid fruit flies. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3B. Amsterdam: Elsevier Science; 195– 208.

Fletcher BS, 1989b. Movements of tephritid fruit flies. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3B. Amsterdam: Elsevier Science; 209–219.

Gallo DNO, Wiendel FM, Silveira Neto S, and Ricardo PLC, 1970. Manual de entomologia. Sao Paulo: Agronomica Ceres.

Gasparich GE, Sheppard WS, Han HY, McPheron BA, and Steck GJ, 1995. Analysis of mitochondrial DNA and development of PCR-based diagnostic molecular marker for Mediterranean fruit fly (*Ceratitis capitata*) populations. Insect Mol Biol 4:61–67.

Gasperi G, Guglielmino CR, Malacrida AR, and Milani R, 1991. Genetic variability and gene flow in geographical populations of (medfly) *Ceratitis capitata* (Wied). Heredity 64:347–356.

Gomulski L, Morandi PA, Brogna S, Bourtzis K, Savakis C, and Gasperi G, 1996. Intron size variation traces the world-wide colonization history of the medfly, *Ceratitis capitata.* In: Proceedings of the XX International Congress of Entomology, Florence, Italy, August 25–31, 1996. 229.

Hagen KS, William WW, and Tassan RL, 1981. Mediterranean fruit fly: the worst may be yet to come. Calif Agric 35:5–7.

Harris EJ, 1989. Hawaiian islands and North America. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3A. Amsterdam: Elsevier Science; 73–81.

Harris H and Hopkinson DA, 1976. Handbook of enzyme electrophoresis in human genetics. Amsterdam: North-Holland.

Haymer DS and McInnis DO, 1994. Resolution of populations of the Mediterranean fruit fly at the DNA level using random primers for the polymerase chain reaction. Genome 37:244–248.

Hooper GHS and Drew RAI, 1989. Australia and South Pacific islands. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3A. Amsterdam: Elsevier Science; 67–72.

Maddison PA and Bartlett BJ, 1989. Contribution towards the zoogeography of the Tephritidae. In: Fruit flies: their biology, natural enemies and control (Robinson AS and Hooper G, eds). Vol. 3A. Amsterdam: Elsevier Science; 27–35.

Malacrida AR, Gasperi G, Baruffi L, Biscaldi GF and Milani R, 1988. Updating of the genetics of *Ceratitis capitata* (Wied.). In: Modern insect control: nuclear techniques and biotechnology. Vienna, Austria: International Atomic Energy Agency; 221–227.

Malacrida AR, Guglielmino CR, D'Adamo P, Torti C, Marinoni F, and Gasperi G, 1996. Allozyme divergence and phylogenetic relationships among species of tephritid flies. Heredity 76:592- -602.

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

Mantel N, 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220.

Mazih A and Debouzie D, 1996. Infestation rate of argan fruit (*Argania spinosa*) by the Mediterranean fruit fly (*Ceratitis capitata*) in relation to phenology and maturation of the fruit. Entomol Exp Appl 81:31–38.

McPheron BA, Gasparich GE, Han HY, Steck GJ, and Sheppard WS, 1994. Mitochondrial DNA restriction map for the Mediterranean fruit fly, *Ceratitis capitata.* Biochem Genet 32:25–33.

Meera Khan P, 1971. Enzyme electrophoresis on cellulose acetate gel: zymogram patterns in man- mouse and man-Chinese hamster somatic cell hybrids. Arch Biochem Biophys 145:470–483.

Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Papadopoulos NT, Carey JR, Katsoyannos BI, and Kouloussis NA, 1996. Overwintering of the Mediterranean fruit fly (Diptera: Tephritidae) in northern Greece. Ann Entomol Soc Am 89:526–534.

Raymond M and Ruosset F, 1995. GENEPOP: population genetics software for exact texts and ecumenicism. J Hered 86:248–249.

Rice WR, 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Robinson AS and Hooper G (eds), 1989. Fruit flies: their biology, natural enemies and control. Vols. 3A and 3B. Amsterdam: Elsevier Science.

Slatkin M, 1985. Rare alleles as indicators of gene flow. Evolution 39:53–65.

Slatkin M, 1987. Gene flow and the geographic structure of natural populations. Science 236:787– 792.

Sneath PHA and Sokal RR, 1973. Numerical taxonomy. San Francisco: W. H. Freeman.

Swofford DL and Selander RB, 1981. BIOSYS-1. A computer program for the analysis of allelic variation in genetics. Urbana, Illinois: University of Illinois, Urbana.

Workman PL and Niswander JD, 1970. Population studies in southwestern Indian tribes. II. Local genetic differentiation in the Papago. Am J Hum Gen 22:24–49.

Wright S, 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:355–420.



Received June 17, 1997
Accepted February 13, 1998

Corresponding Editor: Ross MacIntyre