# Population genetics and migration pathways of the Mediterranean fruit fly *Ceratitis capitata* inferred with coalescent methods

Maria Belen Arias[1,2], Samia Elfekih[3] and Alfried P. Vogler[1,2]

[1] Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, United Kingdom
[2] Department of Life Sciences, Natural History Museum, London, United Kingdom
[3] CSIRO Health & Biosecurity, Black Mountain, Canberra, Australia

## ABSTRACT

**Background**. Invasive species are a growing threat to food biosecurity and cause significant economic losses in agricultural systems. Despite their damaging effect, they are attractive models for the study of evolution and adaptation in newly colonised environments. The Mediterranean fruit fly, *Ceratitis capitata*, as a member of the family Tephritidae, is one of the most studied invasive species feeding on many fruit crops in the tropics and subtropics worldwide. This study aims to determine the global macrogeographic population structure of *Ceratitis capitata* and reconstruct its potential migration routes.

**Method**. A partial mitochondrial cytochrome oxidase I gene from >400 individual medflies and 14 populations from four continents was sequenced and subjected to Bayesian demographic modelling.

**Results**. The Afrotropical populations (Kenya, South Africa and Ghana) harbour the majority of haplotypes detected, which also are highly divergent, in accordance with the presumed ancestral range of medflies in Sub-Saharan Africa. All other populations in the presumed non-native areas were dominated by a single haplotype also present in South Africa, in addition to a few, closely related haplotypes unique to a single local population or regional set, but missing from Africa. Bayesian coalescence methods revealed recent migration pathways from Africa to all continents, in addition to limited bidirectional migration among many local and intercontinental routes.

**Conclusion**. The detailed investigation of the recent migration history highlights the interconnectedness of affected crop production regions worldwide and pinpoints the routes and potential source areas requiring more specific quarantine measures.

**Subjects** Biogeography, Ecology, Genetics, Population Biology
**Keywords** Medfly, Migration, Population genetics, Bayesian method, Macrogeographic patterns

# INTRODUCTION

Globalisation and international economic trade have increased the transportation of species outside their natural ranges. Thus, human activities assist the spread of exotic species and increase the rates at which pest species invade new areas (*Blackburn et al., 2011*; *Karsten et al., 2013*). The arrival of invasive species is frequently associated with biodiversity losses,

changes in ecosystem function, and negative impacts on economy, agriculture and human health (*Mack et al., 2000*). While they have many adverse effects, invasive species also offer unique opportunities to study evolution and adaptation within entirely different environments compared to their ancestral habitats (*Diamantidis, Carey & Papadopoulos, 2008*; *Vogel et al., 2010*).

The Mediterranean fruit fly, *Ceratitis capitata,* also known as medfly, poses a severe economic threat to agriculture, especially fruit production, due to its broad range of more than 260 different host plant species and worldwide distribution (*Malacrida et al., 2007*). Chronological records and global studies based on genetic markers assume that *C. capitata* populations are subdivided into three different groups: an ancestral population in Sub-Saharan Africa, a younger population in the Mediterranean basin, and various recently derived populations in tropical and subtropical America, Australia and Oceania (*Gasperi et al., 1991*; *Malacrida et al., 2007*; *Malacrida et al., 1992*; *Szyniszewska & Tatem, 2014*).

Current medfly management approaches vary between countries, although insecticides (baits and full cover-sprays) are the predominant methods used. However, due to harmful effects of insecticides, the Sterile Insect Technique (SIT) using the release of males subjected to sublethal X-ray irradiation is becoming increasingly common (*Dyck, Hendrichs & Robinson, 2005*). The successful implementation of pest control strategies using SIT relies on information about possible movements and effective population sizes in the regions under management.

Population genetic studies can be used to understand medfly invasion biology by focusing on the degree of subdivision within and among local regions. Additionally, information on demography and dispersal can be inferred from genetic data. However, the interpretation of these data in the medfly is challenging as repeated range expansions and invasions, as well as several cases of regional eradication, have impacted its distribution and genetic diversity. The medfly population genetic structure and invasion routes have been previously studied using various molecular approaches at local (*Bonizzoni et al., 2004*; *Bonizzoni et al., 2001*; *Elfekih, Makni & Haymer, 2010*; *Karsten et al., 2013*) as well as global level (*Bonizzoni et al., 2000*; *Gasperi et al., 2002*; *Malacrida et al., 2007*). Most of the proposed colonisation routes have been calculated based on traditional methods such as genetic distance or the private-allele method of Slatkin's (*Bonizzoni et al., 2000*; *Gasperi et al., 2002*; *Malacrida et al., 1998*). Despite these past efforts, the implementation of coalescence methods to investigate medfly invasion have been limited to one study revealing the origin of medfly in Australia (*Bonizzoni et al., 2004*), and to a another study by *Karsten et al. (2015)*, which used the Approximate Bayesian Computation (ABC) method to show a decrease in genetic diversity outside of Africa, the presumed origin of the introduced range described above. Even though this study provides invaluable genetic information for the medfly colonisation, it provides detailed information on African populations only, therefore, the incorporation of new populations especially in the Palearctic and Neotropical regions is required to improve the current knowledge of medfly dispersal.

In this study, a large-scale phylogeographic analysis was conducted using the cytochrome oxidase gene I (COI) for a pathway analysis of medfly populations across their distribution range. We aim to determine the current macrogeographic population structure of *C. capitata* collected from different populations around the globe, and to reconstruct plausible migration routes using Bayesian coalescence approaches.

## MATERIALS AND METHODS

### Sample collection

Specimens of *C. capitata* were collected from 11 sites across all biogeographic regions where the species occurs (Afrotropical, Palaearctic, Australasian and Neotropical) (Table 1) between 2009 and 2014. Whole specimens were collected via traps in orchards and reared from infested fruit, and all flies were preserved in 80% ethanol at −20 °C until tissue was used for DNA extraction.

Sequences representing populations at further sites (Kenya, Ghana, and Iran) were obtained from GenBank, which was used to increase the sample size at the main collection sites (see Table S1 for details). Below, specimens from the same sampling site will be referred to as a population.

### DNA extraction, sequencing, and alignment

After morphological identification of the collected specimens, genomic DNA was extracted from each specimen using DNeasy Blood & Tissue Spin Column Kit (Qiagen, Valencia, CA, USA). A fragment of the mitochondrial gene cytochrome *c* oxidase subunit I (COI) was amplified using the primers LCO1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′) (*Folmer et al., 1994*). PCRs were conducted in a 20 μl reaction volume, with 0.5 μl of genomic DNA, 0.1 mM dNTPS, 0.5 U/μM BIOTAQ DNA Polymerase, 3 mM MgCl2, 0.3 μM of forward and reverse primers. The PCR program included an initial denaturation step of 94 °C for 5 min followed by 35 cycles of 94 °C, 30 s; annealing at 51 °C for 54 s, 72 °C for 54 s and the final extension at 72 °C for 7 min. PCR products were sequenced bidirectionally using ABI technology. Sequences were aligned in Geneious software v.7.1.7 (*Kearse et al., 2012*) together with sequences retrieved from Genbank.

### Genetic diversity and population structure

Levels of genetic diversity were determined estimating the following parameters in DNAsp v.5 and ARLEQUIN v.3.5.2.1 (*Excoffier & Lischer, 2010*; *Librado & Rozas, 2009*): the number of haplotypes ($k$), number of segregating sites (S), haplotype diversity ($h$) and nucleotide diversity ($\pi$). The median-joining (MJ) network (*Bandelt, Forster & Röhl, 1999*) was used to estimate the genealogical relationships in *C. capitata* haplotypes computed in POPART v.1.7 (*Leigh & Bryant, 2015*). Population genetic structure was estimated by population pairwise $F_{st}$. The significant test statistic was performed using 1,000 permutations, and it was computed in ARLEQUIN v.3.5.2.1.

**Table 1  Collection and sample size for *Ceratitis capitata* included in this study.**  Details of sample sites location, name of host plant where the individuals were collected and number of individuals collected at each sample site.

| Biogeographic Region | Country | Location | Host | Sample size |
|---|---|---|---|---|
| Afrotropical | South Africa | Stellenbosch | Guava | 37 |
| Palearctic | Egypt | – | – | 25 |
| | Israel | Gedera | Fig | 19 |
| | | Lachish | Orange | 19 |
| | | Ness Ziona | Orange | 19 |
| | | Neta'im | Guava | 19 |
| | | Shefayim | Orange | 18 |
| | | Yad Mordechai | Lemon | 15 |
| | Tunisia | Bizerte | Orange | 1 |
| | Greece | Thessaloniki | Apple | 20 |
| | | Aetolia-Acarnania | Orange | 3 |
| | Spain | Valencia | Fig | 21 |
| | | Malaga | Peach | 15 |
| | | Murcia | Orange | 5 |
| Australasian | Australia | Perth | Orange | 24 |
| Neotropical | Guatemala | Santa Barbara | Coffee | 51 |
| | Colombia | Cundinamarca | Peach | 18 |
| | | Nariño | Coffee | 4 |
| | Brazil | Salvador | Guava | 11 |
| | Peru | Ica | Orange | 8 |

## Demographic inferences

The Tajima's *D* (*Tajima, 1989*) and Fu's Fs (*Fu, 1997*) statistic tests were performed to identify deviations from neutral models in ARLEQUIN v.3.5.2.1. The past population dynamic through time for the various *C. capitata* haplogroups was inferred using a Bayesian skyline plot method (BSP). Two independent simulations were run using the Hasegawa-Kishino-Yano (HKY) substitution model and uncorrelated lognormal relaxed molecular clock. Each independent run was performed for $5 \times 10^7$ Markov chain Monte Carlo (MCMC) iterations (sampled every 1000 iterations) and discarding 10% of the trees as burn-in implemented in BEAST v.2.4 (*Drummond & Rambaut, 2007*).

In addition to the considerable variation in mutation and substitution rates between genes and taxa, there is also a substantial disparity between mutation rates estimated directly from population studies and those inferred by phylogenetic (species level) studies (*Ho et al., 2005*). To avoid potential bias defined by the transition between short-term mutation and long-term substitution rate, we compared two molecular rates. The standard invertebrate mitochondrial divergence rate $\mu = 1.15 \times 10^{-8}$ per year (*Brower, 1994*; *Papadopoulou, Anastasiou & Vogler, 2010*), and the mutation rate based on *Drosophila melanogaster* laboratory strain estimations $\mu = 6.2 \times 10^{-8}$ per generation (*Haag-Liautard et al., 2008*). The latter was used to extrapolate a molecular rate of $4.29 \times 10^{-7}$ for *C. capitata* which has an average of 6.92 generations per year (*Diamantidis et al., 2011*). Each

run was validated in TRACER ensuring a minimum of 200 effective samplings for each statistic. The two-run results were combined using LogCombiner v. 2.4.5 (*Drummond & Rambaut, 2007*). Finally, the results were visualised by median of skyline plots using TRACER 1.6.

### Migration rate estimates

Connectivity was explored with the software LAMARC v.2.1.10 which estimates demographic parameters such as theta ($\theta$), population growth (g) and migration rates (M) (*Kuhner, 2006*). Theta values were estimated as $\theta = 2\mu N_e$, where $N_e$ is the effective population size, and $\mu$ represents the mutation rate per nucleotide and generation (see below for details). Migration rate was estimated as $M = m/\mu$, where $m$ is the probability of immigrants per generation and $\mu$ is the mutation rate per site per generation. The migration rate was multiplied by the $\theta$ value of the corresponding recipient population to obtain the migrants per generation value (Nm) (*Kuhner, 2006*). The search strategy consisted of five initial and four final chains; the Bayesian estimation was conducted with ten initial chains with an interval of 20 using a burn-in of 1,000 samples per chain. The analysis results were checked for convergence and effective sample size values (ESS $\geq$ 200) in TRACER.

## RESULTS

### Genetic diversity and population structure in *Ceratitis capitata*

A total of 403 sequences of *C. capitata* collected in 14 sites distributed worldwide were included in the analysis. The final truncated alignment was 538 bp in length corresponding to 179 amino acids of the mitochondrial COI gene. The number of segregating sites (*S*) within populations ranged from 23 in Kenya and one in Greece (Table 2), and the number of haplotypes (*k*) varied between 18 in Kenya and one and two in Tunisia and Greece, respectively. Tunisia was not included in the further analysis of intra-population diversity because of the lack of haplotype variation. Haplotype diversity (*h*) and nucleotide diversity ($\pi$) were much higher in the populations from Kenya and South Africa than anywhere else (Table 2), while the number of unique haplotypes was also far higher, despite the overall large amount of specimens analysed in several of the local populations, such as Spain, Guatemala or Israel (Table 2).

No insertion/deletion or stop codons were detected in the whole data set. Most nucleotide substitutions were synonymous, but six non-synonymous mutations were identified. They corresponded to changes from Methionine to Leucine (Iran), Alanine to Threonine (Peru), Isoleucine to Threonine (Kenya), Proline to Serine (Spain), Valine to Isoleucine (Israel and Australia). All of these changes were mapped to the tips of the haplotype network (Fig. 1).

The median-joining haplotype network contained a total of 58 distinct haplotypes with a low number of 'unsampled' (i.e., more than one mutational step apart) haplotypes (Fig. 1). The network was divided into a reticulated portion mainly consisting of Afrotropical haplotypes and a peripheral star-like portion composed of haplotypes from all other locations (Fig. 1). Overall, the Afrotropical haplotypes were more diverse (35 haplotypes from 64 sequences) than those of other regions combined (23 haplotypes from 339

**Table 2  Population genetic diversity indices and neutrality test statistics for *C. capitata*.** The indices are shown as $n$: number of samples; $k$: number of haplotypes; $S$: number of segregating sites; $h$: haplotype diversity (with standard deviation SD); $\pi$: nucleotide diversity (with standard deviation SD). Tajima's $D$ and Fu's Fs tests were considered statistically significant when *$P$-value<0.05, **$P$-value < 0.01 and ***$P$-value < 0.001.

| Biogeographic region | Population | Code | $n$ | $k$ | $S$ | $h \pm$ SD | $\pi \pm$ SD | Tajima's $D$ | Fu's Fs |
|---|---|---|---|---|---|---|---|---|---|
| Afrotropical | Kenya | KE | 22 | 18 | 23 | $0.969 \pm 0.027$ | $0.0057 \pm 0.0034$ | $-1.94$** | $-15.89$*** |
| | South Africa | SA | 37 | 17 | 16 | $0.941 \pm 0.020$ | $0.0053 \pm 0.0031$ | $-0.85$ | $-8.56$*** |
| | Ghana | GH | 5 | 4 | 4 | $0.900 \pm 0.161$ | $0.0033 \pm 0.0009$ | $-0.41$ | $-1.19$ |
| Palearctic | Egypt | EG | 25 | 3 | 3 | $0.353 \pm 0.112$ | $0.0016 \pm 0.0005$ | $0.28$ | $1.15$ |
| | Israel | IS | 109 | 3 | 4 | $0.072 \pm 0.034$ | $0.0002 \pm 0.0001$ | $-1.58$** | $-1.62$ |
| | Tunisia | TU | 1 | 1 | – | – | – | – | – |
| | Iran | IR | 12 | 3 | 2 | $0.318 \pm 0.164$ | $0.0006 \pm 0.0003$ | $-1.14$ | $-1.18$* |
| | Greece | GR | 29 | 2 | 1 | $0.069 \pm 0.063$ | $0.0001 \pm 0.0001$ | $-1.02$ | $-2.38$ |
| | Spain | SP | 42 | 6 | 5 | $0.592 \pm 0.068$ | $0.0012 \pm 0.0012$ | $-1.69$* | $-3.34$*** |
| Australasian | Australia | AU | 24 | 5 | 4 | $0.377 \pm 0.122$ | $0.0007 \pm 0.0002$ | $0.03$ | $-3.63$* |
| Neotropical | Guatemala | GU | 55 | 10 | 6 | $0.766 \pm 0.039$ | $0.0024 \pm 0.0003$ | $0.25$ | $0.24$ |
| | Colombia | CO | 22 | 3 | 2 | $0.567 \pm 0.051$ | $0.0011 \pm 0.0001$ | $0.15$ | $-0.01$ |
| | Brazil | BR | 12 | 3 | 2 | $0.621 \pm 0.087$ | $0.0013 \pm 0.0002$ | $0.06$ | $-0.22$ |
| | Peru | PE | 8 | 3 | 2 | $0.679 \pm 0.122$ | $0.0014 \pm 0.0003$ | $0.06$ | $-0.22$ |

sequences). The Afrotropical cluster was connected to all others via haplotype Cc_21, which was the most frequent haplotype across the dataset (62.28%) and present in almost all localities. This haplotype occupied a central position in the network from which other haplotypes with local distribution in the Palearctic, Australasian and Neotropical regions were derived. Two specimens from South Africa exhibited haplotype Cc_21 and thus constituted a direct link of Afrotropical and other populations.

Eight haplotypes were shared between at least two localities (Fig. 2, see the identification code in Table 2), of which the haplotypes Cc_42 and Cc_49 were the most dominant, besides the ubiquitous Cc_21. In the Afrotropical cluster, only the haplotypes Cc_04, Cc_06, Cc_13, and Cc_14 were shared between South Africa and Kenya, but none of them was the centre of an expanded genealogy as the common haplotype Cc_21 in the rest of the world. The pairwise $F_{st}$ analysis performed on the 13 localities (Tunisia excluded) showed that the majority of the populations were significantly differentiated (Table 3). Some exceptions were found for neighbouring sites including South Africa and Kenya in the Afrotropical group; Iran, Egypt, Israel and Greece in the Palearctic; or Brazil and Colombia in the Neotropical region. However, some remote sites also presented non-significant differences such as Iran and Australia (Table 3).

## Demographic history

Across the entire dataset, only six of thirteen sites (Tunisia excluded) were significant for Tajima's $D$ and Fu's F$s$ (Table 2). From the Afrotropical cluster, South Africa and Kenya were highly significant and negative for these neutrality tests. Negative values were also found in the Palearctic (Israel, Iran, Spain) and Australasian regions. These findings may indicate either purifying selection acting on protein coding regions or may be due to recent population expansion that favour a non-random variation of haplotypes.
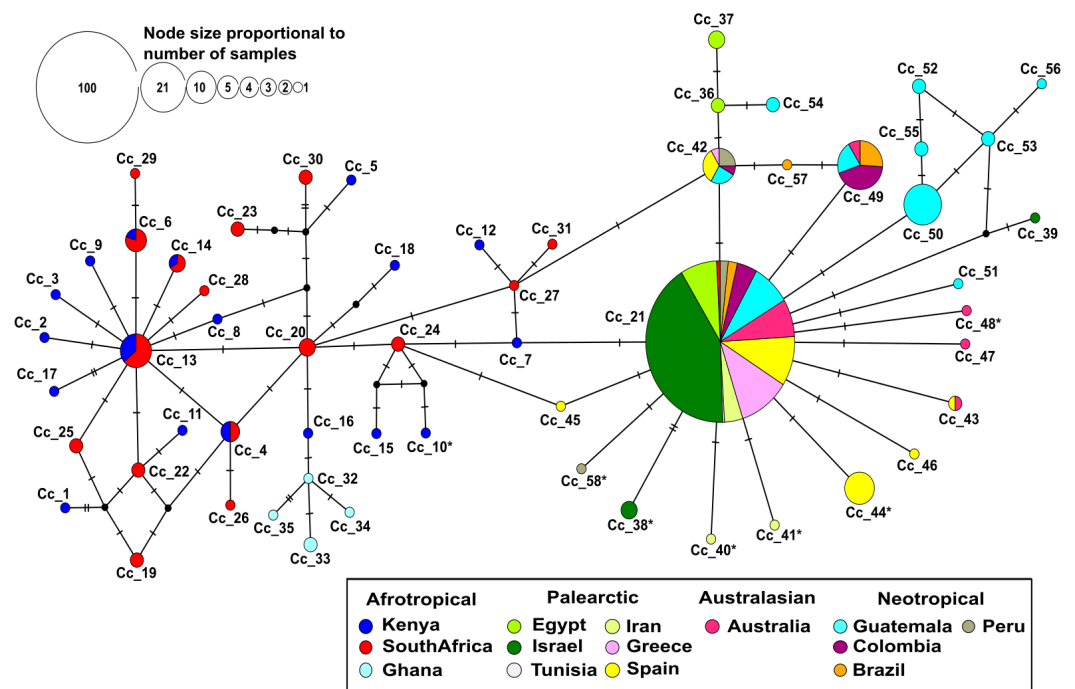
**Figure 1** **Median-joining network based on 403 individuals of the Mediterranean fruit fly generated using 538 bp of mtDNA COI gene, showing location and frequency of haplotypes.** Each circle represents an observed haplotype; the colours reflect sampling location and small black circles indicate unsampled haplotypes inferred from the data. The reticulated network segregated haplotypes according to the different biogeographic region. The more common haplotype in the Afrotropical region cluster is Cc_13 from where singletons are extending outwards. On the other side, the most common haplotype Cc_21 occupies a central position with starburst shape radiation from which the other haplotypes related to Palearctic, Australasian and Neotropical regions are derived. Cc, correspond to *C. capitata*. The * in the haplotype label refers to non-synonymous mutation.

Full-size ◱ DOI: 10.7717/peerj.5340/fig-1

The Bayesian skyline plots exhibited differences in the effective population size calculated among the biogeographical regions (Fig. 3; only the results obtained from the simulations using the corrected mutation rate of *D. melanogaster* are shown). The time to the most recent common ancestor (*tmrca*) was estimated at around 11,600 years ago in the Afrotropical (Fig. 3A). This group also showed a substantial increment (one order of magnitude) in the effective population size after the outset of the Holocene (~10,000 years ago) suggesting a signature of recent expansion which became significant around 3,500 years ago (i.e., when the 95% highest posterior density (HPD) limits no longer includes older estimates), after which the population size was largely stable until the present time (Fig. 3A). In contrast, the Palaearctic group (Fig. 3B) had a lower effective population size and showed more recent date estimations compared to the Afrotropical group, which exhibited significant population expansion only after about 500 years ago. The Australasian and Neotropical groups (Figs. 3C and 3D) remained at a stable population size from about 1,000 years ago and then showed a slight but in significant increment.

The coalescent analysis using Lamarc estimated a $\theta$ ranging from 0.1550 (Afrotropical) to 0.003 (Australasian). The Lamarc results also indicated asymmetric migration between
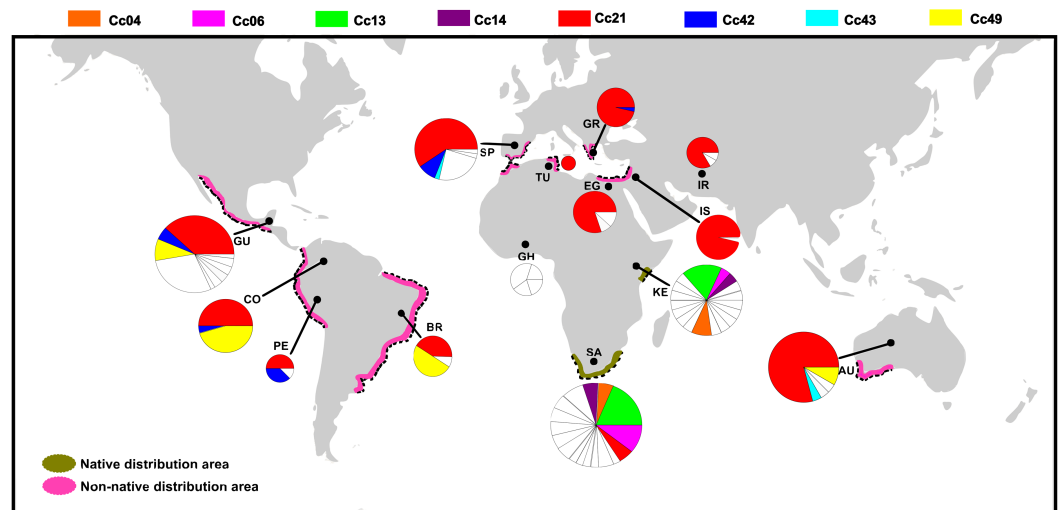
**Figure 2** **Distribution of COI haplotypes across the study area for *Ceratitis capitata*.** The map shows the study locations (country names are abbreviated as in Table 2), and the pie charts indicate the haplotype composition of the population from that location. Each colour represents a shared haplotype found across the study area, and the unique haplotypes (refer to haplotypes found in the samples from one particular population and are absent in the samples from other populations) are uniformly represented in white within pie charts. Native and non-native areas are represented according to *Malacrida et al. (2007)*.
Full-size ◳ DOI: 10.7717/peerj.5340/fig-2

**Table 3** **Pairwise *Fst* values between 13 populations of *C. capitata*.** The *Fst* values for pairwise comparisons among populations calculated from mtDNA data. Significant tests were performed using 1,000 permutations. Bold values are statistically significant.

| | KE | SA | GH | EG | IS | IR | GR | SP | AU | GU | CO | BR | PE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **KE** | | | | | | | | | | | | | |
| **SA** | −0.002 | | | | | | | | | | | | |
| **GH** | **0.058** | **0.074** | | | | | | | | | | | |
| **EG** | **0.346** | **0.302** | **0.492** | | | | | | | | | | |
| **IS** | **0.694** | **0.608** | **0.854** | **0.146** | | | | | | | | | |
| **IR** | **0.315** | **0.278** | **0.474** | −0.009 | 0.107 | | | | | | | | |
| **GR** | **0.516** | **0.439** | **0.772** | **0.079** | −0.007 | 0.054 | | | | | | | |
| **SP** | **0.236** | **0.210** | **0.312** | **0.088** | **0.313** | 0.070 | **0.189** | | | | | | |
| **AU** | **0.332** | **0.289** | **0.471** | 0.004 | **0.125** | -0.024 | 0.062 | **0.074** | | | | | |
| **GU** | **0.140** | **0.131** | **0.187** | **0.169** | **0.416** | **0.153** | **0.280** | **0.112** | **0.149** | | | | |
| **CO** | **0.231** | **0.211** | **0.320** | **0.236** | **0.576** | **0.220** | **0.415** | **0.168** | **0.168** | **0.118** | | | |
| **BR** | **0.188** | **0.177** | **0.269** | **0.297** | **0.689** | **0.280** | **0.542** | **0.196** | **0.228** | **0.115** | −0.051 | | |
| **PE** | **0.154** | **0.142** | **0.224** | **0.187** | **0.621** | 0.168 | **0.454** | 0.055 | **0.166** | 0.075 | 0.161 | 0.181 | |

**Notes.**

Population code: KE, Kenya; SA, South Africa; GH, Ghana; EG, Egypt; IS, Israel; IR, Iran; GR, Greece; SP, Spain; AU, Australia; GU, Guatemala; CO, Colombia; BR, Brazil; PE, Peru.

the Afrotropical and the other populations, whereby the Palearctic received the lowest migrants per generation (Nm = 4.35) and the Australasian population the highest (Nm = 5.47), while migrant flow in the opposite direction was insignificant (Fig. 4A). Migration analyses conducted exclusively on populations from within each biogeographic region
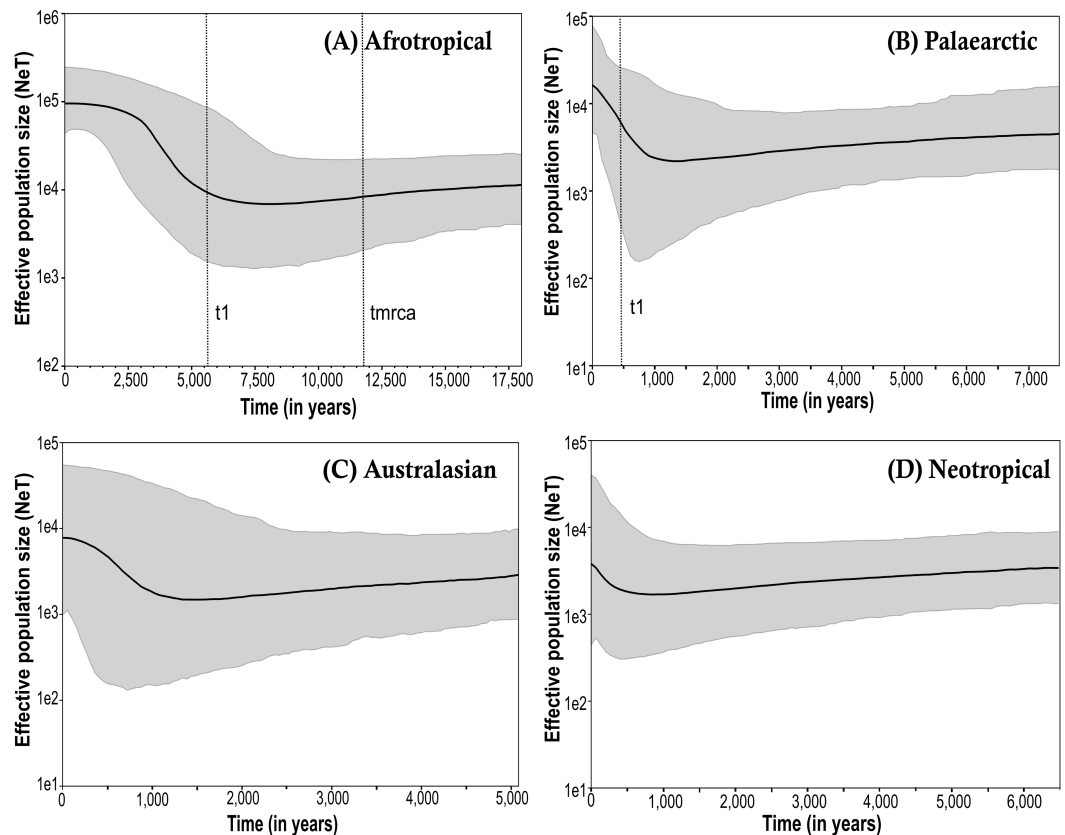
**Figure 3** **Bayesian skyline plot (BSP) estimate of Medfly demographic history for the biogeographic regions (A) Afrotropical; (B) Palaearctic; (C) Australasian and (D) Neotropical.** The $X$-axis is in units of time before present (BP), and the $Y$-axis is equal to the log scale of $N_eT$ (the product of effective population size and the generation time in years). Each BSP plot described the demographic history per biogeographic region represented by a median line (solid line horizontal) with 95% High Posterior Distribution in grey (HPD; equivalent to margins of error). The dashed line t1 is the time of population expansion per biogeographic group. The Time to the most recent common ancestor (tmrca) is represented in the Afrotropical population around 11,600 BP.

Full-size ☒ DOI: 10.7717/peerj.5340/fig-3

showed high levels of unidirectional exchange within the Afrotropical and Neotropical regions (Fig. 4B), such as the remarkably high Nm SA to KE = 133.4 and Nm CO to GU = 4.5, while migrants were notably lower in the reverse direction (Nm KE to SA = 0.429 and Nm GU to CO = 0.032 respectively) (Fig. 4B).

## DISCUSSION

The aim of this study was to describe the current genetic structure and recent demography of *C. capitata* and to provide insights into potential invasion routes leading to its worldwide distribution. Our extensive phylogeographic analysis has revealed a rapid colonisation process over the last 500 years, and a complex genetic structure of *C. capitata* with clear variation between biogeographic regions. The colonisation process of the medfly is well documented by both historical records (*Malacrida et al., 1998*; *Myers et al., 2000*) and
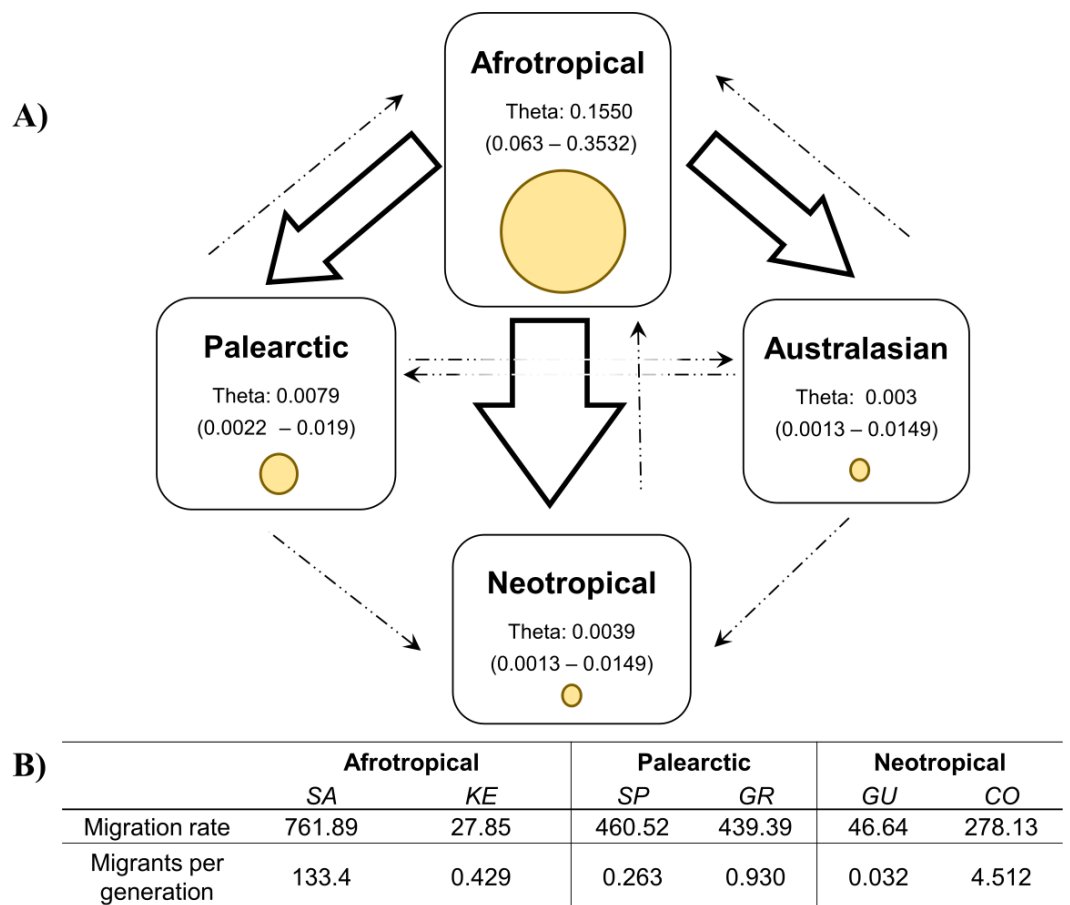
**Figure 4  Values of theta and migration between the biogeographical regions.** The figure (A) contains the theta and migration values among the four biogeographic regions. Yellow circles represent the theta value per biogeographic group and the values in brackets are the 95% HPD confidence. The arrows indicate the direction of the migration and their thickness is the proportion of the migrants per generation. (B) Migration rate and migrants per generation of specific medfly populations within the biogeographic region, country names are abbreviated as in Table 2

Full-size 🖼 DOI: 10.7717/peerj.5340/fig-4

molecular studies (*Barr, 2009*; *Bonizzoni et al., 2004*; *Gasperi et al., 2002*; *Karsten et al., 2015*; *Karsten et al., 2013*). In fact, the recreation of *C. capitata* invasion routes in this study fits broadly with that proposed in the literature, i.e., the medfly populations first migrated from the ancestral Afrotropical region to the Palearctic and then to the Australasian and Neotropical regions (*Malacrida et al., 2007*).

The highest genetic diversities were found in Kenya and South Africa belonging to the Afrotropical cluster. This was expected because the southeastern African countries had been identified as the medfly's ancestral native range (*De Meyer et al., 2002*). Western African areas have been proposed to be part of this large native population source distributed across all of Sub-Saharan Africa (*Gasparich et al., 1997*), but this interpretation conflicts with our finding that the sample from Ghana shows lower genetic variation than Kenya and South Africa, and the *Fst* results are statistically significantly between Ghana and the other two

populations. In the network, the Ghana haplotypes are part of the Afrotropical cluster, but all of them are unique. These findings support the existence of native, but genetically differentiated populations in West Africa, although the number of individuals (a total of five) remains too low to resolve the contradicting literature on the subject of population subdivision in Sub-Saharan Africa (*De Meyer et al., 2002*; *Gasperi et al., 2002*; *Malacrida et al., 1992*).

Biological invasions are often associated with a decrease in genetic diversity of the invasive species due to a small number of founder events in their introduced ranges (*Lockwood, Cassey & Blackburn, 2005*; *Lockwood, Cassey & Blackburn, 2009*). It is therefore unsurprising that we found evidence of a gradual loss of genetic variability from the ancestral Afrotropical region to the Palearctic and all other populations. Low genetic diversity was particularly obvious in the population from Israel, which was represented by the largest number of individuals of all sampled regions, and yet exhibited very low levels of genetic diversity. Similar results were reported for another mitochondrial gene ND4 (NADH subunit 4) in two different populations collected in Israel (*Elfekih, Makni & Haymer, 2010*). Iran and Greece also had low genetic variation compared to other populations, possibly because of limited hosts and climatic ranges suitable for the medfly in these regions. In addition, the constant eradication efforts in these countries, and in particular the use of SIT, might have resulted in occasional population bottlenecks and reduced genetic diversity in these populations. In contrast, the populations from Spain were the most genetically diverse in the Palearctic. The finding may in part be affected by the origin of the Spanish specimens, which were from multiple sites and thus may contain local variation that is not incorporated at most other country samples. In addition, populations in Spain might have a longer phylogeographic history and thus greater diversification, as the likely entry point to the Mediterranean basin.

The curious presence of a shared haplotype (Cc_21) among extremely distant populations suggests a recent connection of all non-native populations. This common haplotype is also present in South Africa, but only as a very small proportion, and it is peripheral to the haplotype network of African haplotypes. While the distribution of Cc_21 indicates a shared history of all non-native populations, the derivation of this haplotype from the South African population is not strongly supported. An individual carrying the Cc_21 haplotype could be the ancestor to all non-native populations, even if this haplotype was rare in the ancestral population, but conceivably the source population could be from elsewhere in Africa where the Cc_21 haplotype is more prevalent. Only more detailed surveys of native African populations will resolve this question.

The interpretation of demographic history results differs among the biogeographic region. The Bayesian Skyline plot result for the Afrotropical region showed it to be the most ancient population dated to about 11,600 years ago. Nevertheless, this time frame is far younger than ages usually associated with the time since species originated, or even the age of a closely similar tephritid fossil found in the Dominican Republic dated to the mid-Miocene to early Eocene (*Norrbom, 1994*). However, the signature of population expansion coincides with a period when the region underwent the African Humid Period which is characterised by major climatic changes that influenced ancient human settlements

(*Manning & Timpson, 2014*). In this context, new strategies for plant use were developing in Africa about 17,000 years ago, although plant domestication was recognised only later at around 4,000 BP (*Marshall & Hildebrand, 2002*). In the Afrotropical region, the significant expansion signature detected about 3,500 years ago by the Bayesian skyline plot is coincident with the plant domestication period in the region. Currently, population size in this region is stable as we can see in the plot but also can be supported by the negative Tajima's $D$ and Fu's values for Kenya and South Africa which are best interpreted as the result of purifying selection, as expected in mitochondrial protein-coding gene evolution (*Meiklejohn, Montooth & Rand, 2007*).

On the other hand, in Spain as one of the first points of colonisation and presumed early origin of Mediterranean populations, the patterns of COI variation may be explained by purifying selection. In contrast, the Palearctic BSP showed a population expansion, as also described before in this region (*Reyes & Ochando, 2004*). Non-synonymous substitutions, which are generally rare in mitochondrial genes, especially in the cytochrome oxidase genes (*Pentinsaari et al., 2016*), were found in all regions, but predominantly in introduced populations characterised by non-significant neutrality tests as expected for star-like topologies. Their position near the tips of the haplotype network suggests that they correspond to neutral variation or slightly deleterious mutations that are maintained in fast expanding populations, rather than adaptive changes affecting, for example, the metabolic rate due to the new environmental conditions exposed (*Castoe et al., 2008*), and thus these changes are consistent with the inference of fast population expansion.

The migration patterns within geographic regions might be affected differently in various parts of the world (Fig. 4). For example, in the Neotropics pest management differs notably among the countries. Guatemala is recognised for the successful establishment of the Mediterranean Fruit Fly Eradication Program (Moscamed Program) which is effectively containing the medfly within Central America (*Aluja & Liedo, 1993*; *Enkerlin et al., 2017*). A notable success of this program was the development of the Sterile Insect Technique (SIT) (*Szyniszewska & Tatem, 2014*). On the other side, Colombia has also recognised the presence of the medfly (*ICA, 2010*) and triggered the National Fruit Fly program focusing on detection, control and eradication methods of medfly based on mass trapping and chemical application (*Conpes, 2008*; *Lasprilla, 2011*). Despite of that, these phytosanitary efforts are not enough to reduce the potential pest risk of Colombian commodities (*PPHIS, 2018*; *Szyniszewska et al., 2016*) and the COI study produced clear evidence for the unidirectional migration from Colombia to Guatemala. Given the differences in their pest control management methods, there is a risk that high migration rate from Colombia now interferes with the successful Guatemalan program established some 40 years ago.

## CONCLUSIONS

The colonisation process of the medfly appears to be associated with a relatively stable demographic structure separating the Afrotropical region and the introduced range (Palearctic, Neotropical and Australasian), but characterised by residual levels of connectivity at regional scales despite considerable distance separating the populations,

such as Egypt and Iran or Brazil and Peru. However, the COI marker used in this study has limitations due to comparatively low variation that may be insufficient to resolve events on the time scale of the medfly dispersal. Yet using an appropriate mutation rate, the demographic analysis produced plausible scenarios associated with the Holocene era, which is closely related to the agriculture and domestication process in the humanity. The inferred migration patterns among populations provide crucial information for the understanding of successful medfly invasions and thus pinpointing where countermeasures are required, in particular in a world connected via agricultural commodities trade. The case of successful containment in Guatemala and the dangers of fruit fly migration from elsewhere in the South and Central American regions illustrate these problems clearly. We used the most basic of molecular markers to study these phenomena, based on short fragment of a single locus, and studied pattern and process of medfly history at global levels based on just 14 local sites. The results are highly plausible and consistent with other studies using diverse approaches. However, the conclusions have to remain tentative, given the limited detail of sampling. Genomic approaches and much denser sampling at regional and global scales will be required to confirm the conclusions drawn here.

## ACKNOWLEDGEMENTS

## ADDITIONAL INFORMATION AND DECLARATIONS

### Competing Interests

The authors declare there are no competing interests.

## Author Contributions

- Maria Belen Arias conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Samia Elfekih conceived and designed the experiments, performed the experiments, approved the final draft.
- Alfried P. Vogler conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

Genbank: MH628362–MH628394.

## Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5340#supplemental-information.

## REFERENCES

**Aluja M, Liedo P. 1993.** *Fruit flies: biology and management.* New York: Springer-VerJag.

**APHIS. 2018.** Supplemental requirements for importation of fresh citrus from Colombia Into the United States. In: *Animal and Plant Health Inspection Service U, editor.* *APHIS–2017–0074.* Washington: Federal Register.

**Reyes A, Ochando MD. 2004.** Mitochondrial DNA variation in Spanish populations of *Ceratitis capitata* (Wiedemann) (Tephritidae) and the colonization process. *Journal of Applied Entomology* **128**:358–364 DOI 10.1111/j.1439-0418.2004.00858.358-364.

**Bandelt HJ, Forster P, Röhl A. 1999.** Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**:37–48 DOI 10.1093/oxfordjournals.molbev.a026036.

**Barr NB. 2009.** Pathway analysis of *Ceratitis capitata* (Diptera: Tephritidae) using mitochondrial DNA. *Journal of Economic Entomology* **102**:401–411 DOI 10.1603/029.102.0153.

**Blackburn TM, Pysek P, Bacher S, Carlton JT, Duncan RP, Jarosik V, Wilson JR, Richardson DM. 2011.** A proposed unified framework for biological invasions. *Trends in Ecology and Evolution* **26**:333–339 DOI 10.1016/j.tree.2011.03.023.

**Bonizzoni M, Guglielmino CR, Smallridge CJ, Gomulski M, Malacrida AR, Gasperi G. 2004.** On the origins of medfly invasion and expansion in Australia. *Molecular Ecology* **13**:3845–3855 DOI 10.1111/j.1365-294X.2004.02371.x.

**Bonizzoni M, Malacrida AR, Guglielmino CR, Gomulski LM, Gasperi G, Zheng L. 2000.** Microsatellite polymorphism in the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Molecular Biology* **9**:251–261.

**Bonizzoni M, Zheng L, Guglielmino C, Haymer D, Gasperi G, Gomulski L, Malacrida A. 2001.** Microsatellite analysis of medfly bioinfestations in California. *Molecular Ecology* **10**:2515–2524.

Arias et al. (2018), *PeerJ*, DOI 10.7717/peerj.5340

14/17

**Brower AV. 1994.** Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**:6491–6495.

**Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD. 2008.** Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLOS ONE* **3**:e2201 DOI 10.1371/journal.pone.0002201.

**Conpes. 2008.** *Política Nacional fitosanitaria y de inocuidad para las cadenas de frutas y otros vegetales.* Consejo Nacional de Política Económica y Social República de Colombia, p 45.

**De Meyer M, Copeland RS, Wharton RA, McPheron BA. 2002.** On the geographic origin of the Medfly *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae). In: Barnes BN, ed. *Proceedings of the 6th International Symposium on fruit flies of economic importance.* Stellenbosch, South Africa, 45–53.

**Diamantidis AD, Carey JR, Nakas CT, Papadopoulos NT. 2011.** Population-specific demography and invasion potential in medfly. *Ecology and Evolution* **1**:479–488 DOI 10.1002/ece3.33.

**Diamantidis AD, Carey JR, Papadopoulos NT. 2008.** Life-history evolution of an invasive tephritid. *Journal of Applied Entomology* **132**:695–705 DOI 10.1111/j.1439-0418.2008.01325.x.

**Drummond AJ, Rambaut A. 2007.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**:214 DOI 10.1186/1471-2148-7-214.

**Dyck VA, Hendrichs J, Robinson S. 2005.** *Sterile insect technique. Principles and practice in area-wide integrated pest management.* Netherlands: Springer.

**Elfekih S, Makni M, Haymer DS. 2010.** Detection of novel mitochondrial haplotype variants in populations of the Mediterranean fruit fly, *Ceratitis capitata*, from Tunisia, Israel and Morocco. *Journal of Applied Entomology* **134**:647–651 DOI 10.1111/j.1439-0418.2009.01500.x.

**Enkerlin WR, Gutiérrez Ruelas JM, Pantaleon R, Soto Litera C, Villaseñor Cortés A, Zavala López JL, Orozco Dávila D, Montoya Gerardo P, Silva Villarreal L, Cotoc Roldán E, Hernández López F, Arenas Castillo A, Castellanos Dominguez D, Valle Mora A, Rendón Arana P, Cáceres Barrios C, Midgarden D, Villatoro Villatoro C, Lira Prera E, Zelaya Estradé O, Castañeda Aldana R, López Culajay J, Ramírez y Ramírez F, Liedo Fernández P, Ortíz Moreno G, Reyes Flores J, Hendrichs J. 2017.** The Moscamed Regional Programme: review of a success story of area-wide sterile insect technique application. *Entomologia Experimentalis et Applicata* **164**:188–203 DOI 10.1111/eea.12611.

**Excoffier L, Lischer HE. 2010.** Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**:564–567 DOI 10.1111/j.1755-0998.2010.02847.x.

**Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994.** DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**:294–299.

**Fu Y. 1997.** Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915–925.

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

**Gasperi G, Bonizzoni M, Gomulski LM, Murelli V, Torti C, Malacrida AR, Guglielmino CR. 2002.** Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*. *Genetica* **116**:1573–6857.

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

**Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. 2008.** Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLOS Biology* **6**:e204 DOI 10.1371/journal.pbio.0060204.

**Ho SY, Phillips MJ, Cooper A, Drummond AJ. 2005.** Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology Evolution* **22**:1561–1568 DOI 10.1093/molbev/msi145.

**ICA. 2010.** *Boletin epidemiologico seccional Narino. Plan Nacional Moscas de la Fruta.* Colombia: Instituto Colombiano Agropecuario.

**Karsten M, Van Vuuren BJ, Addison P, Terblanche JS, Leung B. 2015.** Deconstructing intercontinental invasion pathway hypotheses of the Mediterranean fruit fly (*Ceratitis capitata*) using a Bayesian inference approach: are port interceptions and quarantine protocols successfully preventing new invasions? *Diversity and Distributions* **21**:813–825 DOI 10.1111/ddi.12333.

**Karsten M, Van Vuuren BJ, Barnaud A, Terblanche JS. 2013.** Population genetics of *Ceratitis capitata* in South Africa: implications for dispersal and pest management. *PLOS ONE* **8**:e54281 DOI 10.1371/journal.pone.0054281.

**Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012.** Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649 DOI 10.1093/bioinformatics/bts199.

**Kuhner MK. 2006.** LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22**:768–770 DOI 10.1093/bioinformatics/btk051.

**Lasprilla D. 2011.** Estado actual de fruticultura colombiana y perspectivas para su desarrollo. *Revista Brasileira de Fruticultura* **33**:199–205.

**Leigh JW, Bryant D. 2015.** Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**:1110–1116 DOI 10.1111/2041-210x.12410.

**Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452 DOI 10.1093/bioinformatics/btp187.

**Lockwood JL, Cassey P, Blackburn T. 2005.** The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution* **20**:223–228 DOI 10.1016/j.tree.2005.02.004.

**Lockwood JL, Cassey P, Blackburn TM. 2009.** The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions* **15**:904–910 DOI 10.1111/j.1472-4642.2009.00594.x.

**Mack RN, Simberloff D, Mark Lonsdale W, Evans H, Clout M, Bazzaz FA. 2000.** Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* **10**:689–710.

**Malacrida AR, Gomulski LM, Bonizzoni M, Bertin S, Gasperi G, Guglielmino CR. 2007.** Globalization and fruitfly invasion and expansion: the medfly paradigm. *Genetica* **131**:1–9 DOI 10.1007/s10709-006-9117-2.

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

**Malacrida AR, Marinoni F, Torti C, Gomulski LM, Sebastiani F, Bonvicini C, Gasperi G, Guglielmino CR. 1998.** Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. *The American Genetic Association* **89**:501–507.

**Manning K, Timpson A. 2014.** The demographic response to Holocene climate change in the Sahara. *Quaternary Science Reviews* **101**:28–35 DOI 10.1016/j.quascirev.2014.07.003.

**Marshall F, Hildebrand E. 2002.** Cattle before crops: the beginnings of food production in Africa. *Journal of World Prehistory* **16**:99–143.

**Meiklejohn CD, Montooth KL, Rand DM. 2007.** Positive and negative selection on the mitochondrial genome. *Trends in Genetics* **23**:259–263.

**Myers JH, Simberloff D, Kuris A, Carey JR. 2000.** Eradication revisited: dealing with exotic species. *Trends in Ecology & Evolution* **15**:316–320.

**Norrbom AL. 1994.** New genera of Tephritidae (Diptera) from Brazil and Dominican Amber, with phylogenetic analysis of the tribe Ortalotrypetini. *Insecta Mundi* **8**:1–15.

**Papadopoulou A, Anastasiou I, Vogler AP. 2010.** Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution* **27**:1659–1672 DOI 10.1093/molbev/msq051.

**Pentinsaari M, Salmela H, Mutanen M, Roslin T. 2016.** Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Scientific Reports* **6**:35275 DOI 10.1038/srep35275.

**Szyniszewska AM, Leppla NC, Huang Z, Tatem AJ. 2016.** Analysis of seasonal risk for importation of the mediterranean fruit fly, *Ceratitis capitata* (Diptera: Tephritidae), via air passenger traffic arriving in Florida and California. *Journal of Economic Entomology* **109**:2317–2328 DOI 10.1093/jee/tow196.

**Szyniszewska A, Tatem A. 2014.** Global assessment of seasonal potential distribution of Mediterranean fruit fly, *Ceratitis capitata*. *PLOS ONE* **9**:e111582.

**Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.

**Vogel V, Pedersen JS, Giraud T, Krieger MJB, Keller L. 2010.** The worldwide expansion of the Argentine ant. *Diversity and Distributions* **16**:170–186 DOI 10.1111/j.1472-4642.2009.00630.x.