



# Looking at the big picture: worldwide population structure and range expansion of the cosmopolitan pest *Ceratitis capitata* (Diptera, Tephritidae)

Pablo Deschepper · Terrance N. Todd · Massimiliano Virgilio · Marc De Meyer · Norman B. Barr · Raul Ruiz-Arce

Received: 2 December 2020 / Accepted: 1 July 2021

© The Author(s) 2021

**Abstract** The Mediterranean fruit fly, *Ceratitis capitata* (Weidemann), is considered one of the most significant tephritid pest species worldwide and is an exotic species in most of its range. Here, we investigated polymorphism at 14 microsatellite loci for a total of 126 populations of *C. capitata* from six geographical regions, applying network theory and cluster analyses. Analyses revealed nine distinct modules for the Central American region and one in each of the remaining five regions. Bayesian cluster analysis revealed that the highest level of genetic partitioning corresponds with the presence of 3 well-defined genetic clusters. Our results confirm the African origin for Mediterranean populations based on genetic diversity and suggest a direct invasion of *C. capitata* from the Mediterranean to Central-America. South American populations show links with Central-America, but also exhibit indications of direct admixture with the European cluster. Additionally, the

network analysis proposes a South American origin for the Madeiran and Hawaiian flies. Cluster analysis corroborates the hypothesis of a Mediterranean origin for Australian samples. Our work provides novel insights regarding the migration history of Medfly worldwide.

**Keywords** Population graph · Network analysis · Modularity · Medfly · Invasion history

## Introduction

International trade has surged tremendously in the last few decades (Hulme 2009; Sardain et al. 2019; WTO 2019). The increase in trade and private travel between continents causes a higher risk of introduction of exotic species that can pose serious threats to agriculture (Paini et al., 2016). Phytophagous insects of the dipteran family Tephritidae (the true fruit flies) are considered one of the most impactful pest groups for fruits and vegetables worldwide (Prokopy et al. 1999). One species in particular, the Mediterranean fruit fly or medfly (*Ceratitis capitata*) is considered one of the most notorious invaders. It originates from tropical Sub-Saharan Africa (De Meyer et al. 2004), but during the past century it has spread to other tropical and more temperate regions worldwide (Arias et al. 2018; Malacrida et al. 2007; Nikolouli 2020). Medfly is now established in northern Africa, the Mediterranean,

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10530-021-02595-4>.

---

P. Deschepper (✉) · M. Virgilio · M. De Meyer  
Royal Museum for Central Africa, B3080 Tervuren,  
Belgium  
e-mail: pablo.deschepper@africamuseum.be

T. N. Todd · N. B. Barr · R. Ruiz-Arce  
USDA-APHIS-PPQ-Science and Technology, 22675 N.  
Moorefield Rd. Bldg 6414, Edinburg,  
TX 78541, USA

Latin America, Australia and the Hawaiian islands (Barr 2009; Bonizzoni et al. 2004; Malacrida et al. 2007) and is annually detected in California (Bonizzoni et al. 2001; Carey 1991; Carey et al. 2017).

The success of the medfly can be largely attributed to an extensive list of hosts that includes numerous fruits and vegetables from more than 50 families (Liquido et al. 1990, 2013). The medfly genome is 479 Mb and, in comparison to *Drosophila melanogaster* and *Musca domestica*, it has gene expression patterns that suggest adaptation to a phytophagous biology (Papanicolaou et al. 2016). In addition, *C. capitata* demonstrates a higher phenotypic plasticity towards thermal tolerance in comparison to related *Ceratitis rosa*, which could also contribute to its invasiveness (Nyamukondiwa and Terblanche 2010).

Eradication and pest management programs for this fly are expensive. Estimates suggest that medfly outbreaks in California have cost taxpayers nearly \$500 million (Szyniszewska and Tatom 2014; Ruiz-Arce et al. 2020). In Spain, pest management of medfly in 2017 alone cost €8.5 million (\$9.7 million) to the regional government of Valencia (Dalmau, communication with authors). Pest management of medfly primarily relies on Sterile Insect Technique (SIT) and insecticide use (Enkerlin et al. 2015; Klassen and Curtis 2005).

Population genetic studies can help mitigate the crop damage caused by *C. capitata* by disentangling the genetic structuring of medfly populations. Recently, Ruiz-Arce et al. (2020) have used mitochondrial DNA sequencing methods on 150 *C. capitata* localities (collection maintained by APHIS, USDA), covering largely the worldwide distribution of medfly. They revealed a total of 202 haplotypes in four different haplogroups that show a strong association to geography. Their study complements earlier mitochondrial analyses on more limited collections (Arias et al. 2018; Barr 2009).

Here we applied integrative graph theoretic methods to a smaller subset of the USDA APHIS collection analyzed by Ruiz-Arce et al. (2020). This network approach combined with modularity analysis can be a particularly helpful tool to reveal pathways of gene exchange (Fortuna et al. 2009) and can help to target specific trade routes in the prevention of further medfly invasion.

## Methods

A total of 1214 specimens of *C. capitata* from 126 collections (with numbers of specimens per sample ranging from 1 to 40) have been collected between 1956 and 2016 across 46 countries distributed across six main geographical regions: Sub Saharan Africa (SSA), Europe/Mediterranean (MED), Central (CAM) and South America (SAM), Australia (AUS) and Hawaii (HAW) (see Table S1 for locality, coordinates and year of collection for each population). The geographic regions used in this study are in congruence with delimitation used in previous studies (Barr 2009; Ruiz-Arce et al. 2020). The fly specimens used for this study are a subset of the USDA APHIS collection genotyped in Ruiz-Arce et al. (2020) comprising 1592 flies of 144 fly collections and is based on the patterns observed in the study of Ruiz-Arce et al. (2020). Flies were collected in traps and were identified by the collectors.

### DNA extraction and genotyping

DNA was extracted from ethanol and dry preserved adults following InviMag® Tissue DNA Mini Kit/KF96 (STRATEC Biomedical AG, Birkenfeld, Germany), on an automated magnetic-particle nucleic acid purification system, KingFisher™ Flex (Model# 711, Thermo Scientific, Waltham, Massachusetts). Additional details on extraction methods are reported in Todd et al. (2017). Individual flies were genotyped using 14 polymorphic microsatellite loci as described in Bonizzoni et al. (2000) (ccmic14, ccmic15, ccmic25, ccmic3, ccmic32, ccmic6, ccmic9) and Stratikopoulos et al. (2009) (medflymic30, medflymic43, medflymic67, medflymic74, medflymic78, medflymic81, medflymic92). PCR was performed in multiplex reactions divided among six panels according to Todd et al. (2017). DNA was screened on a 2.5% agarose gel and diluted 1:10 with water. Fragment analyses were conducted on an automated ABI Prism 3100 Genetic Analyzer (Applied Biosystem) and scored with Applied Biosystems PeakScanner v1.0 (Foster City, California) with individuals declared non-amplifiable at a locus after two independent amplification failures.

## Analysis of genetic diversity

The R (R Core Team 2019) package *adegenet* 1.4–2 (Jombart 2008) was used to quantify population genetic variability and differentiation, including number of alleles per locus ( $N_a$ ), observed and expected heterozygosity ( $H_o$ ,  $H_e$ ). A Kruskal–Wallis test was performed using the stats package (R Core Team 2019) in R to assess differences between the geographic regions in regard of  $N_a$ ,  $H_o$  and  $H_e$ . The Benjamini–Hochberg method was applied to correct for the false discovery rate introduced by multiple testing. The absolute number and relative frequency of private alleles at each locus was calculated in R (R Core Team 2019) using the *allele.dist* (*PopGenReport* package, Adamack and Gruber 2014) and *private\_alleles* (*poppr* package, Kamvar et al. 2014) functions respectively. Private alleles were calculated for populations as well as for the main geographical regions. A genotype accumulation curve was plotted using R package *poppr* 2.8.3 (Kamvar et al. 2014) to determine whether enough microsatellite loci were used to allow discrimination between individuals. Deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium were tested for each population across each pair of loci using the log likelihood ratio statistic implemented in the *genepop* package for R (Rousset et al. 2017; R Core Team 2019). Assessment of significance was performed through Markov-chain randomizations based on 10,000 dememorizations, 20 batches, and 5000 iterations per batch. The same software was used to estimate null allele frequencies (per locus and population).

## Structuring of genetic variation

In order to visualize the partitioning of genetic variation among the individuals and geographical regions a principal component analysis (PCA) was carried out using the *ade4* package (Dray and Dufour 2007). For this, data was first scaled and centered using the *dudi.pca* function and five axes were retained (*dudi.pca* function, *ade4* package, Dray and Dufour 2007). In association with the PCA, an analysis of molecular variance (AMOVA) was performed using *Genalex* (Peakall and Smouse 2012) to assess hierarchical partitioning of the observed genetic variation displayed in the PCA to the level of individuals, populations or the six geographical regions. The

number of permutations for the AMOVA was set to 1000.

STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to calculate individual admixture coefficients (Q) across the entire dataset ( $n = 1462$ ). Analyses were based on the admixture model (individuals were allowed to have mixed ancestries from different clusters) with correlated allele frequencies (allele frequencies in different clusters were likely to be similar due to migration or shared ancestry) and the parameter of the Dirichlet distribution of allelic frequencies ( $\lambda$ ) separately inferred for each population. The optimal number of clusters (K) was inferred following Evanno et al. (2005). Since this method only detects the uppermost level of population structure when different hierarchical levels exist, we further investigated the genetic substructuring of our dataset up to  $K = 6$ . For each value of K, three iterations were run for 500,000 generations (with 100,000 generations as burn-in) and the posterior estimates of cluster memberships were visualized using the *pophelper* (Francis 2017) package implemented in R (R Core Team 2019).

## Estimation of network structure

A network approach using georeferenced populations generates an accurate visual representation of the genetic similarity between populations and partitioning of genetic variation (Dyer and Nason 2004; Guimerà and Amaral 2005; Magwene 2001) on a global scale. Since the assembly of a genetic network relies on genetic variability for the estimation of the incidence matrix, we only included samples containing a minimum of 5 specimens (Dyer and Nason 2004; Magwene 2001). This resulted in a subset comprising 95 populations and a total of 1122 specimens included for the network analyses. In population networks, populations are indicated by nodes that can be interconnected by edges or links. Some populations are only poorly connected to others and therefore several links are redundant when describing population connectivity. Based on conditional independence, it is possible to detect these redundant edges and simplify the population graph without losing information on the genetic covariance structure among populations (Dyer and Nason 2004; Fortuna et al. 2009; Magwene 2001). Edge exclusion deviance is a theoretic measure to calculate whether an edge should

be excluded from a fully saturated population network and follows several steps (Magwene 2001).

First, the multilocus Euclidean genetic distance matrix was computed as in Fortuna et al. (2009). It is a Euclidean genetic distance matrix giving more weight to rare alleles. The covariance matrix was calculated from the matrix of distances as in Dyer and Nason (2004) while the steps followed to compute the partial correlation matrix from the covariance matrix are the same as in Magwene (2001). The threshold value for edge exclusion deviance (EED) was calculated for each edge in order to omit edges that are not conditionally independent. Finally the strength ( $s_{ij}$ ) of genetic similarity represented by the retained edges was calculated as:

$$s_{ij} = -\frac{1}{2} \text{Ln} \left[ 1 - (r_{ij})^2 \right]$$

where  $r_{ij}$  is the partial covariation coefficient of the edge. The package *graph4lg* (Savary 2020) implemented in R (R Core Team 2019) was used to build the final matrix used to produce the population graph (Savary 2020). For more in depth information on the procedure to estimate the conditional independence structure of the genetic covariance see Fortuna et al. (2009).

### Modularity analysis

Here, we apply a simulated annealing approach to our population graph to explore different possible states of the population structure (Guimerà and Amaral 2005). We ran 100 replicates of the Guimera and Amaral algorithm for Newman's  $M$  and then obtained the maximum value of  $M$  (Newman 2006). Modularity  $M$  can vary between 0 for non-modular networks and 1 for networks with strong modularity, consisting of dense modules with tightly linked nodes and the absence edges between modules (Newman 2006).

The final population graphs displaying the different modules were constructed using the *igraph* package (Csardi and Nepusz 2006) in R (R Core Team 2019). The population graph is presented in two different ways. In a first layout, populations are georeferenced while a second representation of the population graph positions the nodes using the Fruchterman–Reingold force-directed algorithm (Fruchterman and Reingold 1991). While the geographical layout renders a clear view on the genetic similarity among continents, the

Fruchterman–Reingold network is unconstrained by geography and can display more tightly linked populations in closer proximity (Fruchterman and Reingold 1991).

## Results

The number of scored multilocus genotypes reached a plateau after seven sampled loci, indicating that the genetic variability of *C. capitata* was adequately sampled by the 14 microsatellites markers used in this study (Fig. S2). After scoring genotypes, our dataset contained 2.91% missing alleles.

The 126 *C. capitata* populations ( $n = 1214$ ) yielded a total number of alleles per locus ranging from 10 (medflymic92) to 34 (medflymic43) with an average of  $17.6 \pm 6.5$  alleles per locus. The total number of alleles ranged from 105 (SUD\_Sin) to 17 (GUA\_San) with an average of 45 (SD = 16) (Table S2). Expected heterozygosity ( $H_e$ ) ranged from 0.143 (GUA\_San) to 0.755 (SUD\_Sin) with an average of 0.479 (SD = 0.118), while observed heterozygosity ( $H_o$ ) ranged from 0.171 (ECU\_Tun) to 0.714 (MAL\_Lil) with an average of 0.476 (SD = 0.106) (Table S2). Sub-Saharan Africa had a higher  $N_a$ ,  $H_e$  and  $H_o$  value compared to any of the other regions ( $P \leq 0.05$ ) (Table 1). Significance of pairwise differences among regions of  $N_a$ ,  $H_e$  and  $H_o$  are summarized in Table 2. The average frequency of population-wise private alleles for each region was 0.091, 0.182, 0.107, 0.160 and 0.235 for Hawaii, Australia, the Mediterranean, South-America and Central America respectively, with the Sub-Saharan region scoring the highest with a value of 0.941 (Table S2). Region-wise private alleles were highest for Sub-Saharan Africa (74) and are summarized in Table 1.

Pearson's Chi-squared test showed significant HWE deviations in 170 of 1764 locus-population combinations, corresponding to 9.64% of observations. HW disequilibrium tends to be more present in populations with only a handful of samples, but not restricted to a certain marker only. Linkage disequilibrium was observed in 206 out of 11,466 per-population pairwise tests, corresponding to 1.8% of comparisons.

**Table 1** Average values for number of alleles ( $N_a$ ), observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ) for every region with the number of samples included for every region in the first column

Region	number of samples	$N_a$	$H_o$	$H_e$	private alleles
Sub-Saharan Africa	187	57.12	0.56	0.59	74
Mediterranean	262	44.57	0.5	0.47	6
South America	303	43.68	0.43	0.45	5
Hawaii	137	37.45	0.43	0.4	2
Australia	130	37.09	0.35	0.39	2
Central America	195	36.41	0.44	0.43	5

### Genetic structuring

Eigenvalues of the first three principal components resulting from the principal component analysis (PCA) were 8.56 (33.10% of the variance), 5.21 (20.04% of the variance) and 4.90 (18.87% of the variance) respectively with a drop to 3.94 for the eigenvalue for the fourth principal component (Fig. 1). Sub-Saharan African populations formed a well-defined cluster showing no overlap with populations belonging to other regions along the first three principal components. Australian populations form a separate cluster along the first and third principal component showing the closest proximity to the Mediterranean cluster (Fig. 1a).

Results from the AMOVA show that most of the genetic variation could be attributed to variation within the populations (68%, Table 3). Genetic differences among populations contributed 15% to the total genetic variation while genetic differences among the six regions comprised 17% of the total genetic variation (Table 3).

The STRUCTURE analysis of the 1214 specimens showed a clear  $\Delta K$  peak at  $K = 3$  ( $\Delta K = 838.46$ , Fig. S3) while average  $\Delta K$  for  $K$  values up to 6 was 197 (SD = 325) indicating that the main hierarchical level of population structure is based on three genotype groups (Fig. 2). Individual Bayesian cluster assignment for  $K = 3$  largely reflected the results of the ordinations in the multivariate space of the PCA, recognizing Sub-Saharan African populations as one genetic entity (Figs. 1 and 2).

Within Sub-Saharan Africa, the genetic structuring is markedly homogeneous with the notable exception of MAU\_Wes which exhibits extensive admixture

with populations from the Mediterranean (Fig. 2). All three other sampling locations within Mauritius (MAU\_Eas, MAU\_Nor and MAU\_Sou) do not show elevated signs of European admixture and are primarily assigned to the Sub-Saharan region (Fig. 2).

South American and Hawaiian samples are largely assigned to one genotypic cluster, distancing them from Australian, Mediterranean and Central American samples and agreeing with the PCA where South American and Hawaiian populations are located in close proximity and show a large overlap regarding their 95% inertia ellipses (Fig. 1b).

### Network of spatial genetic variation

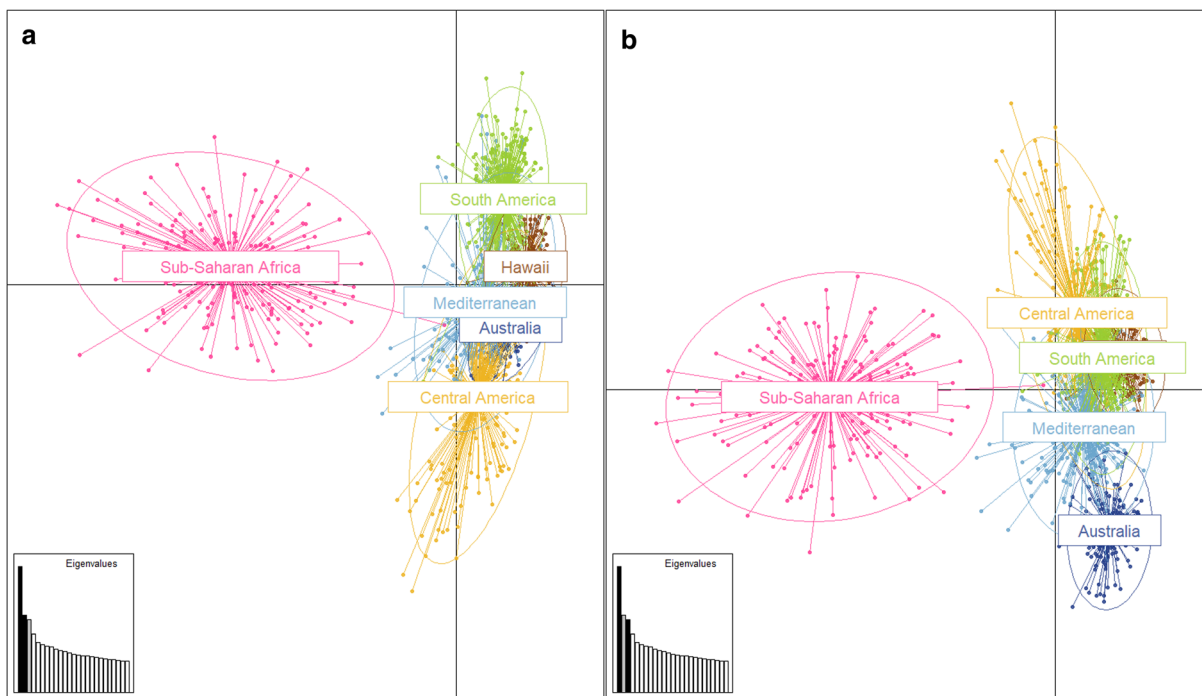
Of the 4465 possible links between nodes, only 325 were retained after applying edge deviance exclusion (Figs. 3 and 4). Node size in Fig. 3a correlates with the participation coefficient of that node, i.e., the number of links the node shares with nodes of a different module. The node size in Fig. 3b is correlated with the number of alleles within the respective population. In Fig. 4, node size correlates with the within-module degree of the node, i.e. the number of edges it shares with nodes within its own module. The total number of intraregional links was 259 (equal to 79.68% of the total number of links) while 66 links (equal to 20.32% of the total number of links) connected populations from different regions. The average number of intraregional links per population was highest for the Sub-Saharan region (SSA: 3.97, SAM: 2.76, CAM: 2.53, EUM: 2.25, AUS: 2.36, HAW: 2.00).

The analysis of modularity revealed 7 distinct modules with a total network modularity of 0.465 (Figs. 3 and 4). One module (dark blue) exclusively

**Table 2** P-values for differences in the genetic parameters between regions

$N_a$	Australia	Central America	Hawaii	Mediterranean	South America
Central America	0.343	–	–	–	–
Hawaii	0.114	0.099	–	–	–
Mediterranean	<b>0.004</b>	<b>0.004</b>	0.523	–	–
South America	<b>0.010</b>	<b>0.036</b>	0.964	0.304	–
Sub-Saharan Africa	<b>0.000</b>	<b>0.000</b>	<b>0.003</b>	<b>0.000</b>	<b>0.000</b>
$H_o$	Australia	Central America	Hawaii	Mediterranean	South America
Central America	<b>0.003</b>	–	–	–	–
Hawaii	<b>0.011</b>	0.892	–	–	–
Mediterranean	<b>0.000</b>	0.146	0.090	–	–
South America	<b>0.018</b>	0.955	0.892	0.143	–
Sub-Saharan Africa	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
$H_e$	Australia	Central America	Hawaii	Mediterranean	South America
Central America	<b>0.001</b>	–	–	–	–
Hawaii	<b>0.001</b>	0.907	–	–	–
Mediterranean	<b>0.000</b>	0.055	0.102	–	–
South America	<b>0.022</b>	0.907	0.945	<b>0.048</b>	–
Sub-Saharan Africa	<b>0.000</b>	<b>0.001</b>	<b>0.009</b>	<b>0.033</b>	<b>0.000</b>

Values smaller than 0.05 are indicated in bold

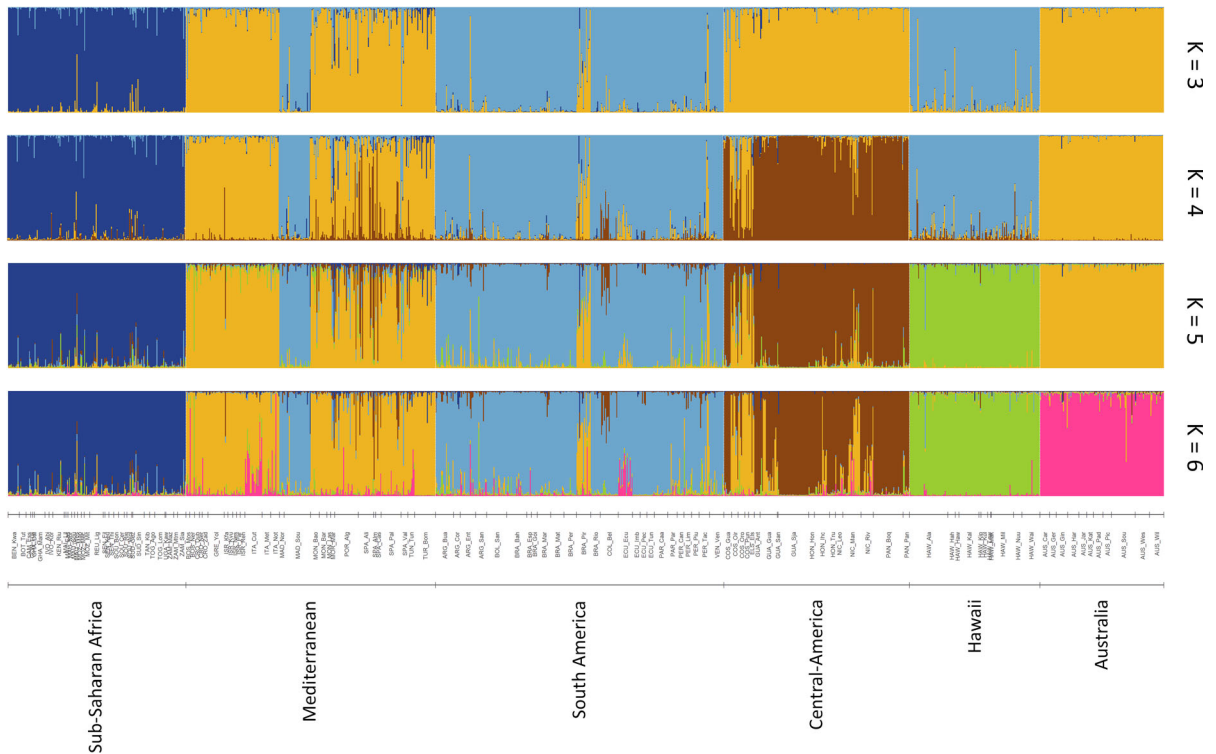


**Fig. 1** Biplot PCA with 95% inertia ellipses. **a** First and second principal component represented, **b** first and second principal component represented. The eigenvalues of the first, second and

third pca axis are 8.56 (33.10% of the variance), 5.21 (20.04% of the variance) and 4.90 (18.87% of the variance) respectively

**Table 3** Results for the AMOVA revealing the partitioning of genetic variation across different levels

Level	df	SS	Est. Var	% variation
Among regions	5	824.246	0.823	17%
Among populations	89	1014.226	0.703	15%
Within populations	1027	3381.750	3.293	68%
Total	1121	5220.223	4.819	100%

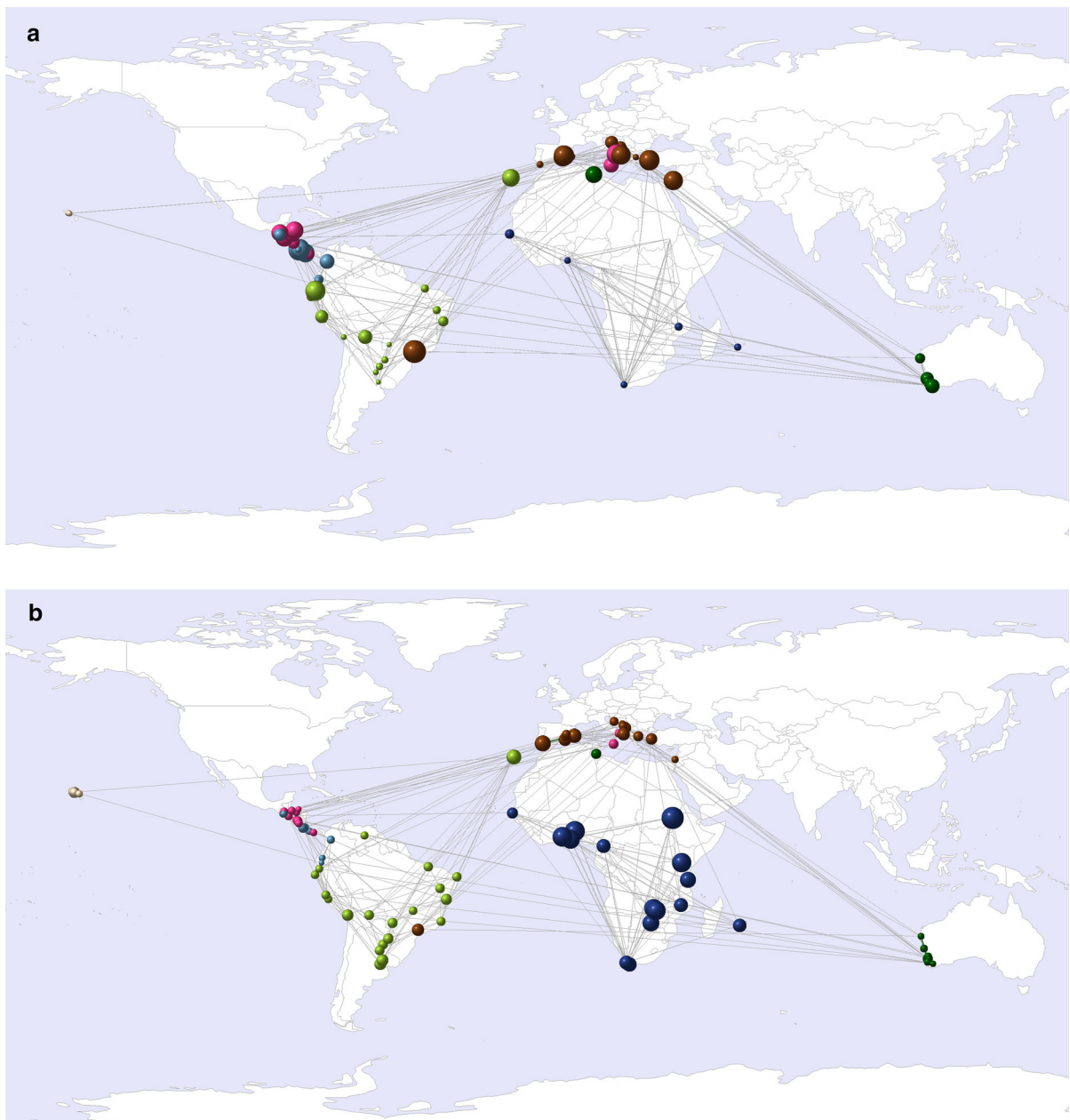
**Fig. 2** Bayesian clustering using STRUCTURE with indication of the six geographical regions for K = 3 to K = 6

comprised populations of Sub-Saharan Africa. The light green module largely coincided with populations of South-America (20 populations out of 22) while the brown module contained most of the populations within the Mediterranean region (15 populations out of 16). Central American populations are mainly confined to the pink (nine Central American populations) and light blue module (six Central American populations). Flies collected in western Australia were all assigned to one module (dark green) that was shared by one population from the Mediterranean region (TUN\_tun). The white module was entirely

represented by all six Hawaiian populations (Figs. 3 and 4).

Sub-Saharan populations are tightly connected with 67 intraregional links. Only five links are shared with populations outside of Sub-Saharan Africa. Four links are shared with the Mediterranean (all connected to TUR\_Bom) while one link is shared with El Salvador (Els\_Els).

The Mediterranean region shares only four links with the Sub-Saharan region through Turkey (TUR\_Bom) which is on its turn is connected with three other populations in the Mediterranean (Figs. 3 and 4). Of the eight links that connect to Madeira, six are shared with the South American continent.



**Fig. 3** Georeferenced *C. capitata* population networks with node size corresponding to the participation coefficient, i.e. the number of links that node shares with nodes of another module

Central America shares 12 links with both the Mediterranean region and South America, one link with Africa (SEN\_Thi) and two with Australia (AUS\_Pad and AUS\_Gin) (Figs. 3 and 4).

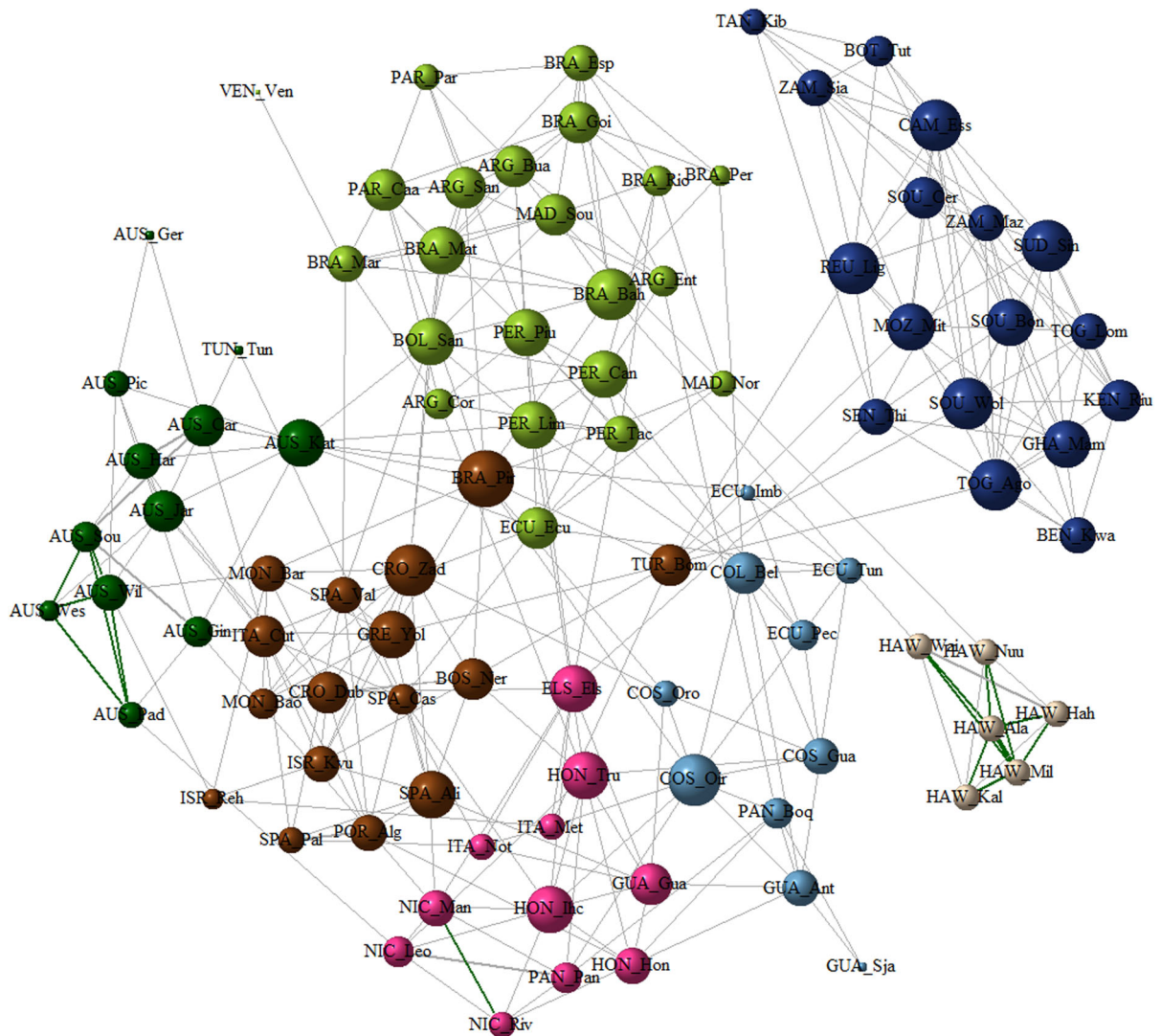
Australia shares eight links with the Mediterranean region (of which AUS\_Jar, AUS\_Har and AUS\_Car

in **a** and node size corresponding to the number of alleles in **b**. Line thickness correlates with genetic similarity. The node color represents the module assignment

show a link with ITA\_Cut), six with South America (AUS\_Kat has a link with BOL\_San, Per\_Lim, ECU\_Imb and ECU\_Ecu) and two with Central America (Figs. 3 and 4).

Hawaiian populations structure in one tight cluster consisting of 14 intraregional links of which 11 have





**Fig. 4** Fruchterman-Reingold representation of *C. capitata* the population network with node size relative to the node's within-module degree, i.e. the number of edges the node shares

an edge strength in the top 5% quantile of the distribution, indicating high genetic similarity. Only two links are shared with populations outside the Hawaiian archipelago, one with Madeira and another with South America.

## Discussion

In this study we have applied an integrative approach, combining network theory and modularity analysis with more conventional population genetic

with nodes within its own module. Line thickness correlates with genetic similarity. The node color represents the module assignment

approaches such as principal component analysis (PCA) and Bayesian clustering (Pritchard et al. 2000) to complement earlier work on the worldwide phylogeography of *Ceratitis capitata* (Bonizzoni et al. 2000; Malacrida et al. 2007; Barr 2009; Barr et al. 2014; Ruiz-Arce et al. 2020). The population networks (Figs. 3 and 4) allow us to assess genetic similarities on a more detailed intraregional scale, as opposed to the work of Ruiz-Arce et al. (2020) who focused on haplotype phylogeny only.

Our findings are in agreement with results of Ruiz-Arce et al. 2020 who suggested a range expansion of

*C. capitata* from Sub-Saharan Africa to other parts of the world. We must stress the fact that our sampling was spread over several decades and included a large number of populations from multiple sources. For this reason, in order to reduce the risk of possible biases, below we focus on the general, large-scale, regional patterns while avoiding inference at finer spatial and temporal scales.

### Sub-Saharan Africa

Our results confirm the African origin of *C. capitata* with Sub-Saharan Africa possessing the largest amount of genetic variation. This is in line with Ruiz-Arce et al. (2020) who found the largest number of mitochondrial haplotypes in this region and results of Karsten et al. (2015) who found the highest allelic richness in the African continent using 11 microsatellite markers. The presence of a large amount of private alleles within Sub-Saharan Africa highlights the high genetic diversity in this region (Fig. S2). Our findings agree with the results of Malacrida et al. (2007) who recognized Kenya is a reservoir of high genetic polymorphism. Only three African countries in our dataset (Sudan: SUD\_Sin, 105 variants; Benin: BEN\_Kwa, 95 variants and Ghana: GHA\_Mam, 93 variants) surpass Kenya in respect of allelic richness.

Sub-Saharan samples form a well-defined cluster in both the PCA (Fig. 1), the structure analysis (Fig. 2) and the analysis of modularity (Figs. 3 and 4), indicating a strong connection between populations from this region.

*C. capitata* is considered as an introduced species to Madagascar, Comoros, Mauritius and La Réunion (CABI 2021) where it was established in La Réunion by 1939 (White et al. 2000) and in Mauritius by 1942 (Seewooruthun et al. 2000). The source of these populations has not been determined but our network analysis revealed links of significant genetic similarity between Réunion and Turkey, Mozambique, South-Africa, Botswana, Zambia, Cameroon, Benin and Senegal (Figs. 3 and 4). Likely, the genotypes of Medfly populations on Mauritius and Réunion are an assemblage of several genotypes of different origins.

The marked admixture of a Mauritian population (MAU\_Wes) with the mainly European haplotype and the links between Réunion, the Mediterranean and South-Africa in the population network (Fig. 3) could be a reflection of the unique position of these islands

along the naval routes to the Far East and Cape of Good Hope (source: transportgeography.org). Similarly, the direct link between South-African populations and west Africa (Senegal, Ghana, Togo and Benin) could be attributed to the presence of naval trading routes connecting Cape of Good Hope with western Africa (Rodrigue 2020). Global travel has undoubtedly facilitated transport of pest species outside of their native range, leaving genetic signatures that can be traced by population genetic analyses (Hulme 2009; Karsten et al. 2013; Malacrida et al. 2007). It has to be noted that the small sample size for each of the Mauritian populations (< 5 individuals) might have influenced the admixture patterns we observed.

Genetic differentiation between the African mainland and islands was already observed in tephritid fruit flies such as the melon fly, *Zeugodacus cucurbitae* (Coquillett 1899), (formerly *Bactrocera cucurbitae*, see Virgilio et al. 2015), *C. rosa* (Virgilio et al. 2013) and *C. capitata* (Bonizzoni et al. 2000; Malacrida et al. 1998) suggesting that geographical isolation can cause considerable genetic divergence between mainland and islands compared to the homogenizing gene flow from the mainland to the islands (Malacrida et al. 2007). More exhaustive sampling on the islands of Réunion and Mauritius is needed to investigate the sources of genetic variation in more detail.

### Europe/Mediterranean

Medfly arrived in the Mediterranean as early as 1842 and is considered as an invasive species in large parts of temperate Europe (Malacrida et al. 2007). Europe has a strong link with the African continent through colonialism from the nineteenth century until decolonization after the world war, boosting trade among continents (Macke et al. 1977) and likely facilitating the movement of *C. capitata*.

After the Sub-Saharan African populations, the European populations exhibit the second highest average allelic richness, although it is significantly lower than in Sub-Saharan Africa (Tables 1, 2 and Table S2). This supports the migration of *C. capitata* out of Africa which was associated with loss of genetic variation by drift and changing selective pressures such as seasonality and temperature changes (Bonizzoni et al. 2000; Malacrida et al. 1992). Loss of genetic variability is often associated with biological

invasions due to founder effects (Arias et al. 2018; Lockwood et al. 2009). Our findings are in line with findings of Ruiz-Arce et al. (2020) who showed that the Mediterranean region had the second highest number of mitochondrial haplotypes after the African region. These results, however, are not from truly independent data sets because the flies included in our study were also included in the Ruiz-Arce et al. (2020) study.

It has been suggested that *C. capitata* found its way into Europe through the Iberian peninsula and subsequently spread further into Western Europe (Bonizzoni et al. 2004; Malacrida et al. 2007). Bayesian cluster analysis shows considerable admixture with the African region for all of the Iberian populations, supporting the Iberian point of entry into Europe. Interestingly, a population in Turkey (TUR\_Bom) exhibits the same signs of admixture (Fig. 2). Moreover, our network analysis confirms the link between Turkey and the African continent with which it shares five links (REU\_Lig, MOZ\_Mit, SEN\_Thi, SOU\_Wol and TOG\_Ago). This finding suggests an eastern Mediterranean entry point that might be considered complementary to the Iberian route hypothesized in Bonizzoni et al. (2004) and Malacrida et al. (2007). Additional data and focused hypothesis testing are now needed to provide more details about the intriguing invasion history of the Mediterranean.

Remarkably, the extensive admixture between South American and Madeiran populations points towards a secondary introduction of *C. capitata* from South America into Madeira. As shown in Ruiz-Arce et al. (2020), haplotype M065 is shared between South American, Hawaiian and Madeiran populations, corroborating our results.

### South America

First records of *C. capitata* in the New World data back to 1905, Brazil (Malacrida et al. 1998). The South American samples form a relatively uniform group in the population graph with the exception of some populations in the north of the continent (ECU\_Imb, ECU\_Pec, ECU\_Tun and COL\_Bel) that share the same module with flies from Costa Rica and Panama (Figs. 3 and 4). Especially COL\_Bel in Colombia exhibits extensive admixture with Central American genotypes (Fig. 2) and shows links with Costa Rica and Guatemala (Figs. 3 and 4). The

population graph (Figs. 3 and 4) shows 12 links between Central-America and northern South America, indicating considerable gene flow between these regions. Similarly, Ruiz-Arce et al. (2020) did not recover the common M001 (81 individuals) and the closely related but less common M040 (two individuals) and M041 (one individual) haplotypes of South America in the Northern part of South America (Ecuador, Venezuela and Colombia), further pointing towards a modest barrier for gene exchange between the north and the rest of the South American continent (See Table S2 in Ruiz-Arce et al. 2020).

Interestingly, a population in Brazil near São Paulo (BRA\_Pir) seems to function as a connector hub between Eastern European and South American populations. The direct influx of genetic variants belonging to the European cluster could explain why BRA\_Pir contains the largest number of alleles (57) of all South American populations (average = 44, SD = 8.03) (Table 2).

### Central America

Medfly was first discovered in Central America in 1955 in Costa Rica (Gallo et al. 1970). Central American samples are genetically closest to the European/Mediterranean samples which is reflected by the large number of links between the two regions (Figs. 3 and 4). Additionally, the large proportion of admixture with the Mediterranean genotype when investigating admixture plots points to an association between Central America and the European continent (Fig. 2). Ruiz-Arce et al. (2020) also suggested introduction of Mediterranean haplotypes into Central America. In their study, Central American samples predominantly showed the M007 haplotype which is only shared with Mediterranean populations and one individual from Senegal. Two of the private haplotypes (M005 and M008) for the region only differed by two and three bases respectively from the M007 haplotype and likely arose within the region (Ruiz-Arce et al. 2020). Another study by Nikolouli (2020) corroborates the link between the European continent and Central America and shows clear clustering of four European populations with four Central American populations, based on eight microsatellite markers. Additionally, a study by Karsten et al. (2015) showed that Guatemala was likely colonized from Greece. Although only one Central-American and one

European mainland population was used in Karsten et al. (2015), this finding supports the European origin of Central-American flies.

Although Central America is small in size, the presence of two different modules that are almost equally represented suggests the presence of genetic structuring beyond the regional level. This is in contrast with findings of Ruiz-Arce et al. (2020) who found no significant structuring beyond the regional level using mitochondrial DNA. More exhaustive sampling in Central America could reveal more details concerning the partitioning of genetic diversity in that area.

### Hawaii

Medfly first colonized Western Hawaii around 1895 (Bonizzoni et al. 2004; Dominiak and Daniels 2012). Bayesian clustering clearly indicates a South American origin for the Hawaiian samples, largely agreeing with the results of Ruiz-Arce et al. (2020) who showed that one haplotype (M065) was primarily shared by Hawaiian and South American populations. Furthermore, results by Nikolouli (2020) suggest a high genetic similarity between populations from Hawaii and South America, applying a similar Bayesian clustering analysis.

Hawaiian populations form a tightly linked cluster with a central hub (HAW\_Ala) (Fig. 4). HAW\_Ala is situated in close proximity (9 km) to the international airport of O'ahu island which might suggest that HAW\_Ala served as the first point of entry of *C. capitata* into the Hawaiian archipelago from which it has spread to the other Hawaiian Islands. Further investigations in respect of fruit transport across the archipelago is needed to unravel the invasion history and further assess the special status of HAW\_Ala.

### Australia

Medfly has been accidentally introduced into Australia at several times in history. The first record of medfly in Australia dates back to 1895 in the area of Perth (Sproule et al. 2001; Dominiak and Daniels 2012). Since then, sporadic medfly outbreaks have occurred between 1952 until 1993 after which medfly has been captured continuously (Bonizzoni et al. 2004). Medfly has expanded its range northwards from Perth, where horticultural crops and the subtropical

climate provide an ideal environment for medfly expansion (Sproule et al. 2001; Bonizzoni et al. 2004).

Our results suggest a secondary colonization event from Mediterranean origin into Australia, in line with the results of Bonizzoni et al. (2004) and Karsten et al. (2015). Especially connections with Central to Eastern Mediterranean populations (Croatia, Montenegro, Greece and Israel) and Tunisia are well represented in the network graphs (Figs. 3 and 4). Furthermore, Bayesian clustering provides a strong indication for a Mediterranean origin of the Australian samples (Fig. 2). Arias et al. (2018) who did not include South American flies, revealed a similar result with most of the Australian samples used in the study sharing the large Cc\_21 haplotype with several countries bordering the Mediterranean basin. Conversely, in the study of Ruiz-Arce et al. (2020), the predominant mitochondrial haplotype (M009) in Australian specimens is shared to a large extent with South-American and only to a lesser extent with Mediterranean flies. Further investigation is needed to assess the impact of direct gene exchange on Australian population dynamics. The links shared by South America and Australia (AUS\_Kat) might be either related to the genetic similarity between a subset of the South American countries (Brazil, Peru and Ecuador) and the Mediterranean samples or might indicate direct gene flow between Australia and South America. The significantly lower  $H_e$  and  $H_o$  compared to all other regions and relatively low  $N_a$  (37.09) suggest that introductions of *C. capitata* in Australia mainly stem from genetically similar regions (Tables 1 and 2).

In summary, presented here is a description of the worldwide population structuring and range expansion of *C. capitata*. We recovered significant genetic groups using three different approaches, revealing a high association between genetic structuring and geography. We hypothesize an African origin for Mediterranean medfly populations from where flies have been introduced in other regions (Australia and Latin America). We uncovered a clear association between the Mediterranean and Central American populations, suggesting a range expansion into Central America directly from the Palearctic rather than colonization of Central America by South American flies. Additionally, we revealed genetic divergence between northern and southern Latin American populations. We caution the reader to consider that while diversity can support colonization time estimates, it is

best interpreted together with other information such as historical reports. This work, together with the contribution by Nikolouli (2020), who used eight microsatellite loci, the work by Karsten et al. (2015) and the more recent study by Ruiz-Arce et al. (2020) who have both used mitochondrial DNA sequencing methods, render the most current view on medfly genetic diversity, geographic genetic structure and invasion pathways. Although we consider the current dataset as an informative instrument for investigating medfly invasion history, we note that it will benefit from resampling of populations and new collections to stay relevant in the future. This is certainly the case for populations that are only represented by a handful of samples in this study. In 2016, a high-quality genome assembly for *C. capitata* has been achieved which could serve as a critical tool for surveying the medfly genome for the invasion potential of different populations and in developing more informative invasion pathway estimation tools.

**Authors' contributions** Drafting of the manuscript: all authors. Acquisition of data: D. T. N. Analysis of data: D. P.

**Funding** Funding for this project was provided by the EU-FFIPM project (Grant ID: 818184).

**Availability of data and material** Collection locations and microsatellite genotypes are available at: <https://doi.org/10.5281/zenodo.3967065>.

**Code availability** R packages and software used are clearly indicated throughout the manuscript.

## Declarations

**Conflict of interest** The authors declare there are no competing interests. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

**Consent for publication** All authors agreed to this publication.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adamack AT, Gruber B (2014) PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol*. <https://doi.org/10.1111/2041-210X.12158>
- Arias MB, Elfekih S, Vogler AP (2018) Population genetics and migration pathways of the Mediterranean fruit fly *Ceratitis capitata* inferred with coalescent methods. *PeerJ*. <https://doi.org/10.7717/peerj.5340>
- Barr NB (2009) Pathway analysis of *Ceratitis capitata* (Diptera: Tephritidae) using mitochondrial DNA. *J Econ Entomol*. <https://doi.org/10.1603/029.102.0153>
- Barr N, Ruiz-Arce R, Armstrong K (2014) Using molecules to identify the source of fruit fly invasions. Trapping and the detection, control, and regulation of Tephritid fruit flies. *Evol Appl*. [https://doi.org/10.1007/978-94-017-9193-9\\_10](https://doi.org/10.1007/978-94-017-9193-9_10)
- Bonizzoni M, Guglielmo CR, Smallridge CJ, Gomulski M, Malacrida AR, Gasperi G (2004) On the origins of medfly invasion and expansion in Australia. *Mol Ecol*. <https://doi.org/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 Mol Biol*. <https://doi.org/10.1046/j.1365-2583.2000.00184.x>
- Bonizzoni M, Zheng L, Guglielmino CR, Haymer DS, Gasperi G, Gomulski LM, Malacrida AR (2001) Microsatellite analysis of medfly bioinvasions in California. *Mol Ecol*. <https://doi.org/10.1046/j.0962-1083.2001.01376.x>
- CABI (2021) *Ceratitis capitata*. Invasive Species Compendium. <http://www.cabi.org/isc>
- Carey JR (1991) Establishment of the Mediterranean fruit fly in California. *Science*. <https://doi.org/10.1126/science.1896848>
- Carey JR, Papadopoulos N, Plant R (2017) The 30-year debate on a multi-billion-dollar threat: Tephritid fruit fly establishment in California. *Am Entomol*. <https://doi.org/10.1093/ae/tmx036>
- Coquillett DW (1899) A new Trypetid from Hawaii. *Entomol News* 10:129–130
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *IntJ Complex Syst* 1695:1–9
- De Meyer M, Copeland RS, Wharton RA, McPherson BA, Barnes BN (2004) On the geographic origin of the Medfly *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae). In: Proceedings of the 6th international symposium on fruit flies of economic importance. Stellenbosch, South Africa. [https://nucleus.iaea.org/sites/naipc/twd/Documents/6thISFFEI\\_Proceedings/DE%20MEYER.pdf](https://nucleus.iaea.org/sites/naipc/twd/Documents/6thISFFEI_Proceedings/DE%20MEYER.pdf)
- Dominiak BC, Daniels D (2012) Review of the past and present distribution of Mediterranean fruit fly (*Ceratitis capitata*, Wiedemann) and Queensland fruit fly (*Bactrocera tryoni*,

- Froggatt) in Australia. *Aust J Entomol.* <https://doi.org/10.1111/j.1440-6055.2011.00842.x>
- Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw.* <https://doi.org/10.18637/jss.v022.i04>
- Dyer RJ, Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. *Mol Ecol.* <https://doi.org/10.1111/j.1365-294X.2004.02177.x>
- Enkerlin W, Gutiérrez-Ruela JM, Cortes AV, Roldan EC, Midgard D, Lira E, Arriaga FJT (2015) Area freedom in Mexico from Mediterranean fruit fly (Diptera: Tephritidae): a review of over 30 years of a successful containment program using an integrated area-wide SIT approach. *Fla Entomol.* <https://doi.org/10.1653/024.098.0242>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Fortuna MA, Albaladejo RG, Fernández L, Aparicio A, Bascompte J (2009) Networks of spatial genetic variation across species. *PNAS.* <https://doi.org/10.1073/pnas.0907704106>
- Francis RM (2017) pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Res.* <https://doi.org/10.1111/1755-0998.12509>
- Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. *Softw Pract Exper.* <https://doi.org/10.1002/spe.4380211102>
- Gallo DNO, Wiendel FM, Neto SS, Ricardo PLC (1970) Manual de entomologia. Agronomica Ceres, San Paolo
- Guimerà R, Amaral LAN (2005) Cartography of complex networks: modules and universal roles. *JSTAT.* <https://doi.org/10.1088/1742-5468/2005/02/P02001>
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J App Ecol.* <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Jombart T (2008) ADEGENET: an R package for the multivariate analysis of genetic markers. *J Bioinform.* <https://doi.org/10.1093/bioinformatics/btn129>
- Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ.* <https://doi.org/10.7717/peerj.281>
- 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.* <https://doi.org/10.1371/journal.pone.0054281>
- Karsten M, van Vuuren BJ, Addison P, Terblanche JS (2015) Deconstructing intercontinental invasion pathway hypothesis of the Mediterranean fruit fly (*Ceratitis capitata*) using a Bayesian inference approach: are port interceptions and quarantine protocols successfully preventing new invasions? *Divers Distrib* 21:813–825
- Klassen W, Curtis CF (2005) History of the sterile insect technique. Sterile insect technique: principles and practice in area-wide integrated pest management. CRC Press, Victoria
- Liquido NJ, Cunningham RT, Nakagawa S (1990) Host plants of Mediterranean fruit fly (Diptera: Tephritidae) on the island of Hawaii (1949–1985 Survey). *J Econ Entomol.* <https://doi.org/10.1093/jee/83.5.1863>
- Liquido NJ, McQuate GT, Suite KA (2013) An encyclopedic bibliography of the host plants of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann). GPDD. <https://doi.org/10.1002/ps.418>
- 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. *Divers Distrib.* <https://doi.org/10.1111/j.1472-4642.2009.00594.x>
- Macke A, Duignan P, Gann LH (1977) Colonialism in Africa, 1870–1960, Vol I, The history and politics of colonialism. Cambridge University Press, Cambridge
- Magwene PM (2001) New tools for studying integration and modularity. *Evol.* <https://doi.org/10.1111/j.0014-3820.2001.tb00823.x>
- Malacrida AR, Gomulski LM, Bonizzoni M, Bertin S, Gasperi G, Guglielmino CR (2007) Globalization and fruitfly invasion and expansion: the medfly paradigm. *Genetica.* <https://doi.org/10.1007/s10709-006-9117-2>
- Malacrida AR, Marinoni F, Torti C, Gomulski LM, Sebastiani F, Bonvicini C, Guglielmino CR (1998) Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. *J Hered.* <https://doi.org/10.1093/jhered/89.6.501>
- Malacrida AR, Guglielmino CR, Gasperi G, Baruffi L, Milani R (1992) Spatial and temporal differentiation in colonizing populations of *Ceratitis capitata*. *Heredity.* <https://doi.org/10.1038/hdy.1992.102>
- Newman MEJ (2006) Modularity and community structure in networks. *PNAS.* <https://doi.org/10.1073/pnas.0601602103>
- Nikolouli K (2020) Genetic structure and symbiotic profile of worldwide natural populations of the Mediterranean fruit fly, *Ceratitis capitata*. *BMC Genet.* <https://doi.org/10.1186/s12863-020-00946-z>
- Nyamukondiwa C, Terblanche JS (2010) Within-generation variation of critical thermal limits in adult Mediterranean and Natal fruit flies *Ceratitis capitata* and *Ceratitis rosa*: thermal history affects short-term responses to temperature. *Phys Entomol.* <https://doi.org/10.1111/j.1365-3032.2010.00736.x>
- Paini DR, Sheppard AW, Cook DC, De Barro PJ, Worner SP, Thomas MB (2016) Global threat to agriculture from invasive species. *PNAS.* <https://doi.org/10.1073/pnas.1602205113>
- Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K et al (2016) The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol.* <https://doi.org/10.1186/s13059-016-1049-2>
- Peakall R, Smouse PE (2012) GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *J Bioinform.* <https://doi.org/10.1093/bioinformatics/bts460>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Prokopy R, Papaj D, Díaz-Fleischer F, Aluja M, Norrbom A (1999) Evolution of fruit fly oviposition behavior. *Ecol Evol.* <https://doi.org/10.1201/9781420074468.ch30>
- Rodrigue JP (2020) The geography of transport systems, 5th edn. Routledge, New York

- Rousset F, Lopez L, Belkhir K (2017) Package “genepop.” <https://cran.r-project.org/web/packages/genepop/genepop.pdf>
- Ruiz-Arce R, Todd TN, Deleon R, Barr NB, Virgilio M, De Meyer M, McPherson BA (2020) Worldwide phylogeography of *Ceratitidis capitata* (Diptera: Tephritidae) using mitochondrial DNA. *J Econ Entomol.* <https://doi.org/10.1093/jee/toaa024>
- Sardain A, Sardain E, Leung B (2019) Global forecasts of shipping traffic and biological invasions to 2050. *Nat Sustain.* <https://doi.org/10.1038/s41893-019-0245-y>
- Savary P (2020) Graph4lg: build graphs for landscape genetics analysis. <https://cran.r-project.org/web/packages/graph4lg/index.html>
- Seewooruthun SI, Shradanand P, Bhubaneswar G, Soonnoo AR, Malini A (2000) Eradication of an exotic fruit fly from Mauritius. [https://inis.iaea.org/search/search.aspx?orig\\_q=RN:36045167](https://inis.iaea.org/search/search.aspx?orig_q=RN:36045167)
- Sroule A, Broughton S, De Lima F, Hardie D, Monzu N, Woods B (2001) The fight against fruit flies in western Australia. Department of Agriculture Western Australia, Perth
- Stratikopoulos EE, Augustinos AA, Pavlopoulos ID, Economou KP, Mintzas A, Mathiopoulos KD, Zacharopoulou A (2009) Isolation and characterization of microsatellite markers from the Mediterranean fruit fly, *Ceratitidis capitata*: cross-species amplification in other Tephritidae species reveals a varying degree of transferability. *Mol Genet Genom.* <https://doi.org/10.1007/s00438-009-0465-3>
- Szyniszewska AM, Tatam AJ (2014) Global assessment of seasonal potential distribution of Mediterranean fruit fly, *Ceratitidis capitata* (Diptera: Tephritidae). *PLoS One.* <https://doi.org/10.1371/journal.pone.0111582>
- The R Core Team (2019) R: A Language and environment for statistical computing. <https://www.R-project.org/>
- Todd TN, Rendon P, Ruiz-Arce R (2017) Inheritance of fifteen microsatellite loci in *Ceratitidis capitata* (Diptera: Tephritidae). *Fla Entomol.* <https://doi.org/10.1653/024.100.0113>
- Virgilio M, Delatte H, Quilici S, Backeljau T, De Meyer M (2013) Cryptic diversity and gene flow among three African agricultural pests: *Ceratitidis rosa*, *Ceratitidis fasciventris* and *Ceratitidis anonae* (Diptera, Tephritidae). *Mol Ecol.* <https://doi.org/10.1111/mec.12278>
- Virgilio M, Jordaens K, Verwimp C, White IM, De Meyer M (2015) Higher phylogeny of frugivorous flies (Diptera, Tephritidae, Dacini): localised partition conflicts and a novel generic classification. *Mol Phylogenet Evol.* <https://doi.org/10.1016/j.ympev.2015.01.007>
- White IM, De Meyer M, Stonehouse JM (2000). A review of native and introduced fruit flies (Diptera, Tephritidae) in the Indian Ocean islands of Mauritius, Réunion and Seychelles. In Price NS, Seewooruthun SI (eds) Proceedings of the Indian Ocean Commission Regional Fruit Fly Symposium, Mauritius, pp. 15–21
- World Trade Organisation (2019) World trade statistical review 2019. [https://www.wto.org/english/res\\_e/statis\\_e/wts2019\\_e/wts2019\\_e.pdf](https://www.wto.org/english/res_e/statis_e/wts2019_e/wts2019_e.pdf)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.