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?

Minette Karsten\textsuperscript{1}, Bettine Jansen van Vuuren\textsuperscript{2}, Pia Addison\textsuperscript{1} and John S. Terblanche\textsuperscript{3}\textsuperscript{*}

\textsuperscript{1}Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa, \textsuperscript{2}Centre for Invasion Biology, Department of Zoology, University of Johannesburg, Johannesburg, South Africa, \textsuperscript{3}Centre for Invasion Biology, Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa

\textbf{ABSTRACT}

\textbf{Aim} Knowledge of how effective interceptions and quarantine measures are in preventing new biological invasions is of paramount importance for maintaining ecosystem function in a rapidly changing world. Here, we determine current macrogeographic population structure and routes of invasion of the Mediterranean fruit fly (Ceratitis capitata) using genetic approaches and reconstruct and test invasion pathway hypotheses in a Bayesian framework.

\textbf{Location} Africa, Australia, Greece, Guatemala and Madeira.

\textbf{Methods} We sampled 323 \textit{C. capitata} individuals from 14 locations worldwide and genotyped all individuals for 11 polymorphic microsatellite markers. We calculated measures of genetic diversity and determined population structure. Moreover, we reconstructed and tested eighteen invasion pathway scenarios in a Bayesian framework using ABC modelling.

\textbf{Results} We show a decrease in genetic diversity outside the native range (Africa) into the introduced range (Australia, Greece, Guatemala and Madeira). The most likely invasion pathway scenario closely matched the historical records, with an initial colonization of Europe from Africa and a secondary colonization of Australia from Europe. Moreover, we show an introduction from Greece to the Americas and, finally, a back introduction into South Africa from Europe.

\textbf{Main conclusions} Given the lack of new introductions into colonized (non-African) locations despite increasing trade, and apart from the initial invasion and establishment of the species outside of Africa, we conclude that quarantine and interception measures have been largely successful to date.

\textbf{Keywords} Agricultural pest, biological invasions, fruit flies, medfly, microsatellites.

\textbf{INTRODUCTION}

With the increase in human population placing rising pressure on food security, together with the predicted effects of climate change on the productivity of agricultural areas (e.g. Godfray \textit{et al.}, 2010; Bebber \textit{et al.}, 2013), research on the biotic and abiotic factors influencing the likelihood, frequency and potential impacts of invasions is urgently required (Blackburn \textit{et al.}, 2011; Kirk \textit{et al.}, 2013). Invertebrate pests move around the world via different human-mediated pathways (reviewed in Hulme, 2009) and after establishment expand their range through natural dispersal...
in the newly colonized region. These human-mediated pathways are closely linked to trade of commodities; transport by air and sea as well as subsequent movement on land by, for example, rail, road and canals being primary movement routes (Hulme, 2009; see also Gaston et al., 2003).

The importance of understanding invasion pathways is made clear in the multiple practical implications thereof, especially surrounding the control of new invasions (i.e. risk management) and their further prevention (Wilson et al., 2009; Estoup & Guillemaud, 2010; Blackburn et al., 2011). For example, reconstruction of invasion pathways can assist in uncovering the native range which can, in turn, provide possibilities for biological control as natural enemies are typically sourced from the native range (Estoup & Guillemaud, 2010). Knowledge of invasion pathways can also be used to assess the effectiveness of current or previously implemented management strategies (e.g. quarantine and/or port interceptions) by assessing the extent of ongoing introductions of invasive species and how this affects genetic diversity of the introduced, newly established population.

After the initial introduction to the new location, the species avoiding or overcoming the geographic barrier faces many additional challenges before successful establishment and further spread can be achieved (Blackburn et al., 2011). Broadly, propagule pressure [number and frequency of individuals introduced at a specific location over time (Lockwood et al., 2009)] is considered the most important driver of successful invasions and not necessarily species-level characteristics (Lockwood et al., 2005; Simberloff, 2009; Blackburn et al., 2013). However, Szücs et al. (2014) showed that although propagule pressure is important in the initial establishment stage, genetic processes are important for the subsequent spread and growth of the population. Predicting which species will become invasive has been an urgent and long-standing priority in invasion biology (e.g. Kimberling, 2004; Van Kleunen et al., 2010; Higgins & Richardson, 2014) together with improved predictive models (Kirk et al., 2013). The number of possible biological input parameters that can be considered in predictive modelling is extensive, but of primary importance are traits characterizing environmental niches, such as physiological parameters (e.g. thermal development requirements or activity limits) and dispersal capacity and/or distance travelled by natural or assisted means (e.g. Berthouly-Salazar et al., 2013). Furthermore, traits of population demographics, which may be linked to environmental niches (e.g. Dixon et al., 2009), are also significant, as rapid population growth rates can assist in establishment and evolutionary adaptation post-introduction (Gilchrist et al., 2008; Rey et al., 2012).

Arthropods (including the Insecta) are generally considered to be understudied in invasion ecology, but especially so on the African continent (Pysek et al., 2008). Some of the most successful invaders and agricultural pest insects worldwide are members of the Tephritidae (for example Bactrocera dorsalis, Aketarawong et al., 2007; Bactrocera invadens (Bactrocera dorsalis), De Meyer et al., 2010), commonly referred to as the ‘true fruit flies’, and comprising more than 5000 species (White & Elson-Harris, 1994). These fruit flies cause high levels of economic losses due to both direct damage of feeding larvae in fruit, as well as indirectly by placing restrictions on the export of fruit to certain trade-partner countries. Regardless of ongoing surveillance and rigorous quarantine measures to ensure fruit and vegetable consignments are pest free, infested consignments are still intercepted at port inspections and a likely source of new invasions (Papadopoulos, 2014). These new introductions can remain undetected for some time (Carey, 2010) and once established can spread rapidly, as in, for example, the recent spread of B. invadens (B. dorsalis) across the African continent (De Meyer et al., 2010). Therefore, dispersal and movement of fruit flies by either natural or human-assisted means are receiving increased attention to determine both the native range and movement patterns of species and to identify sources of new introductions. This information feeds specifically into the development of risk assessments and management of phytosanitary programmes (Leung et al., 2002).

Possibly the most notorious of the Tephritidae flies is Ceratitis capitata (Wiedemann), the Mediterranean fruit fly. Ceratitis capitata is highly invasive world-wide, likely because of its broad larval host range, tolerance of a wide range of climatic conditions and its relatively high dispersal capacity (Lance & McClintin, 2005; Nyamukondiwa et al., 2010). Ceratitis capitata is of quarantine importance world-wide (Reyes & Ochando, 2004; De Meyer et al., 2008) and has spread to many countries outside of its native range, which is believed to be Afrotropical (De Meyer et al., 2002). Although Kenya (East Africa) is the likely native range (Malacrida et al., 2007), there still remains some uncertainty as to the extent of the native range within Africa, as the levels of genetic diversity for South Africa (Karsten et al., 2013) are similar to those estimated for Kenya and do not follow the decline in genetic variability seen in other derived populations world-wide (e.g. Bonizzoni et al., 2004; Malacrida et al., 2007). Although population genetics methods have been used previously to investigate the colonization process, routes of invasion and the underlying evolutionary forces that have shaped the population structure of tephritids (e.g. Bonizzoni et al., 2001, 2004; Gilchrist & Meats, 2009; Aketarawong et al., 2014), several pressing invasion biology questions remain.

Global studies using genetic markers have shown that C. capitata populations can be subdivided into three distinct groups: first, the ancestral population from sub-Saharan Africa; second, a Mediterranean Basin group; and finally, a group from Latin America and the Pacific (Gasperi et al., 2002). The colonization history of C. capitata world-wide is relatively well documented (Gasperi et al., 2002; Malacrida et al., 2007). The proposed invasion pathway and historical chronological order of colonization events include, firstly an introduction of C. capitata to the Mediterranean coast (possibly on the coast of Spain) (De Breme, 1842; Fimiani, 1989).
likely human-mediated (Maddison & Bartlett, 1989) from Africa, after which the rest of the Mediterranean area was colonized (Fimiani, 1989). This initial colonization and spread throughout the Mediterranean area was confirmed using Slatkin’s private allele approach (Gasperi et al., 2002). This is thought to have been followed by a secondary introduction to Australia from Europe in the 1890s (Hooper & Drew, 1989), which was confirmed using microsatellites (Bonizzoni et al., 2004). The next leg of introductions in *C. capitata*’s global colonization was into the Americas (Costa Rica) (Harris, 1989). The source for the colonization of Guatemala remains unclear, although some evidence exist to show that it might have been a southwards spread from Costa Rica after which Mexico was colonized with the flies moving along the coffee belt (Harris, 1989). Given the nature of these records and observations, these colonization dates all likely reflect the earliest record of *C. capitata* in the location, but not necessarily the date of introduction. This incongruity can be due to fruit flies establishing in an area long before they are detected in routine monitoring. Previously, support for proposed routes of invasion has come from genetic diversity estimates, where the area with the highest level of genetic diversity would be assumed to be part of the native range (Malacrida et al., 2007), or from gene flow estimates based on Slatkin’s private allele method (Malacrida et al., 1998). However, these previous estimates are potentially weakened by the reduction of genetic data to a single summary statistic of genetic diversity, which does not necessarily account for stochastic population events (Estoup & Guillemaud, 2010).

Here, we make use of microsatellites, genetic diversity estimates and individual fly sampling at broad geographic scales to investigate the macrogeographic (intercontinental) population structure with a strong focus on sub-Saharan Africa and whether this can inform us of the origins, pathways of invasion and then subsequent success of prevention or intervention strategies for further establishment of *C. capitata* in those countries. We hypothesize that, based on previous studies of global *C. capitata* structure (Malacrida et al., 1998, 2007; Gasperi et al., 2002), there will be a clear decrease in global genetic diversity moving away from East Africa, as well as strong genetic differentiation between native populations in Africa and derived populations elsewhere in the world, possibly due to founder effects and genetic drift.

Moreover, we investigate eighteen distinct but plausible scenarios for the invasion pathway of *C. capitata* using approximate Bayesian computations (ABC). All scenarios considered are based on an ‘out-of-Africa’ approach including information from *C. capitata* macrogeographic studies and historical information of the colonization. The different scenarios were centred on an introduction from Africa (all locations excluding Burgers Hall) to the Mediterranean region [Unsampled location (possibly Spain)] after which the range expanded to Madeira and Greece, with a secondary introduction from the Mediterranean region to Australia. In each scenario, we also test from which location Guatemala was colonized as this information is unavailable in the literature together with testing from where Burgers Hall (South Africa) was recolonized.

**METHODS**

### Sampling and microsatellite genotyping

We sampled 323 *C. capitata* individuals from 14 populations across four biogeographic regions (Afrotropical, Neotropical, Palaeartic and Australasian), representing all major regions from where *C. capitata* are known to occur except the Indian Ocean Islands, the Middle East, South America as well as North America (only with sporadic outbreaks), with more extensive sampling focused on sub-Saharan Africa (Table 1). Flies were collected via trapping in orchards using baits as well as reared from infested fruit. We refer to individuals from the same sampling location (several traps across a few hundred metres) as a population in the rest of this study. After morphological confirmation of the identification of *C. capitata*, DNA was extracted using a DNeasy® Tissue Kit (Qiagen Inc., Hilden, Germany). All individuals collected were genotyped for 11 microsatellite markers following Karsten et al. (2013). In each plate, we included a negative and a positive control to check that plates were read consistently between different runs. Samples were genotyped using an ABI 3130 Automated Sequencer (Applied Biosystems, Foster City, CA, USA). Samples that remained unamplified after two independent reactions were assumed non-amplifiable and therefore not included. Four of the South African sites were part of the Karsten et al. (2013) study. Only a small subset was genotyped again, and all individuals were scored again with the additional samples included.

### Microsatellite analysis

#### Estimates of sample variability

Alleles were scored in Genemapper v3.7 (Applied Biosystems). All microsatellite markers were tested for linkage disequilibrium and deviations from Hardy–Weinberg equilibrium (HWE) using 10,000 permutations in GenePop v4.01 (Raymond & Rousset, 1995; Rousset, 2008). In all cases where multiple testing were performed, we adjusted significance levels using false discovery rates (FDR; QVALUE, Storey, 2002), and these are the values reported. Genetic diversity levels were investigated calculating basic statistics for all populations including expected heterozygosity (*H_e*), observed heterozygosity (*H_o*), number of alleles (*N_a*), number of private alleles (*N_p*), as well as the inbreeding coefficient (*F_is*) (Genetix v4.05.2, Belkhir et al., 1996–2004;
Table 1: The locations of collection sites of *Ceratitis capitata* used in this study as well as sample size (N), number of alleles (N<sub>A</sub>), number of private alleles (N<sub>P</sub>), expected (H<sub>E</sub>) and observed (H<sub>O</sub>) heterozygosity (± standard error), allelic richness (A<sub>R</sub>), the inbreeding coefficient (F<sub>IS</sub>) and the mean null allele frequency (A<sub>n</sub>; Dempster et al., 1977; SD in parentheses).

<table>
<thead>
<tr>
<th>Country</th>
<th>ID</th>
<th>Latitude</th>
<th>Longitude</th>
<th>N</th>
<th>N&lt;sub&gt;A&lt;/sub&gt;</th>
<th>N&lt;sub&gt;P&lt;/sub&gt;</th>
<th>H&lt;sub&gt;E&lt;/sub&gt;</th>
<th>H&lt;sub&gt;O&lt;/sub&gt;</th>
<th>A&lt;sub&gt;R&lt;/sub&gt;</th>
<th>F&lt;sub&gt;IS&lt;/sub&gt;</th>
<th>A&lt;sub&gt;n&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>KEN</td>
<td>−1.29</td>
<td>36.83</td>
<td>27</td>
<td>10.727</td>
<td>0.636</td>
<td>0.796 ± 0.149</td>
<td>0.533 ± 0.143</td>
<td>5.657</td>
<td>0.348</td>
<td>0.139 (0.087)</td>
</tr>
<tr>
<td>Mozambique</td>
<td>MOZ</td>
<td>−19.13</td>
<td>33.43</td>
<td>22</td>
<td>9.818</td>
<td>0.364</td>
<td>0.809 ± 0.113</td>
<td>0.606 ± 0.201</td>
<td>5.698</td>
<td>0.276</td>
<td>0.115 (0.110)</td>
</tr>
<tr>
<td>Senegal</td>
<td>SEN</td>
<td>14.18</td>
<td>−16.56</td>
<td>9</td>
<td>3.909</td>
<td>0</td>
<td>0.588 ± 0.136</td>
<td>0.511 ± 0.247</td>
<td>5.569</td>
<td>0.194</td>
<td>0.061 (0.107)</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burgers Hall</td>
<td>BUR</td>
<td>−25.06</td>
<td>31.05</td>
<td>10</td>
<td>3.273</td>
<td>0.273</td>
<td>0.431 ± 0.224</td>
<td>0.404 ± 0.173</td>
<td>2.877</td>
<td>0.134</td>
<td>0.023 (0.053)</td>
</tr>
<tr>
<td>Levubu</td>
<td>LV</td>
<td>−23.05</td>
<td>30.17</td>
<td>29</td>
<td>11.273</td>
<td>0.636</td>
<td>0.793 ± 0.126</td>
<td>0.575 ± 0.188</td>
<td>5.571</td>
<td>0.292</td>
<td>0.115 (0.103)</td>
</tr>
<tr>
<td>Port Elizabeth</td>
<td>PE</td>
<td>−33.96</td>
<td>25.57</td>
<td>6</td>
<td>4.909</td>
<td>0.182</td>
<td>0.641 ± 0.205</td>
<td>0.530 ± 0.277</td>
<td>3.927</td>
<td>0.274</td>
<td>0.068 (0.090)</td>
</tr>
<tr>
<td>Stellenbosch</td>
<td>CL</td>
<td>−33.88</td>
<td>18.74</td>
<td>30</td>
<td>11.273</td>
<td>1.182</td>
<td>0.814 ± 0.080</td>
<td>0.637 ± 0.188</td>
<td>5.787</td>
<td>0.235</td>
<td>0.099 (0.100)</td>
</tr>
<tr>
<td>Upington</td>
<td>UP</td>
<td>−28.45</td>
<td>21.24</td>
<td>29</td>
<td>9.818</td>
<td>0.455</td>
<td>0.779 ± 0.103</td>
<td>0.611 ± 0.202</td>
<td>5.65</td>
<td>0.233</td>
<td>0.098 (0.083)</td>
</tr>
<tr>
<td>Tanzania</td>
<td>TAN</td>
<td>−6.95</td>
<td>37.53</td>
<td>28</td>
<td>11.273</td>
<td>0.727</td>
<td>0.819 ± 0.091</td>
<td>0.662 ± 0.214</td>
<td>5.138</td>
<td>0.211</td>
<td>0.095 (0.106)</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>ZIM</td>
<td>−17.86</td>
<td>31.04</td>
<td>23</td>
<td>10.273</td>
<td>0.273</td>
<td>0.800 ± 0.113</td>
<td>0.603 ± 0.212</td>
<td>5.228</td>
<td>0.269</td>
<td>0.116 (0.117)</td>
</tr>
<tr>
<td>Australia</td>
<td>AUS</td>
<td>−31.98</td>
<td>115.88</td>
<td>30</td>
<td>3.091</td>
<td>0</td>
<td>0.378 ± 0.232</td>
<td>0.303 ± 0.254</td>
<td>2.26</td>
<td>0.214</td>
<td>0.064 (0.095)</td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>GRE</td>
<td>39.4</td>
<td>21.91</td>
<td>28</td>
<td>2.727</td>
<td>0</td>
<td>0.369 ± 0.228</td>
<td>0.287 ± 0.245</td>
<td>2.147</td>
<td>0.244</td>
<td>0.070 (0.112)</td>
</tr>
<tr>
<td>Madeira</td>
<td>MAD</td>
<td>32.67</td>
<td>−16.84</td>
<td>22</td>
<td>3.364</td>
<td>0.091</td>
<td>0.487 ± 0.176</td>
<td>0.319 ± 0.197</td>
<td>2.718</td>
<td>0.373</td>
<td>0.125 (0.101)</td>
</tr>
<tr>
<td>Americas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guatemala</td>
<td>GUA</td>
<td>14.89</td>
<td>−90.51</td>
<td>30</td>
<td>2.909</td>
<td>0</td>
<td>0.391 ± 0.193</td>
<td>0.359 ± 0.218</td>
<td>2.217</td>
<td>0.102</td>
<td>0.047 (0.070)</td>
</tr>
</tbody>
</table>
Population structure analysis

Individuals were assigned to populations (genetic clusters) using STRUCTURE v2.3.4 (Pritchard et al., 2000; Falush et al., 2003) based on multilocus genotypes without including a priori information. The number of optimal clusters (K) was estimated using an admixture model and correlated allele frequencies which allow allele frequencies to be similar in different populations due to shared ancestry or continued migration. For each possible cluster (K), a number varying between 1 and 14 (total number of sampling localities), we performed 10 independent runs with 1,000,000 MCMC permutations and burn-in set at 100,000, which allowed the different parameters to reach stability and convergence. As this only detects the uppermost level of differentiation in the data, we subsequently investigated the substructure in the data following Coulon et al. (2008) by repeating the analyses until the number of clusters inferred was one. The optimal K (most likely number of clusters) was assessed using two methods. First, we inspected the log-probabilities of the different possible clusters; a high value with limited variance is an indication of the true K (Pritchard et al., 2000). Second, we inspected the second-order rate change of Ln P(X/Y) calculated according to Evanno et al. (2005) implemented in the online resource STRUCTURE HARVESTER (Earl & Von Holdt, 2012). Thereafter, the number of runs (from STRUCTURE) of the chosen K value was averaged in CLUMPP v1.2.2 (Jakobsson & Rosenberg, 2007) and visualized in DISTRUCT v1.1 (Rosenberg, 2004).

We quantified the degree of population differentiation between the genetic clusters identified in STRUCTURE by calculating pairwise Fst values between the clusters in MICRO-SATellite ANALYSER v4.05 (MSA; Dieringer & Schlötterer, 2003) as well as overall Fst values in FRENA v1.0 running 10,000 replications including (INA) and excluding (ENA) null alleles (Chapuis & Estoup, 2007). The hypothesis of isolation by distance (IBD) within the two clusters identified in STRUCTURE was investigated using two separate Mantel tests in ARLEQUIN v3.5.1.2 (Excoffier & Lischer, 2010). In each case, we regressed the sum of squared size differences (Rst) against the geographic distance (km). To further assess the relationships between different populations based on their allele frequency, we used a Principal Coordinate Analysis (PCoA) implemented in the program GENALEX v6.5 (Peakall & Smouse, 2006, 2012). The first three principal axes were plotted in STATISTICA v12 (Statsoft Inc., Tulsa, Oklahoma).

We constructed unrooted neighbour-joining trees and assessed statistical support for each branch using non-parametric bootstrapping (10,000 replicates) using the online POPTREEW (Takezaki et al., 2014) based on Nei’s genetic distance (Dni; Nei et al., 1983), as well as the genetic distance calculated based on the proportion of shared alleles (DST; Bowcock et al., 1994).

Population demography analysis

To obtain in-depth information regarding the invasion pathway of C. capitata, we tested eighteen different hypothetical scenarios and analysed them using the ABC method implemented in the program DIYABC v2.03 (Cornuet et al., 2014). The use of ABC has many advantages including using several summary statistics simultaneously as well as providing probability values with confidence intervals (Estoup & Guillemaud, 2010). Many different scenarios can be tested with the different populations in this study which can become computationally intensive. We therefore narrowed down the number of scenarios to test by utilizing historical information as well as results from our cluster analyses. On the basis of Bayesian cluster analyses, we based the different scenarios on six populations [Africa (without Burgers Hall), Burgers Hall, Madeira, Greece, Guatemala and Australia] as well as including an unsampled population. This unsampled population corresponds to an initial site of colonization of the Mediterranean coast at a location other than locations sampled in this study. We tested 18 different scenarios related to testing (1) from where Guatemala was colonized as this information is unknown as well as (2) from where Burgers Hall (South Africa) was re-colonized (Appendix S1 in Supporting Information). In total, we considered eighteen scenarios to test as hypotheses for the route of invasion, specifically surrounding the overall theme that C. capitata originated in eastern and southern Africa and subsequently spread to the rest of the world. The prior distributions of parameters were uniform and outlined as follows: 1000 < N1 < 1000 000; 10 < N2,N3,N4,N5,N6, N7 < 1000000; 1 < db < 10; 2 < Nib < 100; 10 < t1,t2 < 500; 100 < t3,t4 < 1500; 1000 < t5,t6 < 10000; 0.001 < ra < 0.999; N is the effective population size, Nib is the number of founders in each colonization event, db is the duration of the bottleneck event (number of generations), ra is the admixture rate and t is the timing of an event (generations back in time) (Appendix S2). Prior distributions of parameters for the microsatellite mutation model were set to default values and therefore represented the generalized stepwise mutation model. We included all summary statistics in both the one-sample and two-sample summary statistics. Our timing of events was based on the approximate number of generations C. capitata can support in a year and colonization dates available. We based the number of generations per year on the number of days development takes from egg to egg in the laboratory at 18 °C (Grout & Stoltz, 2007). We assumed that the effective population size for the native populations was larger than those in the introduced range and the bottleneck event (db) occurred after the initial introduction. The newly introduced individuals might take several generations to establish a population, and this parameter was therefore bounded between
one and ten generations. In each test, we simulated 7,000,000 computations and for each scenario computed a posterior probability, including 95% confidence intervals (CI), with a logistic regression (Cornuet et al., 2014). The scenario with the highest posterior probability and non-overlapping 95% CI was chosen as the most likely scenario.

RESULTS

Population genetic diversity

Overall, the C. capitata populations examined showed deviations from HWE. These departures from HWE can possibly be coupled with moderate levels of inbreeding ($F_{IS}$) ranging between 0.102 (Guatemala) and 0.373 (Madeira) (Table 1), the Wahlund effect (reduction in heterozygosity due to hierarchical structuring) and/or the presence of null alleles. The frequency of null alleles in all populations was low to intermediate and ranged between 0.023 and 0.139 (Table 1, Appendix S3). Null alleles are commonly found in taxa with high effective population sizes (Chapuis & Estoup, 2007). The $F_{ST}$ estimation before and after null allele correction (see Population structure section) showed no difference, although this does not suggest that all other analyses are unbiased. Furthermore, no linkage disequilibrium was detected for any of the 11 microsatellite markers. All African populations possessed relatively high levels of genetic diversity based on the expected heterozygosity ($H_{E}$), number of alleles ($N_{A}$) and allelic richness ($A_{R}$) compared to those locations in the introduced range (Australia, Greece, Guatemala, Madeira) (Table 1). All African populations except Senegal had a number of private alleles (0.182–1.182; Table 1). The highest number of private alleles was found in Stellenbosch (South Africa) with many locations (Australia, Greece, Guatemala, Senegal) having no private alleles.

Population structure

Genetic differentiation was measured by overall and pairwise $F_{ST}$ calculations between the genetic clusters inferred by the Bayesian clustering method (Appendix S4). All pairwise $F_{ST}$ comparisons between the different clusters were significant. The overall $F_{ST}$ value before null allele correction ($F_{ST} = 0.143$) as well as after ENA correction ($F_{ST} = 0.142$) indicated significant population differentiation. There was no significant pattern of IBD based on results from the Mantel test for the cluster that includes all populations in the introduced range as well as Burgers Hall ($r = -0.149, P = 0.640$). Moreover, significant IBD was detected over the African continent (excluding Burgers Hall) ($r = 0.736, P = 0.031$).

In the Principal Coordinate Analysis (Fig. 1a), the first 3 axes explained most of the genetic variation (63.22%). The first axis (45.33%) separates samples from Africa with high effective population sizes (Chapuis & Estoup, 2007). The $F_{ST}$ estimation before and after null allele correction (see Population structure section) showed no difference, although this does not suggest that all other analyses are unbiased. Furthermore, no linkage disequilibrium was detected for any of the 11 microsatellite markers. All African populations possessed relatively high levels of genetic diversity based on the expected heterozygosity ($H_{E}$), number of alleles ($N_{A}$) and allelic richness ($A_{R}$) compared to those locations in the introduced range (Australia, Greece, Guatemala, Madeira) (Table 1). All African populations except Senegal had a number of private alleles (0.182–1.182; Table 1). The highest number of private alleles was found in Stellenbosch (South Africa) with many locations (Australia, Greece, Guatemala, Senegal) having no private alleles.

Figure 1 (a) Principal Coordinate Analysis (PCoA) plot for 14 Ceratitis capitata populations (1- Australia, 2- Greece, 3- Guatemala, 4- Madeira, 5- Senegal, 6- Port Elizabeth, 7- Burgers Hall, 8- Upington, 9- Stellenbosch, 10- Zimbabwe, 11- Levubu, 12- Mozambique, 13- Tanzania, 14- Kenya). (b) Unrooted neighbour-joining trees for genetic distance based on shared alleles ($D_s$). The number at each node indicates the bootstrap values after 10 000 bootstrap replicates. Only bootstrap values above 70% are shown. Coloured circles indicate the two clusters that correspond to clusters obtained in STRUCTURE.
investigated by running each cluster separately (Fig. 3). Substructuring in the African cluster (Cluster 1) (Fig. 3a) showed two further clusters based on the optimal $K$ calculated in Evanno et al. (2005). However, no obvious geographic pattern was observed that correlated with the two clusters identified within the African group. Within the world cluster (Cluster 2), four additional clusters were identified (Fig. 3b). The first cluster included Australia, the second individuals from both Greece and Guatemala, the third individuals from Madeira, and the final cluster included all individuals from Burgers Hall. The cluster representing Greece and Guatemala was run again, and these two populations were subsequently clustered separately (Fig. 3c).

**Invasion pathway scenarios from ABC**

To investigate the routes of colonization, eighteen different scenarios were tested using approximate Bayesian computations. One scenario (Scenario 3), based on posterior probability values, was clearly superior to all other scenarios tested (Table 2; Appendix S6). This scenario closely matched the proposed historical routes of invasion (Fig. 4), with an initial colonization of Europe from Africa and a secondary colonization of Australia from Europe. Moreover, we show that Guatemala was likely colonized from Greece and the back introduction into South Africa at Burgers Hall came from the Unsampled location (Europe) included in the analysis. For the most likely scenario, the number of founders in each introduction event ranged between 37.9 and 66.70 and the duration of the bottleneck event was approximately nine generations (Appendix S7). Larger-than-expected effective population sizes (Appendix S7) were observed in some of the introduced populations (Australia, Guatemala, Unsampled location) as well as in Burgers Hall compared to populations from the native range. The timing of events (Appendix S7) was estimated as being much further in the past than indicated by historical records.

**DISCUSSION**

Global biological invasions of pest insects, such as that of *Ceratitis capitata*, are typically characterized by reduced genetic diversity due to a small number of founders colonizing the introduced range. This creates a unique set of genetic and demographic attributes which were investigated here making use of 11 microsatellites for 14 macrogeographic sampling locations of *C. capitata*. The colonization of *C. capitata* world-wide is well documented, and our results indicate high levels of genetic diversity in the native range and a decrease in this diversity in the introduced range. Moreover, clear genetic differentiation was found between Africa and the rest of the world. The reconstructed invasion pathway for *C. capitata* closely matches the proposed hypothesis of an initial colonization of Europe from Africa.
and a secondary colonization of Australia from Europe. We also show that Guatemala was likely colonized from Greece and the reintroduction into Burgers Hall was from the Unsampled location on the Mediterranean coast.

The genetic diversity estimates from this research (expected heterozygosity, allelic richness, number of alleles and number of private alleles) are similar to those found in other studies of *C. capitata* (Bonizzoni et al., 2001, 2004; Gasperi et al., 2002). We found high levels of genetic diversity (expected heterozygosity, number of private alleles, allelic richness) in all African populations, except Senegal and Burgers Hall, with a further decline in the genetic diversity estimates of populations in the introduced range (Australia, Greece, Guatemala, Madeira). The high level of genetic diversity found in this study for *C. capitata* in Africa is possibly due to their large effective population sizes. By contrast, both Senegal and Burgers Hall populations have diversity estimates similar to those of populations in the introduced range. As eastern and southern Africa has been identified as the native range of *C. capitata* (Malacrida et al., 1998, 2007; Gasperi et al., 2002), it is perhaps unsurprising that Senegal (West Africa) has similar levels of genetic diversity as other populations in the introduced range. The diversity estimates for Senegal are in fact higher than those of other locations in the introduced range, which might indicate that flies moved to Senegal from the native range via natural movement as well as by human assistance rather than only by jump dispersal. Interestingly, Burgers Hall, a location within South Africa, also has a degree of genetic diversity similar to that of populations in the introduced range, whilst other locations in South Africa seem to form part of the native range. These results are further supported by STRUCTURE and the

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Posterior probability (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.001 (0.000, 0.546)</td>
</tr>
<tr>
<td>2. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.001 (0.000, 0.549)</td>
</tr>
<tr>
<td>3. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.998 (0.997, 0.999)</td>
</tr>
<tr>
<td>4. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>5. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>6. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>7. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>8. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.001 (0.000, 0.545)</td>
</tr>
<tr>
<td>9. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>10. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>11. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>12. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>13. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>14. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>15. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>16. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>17. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
<tr>
<td>18. Africa-Unsampled-Madeira-Greece;Madeira-Burgers Hall;Greece-Australia;Greece-Guatemala;Madeira-Burgers Hall</td>
<td>0.000 (0.000, 0.545)</td>
</tr>
</tbody>
</table>
of a small number of founders introduced to this region, reduction in genetic diversity here seems to support the idea Mediterranean coast (Maddison & Bartlett, 1989), and the assumed to have travelled human-mediated to the on the African continent that are not pest free. This high level of movement of flies is further aided by limited quarantine restriction within areas (mal) or human travel. This high level of movement of flies is supplied by the continuous trade of goods (formal and informal) or human travel. This high level of movement of flies is therefore likely to have been influenced by human activities, including travel and trade. Hence, the introduction of C. capitata into the Mediterranean region is likely to have been facilitated by human-mediated transport, which may explain the observed high levels of connectivity observed in this region.

Furthermore, we found a clear pattern of genetic differentiation between the native African populations and those from the introduced range broadly forming two groups. The PCoA, STRUCTURE and NJ trees largely support the 'out-of-Africa' hypothesis, grouping all the African populations (Cluster 1) together as well as all the introduced locations elsewhere in the world including Burgers Hall (Cluster 2). Moreover, Cluster 2 was further structured into four groups: Australia, Madeira, Burgers Hall and a group including Greece and Guatemala. This group was subsequently further sub-divided into two separate groups, each group pertaining to a sampling location. These six additional clusters identified in Cluster 2 were supported by the pairwise $F_{ST}$ values. The lack of differentiation throughout Africa indicates high effective population sizes, which may well explain the observed deviations from HWE.

Consequently, it seems reasonable to conclude that there are low levels of connectivity between the African continent and the introduced range except for the reintroduction into Burgers Hall from Europe, although high levels of connectivity exist on a regional scale. This information, in turn, is important for the management of C. capitata world-wide. This pattern of differentiation on a macrogeographic scale and a lack thereof on a more regional scale has also been evident in other tephritid studies including Bactrocera cucurbitae (Virgilio et al., 2010) and Bactrocera oleae (Nardi et al., 2005). The lack of connectivity indicates that quarantine measures for export consignments from Africa to elsewhere in the world are broadly successful in limiting the movement of fruit flies intercontinentally. Moreover, all locations sampled in the introduced range belong to separate genetic clusters, indicating a lack of movement between these locations. However, there is evidence for high levels of movement of fruit flies on the African continent. This does not bode well for the fruit industry on the African continent, and especially in the prevention of new invasions into a previously pest-free region. A case in point is that of Bactrocera invadens (B. dorsalis) which, after its introduction to the African continent in 2003 (Lux et al., 2003), has spread over large parts of the continent despite extensive quarantine and eradication efforts (De Meyer et al., 2010). Information from the reconstruction of the routes of invasion for C. capitata is important to understand the different evolutionary and environmental factors that influence successful invasions, and that can then be incorporated into strategies for the control and prevention of new invasions (Estoup & Guillemaud, 2010). Future studies should therefore focus on including more samples from native and introduced locations for C. capitata as well as higher numbers of molecular markers to infer a more
complete picture of the global invasion pathways and whether this is dynamic over time.

ACKNOWLEDGEMENTS

Microsatellite data were genotyped at the Central Analytical Facility, Stellenbosch University. We are grateful to farmers in South Africa, South Africa, Weidner Enkerlin, Ernie Steiner, Sunday Ekesi, Luis Bota, Louisa Makumbe, Moulid Mwatawala, Nikos Papadopoulos, Ibrahim Solo and Pierre-Alexandre Rastorgueff. This research was funded by THRIP from the National Research Foundation (NRF) of South Africa [(UID) 71909], SAAPPA, SASPA, as well as SATGI. The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by NRF supported research are that of the author (s), and that the NRF accepts no liability whatsoever in this regard. We are grateful for constructive comments and discussion of this work by Matthew Hill, Anthony Clarke, Jaco le Roux and the anonymous referees.

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Schematic representations of the eighteen different scenarios tested using DIYABC v2.03 (Cornuet et al., 2014).

**Appendix S2** Definitions and prior distributions of parameters used in the testing of eighteen invasion scenarios of *Ceratitis capitata* using the ABC method implemented in DIYABC v2.03 (Cornuet et al., 2014).

**Appendix S3** Frequency of null alleles calculated according to Dempster et al. (1977) in FREENA 1.0 (Chapuis & Estoup, 2007) for 323 *Ceratitis capitata* individuals.

**Appendix S4** Pairwise *F*<sub>ST</sub> values calculated in MICROSATellite ANALYSER v4.05 (Dieringer & Schlötterer, 2003) between the six genetic clusters [Australia, Greece, Guatemala, Madeira, Burgers Hall, Africa (all other African locations)] inferred from Bayesian Clustering for *Ceratitis capitata*.

**Appendix S5** The lines indicate the most likely number of clusters (*K*) in STRUCTURE *Ceratitis capitata* populations according to the natural logarithm of the likelihood (Ln P (X/K)) (Pritchard et al., 2000) and the delta *K* method (Evanno et al., 2005).

**Appendix S6** The most likely scenario (3) making use of approximate Bayesian computations (ABC) in DIYABC v2.03 (Cornuet et al., 2014).

**Appendix S7** Means of the posterior probabilities for demographic parameters estimated under Scenario 3.
Minette Karsten is a PhD student at the University of Stellenbosch, South Africa. Her interests include population genetics, invasion biology and the management of fruit flies. She is part of two larger research groups, an Integrated Pest Management group and the Applied Physiological Ecology Lab (http://insectphysiologiecology.blogspot.com/).

Author contributions: J.S.T. and M.K. conceived and designed the experiments. M.K. performed the experiments. J.S.T., B.J.V.V. and P.A. contributed reagents/materials/analysis tools. M.K. wrote the first draft of manuscript, and all authors contributed to interpretation of results and helped writing the manuscript.

Editor: Brian Leung