Molecular Phylogenetics and Evolution 161 (2021) 107160

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomic resolution of the *Ceratitis* FARQ complex (Diptera: Tephritidae)

Yue Zhang [a], Marc De Meyer [b], Massimiliano Virgilio [b], Shiqian Feng [a], Kemo Badji [c], Zhihong Li [a,*]

[a] Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing 100193, China
[b] Royal Museum for Central Africa, Invertebrates Section and JEMU, Tervuren B3080, Belgium
[c] Crop Protection Directorate, Dakar, Senegal

ARTICLE INFO

Keywords:
Ceratitis FARQ complex
Mitochondrial genome
Genome-wide SNPs
Phylogenetic analysis
Species delimitation

ABSTRACT

The *Ceratitis* FARQ complex (formerly FAR complex) includes four frugivorous tephritids, *Ceratitis fasciventris*, *C. anonae*, *C. rosa* and *C. quilicii*, the latter two causing important agricultural losses in Africa. Although FARQ species can be identified on the basis of subtle morphological differences, they cannot be resolved as monophyletic when trying phylogenetic tree reconstructions based on mitochondrial or nuclear gene fragments except for microsatellites. In this study, we used mitogenome and genome-wide SNPs to investigate the phylogenetic relationship within the complex as well as between all four *Ceratitis* subgenera. The analysis of 13 species supported the monophyly of the *Ceratitis* subgenera *Ceratitis*, *Ceratalaspis*, *Pardalaspis*, and recovered *Pterandrus* as paraphyletic but could not properly resolve species within the FARQ complex. Conversely, gene and species tree reconstructions based on 785,484 genome-wide SNPs could consistently resolve the FARQ taxa and provide insights into their phylogenetic relationships. Gene flow was detected by TreeMix analysis from *C. quilicii* to *C. fasciventris*, suggesting the existence of introgression events in the FARQ complex. Our results suggest that genome-wide SNPs represent a suitable tool for the molecular diagnosis of FARQ species and could possibly be used to develop rapid diagnostic methods or to trace the origins of intercepted samples.

1. Introduction

Tephritid pests (Diptera: Tephritidae) attack and infest a variety of economically important fruits and vegetables and represent a major agricultural threat, causing direct and indirect economic losses to small farmers and international traders (Ekesi et al., 2016, Tanga et al., 2018). In Africa, economically important fruit flies mainly belong to the genera *Bactrocera*, *Ceratitis*, *Dacus* and *Zeugodacus* (Ekesi et al., 2016). Within the genus *Ceratitis*, four highly polyphagous and morphologically similar species, *Ceratitis fasciventris* (Bezzi), *C. anonae* Graham, *C. rosa* Karsch and *C. quilicii* De Meyer, Mwatawala & Virgilio, are also referred to as the *Ceratitis* FARQ complex (formerly FAR) (Barr and McPheron, 2006, De Meyer et al., 2016, Virgilio et al., 2013). *C. rosa* and *C. quilicii* are considered the most aggressive pests of the complex, as combined, they infest more than 90 species belonging to 25 families of wild and cultivated plants. However, their host ranges were still not clearly defined until recently (De Meyer et al., 2016), and the two species were both recognized as *C. rosa*. Life-cycle simulation models (Tanga et al., 2018) showed that a number of North American, South American and Southeast Asian regions are potentially suitable for the introduction and establishment of *C. rosa* and *C. quilicii*. In this respect, developing accurate, rapid identification tools for species of the *Ceratitis* FARQ complex would be vital to prevent introductions out of the African distribution of these species. Panels of microsatellite markers (Virgilio et al., 2013, Virgilio et al., 2019) could potentially be used on problematic samples (e.g. immature stages or females). However, more rapid and cost-effective tools should be developed for the routine diagnosis of FARQ species, e.g., in the framework of quarantine or SIT applications.

Even if microsatellite-based population genetics is effective in recovering the species within the FARQ complex as defined by genotypic clusters (Virgilio et al., 2013), phylogenetic relationships within the complex remain unclear, as tree reconstructions based on mitochondrial

* Corresponding author.
E-mail addresses: zyzhangyue@cau.edu.cn (Y. Zhang), marc.de.meyer@africamuseum.be (M. De Meyer), massimiliano.virgilio@africamuseum.be (M. Virgilio), fengshiqian@163.com (S. Feng), kemo2fr@yahoo.fr (K. Badji), lizh@cau.edu.cn (Z. Li).

https://doi.org/10.1016/j.ympev.2021.107160
Received 29 April 2020; Received in revised form 17 March 2021; Accepted 23 March 2021
Available online 29 March 2021

and nuclear gene fragments fail to resolve the FARQ species as monophyletic (Barr and McPheron, 2006, De Meyer et al., 2016, Virgilio et al. 2008). The FARQ complex belongs to the subgenus *Pterandrus* (*sensu* Barr and McPheron, 2006) (De Meyer and Freidberg, 2005), whose classification also proves to be problematic, as it is divided into two different sections, one of which is paraphyletic to the subgenus *Ceratitis s.s.* (Barr and Wiegmann, 2009). It would now be interesting to clarify further evolutionary relationships between FARQ taxa in the context of the subgeneric classification of *Ceratitis*. In this respect, high-throughput sequencing technologies and genomic tools prove to be effective in resolving problematic phylogenies (e.g., see Dupuis et al. 2017). The use of genomic tools, including mitochondrial genomics (mitogenomics) and whole genome sequencing (WGS), proved to be an effective and powerful approach to investigate insect molecular systematics and taxonomy, phylogeography and population dynamics (Burger et al., 2014, Du et al., 2019, Dupuis et al., 2017, Esquerré et al., 2019, Feng et al., 2018, Jia et al., 2012, Kornilios et al., 2020, Liu et al., 2020, Ma et al., 2012, Torricelli et al., 2010, Wang et al., 2013), and whole reference genomes or mitogenomes are publicly available in GenBank (see Table S1 for a list of tephritid mitochondrial genomes).

In this study, we address the molecular taxonomy and phylogeny of the FARQ complex from a phylogenomic perspective with the main objective of clarifying relationships between the FARQ species as well as between the complex and the other *Ceratitis* subgenera.

2. Materials and methods

2.1. Samples collection and DNA extraction

Thirty-six samples of thirteen species, representing four *Ceratitis* subgenera, were collected from nine African countries by protein or trimedlure (TML)-baited traps (Table 1).

Morphological identification was based on a set of multientry identification keys for African frugivorous flies (Diptera, Tephritidae) developed by Virgilio et al. (2014). Total genomic DNA was isolated from each specimen (the abdomen was removed to avoid the pollution of symbiotic bacteria) using the DNeasy DNA Extraction kit (QIAGEN) following the standard operating procedure with minor modifications.

2.2. High-throughput sequencing

Total genomic DNA was sent to BerryGenomics (China) for fragmentation to an average insert size of 450 bp and library construction using the Illumina TruSeq Nano DNA library Prep Kit with input of

Table 1
Taxon sampling in this study.

Genus	Subgenus	Species	Sampling Number	Population	Collection Date	Total Length	Accession Number	Reference
Ceratitis	*Ceratalaspis*	*C. cosyra*	*C. cosyra* SDN1	Singa, Sudan	2009-10-13	16,424	MT036783	Present study
			C. cosyra ZWE1	Zimbabwe	2018-03-21	15,949	MT036784	Present study
		C. pallidula	*C. pallidula* TZA1	Morogoro, Tanzania	2013-06-01	15,959	MT036774	Present study
			C. pallidula TZA2	Morogoro, Tanzania	2013-06-01	15,958	MT036775	Present study
		C. quinaria	*C. quinaria* SDN1	Singa, Sudan	2009-10-13	15,950	MT036788	Present study
	Ceratitis	*C. capitata*	*C. capitata* KEN1	Ruiru, Kenya	2011-05-12	15,980	MT036782	Present study
			C. capitata LAB1	Laboratory strain		15,980	AJ242872	Spanos et al., 2000
	Pardalaspis	*C. bremii*	*C. bremii* GIN1	Conakry, Guinée	2017-07-16	16,252	MT036781	Present study
		C. ditissima	*C. ditissima* NGA1	Nasarawa, Nigeria	2017-08-19	16,334	MT036785	Present study
		C. punctata	*C. punctata* GIN1	Conakry, Guinée	2017-07-16	16,376	MT036786	Present study
	Pterandrus	*C. anoane*	*C. anoane* CMR1	Cameroon	2004-10-18	16,052	MT036765	Present study
			C. anoane CMR2	Cameroon	2004-10-18	16,014	MT036766	Present study
			C. anoane CMR3	Cameroon	2004-10-18	16,021	MT036767	Present study
			C. anoane CMR4	Cameroon	2004-10-18	16,019	MT036768	Present study
		C. fasciventris	*C. fasciventris* BDI1	Nyarubanga, Burundi	2014-01-05	16,027	MT036769	Present study
			C. fasciventris BDI2	Nyarubanga, Burundi	2014-01-05	16,018	MT036770	Present study
			C. fasciventris BDI3	Nyarubanga, Burundi	2014-01-05	16,026	MT036771	Present study
			C. fasciventris BDI4	Nyarubanga, Burundi	2014-01-05	16,022	MT036772	Present study
			C. fasciventris BDI5	Nyarubanga, Burundi	2014-01-05	16,020	MT036773	Present study
			C. fasciventris KEN1	Kenya		16,017	KY436396	Drosopoulou et al., 2017
		C. querita	*C. querita* KEN1	Nairobi, Kenya	2005-05-30	15,990	MT036787	Present study
		C. quilicii	*C. quilicii* TZA1	Langali, Tanzania	2013-05-18	16,016	MT036776	Present study
			C. quilicii TZA2	Langali, Tanzania	2013-05-18	16,021	MT036777	Present study
			C. quilicii TZA3	Langali, Tanzania	2013-05-18	16,025	MT036778	Present study
			C. quilicii TZA4	Langali, Tanzania	2013-05-18	16,021	MT036779	Present study
			C. quilicii TZA5	Langali, Tanzania	2013-05-18	16,021	MT036780	Present study
			C. quilicii ZAF1	Mpumalanga, South Africa	2014-02-20	16,025	MT036791	Present study
			C. quilicii ZAF2	Mpumalanga, South Africa	2014-02-20	16,033	MT036792	Present study
			C. quilicii ZAF3	Mpumalanga, South Africa	2014-02-20	16,022	MT036793	Present study
			C. quiliciiZAF4	Mpumalanga, South Africa	2014-02-20	16,023	MT036794	Present study
			C. quilicii ZAF5	Mpumalanga, South Africa	2014-02-20	16,033	MT036795	Present study
			C. quilicii ZAF6	Mpumalanga, South Africa	2014-02-20	16,018	MT036790	Present study
		C. rosa	*C. rosa* TZA1	Langali, Tanzania	2013-05-18	16,001	MT036796	Present study
			C. rosa TZA2	Langali, Tanzania	2013-05-18	15,993	MT036797	Present study
			C. rosa TZA3	Langali, Tanzania	2013-05-18	15,994	MT036798	Present study
			C. rosa TZA4	Langali, Tanzania	2013-05-18	15,993	MT036799	Present study
			C. rosa TZA5	Langali, Tanzania	2013-05-18	15,999	MT036800	Present study
		C. robivora	*C. rubivora* TZA1	Langali, Tanzania	2013-05-18	16,017	MT036789	Present study
Neoceratitis		*N. asiatica*	*N. asiatica* CHN1	Ningxia, China	2014-06-25	15,481	MF434829	Su et al., 2017

0.15–4.32 μg genomic DNA. Next-generation sequencing was conducted on an Illumina HiSeq 2500 platform (2 × 150 bp paired-end reads). Fastp v0.20.0 (Chen et al., 2018) with default settings was used for preliminary adapter and read trimming and read filtering.

2.3. Mitochondrial genome assembly, annotation and analysis

The 658 bp fragment of *cox1* (DNA barcode region) was used as an anchor for mapping reference strategy assembly. The universal primer pair LCO1490/HCO2198 (Flomer, 1994) was used to amplify the fragments with the following reaction conditions: 1 cycle of denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 53 °C for 30 s, and elongation at 72 °C for 30 s and then a final extension at 72 °C for 10 min. PCR products were purified and sent for Sanger sequencing at BGI (China).

Cleaned paired-end reads assembled to a 658 bp *cox1* fragment using the "Map to Reference" strategy were run in Geneious R.10.0 (https://www.geneious.com) with 1000 iterations to obtain mitochondrial genome reads from data. Parameters were set following Herd et al.'s (2015) recommendation: 1) minimum overlap identity 95%, 2) minimum overlap 50 bp, 3) maximum 10% gaps per read and 4) maximum gap size 20 bp. High quality coverage contigs were generated into a consensus sequence. The overlapping region was manually searched at the head and tail of the consensus sequence and removed to circularize each complete mitochondrial genome.

Thirteen protein-coding genes (13 PCGs) and noncoding RNAs (22 tRNAs and 2 rRNAs) were automatically annotated by MITOS Web Server with the appropriate parameters: Reference= "RefSeq 63 Metazoa" and Genetic Code = "5 Invertebrate" (Bernt et al., 2013). Following the guidelines of Cameron (2014), the accuracy of 13 PCGs and 2 rRNAs was manually checked in Geneious by alignment with homologous genes of previously sequenced tephritid mitochondrial genomes. MitoZ (Meng et al., 2019), a mitogenome toolkit, was used to confirm the *de novo* assembly and annotation accuracy after comparison with the "map to reference strategy" results. Skews were manually calculated with the formula AT skew = (A − T)/(A + T) and GC skew = (G − C)/(G + C) (Perna and Kocher, 1995). A total of 36 annotated specimens of 13 *Ceratitis* species mitochondrial genome sequences were deposited in GenBank.

2.4. Genome-wide SNPs identification

The whole genome of the Mediterranean fruit fly, *Ceratitis capitata* (Papanicolaou et al., 2016) (GenBank accession GCA_000347755.4) was used as a reference genome for paired-end read mapping using Burrows-Wheeler Alignment with BWA-MEM algorithms (Li and Durbin, 2009). Following alignment, Picard Tools (http://broadinstitute.github.io/picard/) was used to mark and remove duplicate reads. The Genome Analysis Toolkit (GATK version 4.1.4.1) (McKenna et al., 2010) was used to call variants via the HaplotypeCaller algorithm in Genomic Variant Call Format (GVCF) mode. The SNP filtering was then implemented by using GATK- VariantFiltration with the following parameters: quality-by-depth ratio (QD) < 2.0, read mapping quality (MQ) < 40.0, probability of strand bias (FS) > 60.0, symmetric odds ratio (SOR) > 3.0, MQRankSum < -12.5 or ReadPosRankSum < -8.0. SNPs in approximate linkage equilibrium were pruned using PLINK v1.90b4 (Purcell et al., 2007) by considering a threshold of R2 > 0.8 and a window size = 1 k bp. Annovar (Wang et al., 2010) was used for SNP annotation, including 1) the distribution of SNPs (exon, intron, 5′ and 3′ untranslated regions, upstream and downstream regions, intergenic regions, and splice sites) and 2) the SNPs in exonic regions that were identified as synonymous or nonsynonymous.

2.5. Phylogenetic analyses

2.5.1. Mitogenomic data

For mitochondrial phylogenetic analyses, the mitochondrial genome of 38 specimens (including *C. capitata* AJ242872 and *C. fasciventris* KY436396 from GenBank) belonging to 13 *Ceratitis* species was used with *Neoceratitis asiatica* as the outgroup. Analyses were repeated on two different alignments obtained using the default parameters of ClustalW (Thompson et al. 1994). Possible incongruences in the alignment of protein-coding genes were verified by translating amino acids in MEGA 7.0 (Kumar et al., 2016) and checking for artifactual stop codons. The first alignment included all 13 protein-coding genes (dataset 1: PCG123) and the second protein-coding genes and ribosomal RNA (*rrnS* and *rrnL*, dataset 2: PCG123 and rRNAs). The rRNA genes were aligned using MAFFT with Q-INS-i iterative refinement methods to consider rRNA structural information (Katoh et al., 2019). For each gene, alignments were verified using Gblocks v0.91b (Castresana, 2000). SequenceMatrix v1.7 (Vaidya et al., 2011) was used to concatenate aligned sequences for each dataset. Mitogenomic data were used for both maximum likelihood (ML) and Bayesian (BI) tree reconstruction following the methods described below. For BI analyses, the best-fit partitioning schemes and substitution models for each gene were determined by PartitionFinder 2.1.1 (Lanfear et al., 2012) for both dataset 1 and dataset 2 using the corrected Akaike information criterion (AICc) greedy algorithm. Detailed partitioning schemes and best-fit models are listed in Tables S5 and S6.

2.5.2. Genome wide SNP data

For the analysis of relationships among *Ceratitis* FARQ complex species, based on genome-wide SNPs, we selected four *C. anonae* samples (*C. anonae* CMR1-CMR4), four *C. fasciventris* samples (*C. fasciventris* BDI1 and BDI3-BDI5), ten *C. quilicii* samples (*C. quilicii* TZA1-TZA5 and *C. quilicii* ZAF2-ZAF6), five *C. rosa* samples (*C. rosa* TZA1-TZA5) and *C. rubivora* as an outgroup for tree reconstructions. The SNPhylo pipeline (Lee et al., 2014) was used to generate multiple alignments from the filtered SNP data. Phylogenetic relationships within the FARQ complex were investigated using both ML and BI tree reconstructions. For ML analyses, we used RAxML (Stamatakis, 2014). Bayesian analyses were conducted with MrBayes v3.2.5 (Ronquist and Huelsenbeck, 2003) using two independent runs with four simultaneous Markov chains, running for 10,000,000 generations and trees sampled every 1000 generations (with the first 25% of trees discarded as burn-in). PartitionFinder, RAxML and MrBayes analyses were run in CIPRES Science Gateway V3.3 (Miller et al., 2010). FigTree v1.4.3 (Rambaut, 2015) and iTOL (Interactive Tree Of Life) v5 (Letunic and Bork, 2019) were used to draw and to annotate phylogenetic trees. Only nodes with bootstrap support > 75% and posterior probability (PP) > 0.95 were considered supported, and all others were discarded from further consideration.

For the FARQ complex, coalescent species delimitation was also implemented in SNAPP (Bryant et al., 2012). In this respect, an input XML file was compiled in BEAUti (Bouckaert et al., 2014) using all genome wide SNPs and default priors and model parameters. An MCMC run was then implemented in BEAST 2 (Bouckaert et al., 2014) for one million generations and sampled every 1000 generations. Run convergence was verified in Tracer v1.7 (Rambaut et al., 2018) by checking that the effective sample size (ESS) of parameters sampled from the MCMC run was > 200. This method was always possible, with the exception of a limited number of theta values for internal branches. DensiTree v2.2.6 (Bouckaert and Heled, 2014) was used to visualize the SNAPP species trees after a 10% MCMC chain burn-in. A maximum-clade-credibility tree was generated with TreeAnnotator (Drummond and Rambaut, 2007) with 10% burn-in and visualized in FigTree v.1.4.3 (Rambaut, 2015).

Split and introgression patterns within the *Ceratitis* FARQ complex were explored in TreeMix (Pickrell and Pritchard, 2012). This approach allows inferring historical relationships between populations on the

basis of covariance of allele frequencies and Gaussian approximation to genetic drift. Python scrpits plink2treemix.py (https://github.com/speciationgenomics/scripts/blob/master) and plink2treemix.py (https://bitbucket.org/nygcresearch/treemix/downloads) were preliminarily used to generate the TreeMix input file. Interspecific introgression was tested using *C. rubiovra* as the outgroup and 0, 1, 2, 3, and 4 migration edges (m). The R package OptM (Fitak, 2020) was used to optimize the number of migration edges using threshold models.

3. Results

3.1. Characterization of the mitochondrial genomes

High-throughput sequencing (HTS) generated an average of 24,727,683 reads per sample (SD = 3,740,156), corresponding to an average coverage of 16.99 (SD = 2.57) (Table S2). All the sequences reported in this study were deposited into the NCBI SRA under the accession number PRJNA701261 (Table S2).

In this study, we obtained 36 complete mitochondrial genomes from 13 species belonging to four subgenera of *Ceratitis*. The length range of the complete mitochondrial genome was from 15,949 bp in *C. cosyra* (Sudan) to 16,424 bp in *C. cosyra* (Zimbabwe). The length variations of 13 PCGs, 22 tRNAs and 2 rRNAs were very minor, while the length variation among different specimens was mainly caused by the variation in the control region (size range from 993 bp in *C. querita* to 1,487 bp in Zimbabwe *C. cosyra*). The arrangement of the 37 genes was identical to the arrangement in *Drosophila yakuba* (Clary and Wolstenholme, 1985), indicating high conservation.

Base composition of the whole mitochondrial genome, i.e., Thirteen PCGs, 22 tRNA genes, 2 rRNA genes, and a control region and comparative analysis of the A + T%, AT skew, and GC skew, are shown in Table S3. All *Ceratitis* species showed high A + T base compositional bias, especially in the control region (the average A + T% of the whole mitochondrial genome, 13 PCGs, 22 tRNA genes, 2 rRNA genes, and the control region was 77.17%, 75.13%, 76.96%, 80.25%, and 90.42%, respectively). All AT skews were positive, while GC skews were negative.

The usage of 13 PCG start codons and stop codons is shown in Table S4. All PCGs shared the common start codons as ATN, except for the *cox1* gene in all 13 species, which started with TCG, and the *nad1* gene in *C. cosyra*, which started with GTT. For Diptera, it is common to share with a special start codon TCG in *cox1*. For stop codons, most PCGs conservatively terminated with TAA. The *cob* gene possessed the unique stop codon TAG, and the *nad1* and *nad5* genes had incomplete termination codons (T in *nad1* and TA in *nad5*).

3.2. Genome-wide SNP identification and annotation

An average of 63,265,735 (SD = 4,941,586) SNPs per specimen were discovered using GATK (Table S5). After filtration, we identified 785,484 shared SNPs within the *Ceratitis* FARQ complex and *C. rubivora* (Table S5). All the samples SNPs data reported in this study were deposited into the The European Bioinformatics Institute - European Variation Archive (EBI-EVA) under the accessions Project ID: PRJEB43696, Analyses ID: ERZ1758389.

Among the SNPs, 71.85% of single base substitutions were transitions (A/G, C/T and A/C), and 28.15% were SNP transversions (A/T, C/G and G/T) (Fig. 1). SNPs recovered in intronic intergenic and exonic regions were 38.12%, 27.43% and 22.18%, respectively(Table 2).

3.3. Phylogenetic relationships among Ceratitis subgenera

The ML and BI analyses of two datasets (dataset 1: PCG123, dataset 2: PCG123 + rRNAs) including 38 mitochondrial genomes from 13 species (Fig. 2) allowed recovery of the subgenera *Ceratalaspis*, *Ceratitis*, and *Pardalaspis* as strongly supported monophyletic groups (posterior probability = 1.00 for BI and bootstrap = 100 for ML in all trees). However, the subgenus *Pterandrus* was paraphyletic because of *C. querita*, which was recovered as a sister species of *C. capitata*, while the other *Pterandrus* representatives (the *Ceratitis* FARQ complex and *C. rubivora*) were monophyletic. All ML and BI tree reconstructions

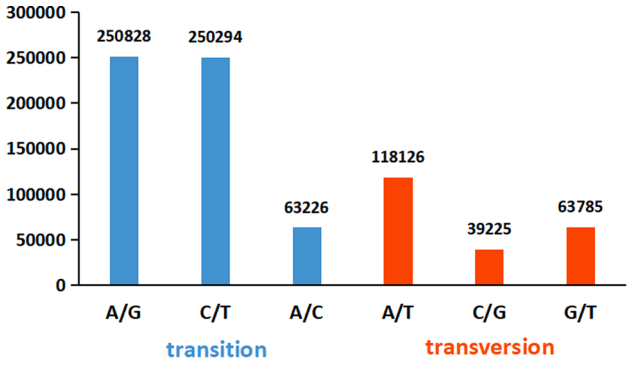

Fig. 1. The transitions and transversions of all the SNPs.

Table 2
Summary of genome-wide SNP annotation information.

Category	SNP Numbers
Number of total shared SNPs	785,484
Upstream	23,609
Exonic	174,188
Nonsynonymous	46,232
Synonymous	126,496
Stop gain	351
Stop loss	17
Unknown	1,092
Intronic	299,437
Splicing	142
Downstream	18,828
Upstream/Downstream	6,824
Intergenic	215,473
UTR3	20,688
UTR5	26,241
UTR5/UTR3	54

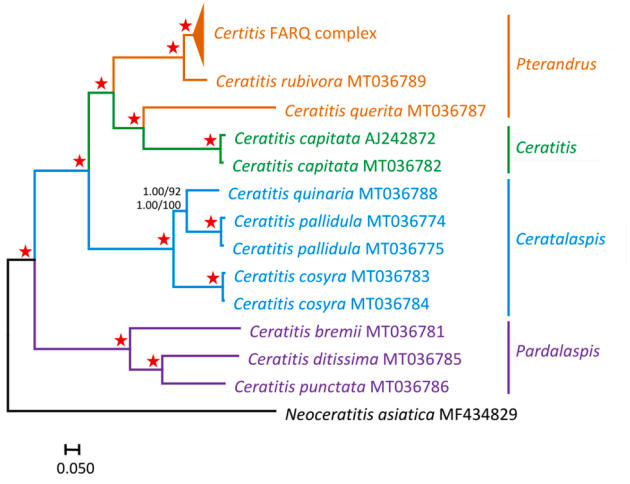

Fig. 2. Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees within four subgenera of *Ceratitis*. *Neoceratitis asiatica* was used as an outgroup. Values above the nodes represent 1) PCG123 Bayesian posterior probabilities for MrBayes; 2) PCG123 bootstrap values for RAxML; 3) PCG123 and 2 rRNAs Bayesian posterior probabilities for MrBayes; and 4) PCG123 and 2 rRNAs bootstrap values for RAxML. '★' indicates posterior probabilities = 1.00 and ML bootstrap = 100 in all trees.

recovered *Pardalaspis* in a basal position to a clade including *Ceratalaspis*, *Ceratitis* and *Pterandrus,* with *Ceratalaspis* basal to the clade, including *Pterandrus* and *Ceratitis*, forming topology as (((*Pterandrus* + *Ceratitis*) + *Ceratalaspis*)) + *Pardalaspis*)).

3.4. Phylogenetic analyses and species delimitation within the FARQ complex

Phylogenetic analyses of the *Ceratitis* FARQ complex based on mitogenomic data (both considering PCGs only or PCGs + rRNA) via ML and BI tree reconstructions are shown in Fig. 3. None of the phylogenetic reconstructions implemented could recover the four morphospecies of the FARQ complex as monophyletic, with most of the supported nodes including mixed samples from different species.

Conversely, the gene trees obtained from the analysis of genomic data including 785,484 polymorphic SNPs allowed recovery of the four species of the FARQ complex as distinct and well-supported monophyletic groups (BS = 100 and PP = 1.00 in all trees, see Fig. 4). *C. anonae* and *C. fasciventris* were recovered as sister clades closely associated with *C. rosa*, with *C. quilicii* in a basal position. The topology of the maximum-clade-credibility tree obtained from the coalescent analysis (Fig. 5) was strongly supported (PP = 1.00 at all nodes) and fully in agreement with the topologies of the gene trees obtained from genomic data (Fig. 4). The species trees visualized in DensiTree were remarkably consistent and further confirmed the close correspondence between gene and species trees.

To explore the patterns of introgression within the FARQ complex, TreeMix software was used to model a maximum-likelihood tree for both species splits and potential admixture events with *C. rubiovra* as an outgroup. The genetic relationships among species revealed by TreeMix distinguished the four clades, which is consistent with the above phylogenetic relationship evidence (Fig. 6), regardless of the absence of migration events, with one migration event. When considering one migration edge, as indicated by OptM (Fig. S1), TreeMix suggested the occurrence of an admixture event between *C. quilicii* and *C. fasciventris*, suggesting the existence of introgression events in the FARQ complex.

4. Discussion

The analysis of phylogenomic relationships between *Ceratitis* subgenera further confirmed the polyphyly of *Pterandrus*, which, according to Barr and Wiegmann (2009), is subdivided into two sections, with the FARQ complex belonging to *Pterandrus* section A (in this study represented by the four species of the FARQ complex and by *C. rubivora*). As in Barr and Wiegmann (2009), we recovered *C. querita* in *Pterandrus* section B, paraphyletic to representatives of the subgenus *Ceratitis s.s.* These results further suggest the need for revision for *Pterandrus*. In this study, taxon sampling for *Ceratalaspis* included only representatives of the *C. cosyra* complex (*C. cosyra*, *C. quinaira* and *C. pallidula*), while we have to take into account that this subgenus was polyphyletic in Barr and McPheron (2006). Conversely, the monophyly of *Pardalaspis* is in agreement with what was observed both in the molecular phylogeny of

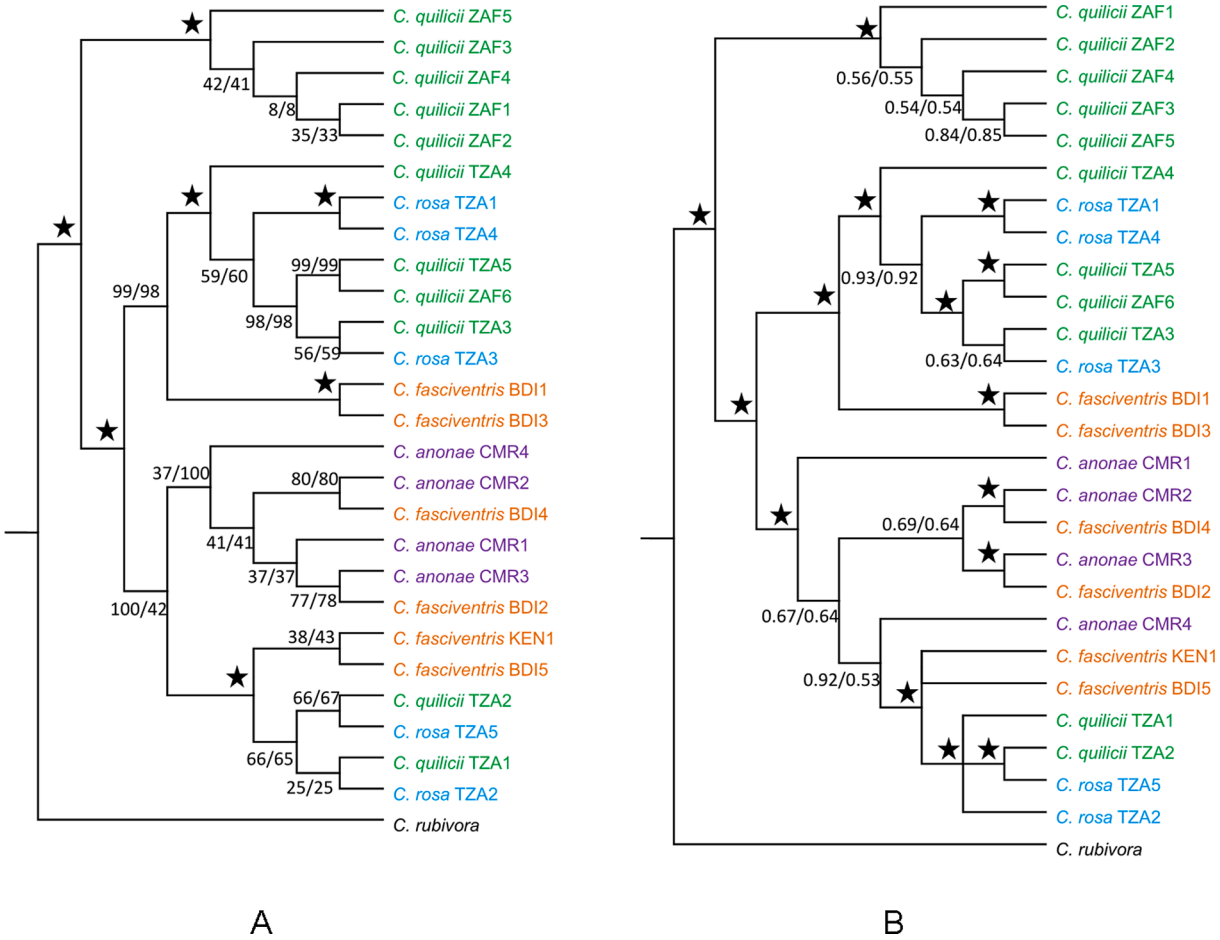

Fig. 3. Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees within the *Ceratitis* FARQ complex. *Ceratitis rubivora* was used as an outgroup based on RAxML (A) and MrBayes (B) analysis. Values above the nodes represent bootstrap values for RAxML using dataset 1: PCG123/dataset 2: PCG123 and 2 rRNAs in (A); Bayesian posterior probabilities for MrBayes using dataset 1: PCG123/dataset 2: PCG123 and 2 rRNAs in (B). '★' indicates posterior probabilities = 1.00 or ML bootstrap = 100 in all trees.

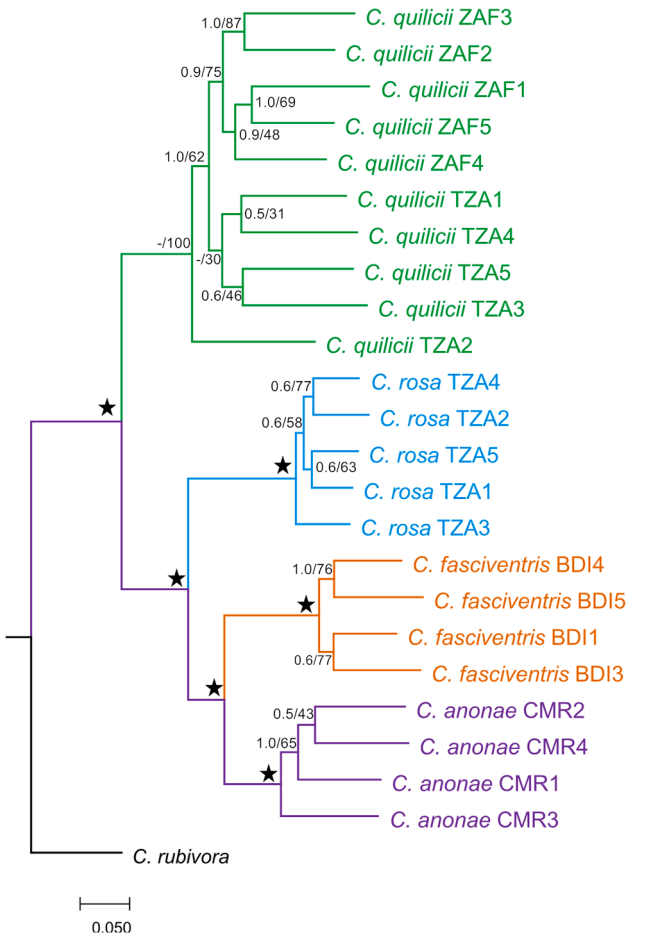

Fig. 4. Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees inferred from genome-wide SNPs within the *Ceratitis* FARQ complex. *Ceratitis rubivora* was used as an outgroup. Values above the nodes represent Bayesian posterior probabilities for MrBayes/bootstrap values for RAxML. '★' indicates posterior probabilities = 1.00 or ML bootstrap = 100 in both trees. '-' indicates nonsupported.

Barr and McPheron (2006), which did not include *C. punctata*, and with the cladistic analysis based on morphological characters and host plant data presented in De Meyer (2005) (which included *C. punctata*).

In our study, the four species belonging to the *Ceratitis* FARQ complex could be identified as distinct monophyletic groups only in the analysis of genome-wide SNPs, while the mitogenomic partition could not resolve them as separate taxonomic groups. These results suggest that incomplete lineage sorting or secondary introgression might have affected the inconsistent tree topologies obtained for the mitochondrial and nuclear partitions (Abbott et al., 2016, Weiss et al., 2017). The introgression between *C. quilicii* and *C. fasciventris* revealed by the TreeMix analyses is consistent with the patterns of interspecific gene flow described in Virgilio et al. (2013) on the basis of microsatellite data. Additionally, the lack of complete genetic isolation between species of the FARQ complex is also supported by the experimental data of Erbout et al. (2008), who showed that *C. fasciventris* and either *C. rosa* or *C. quilicii* (as these two species were identified as one before De Meyer et al. 2016) can generate fertile hybrids under laboratory conditions. It is also vital to continue interspecific mating experiments among species of the FARQ complex to obtain a better understanding of the gene flow discovered within the complex.

First, the analysis of mitogenomic data did not resolve the FARQ species as separate entities. In this respect, we need to consider both the strength and direction of the phylogenetic signals provided by the different partitions considered in this study. In fact, compared to mitochondria, the nuclear partition includes approximately one order of magnitude more data, so the nuclear SNPs might simply have provided a much stronger signal to the resolution of phylogenetic relationships of the FARQ complex. Second, it is well known that mitochondrial and nuclear partitions can provide conflicting signals as a consequence of recent or ongoing speciation (Battey and Klicka, 2017, Irwin et al., 2009, Thielsch et al., 2017, Wang et al., 2014), as maternally inherited mitochondrial genes are more prone to introgression and incomplete lineage sorting. Multilocus sequencing data, for example, genome-wide SNPs, can be a solution to provide unprecedented and accurate insights into species delimitation and the process of speciation (e.g., Esquerré et al., 2019, Hundsdoerfer et al., 2019, Mynhardt et al., 2020, Reyes-Velasco et al., 2020, Weiss et al., 2017). With the possible exception of the microsatellite-based population structure presented in Virgilio et al. (2013), this is the first time that the molecular phylogeny of the species of the FARQ complex can be profitably resolved by a genome-wide SNP study using HTS data. To resolve further the species delimitation controversy and to confirm the species status, a multiple 'species tree' method has also been proposed. SNAPP (Bryant et al., 2012) is widely used for inferring species trees from unlinked biallelic markers (Donkpegan et al., 2020, Gohli et al., 2015, Schmidt-Lebuhn et al., 2017, Stervander et al., 2016, Streicher et al., 2014, Wielstra et al., 2019). This method implements a full coalescent model, bypassing the gene trees and computing species tree likelihoods directly from the markers, integrating over all possible gene trees, rather than sampling them explicitly. Phylogenetic inference based on genomic SNP data concatenation in the coalescent-based species tree (SNAPP tree; Fig. 6) highly supported four major monophyletic clades as separate species within the complex and recovered the same topology as the gene tree (maximum likelihood and Bayesian inference phylogenetic trees; Fig. 4) as (*C. anonae* + *C. fasciventris*) + *C. rosa*) + *C. quilicii*. In the present study, the genomic SNP data corroborated the true status of the four species

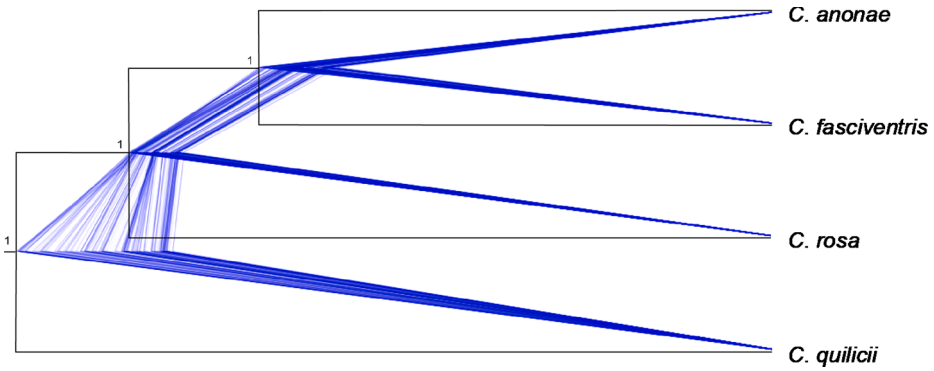

Fig. 5. Species tree estimation of the *Ceratitis* FARQ complex based on the Multi Species Coalescent Model as inferred by SNAPP and drawn in DensiTree. Support values on the nodes indicate SNAPP posterior probabilities.

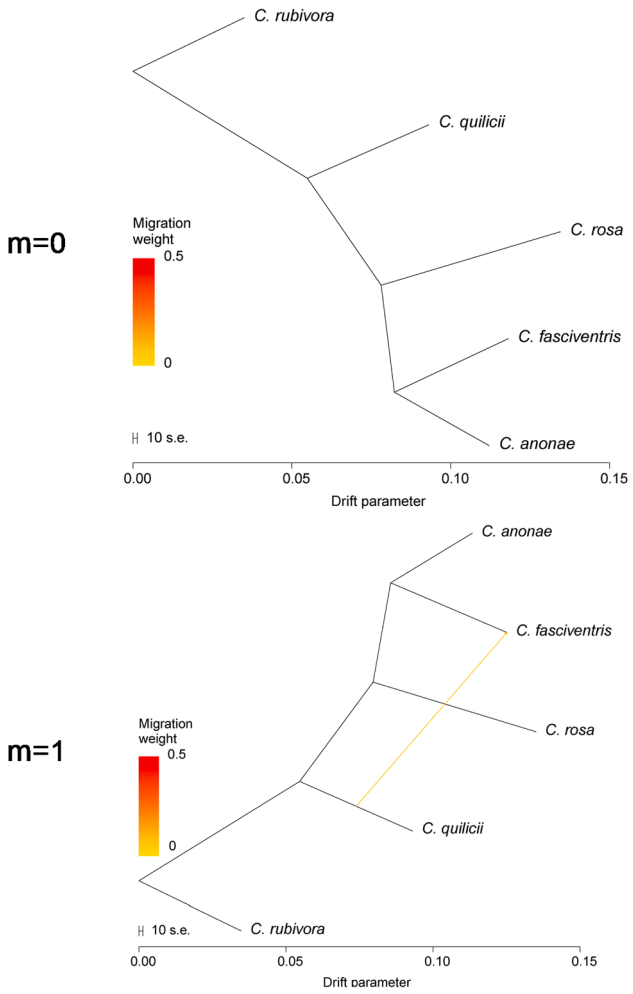

Fig. 6. Phylogenetic network of the genetic relationships among the *Ceratitis* FARQ complex with different possible migration events (m = 0, no migration event; m = 1, one migration event) inferred from TreeMix analysis. *C. rubivora* was used as an outgroup to root the tree. The graph shows the topology and branch lengths according to the drift parameter. Migration arrows were colored according to their weight. The scale bar shows 10 times the average standard error (s.e.) of the entries in the sample covariance matrix W.

within the FARQ complex by providing independent testing to demonstrate the capacity to resolve the deeper node phylogenetic inference although with few representative populations for each species. A coordinated research project on the resolution of cryptic tephritids of economic significance (Hendrichs et al. 2015) showed that, to some extent, the FARQ taxa could also be resolved on the basis of wing morphometrics (Van Cann et al., 2015), cuticular hydrocarbons (CHCs) (Vaníčková et al., 2014, Vaníčková et al., 2015), and larval morphological characters (Steck and Ekesi 2015). The integrative and multidisciplinary analysis of the FARQ species and the analysis of morphological differentiation of the midtibia in adult males resulted in the description of *C. quilicii* sp. nov. (De Meyer et al. 2016). The results of this study also suggest that reduced panels of diagnostic SNPs could be profitably used in the development of rapid and cost-effective tools for the molecular identification of FARQ species.

5. Conclusions

In this study, we used mitogenome and genome-wide SNPs to investigate the phylogenetic relationship within the *Ceratitis* FARQ complex. The mitochondrial genome failed while genome-wide SNPs made comprehensive contributions to resolve the *Ceratitis* FARQ taxa and provide insights into their phylogenetic relationships within the complex inferred by gene and species tree reconstructions. Gene flow was detected by TreeMix analysis from *C. quilicii* to *C. fasciventris*, providing the evidence of the existence of introgression events in the FARQ complex. Genome-wide SNPs detected from more representative geographical populations of each species was necessary to further explore the inter- and intraspecific evolutionary relationships within the complex.

CRediT authorship contribution statement

Yue Zhang: Data curation, Formal analysis, Methodology, Writing - original draft. **Marc De Meyer:** Methodology, Resources, Writing - review & editing. **Massimiliano Virgilio:** Methodology, Resources, Writing - review & editing. **Shiqian Feng:** Methodology, Writing - review & editing. **Kemo Badji:** Resources, Writing - review & editing. **Zhihong Li:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

We thank the Plant Quarantine and Invasion Biological Laboratory of China Agricultural University for providing the platform to conduct molecular experiments and data analysis. We also thank Dr. Ying Wang and Dr. Xudong Zhang from Genek for the guidance on bioinformatics analysis. This manuscript has been written within the collaborative framework of the EU H2020 FF-IPM project (project No. 818184).

Funding

This work was supported by National Key R&D Program of China (2016YFF0203200).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107160.

References

Abbott, R.J., Barton, N.H., Good, J.M., 2016. Genomics of hybridization and its evolutionary consequences. Mol. Ecol. 25, 2325–2332.
Barr, N.B., McPheron, B.A., 2006. Molecular phylogenetics of the genus *Ceratitis* (Diptera: Tephritidae). Mol. Phylogenet. Evol. 38 (1), 216–230.
Barr, N.B., Wiegmann, B.M., 2009. Phylogenetic relationships of *Ceratitis* fruit flies inferred from nuclear CAD and tango/ARNT gene fragments: Testing monophyly of the subgenera *Ceratitis (Ceratitis)* and C. *(Pterandrus)*. Mol. Phylogenet. Evol. 53 (2), 412–424.
Battey, C.J., Klicka, J., 2017. Cryptic speciation and gene flow in a migratory songbird species complex: Insights from the Red-eyed Vireo (*Vireo olivaceus*). Mol. Phylogenet. Evol. 113, 67–75.
Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69 (2), 313–319.
Bouckaert, R.R., Heled, J., 2014. DensiTree 2: Seeing trees through the forest. BioRxiv 012401.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., et al., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10 (4), e1003537.
Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29 (8), 1917–1932.
Burger, T.D., Shao, R., Barker, S.C., 2014. Phylogenetic analysis of mitochondrial genome sequences indicates that the cattle tick, *Rhipicephalus (Boophilus) microplus*, contains a cryptic species. Mol. Phylogenet. Evol. 76, 241–253.
Cameron, S.L., 2014. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Syst. Entomol. 39 (3), 400–411.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17 (4), 540–552.
Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34 (17), i884–i890.
Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22 (3), 252–271.
De Meyer, M., 2005. Phylogenetic relationships within the fruit fly genus *Ceratitis* MacLeay (Diptera: Tephritidae), derived from morphological and host plant evidence. Insect Syst. Evol. 36 (4), 459–479.
De Meyer, M., Freidberg, A., 2005. Revision of the subgenus *Ceratitis* (Pterandrus) Bezzi (Diptera: Tephritidae). Israel J. Entomol. 197–315.
De Meyer, M., Mwatawala, M., Copeland, R.S., Virgilio, M., 2016. Description of new *Ceratitis* species (Diptera: Tephritidae) from Africa, or how morphological and DNA data are complementary in discovering unknown species and matching sexes. Eur. J. Taxon. 233, 1–23.
Donkpegan, A.S., Doucet, J.L., Hardy, O.J., Heuertz, M., Piñeiro, R., 2020. Miocene diversification in the Savannahs precedes tetraploid rainforest radiation in the african tree genus *Afzelia* (Detarioideae, Fabaceae). Front. Plant. Sci. 11, 798.
Drosopoulou, E., Pantelidou, C., Gariou-Papalexiou, A., Augustinos, A.A., Chartomatsidou, T., Kyritsis, G.A., Bourtzis, K., Mavragani-Tsipidou, P., Zacharopoulou, A., 2017. The chromosomes and the mitogenome of *Ceratitis fasciventris* (Diptera: Tephritidae): two genetic approaches towards the *Ceratitis* FAR species complex resolution. Sci. Rep. 7 (1), 1–14.
Du, Z., Hasegawa, H., Cooley, J.R., Simon, C., Yoshimura, J., Cai, W., Sota, T., Li, H., 2019. Mitochondrial genomics reveals shared phylogeographic patterns and demographic history among three periodical cicada species groups. Mol. Biol. Evol. 36 (6), 1187–1200.
Dupuis, J.R., Brunet, B.M.T., Bird, H.M., Lumley, L.M., Fagua, G., Boyle, B., Levesque, R., Cusson, M., Powelle, J.A., Sperling, F.A.H., 2017. Genome-wide SNPs resolve phylogenetic relationships in the North American spruce budworm (*Choristoneura fumiferana*) species complex. Mol. Phylogenet. Evol. 111, 158–168.
Ekesi, S., De Meyer, M., Mohamed, S.A., Virgilio, M., Borgemeister, C., 2016. Taxonomy, ecology, and management of native and exotic fruit fly species in Africa. Annu. Rev. Entomol. 61, 219–238.
Erbout, N., De Meyer, M., Lens, L., 2008. Hybridization between two polyphagous fruit-fly species (Diptera: Tephritidae) causes sex-biased reduction in developmental stability. Biol. J. Linn. Soc. 93 (3), 579–588.
Esquerré, D., Ramírez-Álvarez, D., Pavón-Vázquez, C.J., Troncoso-Palacios, J., Garín, C. F., Keogh, J.S., Leaché, A.D., 2019. Speciation across mountains: Phylogenomics, species delimitation and taxonomy of the *Liolaemus leopardinus* clade (Squamata, Liolaemidae). Mol. Phylogenet. Evol. 139, 106524.
Fitak, R.R., 2020. optM: An R Package to Optimize the Number of Migration Edges Using Threshold Models. Available online: https://cran.r-project.org/web/packages/optM/ (accessed on 18 November 2020).
Gohli, J., Leder, E.H., Garcia-del-Rey, E., Johannessen, L.E., Johnsen, A., Laskemoen, T., Popp, M., Lifjeld, J.T., 2015. The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. Mol. Ecol. 24 (1), 180–191.
Feng, S., Yang, Q., Li, H., Song, F., Stejskal, V., Opit, G.P., Cai, W., Li, Z., Shao, R., 2018. The highly divergent mitochondrial genomes indicate that the booklouse, *Liposcelis bostrychophila* (Psocoptera: Liposcelididae) is a cryptic species. G3: Genes. Genom. Genet. 8 (3), 1039–1047.
Flomer, O., 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotech. 3, 294–299.
Hendrichs, J., Vera, M.T., De Meyer, M., Clarke, A.R., 2015. Resolving cryptic species complexes of major tephritid pests. ZooKeys. 540, 5.
Herd, K.E., Barker, S.C., Shao, R., 2015. The mitochondrial genome of the chimpanzee louse, *Pediculus schaeffi*: insights into the process of mitochondrial genome fragmentation in the blood-sucking lice of great apes. BMC Genomics 16 (1), 661.
Hundsdoerfer, A.K., Lee, K.M., Kitching, I.J., Mutanen, M., 2019. Genome-wide SNP Data Reveal an Overestimation of Species Diversity in a Group of Hawkmoths. Genome Biol. Evol. 11 (8).
Irwin, D.E., Rubtsov, A.S., Panov, E.N., 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). Biol. J. Linn. Soc. 98, 422–438.
Jia, W., Yan, H., Lou, Z., Ni, X., Dyachenko, V., Li, H., Littlewood, D.T.J., 2012. Mitochondrial genes and genomes support a cryptic species of tapeworm within *Taenia taeniaeformis*. Acta trop. 123 (3), 154–163.
Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20 (4), 1160–1166.
Kornilios, P., Thanou, E., Lymberakis, P., Ilgaz, Ç., Kumlutaş, Y., Leaché, A., 2020. A phylogenomic resolution for the taxonomy of Aegean green lizards. Zool. Scr. 49 (1), 14–27.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874.
Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29 (6), 1695–1701.
Lee, T.H., Guo, H., Wang, X., Kim, C., Paterson, A.H., 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15 (1), 162.
Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47 (W1), W256–W259.
Liu, J. X., Zhou, M. Y., Yang, G. Q., Zhang, Y. X., Ma, P. F., Vorontsova, M. S., Li, D. Z., 2020. ddRAD analyses reveal a credible phylogenetic relationship of the four main genera of Bambusa-Dendrocalamus-Gigantochloa complex (Poaceae: Bambusoideae). Mol. Phylogenet. Evol. 106758.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (14), 1754–1760.
Ma, C., Yang, P., Jiang, F., Chapuis, M.P., Shali, Y., Sword, G.A., Kang, L., 2012. Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. Mol. Ecol. 21 (17), 4344–4358.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20 (9), 1297–1303.
Meng, G., Li, Y., Yang, C., Liu, S., 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 47 (11), e63.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: In 2010 gateway computing environments workshop (GCE). Ieee, pp. 1–8.
Mynhardt, S., Bennett, N.C., Bloomer, P., 2020. New insights from RADseq data on differentiation in the Hottentot golden mole species complex from South Africa. Mol. Phylogenet. Evol. 143.
Papanicolaou, A., Schetelig, M.F., Arensburger, P., Atkinson, P.W., Benoit, J.B., 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. 17 (1), 192.
Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41 (3), 353–358.
Pickrell, J., Pritchard, J., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. Nature Precedings 1.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Mallerm, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81 (3), 559–575.
Rambaut, A., Drummond, A. J., 2007. TreeAnnotator. Program and documentation distributed by the author. Website http://beast.bio.ed.ac.uk/TreeAnnotator.
Rambaut, A., 2015. Figtree version 1.4. 0. http//tree.bio.ed.ac.uk/software/figtree/.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67 (5), 901.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574.
Reyes-Velasco, J., Adams, R. H., Boissinot, S., Parkinson, C. L., Campbell, J. A., Castoe, T. A., Smith, E. N., 2020. Genome-wide SNPs clarify lineage diversity confused by coloration in coralsnakes of the Micrurus diastema species complex (Serpentes: Elapidae). Mol. Phylogenet. Evol. 147.
Schmidt-Lebuhn, A.N., Aitken, N.C., Chuah, A., 2017. Species trees from consensus single nucleotide polymorphism (SNP) data: Testing phylogenetic approaches with simulated and empirical data. Mol. Phylogenet. Evol. 116, 192–201.
Spanos, L., Koutroumbas, G., Kotsyfakis, M., Louis, C., 2000. The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. Insect Mol. Biol. 9 (2), 139–144.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.
Steck, G.J., Ekesi, S., 2015. Description of third instar larvae of *Ceratitis fasciventris*, *C. anonae*, *C. rosa* (FAR complex) and *C. capitata* (Diptera Tephritidae). ZooKeys 540, 443.
Stervander, M., Alström, P., Olsson, U., Ottosson, U., Hansson, B., Bensch, S., 2016. Multiple instances of paraphyletic species and cryptic taxa revealed by mitochondrial and nuclear RAD data for Calandrella larks (Aves: Alaudidae). Mol. Phylogenet. Evol. 102, 233–245.
Streicher, J.W., Devitt, T.J., Goldberg, C.S., Malone, J.H., Blackmon, H., Fujita, M.K., 2014. Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. Mol. Ecol. 23 (13), 3273–3291.
Su, Y., Zhang, Y., Feng, S., He, J., Zhao, Z., Bai, Z., Liu, L., Zhang, R., Li, Z., 2017. The mitochondrial genome of the wolfberry fruit fly, *Neoceratitis asiatica* (Becker) (Diptera: Tephritidae) and the phylogeny of *Neoceratitis* Hendel genus. Sci. Rep. 7 (1), 1–8.
Tanga, C.M., Khamis, F.M., Tonnang, H.E.Z., Rwomushana, I., Ekesi, S., 2018. Risk assessment and spread of the potentially invasive *Ceratitis rosa* Karsch and *Ceratitis quilicii* De Meyer, Mwatawala & Virgilio sp. Nov. using life-cycle simulation models: Implications for phytosanitary measures and management. PLoS ONE 13 (1).
Thielsch, A., Knell, A., Mohammadyari, A., Petrusek, A., Schwenk, K., 2017. Divergent clades or cryptic species? Mito-nuclear discordance in a Daphnia species complex. BMC Evol. Biol. 17 (1), 227.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Torricelli, G., Carapelli, A., Convey, P., Nardi, F., Boore, J.L., Frati, F., 2010. High divergence across the whole mitochondrial genome in the "pan-Antarctic" springtail *Friesea grisea*: evidence for cryptic species? Gene 449 (1–2), 30–40.
Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27 (2), 171–180.
Van Cann, J., Virgilio, M., Jordaens, K., De Meyer, M., 2015. Wing morphometrics as a possible tool for the diagnosis of the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera, Tephritidae). ZooKeys 540, 489.
Vaníčková, L., Virgilio, M., Tomčala, A., Břízová, R., Ekesi, S., Hoskovec, M., Kalinová, B., Do Nascimento, R. R., De Meyer, M., 2014. Resolution of three cryptic agricultural pests (Ceratitis fasciventris, C. anonae, C. rosa, Diptera: Tephritidae) using cuticular hydrocarbon profiling. B. Entomol. Res. 104(5), 631-638.

Vaníčková, L., Břízová, R., Pompeiano, A., Ekesi, S., De Meyer, M., 2015. Cuticular hydrocarbons corroborate the distinction between lowland and highland Natal fruit fly (Tephritidae, *Ceratitis rosa*) populations. ZooKeys. 540, 507.

Virgilio, M., Backeljau, T., Barr, N., De Meyer, M., 2008. Molecular evaluation of nominal species in the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera: Tephritidae). Mol. Phylogenet. Evol. 48 (1), 270–280.

Virgilio, M., Delatte, H., Quilici, S., Backeljau, T., De Meyer, M., 2013. Cryptic diversity and gene flow among three African agricultural pests: *Ceratitis rosa*, *Ceratitis fasciventris* and *Ceratitis anonae* (Diptera, Tephritidae). Mol. Ecol. 22 (9), 2526–2539.

Virgilio, M., White, I., De Meyer, M., 2014. A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae). ZooKeys. 428, 97.

Virgilio, M., Daneel, J. H., Manrakhan, A., Delatte, H., Meganck, K., De Meyer, M., 2019. An integrated diagnostic setup for the morphological and molecular identification of the Ceratitis FAR complex (C. anonae, C. fasciventris, C. rosa, C. quilicii, Diptera, Tephritidae). B. Entomol. Res. 109(3), 376-382.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38 (16), e164.

Wang, W., Dai, C., Alström, P., Zhang, C., Qu, Y., Li, S.H., Yang, X., Zhao, N., Song, G., Lei, F., 2014. Past hybridization between two East Asian long-tailed tits (Aegithalos bonvaloti and *A. fuliginosus*). Front. Zool. 11, 40.

Wang, Y., Liu, X., Winterton, S.L., Yan, Y., Chang, W., Yang, D., 2013. Comparative mitogenomic analysis reveals sexual dimorphism in a rare montane lacewing (Insecta: Neuroptera: Ithonidae). PLoS ONE 8 (12).

Weiss, M., Weigand, H., Weigand, A.M., Leese, F., 2017. Genome-wide single-nucleotide polymorphism data reveal cryptic species within cryptic freshwater snail species-the case of the *Ancylus fluviatilis* species complex. Ecol. Evol. 8 (5), 1063–1072.

Wielstra, B., McCartney-Melstad, E., Arntzen, J.W., Butlin, R.K., Shaffer, H.B., 2019. Phylogenomics of the adaptive radiation of Triturus newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle. Mol. Phylogenet. Evol. 133, 120–127.