diversity

MDPI

Article

Nuclear Orthologs Derived from Whole Genome Sequencing Indicate Cryptic Diversity in the *Bemisia tabaci* (Insecta: Aleyrodidae) Complex of Whiteflies

Robert S. de Moya [1,2,*], Judith K. Brown [3], Andrew D. Sweet [4], Kimberly K. O. Walden [1], Jorge R. Paredes-Montero [3,5], Robert M. Waterhouse [6] and Kevin P. Johnson [2]

1 Department of Entomology, University of Illinois Urbana-Champaign, 505 S. Goodwin Ave., Urbana, IL 61801, USA
2 Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820, USA
3 School of Plant Sciences, University of Arizona, 1140 E. South Campus Dr., Tucson, AZ 85721, USA
4 Department of Entomology, Purdue University, 901 W. State St., West Lafayette, IN 47907, USA
5 Facultad de Ciencias de la Vida, Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
6 Department of Ecology and Evolution, University of Lausanne, and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
* Correspondence: rdemoya2@illinois.edu; Tel.: +1-602-694-5505

Received: 15 July 2019; Accepted: 24 August 2019; Published: 29 August 2019

Abstract: The *Bemisia tabaci* complex of whiteflies contains globally important pests thought to contain cryptic species corresponding to geographically structured phylogenetic clades. Although mostly morphologically indistinguishable, differences have been shown to exist among populations in behavior, plant virus vector capacity, ability to hybridize, and DNA sequence divergence. These differences allow for certain populations to become invasive and cause great economic damage in a monoculture setting. Although high mitochondrial DNA divergences have been reported between putative conspecifics of the *B. tabaci* species complex, there is limited data that exists across the whole genome for this group. Using data from 2184 orthologs obtained from whole genome sequencing (Illumina), a phylogenetic analysis using maximum likelihood and coalescent methodologies was completed on ten individuals of the *B. tabaci* complex. In addition, automatic barcode gap discovery methods were employed, and results suggest the existence of five species. Although the divergences of the mitochondrial cytochrome oxidase I gene are high among members of this complex, nuclear divergences are much lower in comparison. Single-copy orthologs from whole genome sequencing demonstrate divergent population structures among members of the *B. tabaci* complex and the sequences provide an important resource to aid in future genomic studies of the group.

Keywords: phylogenomics; hemiptera; read-mapping; cryptic species; pests

1. Introduction:

The *Bemisia tabaci* (Gennadius 1889) (Insecta: Hemiptera: Aleyrodidae) complex of whiteflies has collectively colonized over 600 plant species and includes globally important pests [1,2]. Members of the lineages within this complex have been referred to as races, strains, and biotypes [3–6], or haplotypes and mitotypes [7,8]. Globally, only a few of these genetic variants are known to cause annual economic losses [9–11], whereas most lineages are benign in their native habitats in the absence of irrigated monoculture-agriculture [7,10,12]. Originally described from cultivated tobacco plants in Greece, *B. tabaci* has been defined to include numerous populations that inhabit mostly tropical/subtropical and climatically moderate temperate regions [1,7,8,13,14]. Despite the accumulated evidence of behavioral

differences among selected lineages [3–5,15–18], all but one of these populations have been shown to be morphologically indistinguishable [8,19,20].

Molecular and population studies have demonstrated extensive molecular divergence between different lineages of *B. tabaci* occurring throughout the world [7,8,21–23]. Members of the *B. tabaci* complex can be divided between biogeographic regions into major clades based on mitochondrial cytochrome oxidase I (COI) gene divergences, which can range as high as 24% among clades [7,24]. The latter observation provides the basis for the hypothesis that *B. tabaci* comprises a complex of distinct, mostly cryptic species [7,14,17,21,22,25]. The question surrounding the number of species and their nomenclature has been controversial, with some research groups using a small fragment of the 3′-end of the COI gene (~600 bp) to delineate species boundaries that gave rise to 28 species [26,27]. While the genetic structure has been investigated for selected *B. tabaci* populations [22,28,29], the majority of studies have relied upon comparisons of a partial fragment (658–780 bp) of the COI gene [21,26,30–33]. Mitochondrial sequences evolve at a relatively high rate and have a low effective population size [34,35], which can be useful for understanding gene flow between populations [36–38] and for inferring past demographic events that have shaped the genetic diversity of the complex [39,40]. However, introgression of mitochondrial sequences into different nuclear phylogenetic backgrounds is possible [41], making it important to understand if lineages identified in mitochondrial data are also reflected across the nuclear genome. Only a few studies have extensively sampled nuclear data [22,42,43] for members of the *B. tabaci* complex, and whole genomic data sets have been lacking, precluding the analyses of important genome-level questions.

Numerous biotypes from the *B. tabaci* complex of whiteflies have been described that can be analyzed in a phylogenomic context. Although many biotypes have been described, not all of them may be evolutionary significant lineages or cryptic species. A few biotypes are featured prominently in the literature. For example, the "B" and "Q" biotypes likely have a similar Middle East/North African origin and have become highly invasive in a monoculture setting [7]. The "B" and "Q" types are thought to be a responsible for the displacement of the "A" biotype that was once widespread across North America [44]. Another significant population of whiteflies includes individuals from the sub-Saharan African region and have become an intense focus for scientific studies because of their detrimental impacts on cassava crops [45,46]. Attempts have been made to summarize the many biotypes [7]; however new types are often described and it is difficult to summarize these under a single framework. A review suggests that seven major phylogeographic lineages exist that could be distinct species [7] and the analyses performed in this current study samples six of these major phylogeographic clades. Unfortunately, there exist no international rules for zoological nomenclature regarding the description and reference to biotypes in scientific literature, therefore we refer to major lineages for our current study in the context of geographical acronyms suggested by a previous publication [7].

Next generation sequencing technologies provide a cost-effective method to obtain data for thousands of genes across the nuclear genome [47,48]. Extensive sequence data representing much of the genome is expected to provide needed information at the nuclear genome level to complement the insights gained from published mitochondrial COI datasets. A genome of the "B" biotype [4], has an estimated size of 615 Mbp [49,50]. Thus, the size of the *B. tabaci* genome makes it possible to obtain sequence coverage across the genome using Illumina sequencing technologies and shotgun sequencing at modest cost. This annotated *B. tabaci* genome was also used as a reference for read-mapping among members of the *B. tabaci* species complex to obtain sequences of single-copy orthologs.

To investigate the extent of genomic divergence among lineages of *B. tabaci*, the whole genome was sequenced using the Illumina sequencing platform for nine individuals of the *B. tabaci* complex. Members of the complex were selected to span a broad biogeographic representation of major clades previously recognized [7]. Read-mapping was used to obtain gene sequences for 2184 single-copy orthologs. The resultant data sets were analyzed using phylogenomic approaches (coalescence and maximum likelihood) to assess the concordance between nuclear and mitochondrial datasets and to evaluate overall divergence among major lineages. To assess the stability of ambiguous nodes,

likelihood mapping [51] and quartet sampling [52] were employed to demonstrate the frequency of discordant topologies produced from the concatenated dataset. Distance methods were also employed to compare sequence divergence from the mitochondrial and nuclear data obtained from read-mapping [53].

Automatic barcode gap discovery (ABGD) methods [54] were used to test for the existence of putative species within nuclear and mitochondrial alignments. ABGD methods use barcode gaps that exist between interspecific and intraspecific sequence datasets to assess the existence of unknown species diversity within an alignment [54]. These methods became popular because of the use of COI as a barcode to assess unknown diversity across a wide range of animals. This method has also been used to test for the presence of operation taxonomic units in a phylogenetic dataset containing possible cryptic species [55]. Although this method has traditionally been used on relatively small amounts of sequence data (i.e., COI alignments), the analysis is computationally efficient and has the ability to analyze large amounts of genomic data quickly. The analyses are based on the realistic observation that gaps exist between interspecific and intraspecific sequence divergences [54]. Thus, ABGD analyses have the ability to efficiently quantify these observed gaps and provide an estimate of putative species in genomic datasets.

The goal of this study was to analyze Illumina whole genome data to understand possible species boundaries within the *B. tabaci* complex of whiteflies. This was done through a series of maximum likelihood, coalescent, likelihood mapping, quartet sampling, and ABGD analyses. The genomic analyses of 2,184 genes provide further evidence to complement previous studies that indicate multiple cryptic species exist within the *B. tabaci* complex of whiteflies.

2. Materials and Methods

2.1. Sample Preparation and Sequencing

Based on partial mitochondrial COI DNA sequences, members of the *B. tabaci* complex have been shown to form phylogeographic clades of seven major groups (Table 1) [7]. A range of genetic diversity exists within each group, but divergences among groups are high, potentially marking species boundaries [32]. Nine *B. tabaci* individuals belonging to six phylogeographic groups, previously identified based mostly on COI divergences (Table 1: mitochondrial major clades I-VI) [7], were selected for sequencing. Three whitefly samples were each collected from the Democratic Republic of Congo (Lulimba), Tanzania (Dar es-Salam), and Uganda (Bukoba), respectively, to represent the Sub-Saharan Africa (SSA) clade. One sample from Sudan (Wad Mani) was selected as the representative of the North Africa–Mediterranean–Middle East (NAF–MED–ME) clade. For the Asian–Pacific–Australia (AS–PAC–AU) region, a representative was collected from Hainan, China. To serve as a representative of the Asia major clade, an individual from Ahmedabad in India was selected. To represent the American Tropics/Caribbean (AM-TROP) clade, one each were selected from Puerto Rico, Arizona, and Ecuador (Table 2). A geographical acronym is applied to each phylogeographical major clade [7] and used throughout to refer to the *B. tabaci* lineages (Tables 1 and 2). One adult *B. afer* individual was sequenced and included as the outgroup.

Whitefly adults were collected alive from plants and placed directly into vials containing either 70 or 95% ethanol. Specimens were shipped to the University of Arizona, and immediately transferred to 95% ethanol for storage at −20 °C. Individual adult whiteflies were ground in 1.5 µL tubes. The DNA was extracted by a Qiagen DNA Micro Kit (Qiagen, Valencia, CA, USA) using a 48-h incubation period, and the DNA was suspended in 52 µl of elution buffer. Extracted genomic DNA was quantified with a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) using the manufacturer's recommended protocol and reagents.

Paired-end shotgun genomic libraries were prepared with Hyper Library construction kits from Kapa Biosystems (Roche). The *B. tabaci* libraries were quantified by qPCR and sequenced for 151 cycles from each end of the fragments on an Illumina HiSeq 4000 machine using a HiSeq 4000 sequencing

kit version 1 or a NovaSeq 6000 machine. Paired-end reads were 150 nt in length for the *B. tabaci* specimens. The *B. afer* library was quantified by qPCR and sequenced on one lane for 161 cycles from each end of the fragments on a HiSeq2500 using a HiSeq SBS sequencing kit version 4. Paired-end reads were 160 nt in length for the *B. afer*. Multiplexing was performed to achieve between 240 and 400 Gbp of data per sample. Library preparation and sequencing took place at the W.M. Keck Center (University of Illinois, Urbana, IL, USA). The FASTQ files were generated and demultiplexed with the bcl2fastq v 2.17.1.14 or 2.20 Conversion Software (Illumina). Raw reads were deposited in the sequence read archive (SRA) of the National Center for Biotechnology Information (NCBI) database. Adaptors were trimmed and sequencing quality reports were generated with FastQC v 0.11.7 [71].

Table 1. Classification of the *B. tabaci* sibling species group based on nuclear (lineages), mitochondrial (major clades), and esterase/RAPD-PCR patterns (biotype).

Nuclear Lineages (Current)		Mitochondrial Major Clades [a]		Biotypes [d]
Name	**Acronym**	**Number**	**Name**	
Sub-Saharan Africa	SSA	I	Sub-Saharan Africa—East and South	S
		II	Sub-Saharan Africa—West	
North Africa–Mediterranean–Middle East	NAF–MED–ME	III	North Africa–Mediterranean–Middle East	B, B1, B2 J, L, Q Non-Ug/MS
Asia 2	ASIA	IV	Asia 2	K, P, ZHJ2 ZHJ1 G Cv, I
Asia–Pacific–Australia	AS–PAC–AU	V	Asia–Pacific Islands–Australia	H, M, Na (Nauru), An ZHJ3
American Tropics: North-Central America and Caribbean	AM-TROP	VI	American Tropics: North and Central/ Caribbean	A, A1, A2, C, D, F, O, N, R
		VII	American Tropics: South America [b]	-
Unassigned groups [c]:			Uganda Italy Benin	- T E

[a] Major clade nomenclature based on published classification of the *B. tabaci* species complex [7]. [b] A representative whitefly from this clade was removed due to low-quality alignments. [c] Unstudied groups that represent potential new lineages. [d] Biotypes A–Q [5,20,56]. Biotype O described [57], but later referred to as biotype N [56]. Biotype R identified by Ian Bedford [58]. Biotype S [59]. Biotype T identified in Nebrodi, Italy [60] and further characterized [61]. The MS 'biotype' was identified based on RAPD-PCR patters and COI analysis [62]. Biotypes An and Na from Australia and the Pacific Islands were designated based on RAPD profiles [63]. The biotype Cv [64–66]. The ZHJ1 and ZHJ2 [67–69]. The ZHJ1-3 biotypes [70].

Table 2. Summary of collection information and genomic coverage for each individual sequenced. The last column (% Pairwise Identity) corresponds to the best BLAST hit match from the Megablast search of COI data obtained from read-mapping. Plant host indicates the scientific name of the plant the sample was collected. The color of rows corresponds to the putative species suggested by ABGD analyses.

Geographical Acronym	Country	State/Province	City	Plant Host	Reads	Gbp	Coverage (X)	Mapped Genes	Aligned Genes	Aligned bp	Best Blast Match	% Pairwise Identity
AM-TROP	Puerto Rico	-	-	*Jatropha gossypifolia*	106,701,526	320	52	2192	2183	3,662,977	KX397317	99.7
AM-TROP	Ecuador	Guayas	Isla Puná	*Datura stramonium*	80,297,211	241	39	2192	2183	3,669,262	EU427728	98.6
AM-TROP	United States	Arizona	Tucson	*Gossypium hirsutum*	86,382,401	259	42	2192	2183	3,672,677	AY521259	100
NAF–MED–ME	Sudan	Al Jazirah	Wad Mani	*Phaseolus vulgaris*	94,837,626	285	46	2193	2184	3,672,242	FJ188567	99.0
SSA	DR Congo	-	Lulimba	*Manihot esculenta*	92,877,327	279	45	2193	2184	3,668,485	AF344246	99.7
SSA	Uganda	Kajera	Bukoba	*Manihot esculenta*	133,240,668	400	65	2193	2184	3,668,433	AM040604	99.7
SSA	Tanzania	Dar es-Salam	Dar es-Salam	*Manihot esculenta*	103,951,478	312	51	2193	2184	3,671,055	JQ286457	99.7
AS–PAC–AU	China	Hainan	Hainan	*Gossypium hirsutum*	97,765,484	293	48	2193	2184	3,672,435	HG918196	99.8
ASIA	India	Ahmedabad	New Delhi	*Nicotiana tabacum*	103,114,416	309	50	2193	2184	3,671,735	KF751570	100
B. afer	Australia	New South Wales	Sydney	*Hardenbergia sp.*	86,662,872	277	45	1938	1929	3,002,554	AJ784260	87.9

2.2. Ortholog Prediction and Phylogenomic Analyses

The genome of the "B" biotype has previously been sequenced, assembled, and annotated [49,50]. We used version 1.1 of the genome annotation as a reference sequence for read-mapping of orthologs in newly sequenced individuals [50]. Using the OrthoDB v7 database [72] to subsequently serve as references for read-mapping, 2193 single-copy orthologs were identified. This ortholog set was previously developed for phylogenomics of hemipteroid insects [73]. We read-mapped our newly sequenced *B. tabaci* genomes against the 2193 reference orthologs using Bowtie2 v 2.3.4.1 [74] and SamTools v 1.7 [75], following a general read-mapping pipeline previously published [55]. The sequence alignment map (SAM) files produced by Bowtie2 were compressed to their binary versions (BAM), which were further sorted and indexed using SamTools. All mapped reads were retained in the BAM files by using the -F 4 option in SamTools. Pileup files were created from the BAM files with the mpileup command in SamTools using a maximum depth of 75, minimum base quality of 28, and minimum mapping quality of 10. The pileup files were then converted to VCF and FASTQ files using BCFtools and vcfutils.pl respectively. The same read-mapping pipeline was used for obtaining the sequences of the mitochondrial COI gene for each individual. The single-copy orthologs and COI data from the "B" reference genome used for read-mapping were included in the final alignments and phylogenetic analyses. The ortholog sequences for each individual genome were aligned with PASTA v 1.8.0 by nucleotide sequence using an increased memory capacity of 2048 MB [76]. Alignments were then masked using a 40% gap threshold with trimAl v 1.4 [77]. A concatenated supermatrix of aligned genes (orthologs) was produced with Sequence Matrix v 100.0 [78]. An uncorrected distance matrix was produced from the final concatenated supermatrix of 2184 genes using PAUP* v 4.0 [53].

The supermatrix of aligned nucleotides was analyzed using several different methods for comparison. A concatenated maximum likelihood (ML) analysis was completed with RAxML v 8.2.11 using a GTR + G model of evolution and 100 rapid bootstrap replicates [79]. RAxML does not implement simpler than GTR models of evolution and invariant sites were not considered due to computational non-identifiability between the alpha parameter and gamma distribution. An optimal partitioning scheme was determined with PartitionFinder v 2.1.1 with branch lengths linked, GTR + G model, BIC model selection, rcluster search, and rcluster-max 100 [80]. The resulting partitioning scheme was implemented in RAxML using a GTR + G model and 100 rapid bootstrap replicates.

Because relationships among species could be confounded by incomplete lineage sorting among different genes, a coalescent approach was also implemented to compute a species tree from the individual gene trees [81–83]. All 2184 gene alignments were each individually analyzed in RAxML with a GTR + G model and 100 bootstrap replicates. These derived bipartition trees were summarized in a coalescent analysis with ASTRAL v 5.5.9 using default parameters, and clade support assessed with local posterior probability [81,83]. A coalescent analysis was also performed with ASTRID v 1.4 using multi-locus bootstrapping clade support and input files from concatenated RAxML bipartition files [82].

To provide a comparison with data from the nuclear gene sequences, a 1466 bp fragment of COI from the "Q" biotype (Genbank accession: KJ591614.1) was used as a reference for read-mapping to obtain the COI sequences for each of the genomes that were sequenced. In addition, reads from the *B. afer* library were mapped against a published *B. afer* mitochondrial genome to obtain the COI sequence for this species (Genbank accession: KR819174.1). To verify the identity of each genetic variant, the resulting COI sequences obtained from read-mapping were searched with BLAST against NCBI using Megablast, implemented in Geneious v 11.1.3 (Biomatters, Inc., Newark, NJ, USA) [84–86]. The BLAST results were evaluated based on best hit by percent sequence identity with ties broken by sequence length. The COI sequence from the "B" reference genome [50] (Zhangjun Fei, Boyce Thompson Institute, Cornell University) was included for subsequent ML analyses. The COI sequences were aligned with ClustalW v 2.1 [87] using the default parameters as set in Geneious [86]. The alignment was trimmed to a length of 1459 bp. A ML analysis of the COI data was completed with RAxML using a GTR + G model and 1000 rapid bootstrap replicates. Uncorrected pairwise distances were calculated in PAUP* for the COI alignment (1459 bp). A second ML analysis was performed utilizing a

publicly available dataset of COI data [88]. The available data set was aligned with the mitochondrial sequences obtained from read-mapping. This alignment was then trimmed to a length of 696 bp and a ML analysis was performed with RAxML using parameters as previously described.

To assess frequencies of discordant topologies that are produced from the data set, quartet sampling [52] and likelihood mapping [51] was performed. Quartet sampling methods were implemented using a published available script [52] utilizing the concatenated nucleotide alignment and resulting unrooted nucleotide topology as an input tree for analyses. The number of replicates per branch were set to 200 and the log-likelihood cutoff was set to 2, as suggested by published protocol [52]. Likelihood mapping was performed in IQ-Tree v 1.6.5. [89] using the concatenated nucleotide alignment as the input. All number of quartets possible were drawn, subsequent tree searches were skipped, and a GTR + G model was implemented. The nexus file containing taxon clusters were defined as the following: cluster 1 = *B. afer*, cluster 2 = reference "B" reference genome, cluster 3 = *B. tabaci* sampled from Sudan, and cluster 4 = all other ingroup whiteflies sampled.

As a test for putative species an ABGD test [54] was implemented for both the mitochondrial and nuclear data sets. First the test was conducted using the COI alignment under both Jukes-Cantor (JC69) and simple (i.e., uncorrected) distances. The relative gap width (X) was set to 1 because the program failed to run with the default setting ($X = 1.5$) for the COI data. Remaining settings for the COI analysis were set to the default. Secondly, the ABGD analysis was implemented on the nuclear supermatrix of single-copy orthologs using both JC69 and simple distances with default settings. The *B. afer* specimen was excluded from the alignment for ABGD analyses because of its high divergence from the ingroup.

3. Results

Estimated sequence coverage of the genome for all samples was 39–65X based on calculations using an estimated draft genome size of 615 Mb [49,50], number of reads, and length of reads (Table 2). Read-mapping produced consensus sequences for nearly all of the 2193 single-copy orthologs for members of *B. tabaci* complex, but only 1929 genes mapped using the more divergent *B. afer* genome (Table 2). From the read-mapping results, 2184 genes successfully aligned in PASTA. The final nucleotide supermatrix consisted of 3,673,822 bp of data (109,699 informative). Uncorrected distances from the complete nuclear gene alignment ranged from 0.12%–2.50% between members of the *B. tabaci* complex (Table 3A).

The results of phylogenomic analyses based on the 2184 gene data set were mostly consistent but with some differences found between concatenated ML and coalescent analyses. Samples from Sub-Saharan Africa (SSA) and the American Tropics (AM-TROP), respectively, each form separate monophyletic clades with high support across all analyses (Figure 1a–d). Only one individual each were sampled from AS–PAC–AU region and from the ASIA region, but together are recovered as a highly supported monophyletic lineage across all analyses. Results of concatenated ML analyses recovered that the "B" reference genome was sister to the remaining ingroup *B. tabaci* individuals (Figure 1a,b), but the coalescent (ASTRAL and ASTRID) analyses recovered a sister relationship between the "B" reference genome and Sudan individual (Figure 1c and Figure S1). The coalescent analyses of the whole genome data set are the only phylogenomic analyses supporting a monophyletic NAF–MED–ME group which includes the "B" reference genome and the Sudan individual sampled.

Table 3. Colors of rows correspond to putative species suggested by ABGD analyses. A) Uncorrected pairwise distances among *B. tabaci* individuals based on the nuclear supermatrix of 2184 genes. B) Uncorrected pairwise distances among individuals based on COI data.

(A)	Geographical Acronym	Specimen	1	2	3	4	5	6	7	8	9	10
1		*B. afer*	-									
2	NAF-MED-ME	SRS1302125	0.0755	-								
3	NAF-MED-ME	Sudan	0.0757	0.0057	-							
4	AM-TROP	Arizona	0.0805	0.0164	0.0138	-						
5	AM-TROP	Ecuador	0.0806	0.0163	0.0138	0.0012	-					
6	AM-TROP	Puerto Rico	0.0808	0.0159	0.0136	0.0023	0.0023	-				
7	SSA	DR Congo	0.0808	0.0223	0.0199	0.0221	0.0221	0.0222	-			
8	SSA	Tanzania	0.0830	0.0250	0.0225	0.0246	0.0247	0.0249	0.0051	-		
9	SSA	Uganda	0.0808	0.0221	0.0198	0.0221	0.0221	0.0221	0.0031	0.0052	-	
10	AS–PAC–AU	China	0.0799	0.0161	0.0135	0.0152	0.0153	0.0154	0.0215	0.0241	0.0216	-
11	ASIA	India	0.0791	0.0150	0.0125	0.0143	0.0143	0.0145	0.0205	0.0231	0.0206	0.0093
(B)	**Geographical Acronym**	**Specimen**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1		*B. afer*	-									
2	NAF–MED–ME	SRS1302125	0.2200	-								
3	NAF–MED–ME	Sudan	0.2180	0.0637	-							
4	AM-TROP	Arizona	0.2229	0.1439	0.1508	-						
5	AM-TROP	Ecuador	0.2180	0.1412	0.1487	0.0171	-					
6	AM-TROP	Puerto Rico	0.2229	0.1426	0.1501	0.0178	0.0199	-				
7	SSA	DR Congo	0.2282	0.1594	0.1552	0.1588	0.1602	0.1581	-			
8	SSA	Tanzania	0.2193	0.1489	0.1579	0.1531	0.1545	0.1511	0.0825	-		
9	SSA	Uganda	0.2249	0.1524	0.1627	0.1592	0.1606	0.1572	0.0824	0.0131	-	
10	AS–PAC–AU	China	0.2312	0.1377	0.1412	0.1410	0.1348	0.1403	0.1594	0.1646	0.1681	-
11	ASIA	India	0.2260	0.1419	0.1467	0.1597	0.1576	0.1604	0.1695	0.1681	0.1736	0.1317

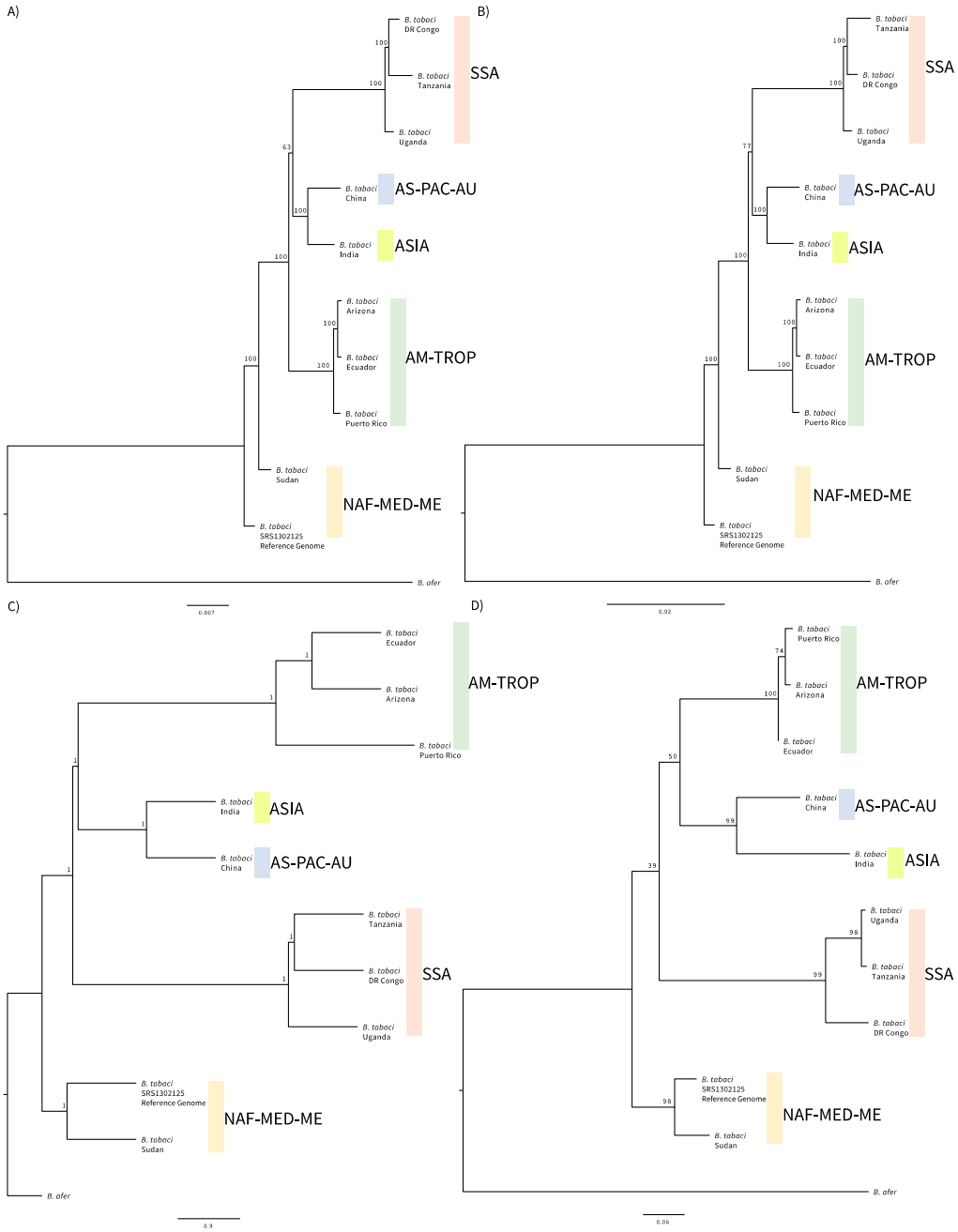

Figure 1. Colors correspond results suggested by automatic barcode gap discovery (ABGD) tests for putative species. (**A**) Result of the concatenated maximum likelihood (ML) analysis of all nucleotide sites from the nuclear supermatrix (lnLikelihood = –6919264.24). Scale bar represents substitutions per site. Clade values are depicted as bootstrap support. (**B**) Result of the partitioned ML analysis of the nuclear supermatrix (lnLikelihood = –6907791.99). Scale bar represents substitutions per site. Clade values are depicted as bootstrap support. (**C**) Result of ASTRAL coalescent analysis of 2184 gene trees. Scale bar represents coalescent units. Clade values are depicted as local posterior probabilities. (**D**) Result of ML analysis of cytochrome oxidase I (COI) sequences obtained from read-mapping (lnLikelihood = –6174.94). Scale bar represents substitutions per site. Clade values are depicted as bootstrap support.

The phylogenetic analysis based on COI sequences was largely concordant with the nuclear tree. However, node support for relationships among these major groups was low, as might be expected from a single, rapidly evolving gene (Figure 1d). When our read-mapped COI sequences were analyzed in a ML framework with publicly available data [69], individuals sampled were recovered within clades of

similar geographic origin (Figure S2). BLAST results also validate read-mapped COI data obtained and the best BLAST matches are reported (Table 2). The uncorrected distances for COI sequences within the *B. tabaci* complex ranged from 1.31–17.0% among the SSA, NAF–MED–ME, ASIA, AS–PAC–AU, and AM-TROP lineages (Table 3B). In general, divergence in COI sequences increased with increasing nuclear divergence (Figure 2). However, COI eventually became saturated, with evidence of multiple substitutions at high divergences with respect to nuclear divergence.

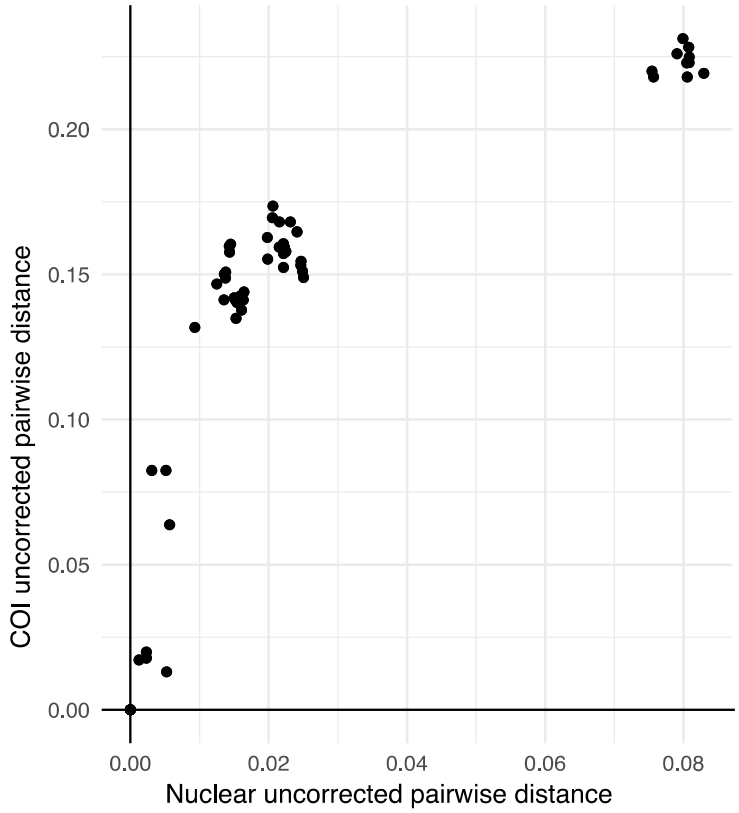

Figure 2. A scatter plot of uncorrected pairwise distances of COI versus uncorrected pairwise nuclear divergence for each whitefly specimen analyzed.

Likelihood mapping and quartet sampling on the concatenated nucleotide alignment showed some discordance which may have contributed to the conflicting coalescent and concatenated topologies. The majority of quartets produced from likelihood mapping (75%) support a sister relationship between the Sudan individual and all other ingroup whiteflies as reported for the ML concatenated analyses. However, 25% of the quartets sampled favor a sister relationship between the reference "B" reference genome and the Sudan individual (Figure 3). Quartet sampling methods also showed discordance at this described node (0.19/0/1) and for the ambiguously supported node for a monophyletic SSA + (AS–PAC–AU + ASIA) (−0.20/0/1) (Figure 4).

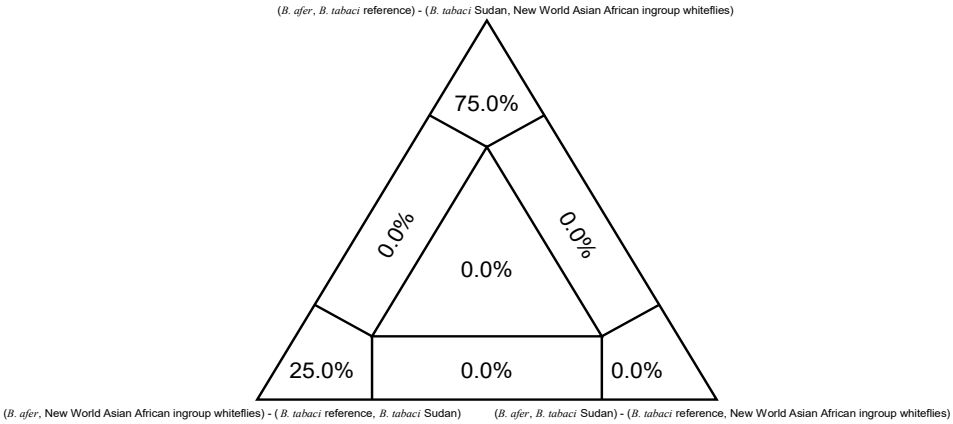

Figure 3. Likelihood map produced from IQ-tree showing the distribution of quartets that support a sister relationship between the reference "B" genome and the Sudan sample produced from the analysis of the concatenated alignment. Results show that 75% of quartets sampled support the Sudan sample together with the remaining samples (not including the reference) while 25% support the Sudan sample together with the reference, when all nucleotide sites are analyzed. These quartets are also relatively decisive in that no quartets provide ambiguous information (i.e., no points in the center or edges of the triangle).

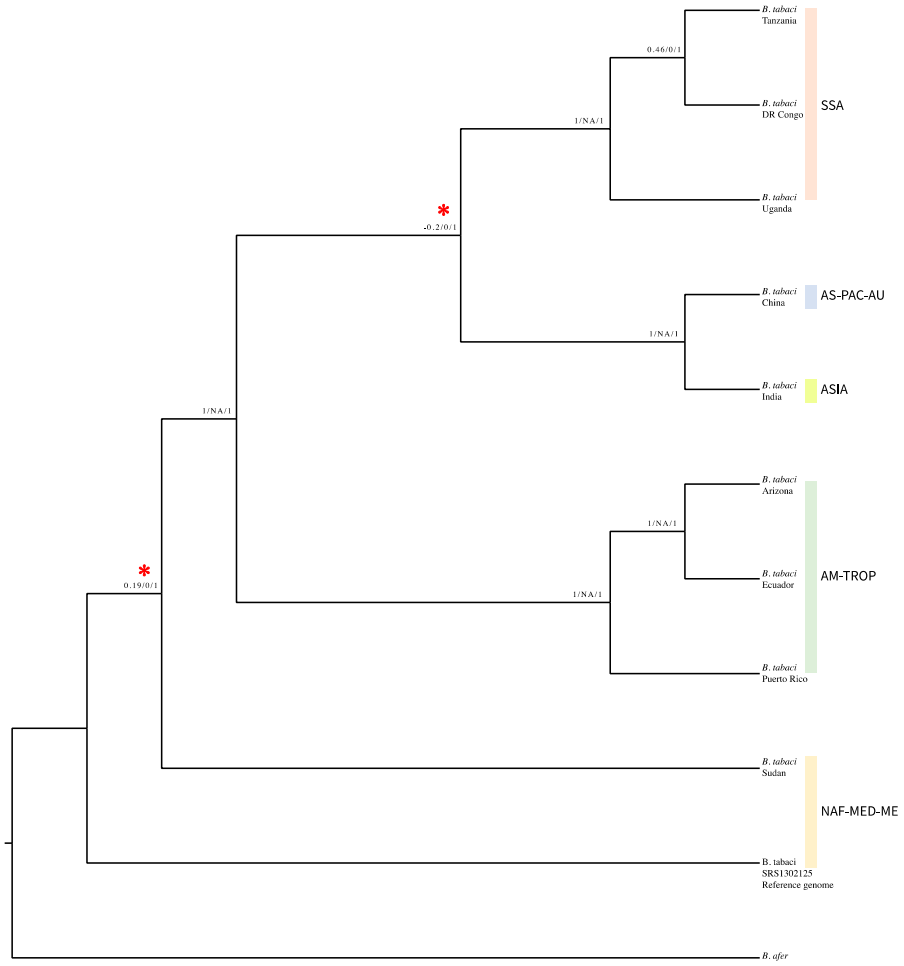

Figure 4. The *B. tabaci* complex of whiteflies showing the results of quartet sampling (QC: quartet concordance (frequency a concordant topology is observed)/QD: quartet differential (scaled fraction of discordant topologies)/QI: quartet informativeness (proportion of replicates that are informative)). Red asterisk indicate nodes of instability as discussed in the text. Nodes represented by 1/NA/1 are most stable and no discordant topologies are detected across quartets sampled (QD = NA).

Results of all the ABGD tests, which uses gaps in mismatch distributions of pairwise distances, suggested the existence of five candidate species in our data set. The candidate species were divided into the following putative species groups: (1) SSA (Uganda, Tanzania, and Democratic Republic of the Congo), (2) NAF–MED–ME ("B" reference genome and Sudan), (3) ASIA (India), (4) AS–PAC–AU (China), and (5) AM-TROP (Arizona, Puerto Rico, Ecuador). Based on these exemplars, the candidate species were consistent, regardless of distance method used for analysis of the COI data and nuclear supermatrix of whole genome data.

4. Discussion

For this study, one *B. afer* and nine *B. tabaci* genomes were sequenced using Illumina high throughput sequencing technologies. Coverage for each genome was approximately 39–65X based upon an estimated draft genome size for *B. tabaci* (615 Mb) [49,50]. This level of coverage allowed for the successful mapping of reads to obtain sequences for over 2184 nuclear single-copy orthologs. Phylogenomic analyses of these sequences revealed evidence of biogeographically partitioned lineages that were also consistent with those identified by the mitochondrial COI gene sequence [7]. Methods to detect gaps (ABGD analysis) in sequence divergence were consistent between mitochondrial and nuclear data, providing additional evidence that these major groups represent multiple species.

The differences among uncorrected pairwise distances of the mitochondrial COI alignment are large relative to nuclear orthologs obtained from whole genome sequencing (Table 3, Figure 2). The mitochondrial divergences between members of the *B. tabaci* complex are high when compared to divergences between other congeneric species of other insect taxa (e.g., aphids: 0.5–7% [90]; psyllids: 3–10% [91,92]; bark lice: 7–9% [93]). Specifically, up to 17% mitochondrial divergence was observed between members of the *B. tabaci* cryptic species complex. The extensive observed inter-clade COI divergence was proposed to reflect an ancient evolutionary age, hypothesized to be about 86 million years ago (MYA) [26]. However, within insects, several haplodiploid groups or those with paternal genome elimination [94–98] have been shown to have substantially higher mitochondrial DNA substitution rates. For instance, in thrips [99], scale insects [100], book lice [101], and parasitic lice [102], divergences among congeneric species range from 10–25%. Whiteflies are also haplodiploid [103] and elevated mitochondrial substitution rates may be an underlying biological artifact. Unlike the mitochondrial COI gene, the total divergence across the nuclear alignment was only 2.5% among the *B. tabaci* exemplars. The highest divergence in COI sequences observed between the *B. tabaci* samples that were not delineated as separate species by the ABGD analysis was 8.25% (Table 3B). This divergence was between the Democratic Republic of the Congo and the two other Sub-Saharan (Tanzania and Uganda) exemplars. However, the equivalent nuclear divergence among these three samples was only 0.3–0.5% (Table 3A). Similarly, the "B" reference genome and the Sudan individual differed by 6.37% for the COI gene, but only 0.57% in nuclear genes. In contrast to sequence divergence measured, lineages that were demarcated as different species by the ABGD analyses consistently showed greater than 13% and 1% divergence in the COI and nuclear data, respectively.

Phylogenetic relationships obtained from the analyses of whole genome and COI data showed some variation in regard to the placement of certain members. A monophyletic relationship between the "B" reference genome and Sudan individual was recovered by coalescent analyses and the ML COI analysis (Figure 1c–d). However, concatenated analyses of nucleotide data favor an alternate topology with the "B" reference genome sister to all remaining *Bemisia* members. The sister taxon of the ASIA and AS–PAC–AU groups also varied between the concatenated and coalescent analyses. The concatenated ML analyses recovered a sister relationship between the two lineages that overlap in the Asian continent (ASIA and AS–PAC–AU) and the SSA lineage, albeit with relatively weak support (Figure 1a,b). However, the coalescent analyses resulted in a sister relationship between the AM-TROP and the ASIA + AS–PAC–AU lineages (Figure 1c).

Likelihood mapping and quartet sampling methods demonstrate the discordance in topologies produced from the concatenated nucleotide alignment, which helps explain the difference in results

between coalescent and ML concatenated results. Likelihood mapping showed 75% of quartets support the topology produced from the ML concatenated analyses, but the remaining 25% support a sister relationship between the "B" biotype reference genome and the Sudan individual (Figure 3). Quartet sampling showed similar discordance for this node with a weak majority of quartets supporting the topology tested (QC = 0.19), but no skew in discordant topologies were detected (QD = 0) (Figure 4), suggesting only one discordant topology is favored similar to the result of likelihood mapping (Figure 3). Quartet sampling methods also display ambiguous support for the monophyly among ASIA, AS–PAC–AU, and SSA individuals with a weak majority of quartets supporting the alternate topology recovered by coalescent analyses (QC = −0.20) (Figure 4). The discordant frequencies detected from quartet methods demonstrate the instability of this node as suggested by coalescent analyses. The conflicting topologies reported between coalescent and concatenated ML analyses can be explained by incomplete lineage sorting and non-random introgression [81–83]. This would not be unexpected given recent analyses based on single nucleotide polymorphism data that suggest historical introgression patterns may have played a role in diversification of the complex [42].

The results of the analyses performed corroborate recent findings that suggest that the Sub-Saharan African group of whiteflies that occur on cassava form a single distinct lineage [45,46]. All analyses recover the three sampled Sub-Saharan individuals together in a monophyletic group, but there is discordance between analyses within the Sub-Saharan clade suggesting incomplete lineage sorting among Sub-Saharan populations. These results are consistent with the recent finding that introgression exists within this lineage [45]. A recent dating analysis based on genomic data estimates a divergence of the Sub-Saharan African population from the "B" (reference genome) to be approximately 5.26 million years ago [46], much younger than previous estimates based on COI data [26]. The nuclear divergences produced from this study may be an accurate reflection of this recent divergence within the complex (5 MYA). Therefore, dating analyses based solely on COI data may overestimate species divergences.

This study used whole genome sequences from shotgun sequencing methods to analyze members of the *B. tabaci* complex within a phylogenetic context. The read-mapping pipeline employed in this study was highly efficient in the recovery of orthologous sequences for members of the *B. tabaci* complex. Of the 2193 genes used as a genomic reference, nearly all were recovered with the read-mapping pipeline (Table 2). The read-mapping pipeline identified 1938 orthologs from the *B. afer* specimen, but this individual was meant to serve as an outgroup for phylogenetic analyses. However, all recovered COI sequences were validated using the described BLAST search protocol including the divergent *B. afer* individual (Table 2). The ML COI analysis on publicly available data [88] also validate read-mapped mitochondrial sequences obtained and are recovered within clades which correspond to geographic origin or biotype (Figure S2).

The results demonstrate that there is nuclear divergence among populations, however divergences among the single-copy nuclear orthologs analyzed are lower relative to COI sequences. The five putative species suggested by ABGD analyses complement phylogenetic analyses, suggesting population structure that divides members of the *B. tabaci* complex. High mitochondrial divergences may be a biological artifact that persists within the complex; thus, analyses of only COI data may overestimate potential diversity when analyzed in a phylogenetic framework. By exploring the nuclear divergences among orthologs identified with read-mapping and comparing them to results produced from analyses of COI data, evidence is growing for future recognition of multiple species within the complex. Although the analyses completed within this study are no panacea for the discord among the perception of *B. tabaci* taxonomy, the published genomic data set will no doubt provide an important resource to complement existing lines of evidence that putative, but undescribed species exist within the *B. tabaci* complex of whiteflies.

5. Conclusions

The *B. tabaci* complex of whiteflies has long been thought to comprise multiple undescribed species, however much disagreement exists on how many species should be recognized. The ABGD analyses suggest that at least five species exist both from the analyses of nuclear orthologs and COI data obtained. However, many populations remain under sampled and there is potential for the existence of unknown *B. tabaci* diversity within the complex. To avoid confusion regarding the recognition of species, caution should be taken regarding official taxonomic description of species, until a reasonable consensus is reached considering present and future evidence.

Supplementary Materials: The following are available online at http://www.mdpi.com/1424-2818/11/9/151/s1. Figure S1: Result of coalescent analysis on 2184 nuclear single-copy orthologs performed with Astrid. Figure S2: Result of maximum likelihood COI analysis of publicly available data [69] and read-mapped mitochondrial sequences obtained.

Author Contributions: R.S.d.M. contributed to methodology, formal analysis, visualization, original draft preparation, review and editing. J.K.B. and J.R.P.-M. contributed to conceptualization, resources, review and editing. A.D.S. contributed to software, methodology, review and editing. R.M.W. contributed to methodology, funding acquisition, review and editing. K.P.J. contributed to funding acquisition, supervision, review and editing. K.K.O.W. contributed to methodology, data curation, review and editing.

Funding: Support for this work was provided to K.P.J. by the National Science Foundation (grant number DEB-1342604). Support for R.M.W. was provided by the Swiss National Science Foundation (grant number PP00P3_170664).

Acknowledgments: We thank P. Gillespie for providing the sample of *B. afer*, and Ali Idris, Everlyne Wosula, Clerisse Casinga, and Peter Sseruwagi for providing the *B. tabaci* from Sudan, Tanzania, DR Congo, and Uganda, respectively. We thank Ray Gill (previously, California Department of Food and Agriculture, CA), for providing morphological identification of all whitefly samples. We also thank the three anonymous reviewers for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mound, L.A.; Halsey, S.H. *Whitefly of the World. A Systematic Catalogue of the Aleyrodidae (Homoptera) with Host Plant and Natural Enemy Data*; John Wiley and Sons: Hoboken, NJ, USA, 1978.
2. Greathead, A.H. *Bemisia tabaci*: A Literature Survey on the Cotton Whitefly with an Annotated Bibliography. In *Host Plants*; Cock, M.J.W., Ed.; CAB International Institutes, Biological Control: Silwood Park, UK, 1986; pp. 17–26.
3. Bird, J. A whitefly-transmitted mosaic of *Jatropha Gossypifolia*. *Tech. Paper* **1957**, *22*, 1–35.
4. Costa, H.S.; Brown, J.K. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. *Entomol. Exp. Appl.* **1991**, *61*, 211–219. [CrossRef]
5. Bedford, I.D.; Briddon, R.W.; Brown, J.K.; Rosell, R.C.; Markham, P.G. Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann. Appl. Biol.* **1994**, *125*, 311–325. [CrossRef]
6. De Barro, P.J.; Trueman, J.W.H.; Frohlich, D.R. *Bemisia argentifolii* is a race of *B. tabaci* (Hemiptera: Aleyrodidae): The molecular genetic differentiation of *B. tabaci* populations around the world. *Bull. Entomol. Res.* **2005**, *95*, 193–203. [CrossRef] [PubMed]
7. Brown, J.K. Phylogenetic Biology of the *Bemisia tabaci* Sibling Species Group. In *Bemisia: Bionomics and Management of a Global Pest*; Stansly, P.A., Naranjo, S.E., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 31–67. ISBN 978-90-481-2460-2.
8. Gill, R.J.; Brown, J.K. Systematics of *Bemisia* and *Bemisia* Relatives: Can Molecular Techniques Solve the *Bemisia tabaci* Complex Conundrum–A Taxonomist's Viewpoint. In *Bemisia: Bionomics and Management of a Global Pest*; Stansly, P.A., Naranjo, S.E., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 5–29. ISBN 978-90-481-2460-2.
9. Oliveira, M.R.V.; Henneberry, T.J.; Anderson, P. History, current status, and collaborative research projects for *Bemisia tabaci*. *Crop Prot.* **2001**, *20*, 709–723. [CrossRef]

10. Sseruwagi, P.; Legg, J.P.; Maruthi, M.N.; Colvin, J.; Rey, M.E.C.; Brown, J.K. Genetic diversity of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non-B biotype that can induce silverleaf symptoms in squash, in Uganda. *Ann. Appl. Biol.* **2005**, *147*, 253–265. [CrossRef]
11. Mugerwa, H.; Rey, M.E.C.; Alicai, T.; Ateka, E.; Atuncha, H.; Ndunguru, J.; Sseruwagi, P. Genetic diversity and geographic distribution of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) genotypes associated with cassava in East Africa. *Ecol. Evol.* **2012**, *2*, 2749–2762. [CrossRef]
12. Barbosa, L.D.F.; Marubayashi, J.M.; Marchi, B.R.D.; Yuki, V.A.; Pavan, M.A.; Moriones, E.; Navas-Castillo, J.; Krause-Sakate, R. Indigenous American species of the *Bemisia tabaci* complex are still widespread in the Americas. *Pest Manag. Sci.* **2014**, *70*, 1440–1445. [CrossRef]
13. Cock, M.J.W. *Bemisia Tabaci, an Update 1986–1992 on the Cotton Whitefly with an Annotated Bibliography*; CAB International Institutes, Biological Control: Silwood Park, UK, 1993.
14. Brown, J.K.; Frohlich, D.R.; Rosell, R.C. The Sweetpotato or Silverleaf Whiteflies: Biotypes of *Bemisia tabaci* or a Species Complex? *Annu. Rev. Entomol.* **1995**, *40*, 511–534. [CrossRef]
15. Costa, A.S.; Russell, R.C. Failure of *Bemisia tabaci* to breed on cassava plants in Brazil (Homoptera: Aleyrodidae). *Cienc. Cult. Saõ Paulo* **1975**, *2*, 388–390.
16. Brown, J.K.; Bird, J. Whitefly-transmitted geminiviruses in the Americas and the Caribbean Basin: Past and present. *Plant Dis.* **1992**, *76*, 220–225. [CrossRef]
17. Perring, T.M.; Cooper, A.D.; Rodriguez, R.J.; Farrar, C.A.; Bellows, T.S. Identification of a whitefly species by genomic and behavioral studies. *Science* **1993**, *259*, 74–77. [CrossRef] [PubMed]
18. Chowda-Reddy, R.; Kirankumar, M.; Seal, S.E.; Muniyappa, V.; Valand, G.B.; Govindappa, M.; Colvin, J. *Bemisia tabaci* Phylogenetic Groups in India and the Relative Transmission Efficacy of Tomato leaf curl Bangalore virus by an Indigenous and an Exotic Population. *J. Integr. Agric.* **2012**, *11*, 235–248. [CrossRef]
19. Bellows, T.S.; Perring, T.M.; Gill, R.J.; Headrick, D.H. Description of a Species of *Bemisia* (Homoptera: Aleyrodidae). *Ann. Entomol. Soc. Am.* **1994**, *87*, 195–206. [CrossRef]
20. Rosell, R.C.; Bedford, I.D.; Frohlich, D.R.; Gill, R.J.; Brown, J.K.; Markham, P.G. Analysis of Morphological Variation in Distinct Populations of *Bemisia tabaci* (Homoptera: Aleyrodidae). *Ann. Entomol. Soc. Am.* **1997**, *90*, 575–589. [CrossRef]
21. Dinsdale, A.; Cook, L.; Riginos, C.; Buckley, Y.M.; Barro, P.D. Refined Global Analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) Mitochondrial Cytochrome Oxidase 1 to Identify Species Level Genetic Boundaries. *Ann. Entomol. Soc. Am.* **2010**, *103*, 196–208. [CrossRef]
22. Hadjistylli, M.; Roderick, G.K.; Brown, J.K. Global Population Structure of a Worldwide Pest and Virus Vector: Genetic Diversity and Population History of the *Bemisia tabaci* Sibling Species Group. *PLoS ONE* **2016**, *11*, e0165105. [CrossRef] [PubMed]
23. De Barro, P.J.; Liu, S.-S.; Boykin, L.M.; Dinsdale, A.B. *Bemisia tabaci*: A Statement of Species Status. *Annu. Rev. Entomol.* **2011**, *56*, 1–19. [CrossRef]
24. Brown, J.K.; Idris, A.M. Genetic Differentiation of Whitefly *Bemisia tabaci* Mitochondrial Cytochrome Oxidase I, and Phylogeographic Concordance with the Coat Protein of the Plant Virus Genus Begomovirus. *Ann. Entomol. Soc. Am.* **2005**, *98*, 827–837. [CrossRef]
25. Frohlich, D.R.; Torres-Jerez, I.; Bedford, I.D.; Markham, P.G.; Brown, J.K. A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* **1999**, *8*, 1683–1691. [CrossRef]
26. Boykin, L.M.; Bell, C.D.; Evans, G.; Small, I.; De Barro, P.J. Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC Evol. Biol.* **2013**, *13*, 228. [CrossRef] [PubMed]
27. Tay, W.T.; Elfekih, S.; Court, L.N.; Gordon, K.H.J.; Delatte, H.; De Barro, P.J. The Trouble with MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive *Bemisia tabaci* (Hemiptera: Aleyrodidae) Cryptic Species Complex. *Genome Biol. Evol.* **2017**, *9*, 2732–2738. [CrossRef] [PubMed]
28. Gawel, N.J.; Bartlett, A.C. Characterization of differences between whiteflies using RAPD-PCR. *Insect Mol. Biol.* **1993**, *2*, 33–38. [CrossRef] [PubMed]
29. De Barro, P.J.; Driver, F.; Trueman, J.W.H.; Curran, J. Phylogenetic Relationships of World Populations of *Bemisia tabaci* (Gennadius) Using Ribosomal ITS1. *Mol. Phylogenet. Evol.* **2000**, *16*, 29–36. [CrossRef] [PubMed]

30. De la Rúa, P.; Simón, B.; Cifuentes, D.; Martinez-Mora, C.; Cenis, J.L. New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *J. Zool. Syst. Evol. Res.* **2006**, *44*, 25–33. [CrossRef]
31. Baoli, Q.; Coats, S.A.; Shunxiang, R.; Idris, A.M.; Caixia, X.; Brown, J.K. Phylogenetic relationship of native and introduced *Bemisia tabaci* (Homoptera: Aleyrodidae) from China and India based on mtCOI DNA sequencing and host plant comparisons. *Prog. Nat. Sci.* **2007**, *17*, 645–654. [CrossRef]
32. Boykin, L.M.; Shatters, R.G.; Rosell, R.C.; McKenzie, C.L.; Bagnall, R.A.; De Barro, P.; Frohlich, D.R. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol. Phylogenet. Evol.* **2007**, *44*, 1306–1319. [CrossRef]
33. Ahmed, M.Z.; Ren, S.-X.; Mandour, N.S.; Maruthi, M.N.; Naveed, M.; Qiu, B.-L. Phylogenetic analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations from cotton plants in Pakistan, China, and Egypt. *J. Pest Sci.* **2010**, *83*, 135–141. [CrossRef]
34. Brown, W.M.; George, M.; Wilson, A.C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 1967–1971. [CrossRef]
35. Ballard, J.W.O.; Whitlock, M.C. The incomplete natural history of mitochondria. *Mol. Ecol.* **2004**, *13*, 729–744. [CrossRef]
36. Latorre, A.; Hernández, C.; Martínez, D.; Castro, J.A.; Ramón, M.; Moya, A. Population structure and mitochondrial DNA gene flow in Old World populations of *Drosophila subobscura*. *Heredity* **1992**, *68*, 15–24. [CrossRef] [PubMed]
37. Besansky, N.J.; Lehmann, T.; Fahey, G.T.; Fontenille, D.; Braack, L.E.O.; Hawley, W.A.; Collins, F.H. Patterns of Mitochondrial Variation within and Between African Malaria Vectors, *Anopheles gambiae* and *An. arabiensis*, Suggest Extensive Gene Flow. *Genetics* **1997**, *147*, 1817–1828. [PubMed]
38. Pearce, R.L.; Wood, J.J.; Artukhin, Y.; Birt, T.P.; Damus, M.; Friesen, V.L. Mitochondrial dna suggests high gene flow in ancient murrelets. *Condor* **2002**, *104*, 84–91. [CrossRef]
39. Harpending, H.C. Signature of Ancient Population Growth in a Low-Resolution Mitochondrial DNA Mismatch Distribution. *Hum. Biol.* **1994**, *66*, 591–600. [PubMed]
40. Schneider, S.; Excoffier, L. Estimation of Past Demographic Parameters from the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. *Genetics* **1999**, *152*, 1079–1089. [PubMed]
41. Toews, D.P.L.; Brelsford, A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **2012**, *21*, 3907–3930. [CrossRef]
42. Elfekih, S.; Etter, P.; Tay, W.T.; Fumagalli, M.; Gordon, K.; Johnson, E.; Barro, P.D. Genome-wide analyses of the *Bemisia tabaci* species complex reveal contrasting patterns of admixture and complex demographic histories. *PLoS ONE* **2018**, *13*, e0190555. [CrossRef]
43. Hsieh, C.-H.; Ko, C.-C.; Chung, C.-H.; Wang, H.-Y. Multilocus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol. phylogenetics Evol.* **2014**, *76*, 172–180. [CrossRef]
44. McKenzie, C.L.; Bethke, J.A.; Byrne, F.J.; Chamberlin, J.R.; Dennehy, T.J.; Dickey, A.M.; Gilrein, D.; Hall, P.M.; Ludwig, S.; Oetting, R.D.; et al. Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) Biotypes in North America After the Q Invasion. *J. Econ. Entomol.* **2012**, *105*, 753–766. [CrossRef]
45. Wosula, E.N.; Chen, W.; Fei, Z.; Legg, J.P. Unravelling the Genetic Diversity among Cassava *Bemisia tabaci* Whiteflies Using NextRAD Sequencing. *Genome Biol. Evol.* **2017**, *9*, 2958–2973. [CrossRef]
46. Chen, W.; Wosula, E.N.; Hasegawa, D.K.; Casinga, C.; Shirima, R.R.; Fiaboe, K.K.M.; Hanna, R.; Fosto, A.; Goergen, G.; Tamò, M.; et al. Genome of the African cassava whitefly *Bemisia tabaci* and distribution and genetic diversity of cassava-colonizing whiteflies in Africa. *Insect Biochem. Mol. Biol.* **2019**, *110*, 112–120. [CrossRef] [PubMed]
47. Metzker, M.L. Sequencing technologies—The next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46. [CrossRef] [PubMed]
48. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [CrossRef] [PubMed]
49. Chen, W.; Hasegawa, D.K.; Arumuganathan, K.; Simmons, A.M.; Wintermantel, W.M.; Fei, Z.; Ling, K.-S. Estimation of the Whitefly *Bemisia tabaci* Genome Size Based on k-mer and Flow Cytometric Analyses. *Insects* **2015**, *6*, 704–715. [CrossRef] [PubMed]

50. Chen, W.; Hasegawa, D.K.; Kaur, N.; Kliot, A.; Pinheiro, P.V.; Luan, J.; Stensmyr, M.C.; Zheng, Y.; Liu, W.; Sun, H.; et al. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol.* **2016**, *14*, 110. [CrossRef] [PubMed]
51. Strimmer, K.; Haeseler, A. von Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6815–6819. [CrossRef] [PubMed]
52. Pease, J.B.; Brown, J.W.; Walker, J.F.; Hinchliff, C.E.; Smith, S.A. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am. J. Bot.* **2018**, *105*, 385–403. [CrossRef] [PubMed]
53. Swofford, D.L. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0. B5. 2001. Available online: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.458.6867 (accessed on 26 August 2019).
54. Puillandre, N.; Lambert, A.; Brouillet, S.; Achaz, G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* **2012**, *21*, 1864–1877. [CrossRef] [PubMed]
55. Sweet, A.D.; Boyd, B.M.; Allen, J.M.; Villa, S.M.; Valim, M.P.; Rivera-Parra, J.L.; Wilson, R.E.; Johnson, K.P. Integrating phylogenomic and population genomic patterns in avian lice provides a more complete picture of parasite evolution. *Evolution* **2018**, *72*, 95–112. [CrossRef] [PubMed]
56. Brown, J.K.; Coats, S.A.; Bedford, I.D.; Markham, P.G.; Bird, J.; Frohlich, D.R. Characterization and distribution of esterase electromorphs in the whitefly, *Bemisia tabaci* (Genn.) (Homoptera: Aleyrodidae). *Biochem. Genet.* **1995**, *33*, 205–214. [CrossRef] [PubMed]
57. Costa, H.S.; Brown, J.K.; Sivasupramaniam, S.; Bird, J. Regional distribution, insecticide resistance, and reciprocal crosses between the A and B biotypes of *Bemisia tabaci*. *Int. J. Trop. Insect Sci.* **1993**, *14*, 255–266. [CrossRef]
58. Perring, T.M. The *Bemisia tabaci* species complex. *Crop Prot.* **2001**, *20*, 725–737. [CrossRef]
59. Banks, G.K.; Bedford, I.D.; Beitia, F.J.; Rodriguez-Cerezo, E.; Markham, P.G. A Novel Geminivirus of *Ipomoea indica* (Convolvulacae) from Southern Spain. *Plant Dis.* **1999**, *83*, 486. [CrossRef] [PubMed]
60. Simón, B.; Cenis, J.L.; Demichelis, S.; Rapisarda, C.; Caciagli, P.; Bosco, D. Survey of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Italy with the description of a new biotype (T) from Euphorbia characias. *Bull. Entomol. Res.* **2003**, *93*, 259–264. [CrossRef] [PubMed]
61. Demichelis, S.; Arnò, C.; Bosco, D.; Marian, D.; Caciagli, P. Characterization of biotype T of *Bemisia tabaci* associated with *Euphorbia characias* in Sicily. *Phytoparasitica* **2005**, *33*, 196–208. [CrossRef]
62. Delatte, H.; Reynaud, B.; Granier, M.; Thornary, L.; Lett, J.M.; Goldbach, R.; Peterschmitt, M. A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. *Bull. Entomol. Res.* **2005**, *95*, 29–35. [CrossRef] [PubMed]
63. De Barro, P.J.; Liebregts, W.; Carver, M. Distribution and identity of biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in member countries of the Secretariat of the Pacific Community. *Aust. J. Entomol.* **1998**, *37*, 214–218. [CrossRef]
64. Qiu, B.-L.; Ren, S.X.; Wen, S.Y.; Mandour, N.S. Population differentiation of three biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in China by DNA polymorphism. *J. South China Agric. Univ.* **2006**, *27*, 29–33.
65. Qiu, B.-L.; Chen, Y.; Liu, L.; Peng, W.; Li, X.; Ahmed, M.Z.; Mathur, V.; Du, Y.; Ren, S. Identification of three major *Bemisia tabaci* biotypes in China based on morphological and DNA polymorphisms. *Prog. Nat. Sci.* **2009**, *19*, 713–718. [CrossRef]
66. Qiu, B.-L.; Dang, F.; Li, S.-J.; Ahmed, M.Z.; Jin, F.-L.; Ren, S.-X.; Cuthbertson, A.G.S. Comparison of biological parameters between the invasive B biotype and a new defined Cv biotype of *Bemisia tabaci* (Hemiptera: Aleyradidae) in China. *J. Pest Sci* **2011**, *84*, 419–427. [CrossRef]
67. Zang, L.-S.; Liu, S.S.; Liu, Y.Q.; Ruan, Y.M.; Wan, F.H. Competition between the B biotype and a non-B biotype of the whitefly, *Bemisia tabaci* (Homoptera: Aleyrodidae) in Zhejang, China. *Biodivers. Sci.* **2005**, *13*, 181–187. [CrossRef]
68. Zang, L.-S.; Chen, W.-Q.; Liu, S.-S. Comparison of performance on different host plants between the B biotype and a non-B biotype of *Bemisia tabaci* from Zhejiang, China. *Entomol. Exp. Appl.* **2006**, *121*, 221–227. [CrossRef]
69. Zang, L.-S.; Tong, J.; Jing, X.; Shu-Sheng, L.; You-Jun, Z. SCAR molecular markers of the B biotype and two non-B populations of the whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Chin. J. Agric. Biotechnol.* **2006**, *3*, 189–194.

70. Li, J.; Tang, Q.; Bai, R.; Li, X.; Jiang, J.; Zhai, Q.; Yan, F. Comparative Morphology and Morphometry of Six Biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from China. *J. Integr. Agric.* **2013**, *12*, 846–852. [CrossRef]
71. Andrews, S. FastQC A Quality Control Tool for High Throughput Sequence Data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 12 July 2018).
72. Kriventseva, E.V.; Rahman, N.; Espinosa, O.; Zdobnov, E.M. OrthoDB: The hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res.* **2008**, *36*, D271–D275. [CrossRef] [PubMed]
73. Johnson, K.P.; Dietrich, C.H.; Friedrich, F.; Beutel, R.G.; Wipfler, B.; Peters, R.S.; Allen, J.M.; Petersen, M.; Donath, A.; Walden, K.K.O.; et al. Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12775–12780. [CrossRef]
74. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
75. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup the Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
76. Mirarab, S.; Nguyen, N.; Guo, S.; Wang, L.-S.; Kim, J.; Warnow, T. PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. *J. Comput. Biol.* **2014**, *22*, 377–386. [CrossRef]
77. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
78. Vaidya, G.; Lohman, D.J.; Meier, R. SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **2011**, *27*, 171–180. [CrossRef]
79. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]
80. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* **2017**, *34*, 772–773. [CrossRef] [PubMed]
81. Mirarab, S.; Reaz, R.; Bayzid, M.S.; Zimmermann, T.; Swenson, M.S.; Warnow, T. ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* **2014**, *30*, i541–i548. [CrossRef] [PubMed]
82. Vachaspati, P.; Warnow, T. ASTRID: Accurate Species TRees from Internode Distances. *BMC Genom.* **2015**, *16*, S3. [CrossRef] [PubMed]
83. Rabiee, M.; Sayyari, E.; Mirarab, S. Multi-allele species reconstruction using ASTRAL. *Mol. Phylogenet. Evol.* **2019**, *130*, 286–296. [CrossRef]
84. Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **2000**, *7*, 203–214. [CrossRef]
85. Morgulis, A.; Coulouris, G.; Raytselis, Y.; Madden, T.L.; Agarwala, R.; Schäffer, A.A. Database indexing for production MegaBLAST searches. *Bioinformatics* **2008**, *24*, 1757–1764. [CrossRef]
86. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
87. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]
88. Boykin, L.M.; Savill, A.; De Barro, P. Updated mtCOI reference dataset for the *Bemisia tabaci* species complex. *F1000Research* **2017**, *6*, 1835. [CrossRef] [PubMed]
89. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
90. Wang, J.-F.; Jiang, L.-Y.; Qiao, G.-X. Use of a mitochondrial COI sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae). *Zookeys* **2011**, *122*, 1–17.
91. Lee, H.-C.; Yang, M.-M.; Yeh, W.-B. Identification of Two Invasive Cacopsylla chinensis (Hemiptera: Psyllidae) Lineages Based on Two Mitochondrial Sequences and Restriction Fragment Length Polymorphism of Cytochrome Oxidase I Amplicon. *J. Econ. Entomol.* **2008**, *101*, 1152–1157. [CrossRef] [PubMed]
92. Percy, D.M. Radiation, Diversity, and Host-Plant Interactions Among Island and Continental Legume-Feeding Psyllids. *Evolution* **2003**, *57*, 2540–2556. [CrossRef] [PubMed]

93. Bess, E.C.; Catanach, T.A.; Johnson, K.P. The importance of molecular dating analyses for inferring Hawaiian biogeographical history: A case study with bark lice (Psocidae: Ptycta). *J. Biogeogr.* **2014**, *41*, 158–167. [CrossRef]
94. Byrne, F.J.; Devonshire, A.L. Biochemical evidence of haplodiploidy in the whitefly *Bemisia tabaci*. *Biochem. Genet.* **1996**, *34*, 93–107. [CrossRef] [PubMed]
95. McMeniman, C.J.; Barker, S.C. Transmission ratio distortion in the human body louse, *Pediculus humanus* (Insecta: Phthiraptera). *Heredity* **2006**, *96*, 63–68. [CrossRef] [PubMed]
96. Andersen, J.C.; Wu, J.; Gruwell, M.E.; Gwiazdowski, R.; Santana, S.E.; Feliciano, N.M.; Morse, G.E.; Normark, B.B. A phylogenetic analysis of armored scale insects (Hemiptera: Diaspididae), based upon nuclear, mitochondrial, and endosymbiont gene sequences. *Mol. Phylogenet. Evol.* **2010**, *57*, 992–1003. [CrossRef] [PubMed]
97. Hodson, C.N.; Hamilton, P.T.; Dilworth, D.; Nelson, C.J.; Curtis, C.I.; Perlman, S.J. Paternal Genome Elimination in *Liposcelis* Booklice (Insecta: Psocodea). *Genetics* **2017**, *206*, 1091–1100. [CrossRef]
98. De la Filia, A.G.; Andrewes, S.; Clark, J.M.; Ross, L. The unusual reproductive system of head and body lice (*Pediculus humanus*). *Med. Vet. Entomol.* **2018**, *32*, 226–234. [CrossRef] [PubMed]
99. Morris, D.C.; Schwarz, M.P.; Crespi, B.J.; Cooper, S.J.B. Phylogenetics of gall-inducing thrips on Australian Acacia. *Biol. J. Linn. Soc.* **2001**, *74*, 73–86. [CrossRef]
100. Park, D.-S.; Suh, S.-J.; Hebert, P.D.N.; Oh, H.-W.; Hong, K.-J. DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). *Bull. Entomol. Res.* **2011**, *101*, 429–434. [CrossRef] [PubMed]
101. Feng, S.; Yang, Q.; Li, H.; Song, F.; Stejskal, V.; Opit, G.P.; Cai, W.; Li, Z.; Shao, R. The Highly Divergent Mitochondrial Genomes Indicate That the Booklouse, *Liposcelis bostrychophila* (Psocoptera: Liposcelididae) Is a Cryptic Species. *G3-Genes Genomes Genet.* **2018**, *8*, 1039–1047. [CrossRef] [PubMed]
102. Johnson, K.P.; Cruickshank, R.H.; Adams, R.J.; Smith, V.S.; Page, R.D.M.; Clayton, D.H. Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). *Mol. PHYLOGENET. Evol.* **2003**, *26*, 231–242. [CrossRef]
103. Horowitz, A.R.; Gerling, D. Seasonal Variation of Sex Ratio in *Bemisia tabaci* on Cotton in Israel. *Environ. Entomol.* **1992**, *21*, 556–559. [CrossRef]