

1 **Unraveling the cryptic *Bemisia tabaci* species complex: global phylogenomic**
2 **analysis reveals evolutionary relationships and biogeographic patterns**

3 Hua-Ling Wang^{1,2#}, Shi-Long Geng¹, Shu-Sheng Liu³, Zhong-Tao Li¹, Stephen
4 Cameron⁴, Teng Lei⁵, Wei Xu⁶, Qing Liu¹, Shuang Zuo¹, Christopher A. Omongo⁷, M.
5 N. Maruthi², Habibu Mugerwa⁸, Xiao-Wei Wang³, Yin-Quan Liu³, Jesús
6 Navas-Castillo⁹, Elvira Fiallo-Olivé⁹, Kyeong-Yeoll Lee¹⁰, Renate Krause-Sakate¹¹,
7 Hélène Delatte¹², James Ng¹³, Susan Seal², John Colvin^{2#}

8 ¹College of Forestry, Hebei Agricultural University, No. 2596, Lekainan Street,
9 Baoding, 071000, Hebei, China

10 ²Natural Resources Institute, University of Greenwich, Kent ME4 4TB, United
11 Kingdom

12 ³The Ministry of Agriculture Key Laboratory of Molecular Biology of Crop
13 Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, 866
14 Yuhangtang Road, Hangzhou 310058, China

15 ⁴Department of Entomology, Purdue University, 915 West State Street, West
16 Lafayette, IN 479074

17 ⁵College of Life Sciences, Taizhou University, Taizhou, 318000, China

18 ⁶State Key Laboratory of Genetic Resources and Evolution & Yunnan Key
19 Laboratory of Biodiversity and Ecological Conservation of Gaoligong Mountain,
20 Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan
21 650223, China

22 ⁷National Crops Resources Research Institute, Namulonge, P.O. Box 7084, Kampala,
23 Uganda

24 ⁸Department of Entomology, University of Georgia, 1109 Experiment Street, Griffin,
25 GA 30223, USA

26 ⁹Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora”
27 (IHSM-UMA-CSIC), Consejo Superior de Investigaciones Científicas, 29750
28 Algarrobo-Costa, Málaga, Spain

29 ¹⁰School of Applied Biosciences, Kyungpook National University, Daegu 702-701,
30 Republic of Korea

31 ¹¹UNESP, Faculdade de Ciências Agrônomicas, 18610-307, Botucatu, Brazil

32 ¹² CIRAD, UMR PVBMT CIRAD, Pôle de Protection des Plantes, 7 chemin de
33 l'IRAT, F-97410 Saint-Pierre

34 ¹³Department of Microbiology and Plant Pathology, University of California,
35 Riverside, California 92521, USA

36 Correspondence should be addressed to Hua-Ling Wang and John Colvin.
37 wang_hual@126.com; j.colvin@greenwich.ac.uk.

38

39 Running Head: **Genome-Wide Phylogeny of *Bemisia tabaci* Complex**

40

41 **Abstract**

42 *Bemisia tabaci* is a complex of cryptic agro-economically important pest species
43 characterized by diverse clades, substantial genetic diversity along with strong
44 phylogeographic associations. However, a comprehensive phylogenomic analysis
45 across the entire complex has been lacking, we thus conducted phylogenomic
46 analyses and explored biogeographic patterns used 680 single-copy nuclear genes
47 (SCNs) obtained from whole-genome sequencing data of 58 globally sourced *B.*
48 *tabaci* specimens. We constructed both concatenation and coalescent trees using 680
49 SCNs, which produced highly supported bootstrap values and nearly identical
50 topologies for all major clades. When comparing these concatenation trees with those
51 constructed using *mtCOI* and mitochondrial genome, we found conflicting
52 phylogenetic relationships, with the later trees recovering fewer major clades. In a
53 separate comparison between concatenation and coalescent trees, particularly those
54 generated using IQ-TREE, were found to delineate population relationships more
55 effectively than RaxML. In contrast, coalescent phylogenies were proficient in
56 elucidating geographical dispersal patterns and the reorganization of biological
57 species. Furthermore, we provided a strict consensus tree that clearly defines
58 relationships within most clades, laying a solid foundation for future research on the
59 evolution and taxonomy of *B. tabaci*. Ancestral range estimates suggested that the
60 ancestral region of the complex is likely situated in equatorial Africa, the Middle East,
61 and Mediterranean regions. Subsequently, expansion occurred into part of the
62 Palearctic and further into the Nearctic, Neotropical, Indomalayan, and Australasian
63 regions. These findings challenge both previous classifications and origin hypotheses,
64 offering a notably more comprehensive understanding of the global distribution,
65 evolutionary history, diversification, and biogeography of *B. tabaci*.

66 **Keywords:** *Bemisia tabaci*; cryptic species; phylogenomic tree; species tree; ancestral
67 geographic ranges

68

69 **INTRODUCTION**

70 The exploration of evolutionary relationships among organisms stands as a pivotal
71 objective and prerequisite for numerous biological investigations, which are
72 commonly illustrated through phylogenetic trees (Darwin, 1859; Dobzhansky, 1973;
73 Schluter, 2000; Grant & Grant, 2011; Boyd *et al.*, 2017; Irisarri *et al.*, 2017). Cryptic
74 species, despite their morphological similarity, represent distinct species and are
75 prevalent across various animal phyla (Bickford *et al.*, 2007; Nygren, 2014). Cryptic
76 species within the domain of pests hold notable importance, given their formidable
77 resistance to various management strategies (Bickford *et al.*, 2007) leading to
78 substantial reductions in agricultural yields. Consequently, considerable time,
79 financial resources, and effort have been allocated to comprehensively control these
80 pest populations.

81 The whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae), represents a prototypical
82 cryptic species complex and is an economically significant pest in tropical and
83 subtropical regions (Bellows *et al.*, 1994; De Barro *et al.*, 2011). Moreover, they are
84 vectors for plant viruses that cause extensive damage to both crops and weeds
85 (Lefeuvre *et al.*, 2011; Legg *et al.*, 2014). Initially considered a singular species with
86 multiple morphologically similar biotypes (Bird, 1957), recent advancements using
87 DNA barcoding with the gene encoding mitochondrial cytochrome oxidase I (*mtCOI*)
88 (Chu *et al.*, 2008; Ma *et al.*, 2009; Firdaus *et al.*, 2013; Díaz *et al.*, 2015;) and
89 reciprocal-crossing experiments have unveiled *B. tabaci* as a cryptic species complex,
90 encompassing >48 *mtCOI*-defined species (Tay *et al.*, 2017; Wang *et al.*, 2010, 2011;
91 Qin *et al.*, 2016). These species are further categorized into distinct genetic clades,
92 including Africa/Middle East/Asia Minor, Asia I, Asia II, Asia III, Asia IV, Asia V,
93 Australia/Indonesia, China, New World, Japan 2, Sub-Saharan Africa (SSA), Uganda,
94 and some unidentified ones (De Barro *et al.*, 2011). The varying species exhibit
95 remarkable diversity in their ecology and behavior, such as the range of their host
96 plant (Zang *et al.*, 2006; Xu *et al.*, 2011), insecticide resistance (Horowitz & Ishaaya,

97 2014), invasiveness (Liu *et al.*, 2007), specificity for plant virus transmission
98 (Fiallo-Olivé *et al.*, 2020), and induction of phytotoxic disorders in host plants
99 (Hogenhout *et al.*, 2008).

100 *Bemisia tabaci* is notorious for vectoring hundreds of plant viruses, predominantly
101 those belonging to the genus *Begomovirus*, causing billions of dollars in economic
102 losses globally to major food crop species (Mar *et al.*, 2017; Fiallo-Olivé and
103 Navas-Castillo, 2019; Menzel *et al.*, 2011; Polston *et al.*, 2014; Amari *et al.*, 2017;
104 Maruthi *et al.*, 2017; Costa *et al.*, 2020). For example, (1) Asia II clade species of *B.*
105 *tabaci* transmitted the cotton leaf curl disease (CLCuD) in Pakistan, causing an
106 economic loss of US\$5 billion from 1992 to 1997 (Briddon, 2003). (2) MED and
107 MEAM1 species from the Africa/Middle East/Asia Minor clade have become
108 invasive globally, with the MED outbreak in China in 2009 transmitting viruses that
109 caused significant tomato yield losses (Naranjo *et al.* 2010). (3) In SSA's
110 cassava-growing regions, SSA clade species carry diseases that lead to yearly
111 economic losses of US1.9–2.7 billion and US 0.1 billion due to cassava mosaic
112 disease (CMD) and cassava brown streak disease (CBSD), respectively (Legg *et al.*,
113 2006; Ndunguru *et al.*, 2015). Given the substantial impact inflicted by species
114 originating from various clades, it is imperative to acquire a comprehensive
115 understanding of their molecular phylogenetic relationships and geographic
116 distribution to effectively implement management strategies.

117 Recent advancements in high-throughput sequencing, in conjunction with enhanced
118 efforts in sample collection, indicate that nuclear genes are effective for exploring
119 phylogenetic relationships within particular clades, surpassing the utility of *mtCOI*
120 and the full mitogenome (Wosula *et al.*, 2017; Vyskočilová *et al.*, 2018; de Moya *et*
121 *al.*, 2019; Elfekih *et al.*, 2021; Mugerwa *et al.*, 2021; Ally *et al.*, 2023). For example,
122 Mugerwa *et al.* (2021) demonstrated the enhanced efficacy of whole-genome single
123 nucleotide polymorphism in resolving phylogenetic relationships within the SSA1
124 clade through comprehensive examination of mating datasets and other biological

125 traits. De Moya *et al.* (2019) utilized 2,184 nuclear orthologs obtained through whole
126 genome sequencing from nine individuals of *B. tabaci* to reveal cryptic diversity
127 within the complex. Additional molecular analyses of the phylogeny of *B. tabaci* have
128 typically relied on one to three single-copy nuclear genes (SCNs) and restriction site
129 associated DNA sequencing (RAD-Seq) which yielded a more resolved phylogeny
130 with high bootstrap support (Wosula *et al.*, 2017; Vyskočilová *et al.*, 2018; Elfekih *et*
131 *al.*, 2021; Mugerwa *et al.*, 2021; Wang *et al.*, 2024). However, the confirmation of the
132 relationship between particular clades remains uncertain, for example, significant
133 disparities in the positioning of the New World and sub-Saharan Africa clades
134 between the phylogenetic trees constructed by different datasets (De Barro *et al.*, 2005;
135 Boykin *et al.*, 2007; Wang *et al.*, 2024). Consequently, this uncertainty hinders a
136 comprehensive understanding of the biogeographic evolution of this elusive species
137 complex. Overall, phylogenetic studies on this cryptic species complex are limited
138 due to inadequate previous sampling in terms of specimen count and region coverage.
139 Therefore, comprehensive taxon sampling and sufficient datasets are crucial for
140 gaining further insights into the evolutionary history of *B. tabaci*.

141 Phylogenetic and biogeographical analyses have gained increasing attention due to
142 their potential in explaining the contemporary distributions of organisms (Edwards,
143 1964; Costello *et al.*, 2015; McGlone, 2005; Riddle, 2005; Cowie & Holland, 2006;
144 Lecture *et al.*, 2016). Hence, understanding the origin and dispersal of this pest is
145 crucial for advancing knowledge about its migration and diversification, as well as
146 offering insights into the development of integrated management practices that
147 effectively handle the rapidly emerging outbreaks of both whitefly and
148 whitefly-transmitted viral diseases worldwide. A previous phylogenetic study by
149 Frohlich *et al.* (1999) on a global *B. tabaci* collection revealed two major groups: Old
150 World and New World. The study traced origin of *B. tabaci* to the Old World group,
151 which was further subdivided into Indian subcontinent, equatorial Africa, and
152 Sahel-region subgroups. Alternative hypotheses regarding the origin and dispersion of
153 *B. tabaci* posit that (i) the Asian region serves as the probable center of origin (Mound,

154 1983); (ii) the Indian subcontinent represents the origin due to its rich diversity of
155 natural predators (Rekha *et al.*, 2005); and (iii) SSA is the likely center of origin,
156 supported by additional genetic diversity within the region (Boykin, 2007; Dinsdale *et*
157 *al.*, 2010). The dispersal and present-day biogeographical distribution of *B. tabaci*
158 species might be attributed to the fragmentation of Gondwanaland and subsequent
159 movements of tectonic plates (Boykin *et al.*, 2013). However, drawing conclusions
160 about migration routes has proven challenging, primarily due to the lack of a robust
161 backbone phylogeny with comprehensive sampling and a phylogeographical analysis.

162 A prevailing hypothesis proposes that using a phylogenomic analysis that is
163 specifically based on Single Copy Nuclear genes (SCNs) at a whole-genome scale can
164 substantially enhance the resolution of phylogenetic relationships at both shallow
165 (Naumann *et al.*, 2001; Wagner *et al.*, 2012) and deep levels (Duarte *et al.*, 2010;
166 Zhang *et al.*, 2012) in plant and in insects (Wiegmann *et al.*, 2000, 2009). The
167 investigation of SCNs has revealed genome skimming as a highly promising
168 methodology (Berger *et al.*, 2017; Vargas *et al.*, 2019). This approach entails the
169 compilation of a limited number of genes obtained from genome skimming datasets
170 with coverage ranging from 2X to 3.5X. In the realm of insect phylogeny, SCN
171 sequences present a promising avenue for the discovery of novel traits (Friedlander *et*
172 *al.*, 1992, 1994; Brower and DeSalle, 1997; Wiegmann *et al.*, 2000, 2009),
173 particularly in elucidating higher taxonomic levels. Nevertheless, the efficacy of
174 SCNs in discerning intragroup relationships within the *B. tabaci* cryptic species
175 remains unexplored.

176 Hence, to investigate the evolutionary relationships within the *B. tabaci* cryptic
177 species and estimate ancestral geographic ranges, we conducted both
178 concatenated-based and coalescent-based phylogenomic analyses using 680 SCNs, as
179 well as mitogenome analysis based on mitogenomes from 58 specimens representing
180 the global distribution of *B. tabaci*. Furthermore, we incorporated the phylogenomic
181 data into a biogeography-based model to estimate the ancestral geographic ranges for
182 key nodes along the backbone of the tree. Our study sought to answer the following

183 questions: (1) What is the phylogenetic relationship among the cryptic species of *B.*
184 *tabaci*? (2) What disparities, if any, are evident in the evolutionary relationships
185 depicted by concatenated-based phylogenetic trees and coalescent-based species trees?
186 (3) What are the origins and ancestral geographic ranges of this pest?

187 **MATERIAL AND METHODS**

188 **Sample Collection**

189 Specimens were obtained either directly from the host during field collections or from
190 laboratory colonies (Wang *et al.*, 2024). In total, specimens were collected from 16
191 sites from countries across four continents as well as from Australia: 1) Asia: China,
192 South Korea, India, Pakistan, Cambodia, and Israel, 2) Africa: Uganda, Malawi,
193 Sudan, Nigeria, Tanzania, Comoros, 3) Europe: Spain, and 4) South America: Brazil,
194 Peru (Table S1).

195 **Library Preparation and Genome Sequencing**

196 We extracted genomic DNA from 53 collected specimens, all preserved in ethanol,
197 using the Qiagen Blood and Tissue kit (Valencia, CA, USA) (Wang *et al.*, 2013). A
198 total of 20 µL of DNA was used for library preparation and DNA sequencing.
199 Genomic DNA was sheared to achieve a mean fragment size of ~300 bp using a
200 Covaris M220 instrument (Covaris, Woburn, USA), followed by the application of T4
201 DNA polymerase (MBI Fermentas, St. Leon-Rot, Germany) to generate blunt ends.
202 Subsequently, adapters were ligated to the ends of the DNA fragments after adding an
203 'A' base to the 3' end of the blunt phosphorylated DNA fragments. The desired
204 fragments were purified through gel-electrophoresis and then selectively enriched and
205 amplified by PCR. During the PCR stage, the index tag was introduced into the
206 adapter as required. A quantified Illumina paired-end library, for Illumina HiSeq
207 sequencing (2 × 150 bp), was prepared (Mugerwa *et al.*, 2021). Additionally, we
208 incorporated other next-generation sequencing (NGS) sequences from seven *B. tabaci*
209 specimens (Mugerwa *et al.*, 2021; Wang *et al.*, 2023) for subsequent analysis.

210 **Sequence Data Quality Control**

211 The raw sequencing data were generated by Illumina base-calling software CASAVA
212 v1.8.2 (http://support.illumina.com/sequencing/sequencing_software/casava.ilmn).
213 FastQC v.0.10.1(<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was
214 used to examine the Illumina sequence data. The raw paired-end reads were trimmed
215 and subjected to quality control using Trimmomatic v0.33
216 (<http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic>)(Bolger *et al.*
217 2014).

218 **SCNs Obtained from the Genomes**

219 To acquire SCN references throughout the entire *B. tabaci* genome, we retrieved
220 genomes associated with *B. tabaci* that were previously published up to 2022 and
221 evaluated their assembly quality. We then extracted SCN references from those
222 genomes that exhibited satisfactory assembly quality. In brief, four draft genomes of
223 *B. tabaci* were analyzed: MEAM1, MED, SSA1, and Asia II 1 (Chen *et al.*, 2016; Xie
224 *et al.*, 2017; Chen *et al.*, 2019; Hussain *et al.*, 2019). Among these, the assembly
225 quality of the MEAM1 genome exhibited superior assembly quality as compared to
226 the others and was selected as a reference for targeted read assembly, thereby
227 providing ample genomic information for the construction of phylogenetic clades.
228 Specifically, we used BUSCO v.2 (Simão *et al.*, 2015) to identify SCNs within the
229 MEAM1 *B. tabaci* genome by assessing insect orthologous datasets retrieved from
230 orthoDB (Zdobnov *et al.*, 2016).

231 **Assembly of Orthologous Sequences**

232 To compile potential SCNs among the quantified Illumina sequence data obtained
233 from various samples, we used aTRAM v.1.0.4 (Allen *et al.*, 2017). Initially, we
234 identified 1,291 SCNs from the genome of *B. tabaci* (MEAM1) (Table S4). These
235 genes were used as the starting sequences for aTRAM. We configured this tool to
236 assemble sequences across divergent taxa, we used the resulting contigs in a
237 post-aTRAM exon stitching process and excluded intron sequences
238 (https://github.com/juliema/phylogenomic_pipeline). The aTRAM exon stitching

239 procedure consisted of two sequential steps: exon positions were first determined
240 using the reference-guided annotation software Exonerate v2.2.0 (Slater & Birney,
241 2005), and then the exons were integrated. To further validate the orthologs of the
242 stitched exons in relation to the original reference sequences, we performed a
243 reciprocal-best-BLAST analysis on each final exon-containing contig (Allen *et al.*,
244 2017).

245 **Phylogenomic Tree Construction**

246 The SCNs extracted from each individual genome underwent alignment using PASTA
247 v1.6.3 (Mirarab *et al.*, 2015). The alignments were subsequently masked by
248 employing a 40% gap threshold utilizing trimAI 1.4 (Capella-Gutiérrez *et al.* 2009).
249 Following this, a concatenated supermatrix comprising the aligned SCNs was
250 generated through Sequence Matrix version 100.0 (Vaidya *et al.*, 2011) for
251 phylogenetic analyses. We initially used a concatenated-based tree analysis to
252 determine the phylogenetic relationships within *B. tabaci*. Codon positions of genes
253 were grouped into distinct partitions to account for rate heterogeneity across genes.
254 PartitionFinder 2 (Lanfear *et al.*, 2016) was used to compute these partitions, resulting
255 in 1,090 data partitions. Subsequently, we merged these data partitions into a single
256 alignment, creating a concatenated supermatrix alignment. This alignment was then
257 used for a partitioned maximum likelihood (ML) analysis using RAxML v.8.1.3 under
258 the GTR+gamma model (Stamatakis, 2014) and IQ-TREE v1.5.5 (Nguyen *et al.*,
259 2015), using the mixed models generated from PartitionFinder 2 (Lanfear *et al.*, 2016).
260 Branch support was estimated through 100 bootstrap replicates.

261 Next, we adopted a coalescent-based species tree estimation approach to infer species
262 relationships, aiming to reconcile gene tree–species tree conflicts arising from
263 incomplete clade sorting (Mirarab *et al.*, 2014; Vachaspati & Warnow, 2015). We
264 used two representative software programs, ASTRAL v.5.6.3 (Mirarab *et al.*, 2014;
265 Mirarab & Warnow, 2015) and ASTRID-2 (Vachaspati & Warnow, 2015), to estimate
266 species trees from a comprehensive set of gene trees. Finally, we derived a strict

267 consensus tree by synthesizing results from concatenation (RAxML and IQ-TREE)
268 and coalescent (ASTRAL and ASTRID) phylogenies, effectively summarizing the
269 phylogenetic outcomes.

270 To assess species determination, two approaches were employed to detect any
271 discrepancies between them. Firstly, a 3.5% *mtCOI* divergence is used to designate
272 different species (Frohlich *et al.*, 1999; Boykin *et al.*, 2007, 2012; Dinsdale *et al.*,
273 2010; Lee *et al.*, 2013). Secondly, we also conducted an automatic barcode gap
274 discovery (ABGD) test (Puillandre *et al.*, 2012) on both the *mtCOI* and nuclear data
275 sets. Initially, the test was performed using the *mtCOI* alignment with both
276 Jukes-Cantor (JC69) and simple distances. The relative gap width (X) was adjusted to
277 1.5. All other settings for the *mtCOI* analysis were kept at their default values.
278 Subsequently, the ABGD analysis was applied to the nuclear supermatrix of
279 single-copy orthologs, utilizing both JC69 and simple distances with default settings
280 (de Moya *et al.*, 2019).

281 **Historical Biogeographical Analyses**

282 BEAUti v1.8.2 (Drummond & Rambaut, 2007) was employed to produce the xml file
283 for BEAST runs using 711,102 alignment nucleotides obtained from 680 SCNs as the
284 input data. The MCMC simulations were executed for 50 million generations, with
285 sampling conducted at every 1000th generation. Convergence across the six runs was
286 verified using Tracer v1.6.0, and the ESS values exceeded 200 in each case. The
287 parameter settings for these runs were in accordance with the ones outlined by
288 Mugerwa *et al.* (2018). Post-processing, we used ETE 2 in Python (Huerta-Cepas *et*
289 *al.*, 2010) to extract only the cryptic species of *B. tabaci* from the resulting BEAST
290 phylogeny. The trimmed BEAST tree was subsequently used as the input tree for
291 BioGeoBEARS v. 1.1.1. Information regarding the current distributions of each
292 cryptic species was sourced from GenBank (based on submitted *mtCOI* sequences)
293 and our collection records. We defined five biogeographical regions according to
294 Udvardy *et al.* (1975): A, Part of Palearctic; B, Indomalayan; C, Australian; D,
295 Nearctic and Neotropical; and E, Western Palearctic plus Afrotropical. To infer the

296 ancestral ranges and colonization history of the *B. tabaci* cryptic species group, we
297 compared three main models: the Dispersal–Extinction–Cladogenesis (DEC) model,
298 which incorporates regional features influencing biogeographic change rates (Ree,
299 2005; Ree & Smith, 2008); the Dispersal-Vicariance (DIVA-like) model (Ronquist,
300 1997) which reconstructs ancestral distributions within a phylogeny without relying
301 on any prior assumptions concerning the nature or form of area relationships; and the
302 BAYAREA-like model which enables the estimation of parameters for a given
303 biogeographic model and facilitates the objective comparison of different
304 biogeographic models (Landis *et al.*, 2013). Each model was implemented with or
305 without the jump dispersal parameter (J), allowing ranges to change to include new
306 areas during speciation (Matzke, 2014). Thus, we compared six distinct models: DEC,
307 DEC + J, DIVA-like, DIVA-like + J, BAYAREA-like, and BAYAREA-like + J. To
308 determine the best-fitting model, a likelihood ratio test was conducted (Ree, 2005).
309 The probabilities of the ancestral states obtained with the best-fitting model were
310 visually represented as pie charts at the nodes of the tree.

311 **Results**

312 **Sampling Strategy**

313 For NGS analysis, we collected 58 *B. tabaci* samples, in addition to two outgroups
314 (*Bemisia afer*). Sample labels were assigned based on *mtCOI* barcode delimitation,
315 geographical distribution, and the order of description (Dinsdale *et al.*, 2010; De
316 Barro *et al.*, 2011; Boykin *et al.*, 2012). The dataset used in this study thus consisted
317 of 60 specimens, which included 33 putative *mtCOI* barcode *B. tabaci* cryptic species
318 and two outgroup species based on *mtCOI* sequences using the 3.5% threshold or 3.5%
319 threshold derived from ABGD analysis (Table S2,3 and Fig. 1A, B). The uncorrected
320 pairwise distances of the *mtCOI* sequences are illustrated in Table S2 and Fig. 1C,
321 which are largely higher than those of SCNs divergence (Table S2), except for the
322 species pair (sub_Saharan Africa 20_Uganda1 and sub_Saharan Africa
323 20_Uganda2/MED_uganda 1 and MED_uganda 2), which has the lowest *mtCOI*

324 pairwise distances as 0.0015 and 0.000 (Table S2). Additionally, we discovered three
325 novel putative species, namely SSA 18, SSA 19, and SSA 20. Notably, SSA 19
326 clustered within the previously identified SSA clades reported by Dinsdale *et al.*
327 (2010), whereas SSA 18 and SSA 20 formed distinct clusters within previously
328 unidentified clades. The sequence identity of *mtCOI* for each unknown species
329 compared to its closest relative in GenBank ranged from 87.65% to 92.98% (Table 1).
330 The corresponding nuclear divergence with the remaining *mtCOI*-defined species
331 varied between 1.99% and 4.52%, 1.56% and 4.54%, as well as 3.43% and 4.51%
332 (Table S2). Notably, we observed two SSA20 samples that exhibited a nuclear
333 divergence of only 0.59%, whereas the *mtCOI* divergence between them was 1.50%.
334 Our sampling encompassed 11 identified clades, along with four unidentified clades,
335 and revealed three novel putative species. This coverage represented 79% (38/48) of
336 the putative *B. tabaci* species, which allows for a more thorough understanding of the
337 diversity and distribution of *B. tabaci* species.

338 **DNA Sequencing and Assembly of SCNs**

339 The genomic libraries for the 58 *B. tabaci* samples and two outgroups yielded an
340 average of 180,230,808 (ranging from 136,146,850 to 264,664,076) raw 150-bp reads
341 per sample. After quality trimming, the clean reads (ranging from 110,870,294 to
342 242,663,488) were aligned to the MEAM1 genome, resulting in allocation rates of
343 66.92–95.78% (7.61% and 7.81% for the two outgroups) and an average depth of
344 15.25–43.03× (1.10–1.20 for the two outgroups), suggesting that the sequencing data
345 provided robust and comprehensive coverage of the MEAM1 genome for most
346 samples, enabling detailed genetic analysis. Sequences from each adult individual
347 covered 64.06–95.62% (5.40%, 6.39% for the two outgroups) of the reference
348 genome, indicating a substantial variation in the extent of genomic coverage among
349 samples. Overall, these results demonstrate the successful alignment of the sequencing data
350 to the reference genome, hereby establishing a robust framework for the subsequent
351 identification and analysis of SCN genes.

352 Of the 1,291 targeted SCNs (Table S4), an average of 1,232 candidate orthologs
353 (range: 930–1289) were assembled, either in part or in whole, from each sample.
354 Eventually, to ensure the reliability and accuracy of the phylogenetic inferences, 680
355 SCNs were selected for phylogenetic analysis after removing potentially
356 non-orthologous contigs and gene sets with >50% missing data. The concatenated
357 alignment contained 711,102 sites. The GC composition was analyzed for each codon
358 position, which indicated that the third codon position had a higher GC composition
359 than the first and second positions. However, the third codon positions were retained
360 for downstream analysis, as the bias was consistent across all taxa (Fig. S1), which is
361 unlikely to introduce systematic errors in the phylogenetic analysis.

362 **Concatenation Trees**

363 Our concatenated analyses resulted in highly supported IQ-TREE (Fig. 2A) and
364 RAxML (Fig. 2B) trees. Both trees revealed the presence of three prominent genetic
365 clades among a total of 14 clades, which included four unknown clades (Fig. 2),
366 suggesting the unexpectedly high diversity of *B. tabaci* cryptic species. These clades
367 consist of species designated as SSA 14, SSA 18, SSA 17, SSA 20, and East
368 Africa—all collected from the SSA region, suggesting a geographical pattern in the
369 distribution of these genetic clades. In the IQ-TREE, all inner nodes received
370 bootstrap values exceeding 95%, except for the pair relationships of sub-Saharan
371 Africa 1_SG3 and sub-Saharan Africa 1_SG2 in Clade a, indicating some uncertainty
372 in their precise relationship. For the RAxML tree, the mean bootstrap support was
373 97.8% and ranged from 69% to 100% (Fig. 2B). Of the total 57 inner nodes, a
374 substantial majority of 52 nodes (91.23%) obtained a bootstrap support of 80% or
375 higher. It is worth mentioning that clade a displayed two branches, namely the
376 connection between sub-Saharan Africa 1_SG1/SG3 and the relationship between
377 sub-Saharan Africa 6/sub-Saharan Africa 1, that received <80% bootstrap support.
378 Conversely, all remaining branches exhibited a bootstrap support exceeding 95%.
379 Clade b exhibited two branches, namely the relationship between Asia
380 I_Cambodia/AI AsiaI_India and the relationship of China 5 with other species, both

381 of which received <80% bootstrap support (Fig. 2B). In clade c, the relationships
382 were strongly supported, as almost all branches received 90% bootstrap support,
383 except for the weaker support observed for the interrelationships among species of
384 MED and MEAM1 (Fig. 2B).

385 In comparing the two trees, branches that received <80% bootstrap support in the
386 RAxML tree had values exceeding 95% bootstrap support in the IQ-TREE while
387 maintaining their topology. The main discrepancies between the RAxML tree and
388 IQ-TREE were associated with the former phylogeny, in which MED (MED_Sudan,
389 MED_Q2_Spain, MED_Israel) exhibited closer relationships to MEAM1 species than
390 to the other MED clade species from Uganda. Moreover, some confusion did arise
391 regarding clade names and the major clades that they belong to. For instance, (i) the
392 China 5 was indeed located at the base of the clade comprising species of Asia I,
393 China, Australia, and some Asia II species. (ii) SSAs appeared in both the
394 Africa/Middle-East/Asia Minor and SSA clades. (iii) The Asia II clade is disrupted by
395 the insertion of a China clade. These discoveries highlight the challenges in
396 classifying and understanding the diverse and dynamic nature of *B. tabaci* species
397 relationships and cladogenesis.

398 **Comparisons of Concatenation Trees with *mtCOI* and Mitochondrial Genome** 399 **Trees**

400 The phylogenetic relationships in *B. tabaci*, as inferred from concatenation trees,
401 showed contradictions with those derived from *mtCOI* and mitochondrial genome
402 (Figs. 1-3). For instance, in comparison to the concatenation trees, the *mtCOI* and
403 mitochondrial genome phylogenies recovered fewer major clades. As for comparing
404 the mitogenome-based (Fig. 3) and concatenation phylogenetic analysis (Figs. 2), East
405 Africa_Uganda1 and East Africa_Uganda2 were clustered outside the clades
406 containing sub-Saharan Africa 19_Uganda1 and sub-Saharan Africa 19_Uganda2, as
407 well as the MEDs and MEAM1s in the former topology. Intriguingly, the
408 concatenation phylogenies yielded a different arrangement, with East

409 Africa_Uganda1 aligning with the Indian Ocean rather than East Africa_Uganda2, and
410 sub-Saharan Africa 19_Uganda1 clustering with sub-Saharan Africa 9 rather than
411 sub-Saharan Africa 19_Uganda2 (Fig. 2). Furthermore, in contrast to the mitogenome
412 phylogeny, our concatenation phylogenies revealed that Asia II 3 and Asia II 9
413 clustered alongside China 1 and China 2. Asia II 6 was positioned at the base of the
414 Asia II clade, which includes Asia II 1, Asia II 7, and Asia II 5 (Fig. 2). The findings
415 are consistent with previous literature, supporting the hypothesis that phylogenies
416 based on *mtCOI* or full mitogenomes often conflict with those derived from whole
417 genome data (Hadjistylli et al., 2016; Wosula et al., 2017; Vyskočilová et al., 2018;
418 Mugerwa et al., 2021; Wang et al., 2024).

419 **Coalescent Phylogenies**

420 Considering the potential influence of incomplete clade sorting on inferred *B. tabaci*
421 phylogeny branches in concatenated tree analyses, we investigated two multispecies
422 coalescent approaches, namely ASTRID and ASTRAL, to elucidate the phylogenetic
423 relationships among *B. tabaci*. In general, the coalescent trees from both ASTRID
424 (Fig. 4A) and ASTRAL (Fig. 4B) approaches exhibited concordance with the trees
425 derived from the concatenated analysis. The ASTRID tree had robust branch support,
426 with values ranging from 80 to 100% for most individual branches. However, one
427 branch (Asia II 7_India) received <80% posterior probability support. Conversely, the
428 ASTRAL tree displayed comparatively lower support values, which were particularly
429 noticeable for potentially conflicting shallow branches dispersed across the phylogeny.
430 This difference highlights the varying sensitivity and resolution of the two coalescent
431 methods. The ASTRAL tree indicated that the Japan 2 was the sister clade to all other
432 *B. tabaci*, whereas Japan 2 was clustered with clade a in the ASTRID tree. The
433 remaining *B. tabaci* were classified into three distinct, strongly supported clades
434 (designated as a–c in Fig. 4). Furthermore, the interrelationships among these three
435 major clades were successfully resolved. Overall, these results contribute to a more
436 nuanced understanding of the phylogenetic relationships and evolutionary history of *B.*

437 *tabaci*, highlighting areas of agreement and discordance between different analytical
438 methods.

439 **Comparative Analysis of Concatenated and Coalescent Phylogenies**

440 In comparison to the concatenated phylogenies, the coalescent species tree analyses
441 resulted in subtle variations (Figs. 3 and 5), which were particularly evident in the
442 positioning of certain species. Notably, within the three major clades, a lone
443 topological discrepancy was observed—an alternative arrangement of New World 1
444 and 2. In the coalescent species trees, these two species were embedded within clade c,
445 albeit with highly supports, whereas in concatenated phylogenies, they were situated
446 within clade b, accompanied by stronger support. Regarding individual clade, China 5
447 exhibited divergence just before Italy3 in the coalescent species trees. Conversely, it
448 was embedded within the clade comprising Asia clades and the Australia clade in the
449 concatenated phylogenies. Notably, the placement of species from Uganda, for
450 example the species SSA 19, presents a conflict in trees constructed using both
451 approaches. An alternative relationship (SSA 1_SG3, SSA 3, Asia II 7) was observed
452 among well-supported clades connected by short branches. For instance, the SSA 3
453 mixed with the SSA 2 in the concatenated species tree, whereas they were placed as a
454 sister group to the SSA2 in coalescent phylogenies. The differing placements of these
455 species in the coalescent trees versus the concatenated phylogenies may reflect
456 distinct adaptations or historical events that have shaped their genetic makeup and
457 distribution. Lastly, the strict consensus of the IQ-TREE, RAxML, ASTRAL, and
458 ASTRID trees highlights a remarkable degree of nearly identical topologies (Fig. 5)
459 with well-resolved relationships within most clades, providing a solid foundation for
460 future evolutionary and taxonomic investigations into *B. tabaci*.

461 **Geographical Separation and Possible Ancestral Origin**

462 The model selection using AIC in BioGeoBEARS favored the BAYAREALIKE+J
463 model, as it exhibits a significantly higher likelihood score (Table 2). Consequently,
464 subsequent analyses primarily concentrated on the outcomes derived from the optimal

465 model. The results indicated that *B. tabaci* from the western Palearctic plus
466 Afrotropical regions were predominantly confined to two major clades (a and c; Fig.
467 6). Similarly, the part of Palearctic or Indomalayan *B. tabaci* were mainly confined to
468 a single major clade (b), except for the Japan 2 species, which did not align with the
469 three aforementioned major clades. The results revealed several dispersal events (Fig.
470 6), underscoring the species' remarkable capacity to colonize novel regions and thrive
471 in diverse environments. One noteworthy example is an ancient migration of *B. tabaci*
472 from the part of Palearctic or Indomalayan region to Australia, giving rise to
473 subsequent speciation events, as evidenced by two taxa from Australia (referred to as
474 Australia and Australia_E) embedded within a clade that predominantly comprises the
475 part of Palearctic and Indomalayan species (clade b). Additionally, the presence of
476 Asia I and Asia II 1 species in both the part of Palearctic and Indomalayan regions
477 suggests extensive movement between these two regions. The identification of these
478 clades and their geographical associations suggests that *B. tabaci* has undergone
479 significant diversification and speciation processes, leading to the formation of
480 genetically distinct populations across different regions.

481 Our ancestral range estimates offered a compelling visualization of the geographical
482 isolation patterns within the *B. tabaci* complex. The optimal biogeographic pathway,
483 as derived from our modeling, indicates a minimum of five dispersal events (Fig. 6,
484 I–V). The distribution of the species initiated in the western Palearctic and
485 Afrotropical region (Fig. 6, I), followed by range extensions into the part of Palearctic
486 (Fig. 6, III) and the Nearctic and Neotropical (Fig. 6, I), further extending into the
487 Indomalayan region (Fig. 6, IV), and ultimately reaching the Australian region (Fig. 6,
488 V). This suggests that the *B. tabaci* complex has undergone significant geographical
489 expansion and adaptation over time. Given that the highest species diversity is
490 concentrated in the middle regions of equatorial North Africa and in the Middle East
491 and Mediterranean regions, it is likely that the ancestral region can be narrowed down
492 to this specific geographic area. The dispersal from the western Palearctic plus
493 Afrotropical region implies migrations to the Americas, as well as the part of

494 Palearctic, contributing to the clades associated with New World, Asia, and China.
495 Subsequently, species indigenous to Asia and China underwent dispersal toward
496 either the Indomalayan or Australian regions. Notably, all major nodes along the
497 phylogenetic tree's backbone trace their ancestral range to the middle regions of
498 equatorial North Africa and the Middle East and Mediterranean regions. The only
499 exception arises with the species Japan 2, which was collected from Korea and
500 showed ancestral range estimates that encompassed the middle regions of equatorial
501 North Africa, the Middle East, and the Mediterranean regions, along with the part of
502 Palearctic, raising questions about its taxonomic status and the completeness of
503 sampling in certain regions. This anomaly could be attributed to either incomplete
504 sampling from the Asia, part of Palearctic or the possibility that Japan 2 is not a *B.*
505 *tabaci* species (Mugerwa *et al.*, 2018). These results provide crucial insights into the
506 evolutionary history, diversification, and biogeography of the *B. tabaci* complex,
507 indicating the importance of integrating genetic and geographical data to understand
508 species distributions and diversity.

509 **Discussion**

510 **Phylogeny Framework of *B. tabaci***

511 Within the *B. tabaci* cryptic species complex, the phylogenetic structure specifically
512 of the SSA clade species have been investigated predominantly relied on
513 genomic-scale SNPs generated through RAD-Seq (Wosula *et al.*, 2017; Vyskočilová
514 *et al.*, 2018; Elfekih *et al.*, 2021; Mugerwa *et al.*, 2021). Additionally, de Moya *et al.*
515 (2019) conducted phylogenetic analysis using 2,184 ortholog genes, which were
516 obtained through whole-genome sequencing of 10 individuals across five putative
517 species. Recently, Wang *et al.* (2024) inferred the relationships among 25
518 representative members of the *B. tabaci* cryptic species complex by utilizing
519 mitogenome, single-copy nuclear genes, and genomic-scale SNPs from RAD-Seq
520 markers. Despite these advancements, there are still significant gaps in our
521 understanding of the broader intragroup relationships within *B. tabaci*. And

522 previously studies have shown that the SCNs derived from the whole genome have
523 demonstrated their ability to enhance phylogenetic resolution and statistical power
524 compared to traditional markers (Berger *et al.*, 2017; Vargas *et al.*, 2019, Allen *et al.*,
525 2017; Lemmon *et al.*, 2012; Mandel *et al.*, 2014; Weitemier *et al.*, 2014; Dodsworth
526 *et al.*, 2019; Zhang *et al.*, 2019, Wagner *et al.*, 2012; Zhang *et al.*, 2012; Wolf *et al.*,
527 2002), presenting a promising avenue for further exploration. We thus employed
528 comprehensive taxon sampling that represented most reported major clades and
529 generating genomic-scale SCNs to enhance our understanding of the intragroup
530 relationships on a broader scale and achieve a well resolved tree for each of the stem
531 nodes along the backbone.

532 Based on a barcode gap distance of 3.3% for 670bp *mtCOI* sequences, which is
533 slightly below the standard threshold of 3.5%, we identified 33 species (see Table S3)
534 in our tested samples using the ABGD method (Puillandre *et al.*, 2012), a specially
535 designed approach for primary species delimitation. It is noteworthy that even if we
536 had applied the standard threshold of 3.5%, our results would still have indicated the
537 presence of 33 species, which further complement phylogenetic analyses, indicating
538 the effectiveness of the ABGD method in delineating species boundaries (de Moya *et*
539 *al.*, 2019; Khedkar *et al.*, 2014). Moreover, de Moya *et al.* (2019) demonstrated
540 consistency in species recognition by ABGD analysis of sequence divergence
541 between mitochondrial and nuclear data, whereas we encountered a conflict between
542 the two approaches. We only detected 1-2 species when utilizing SCNs regardless of
543 how we changed the relative gap width (X). This inconsistency might be attributed to
544 the fact that SCNs are more conserved than the orthologous genes used by de Moya *et*
545 *al.* (2019).

546 By utilizing 711,102 loci in 680 SCNs, the constructed trees resulted in a highly
547 resolved and well-supported nuclear phylogeny that differentiates from the phylogeny
548 based on *mtCOI* (Boykin *et al.*, 2007) or mitogenome. Moreover, the tree uncovered
549 indications of biogeographically partitioned clades that aligned with those identified
550 through analysis of the *mtCOI* gene sequence (Dinsdale *et al.*, 2010). In general, the

551 trees produced by both concatenation and coalescent summary methods were largely
552 compatible and well supported, except for weak support in interrelationships among
553 certain species. Notably, there was minimal discordance among the different
554 approaches used, which can be attributed to our stringent parameters for selecting the
555 SCNs. Overall, compared to the previous phylogenies (Wosula *et al.*, 2017;
556 Vyskočilová *et al.*, 2018; de Moya *et al.*, 2019; Elfekih *et al.*, 2021; Mugerwa *et al.*,
557 2021; Wang *et al.*, 2024), our study included a significantly larger number of
558 individuals (58) to obtain genome scale SCNs, rather than SNPs, mitogenomes, or
559 ortholog genes. This choice was motivated by the fact that SCNs, derived from both
560 parental sources and capturing a broader range of phylogenetically informative sites,
561 can significantly enhance the resolution of phylogenetic relationships and increase
562 statistical power compared to commonly used markers (Wolf *et al.*, 2002; Rokas *et al.*,
563 2003; Zhang *et al.*, 2019, 2012; Duarte *et al.*, 2010; Weber *et al.*, 2014; Allen *et al.*,
564 2017).

565 Furthermore, within the *B. tabaci* species complex, we identified four novel clades
566 across the a and c clades and introduced three new putative species: SSA 18, SSA 19,
567 and SSA 20. These putative species aligned with the Africa/Middle East/Asia Minor,
568 the SSA, and unknown genetic groups, respectively, as previously identified by
569 Dinsdale *et al.* (2010). In summary, our comprehensive sampling of SCNs,
570 encompassing densely represented clades that include both well-established and
571 lesser-known clades, has facilitated the construction of a fully resolved tree. This
572 achievement was previously challenging, particularly for each of the stem nodes along
573 the backbone.

574 **Implications for Naming and Phylogeny Construction Methods**

575 Discrepancies have risen regarding the species names and the major clades to which
576 they belong. These discrepancies can be traced back to the original naming of species,
577 which was primarily based on relationships established through *mtCOI* phylogeny
578 geographical origin and the order of description (De Barro *et al.*, 2011). However,

579 these relationships have shifted in SCNs phylogeny. The overwhelming evidence
580 demonstrates that nuclear genes are highly effective in elucidating phylogenetic
581 relationships within specific clades, outperforming both *mtCOI* and the entire
582 mitogenome (Wosula *et al.*, 2017; Vyskočilová *et al.*, 2018; de Moya *et al.*, 2019;
583 Elfekih *et al.*, 2021; Mugerwa *et al.*, 2021; Ally *et al.*, 2023). Furthermore, the highly
584 resolved phylogenetic structure presented in this study has significantly enhanced our
585 understanding of its evolutionary path and deepened our insights into its evolutionary
586 trajectory. As a result, these findings necessitate revisions in the classification of *B.*
587 *tabaci* (Boykin *et al.*, 2018; Brown *et al.*, 2023) First, our data did not support the
588 clustering of species Asia II 3 and Asia II 9 with the Asia II clade; instead, they
589 grouped with the Clade termed as “China”. Consequently, a revision of the
590 classification of species Asia II 3 and Asia II 9 to a Chinese name or as a separate
591 clade is warranted. Second, the classification of the China 5 species, originally
592 reported from YunNan province, China (Hu *et al.*, 2018), is noteworthy. Whereas
593 concatenation trees placed it within the clade termed as “China”, coalescent
594 approaches positioned it between the clade that includes the species Italy 3 and the
595 clade consisting of species from Asia. Geographically, the latter relationship, in which
596 the China 5 species has a close association with the species Italy 3, seems more
597 plausible, offering insights into potential dispersal events from Europe to Asia. Third,
598 the presence of species referred to as “SSA” within clades of Africa/Middle East/Asia
599 Minor, SSA, and unknown clades highlights the pressing issue of inconsistent naming
600 within the *B. tabaci* species complex (Boykin *et al.*, 2018). A species name holds a
601 wealth of biological information that forms a critical link between its ecology, biology,
602 and life history, thereby informing the development of species-specific management
603 strategies (Boykin *et al.*, 2018; Brown *et al.*, 2023). Now, with the phylogenetic
604 framework decoded using genomic datasets, we are equipped with a reliable reference
605 for making informed species name revision.

606 In comparing the approaches used for constructing the phylogeny of *B. tabaci*, it
607 became evident that concatenation phylogenies based on IQ-TREE might outperform

608 RaxML phylogeny in delineating population relationships. For instance, the IQ-TREE
609 phylogeny consistently clustered MED species together, a result not as pronounced in
610 the RaxML tree. This result from IQ-TREE was expected because specimens labeled
611 as MED share a high nucleotide identity, despite the possibility of containing different
612 species, as noted by Vyskočilová *et al.* (2018). Additionally, IQ-TREE results
613 indicated that the species SSA 3 was embedded within the species SSA 2,
614 corroborating the conclusions of Campbell *et al.* (2023), who proposed the
615 reclassification of the species SSA 3 as a SSA 2 species based on whole-genome
616 SNPs and crossing experiments. Similarly, in an earlier study Elfekih *et al.* (2021) had
617 argued that SSA 4 should also be reclassified as a SSA 2 species, as demonstrated by
618 the concatenation of 14,358 genome-wide SNPs derived from 63 cassava whitefly
619 specimens.

620 In contrast to concatenation phylogenies, coalescent phylogenies are more likely to
621 yield better results in exploring geographical dispersal and the reorganization of
622 biological species (McCormack *et al.*, 2009; Smith *et al.*, 2015; Mirarab & Warnow,
623 2015; Edwards *et al.*, 2016). Consistent with this, McCormack *et al.* (2009) and Smith
624 *et al.* (2015) proposed that “summary” coalescent-based species tree methods could
625 significantly enhance the accuracy of phylogenetic inference. For example, in
626 coalescent phylogenies, species from the New World clade appeared as the sister
627 taxon to the Africa/Middle East/Asia Minor clade, which comprised SSA, MED, and
628 MEAM1 species. This configuration provides compelling evidence for a close
629 evolutionary link between African and Old and New World species. Furthermore, our
630 coalescent phylogenies confirm the distinctness of SSA1_SG3 from SSA1_SG1 and
631 SSA1_SG2, aligning with Mugerwa *et al.*'s (2021) hypothesis that SSA1_SG3
632 represents a unique biological species, while SSA1_SG1 and SSA1_SG2 constitute
633 another, as evidenced by mating datasets.

634 Although slight differences were observed between concatenation and coalescent
635 phylogenies, these results affirm that genome-scale trees constructed through
636 concatenation and coalescent approaches yield accurate gene and species trees

637 (Irisarri *et al.*, 2017). Using both approaches concurrently appears to be instrumental
638 in comprehensively explaining various intriguing biological phenomena.

639 Additionally, our comparative analysis of the topology of concatenation phylogenies
640 and the *mtCOI* or mitogenome tree revealed discrepancies in the displaying the
641 phylogenetic relationships of different cryptic species. For instance,
642 mitochondrial-based phylogeny revealed fewer major clades compared to the
643 phylogenies constructed using SCN data. These results align with existing literature,
644 reinforcing the hypothesis that the *mtCOI* or mitogenome based phylogenies obtained
645 from the full mitogenomes conflicted with those from whole genome (Hadjistylli *et*
646 *al.*, 2016; Wosula *et al.*, 2017; Vyskočilová *et al.*, 2018; Mugerwa *et al.*, 2021; Wang
647 *et al.*, 2024). This discrepancy might arise due to the limitations of
648 mitochondrial-related marker, which may not accurately capture the true diversity
649 among different populations (Mugerwa *et al.*, 2021; Lee *et al.*, 2013) or mtDNA may
650 be less reliable in nonvertebrates compared to vertebrates (Allio *et al.*, 2017). Another
651 possible reason is that mt-associated markers might be more adept at elucidating
652 relationships at lower taxonomic levels, where SCNs could be effectively discriminate
653 both shallow and deep taxonomic levels (Wang *et al.*, 2017; Wiegmann *et al.*, 2000,
654 2009; Mark *et al.*, 2001; Naumann *et al.*, 2001; Wagner *et al.*, 2012; Duarte *et al.*,
655 2010; Zhang *et al.*, 2012). Alternatively, non-random processes like introgressions,
656 and hybridizations may contribute to conflicting topologies between nuclear and
657 mitochondrial phylogenetic trees (Richard *et al.*, 2014).

658 **Evolution and Dispersion of *B. tabaci***

659 Previous hypotheses about the origin and dispersal of *B. tabaci* (Mound, 1983;
660 Frohlich *et al.*, 1999; Dinsdale *et al.*, 2010) suggest origins in the Old World, Asia,
661 the Indian sub-continent, or SSA (Rekha *et al.*, 2005; Boykin *et al.*, 2007; Dinsdale *et*
662 *al.*, 2010; Boykin *et al.*, 2012). In contrast, data presented here suggest that *B. tabaci*
663 most likely originated from the middle regions of equatorial Africa, the Middle East,
664 and/or the Mediterranean regions, all of which were part of the Gondwana landmass
665 and experienced several dispersal events and subsequent diversification. This

666 inference finds substantial support in various observations. (i) The oldest clade in our
667 phylogeny trees are the species from SSA, and this clade shows higher species
668 richness than other clades, once we rule out the uncertainty of the species status of
669 Japan 2 (Lee *et al.*, 2013). Our analysis indicated that this ancient clade subsequently
670 diverged into the two other major clades (b and c). (ii) The four aforementioned
671 regions are physically adjacent. (iii) The high genetic diversity observed in equatorial
672 Africa, further substantiating it as a potential center of origin for *B. tabaci* (Lefeuvre
673 *et al.*, 2011; Mugerwa *et al.*, 2018). (iv) The clustering of certain SSA species with
674 MEAM1 and MED, suggesting geographical links with the Middle East and the
675 Mediterranean. Moreover, intriguing connections were found between different
676 continents, supporting the hypothesis by Boykin *et al.* (2013) attributing the current
677 biogeographical distribution of *B. tabaci* species to the breakup of Gondwanaland and
678 subsequent plate tectonic movements. During the Mesozoic era, it was hypothesized
679 that Western Gondwanaland underwent a fragmentation process, resulting in the
680 separation of South America and Africa between 120 and 84 million years ago
681 (Somoza *et al.*, 2008). Approximately 95 million years ago (Somoza *et al.*, 2008;
682 Veevers *et al.*, 1986; Torsvik *et al.*, 2000), Australia detached from Antarctica, while
683 India disengaged from Madagascar and subsequently migrated northward, eventually
684 colliding with Asia. What is not consistent with this hypothesis, however, is the
685 timing of diversification within the *B. tabaci* species, which was estimated to occur
686 between 60 and 30 million years ago (Cenozoic), a period that challenges the notion
687 of a phylogenetically deep variance event in the *B. tabaci* species complex
688 attributable to continental drift (Boykin *et al.*, 2013).

689 Our preferred BAYAREALIKE+J model, which includes founder speciation,
690 supports a more likely scenario of movement between different continents as plant
691 hosts dispersed, followed by subsequent speciation in the new plant host/environment.
692 However, confidently determining the accuracy of this scenario necessitates
693 associated fossils as calibration points or genome-scale datasets from species that
694 align with the age of the fossil record, a challenge posed by the current diversification

695 time derived from 657-bp *mtCOI* sequences and the use of only one calibration time
696 point.

697 Our biogeographic analysis revealed several notable patterns (Fig. 6). Primarily, we
698 observed that the biogeographic distribution was generally concordant with the major
699 reported clades (Frohlich *et al.*, 1999). For instance, New World species were
700 identified in the Americas (Brown *et al.*, 1995; Marubayashi *et al.*, 2013; Barbosa *et*
701 *al.*, 2014), whereas the Asian clade species (Asia I, Asia II, China) and the SSA clade
702 species were predominantly found in Asia and Africa, respectively. In contrast, the
703 highly invasive MEAM1 and MED species exhibited a wide geographical range, with
704 MED notably absent in Australia. This distribution aligns with the ability of MEAM1
705 and MED to invade new regions and displace indigenous species, often facilitated by
706 the movement of plant material by humans (Brown *et al.*, 1995; Liu *et al.*, 2007;
707 McKenzie *et al.*, 2004). Furthermore, our biogeographic analysis revealed key
708 insights into dispersal pathways. The presence of Australian species in both the
709 Australia clade and the Indonesia clade indicates a likely dispersal route from Asia to
710 Australia, possibly through India and Indonesia. Similarly, the presence of SSA 2 in
711 Europe suggests a dispersal pathway from SSA to Europe. Moreover, the
712 identification of New World 1 in Sudan indicates at least one dispersal event from
713 Africa to the Americas, likely from SSA to America. Additionally, our biogeographic
714 analysis identified a dispersal direction from China to India, with species from India
715 diverging after China 5 (which was collected in China).

716 The BioGeoBEARS analysis provided ambiguous results for the dispersal direction of
717 New World species in America. However, our phylogenetic analysis, which indicated
718 that the species New World 2 diverged after the species New World 1, suggests a
719 route from Latin America into North America. This finding challenges the previously
720 hypothesized invasion route into North America and then into Latin America
721 (Lefevre *et al.*, 2011). The discrepancy may stem from the *mtCOI* tree classification
722 of the species New World 1 as ancestral to the species New World 2, whereas our
723 phylogeny supports the species New World 2 as the ancestral species. To further

724 determine the direction of dispersal events among various species, comprehensive
725 evaluation through in-depth population studies is necessary.

726 **Sampling Gaps and Future Prospects**

727 Our phylogeny substantially expands the taxon sampling for genome-scale
728 phylogenetic studies. However, our study remains unable to cover the complete
729 number of the complex, due to inaccessible data such as Asia II 8 and 11 from India,
730 Asia III from Taiwan or India, China 3, Asia V from China, Italy 1 and 2 from Italy,
731 and SSA 5, 7, 8, 10, and 11 from Africa. The lack of comprehensive sampling
732 hampers our ability to confidently determine the direction of dispersal for some
733 species. Better resolution of these dispersal directions awaits a more extensive and
734 well-supported phylogeny. For instance, the sampling of species found in Italy
735 encompasses only 33% of the total reported species, thereby restricting our capacity
736 to investigate the ancestral range of this clade. Thus, the direction of the dispersal of
737 some species can be rigorously tested only when, and if, more Italian species are
738 included in the analyses. Furthermore, to comprehend the influence of population
739 migration and demography on species dispersal, it is crucial to ascertain the timing of
740 evolutionary events (Guindon *et al.*, 2020). However, in our current investigation, we
741 encountered limitations in deducing divergence time estimates. Although a substantial
742 fossil record exists for Aleyrodidae species (Drohojowska & Szwedo, 2011, 2012,
743 2013), these species were excluded from our taxonomic sampling.

744 **Conclusion**

745 Phylogenomics coupled with biogeographical analyses have revealed the most robust
746 and comprehensive backbone phylogeny for the *B. tabaci* species complex to date
747 along with a series of extensive dispersal events. Given the proven success of SCNs in
748 resolving phylogenetic relationships within cryptic species complexes, future studies
749 exploring biogeographical hypotheses at the species and genus levels could offer
750 further insights into the migrations and dispersals of the whitefly family.

751 **Author Contributions:** Hua-Ling Wang , Shi-Long Geng, Shu-Sheng Liu ,
752 Zhong-Tao Li, Stephen Cameron and John Colvin conceived and designed the
753 experiments; Hua-Ling Wang analyzed the data, Hua-Ling Wang, and Shi-Long Geng
754 wrote the manuscript, and all authors commented on the manuscript; Teng Lei, Wei
755 Xu, Qing Liu, Shuang Zuo, Christopher A. Omongo, M. N. Maruthi, Habibu
756 Mugerwa, Xiao-Wei Wang, Yin-Quan Liu, Jesús Navas-Castillo, Elvira Fiallo-Olivé,
757 Kyeong-Yeoll Lee, Renate Krause-Sakate, H  l  ne Delatte, James Ng, Susan Seal,
758 John Colvin collected experimental data. All authors read and approved the
759 manuscript.

760 **Acknowledgments:** We thank Tao Ye, Shanghai Biozeron Biotechnology Co. Ltd for
761 providing sequencing platform and bioinformatics support. Kevin Gorman,
762 Rothamsted Research UK, for providing whitefly specimens of Mediterranean_Sudan.
763 We acknowledge the support of Jodie Wetherall, Douglas Barry for giving access to
764 the High-Performance Computing facility at the University of Greenwich. We also
765 thank professor Peter Sseruwagi and Joseph Ndunguru, Mikocheni Agriculture
766 Research Institute, Tanzania, for collecting samples from Uganda. We also thank
767 Doctor Andrew Polaszek, Insects Division, Dept of Life Sciences, Natural History
768 Museum, London for providing samples from Grande Comore. We also thank Doctor
769 Paul De Barro, Research Director of the CSIRO, Australia for providing samples from
770 Australia.

771 **Disclosure:** The authors declare they have no conflict of interest.

772 **Funding:** This work was supported by the Bill & Melinda Gates Foundation (African
773 cassava whitefly project, OPP1058938) and the National Natural Science Foundation
774 of China (Grant numbers 31501878).

775 **Data Availability Statement:** *mtCOI* sequences from the *B. tabaci* samples isolated
776 in this study were deposited in GenBank (OR364062-OR364121). Raw sequence data
777 were deposited in the National Center for Biotechnology Information (NCBI)
778 SequenceRead Archive (SRA) database under study SRP455213 and are tied to the

779 BioProject identifier PRJNA997438. Individual BioSample and SRA run data can be
780 found in Supplementary Table S1.

781

782 **References**

- 783 Allen, J.M., Boyd, B., Nguyen, N.P., Vachaspati, P., Warnow, T., Huang, D.I., *et al.* (2017)
784 Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology*, 66,
785 786-798.
- 786 Allio, R., Donega, S., Galtier, N. and Nabholz, B. (2017) Large variation in the ratio of
787 mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and
788 the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, 34,
789 2762-2772.
- 790 Ally HM, Hamss HE, Simiand C, Maruthi MN, Colvin J, Delatte H. (2023) Genetic diversity,
791 distribution, and structure of *Bemisia tabaci* whitefly species in potential invasion and
792 hybridization regions of East Africa. *PLoS One*, 18, e0285967.
- 793 Amari, K., Gonzalez-Ibeas, D., Gómez, P., Sempere, R., Sanchez-Pina, M., Pendon, J., *et al.*
794 (2017) Tomato torrado virus is transmitted by *Bemisia tabaci* and infects pepper and eggplant
795 in addition to tomato. *Plant Disease*, 92, 1139.
- 796 Barbosa, L.daF., Marubayashi, J.M., De Marchi, B.R., Yuki, V.A., Pavan, M.A., Moriones, E., *et*
797 *al.* (2014) Indigenous American species of the *Bemisia tabaci* complex are still widespread in
798 the Americas. *Pest Management Science*, 70, 1440-1445.
- 799 Bellows, T.S.Jr., Perring, T.M., Gill, R.J., Headrick, D.H. (1994) Description of a species of
800 *Bemisia* (Homoptera: Aleyrodidae). *Annals of the Entomological Society of America*, 87,
801 195-206.
- 802 Berger, B.A., Han, J., Sessa, E.B., Gardner, A.G., Shepherd, K.A., Ricigliano, V.A., *et al.* (2017)
803 The unexpected depths of genome-skimming data: A case study examining Goodeniaceae
804 floral symmetry genes. *Applications in Plant Sciences*, 5, 1700042.
- 805 Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., *et al.* (2007) Cryptic
806 species as a window on diversity and conservation. *Trends in Ecology and Evolution*, 22,
807 148-155.
- 808 Bird, J. (1957) A whitefly-transmitted mosaic of *Jatropha gossypifolia*. *Technical Paper of the*
809 *Agricultural Experiment Station of Puerto Rico*, 22, 1-35.
- 810 Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina
811 sequence data. *Bioinformatics*, 30: 2114-2120.
- 812 Boyd, B.M., Allen, J.M., Nguyen, N.P., Sweet, A.D., Warnow, T., Shapiro, M.D. *et al.* (2017)
813 Phylogenomics using target-restricted assembly resolves intrageneric relationships of
814 parasitic lice (Phthiraptera: Columbicola). *Systematic Biology*, 66, 896-911.
- 815 Boykin, L.M., Shatters, Jr.R.G., Rosell, R.C., McKenzie, C.L., Bagnall, R.A., De Barro, P., *et al.*
816 (2007) Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using
817 Bayesian analysis of mitochondrial COI DNA sequences. *Molecular Phylogenetics and*
818 *Evolution*, 44, 1306-1319.

819 Boykin, L.M., Armstrong, K.F., Kubatko, L. and De Barro, P. (2012) Species delimitation and
820 global biosecurity. *Evolutionary Bioinformatics*, 8, 1-37.

821 Boykin, L.M., Bell, C.D., Evans, G., Small, I. and De Barro, P.J. (2013) Is agriculture driving the
822 diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha:
823 Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC*
824 *Evolutionary Biology*, 13, 1-10.

825 Boykin, L.M., Kinene, T., Wainaina, J.M., Savill, A., Seal, S., Mugerwa, H., *et al.* (2018) Review
826 and guide to a future naming system of African *Bemisia tabaci* species. *Systematic*
827 *Entomology*, 43, 427-433.

828 Briddon, R.W. (2003) Cotton leaf curl disease, a multicomponent begomovirus complex.
829 *Molecular Plant Pathology*, 4, 427-434.

830 Brown, J.K., Frohlich, D.R. and Rosell, R.C. (1995) The sweetpotato or silverleaf whiteflies:
831 biotypes of *Bemisia tabaci* or species complex? *Annual Review of Entomology*, 40, 511-534.

832 Brown, J.K., Paredes-Montero, J.R. and Stocks, I.C. (2023) The *Bemisia tabaci* cryptic (sibling)
833 species group - imperative for a taxonomic reassessment. *Current Opinion in Insect Science*,
834 57, 101032.

835 Campbell, L.I., Nwezeobi, J., van Brunschot, S.L. Kaweesi, T., Seal, S.E., Swamy R.A.R., *et al.*
836 (2023) Comparative evolutionary analyses of eight whitefly *Bemisia tabaci* sensu lato
837 genomes: cryptic species, agricultural pests and plant-virus vectors. *BMC Genomics*, 24, 408.

838 Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. (2009) trimAl: A tool for automated
839 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973.

840 Chamala, S., García, N., Godden, G.T., Krishnakumar, V., Jordon-Thaden, I.E., De Smet, R., *et al.*
841 (2015) Markerminer 1.0: A new application for phylogenetic marker development using
842 angiosperm transcriptomes. *Applications in Plant Sciences*, 3, 140011.

843 Chu, D., Wan, F.H., Tao, Y.L., Liu, G.X., Fan, Z.X. and Bi, Y.P. (2008) Genetic
844 differentiation of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotype Q based
845 on mitochondrial DNA markers. *Insect Science*, 15, 115-123.

846 Costa, T.M., Inoue-Nagata, A.K., Vidal, A.H., Ribeiro, S.D.G. and Nagata, T. (2020) The
847 recombinant isolate of cucurbit aphid-borne yellows virus from Brazil is a polerovirus
848 transmitted by whiteflies. *Plant Pathology*, 69, 1042-1050.

849 Costello, M.J., Cheung, A. and De Hauwere, N. (2015) Topography statistics for the surface and
850 seabed area, volume, depth and slope, of the world' s seas, oceans and countries.
851 *Environmental Science and Technology*, 44, 8821-8828.

852 Cowie, R.H. and Holland, B.S. (2006). Dispersal is fundamental to biogeography and the
853 evolution of biodiversity in oceanic islands. *Journal of Biogeography*, 33, 193-198.

854 Chen, W., Hasegawa, D.K., Kaur, N., Kliot, A., Pinheiro, P.V., Luan, J., *et al.* (2016) The draft
855 genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into
856 virus transmission, host adaptation, and insecticide resistance. *BMC Biology*, 14, 110.

857 Chen, W.B., Wosula, E.N., Hasegawa, D.K., Casinga, C., Shirima, R.R., Fiaboe, K.K.M., *et al.*
858 (2019) Genome of the African cassava whitefly *Bemisia tabaci* and distribution and genetic
859 diversity of cassava-colonizing whiteflies in Africa. *Insect Biochemistry and Molecular*
860 *Biology*, 110, 112-120.

861 Darwin, C. (1859) On the origin of species by means of natural selection. London: John Murray.

862 De Barro P.J. (2005) Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region
863 revealed using microsatellite markers. *Molecular Ecology*, 14, 3695-718.

864 De, Barro, P.J., Liu, S.S., Boykin, L.M., Dinsdale, A.B. (2011) *Bemisia tabaci*: a statement of
865 species status. *Annual Review of Entomology*, 56, 1-19.

866 de Moya, R.S., Brown, J.K., Sweet, A.D., Walden, K.K.O., Paredes-Montero, J.R., Waterhouse,
867 R.M., *et al.* (2019) Nuclear orthologs derived from whole genome sequencing indicate cryptic
868 diversity in the *Bemisia tabaci* (Insecta: Aleyrodidae) complex of whiteflies. *Diversity*, 11,
869 151.

870 Díaz, F., Endersby, N.M. and Hoffmann, A.A. (2015) Genetic structure of the whitefly
871 *Bemisia tabaci* populations in Colombia following a recent invasion. *Insect Science*, 22,
872 483-494.

873 Dinsdale, A., Cook, L., Riginos, C., Buckley, Y. and De Barro, P. (2010) Refined global analysis
874 of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial
875 cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the*
876 *Entomological Society of America*, 103, 196-208.

877 Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of evolution. *American*
878 *Biology Teacher*, 75, 87-91.

879 Dodsworth, S., Pokorny, L., Johnson, M.G., Kim, J.T., Maurin, O., Wickett, N.J. *et al.* (2019)
880 Hyb-Seq for flowering plant systematics. *Trends in Plant Science*, 24, 887-891.

881 Drohojowska, J. and Szwed, J. (2011) A new whitefly from Lower Cretaceous Lebanese amber
882 (Hemiptera: Sternorrhyncha: Aleyrodidae). *Insect Systematics and Evolution*, 42, 179-1
883 Drohojowska, J. and Szwed, J. (2012) New Aleyrodidae (Hemiptera: Sternorrhyncha:
884 Aleyrodomorpha) from Eocene Baltic amber. *Journal of Economic Entomology*, 80, 659-677.

885 Drohojowska, J. and Szwed, J. (2013) The first Aleyrodidae from the Lowermost Eocene Oise
886 amber (Hemiptera: Sternorrhyncha). *Zootaxa*, 3636, 319-347.

887 Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling
888 trees. *BMC Evolutionary Biology*, 7, 214-214.

889 Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C. *et al.* (2010)
890 Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza
891 and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology*, 10,
892 61.

893 Edwards, K. C. (1964). The importance of biogeography. *Geography*, 49, 85-97.

894 Edwards, S.V., Xi, Z.X., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., *et al.* (2016)
895 Implementing and testing the multispecies coalescent model: A valuable paradigm for
896 phylogenomics. *Molecular Phylogenetics and Evolution*, 94, 447-462.

897 Elfekih, S., Tay, W., Polaszek, A., Gordon, K., Kunz, D., Macfadyen, S. *et al.* (2021) On species
898 delimitation, hybridization and population structure of cassava whitefly in Africa. *Scientific*
899 *Reports*, 11, 1-11.

900 Fiallo-Olivé, E., Pan, L.L., Liu, S.S. and Navas-Castillo, J. (2020) Transmission of begomoviruses
901 and other whitefly-borne viruses: Dependence on the vector species. *Phytopathology*, 110,
902 10-17.

903 Fiallo-Olivé, E., and Navas-Castillo, J. (2019) Tomato chlorosis virus, an emergent plant virus
904 still expanding its geographical and host ranges. *Molecular Plant Pathology*, 20, 1307-1320.

905

906 Firdaus, S., Vosman, B., Hidayati, N., Jaya Supena, E.D., Visser, R.G.F. and van Heusden,
907 A.W. (2013) The *Bemisia tabaci* species complex: Additions from different parts of the
908 world. *Insect Science*, 20, 723-733.

909 Frohlich, D., Torres-Jerez, I., Bedford, I., Markham, P. and Brown, J.J. (1999) A
910 phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial
911 DNA markers. *Molecular Ecology*, 8, 1683-1691.

912 Grant, P.R. and Grant, B.R. (2011) How and why species multiply: the radiation of Darwin' s
913 Finches. Princeton, NJ: Princeton University Press.

914 Guindon, S. (2020) Rates and Rocks: Strengths and Weaknesses of Molecular Dating Methods.
915 *Frontiers in Genetics*, 11, 526.

916 Hadjistrylli, M., Roderick, G.K. and Brown, J.K. (2016) Global population structure of a
917 worldwide pest and virus vector: genetic diversity and population history of the *Bemisia*
918 *tabaci* sibling species group. *PLoS One*, 11, e0165105.

919 Hogenhout, S.A., Ammar, E.D., Whitfield, A.E. and Redinbaugh, M.G.. (2008) Insect vector
920 interactions with persistently transmitted viruses. *Annual Review of Phytopathology*, 46,
921 327-359.

922 Horowitz, A.R. and Ishaaya, I. (2014) Dynamics of biotypes B and Q of the whitefly *Bemisia*
923 *tabaci* and its impact on insecticide resistance. *Pest Management Science*, 70, 1568-1572.

924 Hu, J., Zhang, X., Jiang, Z., Zhang, F., Liu, Y., Li, Z. *et al.* (2018) New putative cryptic species
925 detection and genetic network analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in China
926 based on mitochondrial COI sequences. *Mitochondrial DNA Part A: DNA Mapping,*
927 *Sequencing, and Analysis*, 29: 474-484.

928 Huerta-Cepas, J., Dopazo, J. and Gabaldón, T. (2010) ETE: a python environment for tree
929 exploration. *BMC Bioinformatics*, 11, 24.

930 Hussain, S., Farooq, M., Malik, H.J., Amin, I., Scheffler, B.E., Scheffler, J.A., *et al.* (2019) Whole
931 genome sequencing of Asia II 1 species of whitefly reveals that genes involved in virus
932 transmission and insecticide resistance have genetic variances between Asia II 1 and
933 MEAM1 species. *BMC Genomics*, 20, 507.

934 Irisarri, I., Baurain, D., Brinkmann, H., Delsuc, F., Sire, J.Y., Kupfer, A. *et al.* (2017)
935 Phylotranscriptomic consolidation of the jawed vertebrate time tree. *Nature Ecology and*
936 *Evolution*, 1, 1370-1378.

937 Khedkar, G.D., Jamdade, R., Naik, S., David, L. and Haymer, D. (2014) DNA Barcodes for the
938 fishes of the Narmada, one of India's longest rivers. *PLoS One*, 9, e101460.

939 Landis, M.J., Matzke, N.J., Moore, B.R. and Huelsenbeck, J.P. (2013) Bayesian analysis of
940 biogeography when the number of areas is large. *Systematic Biology*, 62, 789-804.

941 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. and Calcott, B. (2016) PartitionFinder 2:
942 new methods for selecting partitioned models of evolution for molecular and morphological
943 phylogenetic analyses. *Molecular Biology and Evolution*, 34, 772-773.

944 Ma, W.H., Li, X.C., Dennehy, T.J., Lei, C.L., Wang, M., Degain, B.A., *et al.* (2009) Utility of
945 MtCOI polymerase chain reaction-restriction fragment length polymorphism in
946 differentiating between Q and B whitefly *Bemisia tabaci* biotypes. *Insect Science*, 16,
947 107-114.

948 Nehra, V. (2016). An introduction to biogeography and climate change. *International Journal of*
949 *Research and Scientific Innovation*, 3, 136.

950 Lee, W., Park, J., Lee, G., Lee, S. and Akimoto, S.I. (2013) Taxonomic status of the *Bemisia*
951 *tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent
952 species. *PLoS One*, 8, e63817.

953 Lefeuvre, P., Harkins, G.W., Lett, J.M., Briddon, R.W., Chase, M.W., Moury, B., *et al.* (2011)
954 Evolutionary time-scale of the begomoviruses: evidence from integrated sequences in the
955 *Nicotiana* genome. *PLoS One*, 6, e19193.

956 Lefeuvre, P., Martin, D.P., Harkins, G., Lemey, P., Gray, A.J., Meredith, S., *et al.* (2010) The
957 spread of tomato yellow leaf curl virus from the Middle East to the world. *PLoS Pathogens*, 6,
958 e1001164.

959 Legg, J.P., Shirima, R., Tajebe, L.S., Guastella, D., Boniface, S., Jeremiah, S., *et al.* (2014)
960 Biology and management of *Bemisia* whitefly vectors of cassava virus pandemics in Africa.
961 *Pest Management Science*, 70, 1446-1453.

962 Legg, J.P., Owor, B., Sseruwagi, P. and Ndunguru, J.J. (2006) Cassava mosaic virus disease in
963 East and Central Africa: epidemiology and management of a regional pandemic. *Advances in*
964 *Virus Research*, 67, 355-418.

965 Lemmon, A.R., Emme, S.A. and Lemmon, E.M. (2012) Anchored hybrid enrichment for
966 massively high-throughput phylogenomics. *Systematic Biology*, 61, 727-744.

967 Liu, S.S., De Barro, P.J., Xu, J., Luan, J.B., Zang, L.S., Ruan, Y.M., *et al.* (2007) Asymmetric
968 mating interactions drive widespread invasion and displacement in a whitefly. *Science*, 318,
969 1769-1772.

970 Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., *et al.* (2014) A
971 target enrichment method for gathering phylogenetic information from hundreds of loci: An
972 example from the Compositae. *Applications in Plant Sciences*, 2, 1300085.

973 Mar, T.B., Mendes, I.R., Lau, D., Fiallo-Olivé, E., Navas-Castillo, J., Alves, M.S., *et al.* (2017)
974 Interaction between the New World begomovirus *Euphorbia yellow mosaic virus* and its
975 associated alphasatellite: effects on infection and transmission by the whitefly *Bemisia tabaci*.
976 *Journal of General Virology*, 98, 1552-1562.

977 Mark, S.S., Ronald, W.B., Christophe, D., Heather, M.A., Ole, M., Wilfried, W.J., *et al.* (2001)
978 Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny
979 reconstruction. *Molecular Biology and Evolution*, 18, 132-143.

980 Marubayashi, J.M., Yuki, V.A., Rocha, K.C.G., Mituti, T., Pelegrinotti, F.M., Ferreira, F.Z., *et al.*
981 (2013) At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil.
982 *Journal of Applied Entomology*, 137, 113-121.

983 Maruthi, M.N., Jeremiah, S.C., Mohammed, I.U. and Legg, J.P. (2017) The role of the whitefly,
984 *Bemisia tabaci* (Gennadius), and farmer practices in the spread of cassava brown streak
985 ipomoviruses. *Journal of Phytopathology*, 165, 707-717.

986 Matzke, N.J. (2014). Model selection in historical biogeography reveals that founder-event
987 speciation is a crucial process in island clades. *Systematic Biology*, 63, 951-970.

988 Menzel, W., Abang, M.M. and Winter, S. (2011) Characterization of cucumber vein-clearing virus,
989 a whitefly (*Bemisia tabaci* G.)-transmitted carlavirus. *Archives of Virology*, 156, 2309-2311.

990 McKenzie, C.L., Anderson, P.K. and Villarreal, N. (2004) An extensive survey of *Bemisia tabaci*
991 (Homoptera: Aleyrodidae) in agricultural ecosystems in Florida. *Florida Entomologist*, 87,
992 403-407.

- 993 McCormack, J.E., Huang, H. and Knowles, L.L. (2009) Maximum likelihood estimates of species
 994 trees: How accuracy of phylogenetic inference depends upon the divergence history and
 995 sampling design. *Systematic Biology*, 58, 501-508.
- 996 McGlone, M.S. (2005) Goodbye Gondwana. *Journal of Biogeography*, 32, 739-740.
- 997 Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. and Warnow, T. (2014)
 998 ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30,
 999 541-548.
- 1000 Mirarab, S., Nguyen, N., Guo, S., Wang, L.S., Kim, J. and Warnow, T. (2015) PASTA: ultra-large
 1001 multiple sequence alignment for nucleotide and amino-acid sequences. *Journal of*
 1002 *Computational Biology*, 22, 377-386.
- 1003 Mirarab, S. and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with
 1004 many hundreds of taxa and thousands of genes. *Bioinformatics*, 31, 44-52.
- 1005 Mound, L. (1983). Biology and identity of whitefly vectors of plant pathogens. *Plant Virus*
 1006 *Epidemiology*, 305-313.
- 1007 Mugerwa, H., Seal, S., Wang, H.L., Patel, M.V., Kabaalu, R., Omongo, C.A., *et al.* (2018) African
 1008 ancestry of New World, *Bemisia tabaci*-whitefly species. *Scientific Reports*, 8, 2734.
- 1009 Mugerwa, H., Wang, H.L., Sseruwagi, P., Seal, S. and Colvin, J. (2021) Whole-genome single
 1010 nucleotide polymorphism and mating compatibility studies reveal the presence of distinct
 1011 species in sub-Saharan Africa *Bemisia tabaci* whiteflies. *Insect Science*, 28, 1553-1566.
- 1012 Naranjo, S.E., Castle, S.J., De Barro, P.J. and Liu, S.S. (2010) Population dynamics, demography,
 1013 dispersal and spread of *Bemisia tabaci*. In “*Bemisia*: Bionomics and Management of a Global
 1014 Pest”, eds Stansly PA and Naranjo SE, Springer, pp. 185-226.
- 1015 Naumann, J., Symmank, L., Samain, M.S., Müller, K.F., Neinhuis, C., *et al.* (2011) Chasing the
 1016 hare - Evaluating the phylogenetic utility of a nuclear single copy gene region at and below
 1017 species level within the species rich group *Peperomia* (Piperaceae). *BMC Evolutionary*
 1018 *Biology*, 11, 357.
- 1019 Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djinkeng, A., *et al.* (2015) Analyses
 1020 of twelve new whole genome sequences of *Cassava brown streak viruses* and *Ugandan*
 1021 *cassava brown streak viruses* from East Africa: diversity, supercomputing and evidence for
 1022 further speciation. *PLoS One*, 10, e0139321.
- 1023 Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: a fast and
 1024 effective stochastic algorithm forestimating maximum-likelihood phylogenies. *Molecular*
 1025 *Biology and Evolution*, 32, 268-274.
- 1026 Nygren, A. (2014) Cryptic polychaete diversity: a review. *Zoologica Scripta*, 43: 172-183.
- 1027 Polston, J.E., De Barro, P. and Boykin, L.M. (2014) Transmission specificities of plant viruses
 1028 with the newly identified species of the *Bemisia tabaci* species complex. *Pest Management*
 1029 *Science*, 70, 1547-1552.
- 1030 Puillandre, N., Lambert, A., Brouillet, S. and Achaz, G. (2012) ABGD, Automatic barcode gap
 1031 discovery for primary species delimitation. *Molecular Ecology*, 21, 1864-1877.
- 1032 Qin, L., Pan, L.L. and Liu, S.S. (2016) Further insight into reproductive incompatibility
 1033 between putative cryptic species of the *Bemisia tabaci* whitefly complex. *Insect Science*,
 1034 23, 215-224.
- 1035 Ree, R.H. (2005) Detecting the historical signature of key innovations using stochastic models of
 1036 character evolution and cladogenesis. *Evolution*, 59, 257-265.

- 1037 Ree, R.H. and Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution
1038 by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57, 4-14.
- 1039 Rekha, A., Maruthi, M., Muniyappa, V. and Colvin, J. (2005) Occurrence of three genotypic
1040 clusters of *Bemisia tabaci* and the rapid spread of the B biotype in south India. *Entomologia*
1041 *Experimentalis Et Applicata*, 117, 221-233.
- 1042 Rokas, A., Williams, B.L., King, N. and Carroll, S.B. (2003) Genome-scale approaches to
1043 resolving incongruence in molecular phylogenies. *Nature*, 425, 798.
- 1044 Richard, G.H. and Erica L. (2014) Hybridization, introgression, and the nature of species
1045 boundaries. *Journal of Heredity*, 105, 795-809.
- 1046 Riddle, B.R. (2005) Is biogeography emerging from its identity crisis? *Journal of Biogeography*,
1047 32, 185-186.
- 1048 Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of
1049 historical biogeography. *Systematic Biology*, 46, 195-203.
- 1050 Schluter, D. (2000) The ecology of adaptive radiation. Oxford, UK: Oxford University Press.
- 1051 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015)
1052 BUSCO: assessing genome assembly and annotation completeness with single-copy
1053 orthologs. *Bioinformatics*, 31, 3210-3212.
- 1054 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994) Evolution,
1055 weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of
1056 conserved polymerase chain-reaction primers. *Annals of the Entomological Society of*
1057 *America*, 87, 651-701.
- 1058 Slater, G.S.C. and Birney, E. (2005) Automated generation of heuristics for biological sequence
1059 comparison. *BMC Bioinformatics*, 6, 31.
- 1060 Smith, S.A., Moore, M.J., Brown, J.W. and Yang, Y. (2015) Analysis of phylogenomic datasets
1061 reveals conflict, concordance, and gene duplications with examples from animals and plants.
1062 *BMC Evolutionary Biology*, 15, 150.
- 1063 Somoza, R. and Zafarana, C. B. (2008) Mid-Cretaceous polar standstill of South America, motion
1064 of the Atlantic hotspots and the birth of the Andean cordillera. *Earth and Planetary Science*
1065 *Letters*, 271, 267-277.
- 1066 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1067 large phylogenies. *Bioinformatics*, 30, 1312-1313.
- 1068 Tay, W.T., Elfekih, S., Court, L., Gordon, K.H., Delatte, H. and De Barro, P. (2017) The trouble
1069 with MEAM2: implications of pseudogenes on species delimitation in the globally invasive
1070 *Bemisia tabaci* (Hemiptera: Aleyrodidae) cryptic species complex. *Genome Biology and*
1071 *Evolution*, 9, 2732-2738.
- 1072 Torsvik, T.H., Tucker, R.D., Ashwal, L.D., Carter, L.M., Jamtveit, B., Vidyadharan, K.T., *et al.*
1073 (2000) Late Cretaceous India–Madagascar fit and timing of break-up related magmatism.
1074 *Terra Nova*, 12, 220-224.
- 1075 Udvardy, M.D.F. (1975) A classification of the biogeographical provinces of the world.
1076 Morges(Switzerland): International Union for Conservation of Nature and Natural Resources.
1077 *IUCN Occasional Paper no.18*.
- 1078 Vachaspati, P. and Warnow, T. (2015). ASTRID: accurate species trees from internode distances.
1079 *BMC Genomics*, 16, S3.

1080 Vaidya, G., Lohman, D.J. and Meier, R. (2011) SequenceMatrix: Concatenation software for the
1081 fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, 27,
1082 171-180.

1083 Vargas, O.M., Heuertz, M., Smith, S.A. and Dick, C.W. (2019) Target sequence capture in the
1084 brazil nut family (Lecythidaceae): Marker selection and in silico capture from genome
1085 skimming data. *Molecular Phylogenetics and Evolution*, 135, 98-104.

1086 Veevers, J.J. (1986) Break of Australia and Antractica estimated as midcretaceous (95+/-5ma)
1087 from magnetic and seismic data at the continental margin. *Earth and Planetary Science*
1088 *Letters*, 77, 91-99.

1089 Vyskočilová, S., Tay, W.T., van Brunschot, S., Seal, S. and Colvin, J. (2018) An integrative
1090 approach to discovering cryptic species within the *Bemisia tabaci* whitefly species complex.
1091 *Scientific Reports*, 8, 10886.

1092 Wagner, S.T., Isnard, S., Rowe, N.P., Samain, M.S., Neinhuis, C., *et al.* (2012) Escaping the
1093 lianoid habit: Evolution of shrub-like growth forms in *Aristolochia* subgenus *Isotrema*
1094 (Aristolochiaceae). *American Journal of Botany*, 99, 1609-1629.

1095 Wang, H.L., Yang, J., Boykin, L.M., Zhao, Q.Y., Li, Q., Wang, X.W., *et al.* (2013) The
1096 characteristics and expression profiles of the mitochondrial genome for the Mediterranean
1097 species of the *Bemisia tabaci* complex. *BMC Genomics*, 14, 401.

1098 Wang, H.L., Rao, Q. and Chen, Z.Z. (2023) Identifying potential insecticide resistance markers
1099 through genomic-level comparison of *Bemisia tabaci* (Gennadius) lines. *Archives of Insect*
1100 *Biochemistry and Physiology*, 114, e22034.

1101 Wang, H.L., Lei, T., Wang, X.W., Cameron, S., Navas-Castillo, J., Liu, Y.Q., *et al.* (2024) A
1102 comprehensive framework for the delimitation of species within the *Bemisia tabaci* cryptic
1103 complex, a global pest-species group. *Insect Science*, doi: 10.1111/1744-7917.13361.

1104 Wang, P., Wang, Y.R., Wang, C.D., Zhang, T.T., Saleh, A. Al-Farraj. and Feng, G. (2017) Further
1105 consideration on the phylogeny of the Ciliophora: Analyses using both mitochondrial and
1106 nuclear data with focus on the extremely confused class Phyllopharyngea. *Molecular*
1107 *Phylogenetics and Evolution*, 112, 96-106.

1108 Wang, P., Sun, D.B., Qiu, B.L. and Liu, S.S. (2011) The presence of six cryptic species of the
1109 whitefly *Bemisia tabaci* complex in China as revealed by crossing experiments. *Insect*
1110 *Science*, 18, 67-77.

1111 Wang, P., Ruan, Y.M. and Liu, S.S. (2010) Crossing experiments and behavioral observations
1112 reveal reproductive incompatibility among three putative species of the whitefly *Bemisia*
1113 *tabaci*. *Insect Science*, 17, 508-516.

1114 Weber, C.C., Boussau, B., Romiguier, J., Jarvis, E.D. and Ellegren, H. (2014) Evidence for
1115 GC-biased gene conversion as a driver of between-lineage differences in avian base
1116 composition. *Genome Biology*, 15, 549.

1117 Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., *et al.*
1118 (2014) Hyb-Seq: Combining target enrichment and genome skimming for plant
1119 phylogenomics. *Applications in Plant Sciences*, 2, 1400042.

1120 Wiegmann, B.M., Mitter, C., Regier, J.C., Friedlander, T.P., Wagner, D.M. and Nielsen, E.S.
1121 (2000) Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera.
1122 *Molecular Phylogenetics and Evolution*, 15, 242-259.

1123 Wiegmann, B.M, Trautwein, M.D, Kim, J.W., Cassel, B.K., Bertone, M.A., Winterton, S.L., *et al.*
1124 (2009) Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC*
1125 *Biology*, 7, 34.

1126 Wolf, Y.I., Rogozin, I.B., Grishin, N.V. and Koonin, E.V. (2002) Genome trees and the tree of life.
1127 *Trends in Genetics*, 18, 472-479.

1128 Wosula, E.N, Chen, W., Fei, Z. and Legg, J.P. (2017) Unravelling the genetic diversity among
1129 cassava *Bemisia tabaci* whiteflies using NextRAD sequencing. *Genome Biology*, 9,
1130 2958-2973.

1131 Xie, W., Chen, C., Yang, Z., Guo, L., Yang, X., Wang, D., *et al.* (2017) Genome sequencing of
1132 the sweetpotato whitefly *Bemisia tabaci* MED/Q. *Gigascience*, 6, 1-7.

1133 Xu, J., Lin, K.K. and Liu, S.S. (2011) Performance on different host plants of an alien and an
1134 indigenous *Bemisia tabaci* from China. *Journal of Applied Entomology*, 135, 771-779.

1135 Zang, L.S., Chen, W.Q. and Liu, S.S. (2006) Comparison of performance on different host plants
1136 between the B biotype and a non-B biotype of *Bemisia tabaci* from Zhejiang, China.
1137 *Entomologia Experimentalis Et Applicata*, 121, 221-227.

1138 Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simao, F.A., Ioannidis, P., *et al.*
1139 (2016) OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal,
1140 plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, 45, 744-749.

1141 Zhang, N., Zeng, L., Shan, H. and Ma, H. (2012) Highly conserved low-copy nuclear genes as
1142 effective markers for phylogenetic analyses in angiosperms. *New Phytologist*, 195, 923-937.

1143 Zhang, F., Ding, Y., Zhu, C.D., Zhou, X., Orr, M.C., Scheu, S., *et al.* (2019) Phylogenomics from
1144 low-coverage whole-genome sequencing. *Methods in Ecology*, 10, 507-517.

1145 **Table 1** Three newly reported species identified through *mtCOI* sequences.

<i>B. tabaci</i>	Number	Sequences	GenBank accession	Closest	Identity(%)
SSA18 (1)	1	3a	OR364120	MK445128.1	91.28
SSA19 (2)	2	73;101_3NEW	OR364066;OR364073	MN646951.1	92.98
SSA20 (2)	2	6/493	OR364063;OR364064	MN056066.1	87.65

1146

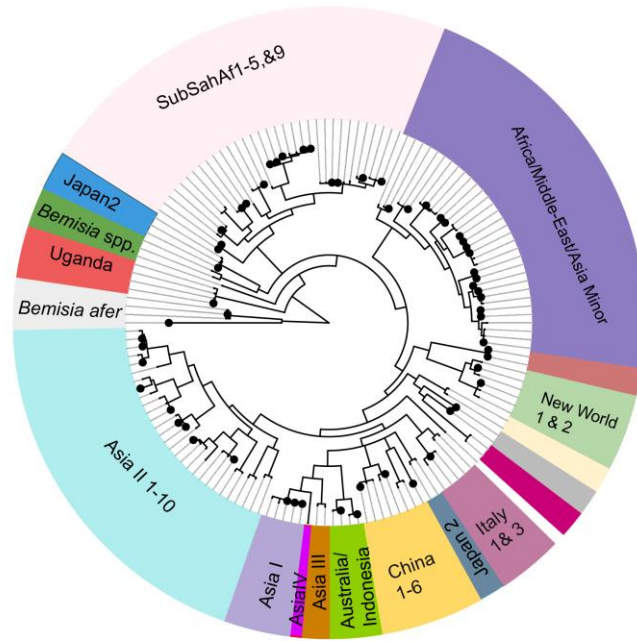
1147 **Table 2** Comparison of different biogeographic models assessed in BioGeoBEARS.

Model	LnL	No. of free parameters	d	e	j	AIC	AIC_wt
DEC	-76.00	2	4.54	1.00E-12	0.000	156.00	8.00E-06
DEC+J	-73.20	3	3.61	6.60E-02	0.012	152.40	4.90E-05
DIVALIKE	-78.65	2	4.88	1.80E+00	0.000	161.30	5.70E-07
DIVALIKE+J	-74.88	3	4.20	1.00E-12	0.013	155.80	9.10E-06
BAYAREALIKE	-81.56	2	3.12	5.00E+00	0.000	167.10	3.10E-08
BAYAREALIKE+JJ	-63.27	3	2.82	4.98E+00	0.018	132.50	1.00E+00

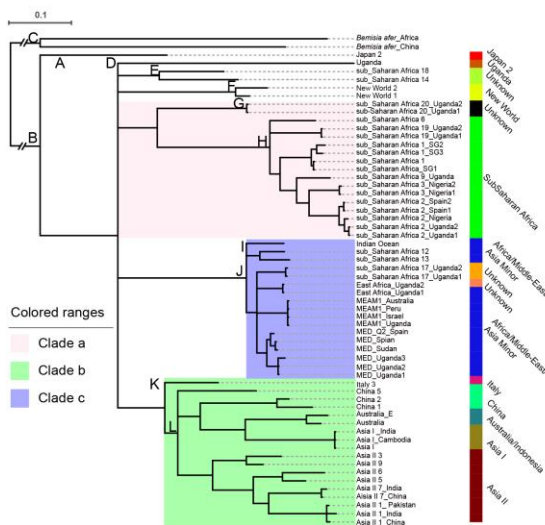
1148 Abbreviations:log-likelihood (LnL), Akaike information criterion (AIC), dispersal matrix (d),
1149 extinctionmatrix(e),jump dispersal or founder-event speciation (j).

1150

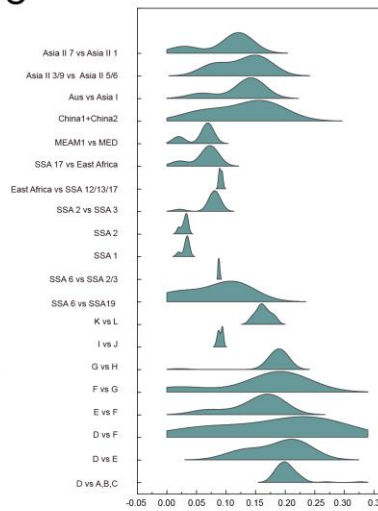
A



B



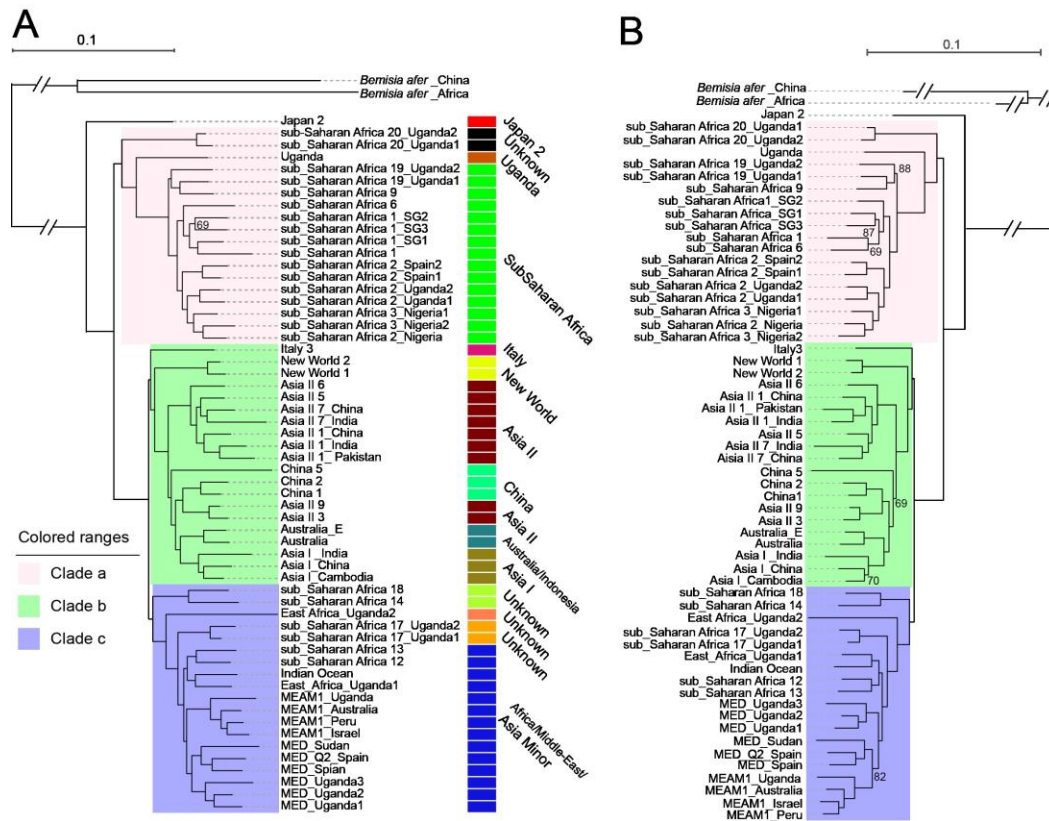
C



1152

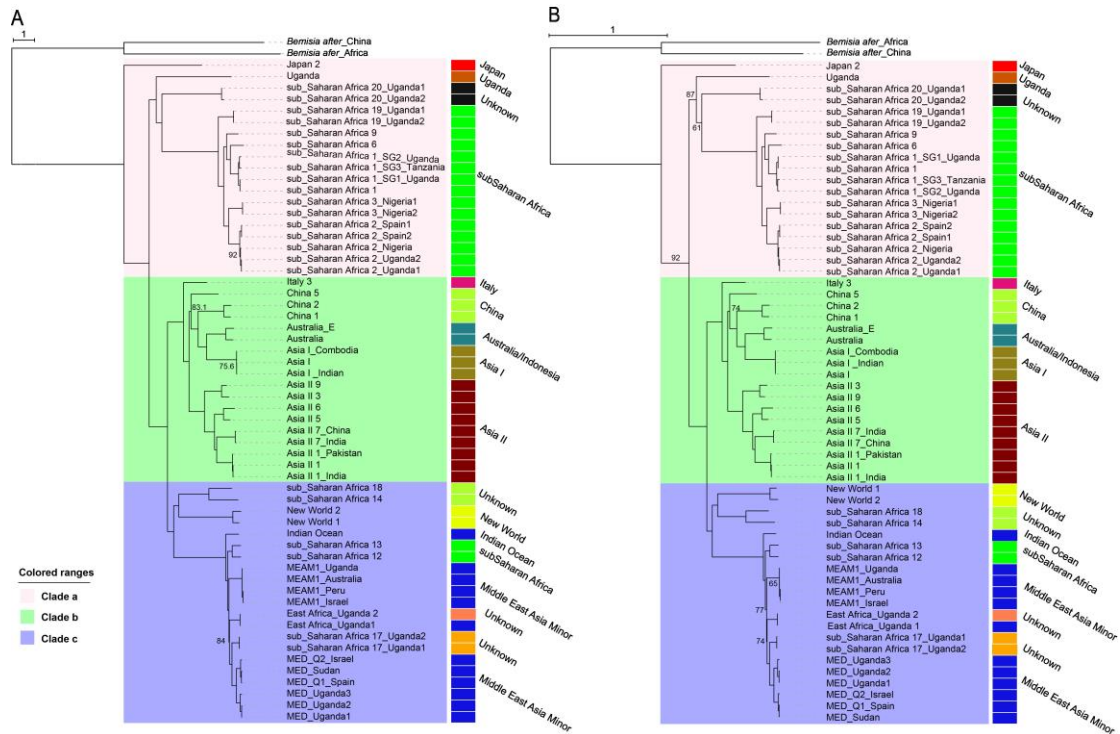
1153 **Fig. 1** Distribution map of samples, phylogeny of the selected species, and pairwise
 1154 comparisons of *mtCOI* uncorrected pairwise nucleotide distances (p-distances). A,
 1155 Phylogeny based on 657 bp of the *mtCOI* using the Bayesian method and distribution
 1156 map of samples used in this study. Black dots at tips indicate samples that were
 1157 selected for genomic sequencing providing a focused subset for subsequent analyses.
 1158 Major genetic clades of interest, labeled a–c, are color-coded. B, Phylogeny based on
 1159 *mtCOI* for the selected species using the Bayesian method, revealing distinct genetic
 1160 clades (labeled A–L). C, Pairwise comparisons of *mtCOI* uncorrected p-distances

1161 among pairs of closely related clades. Note: Aus represents Australia and Australia_E,
 1162 SSA represents sub-Saharan Africa.



1163

1164 **Fig. 2** Phylogenetic trees showing relationships of *B. tabaci* species based on the
 1165 concatenation of 680 SCNs. A, IQ-Tree phylogeny. B, RaxML phylogeny. Colors at
 1166 tips correspond to different clades for both A and B. Bootstrap values of <95% are
 1167 shown on the tree. All other branches were strongly supported (bootstrap values of
 1168 100%). Major genetic clades of interest, labeled a–c, are color-coded.



1169

1170

1171

1172

1173

1174

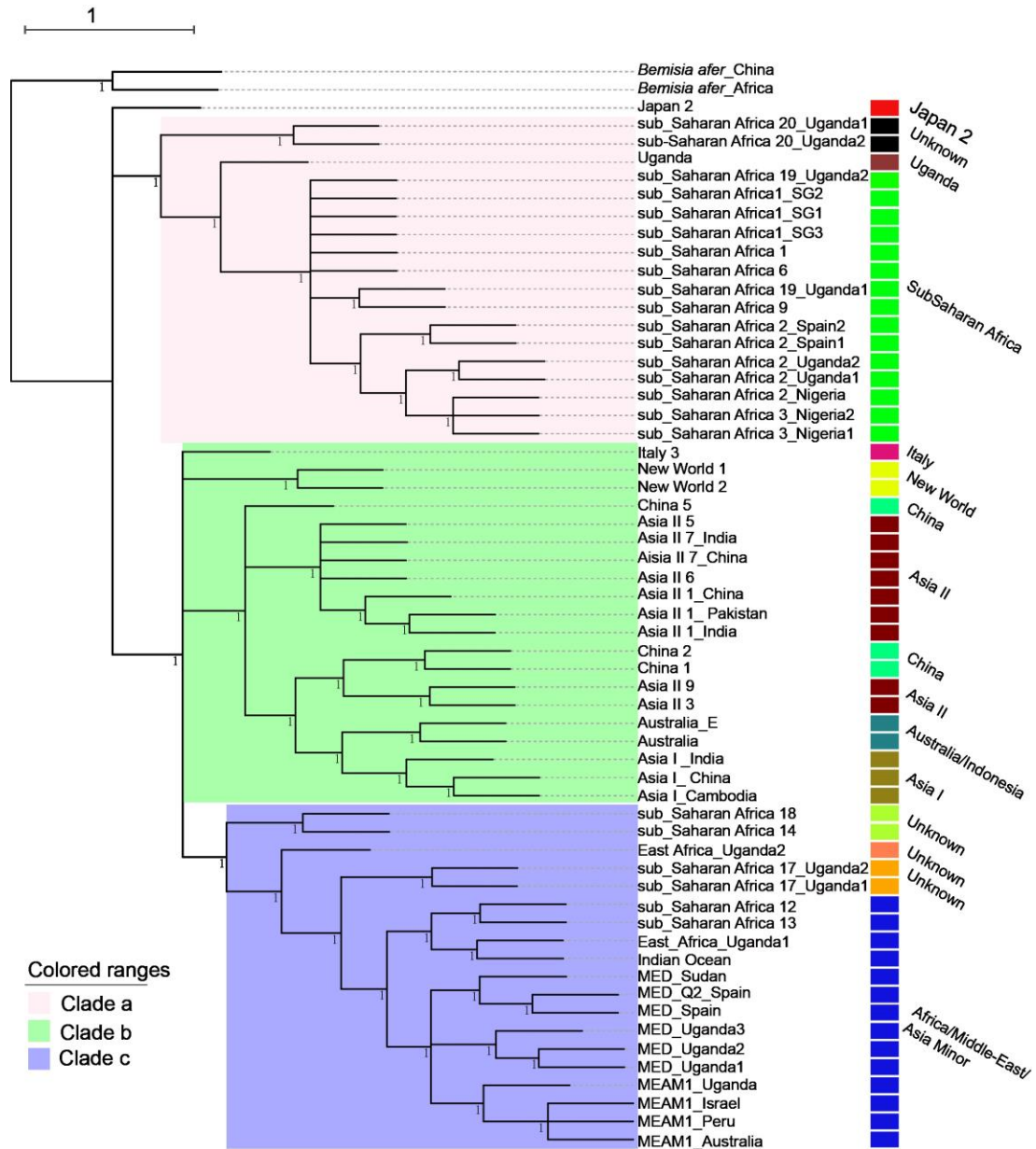
1175

1176

1177

1178

Fig. 3 Phylogenetic trees showing relationships of *B. tabaci* species based on the concatenation of 12 mitochondrial protein-coding genes, as well as 12S rRNA and 16S rRNA genes (totaling 12,725 bp) from mitochondrial genome. A, IQ-Tree phylogeny. B, RaxML phylogeny. Bootstrap values of <95% are shown on the tree.

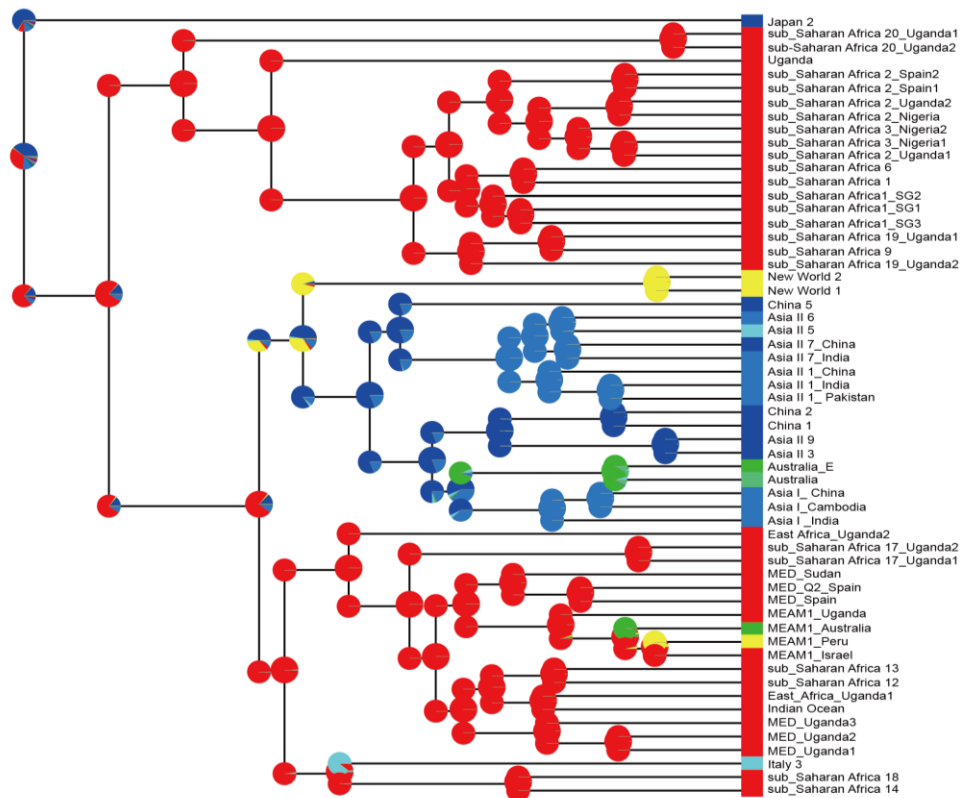
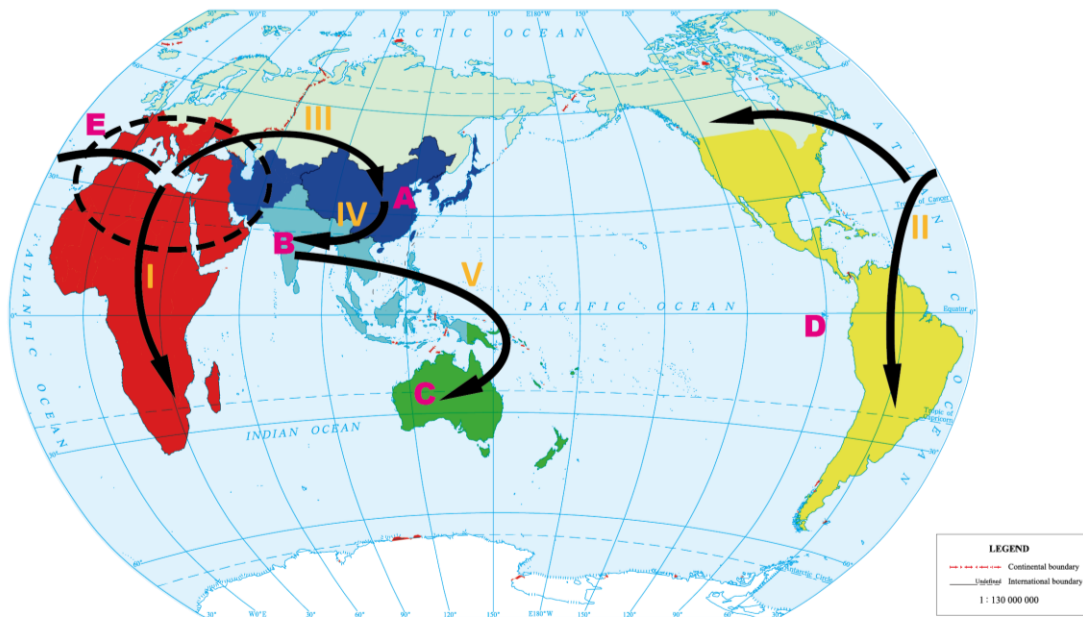


1184

1185 **Fig. 5** The strict consensus of the IQ-TREE, RaxML, ASTRAL, and ASTRID trees
 1186 shows all resolved nodes with a minimum of 95% support across the four trees.
 1187 The resulting tree is presented in the form of a cladogram, in which branch lengths do not
 1188 provide additional information. Major clades a–c are as described in the main text.

1189

1190



1191 **Fig. 6** Ancestral area reconstruction using the BAYAREALIKE+J model in
 1192 BioGeoBEARS. Tip colors represent the current biogeographic area, and pie charts
 1193 indicate predicted ancestral areas during cladogenesis. The map displays the
 1194 biogeographic areas as defined by Udvardy *et al.*, 1975: A, part of Palearctic; B,
 1195 Indomalayan; C, Australian; D, Nearctic and Neotropical; and E Afrotropical plus the
 1196 part of Palearctic.

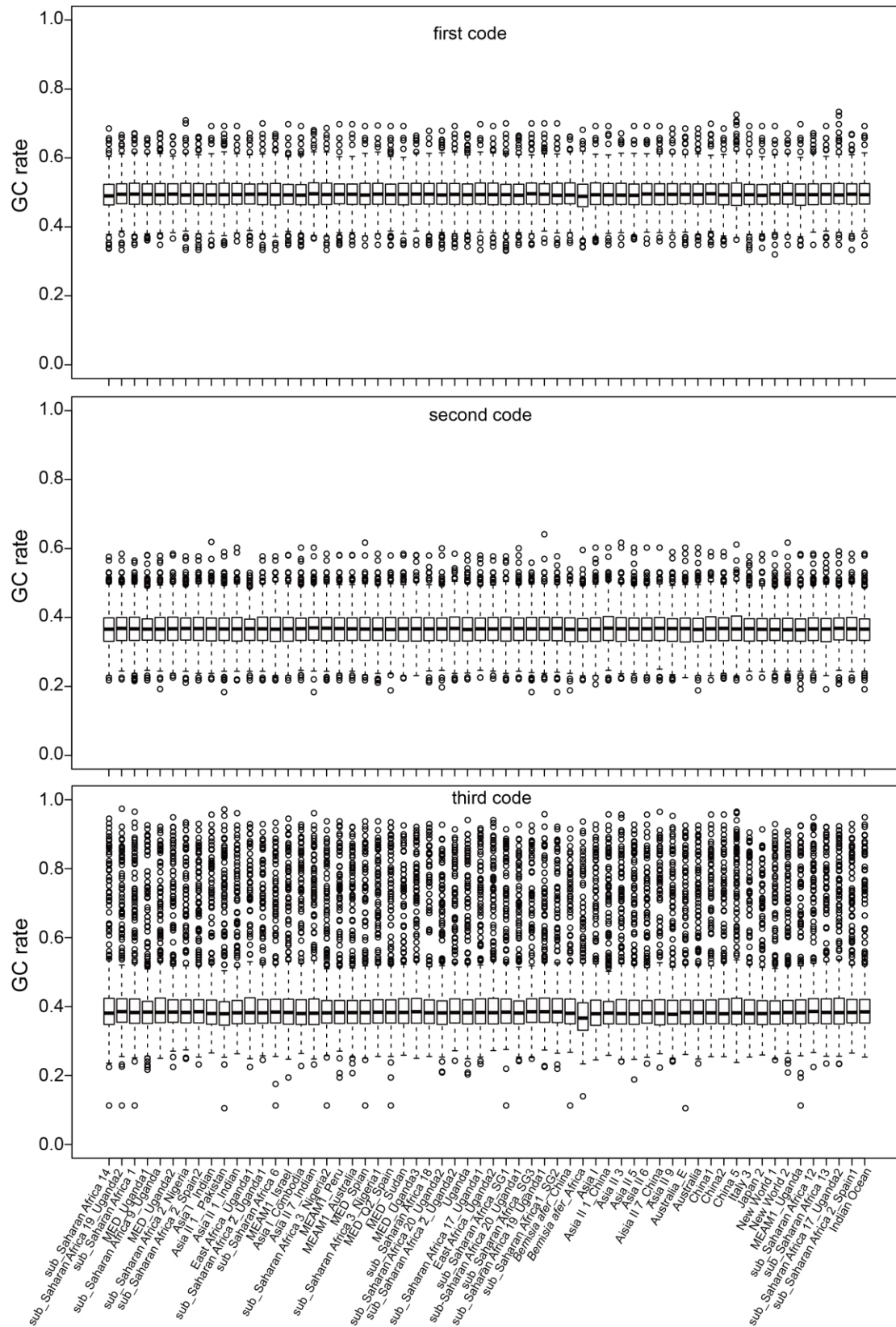
1197 **Supplementary Files:**

1198 **Table S1** Summary of sample information, sequencing results, single-copy nuclear
1199 gene assembly, and GenBank accession numbers.

1200 **Table S2** Percentage of nucleotide divergence in *mtCOI* (657 bp) and in 680 SCNs
1201 among all detected *B. tabaci* specimens.

1202 **Table S3** 33 detected putative species based on ABGD and a nucleotide sequence
1203 divergence threshold of 3.5%.

1204 **Table S4** The 1,291 single-copy nuclear genes used in this study. Gene names that are
1205 in bold indicate the 680 genes used for phylogenetic analysis.



1206

1207 **Fig. S1** GC content of the 680 SCNs for individual specimen analyzed in this study.