African Entomology 2025, 33: e20136 (20 pages)
https://doi.org/10.17159/2254-8854/2025/a20136
RESEARCH ARTICLE

AFRICAN ENTOMOLOGY
ISSN (online) 2224-8854

Multigene phylogeny of South African *Anopheles* mosquitoes

Liezl Whitehead[1], Vaughn R Swart[1]*, Marieka Gryzenhout[2] and Lizette L Koekemoer[3,4]

[1]*Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa.*
[2]*Department of Genetics, University of the Free State, Bloemfontein, South Africa.*
[3]*Wits Research Institute for Malaria, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.*
[4]*Centre for Emerging Zoonotic and Parasitic Diseases, National Institute for Communicable Diseases, a Division of the National Health Laboratory Service, Johannesburg, South Africa.*

Mosquitoes substantially impact human and animal health as vectors of disease and consequently take a heavy toll on the economy. In order to effectively investigate the evolutionary history of vectors of disease and understand their associated biological tendencies, it is vital to correctly identify and classify the relevant species. Since phylogenetic studies on South African species are currently markedly underrepresented in the literature, the current study aimed to investigate the placement of South African *Anopheles* Meigen mosquito species within the genus' extensive taxonomic framework based on the cytochrome oxidase subunit 1 (COI), internal transcribed spacer 2 (ITS2) and 28S ribosomal DNA sequences. Maximum likelihood and Bayesian phylogenetic analyses were performed for each of the COI, ITS2 and 28S DNA datasets, as well as a concatenated analysis for all three DNA regions combined. Upon examination, several phylogenetic findings were corroborated by analyses based on multiple DNA regions. These findings supported the non-monophyly of several taxa relevant to the region (subgenus *Anopheles*, Laticorn Section, and the Funestus Group) and may indicate the non-monophyly of several South African species [*An. coustani* Laveran, *An. tenebrosus* Dönitz, *An. parensis* Gillies, *An. funestus* Giles and *An. longipalpis* C (Theobald) (Type C) (Koekemoer et al. 2009)]. The results reveal numerous challenges within the current systematic framework of the genus *Anopheles* and provide a novel focus on the phylogeny of South African taxa.

CORRESPONDENCE
Vaughn R Swart

EMAIL
SwartVR@ufs.ac.za

DATES
Received: 22 September 2024
Accepted: 23 July 2025

KEYWORDS
COI
ITS2
28S
molecular phylogeny
systematics

SUPPLEMENTARY DATA
Supplementary Table S1 is available online in a separate pdf

INTRODUCTION

Mosquito taxonomy has traditionally relied on morphological characteristics to describe and classify species, which became an integral aspect of their systematics. However, boundaries between closely related mosquito species can be challenging to define, since morphological differences may be life stage-specific or restricted to a particular sex (Edwards 1941; Koekemoer et al. 2002). The morphological diversity within higher taxonomic groups similarly complicates classifications and obscures taxonomic boundaries (Harbach 2007). Furthermore, numerous taxa consist of morphologically similar or indistinguishable groups of species (Gillies and De Meillon 1968; Koekemoer et al. 1999), thereby obscuring their true identification and classification.

Despite its challenges, the morphological approach has nonetheless proved vital in establishing the current culicid systematic framework, which has only recently been revitalised with molecular technology. DNA sequence-based phylogeny has provided a powerful approach to investigating the otherwise cryptic affiliations between species. The technology facilitated deeper investigations into numerous taxa's evolutionary history, providing insight into their biodiversity, biogeography, adaptive radiation, cospeciation and evolutionary rates (Thorne et al. 1998; Huelsenbeck et al. 2000; Egan 2006).

The utilisation of several DNA regions with distinct evolutionary rates can provide resolution at varying levels of a taxon's evolutionary history. Commonly used DNA regions include the mitochondrial cytochrome c oxidase subunit I (COI), nuclear second internal transcribed spacer (ITS2) and nuclear large subunit ribosomal RNA (28S) regions. Here, the relatively high substitution rates of COI and ITS2 can be used for phylogenetic resolution at a group, species or species complex-level (Beebe 2018; Fang et al. 2017; Dassanayake et al. 2008), while the relatively conserved 28S region can provide resolution for deeper phylogenetic relationships (Pawlowski et al. 1996; Friedrich and Tautz 1997; Marinucci et al. 1999). In order to generate more stable phylogenies, investigations often incorporate data from multiple genes, thereby improving phylogenetic resolution, clade support, and the utility of phylogenetic results (Devulder et al. 2005; Baker et al. 1997; Mitchell et al. 2000; Sung et al. 2007). Such a multi-gene approach has also been employed on a small scale to elucidate culicid relationships and lineages (Puslednik et al. 2012; Greni et al. 2018).

Many research efforts have been directed towards the classification and taxonomy of mosquitoes, especially within the genus *Anopheles* Meigen, due to their status as vectors of malaria (Sharp et al. 2007). *Anopheles* was first described as a genus by Meigen (1818) and currently consists of eight subgenera. The internal structure of the genus has been adapted throughout its history, where groups were initially used as the primary divisions of subgenera. This was later adapted by Reid and Knight (1961) where subgenera were divided into sections, which in turn were subdivided into series (Table 1).

A few of these taxonomic subdivisions are relevant to South Africa, with representative species of these taxa occurring in the region. This includes the subgenus *Christya* (currently consisting of two species), the subgenus *Anopheles* (consisting of more than 200 species), two of this subgenus'

Table 1: Major taxonomic divisions of the genus *Anopheles*, adapted from Harbach (2023).

<table>
<tr><th>FAMILY</th><th colspan="21">Culicidae</th></tr>
<tr><th>SUBFAMILY</th><th colspan="21">Anophelinae</th></tr>
<tr><th>GENUS</th><th colspan="21">*Anopheles*</th></tr>
<tr><th>SUBGENUS</th><td colspan="5">*Anopheles*</td><td rowspan="3">*Baimaia*</td><td colspan="6" rowspan="2">*Cellia*</td><td rowspan="3">*Christya*</td><td rowspan="3">*Kerteszia*</td><td rowspan="3">*Lophopodomyia*</td><td colspan="5">*Nyssorhynchus*</td><td rowspan="3">*Stethomyia*</td></tr>
<tr><th>SECTION</th><td colspan="3">Angusticorn</td><td colspan="2">Laticorn</td><td colspan="2">Albimanus</td><td colspan="2">Argyritarsis</td><td rowspan="2">Myzorhynchella</td></tr>
<tr><th>SERIES</th><td>Anopheles</td><td>Cycloleppteron</td><td>Lophoscelomyia</td><td>Arribalzagia</td><td>Myzorhynchus</td><td>Cellia</td><td>Myzomyia</td><td>Neocellia</td><td>Neomyzomyia</td><td>Paramyzomyia</td><td>Pyretophorus</td><td>Albimanus</td><td>Oswaldoi</td><td>Albitarsis</td><td>Argyritarsis</td></tr>
</table>

subdivisions (the Laticorn Section and Myzorhynchus Series) and the subgenus *Cellia* (consisting of more than 230 species), the latter of which is well-represented in South Africa. Representative species of each of *Cellia*'s six series occur in South Africa (Cellia, Myzomyia, Neocellia, Neomyzomyia, Paramyzomyia and Pyretophorus). The subgenera and series of *Anopheles* are largely further subdivided into numerous groups, subgroups and species complexes. The structure of the taxonomic subdivisions relevant to the current discussions are listed in Table 2, while the species relevant to South Africa in the current analyses are listed in Table 3.

Despite extensive research efforts directed towards the systematics of *Anopheles*, current taxonomic divisions may still not always reflect the true evolutionary history of its members. Many morphological and genetic analyses have recovered the non-monophyly of numerous taxonomic divisions, including the subgenus *Anopheles* (Harbach and Kitching 2005; Harbach and Kitching 2016; Wang et al. 2017), its Laticorn Section (Sallum et al. 2002; Harbach and Kitching 2005; Harbach and Kitching 2016; Foster et al. 2017), the subgenus *Cellia* (Gholizadeh et al. 2013; Wang et al. 2017) and its Funestus Group (Norris and Norris 2015), especially within datasets with a greater taxonomic coverage. Harbach and Kitching (2016) has also expressed doubts over the monophyly of the subgenus *Anopheles* and the validity of the internal taxonomic structure of *Cellia*.

Many prior comprehensive phylogenetic studies relied on morphological data as a foundation for their analyses (Reinert et al. 2009; Harbach et al. 2012; Harbach and Kitching 2016). Datasets often focussed on representatives of more established taxa, with the exclusion of many South African species. Only a handful of studies have examined the phylogeny of South African *Anopheles* (Koekemoer et al. 2009; Norris and Norris 2015). These analyses either focused on the relationships of specific taxa or consisted of datasets with a relatively small number if species. Previous analyses also did not examine the placement of the taxa within the broader phylogenetic context of the genus. Therefore, relatively little is known about the genetic diversity, affiliations and evolutionary history of South African *Anopheles* mosquito species in relation to the various subgenera, subdivisions and other species occurring in the world.

The current study therefore aimed to examine the intrageneric relationships of South African *Anopheles* species within the broader systematic framework. DNA sequences of the target regions were obtained from DNA databases (GenBank and BOLD) to include the major *Anopheles* taxonomic divisions and representatives of publicly available South African species. The datasets additionally included sequences obtained from sampled South African *Anopheles* specimens, which were part of a larger research effort to sample and sequence South African mosquitoes. The constructed datasets were used to conduct Bayesian and maximum likelihood phylogenetic analyses for each individual DNA region (COI, ITS2 and 28S), as well as a concatenated dataset consisting of all three target regions combined. The results were examined for consistent and well-supported relationships, which represented a common phylogenetic signal shared between the various DNA regions. The generated results were compared with the relationships recovered by numerous other authors, which uncovered several consistent findings within the available literature.

MATERIALS AND METHODS

Sampling

Mosquitoes were sampled across 24 sampling sites in central South Africa (Free State Province) (Figure 1) representing diverse habitat types (urban, semi-urban, rural and pristine sites, smallholdings and farms) as part of a larger research effort to sample and sequence South African mosquitoes. Various sampling methods were employed, which included a carbon dioxide (CO_2) baited net, a CO_2 baited suction trap based on the CDC light trap (Sudia and Chamberlain 1962), hand collection and sweep netting. A total of 4 197 mosquitoes were sampled across the various sampling sites, where all sampled species belonged to one of five genera (*Aedes* Meigen, *Anopheles*, *Culex* Linnaeus, *Culiseta* Felt and *Mansonia* Blanchard). The sampled individuals were identified based on morphology and samples consisted of 11 subgenera and 26 morphospecies. The vast majority of sampled specimens belonged to *Aedes* (83%) and *Culex* (17%), while less than 1% of specimens collectively belonged to *Anopheles*, *Culiseta* and *Mansonia*.

A relatively low abundance of *Anopheles* was observed in the region, and sampling efforts yielded specimens that were morphologically identified as members of either *An. cydippis* de Meillon or *An. squamosus* Theobald, based on the keys provided by Gillies and Coetzee (1987). Since sampling efforts focused on adult mosquitoes, larval characteristics were not available to differentiate between *An. cydippis* or *An. squamosus* (Coetzee 2020), and these individuals were therefore listed as

Table 2: Taxonomic divisions of the genus *Anopheles* relevant to the current phylogenetic discussions, adapted from Harbach (2023).

Taxon					
Subgenus	**Section**	**Series**	**Group**	**Subgroup**	**Complex**
Anopheles	Angusticorn	Anopheles			Claviger
Anopheles	Angusticorn	Anopheles	Lindesayi		
Anopheles	Angusticorn	Anopheles	Maculipennis		
Anopheles	Angusticorn	Anopheles	Plumbeus		
Anopheles	Angusticorn	Anopheles	Pseudopunctipennis		
Anopheles	Angusticorn	Anopheles	Punctipennis		
Anopheles	Angusticorn	Cycloleppteron			
Anopheles	Laticorn	Arribalzagia			
Anopheles	Laticorn	Myzorhynchus	Albotaeniatus		
Anopheles	Laticorn	Myzorhynchus	Bancroftii		
Anopheles	Laticorn	Myzorhynchus	Barbirostris	Barbirostris	
Anopheles	Laticorn	Myzorhynchus	Coustani		
Anopheles	Laticorn	Myzorhynchus	Hyrcanus		
Anopheles	Laticorn	Myzorhynchus	Hyrcanus	Lesteri	
Baimaia					
Cellia		Cellia	Squamosus		
Cellia		Myzomyia	Demeilloni		
Cellia		Myzomyia	Funestus	Aconitus	
Cellia		Myzomyia	Funestus	Culicifacies	
Cellia		Myzomyia	Funestus	Funestus	
Cellia		Myzomyia	Funestus	Minimus	Fluviatilis
Cellia		Myzomyia	Funestus	Minimus	Minimus
Cellia		Myzomyia	Funestus	Rivulorum	
Cellia		Myzomyia	Marshallii		
Cellia		Myzomyia	Wellcomei		
Cellia		Neocellia	Annularis		
Cellia		Neocellia	Jamesii		
Cellia		Neocellia	Maculatus		
Cellia		Neomyzomyia		Annulipes	
Cellia		Neomyzomyia	Ardensis		
Cellia		Neomyzomyia	Kochi		
Cellia		Neomyzomyia	Leucosphyrus		
Cellia		Neomyzomyia	Punctulatus		
Cellia		Neomyzomyia	Punctulatus	Farauti	
Cellia		Neomyzomyia	Tessellatus		
Cellia		Paramyzomyia	Listeri		
Cellia		Pyretophorus		Gambiae	
Cellia		Pyretophorus		Subpictus	
Cellia		Pyretophorus		Sundaicus	
Kerteszia					
Lophopodomyia					
Nyssorhynchus	Argyritarsis	Albitarsis	Braziliensis		
Nyssorhynchus	Argyritarsis	Argyritarsis	Argyritarsis		
Nyssorhynchus	Argyritarsis	Argyritarsis	Pictipennis		
Nyssorhynchus	Myzorhynchella				
Stethomyia					

Table 3: Species relevant to South Africa included in the current analyses.

Species	Taxon					
	Subgenus	Section	Series	Group	Subgroup	Complex
An. coustani	*Anopheles*	Laticorn	Myzorhynchus	Coustani		
An. tenebrosus	*Anopheles*	Laticorn	Myzorhynchus	Coustani		
An. pharoensis	*Cellia*		Cellia			
An. cf. *cydippis / squamosus*	*Cellia*		Cellia	Squamosus		
An. squamosus	*Cellia*		Cellia	Squamosus		
An. demeilloni	*Cellia*		Myzomyia	Demeilloni		
An. funestus	*Cellia*		Myzomyia	Funestus	Funestus	
An. longipalpis C	*Cellia*		Myzomyia	Funestus	Funestus	
An. parensis	*Cellia*		Myzomyia	Funestus	Funestus	
An. leesoni	*Cellia*		Myzomyia	Funestus	Minimus	
An. rivulorum	*Cellia*		Myzomyia	Funestus	Rivulorum	
An. marshallii	*Cellia*		Myzomyia	Marshallii		Marshallii
An. theileri	*Cellia*		Myzomyia	Wellcomei		
An. maculipalpis	*Cellia*		Neocellia			
An. pretoriensis	*Cellia*		Neocellia			
An. rufipes	*Cellia*		Neocellia			
An. nili	*Cellia*		Neomyzomyia	Ardensis		Nili
An. merus	*Cellia*		Pyretophorus			Gambiae

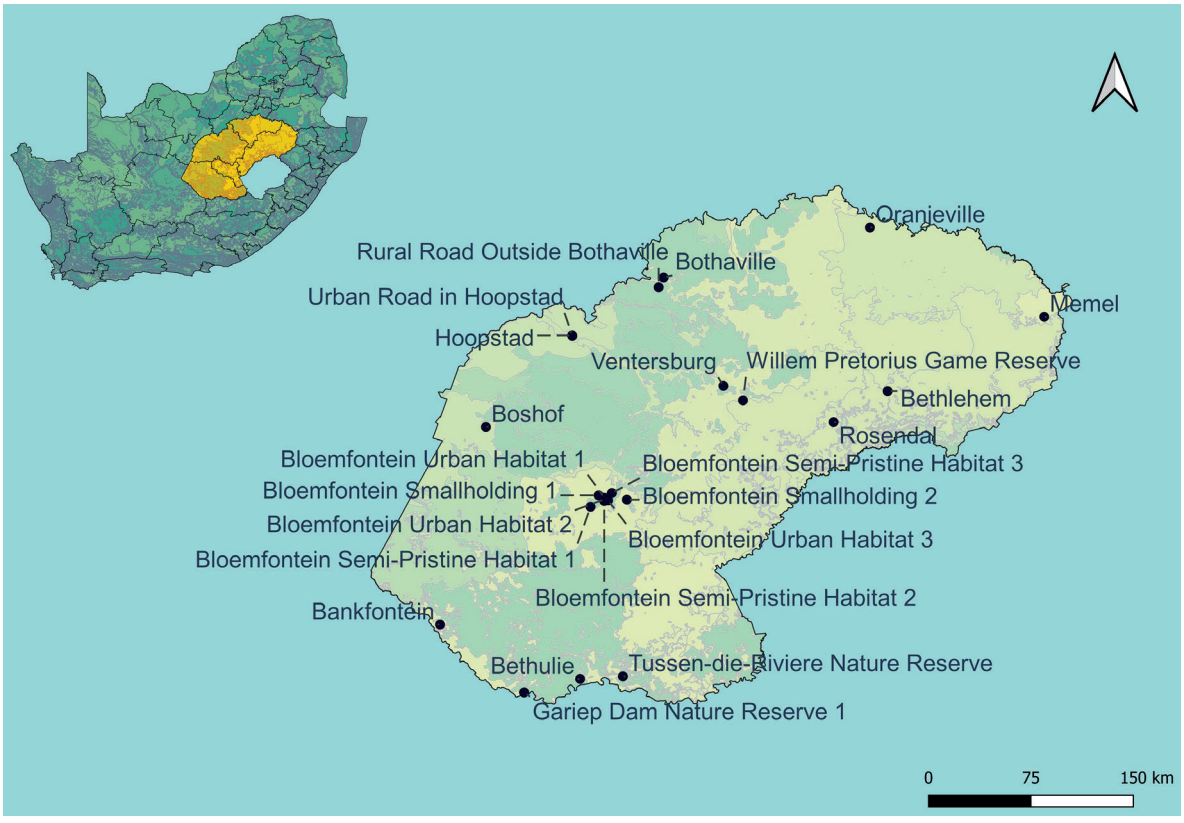

Figure 1: Sampling site localities within the Free State Province, South Africa.

An. cf. *cydippis* / *squamosus*. The coordinates and habitat type (Mucina and Rutherford 2006) of the sampling sites where *Anopheles* specimens were sampled are listed in Table 4.

Sequencing

Sampled South African mosquitoes consisting of several genera and numerous species were sequenced for the COI, ITS2 and 28S DNA regions as part of the larger research effort. For *Anopheles*, one specimen from three different sampling sites were selected for sequencing. Sequencing preparations consisted of DNA extraction with the DNeasy® Blood & Tissue Kit (Qiagen, cat. no. 69504), using either the legs or head of the specimens as the source of DNA, which were manually homogenised prior to DNA extraction. The DNA was amplified with reactions (50 µl) consisting of 1.25 units of GoTaq® Hot Start Polymerase / GoTaq® G2 Hot Start Taq Polymerase (Promega, cat. no. M5001, M7401),

Table 4: Coordinates of the sampling sites where *Anopheles* mosquitoes were collected.

Sampling site	Coordinates	Habitat type	Elevation above sea level	Vegetation type (Mucina and Rutherford 2006)
Rosendal	28°30'12.1" S, 27°55'53.3" E	Semi-urban	1 697 m	Eastern Free State Sandy Grassland
Bloemfontein Semi-Pristine Habitat 1 (Kloofeind)	29°09'37.4" S, 26°02'56.5" E	Semi-pristine	1 429 m	Winburg Grassy Shrubland
Bloemfontein Semi-Pristine Habitat 3 (The Kloof)	29°03'11.3" S, 26°12'50.8" E	Semi-pristine	1 379 m	Winburg Grassy Shrubland
Willem Pretorius Game Reserve	28°20'06.0" S, 27°13'47.4" E	Pristine	1 378 m	Central Free State Grassland

0.2 mM of dNTPs (New England Biolabs® (UK) Ltd, cat. no. N0447S), 1 x 5X Green GoTaq® Flexi Buffer (Promega, cat. no. M8911), 2.5 mM of $MgCl_2$ (Promega, cat. no. A3511), 0.4 µM of each primer and 2 µl of DNA template.

The COI region was amplified with the primer pair listed in Folmer et al. (1994); LCO1490: 5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3' and HCO2198: 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3'. The second region, ITS2, was amplified with the primers used by Djadid et al. (2007); 5.8 s: 5'-ATC ACT CGG CTC GTG GAT CG-3' and 28 s: 5'-ATG CTT AAA TTT AGG GGG TAG TC-3'. The final target region, the domain 1 (D1) portion of the 28S large ribosomal subunit RNA gene, was amplified with the primers provided by Friedrich and Tautz (1997); DIF: 5'-CCC (C/G)CG TAA (T/C)TT AAG CAT AT-3' and DIR: 5'-ACT CTC TAT TCA (A/G)AG TTC TTT (G/C)-3'.

DNA was amplified with a T100™ Thermal Cycler (Bio-Rad) by applying the following cycling conditions: initial denaturation at 95°C for 2 minutes, followed by 40 cycles of denaturation (95 °C for 30 s), annealing (COI: 45 °C; ITS2: 52 °C; 28S: 44 °C for 30 s) and extension (72 °C for 1 min), concluded with a final extension for 5 minutes at 72 °C.

After visualisation of the products with agarose gel electrophoresis and GelRed® nucleic acid dye, the successfully amplified products were purified with the Wizard® SV Gel and PCR Clean-Up System (Promega, cat. no. A9281). These purified products were then used as a template for cycle sequencing using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems™, cat. no. 4337455). Finally, samples were purified with the BigDye XTerminator™ Purification Kit (Applied Biosystems™, cat. no. 4376486) and submitted to the University of the Free State's Department of Genetics for sequencing. The final generated sequences were screened against the GenBank database with the standard nucleotide BLAST function on the NCBI platform (National Center for Biotechnology Information 2020) to corroborate the species identities, as determined through morphological identifications. The generated sequences were then uploaded to BOLD (Molecular Phylogeny of South African Mosquitoes; *Anopheles* specimens: MPSAM015-21, MPSAM022-21, MPSAM061-21) and added to the relevant datasets for the phylogenetic analyses.

Dataset construction

The COI, ITS2 and 28S (D1 domain) and concatenated datasets were constructed mainly from sequences obtained from GenBank (Benson et al. 2012) and BOLD (Ratnasingham and Hebert 2007), however generated sequences from the sampled South African specimens were also included in the datasets. The concatenated dataset consisted of the COI and ITS2 regions and a portion of the adjacent 28S region.

Anopheles sequences were selected to represent available subgenera, sections, series and groups within the public databases. Furthermore, species groups relevant to South Africa were also represented by species from each available subgroup and complex. Accessible representatives of South African species were included (Table 3), consisting of a total of 18 species relevant to the region. The genus was represented by five subgenera, five sections, 13 series, 31 groups and 62 species across the various datasets, and species were represented by three distinct sequences as far as possible. South African species were generally poorly represented within the public DNA databases, with only a small portion of the region's taxa having been sequenced.

Multiple sequence alignment

COI and 28S (D1 Domain) sequences were aligned with the MAFFT version 7 online platform using the default parameters (Katoh et al. 2019). This included the automatic selection of the optimal strategy with a 200PAM / K = 2 scoring matrix and a gap opening penalty of 1.53. All generated alignments were manually inspected and corrected if necessary. Due to the relative fast evolution rate of the ITS2 region and the presence of stretches of varying lengths of tandem repeat sequences in numerous *Anopheles* species (Otsuka 2011), this region was aligned with the use of its more conserved secondary structures (Coleman 2003) as a guide. An online platform, the ITS2 database (Merget et al. 2012), was used to retrieve a list of all available *Anopheles* ITS2 sequences with predicted secondary structure annotations to serve as a template. The custom modelling function of the database with default parameters was then used to generate secondary structures for the provided ITS2 datasets based on these templates. The dataset alignment was based on both the nucleotide sequences and the generated secondary structures, which was executed with 4SALE 1.7.1 using CLUSTAL W 1.83 (Seibel et al. 2006; Seibel et al. 2008). Since most publicly available ITS2 sequences also contained portions of the 5.8S and 28S regions, these regions were identified with the ITS2 database's ITS2-annotation function (version 3.0.13) by employing the dipteran model. The relatively conserved 5.8S and 28S regions were then used to identify any potential misalignments and trimmed from the final alignment for the ITS2 analysis, while the 28S region was retained and partitioned for the concatenated analysis.

The generated COI, ITS2, 28S and concatenated datasets were used to conduct maximum likelihood and Bayesian phylogenetic analyses with RAxML version 8.2.12 (Stamatakis 2014) and MrBayes version 3.2.3 (Huelsenbeck and Ronquist 2001). In order to determine the most appropriate substitution model based on the available shared options for these programmes (GTR+Γ & GTR+I+Γ), the concatenated dataset consisting of all three DNA regions was imported into jModeltest (Posada 2008) version 2.1.10. The likelihood computations were executed with default settings except when BIONJ was selected as the base tree. Best-fitting models were examined according to the Akaike Information Criterion with the correction for small sample sizes (Hurvich and Tsai 1989) and Bayesian Information Criterion (Schwarz 1978). Here, both model selection criteria recovered GTR+I+Γ as the model with a greater likelihood score. Therefore, this model was selected for the execution of all subsequent analyses to ensure consistency between the generated results.

The Bayesian analyses were conducted with a General Time Reversible (GTR) model with a gamma distribution and a proportion of invariable sites (nst = 6; rates = invgamma) across 20 000 000 generations with a burnin percentage of 25%. The same model was selected for the maximum likelihood analyses (-m GTRGAMMAI) with partition-specific estimations of the alpha values, substitution rates, base frequencies and branch lengths (-M). Furthermore, the parsimony inferences and rapid bootstrapping were executed (-p 12345; -f a –x 12345) over 1 000 bootstrap replicates. Since the three codon positions of COI's protein-coding regions have unique evolutionary constraints and evolve at dissimilar rates (Bofkin and Goldman 2007), the single-gene COI analyses were partitioned to account for this independent evolution. Therefore, the substitution rates, base frequencies, alpha values and proportion of invariable sites were estimated independently for each of the three COI codon position partitions. The variables of the remaining ITS2 and 28S regions were estimated for the dataset as a whole since the regions did not consist of any protein-coding sequences. Finally, the concatenated datasets were partitioned for each of the three DNA regions, where the parameters were estimated independently for the COI, ITS2 and 28S sections.

RESULTS

Sequenced specimens

The sequencing attempts of three sampled *Anopheles* specimens produced three COI sequences and one ITS2 sequence, while attempts to sequence the 28S D1 domain were unsuccessful. These sequences were included in the datasets for the COI and ITS2 analyses. Generated sequences were entered as a query within the BLAST search function (National Center for Biotechnology Information 2020), which produced the greatest matches to *An. squamosus* in the GenBank database (Benson et al. 2012), although these percentage identities did not surpass 94% (Table 5). The sampled *An.* cf. *cydippis* / *squamosus* specimens may have represented *An. cydippis* or a closely allied species, since only *An. squamosus* sequences were represented within the database for comparison. Therefore, low identity values could have been expected if the sampled specimens were members of a species other than *An. squamosus*.

Phylogenetic overview

Bayesian and maximum likelihood analyses were conducted for each of the COI, ITS2, 28S and concatenated datasets. During the interpretation of the results, branches with bootstrap support values (BS) of ≥ 70% and posterior probabilities (PP) of ≥ 95% were considered to be significantly supported, based on the thresholds employed by numerous other authors in Bayesian and maximum likelihood phylogenies (Leaché and Reeder 2002; Quenouille et al. 2004; Harris et al. 2004; Miller et al. 2004; Vinuesa et al. 2005; Jiang et al. 2006; Sung et al. 2007; Schuettpelz and Pryer 2007; Hua et al. 2016).

Within the current results, numerous relationships remained consistent and well-supported across various analyses, despite overall topological variations. These associations were examined from a South African perspective and the interpretations were centred on taxa relevant to the region.

The COI dataset was represented by five subgenera and a total of 62 *Anopheles* species. The alignment consisted of 658 sites, including 387 conserved, 271 variable and 236 parsimony-informative positions. As expected for the relatively fast-evolving COI gene region, the gene region yielded a greater degree of phylogenetic support for terminal affiliations, with relatively poor support for supraspecific relationships (Figures 2 and 3). Several terminal clades were well-supported and shared between the phylogenetic methods.

The ITS2 dataset had a similar degree of taxonomic coverage that included four subgenera and 62 *Anopheles* species. The alignment consisted of 2 212 sites, including 600 conserved, 1 292 variable and 1 036 parsimony-informative positions. The analyses produced a relatively well-resolved topology consisting of numerous well-supported clades and several species groupings that remained consistent between the two phylogenetic methods (Figures 4 and 5). However, the most basal portions of the phylogeny were not well-supported, which may have contributed to the non-monophyly of most subgeneric clades.

The 28S dataset consisted of relatively fewer taxa, due to the limited availability of homologous sequences for the region, where the dataset was represented by a total of three subgenera and 14 species. The alignment consisted of 462 sites, including 308 conserved, 149 variable and 129 parsimony-informative positions. Despite the relatively slow evolution rate of the 28S region, the region generally did not yield a greater degree of basal resolution than the ITS2 analyses (Figures 6 and 7). The degree of support for internal relationships was varied, with a few clades being well-supported. Nonetheless, portions of the topology remained consistent between the phylogenetic methods, where the most well-represented subgenus (*Cellia*) was monophyletic.

Finally, the concatenated analyses included a dataset of moderate size, comprising five subgenera and 50 species. The combined phylogenetic signal of the various DNA regions contributed to a greatly supported phylogeny within the concatenated analyses, where a relatively large portion of both terminal, intermediary and basal cades were well-supported (Figures 8 and 9). Here, two well-represented subgenera (*Anopheles* and *Cellia*) were non-monophyletic, where the overall topology remained consistent between the two phylogenetic methods.

The species relevant to South Africa included two members of the subgenus *Anopheles* and 16 species of the subgenus *Cellia*. Several of these species were non-monophyletic in the current analyses, including the subgenus *Anopheles*' *An. coustani* Laveran and the subgenus *Cellia*'s *An. longipalpis* (Theobald). Two other species of the subgenus *Cellia*'s Funestus Subgroup were also sporadically recovered as non-monophyletic (*An. parensis* Gillies and *An. funestus* Giles). The current analyses additionally recovered affiliations of various other South African species within the subgenus *Cellia*. *Anopheles nili* (Theobald), a member of the Neomyzomyia Series, shared a well-supported clade with the Neocellia Series, while sampled individuals of *An.* cf. *cydippis* / *squamosus* formed a well-supported clade distinct from other non-sampled African *An. squamosus* specimens. The current analyses also revealed relationships within *Cellia*'s Myzomyia Series, which included the affiliations of *An. marshallii* (Theobald), *An. theileri* Edwards, *An. dthali* Patton and *An. demeilloni* Evans with the Funestus Group, and *An. leesoni*'s Evans association with the Minimus and Fluviatilis Complexes.

Table 5: BLAST percentage identity values for sequenced *Anopheles* specimens.

Sampled specimens		BLAST results			
Species	Specimen and BOLD Process ID	COI		ITS2	
An. cf. *cydippis / squamosus*	ANO7.1 (MPSAM022-21)	*An. squamosus*	94%		
An. cf. *cydippis / squamosus*	ANO22.1 (MPSAM061-21)	*An. squamosus*	87%		
An. cf. *cydippis / squamosus*	ANO3.1 (MPSAM015-21)	*An. squamosus*	92%	*An. squamosus*	94%

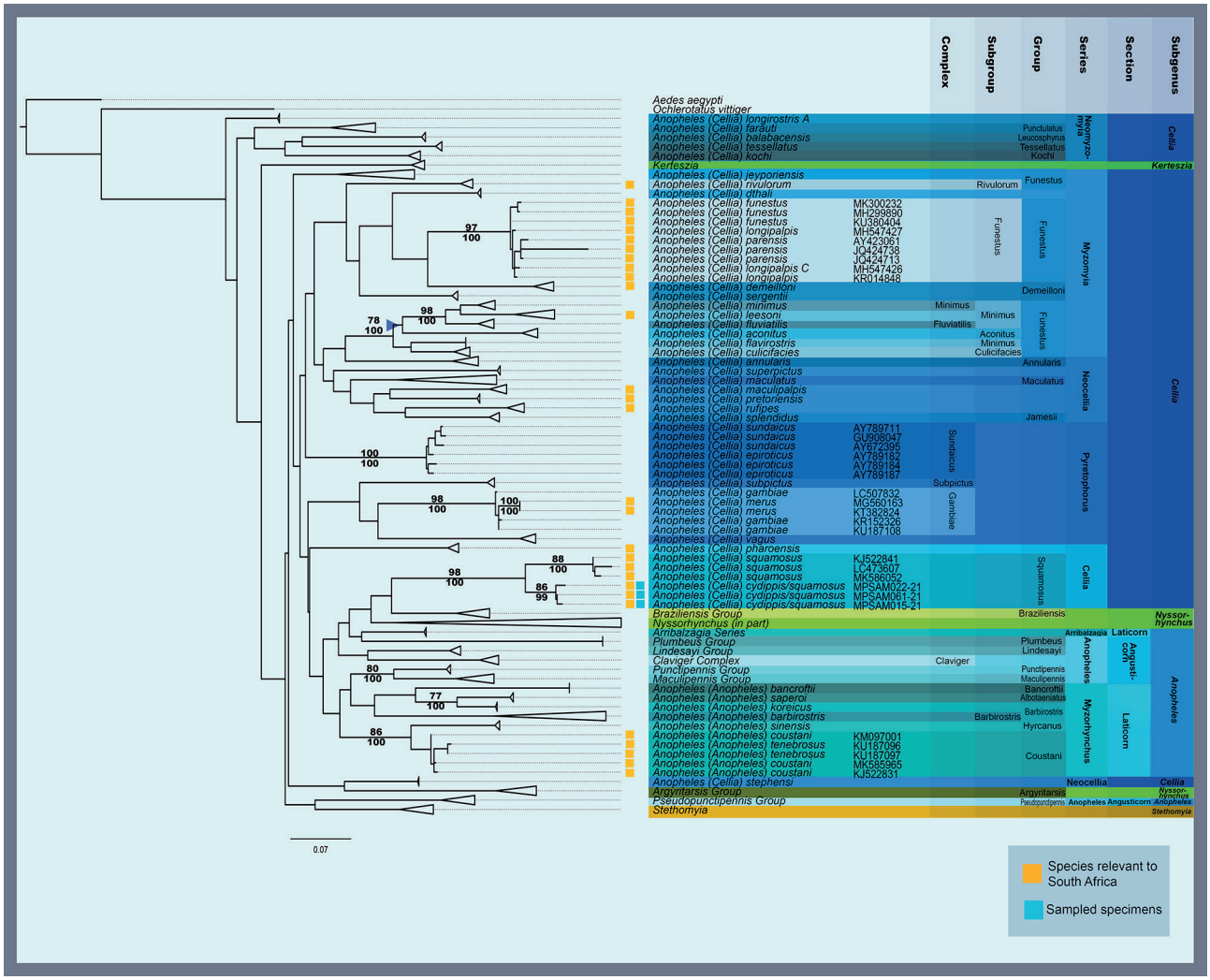

Figure 2: Phylogenetic relationships and taxonomic divisions of *Anopheles* species based on the maximum likelihood analysis of COI sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades.

Anopheles subgenera relevant to South Africa

Subgenus Anopheles

The subgenus was represented by a total of 24 species across the various datasets, including two Coustani Group (Laticorn Section, Myzorhynchus Series) members relevant to South Africa (*An. coustani* and *An. tenebrosus* Dönitz). All analyses consistently recovered the subgenus *Anopheles* as a non-monophyletic assemblage, although the specific topology differed between DNA regions. The COI analyses yielded weak basal support with disjunct clusters of this subgenus. The remaining analyses provided clades with greater support, where some of the ITS2, 28S and concatenated analysis clusters were well-supported. The non-monophyly of *Anopheles* contributed to the polyphyly of two of its sections, Angusticorn and Laticorn, where these sections were often intermixed with one another or affiliated with other taxa. In most analyses, both sections consisted of numerous disjointed clades, while the sections were monophyletic in the 28S results, likely due to the dataset's limited taxonomic coverage.

The COI analyses (Figures 2 and 3) produced a relatively poorly supported basal topology, yet both phylogenetic methods recovered several well-supported terminal clades. This included the sister relationship between the Myzorhynchus Series' (Laticorn Section) Hyrcanus + Coustani Groups (PP 100, BS 86), Myzorhynchus' Albotaeniatus Group + Barbirostris Group in part (*An. koreicus* Yamada & Watanabe) (PP 100, BS 77) and the Anopheles Series' (Angusticorn Section) Punctipennis + Maculipennis Groups (PP 100, BS 80).

The subgenus *Anopheles* once again consisted of several disjointed clades in the ITS2 analyses (Figures 4 and 5). The subclades included groupings of the Arribalzagia Series (Laticorn Section) + Maculipennis Group + Claviger Complex (Angusticorn Section) (PP 100, BS 93), the Pseudopunctipennis + Lindesay Groups (Anopheles Series) (PP 100, BS 100), and a Myzorhynchus clade (in part) consisting of the Barbirostris Group's *An. koreicus* + Albotaeniatus Group + Hyrcanus Group + Coustani Group + Bancroftii Group (PP 98, BS 72).

The 28S analyses yielded relatively poor support for basal relationships (Figures 6 and 7), which may have contributed to the non-monophyletic structure of the subgenus *Anopheles*. The subgenus was represented by relatively few species and consisted of two clusters, where the Laticorn Section's Myzorhynchus Series clade was well-supported (PP 100, BS 92).

Finally, similar to the results from the other analyses, the concatenated analyses produced disjointed groupings of the subgenus *Anopheles* (Figures 8 and 9), where each cluster contained members of both the Angusticorn and Laticorn Sections, thus rendering these sections polyphyletic. One well-supported clade (PP 100, BS 100) largely consisted of Laticorn (Myzorhynchus and Arribalzagia's *An. peryassui* Dyar & Knab) with the inclusion of Angusticorn's Pseudopunctipennis Group. However, this clade also included *Cellia's An. funestus*, which was not recovered in close association with the subgenus *Anopheles*

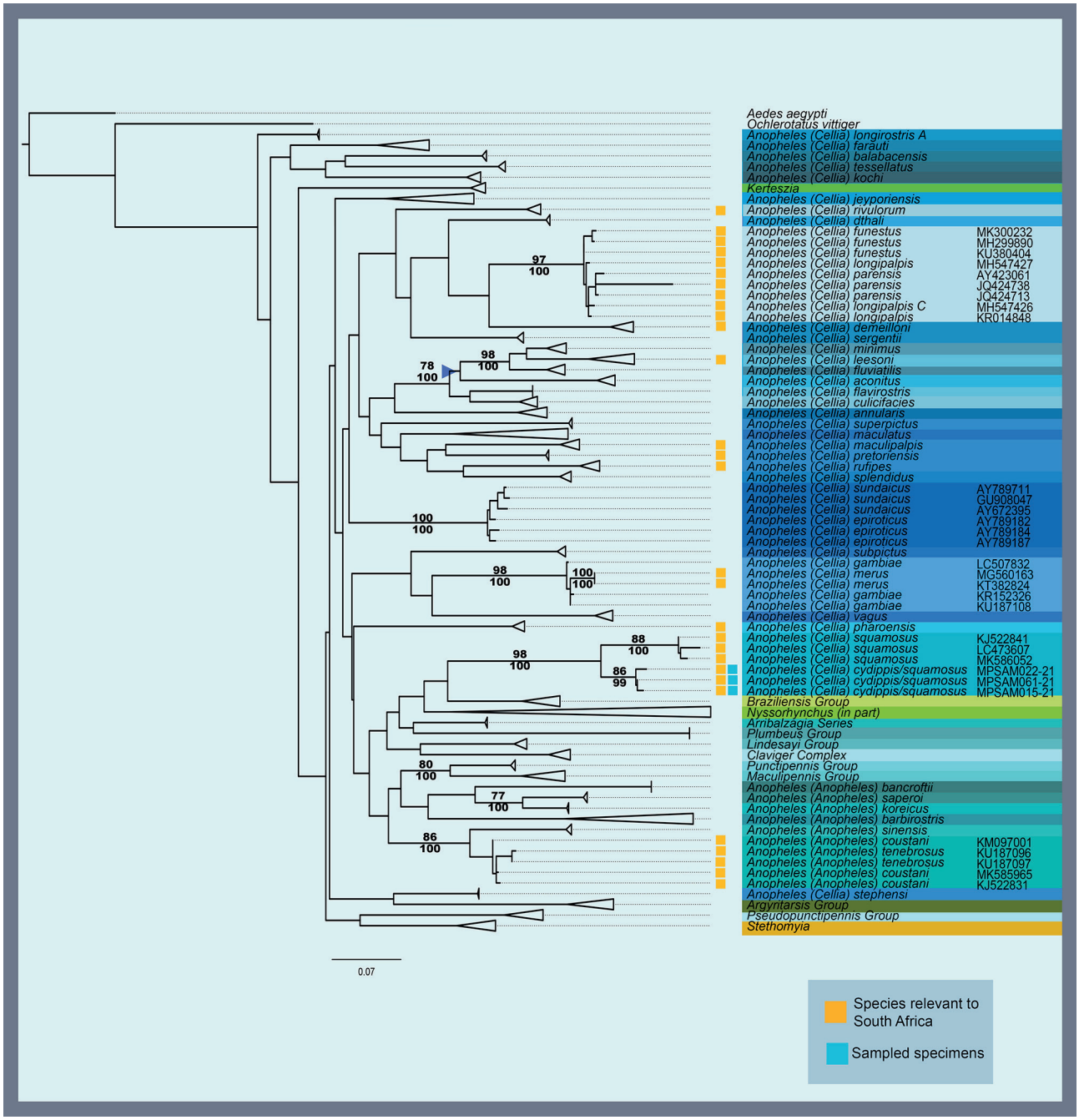

Figure 3: Phylogenetic relationships of *Anopheles* based on the maximum likelihood analysis of COI sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades (cropped figure).

in any of the other analyses. Both the Anopheles (Angusticorn) and Arribalzagia (Laticorn) Series were also polyphyletic, where one Arribalzagia cluster was closely associated with the subgenus *Stethomyia* (PP 100, BS 100). Several of the affiliations recovered in the COI and ITS2 analyses could not be examined in the concatenated analyses, since no DNA sequence data were available for the Barbirostris, Bancroftii, Coustani, Lindesayi, Punctipennis and Albotaeniatus Groups.

Despite the use of independent phylogenetic methods and various DNA regions, several relationships were robustly and frequently recovered by multiple analyses. All analyses that included *An. saperoi* Bohart & Ingram and *An. koreicus* as representatives of the Myzorhynchus Series, recovered their well-supported sister relationship (COI: PP 100, BS 77; ITS2: PP 100, BS 100). *Anopheles koreicus* from the Barbirostris Group was consistently more closely related to the Albotaeniatus Group than *An. barbirostris* van der Wulp, thereby rendering the Barbirostris Group polyphyletic. Furthermore, all analyses that included members of Myzorhynchus' Coustani and Hyrcanus Groups confirmed their well-supported sister relationship (COI: PP 100, BS 86; ITS2: PP 99, BS 83).

Several instances of non-monophyly were included within well-supported and consistently recovered clades. Datasets with greater taxonomic coverage generally revealed unique sets of affiliations, since the subgenus *Anopheles* tended to form at least two distinct well-supported clusters in such analyses. Here, within the concatenated and ITS2 analyses, these clusters either included Laticorn's Myzorhynchus Series and Angusticorn's Pseudopunctipennis Group (ITS2; PP 97, BS 79, concatenated; PP 100, BS 100) or Laticorn's Arribalzagia Series (either in part or as a whole) and Angusticorn's Maculipennis Group (ITS2: PP 100, BS 93). These results represented well-supported non-monophyletic associations for the subgenus *Anopheles*' two sections (Angusticorn and Laticorn), and the Anopheles and

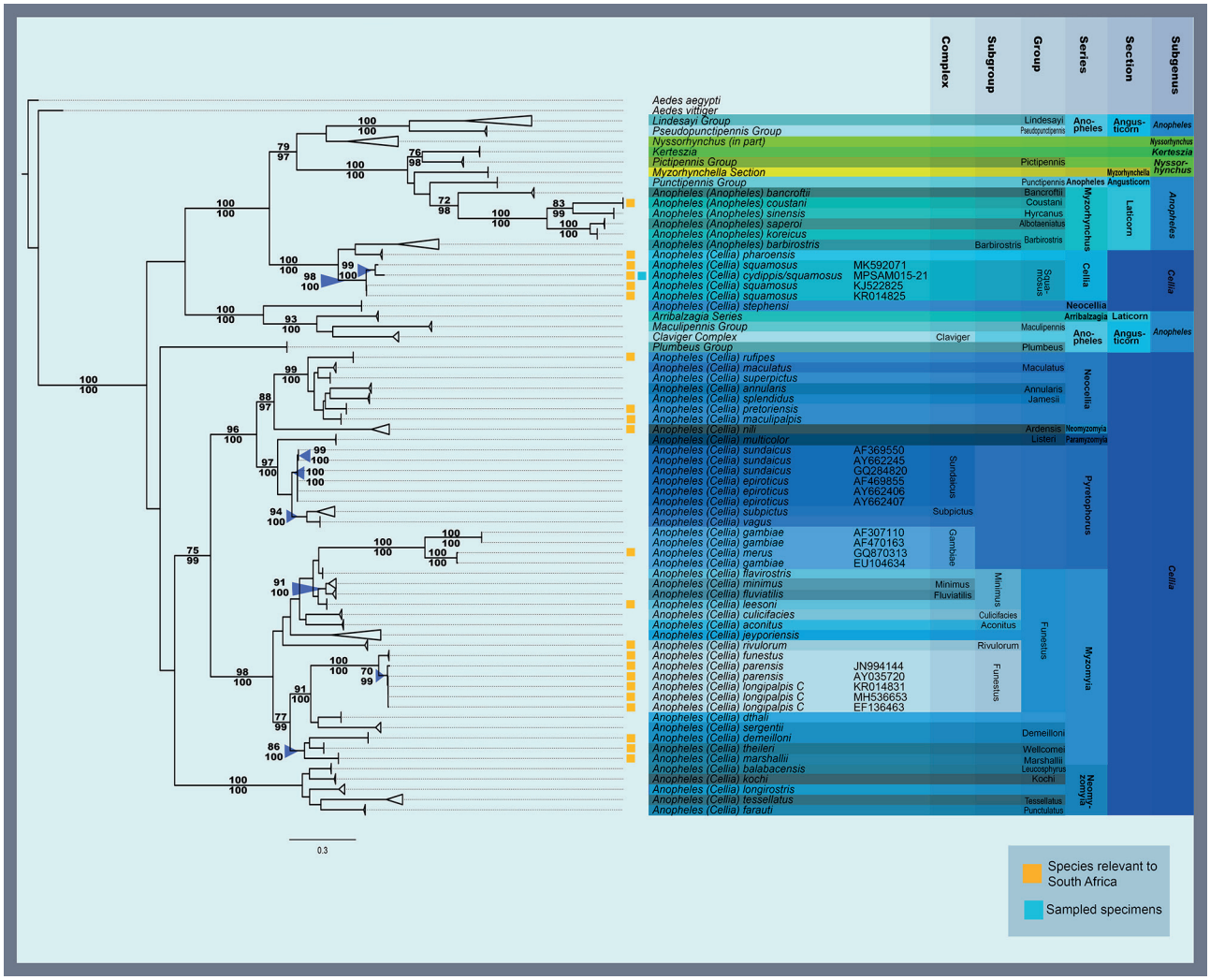

Figure 4: Phylogenetic relationships and taxonomic divisions of *Anopheles* based on the maximum likelihood analysis of ITS2 sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades.

Arribalzagia Series. However, the overall placement of these clades within the greater phylogeny was inconsistent.

In addition to the well-supported affiliations, the non-monophyly of the Anopheles and Myzorhynchus Series was also recovered within several analyses based on different DNA regions. The Anopheles Series was polyphyletic in all analyses with greater taxonomic coverage (COI, ITS2 and concatenated analyses). The Myzorhynchus Series was paraphyletic with the inclusion of the Anopheles Series (in part) (Maculipennis + Punctipennis Groups) in the COI analyses, while the ITS2 results recovered polyphyletic Myzorhynchus clades. However, Myzorhynchus was only represented by two species in the 28S analyses, which formed a well-supported monophyletic grouping (PP 100, BS 92). Furthermore, Myzorhynchus was solely represented by the Hyrcanus Group in the concatenated analyses, where it once again formed a monophyletic grouping. The monophyly of Myzorhynchus in the 28S and concatenated analyses were likely a result of the limited taxonomic coverage.

Within Myzorhynchus, the Coustani Group included two species relevant to South Africa, namely *An. coustani* and *An. tenebrosus*. In the COI results, the Coustani Group was monophyletic, while *An. coustani* itself was paraphyletic with the inclusion of *An. tenebrosus*. However, considering that the Coustani Group was only represented by a single species in the ITS2 analyses and the group was not represented in the 28S and concatenated datasets, its paraphyly could not be confirmed.

Subgenus Cellia

The subgenus *Cellia* was represented by more than 40 species within the *Anopheles* datasets. The datasets included 16 species relevant to South Africa, with species being members of the Neocellia Series, Cellia Series, the Myzomyia Series' Demeilloni, Funestus, Marshallii and Wellcomei Groups, the Neomyzomyia Series' Ardensis Group and the Pyretophorus Series' Gambiae Complex. The subgenus *Cellia* was also represented by three sampled individuals of *An.* cf. *cydippis* / *squamosus* (ANO3.1, ANO7.1, ANO22.1) in the COI and ITS2 datasets. All the analyses that were based on a relatively larger dataset (COI, ITS2 and concatenated analyses) recovered *Cellia* as a polyphyletic assemblage, although the overall topology was inconsistent between DNA regions. The weakly supported COI groupings yielded numerous *Cellia* clades scattered across the tree. However, the support was greatly improved in the ITS2 and concatenated analyses, where *Cellia* was largely included within more coherent and well-supported clusters. Lastly, the subgenus was included in a well-supported monophyletic clade in the 28S results (PP 100, BS 79), where it was represented by a smaller number of species. Representatives of *Cellia* were frequently associated with clusters of the subgenus *Anopheles* in multiple analyses, however the specific associations were often inconsistent.

In the COI results (Figures 2 and 3), *Cellia* consisted of numerous disjoined clusters, likely due to the region's limited support for basal

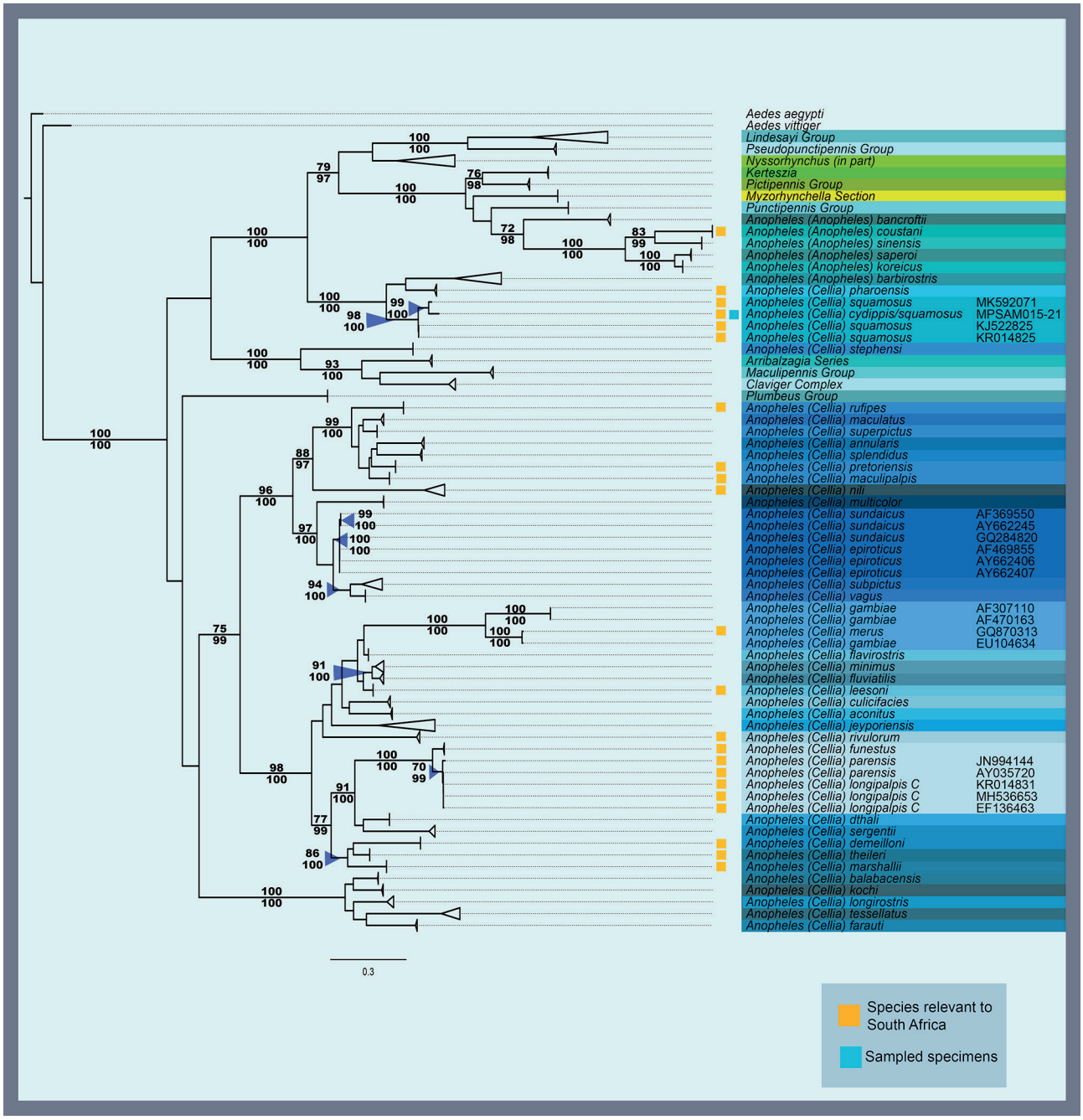

Figure 5: Phylogenetic relationships of *Anopheles* based on the maximum likelihood analysis of ITS2 sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades (cropped figure).

clades. Nonetheless, a greater degree of support was achieved for terminal relationships, where both phylogenetic methods recovered the monophyly of the Pyretophorus Series' Gambiae Complex (PP 100, BS 98) and Sundaicus Complex (PP 100, BS 100), as well as the Myzomyia Series' Funestus Subgroup (PP 100, BS 97). The region also yielded a sister relationship between the Cellia Series' *An. squamosus* and sampled individuals of *An.* cf. *cydippis / squamosus* as members of the Squamosus Group (PP 100, BS 98), and the relationship between the Funestus Group's Minimus Subgroup (in part) and its Aconitus Subgroup (PP 100, BS 78).

In the ITS2 analyses (Figures 4 and 5), *Cellia* once again consisted of disjunct clusters where two clusters were closely affiliated with the subgenus *Anopheles*. Here, a single Neocellia Series species (*An. stephensi* Liston) shared a clade with a subset of the subgenus *Anopheles* (Arribalzagia Series + Maculipennis Group + Claviger Complex) (PP 100, BS 100), while the Cellia Series shared a clade with the subgenus *Anopheles*' Barbirostris Complex (PP 100, BS 100). The remainder of the subgenus *Cellia* clustered in a single clade, where a large portion of the relationships were well-supported. This cluster included a non-monophyletic Pyretophorus Series, divided into two clades. One clade consisted of Pyretophorus' Sundaicus Complex (PP 100, BS 100) + Subpictus Complex + *An. vagus* Dönitz, which as a whole was sister to the Paramyzomyia Series (PP 100, BS 97). However, Pyretophorus' affiliation with Paramyzomyia could not be examined in the other datasets due to the series' limited sequence availability. The ITS2 analyses yielded a well-supported Neocellia Series cluster (in part) (PP 100, BS 99), which was sister to the Neomyzomyia Series' Ardensis Group (PP 97, BS 88). These analyses also recovered a well-supported Myzomyia Series cluster (PP 100, BS 98) that was rendered paraphyletic with the inclusion of Pyretophorus' Gambiae Complex. Within Myzomyia, its Funestus Group was non-monophyletic and incorporated several well-supported subclades. These clades

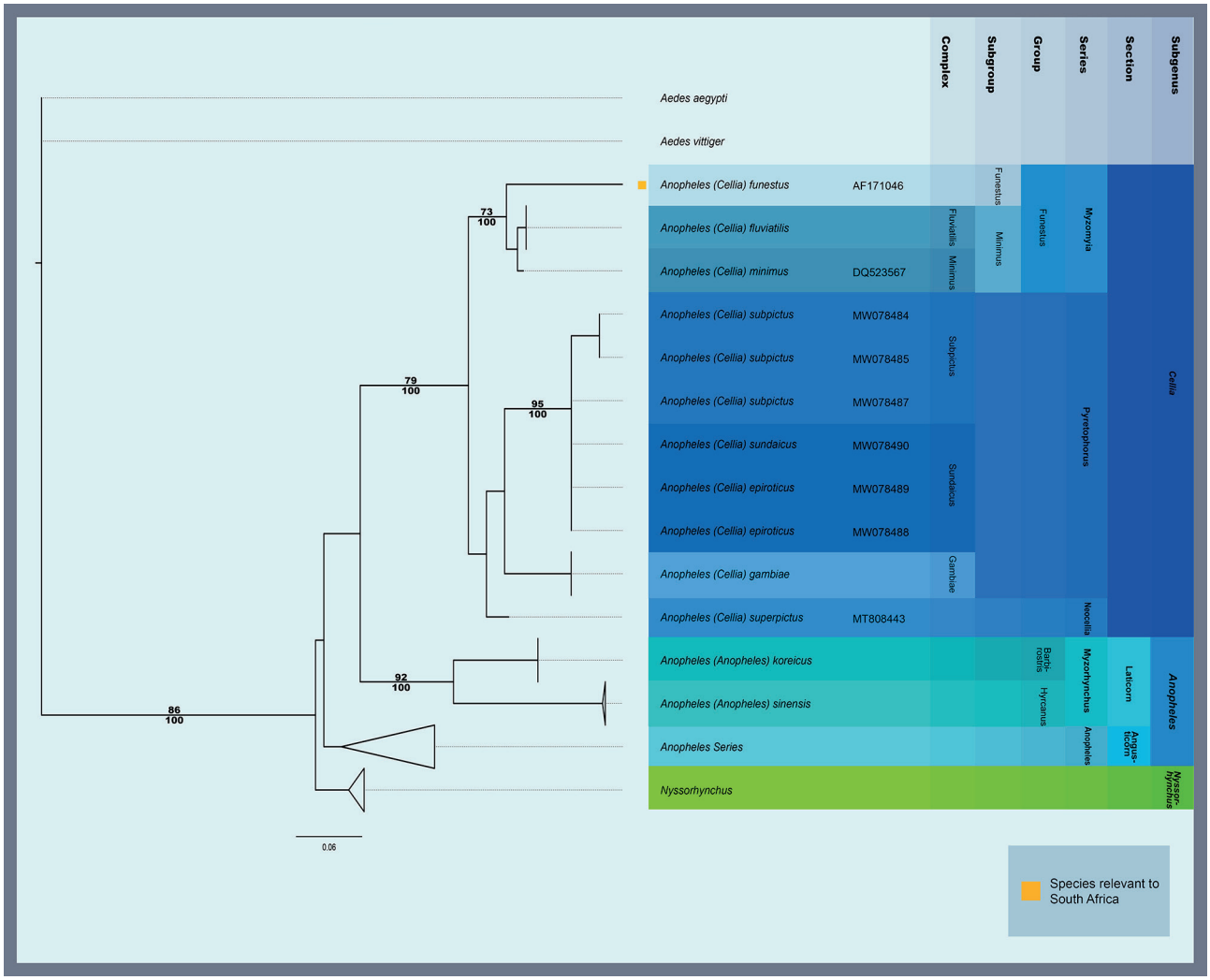

Figure 6: Phylogenetic relationships and taxonomic divisions of *Anopheles* based on the maximum likelihood analysis of 28S sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades.

consisted of its Fluviatilis + Minimus Complexes (PP 100, BS 91), a monophyletic Funestus Subgroup (PP 100, BS 100) and its sister relationship with *An. dthali* + *An. sergentii* (Theobald) (PP 100, BS 91), both species that are not members of the Funestus Group. Myzomyia also included a clade consisting of its Marshallii Group + Wellcomei Group + *An. demeilloni* (PP 100, BS 86). However, due to the limited sequence availability, the Marshallii and Wellcomei Groups were solely represented within the ITS2 datasets, where they were nestled amongst the Demeilloni and Funestus Groups, all as members of the Myzomyia Series. *Anopheles nili* (Ardensis Group) was likewise exclusively represented in the ITS2 analyses, where it was the only member of the Neomyzomyia Series that grouped separately from the rest of the series. This species formed a well-supported relationship with a cluster of the Neocellia Series (PP 97, BS 88). The remaining members of Neomyzomyia nonetheless formed a separate well-supported clade (PP 100, BS 100).

Unlike the previous results, the 28S results (Figures 6 and 7) recovered a well-supported monophyletic cluster of the subgenus *Cellia* (PP 100, BS 79), which was likely due to the limited taxonomic coverage. Here the monophyly of the Funestus Group (PP 100, BS 73) and a clade consisting of Pyretophorus' Subpictus + Sundaicus Complexes (PP 100, BS 95) were well-supported.

In the concatenated analyses (Figures 8 and 9), the subgenus *Cellia* was once again non-monophyletic, similar to the COI and ITS2 results. However, most species belonged to clusters of well-supported clades. This included a grouping consisting of the Neocellia Series + *An. aconitus* Dönitz (Myzomyia Series) + Subpictus Complex (Pyretophorus Series) + Sundaicus Complex (Pyretophorus Series) (PP 100, BS 100), where the subclade consisting of the Subpictus Complex + Sundaicus Complex was also well-supported (PP 100, BS 93). The analyses additionally supported the monophyly of several other taxa, namely the Neomyzomyia Series (PP 96, BS 91), its Punctulatus Group (PP 100, BS 97) and Farauti Complex (PP 100, BS 100). The results from the concatenated analyses also uniquely recovered well-supported affiliations between the Neomyzomyia Series + subgenus *Stethomyia* + Arribalzagia Series (in part) (PP 100, BS 78) and between the Funestus Subgroup (Myzomyia Series) + Myzorhynchus Series (PP 100, BS 92). However, other specific affiliations could not be investigated within the concatenated analyses due to the limited availability of applicable DNA sequences. Nevertheless, this dataset included several unique sets of taxa, where the inclusion of multiple representative species recovered the monophyly of the Myzorhynchus Series' Hyrcanus group, as well as the Neomyzomyia Series' Punctulatus Group (PP 100, BS 97) and Farauti Complex (PP 100, BS 100).

Several affiliations within the current results were supported by the results of multiple analyses and DNA regions. This included the well-supported monophyly of the Funestus Subgroup (COI: PP 100, BS 97; ITS2: PP 100, BS 100) and Gambiae Complex (COI: PP 100, BS 98; ITS2: PP 100, BS 100), which was

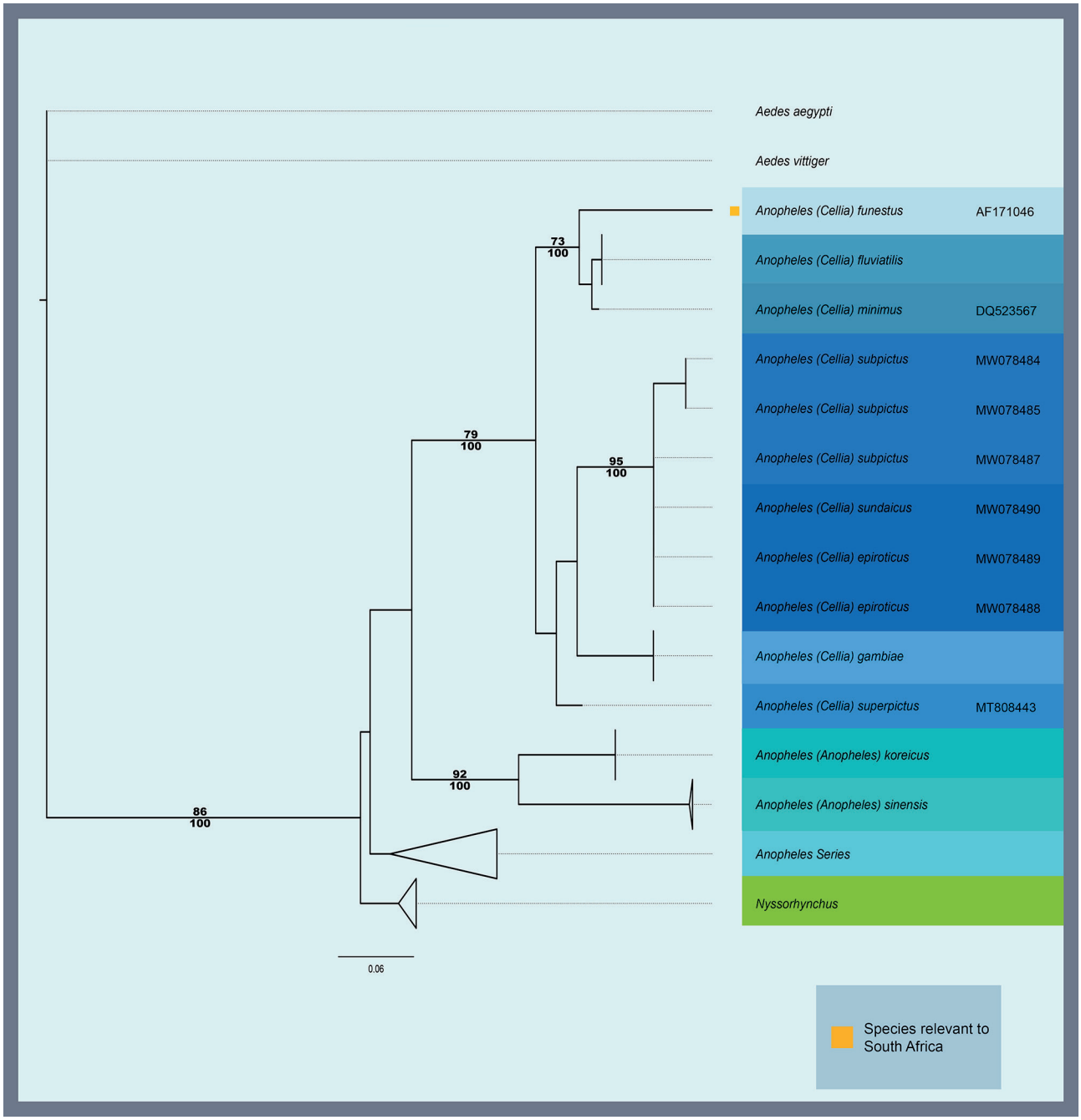

Figure 7: Phylogenetic relationships of *Anopheles* based on the maximum likelihood analysis of 28S sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades (cropped figure).

recovered by all relevant analyses. The analyses also recovered the monophyly of the Sundaicus Complex, which was generally well-supported (COI: PP 100, BS 100; ITS2: PP 100, BS 100; concatenated: BS 88). All analyses, except the poorly supported COI clades, recovered a close relationship between Pyretophorus' Sundaicus and Subpictus Complexes. These taxa formed a single well-supported clade in both the concatenated and 28S analyses (concatenated: PP 100, BS 93; 28S: PP 100, BS 95), and were affiliated with *An. vagus* from the same series in the ITS2 results (PP 100). Moreover, all analyses with a sufficient degree of overall clade support (concatenated, ITS2 and 28S) recovered a close association between the Pyretophorus and Neocellia Series, where these taxa (in part) shared a well-supported clade in the concatenated and ITS2 results (concatenated: PP 100, BS 79, ITS2: PP 100, BS 96). Here, the analyses yielded a relatively distant placement of the Gambiae Complex, as compared to other members of Pyretophorus. Multiple analyses also recovered a well-supported grouping for Myzomyia's Minimus Subgroup (in part) (A*n. minimus* Theobald + *An. leesoni* + *An. fluviatilis* James) (COI: PP 100, BS 98; ITS2: PP 97), while the subgroup as a whole was non-monophyletic due to the relatively distant placement of *An. flavirostris* (Ludlow).

Several taxa were recovered as non-monophyletic in numerous analyses, although they were not necessarily included in well-supported affiliations. The Pyretophorus Series was recovered as a polyphyletic assemblage for all datasets with a relatively larger taxonomic coverage (COI, ITS2 and concatenated analyses), where its members shared well-supported clades with either the Neocellia and Myzomyia (in part) Series (concatenated: PP 100, BS 100), the Paramyzomyia Series + Ardensis Group (Neomyzomyia Series) + Neocellia Series (in part) (ITS2: PP 100, BS 96) or the Myzomyia Series (ITS2: PP 100, BS 98). Neocellia was similarly recovered as a polyphyletic series wherever it was represented by multiple species (COI and ITS2), with one of its

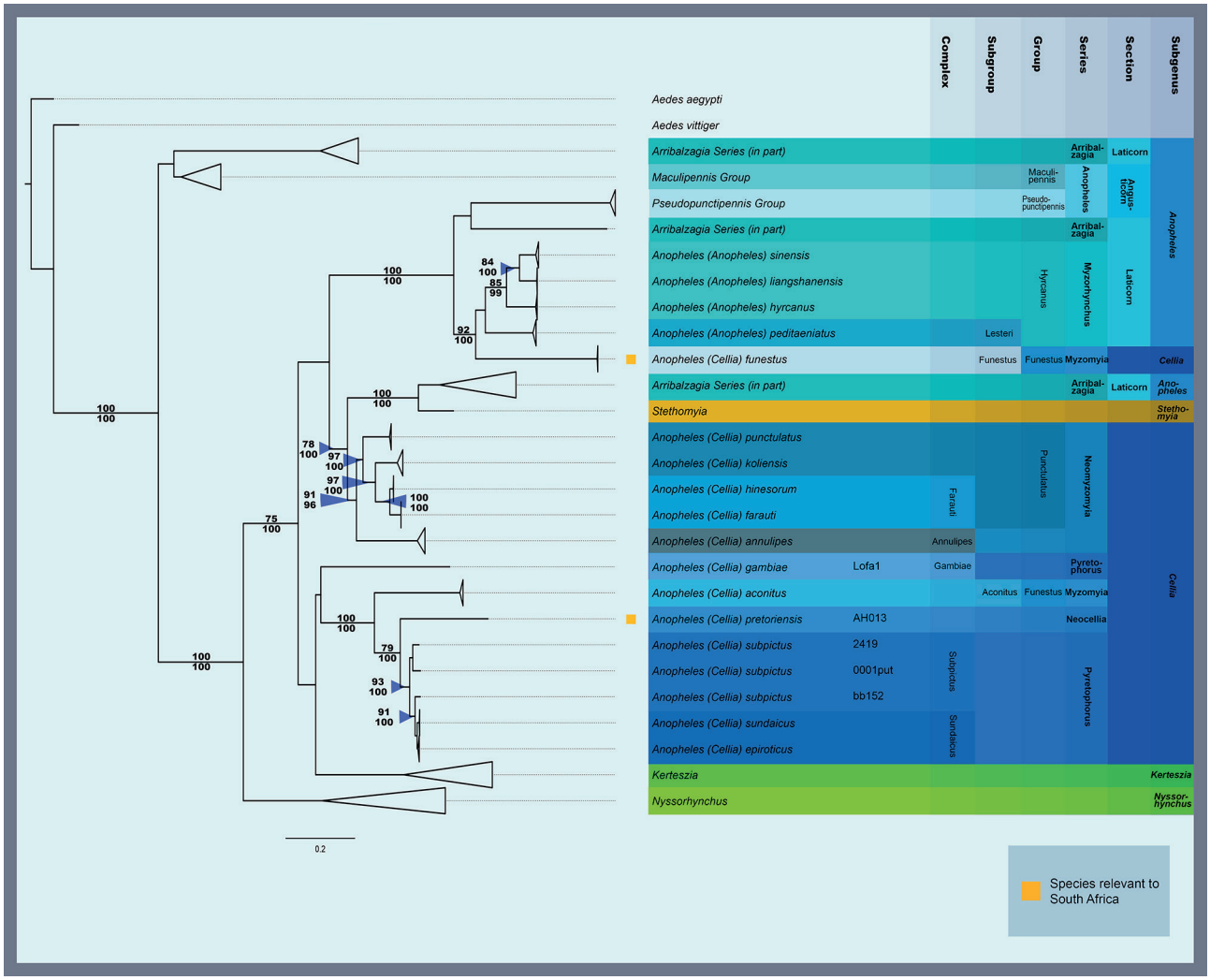

Figure 8: Phylogenetic relationships and taxonomic divisions of *Anopheles* based on the maximum likelihood concatenated analysis of COI, ITS2 and 28S sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades.

species (*An. stephensi*) being situated distantly from all other members of the series. Furthermore, Neocellia's members shared well-supported clades with members of the subgenus *Anopheles* (PP 100, BS 100) or the Pyretophorus (in part) + Paramyzomyia + Neomyzomyia (in part) Series (PP 100, BS 96) in the ITS2 results. Another non-monophyletic taxon was the Cellia Series, which was represented in the COI and ITS2 analyses. Here, it was either polyphyletic (COI analyses) or paraphyletic (ITS2 analyses) with the inclusion of the subgenus *Anopheles*' Barbirostris Subgroup (PP 100, BS 100), however, specific interactions were inconsistent. Myzomyia was also non-monophyletic in several analyses, either as a polyphyletic (COI and concatenated analyses) or paraphyletic assemblage (ITS2 analyses) with the inclusion of the Pyretophorus Series (in part) (PP 100, BS 98).

Furthermore, the Funestus Group (Myzomyia Series) was polyphyletic in all analyses based on relatively larger datasets (concatenated, COI and ITS2). Its Funestus Subgroup shared a well-supported clade with other members of the series in the ITS2 analyses, *An. dthali* + Demeilloni Group + Wellcomei Group + Marshallii Group (PP 99, BS 77), where the remainder of the Funestus Group was paraphyletic with the inclusion of members from the Pyretophorus Series. In the concatenated analyses, the Funestus Group's Aconitus Subgroup similarly formed a well-supported clade with members of the Pyretophorus (in part) and Neocellia Series (PP 100, BS 100). However, the placement of *An. funestus* was unique in the concatenated analyses, sharing a well-supported clade with the subgenus *Anopheles*' Myzorhynchus Series (PP 100, BS 92). These Funestus Group clades were also consistently interspersed with members from the polyphyletic Demeilloni Group in all relevant analyses (COI and ITS2), where *An. sergentii* shared a well-supported subclade with *An. dthali* + Funestus Subgroup (PP 100, BS 91), and *An. demeilloni* with the Wellcomei and Marshallii Groups (PP 100, BS 86) in the ITS2 results.

The various datasets included several representative species relevant to South African, a few of which belong to the Neocellia Series. This included *An. rufipes* (Gough), *An. maculipalpis* Giles and *An. pretoriensis* (Theobald). Although the specific affiliations of these species were inconsistent, they nonetheless grouped within the major cluster consisting of other Neocellia members, where the overarching clade was well-supported in the ITS2 results (PP 100, BS 99). The Neomyzomyia Series was also represented by a single South African species in the ITS2 analyses, *An. nili,* where it shared a well-supported clade with Neocellia (in part) (PP 97, BS 88), rather than with other members of its series.

The Cellia Series was likewise represented by a few South African species in the COI and ITS2 datasets, which consisted of *An. squamosus*, *An. pharoensis* Theobald and sampled individuals of *An.* cf. *cydippis* / *squamosus*. The COI results produced two monophyletic and well-supported sister lineages, with one consisting of locally sampled *An.* cf. *cydippis* / *squamosus* specimens (PP 99, BS 86) and another lineage consisting of non-

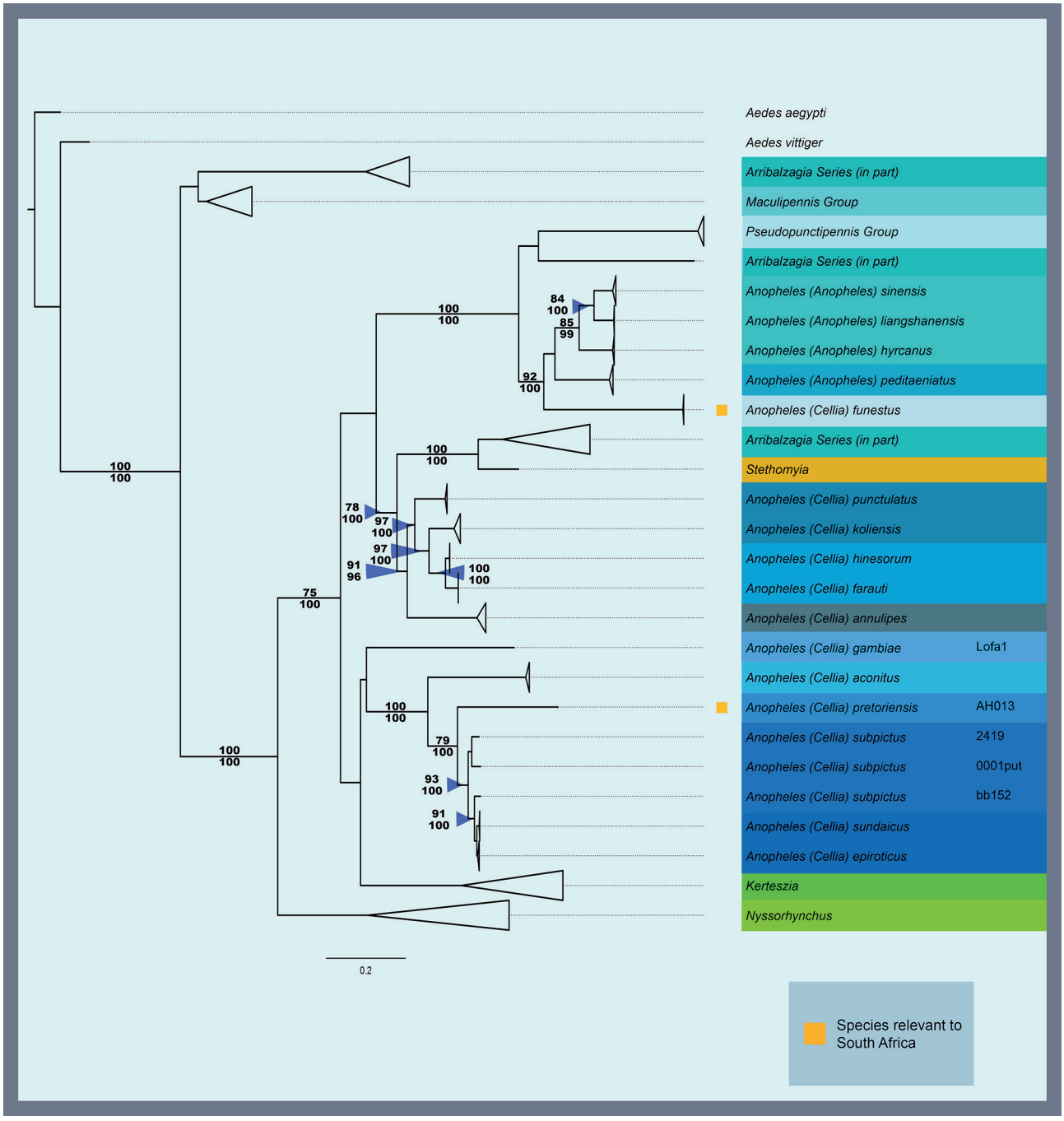

Figure 9: Phylogenetic relationships of *Anopheles* based on the maximum likelihood concatenated analysis of COI, ITS2 and 28S sequences under a GTR + I + Γ model of nucleotide substitution. Branch numbers represent significant support values (BS ≥ 70 or PP ≥ 95) for clades recovered in both the maximum likelihood (BS; upper value) and Bayesian (PP; lower value) analyses. Internal support values were not provided for monophyletic species-level clades (cropped figure).

sampled African *An. squamosus* specimens (PP 100, BS 88). The combined Squamosus Group clade of both sampled and non-sampled specimens was well-supported for both DNA regions (COI: PP 100, BS 98; ITS2: 100, BS 98). However, the placement of the final Cellia Series species, *An. pharoensis*, was inconsistent between the DNA regions.

The Gambiae Complex was represented by two species in the COI and ITS2 analyses, *An. gambiae* Giles and *An. merus* Dönitz, with the latter occurring in South Africa. In each case, *An. gambiae* was paraphyletic with the inclusion of *An. merus*, where the latter species formed a well-supported monophyletic grouping in the COI results (PP 100, BS 100). This Gambiae Complex clade's relationship with the remainder of its Pyretophorus Series was also inconsistent, where it was situated distantly from the rest of the series in the ITS2 and concatenated analyses, sharing a well-supported clade with the Myzomyia Series in the ITS2 results (PP 100, BS 98).

The Myzomyia Series' Funestus and Demeilloni Groups included six species relevant to South Africa, namely *An. rivulorum* Leeson, *An. leesoni*, *An. longipalpis*, *An. parensis*, *An. funestus* and *An. demeilloni*. *Anopheles dthali* and *An. demeilloni* were consistently situated within the vicinity of the Funestus Subgroup. Although *An. dthali* is not a member of either Demeilloni or Funestus Groups, it was consistently situated within the overarching Funestus Group clade. Furthermore, the South African representatives of the Wellcomei and Marshallii Groups (*An. marshallii* and *An. theileri*) were likewise included in the larger Funestus Group cluster in the ITS2 analyses, where these species shared a well-supported clade with *An. demeilloni* (Demeilloni Group) (PP 100, BS 86). *Anopheles rivulorum* (Rivulorum Subgroup) was similarly

associated with other members of its Funestus Group, although the specific associations were inconsistent between analyses.

All relevant analyses recovered a clade consisting of the Funestus Group's Minimus + Aconitus + Culicifacies Subgroups with the addition of Pyretophorus' *An. gambiae* in the ITS2 results. The datasets included *An. leesoni* as a South African member of the Minimus Subgroup, where the subgroup itself was non-monophyletic due to the placement of one of its more basal species, *An. flavirostris*. Nevertheless, the remaining species consistently formed a grouping consisting of the Minimus Subgroup's Minimus Complex + Fluviatilis Complex + *An. leesoni* (COI: PP 100, BS 98; ITS2: PP 97). This relationship between the Minimus and Fluviatilis Complexes was also recovered in the 28S results (BS 85).

Another set of affiliations remained consistent across all applicable analyses, which included the monophyly of the Funestus Subgroup [*An. longipalpis C* (Theobald) (Type C) (Koekemoer et al. 2009), *An. parensis* and *An. funestus*]. Two of the publicly available *An. longipalpis* sequences were not designated as either *An. longipalpis A* or *C*, belonging to the Minimus and Funestus Subgroups, respectively. Both specimens nevertheless clustered with *An. longipalpis C* within the Funestus Subgroup and were, therefore, likely conspecific specimens. The distinction between the various Funestus Subgroup species was generally poor due to the sporadic non-monophyly of *An. parensis* and *An. funestus*, and the consistent non-monophyly of *An. longipalpis*.

DISCUSSION

Phylogenetic overview

Several relationships were robustly recovered by multiple DNA regions and methods of analysis, which consisted of recurring affiliations and numerous non-monophyletic taxa. Several instances of non-monophyly were also associated with well-supported affiliations with other taxa. Consistent findings included the non-monophyly of the subgenus *Anopheles*, polyphyly of its Laticorn Section, and the polyphyly of the subgenus *Cellia*'s Funestus Group, which was associated with numerous other taxa. Furthermore, the current analyses recovered associations between the Coustani and Hyrcanus Groups, and between the Pyretophorus and Paramyzomyia Series. Finally, several analyses recovered instances of paraphyly within numerous *Anopheles* species relevant to South Africa. These results raised doubts about the validity of the subgenus *Anopheles'* numerous formal subdivisions.

Subgenus *Anopheles*

The subgenus *Anopheles* was non-monophyletic in all current analyses, which has previously also been recovered by multiple other authors. The collapsed tree from the morphological analyses of Harbach and Kitching (2016) recovered a polyphyletic subgenus *Anopheles* which incorporated a clade of the subgenera *Stethomyia* + *Lophopodomyia* + *Bamaia* + the anopheline genus *Bironella*. However, these affiliations with the subgenus *Anopheles* were not well-supported. The Bayesian COI sequence analysis of Wang et al. (2017) similarly recovered scattered subgenus *Anopheles* clades affiliated with several other subgenera. Both analyses of Harbach and Kitching (2005) also recovered a non-monophyletic subgenus *Anopheles* with a close relationship to subgenus *Lophopodomyia*. Furthermore, the equal weight analysis of Harbach and Kitching (2005) recovered affiliations with *Bamaia*, *Bironella* and *Stethomyia*, as defined by the current systematic framework. Although representatives of the subgenus *Stethomyia* were unavailable in the current ITS2 and 28S analyses, this affiliation with the subgenus *Anopheles* was still recovered in the COI and concatenated analyses. In both the current COI ML analysis and concatenated analyses, *Stethomyia* shared clades with the subgenus *Anopheles*, which included a well-supported relationship between *Stethomyia* and *Anopheles'* Arribalzagia Series (in part) in the concatenated results (PP 100, BS 100).

On the other hand, several authors recovered *Anopheles* as a monophyletic subgenus. This included the analyses of Gholizadeh et al. (2013) based on ITS2 fragments, where the subgenus was represented by ten morphospecies. The molecular analysis of Foster et al. (2017), based on slow-evolving mitochondrial protein sequences, also recovered a monophyletic subgenus *Anopheles* with a good degree of support. However, this analysis only included nine morphospecies. Lastly, the combined rDNA and mtDNA data analyses with ML and maximum parsimony of Sallum et al. (2002) recovered a monophyletic *Anopheles* subgenus with a significant degree of support, where it was also represented by nine species. These monophyletic findings were consistently recovered by analyses conducted with a relatively small number of representatives. Harbach and Kitching (2016) observed a similar trend and stated that the polyphyly of the subgenus *Anopheles* would likely be recovered with sufficient taxonomic coverage.

Within the subgenus *Anopheles*, the current study's results mainly produced a polyphyletic Laticorn Section intermixed with the Angusticorn Section. Several other studies similarly recovered Laticorn as a non-monophyletic assemblage. In the morphological analyses of Harbach and Kitching (2016), the Laticorn Section was associated with two species from Angusticorn's Cycloleppteron Series. This was similar to the equal weight analysis of Harbach and Kitching (2005), where a Cycloleppteron species was included within the paraphyletic Laticorn Section, while Laticorn was monophyletic in the implied weight analysis. The combined mtDNA and rDNA phylogenetic analyses of Sallum et al. (2002) also recovered three Laticorn species interspersed with Angusticorn clades. Finally, similar findings were observed in the successive weighting morphological analysis of Collucci and Sallum (2007) and the slow-evolving mitochondrial protein sequence analysis of Foster et al. (2017), where Laticorn was paraphyletic with the inclusion of Angusticorn.

All current analyses, which included representatives of the Coustani and Hyrcanus Groups, recovered the well-supported sister relationship between these taxa. This close affiliation was similarly recovered by numerous other authors. Collucci and Sallum (2007) performed phylogenetic analyses based on the morphological characters of anophelines, and their successive weighting (BS 79; Bremer support 1) and implied weights analyses produced a well-supported sister relationship between the Hyrcanus and Coustani Groups. Both the equal and implied weight analysis of Harbach and Kitching (2005) based on morphological data once again recovered a sister relationship between the Coustani and Hyrcanus Groups (Bremer support > 0.8), as did the analyses of Harbach and Kitching (2016). The cladistic analyses of Sallum et al. (2000) likewise recovered this shared relationship in their weighted and unweighted analyses (BS 75, Bremer support 4). This clade was characterised by their shared lateral scales on the clypeus (Collucci and Sallum 2007, Sallum et al. 2000).

Paraphyly in non-sampled South African species of the subgenus Anopheles (An. coustani and An. tenebrosus)

The COI analyses included two representative species of the Coustani Group, with both species occurring within South Africa. Here, *An. coustani* was paraphyletic with the inclusion of *An. tenebrosus*. Similar phylogenetic findings were also recovered by Ciubotariu et al. (2020), where their COI ML analysis nested *An. tenebrosus* within the *An. coustani* clade. Gillies and De Meillon (1968) noted that *An. tenebrosus* and *An. coustani* are morphologically distinct despite their sympatric occurrence, which serves as evidence for their genetic isolation. Despite the

morphological variation of *An. tenebrosus*, the two species can still be distinguished by their leg markings, where the coxae and base of the first hind tarsus of *An. tenebrosus* lack pale scales, and where the forelegs of this species are dark on the apex of the tibia and base of the first tarsus (Gillies and De Meillon 1968). However, since the phylogeny of these species has not yet been examined extensively, additional investigations are needed to clarify the relationship between *An. coustani* and *An. tenebrosus*.

Subgenus *Cellia*

Cellia was represented by numerous species in the current analyses, where the subgenus was largely recovered as a polyphyletic assemblage. However, the subgenus was represented by a well-supported monophyletic clade in the 28S analyses, consisting of a relatively small number of species. The findings of other authors produced conflicting results, where several studies supported *Cellia*'s monophyly. This monophyly was well-supported in the morphological analysis of Harbach and Kitching (2016) and was also recovered in the morphological equal weight and implied weight analyses of Harbach and Kitching (2005). The combined rDNA data analyses of Sallum et al. (2002) and the mitochondrial protein sequence analysis of Foster et al. (2017) produced monophyletic *Cellia* clades. However, the analyses likewise consisted of a small number of species.

On the other hand, at least two studies recovered the non-monophyly of *Cellia*. Within the ITS2 fragment analyses of Gholizadeh et al. (2013), the neighbour-joining tree produced a paraphyletic *Cellia* clade, while the taxon was monophyletic in the ML analysis. The Bayesian COI sequence analysis of Wang et al. (2017) similarly recovered scattered *Cellia* clades, which were weakly associated with several other subgenera. Therefore, the monophyly of *Cellia* seemed to be inconsistent or the product of insufficient taxon sampling. Harbach and Kitching (2016) noted that the internal taxonomic structure of *Cellia* does not reflect its true evolutionary history, which was supported by the recovery of numerous non-monophyletic series in their results.

Multiple analyses in the current study produced non-monophyletic Myzomyia Series groups. Analyses produced polyphyletic and intermixed Funestus and Demeilloni Group clades, where the Marshallii and Wellcomei Groups were incorporated within the overarching Funestus Group clade in the ITS2 results. Other authors similarly recovered various aspects of these affiliations. The morphological analysis of Harbach and Kitching (2016) recovered a polytomous clade within their collapsed tree that included the Demeilloni, Funestus, Marshallii and Wellcomei Groups, amongst several other taxa. The Funestus Group of Norris and Norris (2015) was similarly non-monophyletic, where the COI and ITS2 analyses placed *An. theileri* (Wellcomei Group) within the Funestus clade. The Funestus Group was also polyphyletic in the Bayesian COI analysis of Wang et al. (2017), which may have been the product of the weak basal phylogenetic support. However, in the morphological dataset of Harbach and Kitching (2016), many characters were shared between the members of the Funestus, Demeilloni, Wellcomei and Marshallii Groups. The most exclusive of these characters included the presence of the females' cibarial armature rods, the location of the premental apodeme (removed from the lateral margin) and the structure of the spiracular apparatus in larvae, with the median plate possessing lateral arms. Within the Afrotropical Region, the Funestus Group is widespread and shares a large portion of its range with the Marshallii and Wellcomei Groups, while a smaller portion of this range overlaps with the Demeilloni Group (Gillies and De Meillon 1968). Therefore, considering the shared molecular and morphological features, it is likely that these taxa share a degree of common ancestry.

Anopheles dthali, another member of the Myzomyia Series, was also commonly associated with the Funestus and Demeilloni Groups in the current analyses. This affiliation was partly supported by the ITS2 fragment analyses of Gholizadeh et al. (2013), which produced a close and well-supported relationship between the Demeilloni Group's *An. sergentii* (Demeilloni Group) and *An. dthali*. This affiliation between these species was similarly recovered in the maximum parsimony analysis based on ITS2 DNA sequences of Karimian et al. (2014).

Another frequently recovered Funestus Group clade in the current analyses consisted of the Minimus + Aconitus + Culicifacies Subgroups. This close relationship between the Minimus and Aconitus Subgroups was also recovered in the Bayesian COI sequence analysis of Wang et al. (2017), while multiple analyses supported their affiliation with the Culicifacies Subgroup. The neighbour-joining D3 sequence analysis of Swain et al. (2010) produced a close relationship between the Aconitus and Minimus subgroups, which was sister to a clade of the Funestus Group's *An. jeyporiensis* James + Culicifacies Subgroup. The ML analysis of Yan et al. (2019), based on multiple protein-coding gene sequences, similarly recovered a close relationship between the three subgroups. However, in the studies of Swain et al. (2010) and Yan et al. (2019), these taxa were the only representatives of the Myzomyia Series and the affiliations between these subgroups were inevitable. Even so, the addition of the Rivulorum and Funestus Subgroups in the analyses of Garros et al. (2005) still produced a single clade consisting of the Minimus, Culicifacies and Aconitus Subgroups, where the latter two subgroups were closely related. These subgroups also share numerous morphological features, where the key of Rattanarithikul et al. (2006) listed their markings, upper proepisternal setae and the lack of certain scales as common features between the three subgroups. These subgroups share an Oriental distribution, while the Culicifacies and Minimus subgroups also occur in the Afrotropical Region (Edwards 1932; Coetzee 2020). Therefore, several findings support these taxa's likely shared affiliation.

Furthermore, the generally well-supported affiliations between the Minimus Subgroup's *An. leesoni* and the Minimus and Fluviatilis Complexes were recovered in all relevant analyses. Several other studies recovered comparable results, including the neighbour-joining 28S (D3) sequence analysis of Swain et al. (2010), which produced a Minimus Subgroup clade consisting of the Minimus and Fluviatilis Complex species. Similar results were recovered by the D3 nucleotide sequence and neighbour-joining COII amino acid analyses of Garros et al. (2005), which produced a Minimus Complex + Fluviatilis Complex + *An. leesoni* clade. Finally, the COI and ITS2 sequence analyses of Norris and Norris (2015) also recovered a close relationship between the Minimus Complex and *An. leesoni*. The affiliations between *An. leesoni* and the Minimus and Fluviatilis Complexes were also supported in the morphological dataset of Garros et al. (2005), where these taxa shared three pupal features. This included the presence of three or more 2-Pa setae and the unique branching structure of seta 1-III and 5-III. Therefore, numerous findings support the grouping of *An. leesoni*, the Minimus Complex and the Fluviatilis Complex, which likely reflect a monophyletic subset of the Minimus Subgroup.

The current analyses recovered numerous affiliations of the Pyretophorus Series and its members. Most current analyses recovered the Pyretophorus Series as a polyphyletic assemblage, where it was commonly associated with the Neocellia Series. However, the structure of Pyretophorus differed in the findings of several authors. The monophyly of this series was often associated with less extensive taxonomic coverage, as recovered in the COI and ITS2 sequence analyses of Norris and Norris (2015). The combined mtDNA and rDNA data phylogenetic analyses of Sallum et al. (2002) included only four Pyretophorus species, where the combined mtDNA and one combined rDNA analyses produced a monophyletic series. However, the maximum parsimony

analysis of the combined rDNA data recovered Pyretophorus as a paraphyletic clade that included the Myzomyia Series. The affiliations of Pyretophorus were unresolved in the collapsed cladogram of Harbach and Kitching (2016), since it shared a polytomy with several other *Cellia* series. However, the authors still expressed doubts over the monophyly of Pyretophorus. Therefore, additional investigations are needed to provide further insights into the structure of this series.

The findings of several authors also produced inconclusive results regarding the relationship between Pyretophorus and Neocellia. The relatively close relationship between Pyretophorus and Neocellia was recovered in the ML analyses of Bargues et al. (2006) based on 18S rDNA sequences, where both series were situated as the two most basal clades within the subgenus *Cellia*. The equal weighted and unweighted morphological analyses of Sallum et al. (2000) similarly produced trees where Pyretophorus shared a relatively basal placement within the subgenus *Cellia* clade with Neocellia's *An. superpictus* Grassi, however, another Neocellia species (*An. annularis* van der Wulp) was more closely related to the Cellia Series. Sallum et al.'s (2002) ML and parsimony analyses based on combined 18S and 28S rDNA favoured a close relationship between Pyretophorus and Myzomyia rather than with Neocellia, while the COI sequence analyses (neighbour-joining and ML) of Norris and Norris (2015) also produced no close relationship between Neocellia and Pyretophorus. However, it is worth noting that a portion of Norris and Norris's (2015) phylogeny was weakly supported, which may have obscured any underlying affiliations.

On the other hand, other aspects of Pyretophorus' affiliations were indeed supported by the findings of other authors. This included its affiliation with the Paramyzomyia Series, which was well-supported in the current ITS2 analyses. The equally weighted and unweighted analysis of Sallum et al. (2000) based on morphological data placed both Paramyzomyia and Pyretophorus medially within subgenus *Cellia*. The implied weight analyses of Harbach and Kitching (2005) also produced a clade consisting of a paraphyletic Paramyzomyia with the inclusion of Pyretophorus, where the clade was defined by adult forecoxa presenting with anterior scales, wing vein R_{2+3} with linear dorsal scales, the presence of female cibarial armature with rods and cones, a male tergum IX that is sclerotised as a single sclerite, larval 4-C seta that is relatively strongly developed and larval 9-VII seta that is spine like and at least or greater to half the length of the segment among other characteristics.

Polyphyletic non-sampled South African Cellia species (An. longipalpis C, An. funestus and An. parensis)

The Myzomyia Series' Funestus Subgroup was monophyletic and well-supported in all relevant current analyses, while several analyses recovered the sporadic non-monophyly of two of its members (*An. parensis* and *An. funestus*), as well as the consistent non-monophyly of another member (*An. longipalpis*). However, similar investigations of such non-monophyletic groupings are hard to come by, since phylogenetic studies often include single representatives of each species. These studies nonetheless supported the monophyly of the subgroup, including the Bayesian COI sequence results of Wang et al. (2017) and the 28S (D3) nucleotide and COII amino acid sequence results of Garros et al. (2005). The nucleotide sequence analyses of Norris and Norris (2015) also investigated the interspecific affiliations of the well-supported monophyletic Funestus Subgroup. Their COI neighbour-joining results produced a weakly supported relationship between *An. funestus* and *An. longipalpis*, while the ITS2 maximum parsimony and neighbour-joining results produced a well-supported clade consisting of *An. longipalpis* and *An. parensis*. These species were also included in the phylogenetic analyses of Koekemoer et al. (2009), which were based on ITS2 sequence data. The authors recovered two distinct ITS2 amplicons for *An. longipalpis C*, which consisted of a larger and smaller fragment. These fragments had unique affiliations, where the short *An. longipalpis C* fragment was closely related to *An. parensis* with a high degree of support, while the other fragment was associated with *An. vaneedeni*. However, the third species, *An. funestus* was placed basally to both these groupings. These three taxa would likely benefit from more robust investigations into their affiliations and structure, with the incorporation of additional gene regions and a greater degree of representation within each of these species. This was, however, not possible within the current analyses, due to the limited availability of DNA sequences for the various target regions.

Paraphyly in non-sampled South African Cellia species (An. merus)

The Gambiae Complex of the Pyretophorus Series was represented by publicly available sequences from two different species in the current analyses. The complex as a whole was monophyletic and well-supported in all relevant analyses. However, the paraphyly of one of its members, *An. gambiae*, was recovered with the inclusion of *An. merus*. Previous authors recognised the limits of molecular data when investigating the phylogeny of the Gambiae Complex, where phylogenetic inferences were impeded by introgression, incomplete lineage sorting, incongruent results and the recent diversification of the Gambiae Complex (Besansky et al. 2003). Therefore, authors have instead focused on chromosomal inversions for phylogenetic inferences, which has recovered the close relationship between *An. gambiae* and *An. merus* (Besansky et al. 1994; Kamali et al. 2012). Considering the limits of molecular data, and the limited availability of sequences across the various target regions, no definitive conclusions could be made regarding the relationships of these species.

Affiliations of sampled Cellia species (An. cf. cydippis / squamosus)

All relevant analyses in the current study grouped the sampled *An.* cf. *cydippis / squamosus* specimens with the publicly available *An. squamosus* sequences. This created well-supported monophyletic subclades with distinct sampled and non-sampled lineages in the COI results. However, the specific identity of the sampled specimens was uncertain, since no *An. cydippis* DNA sequences were available for comparison. *Anopheles cydippis* has not previously been documented in the Free State Province (Gillies and De Meillon 1968). Yet, the DNA sequences of Free State specimens were markedly different from other African *An. squamosus* specimens.

This lack of relevant matching sequences within the database highlights the importance of a global effort to continually expand such datasets, to aid in the accuracy and accessibility of molecular data for a wide array of species. However, the lack of molecular data on South African culicids is exemplified by the relative shortage of records on repositories such as Barcode Of Life Data System v4 (Ratnasingham and Hebert 2007) and GenBank (Benson et al. 2012). Since misidentifications of biological specimens are a relatively common occurrence (Bridge et al. 2003; Nilsson et al. 2006; Valkiūnas et al. 2008) and the interpretation of DNA barcoding results in isolation may potentially yield ambiguous results (Meier et al. 2006), the ability to corroborate species identities through both approaches can serve as an invaluable tool to ensure the accuracy and validity of research efforts. Therefore, the sequences of the sampled individuals have been uploaded to a public DNA database as part of a larger effort to sample and sequence South African mosquitoes (BOLD: MPSAM015-21, MPSAM022-21, MPSAM061-21).

CONCLUSIONS

The current study provided a novel investigation into the placement of South African *Anopheles* species and higher taxa within the extensive taxonomic structure of the genus. The larger taxonomic divisions of the genus *Anopheles* were often recovered as non-monophyletic, especially when a relatively high degree of taxonomic coverage was achieved. These results were also frequently supported by the findings of morphological and other molecular studies. This may indicate the need to either subdivide constituent clades of non-monophyletic groupings into separate taxa or to incorporate the affiliated clades within its parent taxon.

Firstly, the subgenus *Anopheles* was non-monophyletic in all current analyses, which contributed to the polyphyly of its Laticorn Section. The non-monophyly of both taxa was also recovered by numerous other authors. Since the current structure of the subgenus *Anopheles* does not seem to reflect the evolutionary history of its taxa, the subgenus may benefit from a comprehensive re-evaluation. Secondly, numerous aspects of the subgenus *Cellia*'s structure were recovered within current results and supported by the findings of other authors. This included the polyphyly of the Funestus Group, where it was associated with numerous other taxa in the current analyses. This group likely does not reflect the natural affiliations of its members, and focused investigations may reveal the constituent monophyletic subdivisions of the Funestus Group.

Other findings indicated the potential non-monophyly of several taxa, which would benefit from further investigations. Here, the current results recovered the non-monophyletic affiliations of several South African species, including the Coustani Group's *An. coustani* and *An. tenebrosus* and the Funestus Subgroup's *An. parensis, An. funestus* and *An. longipalpis C*. Furthermore, the inclusion of Paramyzomyia in the current ITS2 analyses revealed a close and well-supported relationship with Pyretophorus, where affinities between these taxa were also previously recovered by other authors. Finally, the structure and additional affiliations of Pyretophorus were largely inconsistent both within the current analyses and within the available literature, as was the monophyly of the subgenus *Cellia*. The elucidation of the true structure of these taxa would likely benefit from a combined approach incorporating morphological and molecular data from geographically diverse populations, while ensuring extensive taxonomic representation within such analyses.

In conclusion, the current study highlights and supports several findings and research gaps within the phylogenetic literature, revealing numerous challenges within the current systematic framework of the genus *Anopheles*, especially with respect to taxa relevant to South Africa. The study provided a novel large-scale examination into the placement of numerous South African species within the overall phylogeny of the genus. The current study also expanded the available molecular and distribution data for South African *Anopheles* specimens. These investigations help to broaden the evolutionary perspective of South African mosquitoes, contributing to the foundation of available data for further epidemiological, biogeographical and evolutionary investigations.

DATA AVAILABILITY STATEMENT

Sequences generated during the current study are available on BOLD Systems [Project: MPSAM Molecular phylogeny of South African *Anopheles*, *Aedes* and *Culex* (Diptera: Culicidae) based on COI, ITS2 and 28S DNA sequences].

ACKNOWLEDGEMENTS

The authors would like to thank all those who provided guidance and assistance with the current project, including Ruan Booysen who assisted during sampling efforts, Alan Kemp who provided guidance on the practical aspects of culicid research and Thabang Madisha who provided guidance with the molecular aspects of this project. The authors are also very grateful for the partial funding provided by the NRF (NRF reference no: SFH170609238850), the DST and NRF South African Research Chairs Initiative in Vector-borne and Zoonotic Pathogens Research (grant no: U98346) and the Grassland Biodiversity Project from the Department of Zoology and Entomology, UFS. The authors would also like to thank the anonymous reviewers for their valuable contributions to the refinement of this paper.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS APPROVAL STATEMENT

This study was approved by the Biosafety and Environmental Research Ethics Committee of the University of the Free State.

AUTHOR CONTRIBUTIONS

Liezl Whitehead: conceptualisation; data curation; formal analysis; investigation; methodology; project administration; visualisation; writing – original draft. Vaughn R. Swart: conceptualisation; funding acquisition; methodology; project administration; resources; supervision; writing – review and editing. Marieka Gryzenhout: conceptualisation; data curation; methodology; resources; supervision; writing – review and editing. Lizette L. Koekemoer: supervision; writing – review and editing.

ORCID IDs

Liezl Whitehead: https://orcid.org/0000-0001-7839-5346
Vaughn R. Swart: https://orcid.org/0000-0001-7905-5298
Marieka Gryzenhout: https://orcid.org/0000-0002-9224-4277
Lizette L. Koekemoer: https://orcid.org/0000-0003-4236-6345

REFERENCES

Baker RH, DeSalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Systematic Biology. 46(4):654–673. https://doi.org/10.1093/sysbio/46.4.654.

Bargues MD, Latorre JM, Morchon R, Simon F, Escosa R, Aranda C, Sainz S, Fuentes MV, Mas-Coma S. 2006. rDNA sequences of *Anopheles* species from the Iberian Peninsula and an evaluation of the 18S rRNA gene as phylogenetic marker in Anophelinae. Journal of Medical Entomology. 43(3):508–517. https://doi.org/10.1093/jmedent/43.3.508.

Beebe NW. 2018. DNA barcoding mosquitoes: advice for potential prospectors. Parasitology. 145(5):622–633. https://doi.org/10.1017/S0031182018000343.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. Nucleic Acids Research. 41(D1):D36–D42. https://doi.org/10.1093/nar/gks1195.

Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, Fontenille D, Touré YT, Sagnon F. 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. Proceedings of the National Academy of Sciences. 100:10818–10823. https://doi.org/10.1073/pnas.1434337100.

Besansky NJ, Powell JR, Caccone A, Hamm DM, Scott JA, Collins FH. 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proceedings of the National Academy of Sciences. 91:6885–6888. https://doi.org/10.1073/pnas.91.15.6885.

Bofkin L, Goldman N. 2007. Variation in evolutionary processes at different codon positions. Molecular Biology and Evolution. 24:513–521. https://doi.org/10.1093/molbev/msl178.

Bridge PD, Roberts PJ, Spooner BM, Panchal G. 2003. On the unreliability of published DNA sequences. New Phytologist. 160:43–48. https://doi.org/10.1046/j.1469-8137.2003.00861.x.

Ciubotariu II, Jones CM, Kobayashi T, Bobanga T, Muleba M, Pringle JC, Stevenson JC, Carpi G, Norris DE. 2020. Genetic diversity of *Anopheles coustani* (Diptera: Culicidae) in malaria transmission

foci in southern and central Africa. Journal of Medical Entomology. 57:1782–1792. https://doi.org/10.1093/jme/tjaa132.
Coetzee M. 2020. Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae). Malaria Journal. 19:70. https://doi.org/10.1186/s12936-020-3144-9.
Coleman AW. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. TRENDS in Genetics. 19:370–375. https://doi.org/10.1016/S0168-9525(03)00118-5.
Collucci E, Sallum MAM. 2007. Cladistic analysis of the subgenus *Anopheles* (*Anopheles*) Meigen (Diptera: Culicidae) based on morphological characters. Memórias do Instituto Oswaldo Cruz. 102:277–292. https://doi.org/10.1590/s0074-02762007005000026.
Dassanayake RS, Gunawardene YINS, De Silva BDDNK. 2008. ITS-2 secondary structures and phylogeny of *Anopheles culicifacies* species. Bioinformation. 2:456–460. https://doi.org/10.6026/97320630002456.
Devulder G, de Montclos MP, Flandrois JP. 2005. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. International Journal of Systematic and Evolutionary Microbiology. 55:293–302. https://doi.org/10.1099/ijs.0.63222-0.
Djadid ND, Gholizadeh S, Tafsiri E, Romi R, Gordeev M, Zakeri S. 2007. Molecular identification of Palearctic members of *Anopheles maculipennis* in northern Iran. Malaria Journal. 6:6. https://doi.org/10.1186/1475-2875-6-6.
Edwards FW. 1932. Fascicle 194. Diptera. Fam. Culicidae. In: Wytsman P (editor), Genera Insectorum. Brussels: Desmet-Verteneuil. pp 1–258.
Edwards FW. 1941. Mosquitoes of the Ethiopian region. III.-Culicine adults and pupae. London: British Museum (Natural History). https://doi.org/10.1126/science.96.2502.537.
Egan MG. 2006. Support versus corroboration. Journal of Biomedical Informatics. 39:72–85. https://doi.org/10.1016/j.jbi.2005.11.007.
Fang Y, Shi WQ, Zhang Y. 2017. Molecular phylogeny of *Anopheles* hyrcanus group members based on ITS2 rDNA. Parasites & Vectors. 10:417. https://doi.org/10.1186/s13071-017-2351-x.
Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 3:294–299.
Foster PG, de Oliveira TMP, Bergo ES, Conn JE, Sant'Ana DC, Nagaki SS, Nihei S, Lamas CE, González C, Moreira CC, Sallum MAM. 2017. Phylogeny of Anophelinae using mitochondrial protein coding genes. Royal Society Open Science. 4:170758. https://doi.org/10.1098/rsos.170758.
Friedrich M, Tautz D. 1997. Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. Systematic Biology. 46:674–698. https://doi.org/10.2307/2413500.
Garros C, Harbach RE, Manguin S. 2005. Morphological assessment and molecular phylogenetics of the Funestus and Minimus groups of *Anopheles* (*Cellia*). Journal of Medical Entomology. 42:522–536. https://doi.org/10.1093/jmedent/42.4.522.
Gholizadeh S, Djadid ND, Nouroozi B, Bekmohammadi M. 2013. Molecular phylogenetic analysis of *Anopheles* and *Cellia* subgenus anophelines (Diptera: Culicidae) in temperate and tropical regions of Iran. Acta Tropica. 126:63–74. https://doi.org/10.1016/j.actatropica.2012.11.003.
Gillies MT, Coetzee M. 1987. A supplement to the Anophelinae of Africa south of the Sahara. Publications of the South African Institute for Medical Research. 55:1–143.
Gillies MT, De Meillon B. 1968. The Anophelinae of Africa south of the Sahara (Ethiopian zoogeographical region). Publications of the South African Institute for Medical Research. 54:1–343.
Greni SE, Demari-Silva B, de Oliveira TMP, Suesdek L, Laporta GZ, Sallum MAM. 2018. A multi-gene analysis and potential spatial distribution of species of the Strodei subgroup of the genus *Nyssorhynchus* (Diptera: Culicidae). Journal of Medical Entomology. 55:1486–1495. https://doi.org/10.1093/jme/tjy137.
Harbach RE. 2007. The Culicidae (Diptera): a review of taxonomy, classification and phylogeny. Zootaxa. 1668:591–638. https://doi.org/10.11646/zootaxa.1668.1.28.
Harbach, RE. 2023. *Anopheles* classification. Mosquito Taxonomic Inventory. Accessed on 9 March 2025. Retrieved from https://mosquito-taxonomic-inventory.myspecies.info/sites/mosquito-taxonomic-inventory.info/files/Anopheles%20classification_67.pdf.
Harbach RE, Kitching IJ. 2005. Reconsideration of anopheline mosquito phylogeny (Diptera: Culicidae: Anophelinae) based on morphological data. Systematics and Biodiversity. 3:345–374. https://doi.org/10.1017/s147720000500174x.
Harbach RE, Kitching IJ. 2016. The phylogeny of Anophelinae revisited: inferences about the origin and classification of *Anopheles* (Diptera: Culicidae). Zoologica Scripta. 45:34–47. https://doi.org/10.1111/zsc.12137.
Harbach RE, Kitching IJ, Culverwell CL, Dubois J, Linton YM. 2012. Phylogeny of mosquitoes of tribe Culicini (Diptera: Culicidae) based on morphological diversity. Zoologica Scripta. 41:499–514. https://doi.org/10.1111/j.1463-6409.2012.00546.x.
Harris JK, Kelley ST, Pace NR. 2004. New perspective on uncultured bacterial phylogenetic division OP11. Applied and Environmental Microbiology. 70:845–849. https://doi.org/10.1128/aem.70.2.845-849.2004.
Hua YQ, Ding YR, Yan ZT, Si FL, Luo QC, Chen B. 2016. The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. Insect Science. 23:353–365. https://doi.org/10.1111/1744-7917.12326.
Huelsenbeck JP, Rannala B, Larget B. 2000. A Bayesian framework for the analysis of cospeciation. Evolution. 54:352–364. https://doi.org/10.1111/j.0014-3820.2000.tb00039.x.
Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755. https://doi.org/10.1093/bioinformatics/17.8.754.
Hurvich CM, Tsai CL. 1989. Regression and time series model selection in small samples. Biometrika. 76:297–307. https://doi.org/10.1093/biomet/76.2.297.
Jiang ZF, Huang DW, Zhu CD, Zhen WQ. 2006. New insights into the phylogeny of fig pollinators using Bayesian analyses. Molecular Phylogenetics and Evolution. 38:306–315. https://doi.org/10.1016/j.ympev.2005.11.008.
Kamali M, Xia A, Tu Z, Sharakhov IV. 2012. A new chromosomal phylogeny supports the repeated origin of vectorial capacity in malaria mosquitoes of the *Anopheles gambiae* complex. PLoS Pathogens. 8:e1002960. https://doi.org/10.1371/journal.ppat.1002960.
Karimian F, Oshaghi MA, Sedaghat MM, Waterhouse RM, Vatandoost H, Hanafi-Bojd AA, Ravasan NM, Chavshin AR. 2014. Phylogenetic analysis of the oriental-Palearctic-Afrotropical members of *Anopheles* (Culicidae: Diptera) based on nuclear rDNA and mitochondrial DNA characteristics. Japanese Journal of Infectious Diseases. 67:361–367. https://doi.org/10.7883/yoken.67.361.
Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics. 20:1160–1166. https://doi.org/10.1093/bib/bbx108.
Koekemoer LL, Lochouarn L, Hunt RH, Coetzee M. 1999. Single-strand conformation polymorphism analysis for identification of four members of the *Anopheles funestus* (Diptera: Culicidae) group. Journal of Medical Entomology. 36:125–130. https://doi.org/10.1093/jmedent/36.2.125.
Koekemoer LL, Kamau L, Hunt RH, Coetzee M. 2002. A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. The American Journal of Tropical Medicine and Hygiene. 66:804–811. https://doi.org/10.4269/ajtmh.2002.66.804.
Koekemoer LL, Misiani EA, Hunt RH, Kent RJ, Norris DE, Coetzee M. 2009. Cryptic species within *Anopheles longipalpis* from southern Africa and phylogenetic comparison with members of the *An. funestus* group. Bulletin of Entomological Research. 99:41–49. https://doi.org/10.1017/s0007485308006123.
Leaché AD, Reeder TW. 2002. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Systematic Biology. 51:44–68. https://doi.org/10.1080/106351502753475871.
Marinucci M, Romi R, Mancini P, Luca MD, Severini C. 1999. Phylogenetic relationships of seven palearctic members of the *maculipennis* complex inferred from ITS2 sequence analysis. Insect Molecular Biology. 8:469–480. https://doi.org/10.1046/j.1365-2583.1999.00140.x.
Meier R, Shiyang K, Vaidya G, Ng PKL. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology. 55:715–728. https://doi.org/10.1080/10635150600969864.

Meigen JW. 1818. Systematische Beschreibung der bekannten europäischen zweiflügeligen Insekten. Volume 1. Beaufort Sohn, Aachen.

Merget B, Koetschan C, Hackl T, Förster F, Dandekar T, Müller T, Schultz J, Wolf M. 2012. The ITS2 database. JoVE (Journal of Visualized Experiments) 61:e3806. https://doi.org/10.3791/3806.

Miller RE, McDonald JA, Manos PS. 2004. Systematics of *Ipomoea* subgenus *Quamoclit* (Convolvulaceae) based on ITS sequence data and a Bayesian phylogenetic analysis. American Journal of Botany. 91:1208–1218. https://doi.org/10.3732/ajb.91.8.1208.

Mitchell A, Mitter C, Regier JC. 2000. More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Systematic Biology. 49:202–224. https://doi.org/10.1093/sysbio/49.2.202.

Mucina L, Rutherford MC. 2006. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. South African National Biodiversity Institute, Pretoria.

National Center for Biotechnology Information. 2020. Accessed on 1 May 2020. Retrieved from https://www.ncbi.nlm.nih.gov/.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Kõljalg U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS One 1:e59. https://doi.org/10.1371/journal.pone.0000059.

Norris LC, Norris DE. 2015. Phylogeny of anopheline (Diptera: Culicidae) species in southern Africa, based on nuclear and mitochondrial genes. Journal of Vector Ecology 40:16–27. https://doi.org/10.1111/jvec.12128.

Otsuka Y. 2011. Variation in number and formation of repeat sequences in the rDNA ITS2 region of five sibling species in the *Anopheles barbirostris* complex in Thailand. Journal of Insect Science 11: 137. https://doi.org/10.1673/031.011.13701.

Pawlowski J, Szadziewski R, Kmieciak D, Fahrni J, Bittar G. 1996. Phylogeny of the infraorder Culicomorpha (Diptera: Nematocera) based on 28S RNA gene sequences. Systematic Entomology 21:167–178. https://doi.org/10.1046/j.1365-3113.1996.d01-5.x.

Posada D. 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25:1253–1256. https://doi.org/10.1093/molbev/msn083.

Puslednik L, Russell RC, Ballard JWO. 2012. Phylogeography of the medically important mosquito *Aedes* (*Ochlerotatus*) *vigilax* (Diptera: Culicidae) in Australasia. Journal of Biogeography 39:1333–1346. https://doi.org/10.1111/j.1365-2699.2012.02683.x.

Quenouille B, Bermingham E, Planes S. 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. Molecular Phylogenetics and Evolution 31:66–88. https://doi.org/10.1016/s1055-7903(03)00278-1.

Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (www.boldsystems.org). Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Rattanarithikul R, Harrison BA, Harbach RE, Panthusiri P, Coleman RE. 2006. Illustrated keys to the mosquitoes of Thailand IV. *Anopheles*. Southeast Asian Journal of Tropical Medicine and Public Health 37:1–128.

Reid JA, Knight KL. 1961. Classification within the subgenus *Anopheles* (Diptera, Culicidae). Annals of Tropical Medicine & Parasitology 55: 474–488. https://doi.org/10.1080/00034983.1961.11686077.

Reinert JF, Harbach RE, Kitching IJ. 2009. Phylogeny and classification of tribe *Aedini* (Diptera: Culicidae). Zoological Journal of the Linnean Society 157:700–794. https://doi.org/10.1111/j.1096-3642.2009.00570.x.

Sallum MAM, Schultz TR, Wilkerson RC. 2000. Phylogeny of Anophelinae (Diptera: Culicidae) based on morphological characters. Annals of the Entomological Society of America 93:745–775. https://doi.org/10.1603/0013-8746(2000)093[0745:poadcb]2.0.co;2.

Sallum MAM, Schultz TR, Foster PG, Aronstein K, Wirtz RA, Wilkerson RC. 2002. Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. Systematic Entomology 27:361–382. https://doi.org/10.1046/j.1365-3113.2002.00182.x.

Schuettpelz E, Pryer KM. 2007. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56:1037–1050. https://doi.org/10.2307/25065903.

Schwarz G. 1978. Estimating the dimension of a model. The Annals of Statistics 6:461–464. https://doi.org/10.1214/aos/1176344136.

Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M. 2006. 4SALE–a tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics 7:498. https://doi.org/10.1186/1471-2105-7-498.

Seibel PN, Müller T, Dandekar T, Wolf M. 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Research Notes 1:91. https://doi.org/10.1186/1756-0500-1-91.

Sharp BL, Kleinschmidt I, Streat E, Maharaj R, Barnes KI, Durrheim DN, Ridl FC, Morris N, Seocharan I, Kunene S, La Grange JJ. 2007. Seven years of regional malaria control collaboration—Mozambique, South Africa, and Swaziland. The American Journal of Tropical Medicine and Hygiene 76:42–47. https://doi.org/10.4269/ajtmh.2007.76.42.

Stamatakis A. 2014. rAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Sudia WD, Chamberlain RW. 1962. Battery-operated light trap, an improved model. Mosquito News 22:126–129.

Sung GH, Sung JM, Hywel-Jones NL, Spatafora JW. 2007. A multi-gene phylogeny of Clavicipitaceae (Ascomycota, Fungi): Identification of localized incongruence using a combinational bootstrap approach. Molecular Phylogenetics and Evolution 44:1204–1223. https://doi.org/10.1016/j.ympev.2007.03.011.

Swain S, Mohanty A, Tripathy HK, Mahapatra N, Kar SK, Hazra RK. 2010. Molecular identification and phylogeny of Myzomyia and Neocellia series of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). Infection, Genetics and Evolution 10:931–939. https://doi.org/10.1016/j.meegid.2010.05.016.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15:1647–1657. https://doi.org/10.1093/oxfordjournals.molbev.a025892.

Valkiūnas G, Atkinson CT, Bensch S, Sehgal RN, Ricklefs RE. 2008. Parasite misidentifications in GenBank: how to minimize their number? Trends in Parasitology 24:247–248. https://doi.org/10.1016/j.pt.2008.03.004.

Vinuesa P, Silva C, Lorite MJ, Izaguirre-Mayoral ML, Bedmar EJ, Martínez-Romero E. 2005. Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from *rrs*, *atpD*, *recA* and *nifH* sequences, and their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands. Systematic and Applied Microbiology. 28:702–716. https://doi.org/10.1016/j.syapm.2005.05.007.

Wang G, Li C, Zheng W, Song F, Guo X, Wu Z, Luo P, Yang Y, He L, Zhao T. 2017. An evaluation of the suitability of COI and COII gene variation for reconstructing the phylogeny of, and identifying cryptic species in, anopheline mosquitoes (Diptera Culicidae). Mitochondrial DNA Part A. 28:769–777. https://doi.org/10.1080/24701394.2016.1186665.

Yan ZT, Fu WB, Chen B. 2019. Complete mitochondrial genomes of *Anopheles aconitus* and *Anopheles splendidus* and phylogenetics analysis of known mtgenomes in the subgenus *Cellia* (Diptera: Culicidae: Anophelinae). Mitochondrial DNA Part B. 4:1848–1850. https://doi.org/10.1080/23802359.2019.1613185.