PNAS RESEARCH ARTICLE | EVOLUTION

Phylogenomics redefines the evolutionary history of mosquitoes

Mac P. Pierce[a,1], Brian M. Worthington[a,b], Shaolin Han[c], Xue-Bing Ni[a,b], Yunshi Liao[a,c], Marcus H. Shum[a,b], Yi Guan[a,b,d], Edward C. Holmes[b,e], and Tommy T. Lam[a,b,c,1]

Edited by Andrew Clark, Cornell University, Ithaca, NY; received July 17, 2025; accepted August 28, 2025

Mosquitoes have a substantial impact on human and animal health, but their deeper evolutionary relationships have been difficult to resolve. We inferred a time-calibrated phylogenetic history of mosquitoes using conserved genome-wide markers from representatives of major lineages. Our analyses revealed that codon bias and positive selection in subfamily Anophelinae contributed to a substantial level of branch attraction bias between Anophelinae and outgroup taxa, which in our view has misled previous phylogenetic analyses of mosquitoes. Accounting for this systematic phylogenetic bias led to a revised view of mosquito evolution, including the nonmonophyly of subfamily Culicinae. Similarly, we dated the origin of mosquitoes to the mid-Cretaceous (~106 Mya) and most extant genera to after the KPg boundary <66 Mya, 100 My younger than previous estimates and coincident with the origin of *Plasmodium* parasites. Our study provides a foundation for future analyses of the evolution of mosquito-borne disease.

mosquitoes | phylogenomics | evolution

Significance

Our results provide strong phylogenomic evidence that both the timing and pattern of mosquito evolution is very different than previously thought, and that mosquitoes are a relatively young group of organisms. We also found a striking similarity with the pattern and timing of the evolution of *Plasmodium* Malarial parasites, which suggests the evolutionary histories of parasite and vector are tightly linked. Our results contribute to a better understanding of mosquito evolution, provide insight about methodological approaches for studying mosquito phylogenomics, and pave the way for building hypotheses about how mosquitoes evolved as disease vectors.

Mosquitoes transmit a wide range of microbial pathogens that have a substantial impact on human civilization. Mosquito-borne diseases have influenced the course of human history and continue to afflict hundreds of millions of people each year (1, 2). Emerging and re-emerging mosquito-borne diseases have become increasingly problematic in recent decades and will likely become more so in the face of climate change (3, 4). Expanding ranges of mosquito-borne pathogens and invasive mosquito vectors are projected to put billions of people at risk in the coming decades (5–7).

Establishing an accurate evolutionary framework of mosquitoes is an important aspect of the public health response to mosquito-borne diseases. For example, previous studies have utilized time-calibrated mosquito phylogenies to study the evolutionary dynamics of genes involved in vectorial capacity, the evolution of host-use across mosquito lineages, and changes in the speciation rate of mosquitoes in response to climate change and atmospheric CO_2 levels (8–10). The conclusions gained from these and related studies improve our understanding of the genetic, evolutionary, and ecological forces that shape vectorial capacity in mosquitoes and can be leveraged to improve public health strategies for mosquito control. However, an accurate time-resolved phylogenetic history of mosquitoes has remained elusive. Previous phylogenetic studies have produced conflicting genus and tribe level relationships, and discordant temporal or phylogenetic signals between mitochondrial and nuclear markers, codon positions, and substitution models are frequently reported (9, 11–15).

Nonetheless, all previous phylogenetic studies agree on two fundamental aspects of mosquito evolution. First, all mosquito species are classified into two subfamilies, Anophelinae and Culicinae, and subfamily Anophelinae has consistently been identified as an ancient lineage sister to all other genera (9, 12). Second, mosquitoes are considered to have an ancient origin. Most hypotheses for mosquito evolution favor a Triassic origin (~220 Mya), with an early to mid-Jurassic split between the Anophelinae and Culicinae subfamilies (180 to 200 Mya) (9, 12–14). Major extant lineages, including *Aedes*, *Culex*, and *Anopheles*, are held to have appeared by at least 100 Mya (Fig. 1*A*) (9, 12–14). According to this current evolutionary framework, the generic diversity of mosquitoes was mostly established prior to the KPg (Cretaceous–Paleogene) boundary and subsequent diversification in the last 65 My has mainly occurred at the subgenus and species level.

However, several lines of evidence suggest that the ancient origin of mosquito evolution may be considerably overestimated. An insect-wide phylogenomic analysis obtained estimates for the origin of mosquitoes that are 100 My more recent, an evolutionary gap that is also reflected in the fossil record (16). Specifically, the oldest mosquito fossils are dated to only 72.1 to 99.6 Mya and comprise a single stem group–the Burmaculicinae–basal to all modern Anopheline and Culicine mosquitoes (17–21). Notably, the oldest crown group mosquito fossils from recognizable extant genera occur

Author affiliations: [a]State Key Laboratory of Emerging Infectious Diseases, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China; [b]Laboratory of Data Discovery for Health Limited, Hong Kong Special Administrative Region, China; [c]Centre for Immunology and Infection Limited, Hong Kong Special Administrative Region, China; [d]Shanghai Institute of Virology, School of Medicine, Shanghai Jiao Tong University, Shanghai 200025, China; and [e]School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2006, Australia

Author contributions: M.P.P. and T.T.L. designed research; M.P.P. performed research; M.P.P., B.M.W., S.H., X.-B.N., Y.L., M.H.S., Y.G., E.C.H., and T.T.L. analyzed data; B.M.W. and E.C.H. guidance for analysis and interpretation, manuscript editing; S.H., X.-B.N., Y.L., M.H.S., and Y.G. guidance for analysis and interpretation; T.T.L. funding and supervision, guidance for analysis and interpretation; and M.P.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: macppierce@gmail.com or ttylam@hku.hk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2519291122/-/DCSupplemental.

Published October 6, 2025.

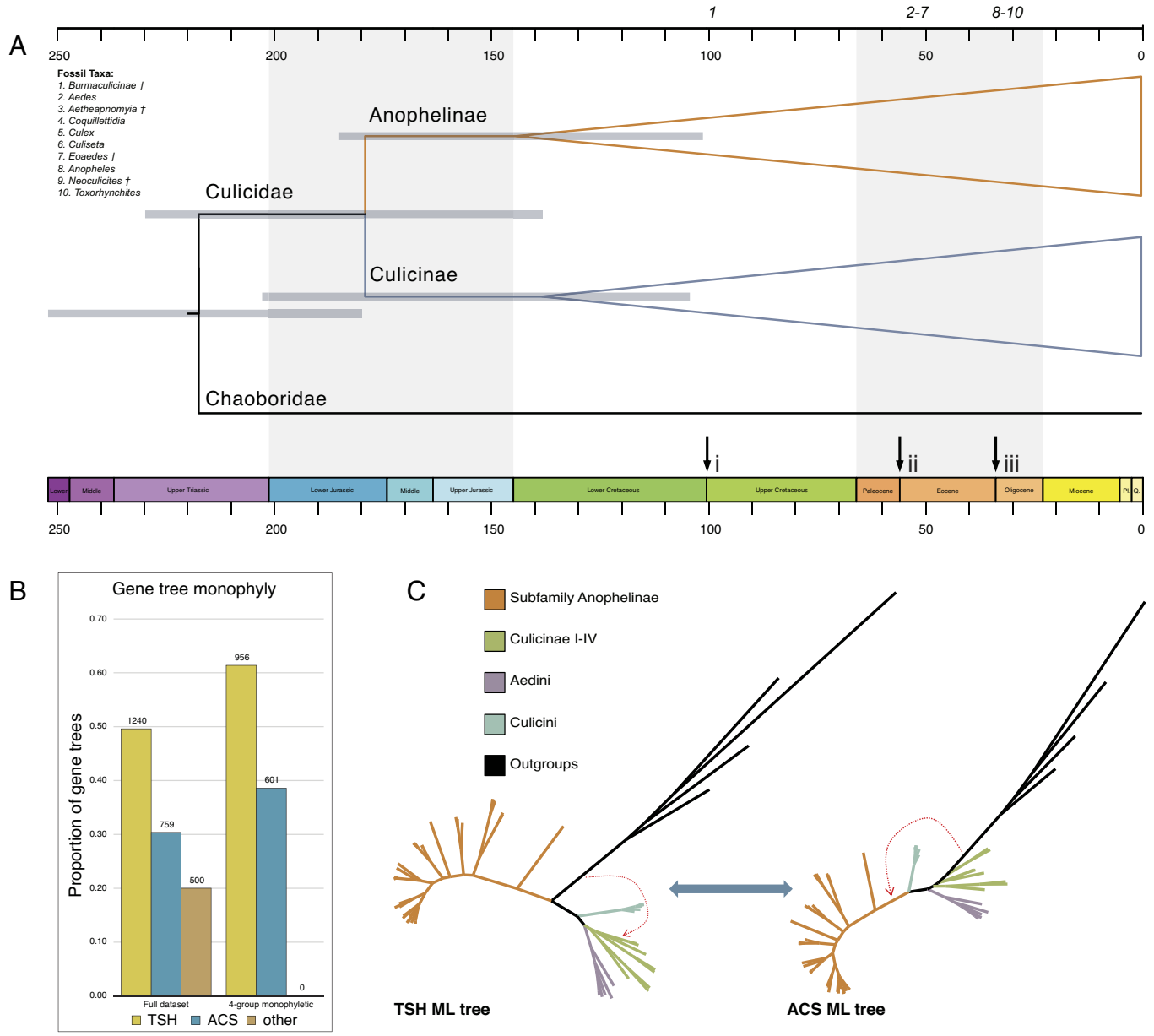

Fig. 1. Incongruence in the mosquito phylogeny. (*A*) Current two-subfamily hypothesis (TSH) of mosquito evolution. Node bars reflect the earliest and latest estimates based on recent molecular studies (9, 13, 14). Numbers at the top scale bar refer to fossils, listed at left. † indicates an extinct group. Numerals at bottom scale bar indicate first fossil appearances of: i) the mosquito lineage Burmaculicinae ii) unambiguous fossils of subfamily Culicinae; iii) unambiguous fossils of subfamily Anophelinae. (*B*) Proportion of gene trees supporting the TSH or ACS hypothesis, or neither. (*C*) Unrooted TSH and ACS trees, red arrow indicating the change in outgroup position between them.

no earlier than the Eocene, at most 56 Mya and possibly younger (22). Although the sparse record of mosquito fossils may be explained by the general rarity with which soft-bodied organisms are preserved, the comparatively rich fossil record of closely related midges with similar biological and ecological traits provides a counterweight to this argument (23, 24).

In order to reassess the pattern and timing of early mosquito evolution given these discrepancies, we performed a phylogenomic analysis using genome-wide nuclear markers from representatives of major mosquito clades. Mosquitoes exhibit rapid gene turnover, which varies across gene families and between mosquito lineages (8, 25). Additionally, phylogenomic analyses can be strongly influenced by low or nonrandom taxon-occupancy, the use of nonhomologous loci, and differences in evolutionary rate between lineages (26). To address these issues we used two datasets of conserved genomic markers for evolutionary analyses, based on benchmarking universal single-copy orthologs (BUSCOs) and ultraconserved elements (UCEs), respectively (27, 28). Through comprehensive phylogenetic analysis we identified the presence of two strongly supported but competing mosquito topologies, which were the result of systematic bias induced by codon bias and positive selection. Dissecting these conflicting phylogenetic signals and their drivers revealed evidence for a revised evolutionary pattern and timing of mosquitoes distinct from previous hypotheses.

Results

Incongruence in the Mosquito Phylogeny. Maximum likelihood (ML) phylogenetic analysis of the full set of BUSCO genes resulted in a species tree supporting the accepted two-subfamily structure of mosquitoes (*SI Appendix*, Fig. S1). However, there was substantial

incongruence between gene trees and the species tree (*SI Appendix*, Fig. S2). In particular, we observed a strong phylogenetic signal for the nonmonophyly of Culicinae and a sister relationship between subfamily Anophelinae and the genus *Culex*, in conflict with the accepted subfamily structure (9, 12–14). We designate these conflicting phylogenetic signals as the "two-subfamily hypothesis" (TSH) and the "Anophelinae-*Culex* sister" hypothesis (ACS), respectively. To reduce noise stemming from poorly resolved gene trees, we filtered out gene trees with nonmonophyletic groups for Anophelinae, *Aedes*, *Culex*, or outgroup taxa. The remaining "four-group monophyletic" (4GM) gene trees were entirely split between the alternate TSH and ACS topologies, indicative of systematic bias (Fig. 1*B*). We then constructed two subsets of 4GM loci based on whether their gene trees supported the TSH (N = 956) or ACS (N = 601) topology. Gene tree discordance patterns were mostly similar between the subsets of TSH and ACS loci, except for the placement of *Uranotaenia*, which was identified as a rogue taxon in 18% of loci and was therefore removed from the sequence alignment (*SI Appendix*, Figs. S3 and S4). ML analysis of the subsets of TSH and ACS loci demonstrated that while the outgroup-rooted phylogenetic trees of the two subsets showed substantially different evolutionary patterns, unrooted internal mosquito relationships were identical, and therefore incongruence between the TSH and ACS trees was based entirely on outgroup position (Fig. 1*C* and *SI Appendix*, Figs. S5 and S6).

In order to assess the impact of the phylogenetic inference method and model choice on topology, we analyzed the datasets using several different models as well as Bayesian inference and distance-based approaches. In all cases, the species trees retained the same topologies. At the gene tree level, in some cases there were changes in topology, but these were infrequent and the overall proportions of gene trees supporting the TSH or ACS trees remained similar. We therefore concluded that our topological observations were robust to analysis method and model choice.

We further assessed support for the conflicting topologies and for internal node mosquito relationships with four-cluster likelihood mapping (FcLM schemes A and B, *SI Appendix*, Fig. S7 *A* and *B*) (29). The majority of quartets favored the ACS topology over the TSH topology (54.8% vs. 45.2%), and ACS loci had less conflict (disagreement between possible quartet topologies) and discordance (inability to resolve relationships) compared to TSH loci (Dataset S5*A*). The set of internal mosquito relationships shared by the ACS and TSH topologies were highly supported across all datasets, but TSH loci showed substantially greater discordance indicative of a weaker phylogenetic signal (Dataset S5*B*). These observations indicate the presence of a widespread systematic error affecting the phylogenetic analysis of a large proportion of genes resulting in strong support for two incongruent phylogenetic hypotheses (26).

Branch Attraction Bias Drives Incongruence. We assessed differences between TSH and ACS loci to identify the factors underlying the incongruent topologies. GC content was significantly lower in TSH vs. ACS loci (P-value < 0.001), although the difference was restricted to Anophelinae and *Culex*, which had the highest GC content of all groups (Fig. 2 *A* and *B* and *SI Appendix*, Fig. S8). Heterogeneity in GC content can mislead phylogenetic inference due to attraction between groups with similar base compositions (30). Although incongruence was based only on outgroup position, the clustering of high-GC Anophelinae and *Culex* in ACS loci suggested base compositional attraction (BCA) may be involved. Evolutionary distance was significantly higher in TSH vs. ACS loci (Fig. 2*A* and *SI Appendix*, Fig. S8). Branch length values of TSH loci were higher across all clades, but highest in Anophelinae and outgroups (Fig. 2*C*). Long-branch attraction (LBA) between rapidly evolving lineages is a common occurrence and frequently involves outgroup taxa (31). LBA between Anophelinae and outgroups could explain the observed incongruence but would also imply that the universally accepted two-subfamily structure of mosquitoes is incorrect.

Common methods to identify artifacts of attraction bias include removing the affected clades entirely or in tranches and repeating analyses (32, 33). Since both alternate attraction-based scenarios involved subfamily Anophelinae, we performed several analyses by removing Anopheline sequences and assessing the impact on outgroup position. Initial analyses performed using FcLM (scheme C, *SI Appendix*, Fig. S7*C*) resulted in near-zero support for a quartet topology consistent with TSH, while the majority of quartets across all datasets supported a quartet topology consistent with ACS (Dataset S5*C*). Similarly, ML analysis excluding all Anophelines resulted in a tree topology consistent with ACS (Fig. 2*D* and *SI Appendix*, Fig. S9*A*), an effect also seen at the gene tree level (*SI Appendix*, Fig. S9*B*). Sequential removal of the fastest-evolving Anopheline sequences in 10% tranches showed increasing support for ACS and decreasing support for TSH, based on both FcLM and gene tree analysis (*SI Appendix*, Fig. S10 *A* and *B*). Notably, when the fastest 50% of Anopheline sequences had been removed, the resulting ML tree precisely matched the ACS tree (*SI Appendix*, Fig. S11). These results demonstrate that the TSH topology is likely an artifact of attraction bias between Anophelinae and outgroup taxa consistent with LBA, and strongly suggest that subfamily Culicinae is paraphyletic.

Topological Support Is Structured by Codon Position. Previous mosquito phylogenetic studies have shown that sequence type and codon position can influence tree topology, and have encouraged the use of amino acids and codon position two (9, 13, 14). In our case, support for the ACS topology was mostly based on the third codon position whereas codon positions one, two, and amino acids favored the TSH topology, as demonstrated by ML, gene tree, and FcLM (scheme A) analyses (*SI Appendix*, Figs. S12 and S13 and Dataset S5*A*). A recent study similarly found a phylogenetic signal for nonmonophyly of Culicinae based on codon position three, which was then removed due to concerns of substitution saturation and the baseline assumption of Culicine monophyly (9). Saturation diminishes phylogenetic signal and is a particular problem at the third codon position because substitutions are mostly synonymous (34). We therefore used multiple analyses to assess whether the ACS topology was an artifact of substitution saturation.

Contrary to the expected effects of saturation, we found that codon position three had higher quality phylogenetic signal compared to amino acids and the other codon positions, with all pairwise genetic distances below the 0.75 substitutions/site (range 0.31 to 0.69) saturation cut-off (35). Specifically, the set of internal mosquito relationships shared between TSH and ACS (FcLM scheme B) were supported across all datasets (codon positions and amino acids), but codon position three had the least conflict and discordance (Dataset S5*B*). Codon position three also had the strongest correlation between patristic and uncorrected distances, consistent with unsaturated data (*SI Appendix*, Fig. S14) (36, 37). We also found that support for the ACS topology was associated with better phylogenetic signal and fewer substitutions in codon position three. ACS loci had fewer unresolvable quartets and were better able to resolve relationships among mosquito taxa compared to TSH loci (*SI Appendix*, Fig. S15*A*). Additionally, although TSH and ACS loci had the same proportion of informative sites at codon position three, several factors including pairwise identity, branch length, entropy,

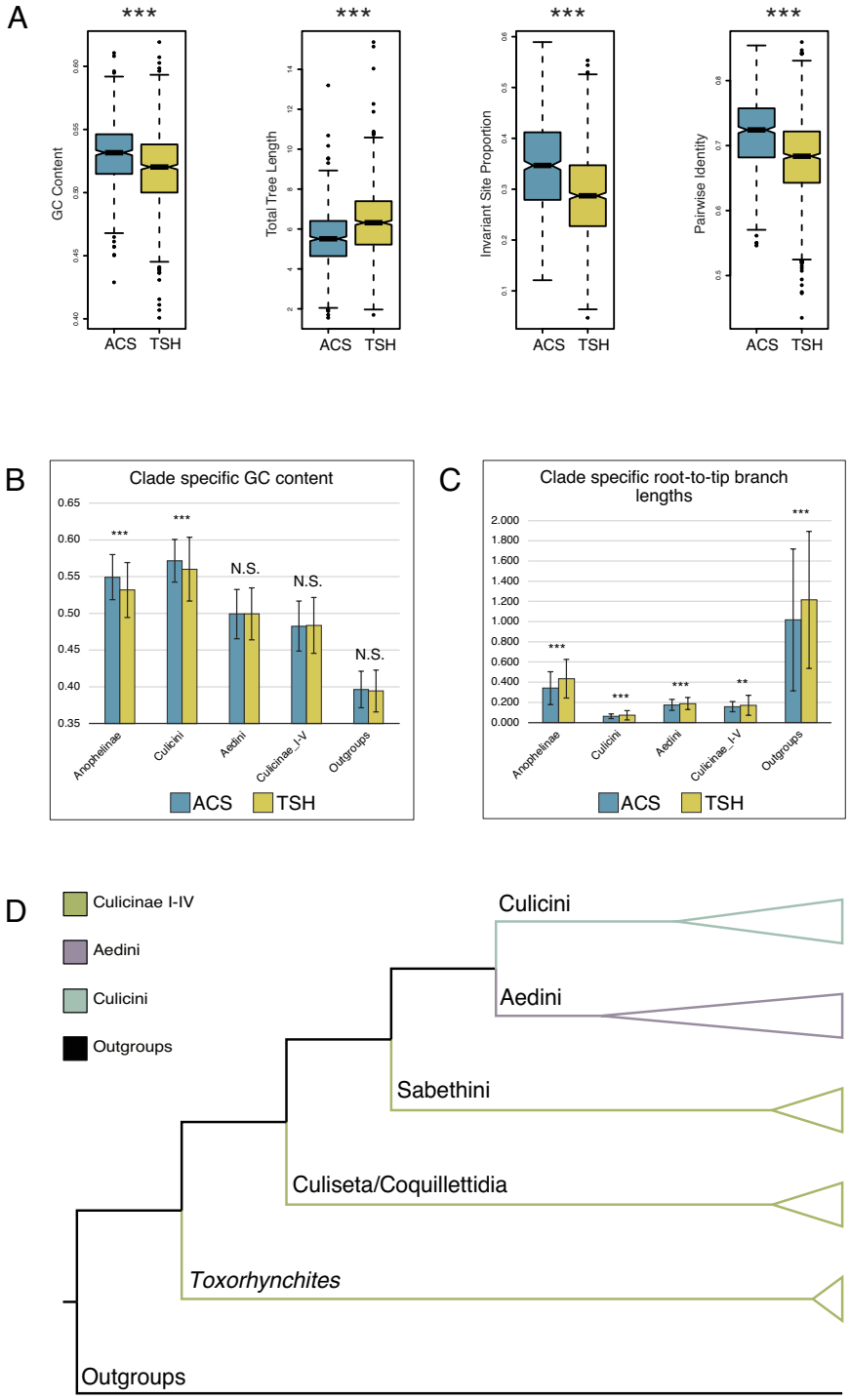

Fig. 2. LBA between Anophelinae and outgroup taxa. (*A*) Boxplots comparing GC content and evolutionary distance between locus alignments and trees based on their topological support. (*B*) Clade-specific GC content, comparing loci based on their topological support. (*C*) Clade-specific root-to-tip branch length, comparing loci based on their topological support. (*D*) ML tree of mosquitoes with all Anopheline taxa removed. Based on the full dataset, partitioned by locus, using a GTR+F+I+R4 substitution model and 1,000 replicates of the SH-like approximate likelihood ratio test. All support values are 100/100. Clades have been collapsed for simplification.

and patristic versus uncorrected distances all indicated that the substitution rate was higher and phylogenetic signal was lower in the third codon position of TSH loci (*SI Appendix*, Fig. S15*B*). Last, FcLM (scheme C) analysis demonstrated that when Anopheline taxa were excluded, all codon position and amino acid datasets had greater support for a quartet topology consistent with ACS than one consistent with TSH, and that the level of conflict and discordance was lowest for codon position three (Dataset S5*C*).

These results strongly indicate that ACS support is not a result of saturation, and that in these data amino acids and codon positions one and two are particularly sensitive to the effects of branch attraction bias. This is likely due to the fact that within these conserved loci, codon position three contains the majority of informative sites and thus phylogenetic signal. Conversely, the high level of conservation in the other positions and in amino acids may cause their phylogenetic signal, which is based on a small number of informative

sites, to be disproportionately sensitive to the effects of branch attraction-induced bias.

Codon Bias in Subfamily Anophelinae. Given that ACS support was derived mainly from codon position three, we reassessed FcLM (scheme A) support for outgroup placement based only on this position. Taxon-specific support values were split entirely between the alternate TSH and ACS topologies but varied widely across taxa, including between 9 to 92% of quartet support for TSH in subfamily Anophelinae, indicating that the mechanism driving branch attraction bias in codon position three varies across a relatively shallow phylogenetic scale (*SI Appendix*, Fig. S16*A*). Because GC content also differed between TSH and ACS loci, we evaluated whether codon bias in Anophelinae was a driving force of branch attraction bias. In Anopheline loci, there was a strong correlation between the effective number of codons (ENc) and the GC content of codon position 3 (GC3), indicating that Anopheline codons are subject to strong bias due to GC compositional constraint (*SI Appendix*, Fig. S16*B*) (38). Moreover, ENc, GC3, root to tip branch length (RtTBL), and ACS support were strongly correlated in Anopheline taxa (Fig. 3 *A* and *B* and *SI Appendix*, Fig. S16 *C–E*). Anopheline taxa with high ACS support also had high GC3 content and shorter RtTBL values, while the reverse was true for species with high TSH support. Basal Anophelines tended to support the ACS topology, suggesting that Anopheline taxa have shifted away from GC-rich codon preference (*SI Appendix*, Fig. S17). Analysis of relative synonymous codon usage (RSCU) across all taxa supported this view. Outgroup taxa were heavily biased toward AT-rich codons, whereas there was a preference for GC-rich codons across mosquitoes (*SI Appendix*, Fig. S18). GC preference was highest in *Culex* as well as those Anopheline taxa with high ACS support, which had overall similar codon usage values (*SI Appendix*, Fig. S19). This implies that GC-rich codon preference is the ancestral mosquito state, and that subsequent shifts toward stronger AT-preference have occurred to varying degrees in Anophelines. Consistent with this, there was a clear trend of increased usage of AT-rich codons in the fastest evolving mosquito loci, which was strongest in Anopheline taxa and weakest in *Culex* (Fig. 3*D*).

Ancestral sequence analysis provided additional evidence of an increase in AT-content in Anophelines. GC3 content was significantly higher in the ancestral sequence of the MRCA of subfamily Anophelinae compared to contemporary Anophelines, and there were around twice as many GC-to-AT substitutions than AT-to-GC substitutions (Fig. 3*C* and *SI Appendix*, Fig. S20*A*). There were also significantly more mutations in TSH than ACS loci, indicating that the increase in AT-content impacted phylogenetic inference. ML phylogenetic analysis using the ancestral sequences generated increased support for the ACS topology in both the species and gene trees (*SI Appendix*, Fig. S20 *B* and *C*). Collectively, these results indicate that a shift in codon bias acting heterogeneously across Anophelines has contributed to branch attraction bias between Anophelinae and the outgroups. Because this shift involves compositional bias and an increase in substitutions, both LBA and BCA between Anophelinae and the outgroups may have occurred.

Positive Selection Contributes to Incongruence. The codon position two and amino acid datasets had near exclusive preference for the TSH topology. Yet, both showed signs of branch attraction bias based on FcLM (scheme C) analysis (Dataset S5*C*). We reasoned that strong positive selection in subfamily Anophelinae may have contributed to branch attraction bias, which may be detected through site-specific phylogenetic analysis (39). To assess whether this had occurred in our dataset we calculated the difference in site-specific likelihood support (ΔSSLS) based on topology testing of loci using the TSH and ACS topologies in codon position two, which is only subject to nonsynonymous mutations. We detected a significant preference for the ACS topology ($|\Delta SSLS| > 0.1$) in a substantial proportion of sites, demonstrating the presence of a latent ACS phylogenetic signal within codon position two that was indiscernible in our previous sequence-based analyses (*SI Appendix*, Fig. S21). High ΔSSLS values ($|\Delta SSLS| > 0.5$) were mostly associated with TSH support whereas moderate support values ($0.5 > |\Delta SSLS| > 0.1$) favored the ACS topology, consistent with the expected effects of selection. We verified this by performing selection analyses, finding that Anopheline branches were subject to positive selection at a greater proportion of loci than other clades, and also in TSH compared to ACS loci (*SI Appendix*, Fig. S22 *A* and *B*). Codon-based selection analysis had similar results, with significantly more sites under selection in Anophelines and in TSH loci (*SI Appendix*, Fig. S22 *C* and *D*). Additionally, ΔSSLS for the TSH topology was far stronger in codons found to be under selection in Anophelines than it was when averaged across all sites, particularly for codon positions one and two, demonstrating a direct relationship between positive selection in Anophelinae and support for the TSH topology (*SI Appendix*, Fig. S23).

The impact of selection on topological support was further investigated by removing codons from the alignment that were positively selected in Anophelinae and performing gene tree, FcLM (scheme A), and topology testing analyses. Although only 4.4% of total codons were found to be under selection in Anophelinae with statistical significance, removing these sites increased support for the ACS topology across all analyses, demonstrating that these sites significantly influenced phylogenetic inference despite their relatively low number (*SI Appendix*, Fig. S24 *A–C*). We then removed codons based on their d_N/d_S values in sequential tranches starting from 5% up to a total of 25% of codons. This revealed that as tranches of codons with the highest d_N/d_S values were removed in succession, support for the ACS topology increased across all analyses (*SI Appendix*, Fig. S24 *A–C*). Hence, these results demonstrate that strong positive selection in Anophelinae directly contributes to topological incongruence and support for the TSH topology, and further substantiates the ACS topology as the true mosquito phylogeny. Although previous mosquito phylogenetic studies have emphasized the use of codon position two and amino acids, our findings suggest that these datasets may be more sensitive to bias in highly conserved genes, likely due to their comparatively low phylogenetic signal.

Timing the Early Evolution of Mosquitoes. Our phylogenetic results of BUSCO loci revealed considerable systematic bias linked to evolutionary rate in Anophelinae. We therefore utilized a secondary dataset for divergence estimation based on UCEs. ML phylogenetic analysis of the UCE dataset produced a species tree precisely matching the ACS topology, demonstrating that these highly conserved loci are advantageous for resolving deep mosquito phylogenetic relationships (*SI Appendix*, Fig. S25). We estimated divergence times using BEAST2 by performing independent runs with different sets of UCE loci, prior distributions, and constraint trees based on the TSH and ACS topologies (*SI Appendix*, Fig. S26 and Datasets S4 and S5) (40). To calibrate the analyses, we used seven MRCA prior distributions based on the best available evidence from the fossil record and previous phylogenetic studies and spanning both early and late divergence times (Dataset S4) (9, 12–14, 16, 20–22, 41, 42). Three were based on the ancestral nodes of mosquitoes, mosquitoes plus their nearest relative

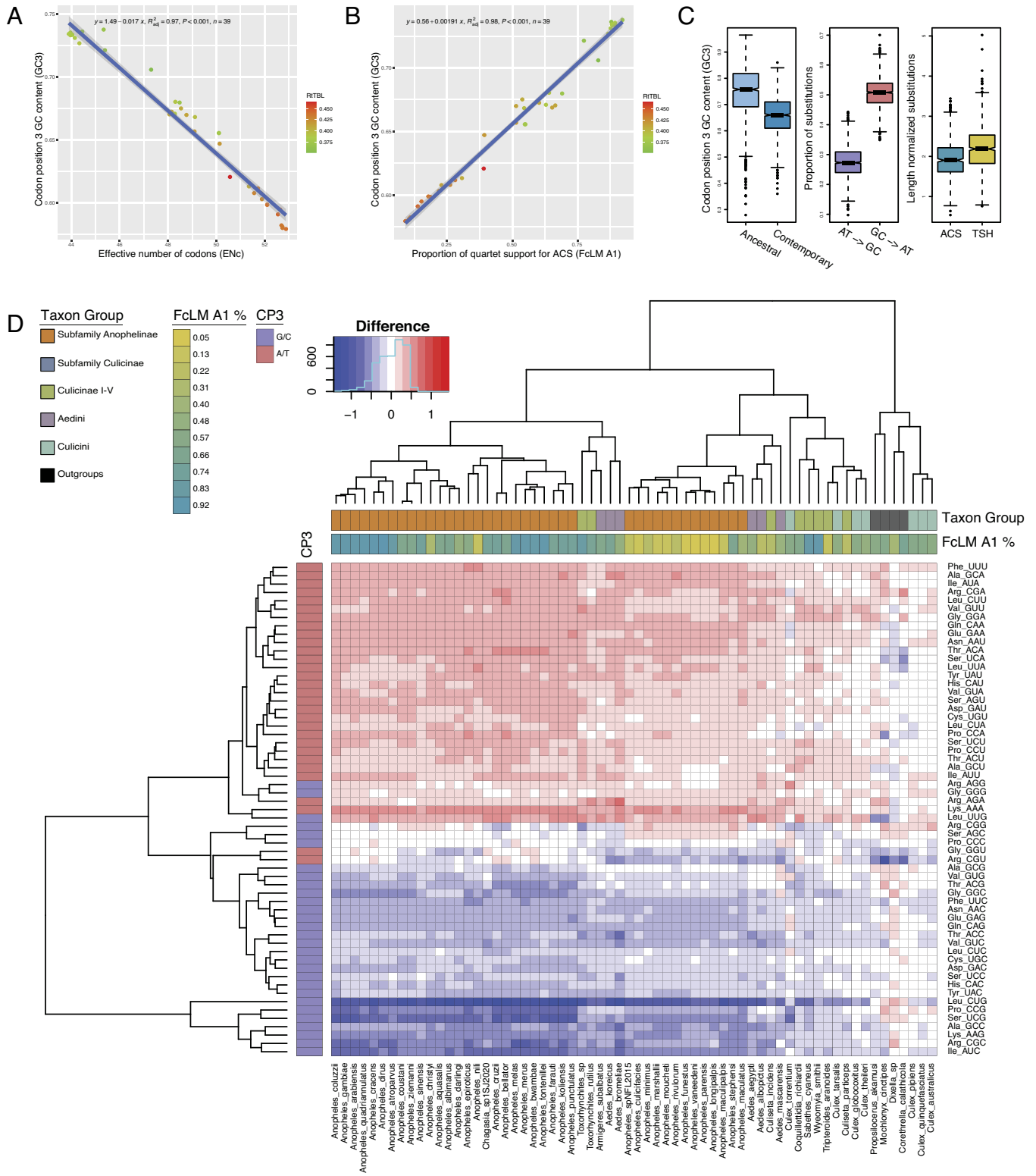

Fig. 3. Codon usage bias contributes to branch attraction bias. (*A*) Relationship between codon position three GC content (GC3) and ENc in Anopheline taxa. (*B*) Relationship between GC3 and support for the ACS phylogeny in the form of FcLM A1 support in Anopheline taxa. (*C*) Comparison of: GC3 between contemporary Anopheline taxa and the estimated ancestral most recent common ancestor (MRCA) of Anophelinae; proportion of AT to GC and GC to AT mutations in Anopheline taxa; number of mutations in Anopheline loci that supported the ACS or TSH phylogeny, normalized to locus length. (*D*) Difference in RSCU between the 100 fastest and 100 slowest evolving loci for all taxa. Red and blue indicate preferential RSCU in fast and slowly evolving loci, respectively. Color gradient on the vertical axis indicates whether a given codon ends in GC or AT. Color gradients on the horizontal axis indicate clade membership of taxa and support for the ACS phylogeny based on FcLM A1 support.

Chaoboridae, and the superfamily Culicoidea. The remaining four calibrations utilized extant genera that have valid fossils and were represented in our study.

Estimated divergence times were similar across datasets, prior distributions, and ACS vs. TSH constraint trees (*SI Appendix*, Fig. S27). Our dating estimates painted a much younger picture

of mosquito evolutionary history than previous studies, which have almost universally proposed a Jurassic or Triassic origin of mosquitoes (Fig. 4) (9, 12–14), but were largely consistent with previous studies of higher-level taxonomic groups such as flies and insects (16, 41). We estimated that the split between mosquitoes and their nearest relative Chaoboridae occurred in the late Cretaceous around 106 Mya (95% HPD: 102.3 to 110.8) similar to the estimate of 118 Mya by Bertone et al. (41). The MRCA of the crown group mosquitoes was dated to 73 Mya (95% HPD: 69.2 to 77.4), close to the estimate of 75 Mya of ref. 16 and over 100 My younger than previous studies (16). Except for Mansonia all mosquito lineages in our study were of Cenozoic origin, emerging after the mass extinction event at the KPg boundary. This suggests that early diversification of mosquito lineages was likely driven by the post-KPg radiation of mammals and birds, their predominant hosts (9, 43, 44). Notably, divergences between major mosquito clades coincided with important global climatic fluctuations including the Paleocene–Eocene Thermal Maximum and Early Eocene Climatic Optimum (EECO), which are associated with massive faunal turnover in mammals (45).

Discussion

Our study revealed a phylogenetic hypothesis for mosquito evolution in which subfamily Culicinae is nonmonophyletic, subfamily Anophelinae is the sister group of tribe Culicini, and mosquitoes are around 100 My younger than previously thought (Fig. 4). We found that the incongruence between this topology (termed ACS) and the currently accepted two-subfamily framework, supported by previous morphological and molecular phylogenetic studies, was due to branch attraction bias caused by extensive positive selection and codon bias in subfamily Anophelinae. However, the ACS phylogenetic signal was both subtle and difficult to resolve, varying substantially across codon positions, loci, and taxa, and was only supported by a minority of the total sites. Although we were able to generate substantial evidence for the validity of the ACS phylogenetic signal, we were only able to detect it through the combined use of highly conserved genetic markers and analyses explicitly designed to detect phylogenetic discordance and test for phylogenetic bias. This explains why previous mosquito phylogenetic studies, which have

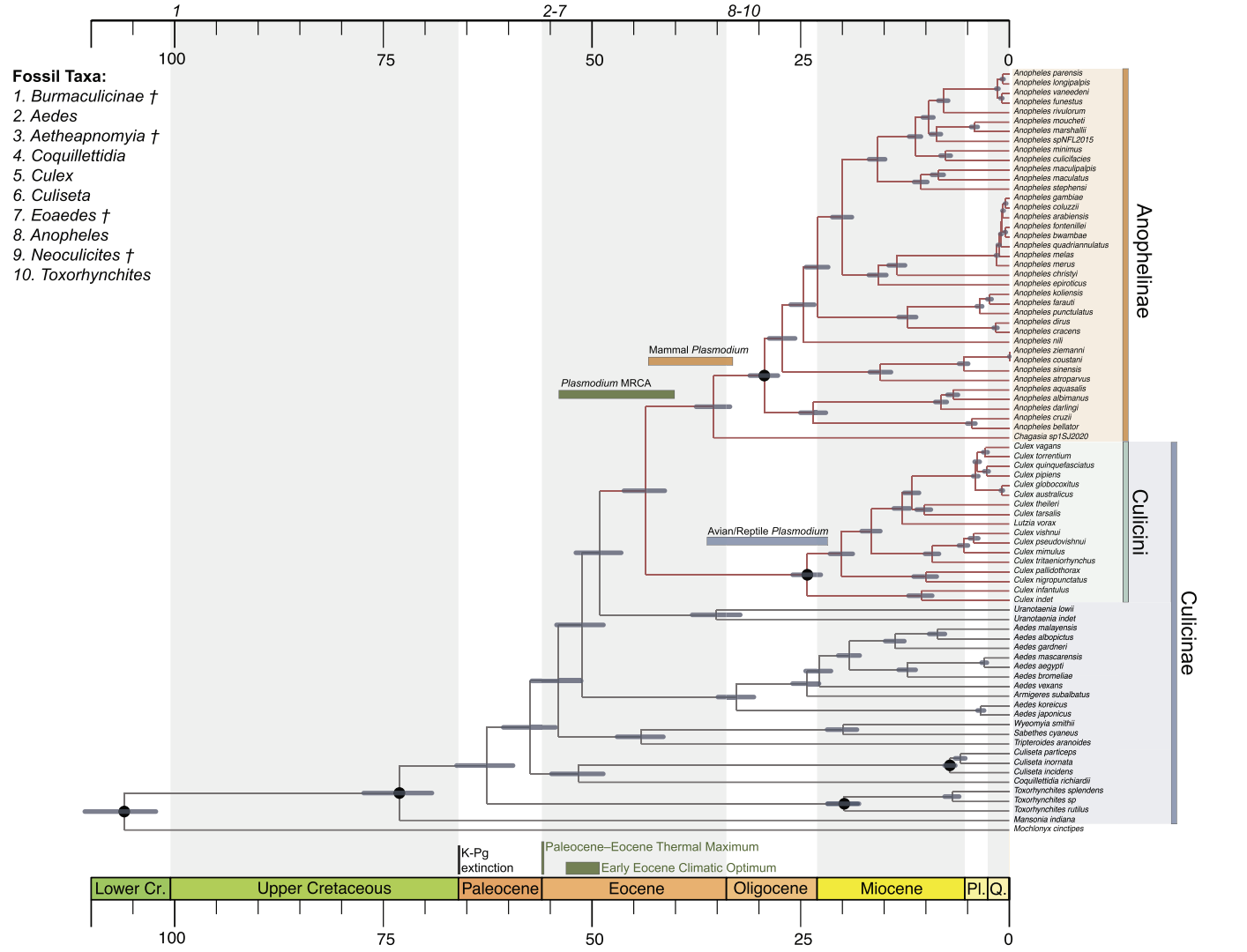

Fig. 4. Dated phylogenetic tree of mosquito relationships. Based on the BEAST2 run using the longest 100 UCE loci that supported the ACS tree, using log-normal priors and the ACS constraint tree, and run for three replicates of 1 billion generations each. Calibration points are marked with black circles (some outgroup taxa not shown for clarity). The top scale bar shows the ages of all currently identified mosquito fossils, listed at *Left*. The bottom scale bar shows geological epochs and important events in Earth's history. Vertical colored bars at *Right* show the currently accepted subfamilies. Red branches highlight the sister relationship of Anophelinae and Culicini. Horizontal colored bars in the center plot represent the 95% credibility intervals of the ages of indicated Plasmodium clades from previous studies (46, 47).

mostly been based on morphological traits or less conserved genetic markers, have either not detected this signal or misattributed it to phylogenetic bias due to saturation.

Phylogenomic conflict between morphological and molecular analyses is relatively common and can arise from rapid morphological radiation (48, 49). If Anopheline ancestors went through a period of significant morphological innovation, as suggested by the strong level of positive selection in subfamily Anophelinae demonstrated here, previous morphology-based phylogenetic studies may have misinterpreted this as ancient divergence. Mosquito molecular studies have typically been based on mitochondrial genomes, which can have both a higher mutation rate and stronger AT bias compared to nuclear genomes (50, 51). Both factors were associated with the phylogenetic incongruence observed here, suggesting that previous molecular studies based on mosquito mitochondrial genomes may have been impacted by the effects of branch attraction bias. The only previous mosquito phylogenetic study which used a large dataset of both nuclear genes and species also identified a phylogenetic signal for the nonmonophyly of subfamily Culicinae, but attributed this to the effect of substitution saturation in codon position three (9). Saturation is characterized by the accumulation of multiple substitutions at a given site, which leads to poor phylogenetic signal and obscures true evolutionary relationships. This is often associated with codon position three due to its relatively high rate of synonymous substitutions. However, we found that the phylogenetic signal for the ACS topology had a stronger and less discordant phylogenetic signal than the TSH topology, indicating that it was not caused by saturation.

We also found that codon position three had a stronger phylogenetic signal compared to the other codon positions and amino acids, resulting in a greater ability to resolve relationships. We hypothesized that this may be due to the highly conserved nature of these genes, in which a large proportion of sites in codon positions one, two, and amino acids, are either invariant or nearly invariant. Because invariant sites do not contain any phylogenetic information content, the phylogenetic signal from these datasets is based on a relatively small set of informative sites. Therefore, positive selection at even a small number of sites may result in a strong misleading signal for the wrong tree, especially when this selection results in convergent substitutions between otherwise unrelated groups (39). Through site-specific analysis, we were able to demonstrate that a large proportion of codon position two sites supported the ACS topology, despite the codon position two dataset strongly supporting the TSH topology at the sequence level. We then found that site-specific support for the TSH topology was stronger in codons that had undergone positive selection, and that removing even a small proportion of these codons reversed phylogenetic signal and increased support for the ACS tree.

By contrast, codon position three contains very few uninformative sites and may therefore be more robust to bias. Although codon position three is more sensitive to substitution saturation as described above, this is only problematic if it misleads phylogenetic inference by obscuring the true relationships. We were able to demonstrate that codon position three sites in loci that supported the ACS topology had a lower substitution rate and better phylogenetic signal compared to loci that supported the TSH topology, indicating that the ACS loci were in fact less saturated. Consequently, to the extent that saturation occurred in codon position three, it contributed to branch attraction bias and to the TSH topology. However, because this false signal was minimal in the codon position three dataset, which showed strong support for the ACS topology across analyses, saturation did not pose a major problem by obscuring the true phylogenetic relationships. This suggests that codon position three may be especially useful when using conserved data, and that the specific phylogenetic impact of saturation should be explicitly tested before removing codon position three from analyses as is frequently done.

Our study also revealed a timing of mosquito evolution in which mosquitoes originated in the Cretaceous and radiated in the Paleogene, contrasting sharply with current estimates in which mosquitoes emerged 150 to 250 Mya (9, 13, 14). Previous studies have typically calibrated the two mosquito subfamilies using the Cretaceous fossils *Priscoculex* (Anophelinae) and *Paleoculicis* (Culicinae), which are known from amber deposits dated to the late Cretaceous approximately 72 to 99 My old (18, 19). In our view the taxonomic assignment of these fossils was questionable given their morphological descriptions, and we considered them to represent stem mosquito lineages rather than crown group members within either subfamily. Confirming this view, sometime after our analyses both species were indeed synonymized within the stem subfamily Burmaculicinae (21). Given these recent taxonomic changes, all pre-Eocene (>55.8 Ma) mosquito taxa known from fossil evidence are now classified into the subfamily Burmaculicinae, a stem lineage considered to be sister to all other extant and fossil mosquitoes, while all fossils from subfamilies Anophelinae and Culicinae are no older than 33.9 Ma and 55.8 Ma, respectively (17, 22). Therefore current divergence estimates which suggest a Triassic origin of mosquitoes are in disagreement with the mosquito fossil record (21). The discrepancies between our results and those of previous studies therefore appear to be caused by differences in calibration choices, and our estimates of a Cretaceous origin and Paleocene–Eocene diversification of mosquito lineages are more consistent with the updated mosquito fossil record than currently accepted estimates (9, 13, 14, 21).

Our findings provide a framework for generating insights and reevaluating previous conclusions about several aspects of mosquito evolution. This includes the relationship between mosquito speciation and climate, the biogeographic origin of mosquito lineages, evolutionary rates, and important vector characteristics like host-use and the relationship between mosquito vectors and pathogens (8–10, 52). For example, one important previous study found that mammal-feeding behavior has evolved repeatedly in mosquitoes (9). When this is considered in the context of our findings, it appears that mammal-feeding behavior is the ancestral state of mosquitoes and that the use of avian, reptilian, and amphibian hosts emerged independently in some groups. Similar to our findings for mosquitoes, the evolutionary radiations of both mammals and birds mainly occurred following the KPg boundary, suggesting that evolutionary pressures surrounding host-use may have played a major role in shaping the early evolution of mosquito lineages (43, 53). The same study identified that several important mosquito lineages, including subfamily Anophelinae and the tribes Aedini and Culicini, have basal clades of species that are endemic to the New World (9). In our study these three groups plus *Uranotaenia* shared a common ancestor 51.3 Mya (95% HPD: 48.6 to 54.3), conforming almost exactly to the EECO which occurred between 49 to 54 Mya, and which had a major impact on biodiversity (Fig. 4) (45). Our results therefore suggest that these mosquito lineages originated in the New World and subsequently diversified globally, which would explain the presence of basal clades of New World endemics in these groups. A New World origin of Anophelines 35.4 Mya (95% HPD: 33.5 to 37.5) would also explain the Australasian distribution of the Anopheline lineage Bironella, which is likely a basal group sister to the genus *Anopheles*, and could have reached this area via Antarctica prior to the total glaciation of the continent due to the Late Cenozoic Ice Age at the Eocene–Oligocene Boundary which began 34 Mya (54).

Our results also shed light on vector evolution in mosquitoes, particularly in regard to *Plasmodium* malarial parasites. The

Plasmodium phylogeny is largely divided between two clades–those that primarily infect mammals and those that infect birds or reptiles (46, 47, 55). Anopheline mosquitoes are the near-exclusive vectors of mammalian *Plasmodium* (56). Avian and reptilian *Plasmodium* parasites have been detected in a variety of mosquito hosts, but *Culex* is by far the most commonly identified vector (57, 58). This structure mirrors the predominantly mammal feeding behavior of Anophelinae and the mosaic host-preference pattern of *Culex* which feeds widely on birds, reptiles, and mammals (9). Our mosquito phylogeny contained a sister relationship between Anophelinae and *Culex*, suggesting that *Plasmodium* parasites evolved in the common ancestor of these clades. Our divergence dating estimates supported this scenario, with a striking similarity to previous estimates of *Plasmodium* evolution (Fig. 4) (46, 47). The MRCA of *Plasmodium* is dated to 47 Mya, overlapping the ancestral branch of Anophelinae and *Culex* 43 to 49 Mya. We dated the split between Anophelinae and *Culex* to 43.6 Mya (95% HPD: 41.3 to 46.2), nearly identical to the estimated divergence of mammal and saurian *Plasmodium* at 45 Mya. Likewise, our estimates for the MRCAs of Anophelinae and *Culex* were 35.4 (95% HPD: 33.5 to 37.5) and 24.2 Mya (95% HPD: 22.6 to 26.0), respectively, similar to the MRCAs of mammalian and saurian *Plasmodium* of 38 and 28 Mya. We therefore infer that malarial infection of mosquitoes originated in the Eocene with the introduction of a Haemosporidian ancestor of *Plasmodium* into the ancestral lineage of Anophelinae and *Culex*. Codiversification of mosquito and parasite lineages followed, likely driven by the radiation of their vertebrate hosts. In this case, the evolution of host-preference in mosquitoes may have driven the host range of *Plasmodium*, particularly given that early Anopheline and *Culex* species were likely mammal and bird/reptile specialists, respectively (9).

Establishing an accurate phylogenetic history for mosquitoes is fundamental to understanding their evolution as vectors, which is especially important given that mosquito-borne diseases have been among the leading causes of human morbidity and mortality throughout history (2). Our study suggests that the evolutionary history of mosquitoes is both younger and more complex than previously thought, due to inherent challenges in analyzing mosquito molecular data. Based on this we recommend that future studies investigating deep phylogenetic relationships in mosquitoes utilize large datasets of conserved genetic markers, and include analyses explicitly designed to reveal discordance and test for phylogenetic bias. However, an important limitation of our study was that relatively few species were included and several groups did not have any representatives in our dataset, including the tribes Aedeomyiini, Ficalbiini, Hodgesiini, and Orthopodomyiini. Additionally, we had representatives of only 14 out of the more than 100 mosquito genera currently recognized (59). Including additional species and lineages is unlikely to impact the finding of Culicine nonmonophyly, since this was demonstrated to be the result of branch attraction bias between Anophelinae and outgroup taxa. However, it would affect the relationships between mosquito lineages in the ACS phylogeny, and consequently the evolutionary insights derived from it. For example, the striking similarity between mosquito and *Plasmodium* evolution is due in part to the sister relationship between Anophelinae and Culicini in the ACS phylogeny. Therefore, validating this relationship in future studies by including a greater taxonomic sampling will be critical.

More broadly, our results demonstrate the significance of the problem of systematic bias for phylogenetic studies focused on deeper evolutionary relationships. In the case of mosquitoes, this bias has led to a strongly supported but inaccurate picture of evolutionary relationships spanning a multitude of studies conducted across multiple decades. This same pattern may occur in other groups, and it is therefore advisable that deep-scale phylogenetic studies employ methods to explicitly test for the presence of latent phylogenetic signals. In our case, we were able to initially detect this signal through discordance analysis, and further understand its nature through comparative analysis of conflicting gene alignments and datasets based on different codon positions. This approach allowed us to develop and test phylogenetic hypotheses using a variety of methods to evaluate the competing signals. We suggest that future studies may benefit from a similar approach, particularly for assessing deeper evolutionary relationships.

Materials and Methods

Data Generation. We extracted BUSCO and UCE loci from mosquito and outgroup species with publicly available reference genomes, whole genome sequencing or transcriptome datasets (Dataset S1). The keyword "Culicidae" was used to search the NCBI genome and SRA databases, and data were downloaded from all available mosquito species (as of May 2023). For outgroup selection we extracted data from the genome assemblies of four taxa: *Mochlonyx cinctipes* (Chaoboridae), *Propsilocerus akamusi* (Chironomidae), *Corethrella calathicola* (Corethrellidae), and *Dixella* sp (Dixidae). Together with mosquitoes, each of the four families within the superfamily Culicoidea (Dixidae, Chaoboridae, Corethrellidae, Culicidae) and the two superfamilies within the infraorder Culicomorpha (Chironomoidea, Culicoidea) were represented. BUSCO and UCE sequences were then extracted computationally.

To obtain a more complete dataset to maximize the taxonomic diversity for divergence analysis, we sequenced UCEs from an additional 15 species (nine *Culex* sp., two *Aedes* sp., and four others) collected in Hong Kong and Singapore (Dataset S1). Mosquito specimens were collected from urban and periurban sites and nature reserves in Hong Kong SAR and Singapore. Genomic DNA was extracted destructively using the Qiagen DNeasy® Blood & Tissue Kit. Species identification was done through morphological examination and DNA barcoding of the Cytochrome Oxidase I gene using the primers described in ref. 60.

Dual-indexed libraries were prepared for each specimen using the KAPA Hyper Prep kit according to the protocol described in ref. 61. Individual libraries were pooled in sets of eight and used as input for target enrichment, which was done using the Diptera 2.7Kv1 probe set (Daicel Arbor Biosciences), hybridization settings of 65 °C for 22 h, and otherwise done according to the manufacturer's protocol (62). Post-enrichment libraries were sequenced in-house on an Illumina MiSeq instrument and externally on an Illumina NovaSeq 6000 instrument by Novogene (63). In total, we sequenced UCEs from 15 mosquito species, 11 of which were *Culex* sp., 2 were *Aedes* sp., and 4 represented other species groups.

Sequence Assembly. Sequencing reads from SRAs and newly sequenced specimens were trimmed and de novo assembled into contigs using metaSPAdes (64, 65). Single-copy orthologous genes were then extracted from de novo assembled SRA contigs and reference genome assemblies using the program BUSCO and the diptera_odb10 database (27, 66). Species-specific datasets were generated and any taxa with less than 40% recovery of the 3,285 targeted BUSCO loci were removed from all downstream analyses. The amino acid sequences were aligned and back-translated into codon alignments, trimmed to a 75% threshold, and loci with less than 85% taxon occupancy were removed. UCE loci were extracted from de novo assembled contigs and reference genome assemblies with PHYLUCE v1.7.1 (67). Paralogous loci were removed, and taxa with less than 40% recovery of the 2,711 targeted UCE loci were removed from all downstream analyses. Sequences were then aligned, trimmed to a 75% threshold, and loci with less than 85% taxon occupancy were removed. After filtering, the datasets comprised 4.3 Mb from 2,499 BUSCO loci of 65 individuals and 895 Kb from 1,526 UCE loci of 83 individuals (Dataset S2).

Phylogenetic Analyses.

Partitioning schemes and ML phylogenetic analyses. Phylogenetic analyses were done using concatenated supermatrices, individual locus alignments, and multiple partitioning schemes. BUSCO alignments were partitioned by locus and codon position and analyzed as nucleotides and amino acids. UCE alignments were partitioned by locus and using the SWSC-EN method, and analyzed as nucleotides (68).

ML phylogenetic inference was done with IQ-TREE v2.2 (69). GTR+F+I+R4, MFP, and MFP+MERGE were used as substitution models for nucleotide data and a LG4X model was used for amino acid data (70). Branch support was calculated

using 1,000 replicates of the ultrafast bootstrap method and SH-like approximate likelihood ratio test (71, 72). Analyses were conducted using the High Performance Cluster at The Laboratory of Data Discovery for Health ($D^2$4H, https://www.d24h.hk/).

Discordance analysis on gene trees. BUSCO gene trees were rooted on the outgroup taxa and poorly supported nodes were collapsed with Newick utilities (73). These were used as input for discordance analysis using gene concordance factors in IQ-TREE and the program ASTRAL v5.7.8 (74, 75).

Subsetting BUSCO loci based on topology. BUSCO loci were filtered by removing loci whose gene trees were nonmonophyletic for subfamily Anophelinae, tribe Culicini, tribe Aedini, or the outgroups. The remaining loci were subset into groups based on their topological support for Culicine monophyly. Loci with a monophyletic grouping of subfamily Anophelinae and outgroup taxa were designated as TSH, and loci that included an unrooted monophyletic grouping of subfamily Anophelinae and tribe Culicini were designated as ACS. All filtered loci fit one of these two criteria. Hereafter, TSH and ACS refer to the alternate phylogenetic hypotheses and the subsets of loci that support them.

Rogue taxon identification. Rogue taxa were identified in BUSCO gene trees with the program RogueNaRok (76). *Uranotaenia lowii*, the sole representative of tribe Uranotaeniini, was identified as rogue in a large proportion of gene trees and was therefore removed from downstream analyses except where specified.

FcLM. Three phylogenetic hypotheses (A-C) were assessed using FcLM in IQ-TREE (29). Scheme A evaluated support for outgroup position relative to major mosquito clades and was used to assess support for the alternate phylogenetic hypotheses (TSH and ACS) across different datasets, taxa, and analyses. Scheme B evaluated support for internal mosquito relationships between major clades and was used as a control for evaluating the impact of analyses that involved pruning sequences or sites from datasets. Scheme C evaluated the impact of excluding subfamily Anophelinae from analysis, and was used to test the alternative hypotheses of LBA and BCA. The taxonomic compositions and clade definitions of the three FcLM schemes are given in Dataset S3.

Locus and taxon statistics. A variety of standard statistics for alignments, gene trees, and taxa were calculated using a combination of programs and custom scripts including PHYLUCE v1.7.1, IQ-TREE2, DiscoVista (77), PhyKIT (78), Sliding-Window Site Characteristics, and the python package CodonU (79). Comparative statistics were plotted using the boxplot function in R and statistical significance was determined with Mann–Whitney *U* tests and two-way ANOVA. For correlative statistics, plots were made using ggplot2 (80) and linear regression models were applied with the R package ggpmisc (81).

Evaluating alternate attraction-based hypotheses. The alternate hypotheses of LBA and BCA were assessed using the Siddal and Whiting (SAW) method and removal of fast-evolving proteins (RFP) (33, 82). For SAW analyses, all Anopheline taxa were removed entirely from alignments. For RFP analyses, the fastest evolving sequences of each Anopheline taxon, using RtTBL as a proxy for evolutionary rate, were sequentially removed from alignments in five iterations of 10% tranches, and the resulting alignments were analyzed using ML, FcLM, and gene tree analysis.

RSCU. RSCU was calculated from all loci of each taxon using CodonW v.1.4.4 (http://codonw.sourceforge.net//). The difference in RSCU values between the top and bottom 100 loci of each taxon based on RtTBL values was also calculated. Values were plotted using the heatmap and heatmap.2 functions in R.

Ancestral sequence analysis. Ancestral sequences, GC content, and substitutions were estimated for the basal nodes of subfamily Anophelinae and tribes Aedini and Culicini using the program TreeTime (83). The basal node sequences were used for ML phylogenetic analysis.

Site-specific log-likelihoods. SSLS was calculated for ultrametric TSH and ACS topologies using IQ-TREE2 and the R package "ape" (39, 84, 85). The SSLS values of the TSH topology were subtracted from those of the ACS topology to generate the difference in SSLS (ΔSSLS), such that positive and negative values reflected site-specific support for the ACS and TSH topologies, respectively.

Selection analyses. Selection analyses were done using the adaptive Branch-Site Random Effects Likelihood and Mixed Effects Model of Evolution (MEME) methods implemented in the HyPhy software package (86–88). Multiple runs were done for each locus using the alternate ACS and TSH topologies and different test branches.

Site-removal analyses. Codons under positive selection in subfamily Anophelinae were removed from alignments in six iterations using the perl script "deleteAlignColumn.pl" (available from https://www.biostars.org/p/55555/). The first iteration removed codons with P-value < 0.1 from MEME analyses, and the remaining five iterations removed codons based on their nonsynonymous substitution rate in tranches of 5% up to a maximum of 25% of total codons. The resulting alignments were used for ML analysis with FcLM and topology testing of the ACS and TSH topologies. Double-delta log-likelihood (ΔΔLL) values were calculated by subtracting the ΔLL of the TSH topology from those of the ACS topology, such that positive values of ΔΔLL reflected support for ACS, while negative values indicated support for TSH.

Divergence dating. Divergence dating was done with BEAST v2.7.3 using 12 analyses each with three replicate runs of 500 million generations and sampling every 10,000 generations (40). The 12 analyses were divided across three datasets each with 100 UCE loci, log-normal versus uniform MRCA prior distributions, and a fixed ACS versus TSH constraint tree. All analyses were run using locus-partitioned alignments with linked trees, a GTR substitution model with four gamma rate categories and estimated substitution rate, Birth–Death model for the tree prior, log-normal distributions with a mean of 0.1 for the ucld.mean parameters, and optimized relaxed clock models (89). Constraint trees fitting the ACS and TSH topologies were generated by rescaling branch lengths using the program "chronopl" in the R package ape, setting a minimum age of 99 My for the mosquito MRCA node (85).

Seven MRCA calibrations were used based on divergence estimates from previously published studies and fossilized mosquito specimens. The root node calibration of Culicoidea was chosen to span the estimated age of the oldest mosquito fossils (~99 Ma) and the estimated age of the insect order Diptera (~250 Ma) from previously published studies (16, 41, 42). A second calibration was used for the clade including mosquitoes and their nearest relative Chaoboridae, based on the ages of the oldest known stem lineage mosquito and Chaoborid fossils, at ~99 Ma and ~187 Ma, respectively (17, 23). The prior distribution of the calibration of family Culicidae spanned the ages of fossilized crown group mosquitoes from the Oligocene to Eocene and the upper age range of mosquitoes based on recent estimates (95% CI 37.9 to 213 Ma), with the mean chosen to reflect the age of the oldest stem lineage mosquito fossils (subfamily Burmaculicinae) at 99 Ma (9, 18, 19, 22). The remaining four calibrations were based on the extant clades with known fossil representatives, with the distributions chosen based on the geological strata from which the fossils originated. Details regarding MRCA prior calibrations are provided in Dataset S4.

Log files were examined for convergence between runs and appropriate ESS values in Tracer v1.7.1 (90). Tree files from replicates were combined using a burn-in of 50% with LOGCOMBINER, and TREEANNOTATOR was used to generate a time-calibrated phylogeny (91).

Data, Materials, and Software Availability. Short-read UCE sequencing data have been deposited at the NCBI Sequence Read Archive Database (https://www.ncbi.nlm.nih.gov/sra/) under the BioProject accession code PRJNA1309844 (63).

ACKNOWLEDGMENTS. We thank Julie Chu, David Pierce, Mary Phillips, and the members of the Guan lab in general for their support. We also thank two anonymous reviewers for their suggestions on improving the manuscript. This work was supported by the InnoHK initiative (C2i) from the Innovation and Technology Commission, and the Health and Medical Research Fund (18170432) from the Health Bureau, of the Hong Kong Special Administrative Region Government.

1. L. H. V. Franklinos, K. E. Jones, D. W. Redding, I. Abubakar, The effect of global change on mosquito-borne disease. *Lancet Infect. Dis.* **19**, e302–e312 (2019).
2. T. S. Athni *et al.*, The influence of vector-borne disease on human history: Socio-ecological mechanisms. *Ecol. Lett.* **24**, 829–846 (2021).
3. A. M. Kilpatrick, S. E. Randolph, Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* **380**, 1946–1955 (2012).
4. F. J. Colon-Gonzalez *et al.*, Projecting the risk of mosquito-borne diseases in a warmer and more populated world: A multi-model, multi-scenario intercomparison modelling study. *Lancet Planet. Health* **5**, e404–e414 (2021).
5. M. U. G. Kraemer *et al.*, Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nat. Microbiol.* **4**, 854–863 (2019).
6. J. P. Messina *et al.*, The current and future global distribution and population at risk of dengue. *Nat. Microbiol.* **4**, 1508–1515 (2019).
7. M. E. Sinka *et al.*, A new malaria vector in Africa: Predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24900–24908 (2020).
8. D. E. Neafsey *et al.*, Mosquito genomics. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. *Science* **347**, 1258522 (2015).

9. J. Soghigian *et al.*, Phylogenomics reveals the history of host use in mosquitoes. *Nat. Commun.* **14**, 6252 (2023).
10. C. Tang, K. E. Davis, C. Delmer, D. Yang, M. A. Wills, Elevated atmospheric CO(2) promoted speciation in mosquitoes (Diptera, Culicidae). *Commun. Biol.* **1**, 182 (2018).
11. R. E. Harbach, The Culicidae (Diptera): A review of taxonomy, classification and phylogeny*. *Zootaxa* **1668**, 591–638 (2007).
12. K. R. Reidenbach *et al.*, Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. *BMC Evol. Biol.* **9**, 298 (2009).
13. C. Lorenz, J. M. P. Alves, P. G. Foster, L. Suesdek, M. A. M. Sallum, Phylogeny and temporal diversification of mosquitoes (Diptera: Culicidae) with an emphasis on the Neotropical fauna. *Syst. Entomol.* **46**, 798–811 (2021).
14. A. F. da Silva *et al.*, Culicidae evolutionary history focusing on the Culicinae subfamily based on mitochondrial phylogenomics. *Sci. Rep.* **10**, 18823 (2020).
15. N. Zadra, A. Rizzoli, O. Rota-Stabelli, Chronological incongruences between mitochondrial and nuclear phylogenies of *Aedes* mosquitoes. *Life* **11**, 181 (2021).
16. B. Misof *et al.*, Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
17. A. Borkent, D. A. Grimaldi, The cretaceous fossil *Burmaculex antiquus* confirmed as the earliest known lineage of mosquitoes (Diptera: Culicidae). *Zootaxa* **4079**, 457–466 (2016).
18. G. O. Poinar, T. Zavortink, T. Pike, P. Johnston, Paleoculicis minutus (Diptera: Culicidae) n. gen., n. sp., from Cretaceous Canadian amber, with a summary of described fossil mosquitoes. *Acta Geolog. Hispanica* **35**, 119–128 (2000).
19. G. Poinar, T. J. Zavortink, A. Brown, Priscoculex burmanicus n. gen. et sp. (Diptera: Culicidae: Anophelinae) from mid-Cretaceous Myanmar amber. *Histor. Biol.* **32**, 1157–1162 (2019).
20. R. E. Harbach, Libanoculex intermedius is not a mosquito (Diptera:Culicidae): It is a chaoborid (Chaoboridae). *Zootaxa* **5424**, 139–144 (2024).
21. R. Szadziewski *et al.*, Burmaculex (Diptera: Culicidae) from mid-Cretaceous Burmese amber, with the first description of a male and three new species. *Palaeoentomology* **7**, 539–553 (2024).
22. W. Gika, R. E. Harbach, E. E. Perkovsky, Mosquitoes (Diptera: Culicidae) in Eocene amber from the Rovno region, Ukraine. *Zootaxa* **5016**, 257–270 (2021).
23. A. Borkent, World catalog of extant and fossil Chaoboridae (Diptera). *Zootaxa* **3796**, 469–493 (2014), 10.11646/zootaxa.3796.3.4.
24. L. A. Parry *et al.*, Soft-bodied fossils are not simply rotten carcasses–Toward a holistic understanding of exceptional fossil preservation: Exceptional fossil preservation is complex and involves the interplay of numerous biological and geological processes. *Bioessays* **40**, 1700167 (2018).
25. L. Ruzzante, M. Reijnders, R. M. Waterhouse, Of genes and genomes: Mosquito evolution and diversity. *Trends Parasitol.* **35**, 32–51 (2019).
26. J. L. Steenwyk, Y. Li, X. Zhou, X. X. Shen, A. Rokas, Incongruence in the phylogenomics era. *Nat. Rev. Genet.* **24**, 834–850 (2023).
27. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simao, E. M. Zdobnov, BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
28. B. C. Faircloth *et al.*, Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* **61**, 717–726 (2012).
29. K. Strimmer, A. von Haeseler, Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6815–6819 (1997).
30. P. G. Foster, Modeling compositional heterogeneity. *Syst. Biol.* **53**, 485–495 (2004).
31. J. Bergsten, A review of long-branch attraction. *Cladistics* **21**, 163–193 (2005).
32. M. E. Siddall, M. F. Whiting, Long-branch abstractions. *Cladistics* **15**, 9–24 (1999).
33. H. Brinkmann, M. van der Giezen, Y. Zhou, G. Poncelin de Raucourt, H. Philippe, An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst. Biol.* **54**, 743–757 (2005).
34. X. Xia, Z. Xie, M. Salemi, L. Chen, Y. Wang, An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **26**, 1–7 (2003).
35. R. C. Lewontin, Inferring the number of evolutionary events from DNA coding sequence differences. *Mol. Biol. Evol.* **6**, 15–32 (1989).
36. H. Philippe *et al.*, Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
37. O. Jeffroy, H. Brinkmann, F. Delsuc, H. Philippe, Phylogenomics: The beginning of incongruence? *Trends Genet.* **22**, 225–231 (2006).
38. F. Wright, The "effective number of codons" used in a gene. *Gene* **87**, 23–29 (1990).
39. T. A. Castoe *et al.*, Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8986–8991 (2009).
40. R. Bouckaert *et al.*, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
41. M. A. Bertone, G. W. Courtney, B. M. Wiegmann, Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes. *Syst. Entomol.* **33**, 668–687 (2008).
42. B. M. Wiegmann *et al.*, Episodic radiations in the fly tree of life. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5690–5695 (2011).
43. D. M. Grossnickle, S. M. Smith, G. P. Wilson, Untangling the multiple ecological radiations of early mammals. *Trends Ecol. Evol.* **34**, 936–949 (2019).
44. E. D. Jarvis *et al.*, Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
45. M. O. Woodburne, G. F. Gunnell, R. K. Stucky, Climate directly influences Eocene mammal faunal dynamics in North America. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13399–13403 (2009).
46. M. A. Pacheco *et al.*, A phylogenetic study of *Haemocystidium* parasites and other Haemosporida using complete mitochondrial genome sequences. *Infect. Genet. Evol.* **85**, 104576 (2020).
47. M. A. Pacheco *et al.*, Mode and rate of evolution of haemosporidian mitochondrial genomes: Timing the radiation of avian parasites. *Mol. Biol. Evol.* **35**, 383–403 (2018).
48. J. N. Keating, R. J. Garwood, R. S. Sansom, Phylogenetic congruence, conflict and consilience between molecular and morphological data. *BMC Ecol. Evol.* **23**, 30 (2023).
49. C. Parins-Fukuchi, G. W. Stull, S. A. Smith, Phylogenomic conflict coincides with rapid morphological innovation. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023058118 (2021).
50. C. Haag-Liautard *et al.*, Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol.* **6**, e204 (2008).
51. R. Allio, S. Donega, N. Galtier, B. Nabholz, Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772 (2017).
52. M. A. Pacheco, A. A. Escalante, Origin and diversity of malaria parasites and other haemosporida. *Trends Parasitol.* **39**, 501–516 (2023).
53. S. Claramunt, J. Cracraft, A new time tree reveals Earth history's imprint on the evolution of modern birds. *Sci. Adv.* **1**, e1501005 (2015).
54. C. L. O'Brien *et al.*, The enigma of Oligocene climate and global surface temperature evolution. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25302–25309 (2020).
55. U. Böhme *et al.*, Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res.* **28**, 547–560 (2018).
56. A. Molina-Cruz, T. Lehmann, J. Knockel, Could culicine mosquitoes transmit human malaria? *Trends Parasitol.* **29**, 530–537 (2013).
57. D. Santiago-Alarcon, V. Palinauskas, H. M. Schaefer, Diptera vectors of avian haemosporidian parasites: Untangling parasite life cycles and their taxonomy. *Biol. Rev. Camb. Philos. Soc.* **87**, 928–964 (2012).
58. M. Ferraguti, J. Martínez-de la Puente, S. Ruiz, R. C. Soriguer, J. Figuerola, Landscape and mosquito community impact the avian Plasmodium infection in Culex pipiens. *iScience* **27**, 109194 (2024).
59. R. E. Harbach, *Composition and Nature of the Culicidae (mosquitoes)* (CAB International, Boston, MA, 2024), pp. pages cm.
60. O. Folmer, M. Black, W. Hoeh, R. Lutz, R. Vrijenhoek, DNA primers for amplification of mitochondrial cytochrome c oxidase subunit i from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
61. T. C. Glenn *et al.*, Adapterama i: Universal stubs and primers for 384 unique dual-indexed or 147, 456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* **7**, e7755 (2019).
62. B. C. Faircloth, Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods Ecol. Evol.* **8**, 1103–1112 (2017).
63. M. P. Pierce *et al.*, Data from "Phylogenomic analysis of mosquitoes from Hong Kong using ultraconserved elements (UCEs)." NCBI. http://www.ncbi.nlm.nih.gov/bioproject/1309844. Deposited 23 August 2025.
64. A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics* **70**, e102 (2020).
65. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
66. M. Manni, M. R. Berkeley, M. Seppey, E. M. Zdobnov, BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* **1**, e323 (2021).
67. B. C. Faircloth, Phyluce is a software package for the analysis of conserved genomic loci. *Bioinformatics* **32**, 786–788 (2016).
68. V. A. Tagliacollo, R. Lanfear, Estimating improved partitioning schemes for ultraconserved elements. *Mol. Biol. Evol.* **35**, 1798–1811 (2018).
69. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
70. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
71. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
72. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
73. T. Junier, E. M. Zdobnov, The newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 (2010).
74. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
75. B. Q. Minh, M. W. Hahn, R. Lanfear, New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* **37**, 2727–2733 (2020).
76. A. J. Aberer, D. Krompass, A. Stamatakis, Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Syst. Biol.* **62**, 162–166 (2013).
77. E. Sayyari, J. B. Whitfield, S. Mirarab, Discovista: Interpretable visualizations of gene tree discordance. *Mol. Phylogenet. Evol.* **122**, 110–115 (2018).
78. J. L. Steenwyk *et al.*, PhyKIT: A broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* **37**, 2325–2331 (2021).
79. S. Choudhuri, K. Sau, CodonU: A Python package for codon usage analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **21**, 36–44 (2024).
80. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016). https://ggplot2.tidyverse.org.
81. P. Aphalo, ggpmisc: Miscellaneous Extensions to 'ggplot2' (R package version 0.6.2.9001, 2025). https://docs.r4photobiology.info/ggpmisc/.
82. M. E. Siddall, M. F. Whiting, Long-branch abstractions. *Cladistics Int. J. Willi Hennig Soc.* **15**, 9–24 (1999).
83. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
84. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
85. E. Paradis, K. Schliep, Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
86. S. L. Pond, S. D. Frost, S. V. Muse, HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
87. M. D. Smith *et al.*, Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
88. B. Murrell *et al.*, Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
89. A. J. Drummond, S. Y. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
90. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
91. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).