Syst. Biol. 0(0):1–17, 2022

https://doi.org/10.1093/sysbio/syac019
Advance Access Publication 7 March 2022

A Genomic Perspective on the Evolutionary Diversification of Turtles

SIMONE M. GABLE[1], MICHAEL I. BYARS[1], ROBERT LITERMAN[2] AND MARC TOLLIS[1,*]

[1]*School of Informatics, Computing, and Cyber Systems, Northern Arizona University, PO Box 5693, Flagstaff, AZ 8601, USA and* [2]*Department of Biological Sciences, University of Rhode Island, 120 Flagg Road, Kingstown, RI 0288, USA;*
Simone M. Gable and Michael I. Byars contributed equally to this article.
[*]*Correspondence to be sent to: School of Informatics, Computing, and Cyber Systems, Northern Arizona University, PO Box 5693, Flagstaff, AZ 86011, USA;*
E-mail: marc.tollis@nau.edu.

Received 19 October 2021; reviews returned 28 February 2022; accepted 1 March 2022
Associate Editor: Matthew Hahn

Abstract.—To examine phylogenetic heterogeneity in turtle evolution, we collected thousands of high-confidence single-copy orthologs from 19 genome assemblies representative of extant turtle diversity and estimated a phylogeny with multispecies coalescent and concatenated partitioned methods. We also collected next-generation sequences from 26 turtle species and assembled millions of biallelic markers to reconstruct phylogenies based on annotated regions from the western painted turtle (*Chrysemys picta bellii*) genome (coding regions, introns, untranslated regions, intergenic, and others). We then measured gene tree-species tree discordance, as well as gene and site heterogeneity at each node in the inferred trees, and tested for temporal patterns in phylogenomic conflict across turtle evolution. We found strong and consistent support for all bifurcations in the inferred turtle species phylogenies. However, a number of genes, sites, and genomic features supported alternate relationships between turtle taxa. Our results suggest that gene tree-species tree discordance in these data sets is likely driven by population-level processes such as incomplete lineage sorting. We found very little effect of substitutional saturation on species tree topologies, and no clear phylogenetic patterns in codon usage bias and compositional heterogeneity. There was no correlation between gene and site concordance, node age, and DNA substitution rate across most annotated genomic regions. Our study demonstrates that heterogeneity is to be expected even in well-resolved clades such as turtles, and that future phylogenomic studies should aim to sample as much of the genome as possible in order to obtain accurate phylogenies for assessing conservation priorities in turtles. [Discordance; genomes; phylogeny; turtles.]

Next-generation sequencing has greatly increased the numbers of sampled loci for phylogenomic analyses (Lemmon et al. 2012; Faircloth et al. 2013; Edwards et al. 2016), shedding new light on the branching order of diversification across the history of life. However, phylogenetic studies based on genome-scale data often contain a great deal of heterogeneity in the inferred patterns (Hime et al. 2021; Lopes et al. 2021; Morales-Briones et al. 2021; Singhal et al. 2021), particularly in the form of discordance among sites and gene trees (Kumar et al. 2012). Gene tree-species tree discordance occurs when gene trees are in conflict with the underlying species relationships, whereas site discordance occurs when a subset of sites support different bifurcations in a species tree, and both factors can lead to biased phylogenetic estimates. Understanding patterns and potential sources of phylogenomic heterogeneity will not only help reconcile phylogenetic arguments and reduce bias in tree building, but also provide better models with which to understand trait evolution and help define conservation priorities through comparative methods (Garland et al. 2005; Colston et al. 2020).

Biological factors driving phylogenomic heterogeneity include: differing evolutionary rates across the genome, intralocus recombination, incomplete lineage sorting, and hidden paralogy (Pamilo and Nei 1988; Galtier and Daubin 2008). Artifacts that drive heterogeneity include sequencing errors and substitution model violations such as genetic saturation, compositional heterogeneity, and codon usage bias (Foster 2004; Cooper 2014; Cox et al. 2014). Accounting for gene and site discordance in phylogenomic estimation often requires partitioning schemes which try to account for differing evolutionary rates and model parameters (Kubatko and Degnan 2007), methods that are consistent with the multispecies coalescent accounting for population-level processes such as incomplete lineage sorting (Edwards et al. 2016), as well as strict filtering of sequence alignments that aim to reduce potential errors. The application of hundreds or even thousands of loci is often an attempt to overcome biases stemming from or driving phylogenomic conflict, and these studies often boast strong branch support as measured by bootstrap replicates or posterior probabilities, likely as a result of reduced sampling variance in large data sets (Minh, Hahn et al. 2020). However, competing phylogenomic studies of the same clades that use different data sets and tree building methods can produce conflicting results, despite high levels of statistical support (see modern birds: Jarvis et al. 2014; Prum et al. 2015).

An emerging advantage of the genomic era is the theoretical ability to access the whole genome in order to sample loci representing a diversity of evolutionary rates and coalescent histories (Wolf et al. 2002; Rokas et al. 2003). The most widely used sampling methods include the use of large numbers of coding sequences (CDS; Singhal et al. 2021), anchored hybrid enrichment loci (AHE; Lemmon et al. 2012), ultraconserved elements (UCEs; Faircloth et al. 2013), and transcriptomes (Irisarri et al. 2017), but these are each still reduced representations of a relatively small proportion of the genome (Lynch 2007). Meanwhile, the development of bioinformatics tools and

sequencing technologies have enabled chromosome-scale genome assemblies for nonmodel organisms (see http://dnazoo.org, Dudchenko et al. 2017), potentially providing insights to the genomic regions or types of loci that drive heterogeneous phylogenetic results. Thus, by accessing complete genomes and studying heterogeneity closely, we can more fully comprehend conflicting results across gene trees, obtain more accurate perspectives of species relationships, and gain a fuller understanding of molecular evolution across the genome.

Here, we examined the importance of genome-wide heterogeneity in the phylogenomic estimation of turtles (Order Testudines). Turtles are a near-globally distributed and morphologically distinct clade of shelled reptiles, with a rich fossil history dating from the Triassic Period (Gaffney 1980; Joyce 2007; Joyce et al. 2021). Despite their persistence in the fossil record and across modern ecosystems, almost half of turtle species are listed in the International Union for Conservation of Nature (IUCN) as endangered, critically endangered, or vulnerable to extinction (IUCN 2021). The approximately 350 extant species of turtles are classified into two main groups, Pleurodira ("side-necked") and Cryptodira ("hidden-necked"), which are further divided into 8 superfamilies and 14 families (Uetz et al. 2021). Early molecular phylogenetic results disagreed in the placement of some turtle taxa; in particular, the position of the monotypic big-headed turtle (*Platysternon*) and its relationships with or within Testudinoidea (Shaffer et al. 1997; Krenz et al. 2005). Other questions about deeper turtle relationships such as the position of soft-shell turtles (Trionychoidea), the Americhelydiae, and the position of turtles in the amniote phylogeny persisted into the next-generation sequencing era (Chiari et al. 2012; Crawford et al. 2012, 2015; Brown and Thomson 2017). Since then, phylogenomic studies of turtles have produced fully resolved species trees with 100% statistical support at every node and near complete agreement across studies as measured by bootstrap replicates or posterior probabilities (Crawford et al. 2015; Shaffer et al. 2017).

Phylogenomic heterogeneity has been examined in vertebrate groups such as mammals (Tarver et al. 2016; Liu et al. 2017), squamates (Burbrink et al. 2020; Singhal et al. 2021), birds (Jarvis et al. 2014; Prum et al. 2015), and amphibians (Hime et al. 2021), but there has not yet been an investigation into how gene tree-species tree discordance across the genome affects the inference of relationships between the major turtle lineages. Although phylogenetic conflict often belies more traditional means of measuring certainty via bootstrap support or posterior probabilities (Minh, Hahn et al. 2020), the study of turtle systematics has not yet truly taken advantage of the genomic era; this is despite ample genomic resources being available for a wide range of turtle species (Shaffer et al. 2013; Wang et al. 2013; Tollis et al. 2017; Quesada et al. 2019). Because evolutionary distinctiveness is linked to conservation status in turtles (Colston et al. 2020), and many turtle species have adaptations of physiological and developmental importance such as longevity, resistance to hypoxia, cold tolerance, and the anatomical changes responsible for the turtle shell, a study using genome-scale data to measure phylogenomic heterogeneity is needed in order to obtain the most accurate models of turtle evolution. Our goals in this study were threefold: (1) to revisit discoveries about higher turtle systematics using substantial proportions of the genome; (2) to measure heterogeneity at genes and sites in important events in turtle evolution; and (3) determine what drives heterogeneity in turtle phylogenomics in terms of mechanisms and locus type.

To accomplish these goals, we generated two distinct yet overlapping genomic data sets for turtles. The first comprises 5310 high-confidence single-copy orthologs bioinformatically extracted from the genome assemblies of 19 turtles plus three outgroups, aligned and phylogenetically analyzed using multilocus coalescent-consistent and concatenated partition-based methods. The second data set (26 species, including all turtle species from the first method) consisted of 1,655,675 biallelic parsimony-informative sites that we extracted using mapped Illumina sequence reads. We assembled the reads into contigs that were further mapped to a reference genome (the western painted turtle, *Chrysemys picta bellii*). Based on genome annotations, we partitioned sites by locus type including coding regions (CDS), introns, 5′-UTR, 3′-UTR, intergenic, pseudogenes, lncRNA, and smRNA, and reconstructed separate phylogenies using each locus type. With the first data set we assessed patterns and sources of heterogeneity in turtle phylogenomics using a large number of DNA sequence alignments, whereas with the second data set we compared topologies reconstructed with parsimony-informative sites extracted from different annotated regions across the turtle genome. Taken together, this scope of analyses is unprecedented in turtle phylogenomics. We confirm that heterogeneity in phylogenomic data sets is to be expected, even in well-resolved clades such as turtles, and suggest that a combination of processes is driving the incongruence between previous studies of turtle relationships.

Materials and Methods

Phylogenomic Analysis of High-Confidence Single-Copy Orthologs

We downloaded the complete genome assemblies of 19 turtle species plus three outgroup taxa from NCBI. The species, assembly accession numbers, assembly lengths, and assembly contiguities measured by scaffold N50 are shown in Table 1. Using BUSCO v3 (Waterhouse et al. 2018), we extracted the nucleotide sequences of 5310 OrthoDB v9 tetrapod orthologs (Zdobnov et al. 2017) from each genome assembly. BUSCO orthologs are high-confidence genes that persist in eukaryotic genomes as single copy, which precludes downstream problems in

TABLE 1. Genomic data for turtles analyzed in this study

Suborder	Superfamily/ Clade	Family	Species name	Assembly accession	Assembly length (Gb)	Scaffold N50 (Mb)	Trimmed read depth[a]
Cryptodira	Chelonioidea	Cheloniidae	*Chelonia mydas*	GCF_000344595.1	2.2	3.9	12.8×
		Dermochelyidae	*Dermochelys coriacea*	GCA_006547105.1	2.1	0.12	31.3×
	Chelydridae	Chelydridae	*Chelydra serpentina*	GCA_007922165.1	2.5	21	16.1×
	Kinosternoidea	Dermatemydidae	*Dermatemys mawii*	GCA_007922305.1	1.9	34.4	39.7×
		Kinosternidae	*Sternotherus carinatus*	NA	NA	NA	45.9×
	Testudinoidea	Emydidae	*Chrysemys picta*	GCA_000241765.3	2.4	16	10.3×
			Malaclemys terrapin	GCA_001728815.2	2.4	0.44	9.3×
			Terrapene mexicana	GCF_002925995.2	2.6	24.2	12.4×
			Emys orbicularis	NA	NA	NA	1.3×
			Trachemys scripta	NA	NA	NA	40.2×
			Actinemys marmorata	NA	NA	NA	41.4×
		Geoemydidae	*Cuora amboinensis*	GCA_004028625.2	2.1	0.25	23.1×
			Cuora mccordi	GCA_003846335.1	2.4	32.6	70.6×
			Mauremys reevesii	NA	NA	NA	34.1×
			Batagur trivittata	NA	NA	NA	1.1×
		Platysternidae	*Platysternon megacephalum*	GCA_003942145.1	2.3	7.2	16.7×
		Testudinidae	*Chelonoidis abingdonii*	GCA_003597395.1	2.3	1.3	17.3×
			Gopherus agassizii	GCA_002896415.1	2.2	0.228	40.5×
			Aldabrachelys gigantea	NA	NA	NA	21.9×
	Trionychoidea	Carettochelyidae	*Carettochelys insculpta*	GCA_007922185.1	2.3	45.9	49.7×
		Trionychidae	*Apalone spinifera*	GCA_000385615.1	1.9	2.3	10.3×
			Pelodiscus sinensis	GCF_000230535.1	2.2	3.4	21.1×
Pleurodira	Chelidoidea	Chelidae	*Emydura subglobosa*	GCA_007922225.1	2	44.8	40.6×
			Mesoclemmys tuberculata	GCA_007922155.1	2	46.4	49.0×
	Pelomedusoidea	Pelomedusidae	*Pelusios castaneus*	GCA_007922175.1	2	14.1	47.5×
		Podocnemididae	*Podocnemis expansa*	GCA_007922195.1	2.4	37.1	10.5×

Gb = gigabase; Mb = megabase; NA = no genome assembly was analyzed in the current study.
[a]Short Read Archive Accession numbers are in Supplementary material available on Dryad.

phylogenetics stemming from gene duplication such as hidden paralogy (Waterhouse et al. 2018). We aligned each set of orthologs with MAFFT v7.475 (Katoh and Standley 2013), and removed erroneous columns in the alignments with the heuristic method (*automated 1*) in TrimAL v1.4.1 (Capella-Gutierrez et al. 2009). Summary statistics including number of taxa, alignment length, missing data percentage, number and proportion of variable sites, number and proportion of parsimony-informative sites, and GC content for each alignment were estimated with AMAS v3.04 (Borowiec 2016). Trimmed alignments were filtered at a cut-off length ≥1500 bp and a minimum of 50% taxa representation. We estimated the average pairwise Kimura 2-parameter corrected distance for each alignment with MEGAX (Kumar et al. 2018).

Gene trees were reconstructed for each locus using RAxML v1.8 (Stamatakis 2014). For each ortholog, we generated 10 maximum likelihood trees with the GTRGAMMA substitution model, and performed 100 rapid bootstrap replicates on the best tree. We then collected the best tree for each ortholog to use as evidence to construct a species tree with ASTRAL-III (Zhang et al. 2018), assessing branch support with local posterior probabilities (Sayyari and Mirarab 2016). We used Snakemake (Mölder et al. 2021) to automate a reproducible bioinformatics pipeline for BUSCO extraction, filtering, alignment, gene tree and species tree estimation, available at https://github.com/mibyars/busco_phylo_pipeline. We also concatenated the single-copy orthologs into a supermatrix with AMAS and performed a partitioned maximum likelihood phylogenetic reconstruction with IQ-TREE v2.1.2 (Minh, Schmidt et al. 2020). We used ModelFinder (Kalyaanamoorthy et al. 2017) set to TEST-MERGE to combine partitions where necessary. The best-fit partition scheme for the concatenated supermatrix included 279 partitions. We assessed branch support on the concatenated partitioned maximum likelihood tree using 10,000 ultrafast bootstrap replicates and the Shimodaira-Hasegawa-like (SH-like) approximate likelihood ratio test (aLRT, Shimodaira 2002).

Phylogenomic Analysis of Biallelic Markers

In addition to single-copy orthologs from turtle and outgroup genome assemblies, we constructed sets of biallelic single nucleotide polymorphisms (SNPs) using paired-end sequence data from 26 turtle species downloaded from the NCBI Short Read Archive (SRA, Table 1, Supplementary material available on Dryad at https://doi.org/10.5061/dryad.tdz08kq14). We assessed each SRA data set with FastQC v0.115 (S. Andrews—http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed the raw reads with BBDuk v.37.41 (B. Bushnell—sourceforge.net/projects/bbmap/, last accessed December 2020). We used Site Identification from Short Read Sequences (SISRS) to generate de novo orthologs from the processed reads (Schwartz et al. 2015; Literman and Schwartz 2021).

After subsampling all sequences to a target depth of ~10× per taxon (except for two low-coverage species, Table 1, Supplementary material available on Dryad), we assembled a composite genome using Ray v2.2.3-devel (Boisvert et al. 2010) with default parameters and a *k* value of 31. We then used SISRS to map the trimmed reads from each species to the composite genome, retaining only uniquely mapped reads to avoid false positives from duplicated or repetitive regions. Bases in the composite genome were replaced according to mapped data from each species if sites were covered by at least three reads and if there was no within-taxon variation; sites not meeting these criteria were masked as "N."

For downstream ortholog annotation, we mapped each scaffold from the composite sequences against the reference genome for the western painted turtle (*C. picta bellii*, GCA_000241765.3) with Bowtie2 v2.3.4 (Langmead and Salzberg 2012), removing contigs that either failed to map or mapped to multiple positions, and removing sites resulting from overlapping coverage of independent scaffolds. We then annotated the mapped composite genome sites parsed from the painted turtle genome as belonging to the following genomic regions (locus types) with BEDTools v2.2.6 (Quinlan and Hall 2010): (1) CDS; (2) 3′-UTR; (3) 5′-UTR; (4) introns; (5) other genic regions not annotated as CDS, UTR, or intronic; (6) long-noncoding RNAs (lncRNAs); (7) pseudogenes; (8) smRNAs (including miRNAs, ncRNAs, rRNAs, scRNAs, smRNAs, snoRNAs, snRNAs, tRNAs, and vaultRNAs); and (9) intergenic regions. We defined "intergenic" as all regions not included within the other locus types.

We used SISRS to identify parsimony-informative sites along the mapped composite contigs and to remove invariant sites and singletons, as well as any sites containing "Ns" (i.e., polymorphic within species) and overlapping indels or gaps. For the nine annotated genomic regions as well as a set of the complete mapped data (=10 locus types), we concatenated biallelic parsimony-informative sites and inferred locus type phylogenies via SVDquartets (Chifman and Kubatko 2014) in PAUP* v4.0a (build 168) (Swofford 2003) with exhaustive quartet sampling using 1000 bootstrap replicates (parameters: evalq = all bootstrap = standard) and estimated majority consensus trees.

Heterogeneity Analyses

To measure topological conflicts across genealogies in terms of their dissimilarity to the species tree, we calculated Robinson–Foulds (RF) distances (Robinson and Foulds 1981) with PhyKIT (Steenwyk et al. 2021). First, we measured the RF distance between each inferred single-copy ortholog gene tree and the inferred species tree. In order to take into account missing species across some of the gene trees, we normalized the RF distances by taking the raw RF distance and dividing it by $2(n-3)$ where n is the number of tips in the phylogeny. We also calculated multiple pairwise RF distances between the trees inferred from SISRS data mapped to each genome annotation (CDS, introns, intergenic, etc.). Finally, we calculated the RF distances of each SISRS phylogeny from the topology inferred with all mapped sites across the turtle genome.

We computed gene and site concordance factors for each branch in the turtle phylogeny using the trimmed and length-filtered single-copy ortholog alignments, their gene trees, and the ASTRAL tree with IQ-TREE (Minh, Hahn et al. 2020). A branch's gene concordance factor represents the percentage of decisive gene trees that also contain that branch, and a branch's site concordance factor is the percentage of decisive alignment sites supporting that branch. We also computed gene concordance factor and site concordance factor for each branch in the concatenated single-copy ortholog tree by computing individual locus trees based on the partitions in IQ-TREE. Locus trees were inferred with default settings, and the species tree was inferred using an edge-linked proportional partition model with an SH-like aLRT and ultrafast bootstrap analysis of 10,000 replicates each (-aLRT 10,000; -B 10,000). The greedy algorithm of PartitionFinder (Lanfear et al. 2012) was applied to find the best-fit partitioning scheme which was then used in the subsequent tree reconstruction step (-m TESTMERGE). We also calculated site concordance factors for every branch in the SVDquartets trees inferred from biallelic parsimony-informative sites within locus types.

Discordance between gene trees and species trees can be caused by introgression, incomplete lineage sorting, erroneously inferred gene trees stemming from model violations and/or other artifacts, or combinations of all of these factors, and we aimed to determine how these issues affect phylogenomic inference in turtles. One expectation of incomplete lineage sorting is a roughly equal proportion of topologies supporting alternative relationships, whereas introgression would cause some minor gene trees to be more frequent than others, reflecting the direction of gene flow (Huson et al. 2005; Green et al. 2010). We collected the two most common minor topologies and each node (gDF1 and gDF2 resulting from the concordance factor analysis in IQ-TREE) and used a chi-squared test to determine if they were similar in terms of their proportions of all gene trees, where a rejection of the null hypothesis was interpreted as the presence of introgression in addition to incomplete lineage sorting and potential model violations. Nodes where we failed to reject the null hypothesis were interpreted as divergence patterns in turtle evolution driven primarily by incomplete lineage sorting or model violations, or a combination of both.

Analysis of Potential Model Violations Via Substitutional Saturation, Compositional Heterogeneity, and Codon Usage Bias

Phylogenetic signal can be obscured by the accumulation of multiple substitutions at the same site

over time, leading to model violations and increased gene tree-species tree discordance, particularly at deep evolutionary timescales (Jeffroy et al. 2006; Philippe et al. 2011). Therefore, to further distinguish between the effects of incomplete lineage sorting and model violations at nodes in the turtle species tree, we estimated the levels of substitutional saturation in the single-copy ortholog data set. First, we calculated the slope of a linear regression between the computed raw pairwise distances and corrected pairwise distances under a TN93 substitution model for 685 single-copy ortholog alignments with complete taxon sampling using the ape 5.5 R package (Paradis and Schliep 2019). For genes with high levels of saturation, the slope of this regression will be closer to 0 whereas less saturated genes will have a slope approaching 1 (Philippe et al. 1994). We then plotted a histogram of all slopes and characterized each alignment as having slopes above ("unsaturated") and below ("saturated") the mean. To determine the effect of substitutional saturation on phylogenomic inference in turtles, we compared ASTRAL species trees inferred from the RAxML trees inferred from unsaturated and saturated genes.

To analyze substitutional saturation at codon positions and assess their effects on phylogenomic inference in turtles, we realigned the 685 single-copy orthologs with complete taxon sampling using MACSE (Ranwez et al. 2011), and concatenated alignments that were partitioned based on (1) first and second codon positions and (2) third codon positions with AMAS. We then computed the corrected and uncorrected pairwise distances and estimated the regression slopes for the two codon-based partitions in R as above. We also further separated these codon partitions according to whether they belonged to the unsaturated or saturated gene sets, inferred maximum likelihood phylogenies on the four resulting data sets in IQ-TREE with model testing and 1000 ultrafast bootstrap replicates, and calculated site concordance factors for each inferred tree from codon partitions. Finally, to account for substitutional saturation we computed a maximum likelihood tree from the concatenated and partitioned MACSE amino acid alignments of the 685 single-copy orthologs using model testing in IQ-TREE. Maximum likelihood tree topologies, branch support, and concordance factors from codon-partitioned saturated and unsaturated alignments as well as the amino acids were compared with the ASTRAL and concatenated supermatrix trees based on single-copy orthologs $\geq$1500 bp.

A chi-squared test for lineage-specific compositional heterogeneity was computed before each IQ-TREE run, where the frequency of each nucleotide in a given taxon was compared with the overall frequency of each nucleotide in the entire data set. We compared the results of this test across IQ-TREE analyses of the saturated and unsaturated concatenated partitions based on codon position. To analyze codon usage bias in our single-copy ortholog data set, we estimated how often a codon was observed relative to its predicted occurrence in the absence of codon usage bias, by computing the relative synonymous codon usage (RSCU) of the concatenated 685 complete-taxon MACSE alignments using CodonW v1.4.4 (Peden 1999) followed by factor analysis in FactoMineR (Lê et al. 2008) in R.

Divergence Time Estimation

We estimated divergence times with single-copy ortholog alignments using Bayesian modeling implemented in BEAST v.2.6.3 (Bouckaert et al. 2019). We subsampled 685 alignments with complete taxon sampling, from which we further randomly sampled three replicate sets of 10 genes. Replicate sets were partitioned by gene, and based on model testing with ModelFinder, we applied the HKY + Gamma + Invariant site model to all partitions, and utilized a relaxed lognormal molecular clock and Calibrated Yule tree prior. To calibrate the time trees, we used 11 calibration priors from the literature (Joyce et al. 2013) which overlapped with our taxon sampling (Supplementary material available on Dryad). To account for uncertainty in fossil dating, we used hard minimum constraints and set soft maximum constraints by placing the maximum ages in the 97.5% quantile of a lognormal prior distribution. Markov chain Monte Carlo (MCMC) analyses were sampled every 50,000 generations. For each replicate, we assessed convergence of parameter estimates across the MCMC by monitoring effective sample sizes (i.e., ESS value $\geq$200) with Tracer v.1.7.2 (Rambaut et al. 2018), and we estimated a maximum clade credibility tree with TreeAnnotator v2.6.3.

We also estimated divergence times in turtle evolution using penalized likelihood with r8s v1.8.1 (Sanderson 1997, 2002, 2003), using the same fossil calibrations as above for minimum and maximum node ages; the main difference in this analysis was we fixed the time to most recent common ancestor (TMRCA) for amniotes at 312 million years ago (Ma) (Donoghue and Benton 2007) and the TMRCA for Testudines at 220 Ma (Thomson et al. 2021). First, we concatenated the 685 complete partitions and inferred a maximum likelihood tree with model testing for each partition as above in IQ-TREE. We then used the inferred phylogeny with branches in terms of substitutions per site and the 1,836,182 concatenated bases for penalized likelihood estimation, with cross validation to optimize the smoothing parameter which quantifies the deviation from the molecular clock.

We used linear regression to determine the relationship between the estimated divergence time for each node and the concordance factors computed with the inferred trees from each locus type in the SISRS analysis, the gene trees and the species tree based on single-copy orthologs, and the locus trees and partitioned maximum likelihood tree based on single-copy orthologs. We also tested for a correlation between the concordance factors and the estimated rates of molecular evolution at each branch. We repeated these regressions using estimated divergence times from Thomson et al. (2021).

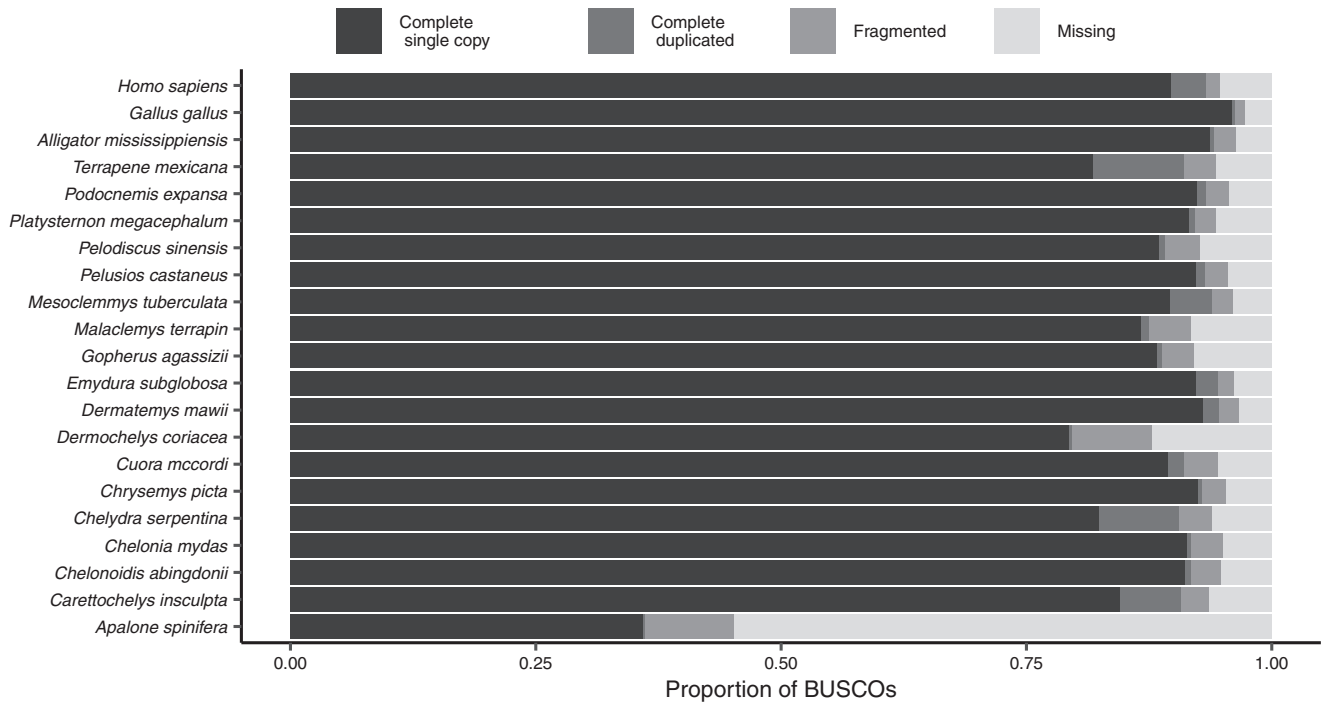

FIGURE 1. Summary of single-copy orthologs. Percent of tetrapod orthologs from orthoDBv9 present in each genome assembly as complete single copy, complete duplicated, fragmented, or missing.

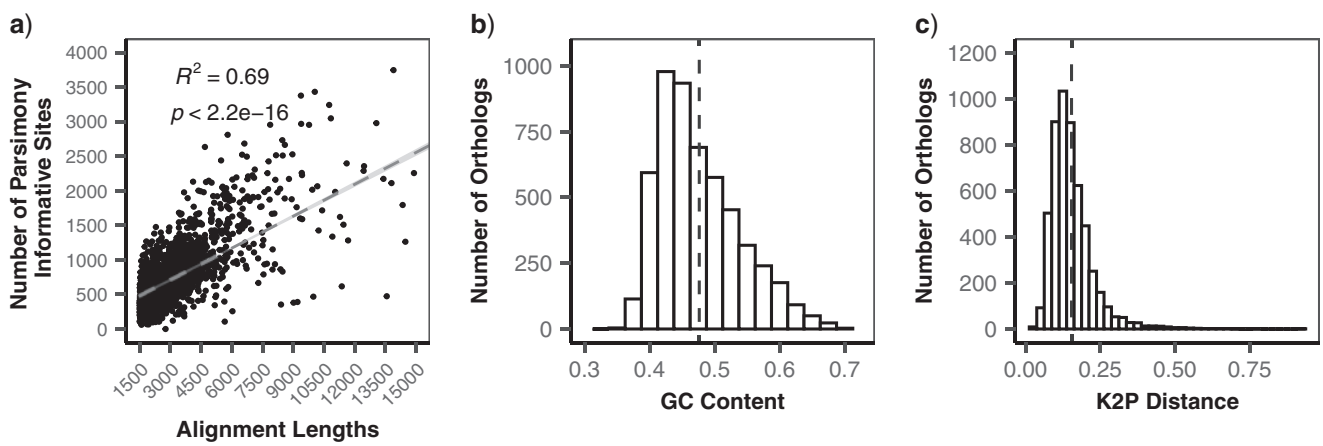

FIGURE 2. Description of single-copy ortholog alignments. a) Scatterplot showing the relationship between alignment length and the number of parsimony-informative sites for trimmed alignments ≥1500 bp. b) Histogram showing the distribution of GC content in all single-copy orthologs after trimming. Solid vertical line represents the mean value (0.476). c) Histogram showing the distribution of average pairwise Kimura 2-parameter (K2P) distances for single-copy orthologs after trimming. Solid vertical line represents the mean value (0.154).

RESULTS

A Very Large Number of Informative Sites for Turtle Phylogenomics

We analyzed publicly available genome assemblies to extract single-copy orthologs from 19 turtles representing 13 out of 14 extant families and 7 of 8 superfamilies, plus 3 outgroup taxa (Table 1). On average, 4609 (83%, range 1911–5099) conserved tetrapod orthologs were complete in the genome assemblies, 112 (2.1%, range 7–496) were duplicated, 170 (3.2%, range 55–485) were fragmented, and 419 (7.9%, range 143–2907) were missing (Fig. 1). Fifty-five percent of conserved tetrapod orthologs were missing from the genome assembly of the trionychid *Apalone spinifera*; therefore, we omitted this species from downstream analyses of single-copy orthologs due to concerns about assembly quality (Waterhouse et al. 2018). After this, the analyzed single-copy ortholog data set contained 21 total taxa (18 turtle species and three outgroups). Each single-copy ortholog was present in an average of 19 out of 21 species, averaged 1902 bp in length, and contained an average of 53% variable sites and 26% parsimony-informative sites. Alignment length was correlated with the number

TABLE 2. Summary of short read data from 26 turtle species mapped to the western painted turtle (*Chrysemys picta*) genome annotations

Locus type	Annotated bases in the genome	Bases mapped by SISRS	Number of biallelic parsimony-informative sites[a]	Percent of annotated bases mapped
All mapped	2,365,766,571	302,355,876	1,655,675	12.78
Intergenic	1,382,145,112	161,445,195	746,724	11.68
Intronic	812,010,672	115,297,781	533,336	14.20
lncRNA	64,450,450	9,451,244	52,207	14.66
Genic other	65,764,053	9,052,113	36,604	13.76
CDS	33,486,790	5,718,074	136,823	17.08
3′-UTR	8,034,514	1,510,783	15,150	18.80
5′-UTR	5,096,255	708,214	6519	13.90
Pseudogenic	670,023	59,058	427	8.81
smRNA	133,894	20,076	137	14.99

[a]When allowing up to two missing species per site, see Supplementary material available on Dryad for numbers of sites with varying levels of missing data.

of parsimony-informative sites ($R^2 = 0.69$, $P = 2.2e\text{-}16$, Fig. 2a). GC content was consistent across the single-copy orthologs and averaged 48% (standard deviation 6.2%, Fig. 2b), as was average pairwise Kimura 2-parameter distance (mean 0.154, median 0.139, standard deviation 0.072, Fig. 2c). The subset of 2513 alignments ≥1500 bp in length had very similar characteristics to the complete set of loci; the average large locus length was 1903 bp and the relationship between alignment length and the number of parsimony-informative sites held ($R^2 = 0.73, P = 2.2e\text{-}16$).

We downloaded 38 SRA data sets from 26 turtle species representing 7 of 8 superfamilies and 14 families (Table 1; Supplementary material available on Dryad), including the 19 species with complete genome assemblies from the first data set. Based on FastQC results, we selected 32 of the SRA data sets for further analysis ranging from ∼1 to 70× posttrimming coverage. The composite genome consisted of 5,555,666 contigs with an N50 of 2.4 Mb. We successfully mapped 302,355,876 bases to 2,110,354 contigs in the composite genome which covered 80% of the painted turtle genome (Table 2), and called 1,655,675 parsimony-informative biallelic SNPs when allowing missing data for up to two species at a given site. The number of parsimony-informative biallelic SNPs was much larger in mapped data sets containing only cryptodires (2,787,072 bases) and testudinoideans (4,696,637 bases) (Supplementary material available on Dryad), as expected based on Literman and Schwartz (2021). We collected the highest number of biallelic SNPs from intergenic regions (746,724 bases), followed by introns (533,336 bases) and CDS (136,823 bases). The relatively low number of parsimony-informative biallelic SNPs in pseudogenic (427) and smRNA regions (137) is likely due to a small number of annotations for these features in the western painted turtle genome.

Bioinformatic Analysis of Turtle Genomes Yielded Well-Resolved Species Relationships

The inferred relationships among turtle lineages based on single-copy orthologs were consistent with previous studies using phylogenomic markers (Crawford et al. 2015; Shaffer et al. 2017), with full statistical support for all resulting branches based on posterior probabilities in the ASTRAL tree, and bootstrap replicates and SH-like approximate likelihood in the supermatrix tree (Fig. 3, Figs. S1 and S2 of the Supplementary material available on Dryad). Hereafter we refer to these results collectively as the turtle species tree. The turtle species tree supports the reciprocal monophyly of Pleurodira and Cryptodira (Fig. 4). Within cryptodires, there is a split between the Trionychoidea (soft-shelled turtles) and the Durocryptodira (Crawford et al. 2015), and a further subdivision of Durocryptodira into the Testudinoidea (pond turtles, big-headed turtles, geoemyds, and tortoises) and Americhelydia (sea turtles, snapping turtles, mud turtles, and *Dermatemys*; Joyce et al. 2013). Within the Testudinoidea, the turtle species tree supports Testuguria (geoemyds and tortoises; Joyce et al. 2004), and a *Platysternon* + Emydidae clade ("Platyemidae"; Barley et al. 2010; Crawford et al. 2015).

SVDquartets trees from the genome-wide, all-mapped data as well as from CDS, introns, and intergenic regions were consistent with the species tree based on single-copy orthologs (Fig. 4, Figs. S3–S11 of the Supplementary material available on Dryad). When there was disagreement between trees from mapped regions and the species tree, the difference was often in the placement of *Platysternon* (Fig. 5). For instance, when allowing zero missing gapless parsimony-informative sites from the all-turtle data, 5′ UTR placed *Platysternon* as the outgroup of Geoemydidae + Testudinidae (435 sites). When allowing up to two missing gapless parsimony-informative sites from the all-turtle data, pseudogenic regions (426 sites) also placed *Platysternon* as the outgroup of Geoemydidae + Testudinidae, whereas smRNA (137 sites) agreed with the species tree and all-mapped topologies. Meanwhile, at multiple levels of allowed missing data the Testudinoidea-only mapped sites weakly supported *Platysternon* as the sister taxon to Geoemydidae + Testudinidae + Emydidae (Supplementary material available on Dryad).

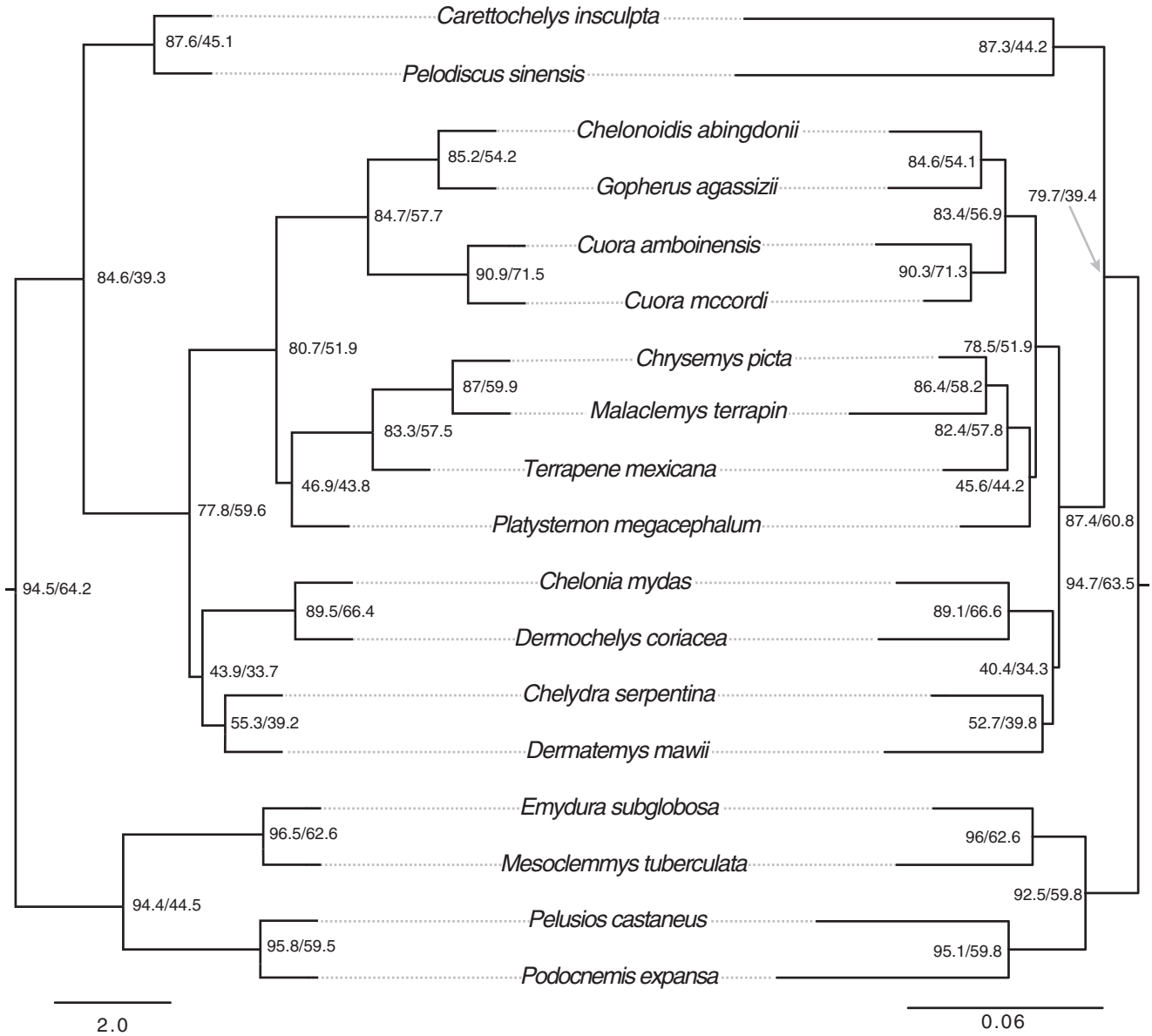

FIGURE 3. Species trees for turtles based on high-confidence single-copy orthologs. We used 2513 single-copy orthologs ≥1500 bp in length to infer a multilocus coalescent-consistent tree based on gene genealogies (ASTRAL, left) and a maximum likelihood tree (ML, right) from a concatenated and partitioned supermatrix in IQ-TREE. All branches received 100% support from local posterior probabilities (ASTRAL), 1000 bootstrap replicates (IQ-TREE), and approximate likelihood ratio test (IQ-TREE). Node labels indicate gene and site concordance factors before and after the slash, respectively. Branch lengths are in terms of coalescent units (ASTRAL) and substitutions per site (ML). Outgroups (*Gallus, Alligator, Homo*) are not shown.

Heterogeneity is Prevalent at Well Resolved Nodes in the Turtle Phylogeny

Although results from our single-copy orthologs and biallelic sites analyses of turtles are consistent and largely in agreement with other studies based on 100% branch support (Crawford et al. 2015; Shaffer et al. 2017), some genomic regions, either from single-copy orthologs or from genome feature annotations, yielded genealogies conflicting with the turtle species tree. For instance, based on normalized RF distances most gene trees from single-copy orthologs were similar to the species tree, although there was a long tail to this distribution indicating gene trees with greater dissimilarity (Fig. 6a). SVDquartets trees using mapped data from CDS, introns, and intergenic regions were consistent with both the single-copy ortholog species tree and a SVDquartets tree derived from the all-mapped data. In contrast, mapped pseudogenic DNA, 5′ UTR, and smRNA deviated from the all-mapped data in terms of their RF distances (Fig. S12 of the Supplementary material available on Dryad).

We also found a high degree of heterogeneity based on gene and site concordance factors throughout the

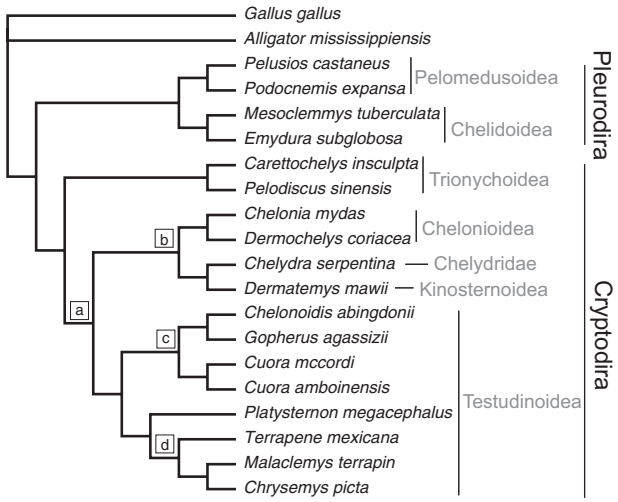

FIGURE 4. The phylogenetic relationships of higher turtle taxa. Cladogram depicting the turtle species tree in agreement with results from ASTRAL and concatenated partitioned maximum likelihood analyses as well as most SVDquartets analyses using the mapped data (although these analyses lacked outgroups). Superfamilies or clades are indicated by horizontal shaded text. Additional key clades are labeled with letters in boxes. a) Durocryptodira, b) Americhelydia, c) Testuguria, and d) Emydidae. Additional outgroup *Homo sapiens* not shown.

turtle phylogeny (Figs. 3 and 6, Figs. S1 and S2 of the Supplementary material available on Dryad). For instance, although local posterior probabilities were uniformly 100% across branches of the ASTRAL tree, gene concordance factors ranged from 44 to 97 and site concordance factors ranged from 30 to 71 (Fig. 6b–d). Gene and site concordance factors based on the ASTRAL tree were correlated with each other (Fig. 6b; $R^2 = 0.55$, $P = 0.0002$) and with branch lengths (Fig. 5c and d; $R^2 = 0.88$, 3.88e-09 and $R^2 = 0.49$, $P = 0.0007$, respectively). We also estimated gene and site concordance factors based on the locus trees with respect to the maximum likelihood tree inferred from the concatenated supermatrix. A similar range of gene concordance factors (34–92) and site concordance factors (30–71) was found for these locus trees, despite 100% bootstrap support, as well as a strong correlation between the two concordance measures ($R^2 = 0.60$, $P = 0.0001$). Gene and site concordance factors based on the concatenated partitioned maximum likelihood tree were also correlated with branch lengths ($R^2 = 0.90$, 1.02e-09 and $R^2 = 0.58$, $P = 0.0001$, respectively).

The ancestral branch leading to turtles was assigned a relatively large gene concordance factor and site concordance factor across analyses based on the ASTRAL, supermatrix, and SVDquartets trees, and there was high concordance for reciprocally monophyletic Pleurodira and Cryptodira across genes and sites (Figs. 3 and 4, Figs. S3–S10 of the Supplementary material available on Dryad). The placement of Trionychoidea as sister taxon to all other cryptodires was assigned a relatively large gene concordance factor but small site concordance factor, suggesting that there are a large number of sites supporting alternative bifurcations. One area with particularly high discordance (i.e., relatively small gene concordance factor and site concordance factor across all data sets) was the placement of the big-headed turtle (*Platysternon*) with respect to Emydidae. Our chi-squared test rejected introgression at this node (Table 3), favoring incomplete lineage sorting or potential model violations as sources of discordance in the position of *Platysternon*. There was also a high degree of discordance associated with the americhelydian taxa (Chelonioidea + Chelydridae + Kinosternoidea), particularly with the ancestral branch leading to Americhelydia as well as with the sister taxon relationship between *Dermatemys* (Kinosternoidea) and *Chelydra* (Chelydridae). Based on single-copy orthologs, minor gene tree frequencies were significantly different at these nodes (Table 4), suggesting that introgression in addition to incomplete lineage sorting and/or model violations may be driving phylogenomic discordance in these groups.

Patterns of Substitutional Saturation, Compositional Heterogeneity, and Codon Usage Bias in Turtle Phylogenomics

We measured the slope of the line in a regression between uncorrected and corrected genetic distances for 685 single-copy ortholog alignments with complete taxon sampling. The slopes ranged from $y = 0.47x$ (most saturated) to $y = 0.91x$ (least saturated), but were skewed toward low levels of saturation, with a mean of $y = 0.76x$, a first quartile of $y = 0.72x$, and a third quartile of $y = 0.81x$ (Fig. S13 of the Supplementary material available on Dryad). When analyzed by codon position, we found that first and second codon positions were unsaturated ($y = 0.86x$) whereas the slope of the line between corrected and uncorrected distances at third codon positions were consistent with saturation ($y = 0.57x$) (Fig. S14 of the Supplementary material available on Dryad). Saturation did not have an effect on the topology of the species tree, as ASTRAL trees derived from the most and least saturated sets of genes were identical (Supplementary material available on Dryad). In addition, all codon partitions regardless of their saturation levels produced maximum likelihood trees with intra-Testudines relationships that were identical in topology to the species tree, and the maximum likelihood tree estimated based on amino acid partitions was also identical in topology to the species tree (Supplementary material available on Dryad). There was some disagreement in the placement of turtles in the amniote phylogeny that appeared to be driven by substitutional saturation; for instance, the maximum likelihood trees based on first and second codon positions from saturated genes and all sets of third codon positions supported turtles as the sister taxon to crocodilians (Supplementary material available on Dryad; see Discussion).

We found that compositional heterogeneity was most prominent in the saturated third position and first and second position data sets, as well as in the saturated first

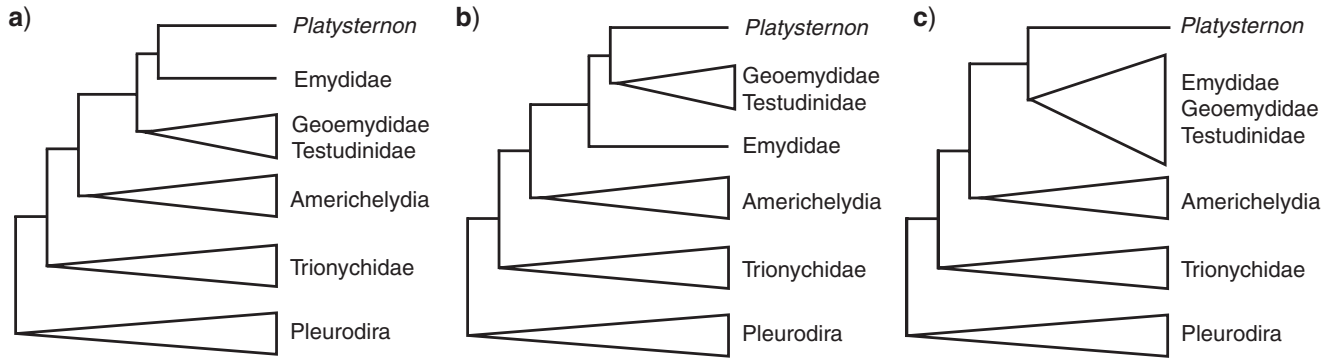

FIGURE 5. Alternative topologies in turtle phylogenetics based on markers from different regions of the genome. a) *Platysternon* as the sister taxon to Emydidae to the exclusion of Testuguria (Geoemydidae + Testudinidae), consistent with the turtle species tree (current study), Crawford et al. (2015), and Shaffer et al. (2017). b) *Platysternon* as the sister taxon to Testuguria with the exclusion of Emydidae, consistent with 17% of the single-copy ortholog genes trees, mapped 5′-UTR parsimony-informative sites data with zero missing data, mapped pseudogene parsimony-informative sites when allowing two missing taxa per site, and Krenz et al. (2005). c) *Platysternon* as the sister taxon to Geoemydidae + Testudinidae + Emydidae, consistent with 20% of single-copy ortholog gene trees and Shaffer et al. (1997).

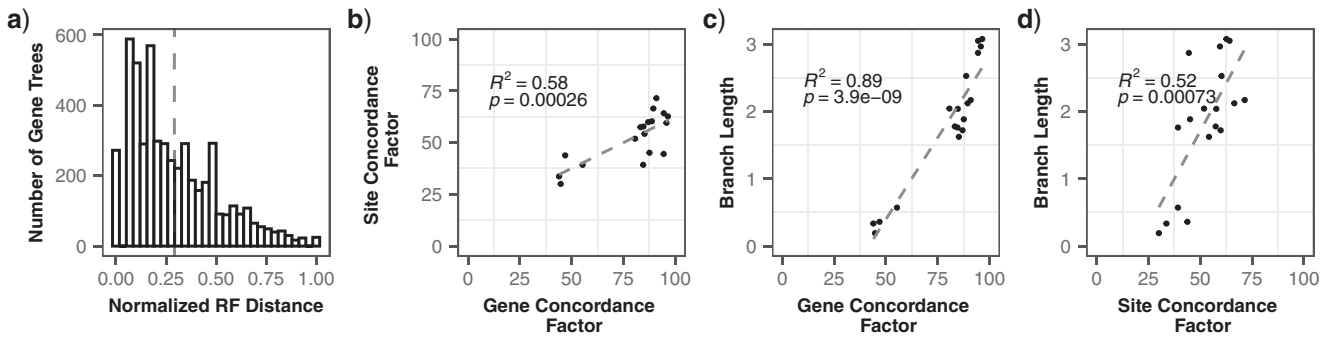

FIGURE 6. Phylogenomic heterogeneity in turtles. a) Histogram of RF distances between gene trees and species tree. Dashed vertical line represents the mean value (0.292). b, c) Concordance factors calculated using single-copy ortholog alignments, their gene trees, and the ASTRAL species tree. b) Scatterplot depicting the relationship between gene and site concordance factors at each node. c) Scatterplot depicting the relationship between gene concordance factor and branch length. d) Scatterplot depicting the relationship between site concordance factor and branch length. Results based on the concatenated supermatrix maximum likelihood tree are in Supplementary material available on Dryad.

TABLE 3. Gene concordance factors for clades represented by nodes in the ASTRAL species tree phylogeny where the minor gene tree frequencies were similar according to a chi-squared test

Clade description	Number of possible supporting gene trees	Gene concordance factor	Percent of gene trees supporting first alternative topology	Percent of gene trees supporting second alternative topology
Durocryptodira	2387	88.73	0.54	0.21
Chelonioidea	1755	89.46	1.48	0.8
Testudinoidea	2435	80.66	0.57	0.7
Testuguria	2255	84.7	0.62	0.8
Testudinidae	1989	85.22	4.17	5.38
Cuora	1932	90.94	2.69	2.07
Emydidae + Platysternidae	2270	46.92	17.05	19.52
Malachlemys + *Chrysemys*	1730	86.99	5.03	4.39
Pelomedusoidea	2195	95.81	1.32	1.14
Chelidoidea	2138	96.49	0.8	1.08

and second position data set (Supplementary material available on Dryad). The unsaturated alignment of first and second codon positions exhibited the least amount of compositional heterogeneity, with five species differing significantly from the rest of the alignment; however, these species were distributed randomly across the turtle species tree. Correspondence analyses of the RSCU revealed that some codons, particularly CCG (Proline), ACG (Threonine), and GCG (Alanine) were favored in different species (Fig. S15 of the Supplementary material available on Dryad). However, there was little evidence for a phylogenetic signal in this pattern within turtles

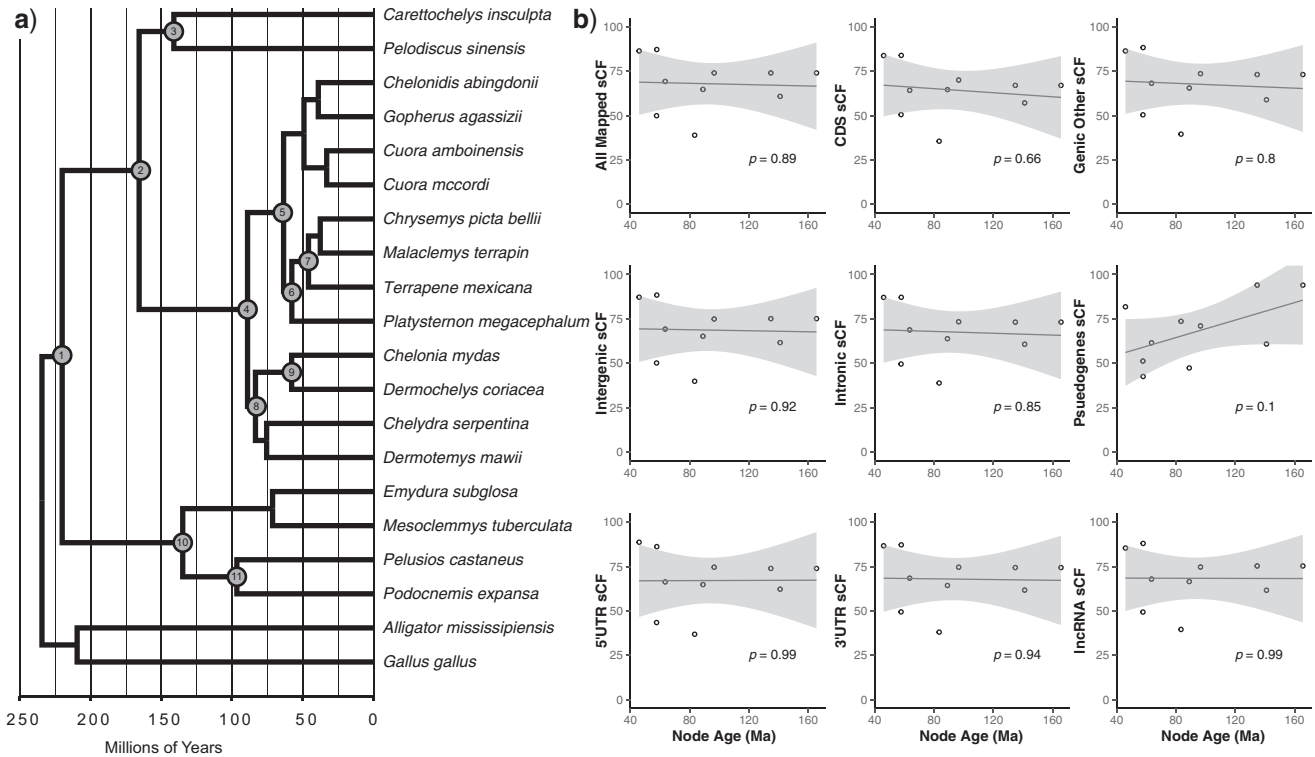

FIGURE 7. Site concordance and estimated node ages in turtle evolution. a) Chronogram showing estimated divergence times in millions of years based on penalized likelihood and fossil calibrations of labeled nodes. 1) Testudines, 2) Cryptodira, 3) Trionychoidea, 4) Durocryptodira, 5) Geoemydidae, 6) Platyemidae, 7) Emydidae, 8) Americhelydia, 9) Chelonioidea, 10) Pleurodira, and 11) Pelomedusoidea. b) Site concordance factors (sCFs) for each node in phylogenies inferred from nine locus types and the estimated node age in millions of years ago (Ma) based on penalized likelihood.

because codon usage bias was dispersed randomly among species, and the only outlier along the second dimension was an outgroup taxon (*Gallus*).

Heterogeneity at Different Timescales in Turtle Evolution

The estimated age of the root of the turtle species tree was consistently ~160 Ma (Middle Jurassic) across BEAST runs, which is much closer to the minimum constraint than Thomson et al. (2021) (Supplementary material available on Dryad), but similar to estimates from Chiari et al. (2012) and Pereira et al. (2017). Across locus types from the SISRS analyses, we found that site concordance factors were not strongly correlated with estimated node ages (Fig. 7, Fig. S16 of the Supplementary material available on Dryad). For most locus types the relationship between site concordance factor and divergence time at a given node trended slightly negatively, suggesting a pattern of higher concordance at younger nodes, although these comparisons were nonsignificant (Fig. 7b, $P > 0.1$). The one exception to this trend was for pseudogenes, for which the relationship between site concordance factors and estimated divergence times trended positive albeit without significance or with only moderate significance ($P = 0.072$ using Thomson et al.'s (2021) divergence time estimates). The relationship between site concordance factors and estimated DNA substitution rates trended slightly positive for all locus types from the SISRS analysis, but was nonsignificant in all comparisons (Fig. S17 of the Supplementary material available on Dryad).

We also found a nonsignificant negative trend in the relationship between site concordance factor and divergence time at a given node based on both the ASTRAL and supermatrix trees based on single-copy orthologs (Fig. 8a,b; Fig. S18 of the Supplementary material available on Dryad). However, we found a positive (albeit nonsignificant) trend in the relationship between the estimated divergence time at a given node and its gene concordance factor when using both trees derived from single-copy orthologs (Fig. 8d,e; Fig. S18 of the Supplementary material available on Dryad). There was no clear relationship between the site concordance factor of a given node and the estimated substitution rate at the preceding branch (Fig. 8c), and a moderately significant positive trend ($P \leq 0.05$) for the relationship between the gene concordance factor of a given node and the estimated substitution rate at the preceding branch (Fig. 8f; Fig. S18 of the Supplementary material available on Dryad).

DISCUSSION

Heterogeneity has been recognized as an important feature of phylogenomic data sets, and the exploration of

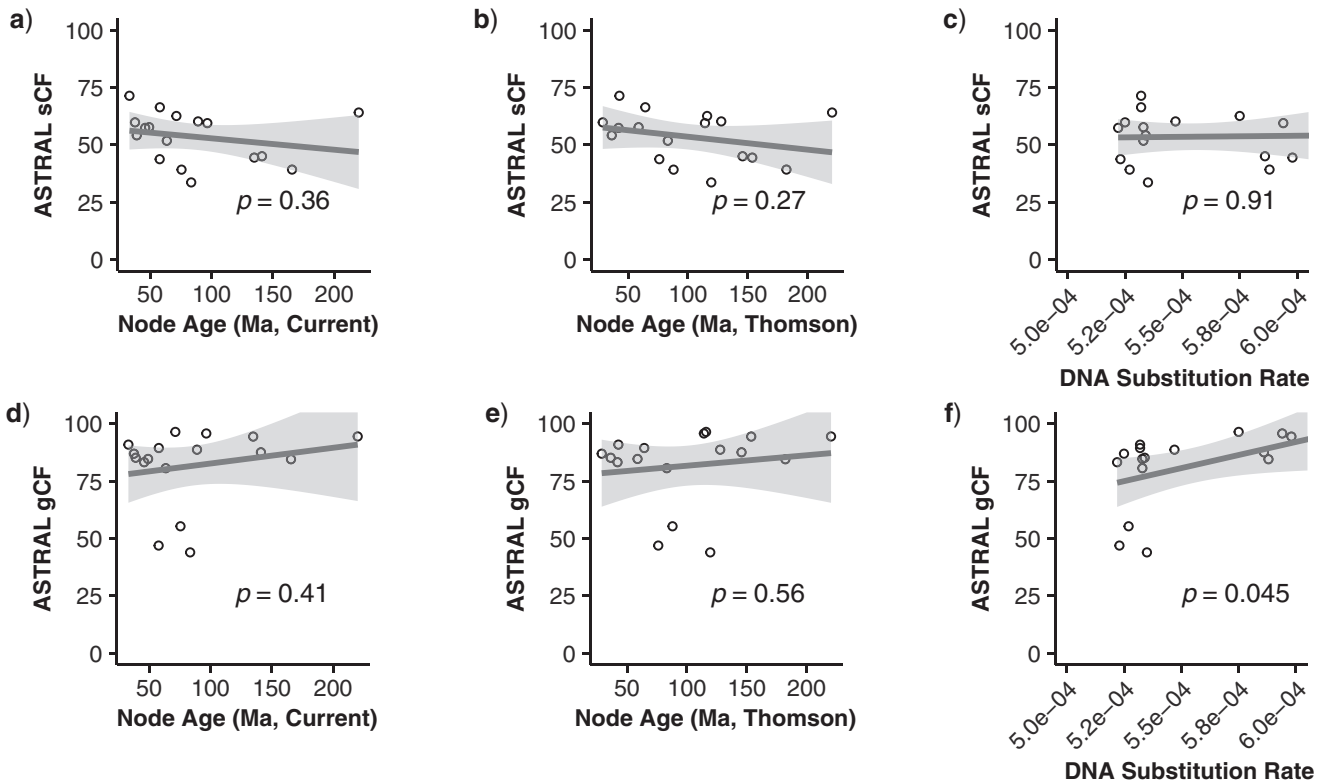

FIGURE 8. Concordance at genes and sites, estimated node ages, and DNA substitution rates in turtle evolution. Results shown are from comparisons of single-copy ortholog alignments and their gene trees to the species tree inferred using ASTRAL. There was a negative trend between site concordance factors (sCFs) and estimated node age (in millions of years ago or Ma) based on penalized likelihood (PL) from the current study (a) and the estimated node ages in Ma from Thomson et al. (2021) (b), and no discernible trend between sCF and estimated DNA substitution rates in terms of substitutions per site per million years (c). There was a positive trend between gene concordance factors (gCFs) and estimated node ages (d,e) and a moderately significant positive relationship between gene concordance factors and estimated DNA substitution rates (f). Results from comparisons of the locus trees and the maximum likelihood tree inferred from the concatenated supermatrix are shown in Supplementary material available on Dryad.

the patterns and sources of gene tree and site discordance has become an essential step when reconstructing the branching order of life with molecular data (Kumar et al. 2012; Minh, Hahn et al. 2020). To our knowledge, although the role of gene tree-species tree discordance in estimating the position of turtles within the amniote phylogeny has been the focus of a few studies (Chiari et al. 2012), there has not been an assessment of phylogenomic heterogeneity and its effects on the assessment of intraturtle relationships until now. Here, we have constructed a robust phylogenomic data set for turtles based on whole genomes in order to sample the major turtle lineages. We identified splits in the turtle phylogeny and regions of the turtle genome associated with gene tree and site heterogeneity, and investigated how potential sources of this discordance such as introgression, incomplete lineage sorting, and technical issues leading to model violations may have contributed to conflicting results in phylogenetic estimation of turtles. The results of our study demonstrate that heterogeneity can be rampant, even in phylogenomic data sets which have yielded fully resolved species trees.

Protein-coding genes such as the single-copy orthologs used in the current study are common markers used for phylogenetics, due to their straightforward amplification via polymerase chain reaction (PCR) or RNA-Seq, as well as relatively straightforward methods for their orthology assignment and alignment. However, the use of CDS in phylogenetics is complicated by difficulties with accurate evolutionary modeling due to multiple hits at ancient divergences, a diversity of substitution rates across the genome, and selective constraint potentially limiting the number of variant sites. These problems often scale with the size of phylogenomic data sets and can lead to biased results (Philippe et al. 2011). Therefore, in addition to the single-copy orthologs we also generated a complementary data set for turtle phylogenomics consisting of millions of parsimony-informative biallelic sites using SISRS. Employment of these sites has multiple advantages: (1) they are easily extracted from multiple sequence alignments; (2) they do not require complicated evolutionary modeling (i.e., they either support or do not support a splitting event); and (3) the sheer number of potential sites allows strict quality filtering without loss of signal (Literman and Schwartz 2021). SISRS resulted in a considerable number of sites collected from coding, noncoding, and regulatory parts of the genome, including intergenic regions which

yielded almost 750,000 parsimony-informative sites for turtles. These locus types sample a wide range of substitution rates, whereas many other phylogenomic methods focus on a much smaller range of locus types (such as UCEs, AHE, or coding genes). Our results show that many different regions of the turtle genome yield consistent support for the turtle species tree, leaving a solid phylogenetic hypothesis for higher turtle systematics. However, inferences from some regions such as pseudogenes and smRNA were based on a relatively small number of parsimony-informative sites, highlighting the need for more comprehensive annotations of noncoding portions of turtle genomes.

The relationships between turtles and other reptiles has been a longstanding debate in evolutionary biology, due to conflicting results between studies that use fossils and molecules (Benton 2005; Chiari et al. 2012; Brown and Thomson 2017). Paleontological studies have placed turtles either at the root of the amniote phylogeny, or as sister taxa with various other reptilian groups, some of which are long extinct (Gaffney 1980). In the genomic era, analyses of nucleotides and protein sequences have been mostly in agreement, placing turtles as the sister taxon to archosaurs (birds and crocodilians) (Shaffer et al. 2013; Wang et al. 2013). Based on 2513 long single-copy orthologs, our ASTRAL tree results supported a turtles + archosaurs clade, whereas the maximum likelihood tree based on the concatenated and partitioned supermatrix supported a turtles + crocodilians clade that excludes birds. Several lines of evidence suggest that the latter result is due to poorly fit evolutionary models. First, the maximum likelihood phylogeny inferred using partitioned amino acids, which accounts for saturation because proteins are less degenerate than nucleotides, yielded the turtles + archosaurs relationship. Second, we found that first and second codon positions from nonsaturated genes supported turtles + archosaurs, whereas first and second codon positions from saturated genes as well as all third codon positions agreed with turtles + crocodilians (Supplementary material available on Dryad). Brown and Thomson (2017) found that duplicated genes in the data set published by Chiari et al. (2012) data set were more likely to reject a turtles + archosaurs clade. Meanwhile, our reliance on single-copy orthologs from OrthoDB greatly reduces the role of paralogy in the conflict at this node, and paralogy might not be a concern for certain scenarios of phylogenetic inference (Smith and Hahn 2021). We suggest that the correct placement of turtles within the amniote phylogeny using genomic data has been hampered by substitutional saturation at such ancient divergences, and that when accounting for these artifacts the data support a turtles + archosaurs clade (Archelosauria; see Crawford et al. 2015).

When gene tree-species tree discordance is driven largely by technical errors such as saturation, it should be most apparent at older nodes in phylogeny, which are more affected by homoplasy, poor alignments, or model misspecifications. However, with the exception of the turtle-archosaur split, across both single-copy ortholog and SISRS data sets we found no significant correlation between concordance factors and estimated node ages. A slightly negative relationship between site concordance factors calculated from the single-copy ortholog data set and node age suggests that these deeper splits in turtle evolution might be affected by multiple substitutions; however, this pattern is obscured by a slightly positive relationship between gene concordance factors and node age using the same data. In addition, inferred turtle relationships were identical when analyzing different sets of genes separated by levels of substitutional saturation and partitioned by codon position. Thus, although it may be possible that multiple substitutions or other model violations are driving discordance at ancient nodes in the turtle phylogeny, their effects are either extremely weak or nonexistent within and among most turtle lineages, and the observed patterns may have more to do with errors in rate estimation using molecular data and fossil calibrations. Another potential source of discordance from model violations is intragenic recombination, particularly for long loci with more potential breakpoints, although simulations and empirical data sets show that species tree methods are robust to it (Edwards et al. 2016; Karin et al. 2020).

Our results suggest that population-level processes such as introgression and/or incomplete lineage sorting drive gene tree-species tree discordance at multiple stages of turtle diversification, leading to confusion in the literature about the placement of at least some taxa. In particular, the phylogenetic position of *Platysternon* has been highly contentious (Joyce et al. 2004, 2021), with early molecular studies in disagreement (Shaffer et al. 1997; Krenz et al. 2005; Parham et al. 2006; Barley et al. 2010). *P. megacephalum* is currently listed as Critically Endangered by the IUCN Red List, and most aspects of its biology and reproduction have not been documented until recently (Sung et al. 2014). Our turtle species tree results mirror Crawford et al. (2015) and Shaffer et al. (2017), fully supporting *Platysternon* as the sister taxon to Emydidae. However, we found that only 47% of single-copy ortholog gene trees are in agreement with a *Platysternon* + Emydidae clade, and 37% of the remaining gene trees support only two alternative relationships: of these, 17% of gene trees support *Platysternon* as the sister taxon to Testuguria (Geoemydidae + Testudinidae) following Krenz et al. (2005), and 20% support *Platysternon* as the sister taxon to Testudinoidea (Emydidae + Geoemydidae + Testudinidae) following Shaffer et al. (1997). The proportion of gene trees supporting these alternative topologies was similar according to our chi-squared test (Table 3), and when we analyzed gene trees according to their levels of substitutional saturation, we saw no effect on the placement of *Platysternon*. This suggests that biological factors such as incomplete lineage sorting are a previously unrecognized driver of conflict in the phylogenetic placement of *P. megacephalum*. The internal branch of the turtle species tree leading to *Platysternon* + (*Terrapene* + *Malaclemys* + *Chrysemys*) is short in estimated absolute time as well as ASTRAL

TABLE 4. Gene concordance factors for clades represented by nodes in the ASTRAL species tree phylogeny where the minor gene tree frequencies were statistically different according to a chi-squared test

Clade description	Number of possible supporting gene trees	Gene concordance factor	Percent of gene trees supporting first alternative topology	Percent of gene trees supporting second alternative topology
Archosauria	2144	44.73	10.21	43.38
Testudines	2254	94.5	0.35	1.24
Cryptodira	2390	84.56	1.63	5.98
Trionychoidea	1920	87.6	2.24	3.85
Americhelydia	2335	43.94	17.69	13.83
Kinosternoidea + Chelydridae	1970	55.33	14.01	9.95
Emydidae	1924	83.26	3.12	1.87
Pleurodira	2396	94.45	0.29	1.34

coalescent units, suggesting rapid speciation events and/or small effective population sizes (N_e) in the evolutionary history of *P. megacephalum*. These factors are thought to contribute to the anomaly zone, which is a problematic scenario in phylogenetic inference where some gene trees are more probable than the species tree (Liu and Edwards 2009; Linkem et al. 2016).

Another subject of conflict across previous phylogenetic studies of turtles is the relationship between Chelonioidea, Kinosternoidea, and the Chelydridae, all of which are now considered to be in the monophyletic Americhelydia (Barley et al. 2010; Joyce et al. 2013; Crawford et al. 2015). We found consistent support for Americhelydia across data sets and methods, as well as a sister taxon relationship between Chelydridae and Kinosternoidea, yet with considerable gene tree-species tree discordance with single-copy orthologs (Fig. 3) and site discordance with SISRS data (Figs. S3–S11 of the Supplementary material available on Dryad). For instance, only 44% of decisive gene trees from single-copy orthologs contained an Americhelydia clade, and only 55% of decisive gene trees supported Chelydridae + Kinosternoidea. As with *Platysternon*, comparison of trees when accounting for saturation or codon position did not affect the placement of americhelydians. The frequencies of minor gene trees disagreeing with these splits were statistically different (Table 4), suggesting both introgression and incomplete lineage sorting may have played a role in the diversification of these taxa. To our knowledge, the role of hybridization during the speciation events at the deeper splits of the turtle phylogeny has not been investigated, although contemporary hybridization between sea turtle genera with up to ~65 Ma of divergence is well documented (Karl et al. 1995; Arantes et al. 2020). As these occurrences are associated with low reproductive success (Arantes et al. 2020), it will be important to determine if episodic introgression is a feature of the long-term evolution of turtles.

Some recent molecular analyses of turtle relationships are limited in terms of the size of the molecular data set, yet have more complete taxonomic sampling than our study, including up to 85% of extant species (Pereira et al. 2017; Colston et al. 2020; Thomson et al. 2021). A denser sampling of lineages maximizes the molecular signal for testing hypotheses about past diversification events as well as the relationship between evolutionary distinctiveness and extinction risk. In contrast, our study is limited in terms of taxonomic sampling due to our reliance on complete genomic sequences and assemblies for turtles, which are lagging compared with other amniote groups. As our results demonstrate widespread discordance in turtle phylogenomic data sets, additional lineages should be targeted for genomic sequencing in order to improve our ability to accurately estimate the evolutionary distinctiveness of certain species. These studies will be crucial in setting future conservation priorities for turtles.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.tdz08kq14.

FUNDING

This work was supported by the National Science Foundation [1812291 to R.L.] and by startup funds provided to M.T. by the School of Informatics, Computing, and Cyber Systems at Northern Arizona University.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Monsoon computing cluster at Northern Arizona University (https://nau.edu/high-performance-computing/) for providing the computational resources necessary to carry out this study. They would also like to extend their thanks to the associate editor and two anonymous reviewers for suggestions which greatly improved our manuscript.

REFERENCES

Arantes L.S., Ferreira L.C.L., Driller M., Repinaldo Filho F.P.M., Mazzoni C.J., Santos F.R. 2020. Genomic evidence of recent

hybridization between sea turtles at Abrolhos Archipelago and its association to low reproductive output. Sci. Rep. 10:12847.
Barley A.J., Spinks P.Q., Thomson R.C., Shaffer H.B. 2010. Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life. Mol. Phylogenet. Evol. 55:1189–1194.
Benton M. 2005. Vertebrate paleontology. Oxford: Blackwell Science Ltd.
Boisvert S., Laviolette F., Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J. Comput. Biol. 17:1519–1533.
Borowiec M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 4:e1660.
Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., Maio N.D., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., Plessis L. du, Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 15:e1006650.
Brown J.M., Thomson R.C. 2017. Bayes factors unmask highly variable information content, bias and extreme influence in phylogenomic analyses. Syst. Biol. 66:517–530.
Burbrink F.T., Grazziotin F.G., Pyron R.A., Cundall D., Donnellan S., Irish F., Keogh J.S., Kraus F., Murphy R.W., Noonan B., Raxworthy C.J., Ruane S., Lemmon A.R., Lemmon E.M., Zaher H. 2020. Interrogating genomic-scale data for squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. Syst. Biol. 69:502–520.
Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. 2009. trimAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.
Chiari Y., Cahais V., Galtier N., Delsuc F. 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biol. 10:1–15.
Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30:3317–3324.
Cooper E.D. 2014. Overly simplistic substitution models obscure green plant phylogeny. Trends Plant Sci. 19:576–582.
Colston T.J., Kulkarni P., Jetz W., Pyron R.A. 2020. Phylogenetic and spatial distribution of evolutionary diversification, isolation, and threat in turtles and crocodilians (non-avian archosauromorphs). BMC Evol. Biol. 20:81.
Cox C.J., Li B., Foster P.G., Embley T.M., Civáò P. 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63:272–279.
Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., Glenn T.C. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8:783–786.
Crawford N.G., Parham J.F., Sellas A.B., Faircloth B.C., Glenn T.C., Papenfuss T.J., Henderson J.B., Hansen M.H., Simison W.B. 2015. A phylogenomic analysis of turtles. Mol. Phylogenet. Evol. 83:250–257.
Donoghue P.C.J., Benton M.J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol. Evol. 22:424–431.
Dudchenko O., Batra S.S., Omer A.D., Nyquist S.K., Hoeger M., Durand N.C., Shamim M.S., Machol I., Lander E.S., Aiden A.P., Aiden E.L. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95.
Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.
Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLOS One. 8:e65923.
Foster P.G. 2004. Modeling compositional heterogeneity. Syst. Biol. 53:485–495.
Galtier N., Daubin V. 2008. Dealing with incongruence in phylogenomic analyses. Philos. Trans. R. Soc. B Biol. Sci. 363:4023–4029.
Garland T., Bennett A.F., Rezende E.L. 2005. Phylogenetic approaches in comparative physiology. J. Exp. Biol. 208:3015–3035.
Gaffney E. 1980. Phylogenetic relationships of the major groups of amniotes. In: Panchen A., editor. The terrestrial environment and the origin of land vertebrates. London: Academic Press. p. 593–610.
Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.-Y., Hansen N.F., Durand E.Y., Malaspinas A.-S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Ž., Gušic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., Rasilla M. de la, Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S. 2010. A draft sequence of the neandertal genome. Science. 328:710–722.
Hime P.M., Lemmon A.R., Lemmon E.C.M., Prendini E., Brown J.M., Thomson R.C., Kratovil J.D., Noonan B.P., Pyron R.A., Peloso P.L.V., Kortyna M.L., Keogh J.S., Donnellan S.C., Mueller R.L., Raxworthy C.J., Kunte K., Ron S.R., Das S., Gaitonde N., Green D.M., Labisko J., Che J., Weisrock D.W. 2021. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. Syst. Biol. 70:48–66.
Huson D.H., Klöpper T., Lockhart P.J., Steel M.A. 2005. Reconstruction of reticulate networks from gene trees. Berlin: Springer. p. 233–249.
Irisarri I., Baurain D., Brinkmann H., Delsuc F., Sire J.-Y., Kupfer A., Petersen J., Jarek M., Meyer A., Vences M., Philippe H. 2017. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat. Ecol. Evol. 1:1370–1378.
IUCN 2021. The IUCN red list of threatened species. Version 2021-2. Available from: URL https://www.iucnredlist.org.
Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Núñez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldså J., Orlando L., Barker F.K., Jønsson K.A., Johnson W., Koepfli K.-P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 346:1320–1331.
Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.
Joyce W.G. 2007. Phylogenetic relationships of mesozoic turtles. Bull. Peabody Mus. Nat. Hist. 48:3–102.
Joyce W.G., Anquetin J., Cadena E.-A., Claude J., Danilov I.G., Evers S.W., Ferreira G.S., Gentry A.D., Georgalis G.L., Lyson T.R., Pérez-García A., Rabi M., Sterli J., Vitek N.S., Parham J.F. 2021. A nomenclature for fossil and living turtles using phylogenetically defined clade names. Swiss J. Palaeontol. 140:5.
Joyce W.G., Parham J.F., Gauthier J.A. 2004. Developing a protocol for the conversion of rank-based taxon names to phylogenetically defined clade names, as exemplified by turtles. J. Paleontol. 78:989–1013.
Joyce W.G., Parham J.F., Lyson T.R., Warnock R.C.M., Donoghue P.C.J. 2013. A divergence dating analysis of turtles using fossil calibrations: an example of best practices. J. Paleontol. 87:612–634.
Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate

phylogenetic estimates. Nat. Methods. 14:587–589.
Karin, B.R., Gamble T., Jackman, T.R. 2020. Optimizing phylogenomics with rapidly evolving long exons: comparison with anchored hybrid enrichment and ultraconserved elements. Mol. Biol. Evol. 37:904–922.
Karl S.A., Bowen B.W., Avise J.C. 1995. Hybridization among the ancient mariners: characterization of marine turtle hybrids with molecular genetic assays. J. Hered. 86:262–268.
Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
Krenz J.G., Naylor G.J.P., Shaffer H.B., Janzen F.J. 2005. Molecular phylogenetics and evolution of turtles. Mol. Phylogenet. Evol. 37:178–191.
Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.
Kumar S., Filipski A.J., Battistuzzi F.U., Kosakovsky Pond S.L., Tamura K. 2012. Statistics and truth in phylogenomics. Mol. Biol. Evol. 29:457–472.
Kumar S., Stecher G., Li M., Knyaz C., Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35:1547–1549.
Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.
Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359.
Lê S., Josse J., Husson F. 2008. FactoMineR: an R Package for multivariate analysis. J. Stat. Softw. 25:1–18.
Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61:727–744.
Linkem C.W., Minin V.N., Leaché A.D. 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata:Scincidae). Syst. Biol. 65:465–477.
Literman R., Schwartz R. 2021. Genome-scale profiling reveals non-coding loci carry higher proportions of concordant data. Mol. Biol. Evol. 38:2306–2318.
Liu L., Edwards S.V. 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58:452–460.
Liu L., Zhang J., Rheindt F.E., Lei F., Qu Y., Wang Y., Zhang Y., Sullivan C., Nie W., Wang J., Yang F., Chen J., Edwards S.V., Meng J., Wu S. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. Proc. Natl. Acad. Sci. USA 114:E7282–E7290.
Lopes F., Oliveira L.R., Kessler A., Beux Y., Crespo E., Cárdenas-Alayza S., Majluf P., Sepúlveda M., Brownell R.L. Jr., Franco-Trecu V., Páez-Rosas D., Chaves J., Loch C., Robertson B.C., Acevedo-Whitehouse K., Elorriaga-Verplancken F.R., Kirkman S.P., Peart C.R., Wolf J.B.W., Bonatto S.L. 2021. Phylogenomic discordance in the eared seals is best explained by incomplete lineage sorting following explosive radiation in the southern hemisphere. Syst. Biol. 70:786–802.
Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.
Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37:2727–2733.
Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534.
Mölder F., Jablonski K.P., Letcher B., Hall M.B., Tomkins-Tinch C.H., Sochat V., Forster J., Lee S., Twardziok S.O., Kanitz A., Wilm A., Holtgrewe M., Rahmann S., Nahnsen S., Köster J. 2021. Sustainable data analysis with Snakemake. Available from: URL https://f1000research.com/articles/10-33.
Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A., Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. Syst. Biol. 70:219–235.
Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.
Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.
Parham J.F., Feldman C.R., Boore J.L. 2006. The complete mitochondrial genome of the enigmatic bigheaded turtle (*Platysternon*): description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evol. Biol. 6:11.
Peden J. 1999. Analysis of codon usage [PhD dissertation]. [Nottingham (UK)]: University of Nottingham.
Pereira A.G., Sterli J., Moreira F.R.R., Schrago C.G. 2017. Multilocus phylogeny and statistical biogeography clarify the evolutionary history or major lineages of turtles. Syst. Biol. 113: 59–66.
Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLOS Biol. 9:e1000602.
Philippe H., Sörhannus U., Baroin A., Perasso R., Gasse F., Adoutte A. 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. J. Evol. Biol. 7: 247–265.
Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 526:569–573.
Quesada V., Freitas-Rodríguez S., Miller J., Pérez-Silva J.G., Jiang Z.-F., Tapia W., Santiago-Fernández O., Campos-Iglesias D., Kuderna L.F.K., Quinzin M., Álvarez M.G., Carrero D., Beheregaray L.B., Gibbs J.P., Chiari Y., Glaberman S., Ciofi C., Araujo-Voces M., Mayoral P., Arango J.R., Tamargo-Gómez I., Roiz-Valle D., Pascual-Torner M., Evans B.R., Edwards D.L., Garrick R.C., Russello M.A., Poulakakis N., Gaughran S.J., Rueda D.O., Bretones G., Marquès-Bonet T., White K.P., Caccone A., López-Otín C. 2019. Giant tortoise genomes provide insights into longevity and age-related disease. Nat. Ecol. Evol. 3:87–95.
Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842.
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.
Ranwez V., Harispe S., Delsuc F., Douzery E.J.P. 2011. MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. PLOS One. 6:e22594.
Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.
Rokas A., Williams B.L., King N., Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425:798–804.
Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.
Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.
Sanderson M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19:301–302.
Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33:1654–1668.
Schwartz R.S., Harkins K.M., Stone A.C., Cartwright R.A. 2015. A composite genome approach to identify phylogenetically informative data from next-generation sequencing. BMC Bioinform. 16: 193.
Shaffer H.B., McCartney-Melstad E., Near T.J., Mount G.G., Spinks P.Q. 2017. Phylogenomic analyses of 539 highly informative loci dates a fully resolved time tree for the major clades of living turtles (Testudines). Mol. Phylogenet. Evol. 115:7–15.
Shaffer H.B., Meylan P., McKnight M.L. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Syst. Biol. 46:235–268.

Shaffer H.B., Minx P., Warren D.E., Shedlock A.M., Thomson R.C., Valenzuela N., Abramyan J., Amemiya C.T., Badenhorst D., Biggar K.K., Borchert G.M., Botka C.W., Bowden R.M., Braun E.L., Bronikowski A.M., Bruneau B.G., Buck L.T., Capel B., Castoe T.A., Czerwinski M., Delehaunty K.D., Edwards S.V., Fronick C.C., Fujita M.K., Fulton L., Graves T.A., Green R.E., Haerty W., Hariharan R., Hernandez O., Hillier L.W., Holloway A.K., Janes D., Janzen F.J., Kandoth C., Kong L., de Koning A.P.J., Li Y., Literman R., McGaugh S.E., Mork L., O'Laughlin M., Paitz R.T., Pollock D.D., Ponting C.P., Radhakrishnan S., Raney B.J., Richman J.M., St John J., Schwartz T., Sethuraman A., Spinks P.Q., Storey K.B., Thane N., Vinaø T., Zimmerman L.M., Warren W.C., Mardis E.R., Wilson R.K. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome Biol. 14:R28.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492–508.

Singhal S., Colston T.J., Grundler M.R., Smith S.A., Costa G.C., Colli G.R., Moritz C., Pyron R.A., Rabosky D.L. 2021. Congruence and conflict in the higher-level phylogenetics of squamate reptiles: an expanded phylogenomic perspective. Syst. Biol. 70:542–557.

Smith, M.L., Hahn, M.W. 2021. New approaches for inferring phylogenies in the presence of paralogs. Trends Genet. 37: 174–187.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Steenwyk J.L., Buida T.J. III, Labella A.L., Li Y., Shen X.-X., Rokas A. 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics. 37: 2325–2331.

Sung Y-H., Hau B.C.U., Karraker N.E. 2014. Reproduction of endangered big-headed turtle, *Platysternon megacephalum* (Reptilia: Testudines: Platysternidae). Acta Herpetol. 9: 237–247.

Swofford D.L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.

Tarver J.E., dos Reis M., Mirarab S., Moran R.J., Parker S., O'Reilly J.E., King B.L., O'Connell M.J., Asher R.J., Warnow T., Peterson K.J., Donoghue P.C.J., Pisani D. 2016. The interrelationships of placental mammals and the limits of phylogenetic inference. Genome Biol. Evol. 8:330–344.

Thomson R.C., Spinks P.Q., Shaffer H.B. 2021. A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. Proc. Natl. Acad. Sci. USA 118:e2012215118.

Tollis M., Denardo D.F., Cornelius J.A., Dolby G.A., Edwards T., Henen B.T., Karl A.E., Murphy R.W., Kusumi K. 2017. The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. PLoS One 12:e0177708.

Uetz P., Freed P., Aguilar R., Hošek J. 2021. The reptile database. Available from: http://www.reptile-database.org/.

Wang Z., Pascual-Anaya J., Zadissa A., Li W., Niimura Y., Wang J., Huang Z., Li C., White S., Xiong Z., Fang D., Wang B., Ming Y., Chen Y., Zheng Y., Kuraku S., Pignatelli M., Herrero J., Beal K., Nozawa M., Li Q., Wang J., Zhang H., Yu L., Shigenobu S., Wang J., Liu J., Flicek P., Searle S., Wang J., Kuratani S., Yin Y., Aken B., Zhang G., Irie N. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat. Genet. 45:701–706.

Waterhouse R.M., Seppey M., Simão F.A., Manni M., Ioannidis P., Klioutchnikov G., Kriventseva E.V., Zdobnov E.M. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35:543–548.

Wolf Y.I., Rogozin I.B., Grishin N.V., Koonin E.V. 2002. Genome trees and the tree of life. Trends Genet. 18:472–479.

Zdobnov E.M., Tegenfeldt F., Kuznetsov D., Waterhouse R.M., Simão F.A., Ioannidis P., Seppey M., Loetscher A., Kriventseva E.V. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 45:D744–D749.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform. 19:153.