Draft genomes of a male and female Australian jacky dragon (*Amphibolurus muricatus*)

Ran Tian[1,2,‡], Hao Dong[1,‡], Fan Zhang[1], Hao Yu[1], Enqing Pei[1], Chengcheng Shi[3,4], Guangyi Fan[3,4], Sarah L. Whiteley[5,6], Clare E. Holleley[6], Inge Seim[1,7], Arthur Georges[5,8*]

[1] Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China
[2] Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China
[3] BGI-Qingdao, Qingdao 266555, China
[4] BGI-Shenzhen, Shenzhen 518083, China
[5] Institute for Applied Ecology, University of Canberra ACT 2617, Australia
[6] Australian National Wildlife Collection, CSIRO, Canberra ACT 2601, Australia
[7] School of Biology and Environmental Science, Queensland University of Technology, Brisbane Qld 4000, Australia
[8] AusARG consortium, Bioplatforms Australia, Macquarie University NSW 2109, Australia

[*]Corresponding author: Institute for Applied Ecology, University of Canberra, ACT 2601, Australia. E-mail: georges@aerg.canberra.edu.au
[‡]These authors contributed equally to this work and share first authorship.

ABSTRACT

Australia is remarkable for its lizard diversity, with very high endemicity because of continental-scale diversification and adaptive radiation during prolonged isolation. We here employed stLFR linked-read technology to generate male and female draft genomes of the jacky dragon *Amphibolurus muricatus*, an Australian dragon lizard (family Agamidae; the agamids). The assemblies are 1.8 Gb in size and have a repeat content (39%) and GC content (42%) similar to other dragon lizards. The longest scaffold was 39.7 Mb (female) and 9.6 Mb (male), with corresponding scaffold N50 values of 6.8 Mb and 1.6 Mb. The BUSCO (Sauropsida database) completeness percentages were 90.2% and 88.8% respectively. Phylogenetic comparisons show that Australian and Asian agamids split from a common ancestor about 80 million years ago, while the Australian genera *Amphibolurus*, *Pogona*, and the basal *Intellagama* split ~37 million years ago. The draft *A. muricatus* assemblies will be a valuable resource for understanding lizard sex determination and the evolution and conservation of Australian dragon lizards.

Keywords: agamid lizard; Agamidae; squamate; nuclear genome; genome assembly

Introduction

The Australian jacky dragon *Amphibolurus muricatus* (**Figure 1**) is a lizard that is widespread in dry sclerophyll forests of south-eastern and eastern Australia (Cogger 2014). It is a model species for biogeography (Pepper et al. 2014), evolutionary biology (Warner et al. 2013; Warner and Shine 2008), social behaviour (Peters and Evans 2003; Woo and Rieucau 2013) and development (Whiteley et al. 2021; Esquerré et al. 2014).

Species in the genus *Amphibolurus* and *Chlamydosaurus* are a major clade in the Australian radiation of the Agamidae (Hugall et al. 2008). The draft assembly of *A. muricatus*, together with that of *Pogona vitticeps* (Georges et al. 2015), represents the first foray into generating the necessary high-quality genomes for the Agamidae. In particular, *A.muricatus* occupies mesic habitats and so is intermediate between the Australian water dragon *Intellagama lesueurii* and the forest dragon *Lophosaurus boydii* that occupy hydric habitats, and the central bearded dragon *Pogona vitticeps* and the Lake Eyre dragon *Ctenophorus maculosus,* for example, that occupy more xeric habitats. As such, it is one of several species important for understanding genomic adaptation to the progressive aridity that has occurred in Australia in the past 15 Myr. *Amphibolurus muricatus* is also of particular interest because it has temperature-dependent sex determination (TSD) (Harlow and Taylor 2000) and it is unclear as to whether this arises from classical TSD or a combination of genetic and environmental influences (Whiteley et al. 2021). Studies of the underlying mechanisms of TSD require a genome assembly and knowledge of genome organisation to identify genes on the sex chromosomes of species with genotypic sex determination (GSD) and their chromosomal and gene homology in closely related TSD species. This is particularly so in species with TSD that show evidence of cryptic residual or de novo genotypic influence on offspring sex ratios, as is suspected for *A. muricatus* (Whiteley et al. 2021).

Here, we generated draft, annotated genome assemblies for a male and a female *A. muricatus*. We used transcriptomes sequenced and assembled for *A. muricatus* and published genomes (*Anolis*, *Varanus* and *Pogona*) to annotate the assemblies. Our assemblies will provide a resource to increase capacity and accelerate the progress of studies into the evolution, ecology, and conservation of Australian dragon lizards.

Materials and methods

Sample collection

In attempt to reduce the high heterozygosity that presented difficulties in the assembly of the genome of *Pogona vitticeps* (Georges et al., 2015), we generated inbred lines of *A. muricatus*. The founding male and female pair were sourced from the wild and bred in captivity. The two animals used to generate the genome were obtained from the fourth generation of the inbred pedigree produced by sib-sib matings and back crossing (see **Figure S1** for the complete pedigree). The male (AA069033) and female (AA069032) individuals used for the genome and transcriptome sequencing were humanely euthanised via intraperitoneal injection of sodium pentobarbitone (60 mg/ml in isotonic saline). Organs were rapidly dissected and snap frozen in liquid nitrogen.

DNA extraction

High molecular weight DNA was extracted from liver (female) and blood (male). DNA for 10x Genomics linked-read sequencing was extracted using a QIAGEN Gentra Puregene DNA Isolation kit (to extract >100 kb DNA fragments). DNA for stLFR linked-read sequencing was extracted using a QIAGEN MagAttract HMW DNA Kit (to extract >40 kb DNA fragments). DNA yield and quality was assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and a Qubit fluorometer (Thermo Fisher Scientific) and pulse-field gel electrophoresis.

Genome size and heterozygosity estimation

We used the paired-end stLFR sequencing reads to estimate the genome size and heterozygosity of the female and male *A. muricatus* samples. Briefly, *k*-mers (*k*=17, as employed in a previous study on *P. vitticeps* (Georges et al. 2015)) in reads were counted and converted to histogram files using Jellyfish (v2.2.10) (Marcais and Kingsford 2011), followed by analysis using GenomeScope v2.0 (Vurture et al. 2017; Ranallo-Benavidez et al. 2020).

Assembly 1.0: A 10x Genomics linked-read sequencing assembly

Male and female *A. muricatus* genome sequencing libraries were constructed on the Chromium system (10x Genomics, Pleasanton, CA, USA) by the Ramaciotti Centre for Genomics (Sydney, Australia). The Chromium instrument enables unique barcoding of long stretches of DNA on gel beads. The barcodes allow later reconstruction of long DNA

fragments from a series of short DNA fragments with the same barcode (i.e., linked-reads). After barcoding, DNA was sheared into smaller fragments and sequenced on the NovaSeq 6000 platform (Illumina, CA, USA) to generate 151 bp paired-end (PE) reads. A total of 904.9 M raw 10x Genomics Chromium linked-reads were generated. Raw 10x data were assembled with Supernova v2.1.1 (Weisenfeld et al. 2017) and a FASTA file was generated using the 'pseudohap style' option in Supernova mkoutput. All female (~450 M) and male (~550 M) read pairs were used (female sequencing depth *ca* 50.3×; male, *ca* 47.8×). The resulting assemblies were further scaffolded with ARKS v1.0.3 (Coombe et al. 2018), reusing the 10x reads and the companion LINKS program (v1.8.7) (Warren et al. 2015). ARKS employs a *k*-mer approach to map linked barcodes to the contigs in the initial Supernova assembly to generate a scaffold graph with estimated distances for LINKS input. These assemblies were denoted AmpMurF_1.0 (female) and AmpMurM_1.0 (male). We used GapCloser v1.12 (part of SOAPdenovo2) (Luo et al. 2012) to fill gaps in the assembly. GapCloser was run using the parameter -l 150) and clean 10x Genomics reads PE reads.

Assembly 1.1: Further scaffolding of assembly 1.0 using RNA-seq data

We attempted to improve the v1.0 genome assemblies' contiguity using RNA-sequencing reads. RNA-seq reads (from brain, ovary, and testis; see below) were filtered (i.e., cleaned) to remove adapters and low-quality reads using Flexbar v3.4.0 and used to further re-scaffold the v1.0 assemblies (FASTA files before gapclosing) with P_RNA_scaffolder (Zhu et al. 2018). The default Flexbar settings discards all reads with any uncalled bases. A final round of scaffolding was performed on the resulting assemblies using L_RNA_scaffolder (Xue et al. 2013). These assemblies were denoted AmpMurF_1.1 (female) and AmpMurM_1.1 (male). As before, GapCloser and clean 10x Genomics reads were used to fill gaps.

Assembly 2.0: Further scaffolding of assembly 1.0 using SLR-superscaffolder

As an alternative approach, we attempted to improve the v1.0 genome assemblies' contiguity using SLR-superscaffolder (Guo et al. 2021). Briefly, SLR-superscaffolder employs single-tube long fragment read (stLFR) sequencing (Wang et al. 2019) reads (see section below) to generate hybrid genome assemblies. The software was run with default parameters except for PE_SEED_MIN=300 (minimum contig size to fill; default 1000). These assemblies were denoted AmpMurF_2.0 (female) and AmpMurM_2.0 (male). GapCloser and clean stLFR reads (with the barcode removed using https://github.com/BGI-Qingdao/stLFR_barcode_split) were used to fill gaps.

Assembly 3.0: An stLFR linked-read sequencing assembly

We also generated independent assemblies for the individuals sequenced on the 10x Genomics Chromium system using stLFR sequencing (Wang et al., 2019). BGI (Brisbane, Australia) generated ~100× coverage 100-bp paired-end reads (plus a 42-bp stLFR barcode on the right/_2 read) per individual. Low-quality reads, PCR duplicates, and adaptors were removed using SOAPnuke v1.5 (Chen et al. 2018). All female (~1,517 M) and male (~1,427 M) read pairs were utilised. The stLFRdenovo pipeline (https://github.com/BGI-biotools/stLFRdenovo), which is based on Supernova v2.11 and customised for stLFR data, was used to generate a *de novo* genome assembly. The stLFRdenovo tool 'FillGaps' was used to fill gaps.

RNA-seq and transcriptome assembly

Raw data 125 bp PE reads, generated on an Illumina HiSeq 2500 instrument was filtered using Flexbar v3.4.0 (Roehr et al. 2017; Dodt et al. 2012) with default settings (eliminates reads with any uncalled bases). Any residual ribosomal RNA reads (the majority removed by poly(A) selection prior to sequencing library generation) were removed using SortMeRNA v2.1b (Kopylova et al. 2012) against the SILVA v119 ribosomal database (Quast et al. 2013). Tissue transcriptomes were de novo assembled using Trinity v2.11.0 (Haas et al. 2013; Grabherr et al. 2011; Henschel et al. 2012) and assessed using BUSCO v5.0.0.

Genome annotation

We identified repetitive elements by integrating homology and de novo prediction data. Protein-coding genes were annotated using homology-based prediction, de novo prediction, and RNA-seq-assisted prediction methods.

Homology-based transposable elements (TE) annotations were obtained by interrogating a genome assembly with known repeats in the Repbase database v16.02 (Bao et al. 2015) using RepeatMasker v4.0.5 (DNA-level) (parameters: -species lizard -e rmblast -xsmall -s -gff -pa 4) (Tarailo-Graovac and Chen 2009) and RepeatProteinMask (protein-level; implemented in RepeatMasker). De novo TE predictions were obtained using RepeatModeler v1.1.0.4 (parameters: -pa 4 -database lizard -LTRStruct) (Smit and Hubley 2010) and LTRharvest v1.5.8 (default parameters) (Ellinghaus et al. 2008) to generate database for a RepeatMasker run. Tandem Repeat Finder v4.07 (2 7 7 80 10 50 500 -m -f -d) with default settings (i.e., MaxPeriod parameter, the maximum period size,) (Benson 1999)

was used to find tandem repeats (TRs) in the genome. A non-redundant repeat annotation set was obtained by combining the above data.

Protein-coding genes were annotated using homology-based prediction, de novo prediction, and RNA-seq-assisted [generated from ovary, testis, and brain (both sexes)] prediction methods. Sequences of homologous proteins from three lizards [*Anolis carolinensis* (green anole) assembly AnoCar2.0 (RefSeq assembly GCF_000090745.1) (Alfoldi et al. 2011); *Varanus komodoensis* (Komodo dragon) assembly ASM479886v1 (GCA_004798865.1) (Lind et al. 2019); and *Pogona vitticeps* (central bearded dragon) assembly pvi1.1 (GCF_900067755.1)] (Georges et al. 2015) were downloaded from NCBI. These protein sequences were aligned to the repeat-masked genome using BLAT v0.36 (Kent 2002). GeneWise v2.4.1 (Birney et al. 2004) was employed to generate gene structures based on the alignments of proteins to a genome assembly. De novo gene prediction was performed using AUGUSTUS v3.2.3 (Stanke et al. 2006), GENSCAN v1.0 (Burge and Karlin 1997), and GlimmerHMM v3.0.1 (Majoros et al. 2004) with a human training set. Transcriptome data (clean reads) were mapped to the assembled genome using HISAT2 v2.1.0 (Kim et al. 2019) and SAMtools v1.9 (Li et al. 2009), and coding regions were predicted using TransDecoder v5.5.0 (Grabherr et al. 2011; Haas et al. 2013). A final non-redundant reference gene set was generated by merging the three annotated gene sets using EvidenceModeler v1.1.1 (EVM) (Haas et al. 2008) and excluding EVM gene models with only ab initio support. The gene models were translated into amino acid sequences and used in local BLASTp (Camacho et al. 2009) searches against the public databases Kyoto Encyclopedia of Genes and Genomes (KEGG; v89.1) (Kanehisa and Goto 2000), NCBI non-redundant protein sequences (NR; v20170924) (O'Leary et al. 2016), Swiss-Prot (release-2018_07) (UniProt Consortium 2012), and InterPro (v69.0) (Mitchell et al. 2019).

Phylogeny and divergence time estimation

In addition to *A. carolinensis*, *V. komodoensis* and *P. vitticeps* (see section above), the genome and sequences of homologous proteins from *Gekko japonicus* (Schlegel's Japanese gecko) assembly Gekko_japonicus_V1.1 (GCA_001447785.1) (Liu et al. 2015) and *Crotalus tigris* (tiger rattlesnake) assembly ASM1654583v1 (GCA_016545835.1) (Margres et al. 2021) were downloaded from NCBI. The genome and annotations of *Ophisaurus gracilis* (Anguidae lizard) were downloaded from GigaDB (Song et al. 2015a; Song et al. 2015b). No gene annotation data were available for three species: *Intellagama lesueurii* (Australian water dragon; assembly EWD_hifiasm_HiC generated as part of the AusARG consortium and

(downloaded from DNA Zoo (Dudchenko et al. 2018; Cheng et al. 2021; Dudchenko et al. 2017)) and the Chinese agamid lizards *Phrynocephalus przewalskii* (Przewalski's toadhead agama) (Gao et al. 2019) and *Pyrocephalus vlangalii* (Ching Hai toadhead agama) (Gao et al. 2019) (CNGBdb accession no. CNP0000203). Their protein-coding genes were annotated using homology-based prediction, de novo prediction, and RNA-seq-assisted prediction methods (see genome annotation section above).

We identified 1:1 orthologs by interrogating the predicted proteins from the gene models of 13 squamate species using SonicParanoid v1.3.0 (Cosentino and Iwasaki 2019). The corresponding coding sequences (CDS) for each species were aligned using PRANK v100802 (Loytynoja and Goldman 2005) and filtered by Gblocks v0.91b (Talavera and Castresana 2007) to identify conserved blocks (removing gaps, ambiguous sites, and excluding alignments less than 300 bp in size), leaving 3,130 genes. Maximum-likelihood (ML) phylogenetic trees were generated using RaxML v7.2.8 (Stamatakis 2006) and IQ-Tree v2.1.3 (Minh et al. 2020) with three CDS data sets: the whole coding sequence (whole-CDS), first codon positions, and fourfold degenerate (4d) sites. Identical topologies and similar support values were obtained (1,000 bootstrap iterations were performed). The divergence time between species was estimated using MCMCTree [a Bayesian molecular clock model implemented in PAML v4.7 (Yang 2007)] with the JC69 nucleotide substitution model, and the whole-CDS ML tree and concatenated whole-CDS supergenes as inputs. We used 100,000 iterations after a burn-in of 10,000 iterations. MCMCTree calibration points (million years ago; Mya) were obtained from (Oliver and Hugall 2017) (crown age of Australian agamids 27.1 Mya, with 95% CI 20.1-37.7) and TimeTree (Kumar et al. 2017): *G. japonicus*-*P. przewalskii* (190–206 Mya), *V. komodoensis*-*O. gracilis* (121–143 Mya), *V. komodoensis*-*C. tigris* (156–174 Mya), *V. komodoensis*-*A. carolinensis* (155–175 Mya), *I. lesueurii*-*A. carolinensis* (139–166 Mya), *I. lesueurii*-*P. przewalskii* (73–93 Mya), *I. lesueurii*-*A. muricatus* (25.5–42.4 Mya), *P. vitticeps*-*A. muricatus* (20.2–34.6 Mya).

Results and discussion

Draft genome assembly and comparisons with other squamates

We used stLFR reads, Jellyfish (Marcais and Kingsford 2011), and GenomeScope (Vurture et al. 2017; Ranallo-Benavidez et al. 2020) to estimate genome size and heterozygosity. The size of the *A. muricatus* genome was estimated to be around 1.5 Gb. The genome-wide heterozygosity from our inbred (four generations) *A. muricatus* lines was estimated to range from 0.66% (female) to 0.73% (male), slightly lower than the central bearded dragon (*Pogona vitticeps*) (0.85%) (Georges et al. 2015).

We generated four genome assemblies per sample. The v1.0 assemblies were generated using 10x Genomics Chromium data and the Supernova assembler and further refined using ARKS and LINKS. The v1.1 assemblies employed P_RNA_scaffolder (uses RNA-seq reads from brain, ovary, and testis) (**Table S1**) and L_RNA_scaffolder (uses Trinity transcriptome assemblies) (**Tables S2** and **S3**) to improve the v1.0 assemblies, while the v2.0 assemblies used SLR-superscaffolder and stLFR reads to improve the v1.0 assemblies. Finally, the 3.0 assemblies were generated using stLFR reads alone and Supernova. While re-scaffolding of the assemblies generated using 10x Genomics Chromium sequencing improved the initial v1.0 assembly (in particular, SLR-superscaffolder), assembly using stLFR data alone gave the best assembly result (**Table S4**).

For the remainder of the manuscript, conclusions will be drawn from the *A. muricatus* v3.0 assemblies. A comparison with seven other agamid genome assemblies is shown in **Table 1**. The final, v3.0 assemblies have a total scaffold length (i.e., containing gaps) of 1.87 (female) and 1.80 (male) Gb. The longest scaffold was 39.7 Mb (female; AmpMurF_3.0) and 9.6 Mb (male; AmpMurM_3.0), and the corresponding scaffold N50 values of 6.9 and 1.6 Mb. The contig N50s were 67.2 kb (AmpMurF_3.0) and 59.3 kb (AmpMurM_3.0). The N50 values are lower than most recent squamate assemblies but similar to the agamids central bearded dragon (*P. vitticeps*; scaffold N50 2.29 Mb and contig N50 31.30 kb) (Georges et al. 2015), Przewalski's toadhead agama (*P. przewalskii*; scaffold N50 6.88 MB and contig N50 56.43 kb) (Gao et al. 2019), and Ching Hai toadhead agama (*P. forsythii*; scaffold N50 2.40 Mb and contig N50 31.29 kb) (Gao et al. 2019) (**Figure 2**). The BUSCO (v5.0.0 with the 7,480-gene sauropsida_odb10 dataset) metrics of the *A. muricatus* assemblies compared well to other squamate assemblies, including agamids from Australia [*P. vitticeps* (Georges et al. 2015) and *I. lesueurii* (Australian water dragon)], China (toad-headed agamas of genus

Phrynocephalus sp. (Gao et al. 2019; Jin et al. 2022; Qi et al. 2023)), and the widely distributed oriental garden lizard (*C. versicolor*) (Wang et al. 2023) (**Figure 3** and **Table S5**).

Genome annotation

The *A. muricatus* assemblies are composed of ~38% repeat elements and have a GC content of ~42% (**Tables S6-S8**), similar to that of *P. vitticeps* (Georges et al. 2015) – with LINEs being the predominant subtype. Protein-coding genes were annotated by combining transcriptome evidence with homology-based (*A. carolinensis*, *V. komodoensis*, and *P. vitticeps*) and de novo gene prediction methods. Gene statistics (**Table S9**) (see (Georges et al. 2015)) and gene set BUSCO scores (**Table S10**) are comparable to other squamates. Using ab inito, transcriptome, and homology-based prediction methods, we functionally annotated 21,655 (95.12%) and 21,799 (94.70%) protein-coding genes in the female and male assembly (**Tables S11 and S12**) and recovered 89.2% and 88.2% of 7,480 sauropsid (i.e., non-avian reptiles and birds) benchmarking universal single-copy orthologs (BUSCOs), respectively.

Phylogenetic relationships

To construct a time-calibrated species tree (**Figure 4**), we identified 3,130 single-copy orthologs from the female *A. muricatus* assembly and 12 other squamate species. In addition to *A. muricatus*, seven agamid species currently have genome assemblies: two Australian dragon lizards (*P. vitticeps* and *I. lesueurii*), four toad-headed agamas (genus *Phrynocephalus*), and the oriental garden lizard (*C. versicolor*) (Gao et al. 2019; Georges et al. 2015; Qi et al. 2023; Wang et al. 2023; Jin et al. 2022). Our analysis shows that the agamids shared an ancestor about 79.8 Mya [74.5–86.1 Mya 95% credibility interval (CI)]. We estimate that the three Australian dragon lizard species split from a common ancestor about 37.2 Mya (95% CI 30.6–42.2), while the lineages leading to *A. muricatus* and *P. vitticeps* diverged 24.6 Mya (95% CI 19.6–31.7). These observations agree with previous dating from a small set of genes and fossil data (Hugall et al. 2008; Oliver and Hugall 2017).

Conclusions and perspectives

In this study, we generated the first annotated genome assemblies of *Amphibolurus muricatus*. Overall, the assemblies are similar in quality to a range of squamate genomes and will be immediately useful for the understanding of agamid lizard evolution, ecology, and conservation.

Data availability

A. muricatus raw 10x Genomics genome and transcriptome sequencing reads have been deposited to the NCBI Short Read Database (BioProject ID: PRJNA767251). Raw stLFR genome sequencing reads have been deposited at the China National GeneBank Nucleotide Sequence Archive (CNSA: https://db.cngb.org/cnsa) under accession number CNP0004768. The male and female *A. muricatus* genome assemblies are available at Zenodo (Tian et al. 2023a). Gene annotation files and associated FASTA files for *A. muricatus* (assembly AmpMurF_3.0 and AmpMurM_3.0), *I. lesueurii*, *P. przewalskii*, and *P. vlangalii* are available at Zenodo (Tian et al. 2023b). *A. muricatus* transcriptome assemblies are available at Zenodo (Tian et al. 2021). Various scripts used for data processing and analyses are available on GitHub at https://github.com/sciseim/JackyDragon.

Acknowledgements

We thank Dr Wendy Ruscoe and Jacqui Richardson for their assistance in generating the inbred line of *A. muricatus* and for animal husbandry.

Conflict of interest

The authors declare there is no conflict of interest.

Ethics Approvals

All sampling and breeding experiments were conducted with approval of the Animal Ethics Committee of the University of Canberra and in accordance with their Standard Operating Procedures.

Funding

Financial support for this work was provided by an Australian Research Council Discovery Grant (DP170101147; to A.G. and C.E.H.), a specially-appointed Professor of Jiangsu Province grant (to I.S.), the Jiangsu Science and Technology Agency (to I.S.), the Jiangsu Foreign Expert Bureau (to I.S.), the Jiangsu Provincial Department of Technology (grant JSSCTD202142 to I.S.), the National Natural Science Foundation of China (grant 32270441 to R.T.), the National Key Programme of Research and Development, Ministry of Science and Technology (grant 2022YFF1301601 to R.T.), and the Young Elite Scientists Sponsorship Program of the China Association for Science and Technology (grant 2023QNRC001 to R.T.).

Literature cited

Alfoldi, J., F. Di Palma, M. Grabherr, C. Williams, L. Kong *et al.*, 2011 The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477 (7366):587-591.

Bao, W., K.K. Kojima, and O. Kohany, 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.

Benson, G., 1999 Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27 (2):573-580.

Birney, E., M. Clamp, and R. Durbin, 2004 GeneWise and Genomewise. *Genome Res* 14 (5):988-995.

Burge, C., and S. Karlin, 1997 Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268 (1):78-94.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Chen, Y., Y. Chen, C. Shi, Z. Huang, Y. Zhang *et al.*, 2018 SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7 (1):1-6.

Cheng, H., G.T. Concepcion, X. Feng, H. Zhang, and H. Li, 2021 Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 18 (2):170-175.

Cogger, H., 2014 *Reptiles and amphibians of Australia*: CSIRO publishing.

Coombe, L., J. Zhang, B.P. Vandervalk, J. Chu, S.D. Jackman *et al.*, 2018 ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics* 19 (1):234.

Cosentino, S., and W. Iwasaki, 2019 SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics* 35 (1):149-151.

Dodt, M., J.T. Roehr, R. Ahmed, and C. Dieterich, 2012 FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* 1 (3):895-905.

Dudchenko, O., S.S. Batra, A.D. Omer, S.K. Nyquist, M. Hoeger *et al.*, 2017 De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356 (6333):92-95.

Dudchenko, O., M.S. Shamim, S.S. Batra, N.C. Durand, N.T. Musial *et al.*, 2018 The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *BioRxiv*:254797.

Ellinghaus, D., S. Kurtz, and U. Willhoeft, 2008 LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

Esquerré, D., J.S. Keogh, and L.E. Schwanz, 2014 Direct effects of incubation temperature on morphology, thermoregulatory behaviour and locomotor performance in jacky dragons (Amphibolurus muricatus). *Journal of Thermal Biology* 43:33-39.

Gao, W., Y.B. Sun, W.W. Zhou, Z.J. Xiong, L. Chen *et al.*, 2019 Genomic and transcriptomic investigations of the evolutionary transition from oviparity to viviparity. *Proc Natl Acad Sci U S A* 116 (9):3646-3655.

Georges, A., Q. Li, J. Lian, D. O'Meally, J. Deakin *et al.*, 2015 High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard Pogona vitticeps. *Gigascience* 4:45.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29 (7):644-652.

Guo, L., M. Xu, W. Wang, S. Gu, X. Zhao *et al.*, 2021 SLR-superscaffolder: a de novo scaffolding tool for synthetic long reads using a top-to-bottom scheme. *BMC Bioinformatics* 22 (1):158.

Haas, B.J., A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood *et al.*, 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8 (8):1494-1512.

Haas, B.J., S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen *et al.*, 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9 (1):R7.

Harlow, P.S., and J.E. Taylor, 2000 Reproductive ecology of the jacky dragon (Amphibolurus muricatus): an agamid lizard with temperature dependent sex determination. *Austral Ecology* 25 (6):640-652.

Henschel, R., M. Lieber, L.-S. Wu, P.M. Nista, B.J. Haas *et al.*, 2012 Trinity RNA-Seq assembler performance optimization, pp. 45 in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*. Association for Computing Machinery, Chicag, IL, USA.

Hugall, A.F., R. Foster, M. Hutchinson, and M.S. Lee, 2008 Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. *Biological Journal of the Linnean Society* 93 (2):343-358.

Jin, Y., D. Aguilar-Gomez, Y.C.B. D, T.A. Square, J. Li *et al.*, 2022 Population Genomics of Variegated Toad-Headed Lizard Phrynocephalus versicolor and Its Adaptation to the Colorful Sand of the Gobi Desert. *Genome Biol Evol* 14 (7).

Kanehisa, M., and S. Goto, 2000 KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28 (1):27-30.

Kent, W.J., 2002 BLAT--the BLAST-like alignment tool. *Genome Res* 12 (4):656-664.

Kim, D., J.M. Paggi, C. Park, C. Bennett, and S.L. Salzberg, 2019 Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37 (8):907-915.

Kopylova, E., L. Noe, and H. Touzet, 2012 SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28 (24):3211-3217.

Kumar, S., G. Stecher, M. Suleski, and S.B. Hedges, 2017 TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 34 (7):1812-1819.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078-2079.

Lind, A.L., Y.Y.Y. Lai, Y. Mostovoy, A.K. Holloway, A. Iannucci *et al.*, 2019 Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nat Ecol Evol* 3 (8):1241-1252.

Liu, Y., Q. Zhou, Y. Wang, L. Luo, J. Yang *et al.*, 2015 Gekko japonicus genome reveals evolution of adhesive toe pads and tail regeneration. *Nat Commun* 6:10033.

Loytynoja, A., and N. Goldman, 2005 An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A* 102 (30):10557-10562.

Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang *et al.*, 2012 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1 (1):18.

Majoros, W.H., M. Pertea, and S.L. Salzberg, 2004 TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20 (16):2878-2879.

Marcais, G., and C. Kingsford, 2011 A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27 (6):764-770.

Margres, M.J., R.M. Rautsaw, J.L. Strickland, A.J. Mason, T.D. Schramer *et al.*, 2021 The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc Natl Acad Sci U S A* 118 (4).

Minh, B.Q., H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams *et al.*, 2020 IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 37 (5):1530-1534.

Mitchell, A.L., T.K. Attwood, P.C. Babbitt, M. Blum, P. Bork *et al.*, 2019 InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47 (D1):D351-D360.

O'Leary, N.A., M.W. Wright, J.R. Brister, S. Ciufo, D. Haddad *et al.*, 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44 (D1):D733-745.

Oliver, P.M., and A.F. Hugall, 2017 Phylogenetic evidence for mid-Cenozoic turnover of a diverse continental biota. *Nat Ecol Evol* 1 (12):1896-1902.

Pepper, M., M.D. Barquero, M.J. Whiting, and J.S. Keogh, 2014 A multi-locus molecular phylogeny for Australia's iconic Jacky Dragon (Agamidae: Amphibolurus muricatus): Phylogeographic structure along the Great Dividing Range of south-eastern Australia. *Molecular Phylogenetics and Evolution* 71:149-156.

Peters, R.A., and C.S. Evans, 2003 Introductory tail-flick of the Jacky dragon visual display: signal efficacy depends upon duration. *Journal of Experimental Biology* 206 (23):4293-4307.

Qi, Y., W. Zhao, Y. Zhao, C. Niu, S. Cao *et al.*, 2023 Chromosome-level genome assembly of Phrynocephalus forsythii using third-generation DNA sequencing and Hi-C analysis. *DNA Res* 30 (2).

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer *et al.*, 2013 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41 (Database issue):D590-596.

Ranallo-Benavidez, T.R., K.S. Jaron, and M.C. Schatz, 2020 GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* 11 (1):1432.

Roehr, J.T., C. Dieterich, and K. Reinert, 2017 Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* 33 (18):2941-2942.

Smit, A.F., and R. Hubley, 2010 RepeatModeler Open-1.0. 2008-2015. Available at http://www.repeatmasker.org.

Song, B., S. Cheng, Y. Sun, X. Zhong, J. Jin *et al.*, 2015a Anguidae lizard (Ophisaurusgracilis) genome assembly data. *Dryad Digital Repository. doi* 10:100119.

Song, B., S. Cheng, Y. Sun, X. Zhong, J. Jin *et al.*, 2015b A genome draft of the legless anguid lizard, Ophisaurus gracilis. *Gigascience* 4:17.

Stamatakis, A., 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22 (21):2688-2690.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack *et al.*, 2006 AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34 (Web Server issue):W435-439.

Talavera, G., and J. Castresana, 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56 (4):564-577.

Tarailo-Graovac, M., and N. Chen, 2009 Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4:Unit 4 10.

Tian, R., H. Guo, C. Yang, G. Fan, S.L. Whiteley *et al.*, 2021 Assembled transcriptomes of ovary, testis, and brain (male and female) of Amphibolurus muricatus (jacky dragon) generated using Trinity v2.11.0, https://doi.org/10.5281/zenodo.5523684. Zenodo.

Tian, R., H. Guo, C. Yang, C. Shi, G. Fan *et al.*, 2023a Draft de novo genome assemblies of a male and female Amphibolurus muricatus (jacky dragon), https://doi.org/10.5281/zenodo.8211053. Zenodo.

Tian, R., H. Guo, C. Yang, C. Shi, G. Fan *et al.*, 2023b Gene annotations of Amphibolurus muricatus (jacky dragon), Intellagama lesueurii (Australian water dragon), Phrynocephalus przewalskii (Przewalski's toadhead agama), and Phrynocephalus vlangalii (Ching Hai toadhead agama), https://doi.org/10.5281/zenodo.8210940. Zenodo.

UniProt Consortium, 2012 Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40 (Database issue):D71-75.

Vurture, G.W., F.J. Sedlazeck, M. Nattestad, C.J. Underwood, H. Fang *et al.*, 2017 GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33 (14):2202-2204.

Wang, O., R. Chin, X. Cheng, M.K.Y. Wu, Q. Mao *et al.*, 2019 Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res* 29 (5):798-808.

Wang, Q., F. He, R.Y. Huang, X. Yang, D. Yang *et al.*, 2023 Draft genome of the oriental garden lizard (Calotes versicolor). *Front Genet* 14:1091544.

Warner, D., and R. Shine, 2008 The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* 451 (7178):566-568.

Warner, D.A., T. Uller, and R. Shine, 2013 Transgenerational sex determination: the embryonic environment experienced by a male affects offspring sex ratio. *Scientific reports* 3 (1):1-4.

Warren, R.L., C. Yang, B.P. Vandervalk, B. Behsaz, A. Lagman *et al.*, 2015 LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* 4:35.

Weisenfeld, N.I., V. Kumar, P. Shah, D.M. Church, and D.B. Jaffe, 2017 Direct determination of diploid genome sequences. *Genome Res* 27 (5):757-767.

Whiteley, S.L., A. Georges, V. Weisbecker, L.E. Schwanz, and C.E. Holleley, 2021 Ovotestes suggest cryptic genetic influence in a reptile model for temperature-dependent sex determination. *Proceedings of the Royal Society B* 288 (1943):20202819.

Woo, K., and G. Rieucau, 2013 Efficiency of aggressive and submissive visual displays against environmental motion noise in Jacky dragon (Amphibolurus muricatus). *Ethology Ecology & Evolution* 25 (1):82-94.

Xue, W., J.T. Li, Y.P. Zhu, G.Y. Hou, X.F. Kong *et al.*, 2013 L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* 14:604.

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24 (8):1586-1591.

Zhu, B.H., J. Xiao, W. Xue, G.C. Xu, M.Y. Sun *et al.*, 2018 P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. *BMC Genomics* 19 (1):175.

Figure legends

Figure 1 Photograph of an adult male jacky dragon (*Amphibolurus muricatus*). Image credit: David Cook Wildlife Photography.

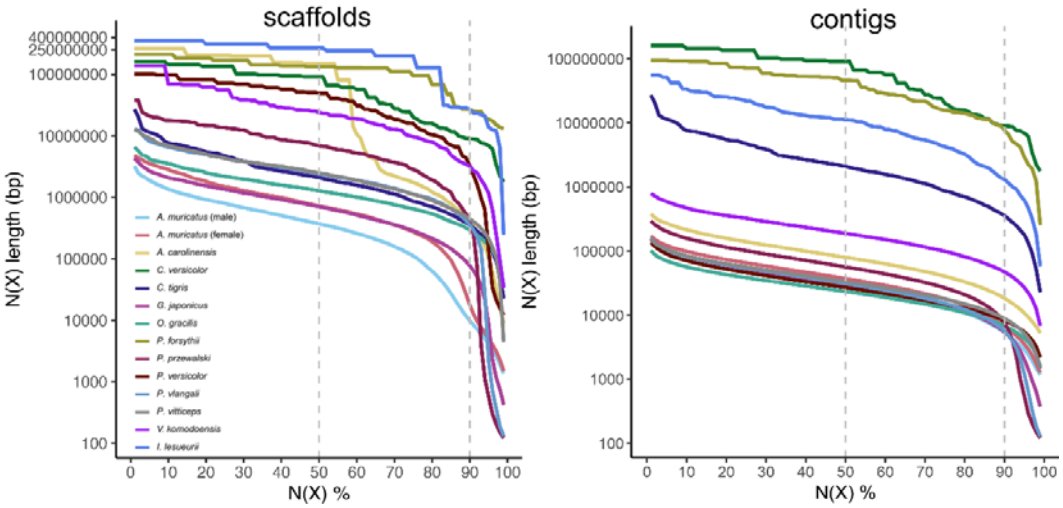

Figure 2 Comparison of the contiguity of two *A. muricatus* assemblies and 12 publicly available squamate assemblies. N(x)% graphs show the (A) contig and (B) scaffold lengths (*y*-axis), where x% (*x*-axis) of the genome assembly consist of scaffolds and contigs of at least that size. Dashed, grey lines denote N50 and N90 values. AmpMurF_3.0 and AmpMurM_3.0. denotes the female and male *A. muricatus* assembly, respectively

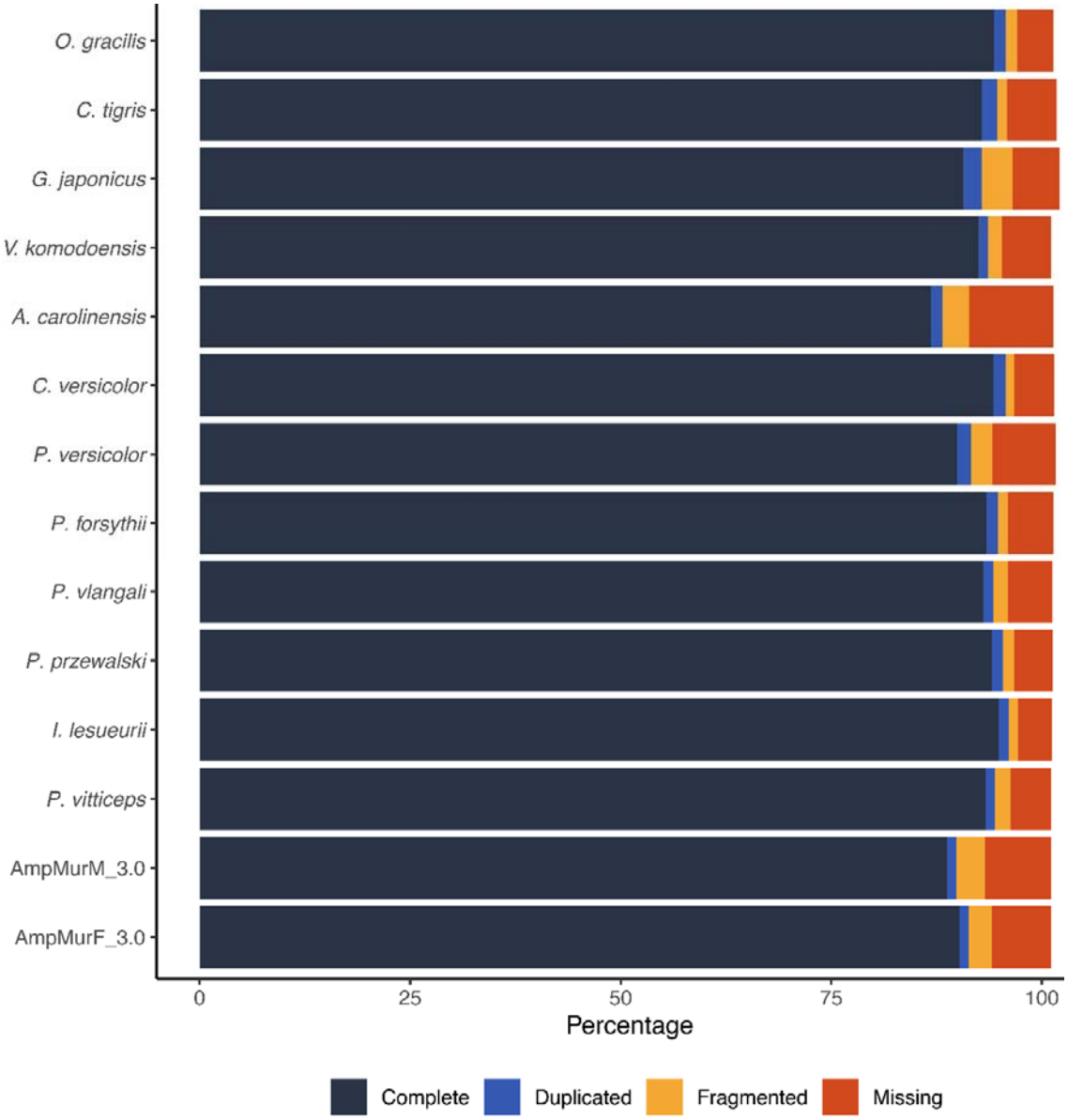

Figure 3 BUSCO assessment of genome assemblies from 13 squamate species. All assemblies were examined using the same version and library of BUSCO (v5.0.0 with the 7,480-gene sauropsida_odb10 dataset). AmpMurF_3.0 and AmpMurM_3.0. denotes the female and male *A. muricatus* assembly, respectively.

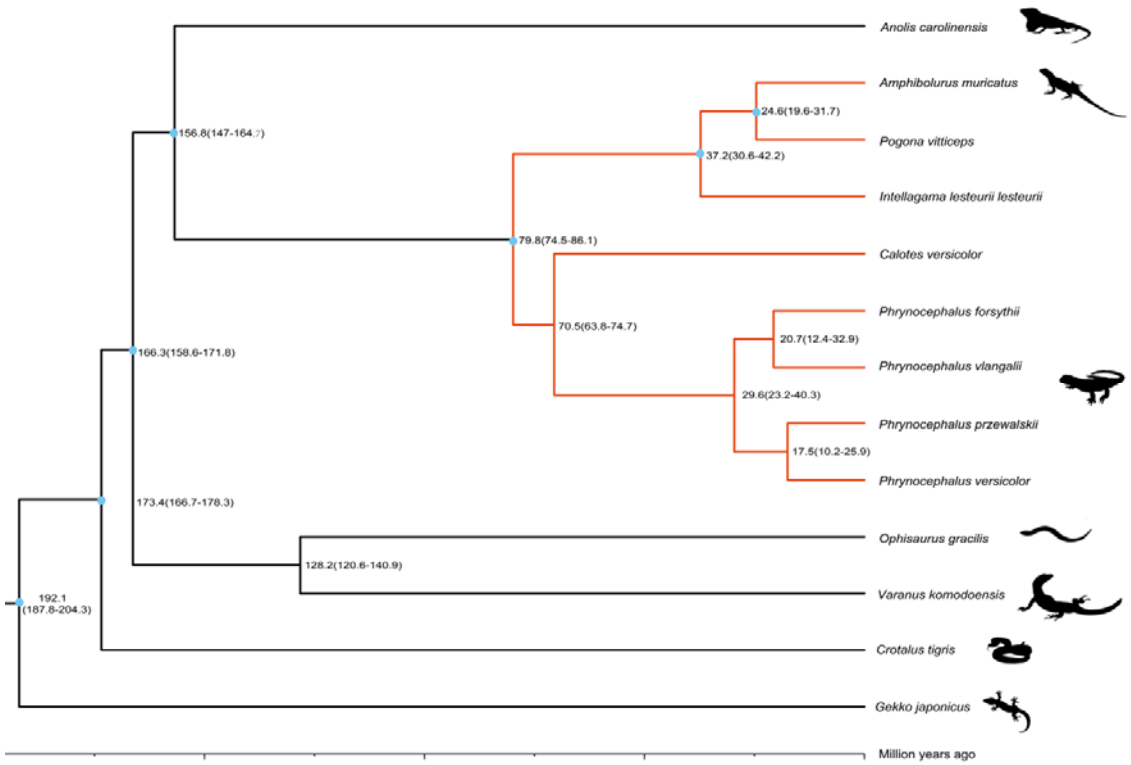

Figure 4 Inferred phylogeny of 13 squamate species based on whole-coding sequences of 3,130 1:1 orthologs. Numbers at nodes represent the estimated divergence time from present (million years ago; Mya) between lineages. Agamid (family Agamidae) lineages are indicated in red.

Tables

Table 1 Agamidae genome assembly statistics.
Lengths in base pairs (bp). AmpMurF_3.0 and AmpMurM_3.0 denotes the female and male *Amphibolurus muricatus* assemblies, respectively. NA denotes not applicable. Assemblies with the suffix _HiC are available at DNA Zoo; CalVer_1.0 (CNP0003598), Phprz_v1.0 (CNP0000203), and Phvla_v1.0 (CNP0000203) are available at the China National GeneBank Database; the remainder are available via NCBI.

Species name	*Amphibolurus muricatus*	*Amphibolurus muricatus*	*Pogona vitticeps*	*Intellagama lesueurii*	*Phrynocephalus przewalskii*	*Phrynocephalus vlangalii*	*Phrynocephalus forsythii*	*Phrynocephalus versicolor*	*Calotes versicolor*
Common name	Jacky dragon	Jacky dragon	Central bearded dragon	Australian water dragon	Przewalski's toadhead agama	Ching Hai toadhead agama	Moustache toadhead agama	Variegated toadhead agama	Oriental garden lizard
Assembly	AmpMurF_3.0	AmpMurM_3.0	pvi1.1	EWD_hifiasm_HiC	Phprz_v1.0	Phvla_v1.0	ASM2928247v1	PhrVer	CalVer_1.0
Reference	This study	This study	Georges *et. al* (2015)	NA	Gao *et al.*, 2019	Gao *et al.*, 2019	Qi *et al.* 2023	Jin *et al.* 2022	Wang *et al.* 2023
Contig number	145,095	95,472	636,780	1,457	592,043	506,434	944	102,689	104
Contig length	1,804,035,661	1,752,355,218	1,747,541,145	1,816,116,528	1,738,531,458	1,773,991,985	1,816,858,791	1,592,519,988	1,614,101,240
Contig N50 (bp)	67,166	59,294	31,297	11,213,063	56,431	31,286	46,215,561	27,024	91,599,063
Contig max length	645,479	773,372	295,776	55,971,000	547,676	368,251	95,027,530	275,221	160,793,834
Scaffold number	97,556	45,762	545,310	619	543,179	411,697	644	4,557	104
Scaffold length	1,868,109,324	1,804,786,947	1,816,116,151	1,816,535,528	1,763,820,185	1,820,170,410	1,816,888,593	1,603,387,288	1,614,101,240
Scaffold N50 (bp)	6,818,063	1,568,728	2,290,546	268,945,440	6,884,165	2,396,942	133,255,571	49,226,030	91,599,063
Scaffold max length	39,679,044	9,582,854	14,681,335	353,375,234	37,740,249	13,874,271	212,878,235	102,350,298	160,793,834
Gaps (bp)	64,073,663	52,431,729	68,575,006	419,000	25,288,727	46,178,425	29,802	10,867,300	0
Gaps (%)	3.43	2.91	3.78	0.02	1.43	2.54	0.0016	0.68	-
GC content (%)	41.75	41.67	41.87	41.94	42.94	43.07	43.32	42.91	43.79