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

Ran Tian[1,2], Han Guo[1], Chen Yang[1], 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, Canberra 2617, Australia
[6] Australian National Wildlife Collection, CSIRO, Canberra 2601, Australia
[7] School of Biology and Environmental Science, Queensland University of Technology, Brisbane 4000, Australia
[8] AusARG consortium, Bioplatforms Australia, Research Park Drive, Macquarie University, Macquarie Park 2109, Australia

[*]Corresponding author: Institute for Applied Ecology, University of Canberra, ACT 2601, Australia. E-mail: georges@aerg.canberra.edu.au

ABSTRACT

Australia is remarkable for its lizard diversity, with very high endemicity because of continental-scale diversification and adaptive radiation. We employed 10X Genomics Chromium linked-reads technology to generate male and female draft genomes of the jacky dragon (*Amphibolurus muricatus*), an Australian dragon lizard (family Agamidae). The assemblies are 1.8 Gb in size and have a repeat content (38%) and GC content (42%) similar to other dragon lizards. The contig N50 values for the assemblies were 37.2 kb (female) and 28.8 kb (male), with corresponding scaffold N50 values of 720.5 kb and 369 kb. The longest scaffold was 6.5 Mb in each assembly. The BUSCO completeness percentages were 92.2% and 90.8% respectively. These statistics are comparable to other lizard genomes assembled using similar technology. Phylogenetic comparisons show that Australian dragon lizard species split from a common ancestor around 33.4 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 lizard 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 and Shine, 2008; Warner et al., 2013), social behaviour (Peters and Evans, 2003; Woo and Rieucau, 2013), and development (Esquerré et al., 2014; Whiteley et al., 2021).

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 employ high-throughput linked-read sequencing (10X Genomics) (Zheng et al., 2016) to generate draft, annotated genome assemblies for a male and a female *A. muricatus* that are comparable in contiguity and completeness to other published Agamid genomes generated using Illumina short-read technologies (Georges et al., 2015). We used transcriptomes sequenced and assembled for *A. muricatus* and published assemblies (*Anolis*, *Varanus*, *Pogona*) to annotate the genomes. 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

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) individual 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.

Linked-read whole-genome sequencing and de novo assembly

High molecular weight DNA was extracted from liver (female) and blood (male) using the Gentra Puregene DNA Isolation kit (Qiagen). DNA yield and quality was assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and a Qubit fluorometer (Thermo Fisher Scientific). 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.

To estimate the genome sizes and heterozygosity of the female and male *A. muricatus* samples, we used GenomeScope (Vurture et al., 2017). Briefly, we employed KmerGenie v1.7051 (Chikhi and Medvedev, 2014) and short-read sequencing data (paired-ends reads generated above) to determine the optimal *k*-mer-value range, indicating that a minimum *k*-mer length of 31 and 41 for the male and female, respectively. Next, *k*-mers in the paired-end sequencing reads were counted and converted to histogram files using Jellyfish (v2.2.10) (Marcais and Kingsford, 2011), followed by analysis using GenomeScope v1.0.

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 utilised (female sequencing depth *ca* 50.3*x*; male, *ca* 47.8*x*). The resulting assembly was 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. Next, 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 assembly 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 assembly using L_RNA_scaffolder (Xue et al., 2013). 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 cleaned 10X Genomics reads PE reads. Genome assemblies were assessed using BUSCO 5.0.0_cv1 (Seppey et al., 2019).

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 (Dodt et al., 2012; Roehr et al., 2017) with default settings (removes 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 (Grabherr et al., 2011; Haas et al., 2013; Henschel et al., 2012) and assessed using BUSCO 5.0.0_cv1 (Seppey et al., 2019).

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) (Tarailo-Graovac and Chen, 2009) and RepeatProteinMask (protein-level; implemented in RepeatMasker). De novo TE predictions

were obtained using RepeatModeler v1.1.0.4 (Smit and Hubley, 2010) and LTRharvest v1.5.8 (Ellinghaus et al., 2008) to generate database for a RepeatMasker run. Tandem Repeat Finder (v4.07) (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 (cleaned 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), TrEMBL (Translation of EMBL [nucleotide sequences that are not in Swiss-Prot]; release-2018_07) (O'Donovan et al., 2002), 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 was 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 (Cheng et al., 2021; Dudchenko et al., 2017; Dudchenko et al., 2018)) and the Chinese agamid lizards *Phrynocephalus przewalskii* (Przewalski's toadhead agama) (Gao et al., 2019) and *Phrynocephalus 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 4,242 high-confidence 1:1 orthologs by interrogating the predicted proteins from the gene models of ten 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 4,242 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 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 GenomeScope (Vurture et al., 2017) to obtain a *k*-mer-based estimate of genome size and heterozygosity from paired-end sequence reads. The size of the *A. muricatus* genome is estimated to be around 1.8 Gb (**Figure S2**). The genome-wide heterozygosity from our inbred *A. muricatus* lines was estimated to range from 0.58% (female) to 0.78% (male) (**Figure S2**), slightly lower than the central bearded dragon (*Pogona vitticeps*) (0.85%) (Georges et al., 2015).

10X Genomics sequencing data were assembled into contigs, then oriented and merged into scaffolds. Combining ARKS and P_RNA_scaffolder (employs RNA-seq reads from brain, ovary, and testis) (**Table S1**) and L_RNA_scaffolders (employs Trinity transcriptome assemblies) (**Tables S2** and **S3**), the scaffold N50 size of the female *A. muricatus* (named AmuF1.1) assembly improved by 38.3% (from 371.5 kb to 720.5 kb), while the BUSCO completeness percentage improved from 84.1% to 92.2% (**Table S4**). The male assembly (AmuM1.1) saw a similar improvement, with the scaffold N50 and BUSCO score increasing from 180.2 kb to 369.9 kb and from 78.9% to 90.8%, respectively. The final assemblies had a total scaffold length (contain gaps) of 1.84 Gb (female) and 1.83 Gb (male) with a longest scaffold *ca* 6.5 Mb in each assembly (**Table 1**). The contig N50 values for the assemblies were 37.2 kb (female) and 28.8 kb (male). The scaffold and contig N50 values are similar to those of other squamate genome assemblies (**Figure 2**), with the exception of the chromosome-assigned assemblies of Australian water dragon (*Intellagama lesueurii*; scaffold N50 268.9 Mb), tiger rattlesnake (*Crotalus tigris*; scaffold N50 2.1 Mb) (Margres et al., 2021), green anole (*Anolis carolinensis*; scaffold N50 150.1 Mb), and Komodo dragon (*Varanus komodoensis*; scaffold N50 23.8 Mb) (Lind et al., 2019). The BUSCO metrics of the *A. muricatus* assemblies indicate that they compare well to other squamate assemblies, including agamids from Australia [*P. vitticeps* (Georges et al., 2015) and *I. lesueurii* (Australian water dragon)] and China (toad-headed agamas of genus *Phrynocephalus* sp. (Gao et al., 2019)) (**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 S4 and S6**), 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 S8**) (see (Georges et al., 2015)) and gene set BUSCO scores (**Table S9**) are comparable to other squamates. Using ab inito, transcriptome, and homology-based prediction methods, we functionally annotated 18,197 (85.0%) and 17,360 (88.0%) protein-coding genes in the female and male assembly (**Tables S10 and S11**) and recovered 94.7% and 93.1% of 3,354 vertebrate benchmarking universal single-copy orthologs (BUSCOs) (Seppey et al., 2019), respectively.

Phylogenetic relationships

To construct a time-calibrated species tree (**Figure 4**), we identified 4,242 high-confidence single-copy orthologs from the female *A. muricatus* assembly and nine other squamate species. There are currently five agamid lizard genome assemblies: three Australian dragon lizard assemblies (*A. muricatus*, *P. vitticeps*, and *I. lesueurii*) and two toad-headed agama assemblies (genus *Phrynocephalus*) (Gao et al., 2019; Georges et al., 2015). Our analysis shows that the five agamid species shared an ancestor about 85.7 Mya [81.3-88.2 Mya 95% credibility interval (CI)]. We estimate that the three Australian dragon lizard species split from a common ancestor about 33.4 Mya (95% CI 28.8-39.1), while the lineages leading to *A. muricatus* and *P. vitticeps* diverged 23.6 Mya (95% CI 19.4-28.2). These observations are in agreement with an appraisal from a small set of mitochondrial and nuclear genes (Hugall et al., 2008).

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 researchers. Nevertheless, it is appreciated that the assemblies can be further improved. Such efforts will be particularly important for future studies on squamate chromosome evolution and sex determination. Single-tube Long Fragment Read (stLFR) sequencing (Fan et al., 2019; Wang et al., 2019) and re-scaffolding using SLR-superscaffolder (Guo et al., 2021), and chromatin conformation capture (Hi-C) (Lieberman-Aiden et al., 2009) sequencing to generate a chromosome-level assembly is planned by the AusARG initiative of Bioplatforms Australia (https://bioplatforms.com). In conclusion, the *A. muricatus* assemblies described here should prove valuable for the understanding of agamid lizard evolution, ecology, and conservation.

Data availability

A. muricatus raw genome and transcriptome reads have been deposited to the NCBI Short Read Database (BioProject ID: PRJNA767251). The male and female *A. muricatus* assemblies are available at Zenodo (Tian et al., 2021b). Gene annotation files and associated FASTA files for *A. muricatus* (assembly AmuF1.1), *I. lesueurii*, *P. przewalskii*, and *P. vlangalii* are available at Zenodo (Tian et al., 2021c). *A. muricatus* transcriptome assemblies are available at Zenodo (Tian et al., 2021a). 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.), and the Jiangsu Foreign Expert Bureau (to I.S.).

Literature cited

Alfoldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C.B., Glor, R.E., Jaffe, J.D., *et al.* (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature *477*, 587-591.

Bao, W., Kojima, K.K., and Kohany, O. (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*, 573-580.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*, 988-995.
Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol *268*, 78-94.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods *18*, 170-175.
Chikhi, R., and Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. Bioinformatics *30*, 31-37.
Cogger, H. (2014). Reptiles and amphibians of Australia (CSIRO publishing).
Coombe, L., Zhang, J., Vandervalk, B.P., Chu, J., Jackman, S.D., Birol, I., and Warren, R.L. (2018). ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics *19*, 234.
Cosentino, S., and Iwasaki, W. (2019). SonicParanoid: fast, accurate and easy orthology inference. Bioinformatics *35*, 149-151.
Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) *1*, 895-905.
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., *et al.* (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92-95.
Dudchenko, O., Shamim, M.S., Batra, S.S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., St Hilaire, B.G., Yao, W., and Stamenova, E. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. BioRxiv, 254797.
Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics *9*, 18.
Esquerré, D., Keogh, J.S., and Schwanz, L.E. (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.
Fan, H., Wu, Q., Wei, F., Yang, F., Ng, B.L., and Hu, Y. (2019). Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. Genome Biol *20*, 267.
Gao, W., Sun, Y.B., Zhou, W.W., Xiong, Z.J., Chen, L., Li, H., Fu, T.T., Xu, K., Xu, W., Ma, L., *et al.* (2019). Genomic and transcriptomic investigations of the evolutionary transition from oviparity to viviparity. Proc Natl Acad Sci U S A *116*, 3646-3655.
Georges, A., Li, Q., Lian, J., O'Meally, D., Deakin, J., Wang, Z., Zhang, P., Fujita, M., Patel, H.R., Holleley, C.E., *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., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644-652.
Guo, L., Xu, M., Wang, W., Gu, S., Zhao, X., Chen, F., Wang, O., Xu, X., Seim, I., Fan, G., *et al.* (2021). SLR-superscaffolder: a de novo scaffolding tool for synthetic long reads using a top-to-bottom scheme. BMC Bioinformatics *22*, 158.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., *et al.* (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc *8*, 1494-1512.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol *9*, R7.

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

Henschel, R., Lieber, M., Wu, L.-S., Nista, P.M., Haas, B.J., and LeDuc, R.D. (2012). Trinity RNA-Seq assembler performance optimization. Paper presented at: Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond (Chicag, IL, USA: Association for Computing Machinery).

Hugall, A.F., Foster, R., Hutchinson, M., and Lee, M.S. (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*, 343-358.

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

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

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

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

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lind, A.L., Lai, Y.Y.Y., Mostovoy, Y., Holloway, A.K., Iannucci, A., Mak, A.C.Y., Fondi, M., Orlandini, V., Eckalbar, W.L., Milan, M., *et al.* (2019). Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. Nat Ecol Evol *3*, 1241-1252.

Liu, Y., Zhou, Q., Wang, Y., Luo, L., Yang, J., Yang, L., Liu, M., Li, Y., Qian, T., Zheng, Y., *et al.* (2015). Gekko japonicus genome reveals evolution of adhesive toe pads and tail regeneration. Nat Commun *6*, 10033.

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

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

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

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

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

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol *37*, 1530-1534.
Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., *et al.* (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res *47*, D351-D360.
O'Donovan, C., Martin, M.J., Gattiker, A., Gasteiger, E., Bairoch, A., and Apweiler, R. (2002). High-quality protein knowledge resource: SWISS-PROT and TrEMBL. Brief Bioinform *3*, 275-284.
O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., *et al.* (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res *44*, D733-745.
Pepper, M., Barquero, M.D., Whiting, M.J., and Keogh, J.S. (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 Evans, C.S. (2003). Introductory tail-flick of the Jacky dragon visual display: signal efficacy depends upon duration. Journal of Experimental Biology *206*, 4293-4307.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glockner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res *41*, D590-596.
Roehr, J.T., Dieterich, C., and Reinert, K. (2017). Flexbar 3.0 - SIMD and multicore parallelization. Bioinformatics *33*, 2941-2942.
Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol *1962*, 227-245.
Smit, A.F., and Hubley, R. (2010). RepeatModeler Open-1.0. 2008-2015. Available at http://www.repeatmasker.org.
Song, B., Cheng, S., Sun, Y., Zhong, X., Jin, J., Guan, R., Murphy, R., Che, J., Zhang, Y., and Liu, X. (2015a). Anguidae lizard (Ophisaurusgracilis) genome assembly data. Dryad Digital Repository doi *10*, 100119.
Song, B., Cheng, S., Sun, Y., Zhong, X., Jin, J., Guan, R., Murphy, R.W., Che, J., Zhang, Y., and Liu, X. (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*, 2688-2690.
Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res *34*, W435-439.
Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol *56*, 564-577.
Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics *Chapter 4*, Unit 4 10.
Tian, R., Guo, H., Yang, C., Fan, G., Whiteley, S.L., Holleley, C.E., Seim, I., and Georges, A. (2021a). 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., Guo, H., Yang, C., Fan, G., Whiteley, S.L., Holleley, C.E., Seim, I., and Georges, A. (2021b). Draft de novo genome assemblies of a male and female Amphibolurus muricatus (jacky dragon), https://doi.org/10.5281/zenodo.5523788 (Zenodo).
Tian, R., Guo, H., Yang, C., Fan, G., Whiteley, S.L., Holleley, C.E., Seim, I., and Georges, A. (2021c). 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.5523656 (Zenodo).
UniProt Consortium (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res *40*, D71-75.
Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202-2204.
Wang, O., Chin, R., Cheng, X., Wu, M.K.Y., Mao, Q., Tang, J., Sun, Y., Anderson, E., Lam, H.K., Chen, D., *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*, 798-808.
Warner, D., and Shine, R. (2008). The adaptive significance of temperature-dependent sex determination in a reptile. Nature *451*, 566-568.
Warner, D.A., Uller, T., and Shine, R. (2013). Transgenerational sex determination: the embryonic environment experienced by a male affects offspring sex ratio. Scientific reports *3*, 1-4.
Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J., and Birol, I. (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience *4*, 35.
Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. (2017). Direct determination of diploid genome sequences. Genome Res *27*, 757-767.
Whiteley, S.L., Georges, A., Weisbecker, V., Schwanz, L.E., and Holleley, C.E. (2021). Ovotestes suggest cryptic genetic influence in a reptile model for temperature-dependent sex determination. Proceedings of the Royal Society B *288*, 20202819.
Woo, K., and Rieucau, G. (2013). Efficiency of aggressive and submissive visual displays against environmental motion noise in Jacky dragon (Amphibolurus muricatus). Ethology Ecology & Evolution *25*, 82-94.
Xue, W., Li, J.T., Zhu, Y.P., Hou, G.Y., Kong, X.F., Kuang, Y.Y., and Sun, X.W. (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*, 1586-1591.
Zheng, G.X., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., *et al.* (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol *34*, 303-311.
Zhu, B.H., Xiao, J., Xue, W., Xu, G.C., Sun, M.Y., and Li, J.T. (2018). P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. BMC Genomics *19*, 175.

Figure legends

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

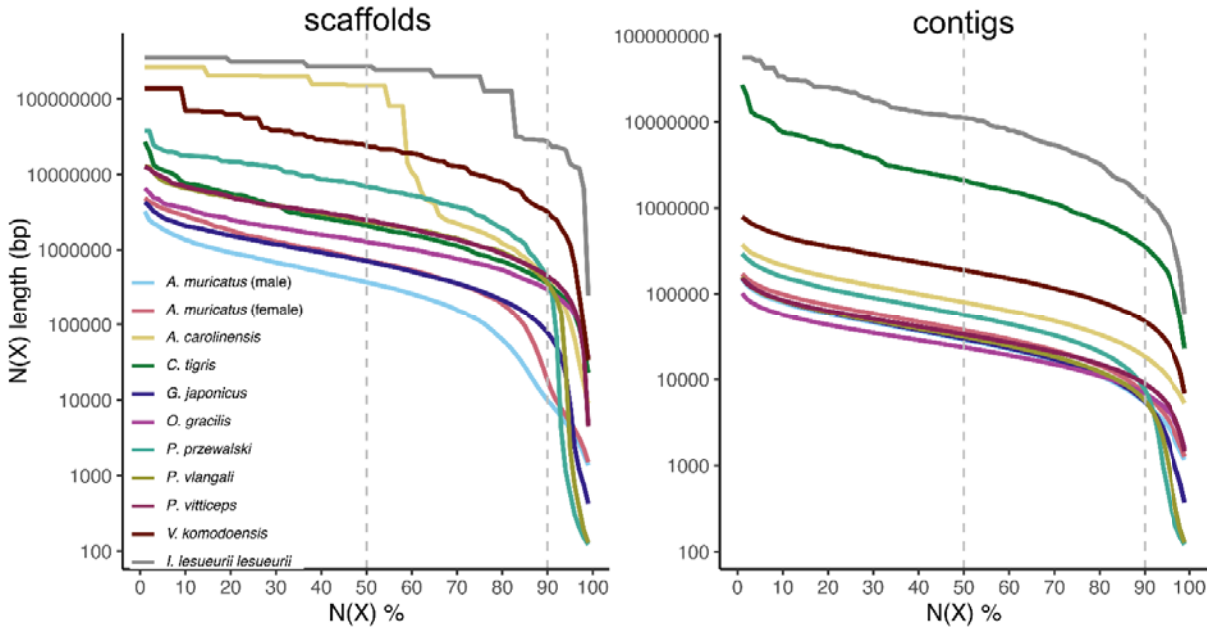

Figure 2 Comparison of the contiguity of two *A. muricatus* assemblies and nine publicly available squamate assemblies. N(x)% graphs show the (A) contig and (B) scaffold lengths (*y*-axis), where x% (*x*-axis) of the genome assembly consist of scaffolds and contigs of at least that size. Dashed, grey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

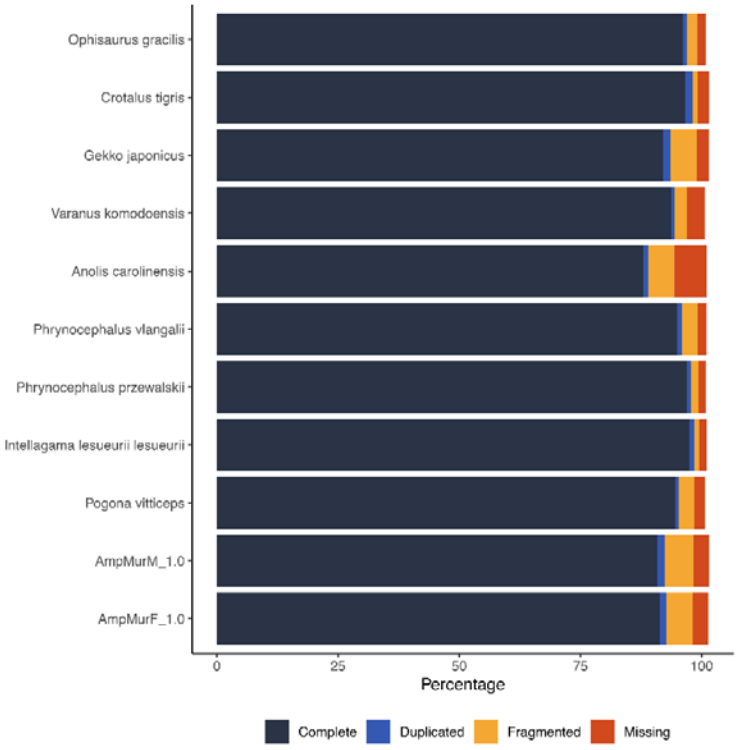

Figure 3 BUSCO assessment of assemblies from ten squamate species. All genome assemblies were examined using the same version and library of BUSCO (5.0.0_cv1 with the 3,354-gene vertebrata_odb10 dataset). AmuF_1.1 and AmuM_1.1 denotes the female and male *A. muricatus* assembly, respectively.

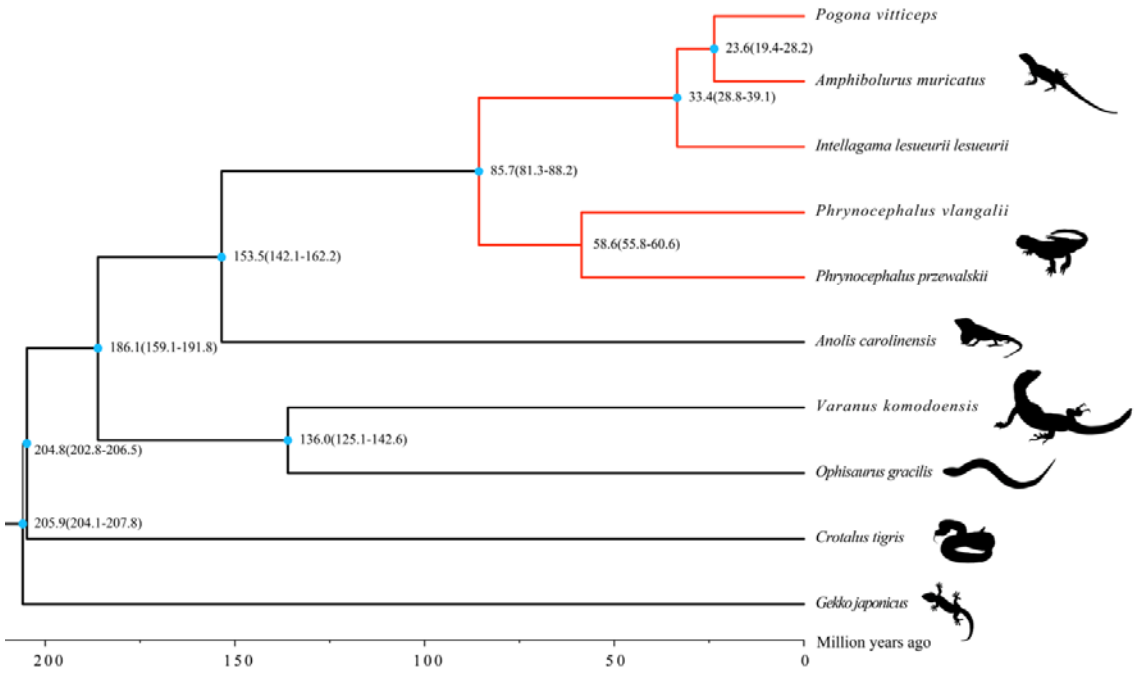

Figure 4 Inferred phylogeny of ten squamate species based on whole-coding sequences of 4,242 1:1 orthologs. Numbers at nodes represent the estimated divergence time from present (million years ago; Mya) between lineages. Agamid (family Agamidae) lineages are indicated in red.

Tables

Table 1 *A. muricatus* genome assembly statistics. Lengths in base pairs (bp).

Assembly methods	**Female (AmuF.1.1)**	**Male (AmuM.1.1)**
Contig number	124,200	151,787
Contig length	1,750,545,991	1,741,048,453
Contig N50 (bp)	37,220	28,761
Contig max length	348,284	288,200
Scaffold number	57,227	73,856
Scaffold length	1,841,491,868	1,833,283,242
Scaffold N50 (bp)	720,518	369,860
Scaffold max length	6,534,950	6,446,322
Gaps (bp)	90,945,877	92,234,789
Gaps (%)	4.94	5.03
GC content (%)	41.77	41.70