Tail regeneration in the Tokay Gecko (Gekko gecko)

Nurhidayat, L.

Chapter 6. A High-Quality Genome of The Tokay Gecko (*Gekko gecko*)

Luthfi Nurhidayat[1,2], Vladimir Benes[3], Herman P. Spaink[1], Michael K. Richardson[1]

1. Institute of Biology Leiden, Leiden University, Sylvius Laboratory, Sylviusweg 72, 2333 BE, Leiden, The Netherlands.
2. Faculty of Biology, Universitas Gadjah Mada, Jalan Teknika Selatan Sekip Utara, 55281, Yogyakarta, Indonesia
3. Genomics Core Facility, European Molecular Biology Laboratory Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany

---

Manuscript in preparation.

Abstract

The tokay gecko is an ideal model organism for studying regeneration biology, as it can regenerate its tail, which is a complex structure. Sequencing of the tokay gecko genome holds great potential for advancing research on regeneration, including therapeutic applications, and on other biological research that uses the tokay gecko as a model. In this chapter, we describe a high-quality, scaffold-level genome assembly for the tokay gecko using a hybrid approach combining Illumina and Nanopore sequencing. The assembly has 4,208 scaffolds with an N50 of 190.8 Mbp and L50 of 6, and it reaches 94.8% completeness of BUSCO genes. By integrating two annotation pipelines, BRAKER2 and funannotate, we identified 27,350 genes with 172,817 distinct exons and 30,229 predicted transcripts. As a reference, we also asked Ensembl run our genome assembly through their genebuild pipeline. This predicted 22,009 genes with 35,820 transcripts built from 232,387 exons. Our preliminary synteny analysis showed the tokay gecko assembly is comparable in completeness and contiguity to chromosome-level assemblies of both Townsend's least gecko *Sphaerodactylus townsendi* and the leopard gecko *Eublepharis macularius*. In summary, we were able to produce a high-quality, scaffold-level genome assembly with relatively high-quality annotation. This genome annotation needs some improvements, including reducing fragmentation and duplications.

Introduction

The tokay gecko (*Gekko gecko*) is a lizard belonging to the Gekkonidae, and known for its distinctive appearance with vibrant blue-gray skin covered with orange or red spots (Kongbuntad et al., 2016). Native to Southeast Asia and the Western Pacific, tokay geckos are nocturnal and arboreal, and are commonly found in urban areas, forests, and agricultural lands (Das, 2015). They exhibit territorial behavior and possess adhesive toe pads that enable them to climb vertical surfaces with ease (Cobos and Higham, 2022).

One of the most remarkable features of tokay geckos and some other lizards is their ability to undergo tail autotomy, the voluntary shedding of their tails as a defense mechanism against predators (Rumping and Jayne, 1996). Upon autotomy, the detached tail continues to wriggle, diverting the predator's attention while the gecko escapes. Following tail loss, tokay geckos can regenerate a functional tail, complete with bone, muscle, skin, and nerves. This process involves the formation of a specialized structure called the blastema, which gives rise to the replacement tail unlike axolotls which can regenerate their limbs and tails (Franklin et al., 2017; McCusker et al., 2015; McHedlishvili et al., 2012; Polikarpova et al., 2022), lizards only regenerate their tails (Alibardi, 2019; Lozito and Tuan, 2017). Humans and other mammals have a much more limited regeneration capacity. As an amniote, geckos are closer phylogenetically to mammals than are axolotls and as this thesis argues, may be a better model for human translation research in regeneration biology.

By investigating the cellular and molecular processes involved in tail regeneration, researchers can uncover key mechanisms that drive tissue regeneration in vertebrates. Additionally, tokay geckos have a number of advantages as experimental animals. The tokay gecko has an excellent adaptation to the urban environment as shown by the low genetic diversity in Indonesia despite the geographic isolation of the islands (Kadafi et al., 2024). This means they are relatively easy to maintain in laboratory settings, making them accessible for experimental studies. They also have a larger body size compared to the other lizards, which in our experience makes it easier to sample tissues and perform experiments. Furthermore, they have the advantage of being are oviparous (egg-laying) with the freshly-laid egg containing a very early stage of embryo. Finally, they are able to regenerate multiple times throughout their lifespan.

Several lizard genomes have been sequenced and their implications for regeneration research explored. For instance, a study of Schlegel's Japanese Gecko (*Gekko japonicus*) genome, a species closely related to the tokay gecko, described the evolution of gene families potentially involved in tail regeneration (Liu et al., 2015). Additionally, those genome data have provided insights into the evolution of adhesive toe pads, and the sex determination mechanisms (Liu et al., 2015). In addition, the lizard genomes give a perspective on evolutionary aspect of high-altitude adaptation in lizards

(Li et al., 2023), population structure (Davalos-Dehullu et al., 2023), chromosome transition (Pinto et al., 2022), and asymmetric fates of duplicated genes (Hara et al., 2018).

Genome sequencing of the tokay gecko holds immense potential for advancing research on regeneration and other biological processes. It may also provide valuable insights into the genetic basis of regeneration. Thus, by identifying genes and regulatory elements associated with tail regeneration, researchers can unravel the molecular pathways underlying this phenomenon. Furthermore, genome sequencing facilitates comparative genomics studies, allowing researchers to understand the evolution of geckos, as well as the conservation and divergence of regenerative mechanisms across species. Ultimately, a fully annotated genome of the tokay gecko will serve as a foundation for future studies aiming to translate them to therapeutic applications in humans and other organisms.

Materials and Methods

Ethics statement

The tokay gecko (*Gekko gecko*) is listed as 'least concern' in the International Union for Conservation of Nature *Red List* (IUCN, https://www.iucnredlist.org/species/195309/2378260) and also listed in Appendix II in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES; https://cites.org/eng/taxonomy/term/3919). In Indonesia, the tokay gecko is not a protected species. All animals were collected in Yogyakarta, Indonesia under a license issued by the Ministry of Environment and Forestry, the Republic of Indonesia (permit number SK.83/KSDAE/SET/KSA.2/5/2021 signed on 7 May 2021) and a recommendation letter issued by The Indonesian Institute of Sciences (Lembaga Ilmu Pengetahuan Indonesia/LIPI; Recommendation Number: B- 2158 /IV/KS.01.04/3/2021 signed on 19 March 2021). The transport of samples from Indonesia to the Netherlands was conducted under CITES permit numbers 05717/IV/SATS-LN/2021 and 09160/IV/SATS-LN/2022 issued by the Ministry of Environment and Forestry, the Republic of Indonesia.

All experimental and surgical procedures needed for samples collection were done at Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia with the approval from The Ethical Committee of the Integrated Laboratory for Research and Testing (Laboratorium Penelitian dan Pengujian Terpadu/LPPT) Universitas Gadjah Mada (Ethical Clearance number: 00014/04/LPPT/IV/2021 signed on 30 April 2021). Tail autotomy was performed under ketamine anesthesia.

Animal collection and maintenance

Thirty-nine adult tokay geckos were captured in the vicinity of the Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia. They were acclimatized for two weeks,

two males with one female, in a stainless-steel terrarium and fed various species of live crickets and sometimes cockroaches, twice a day. They were maintained at ambient temperature and sprayed once a day with water. The geckos were also basked under natural sunlight once every two days for 1 h in the period 09:00–11:00. Twenty-one were used for the harvesting of regeneration blastemas under anesthesia. Of these, three were euthanized for the harvesting of internal organs. The remaining 18 geckos were bred to produce eggs for embryo sampling.

Experimentation and sample collection

Regenerated tail samples

Twenty-one adult tokay geckos were anaesthetized with intramuscular ketamine injection (50–75 mg/Kg body weight) administered in the thigh prior to the autotomy procedure. Autotomy was performed by gently restraining the gecko manually, while leaving the tail accessible. Using the thumb and index finger, a firm pinch was applied to the required autotomous region (every segment after the 5th caudal vertebra) of the tail causing the gecko to spontaneously shed its tail. The wound was painted with Betadine® iodine antiseptic solution but no dressing was applied. After the autotomy procedure was performed, the geckos were allowed to regenerate a new tail. The geckos were distributed into experimental groups based on the number of days post-autotomy, each consisting of three geckos. After ketamine-anesthesia, the tails of geckos in each group were re-autotomized 2 cm cranially from the previous autotomy site at 4 and 16 dpa (days post-autotomy). The regenerated tail samples were rinsed and further handled in phosphate-buffered saline (PBS, 4 °C) and subsequently fixed with cold RNA*later*® stabilization solution (Invitrogen).

Embryo samples

Eighteen adult tokay geckos were grouped and bred as described above. The terraria were checked daily for the presence of gecko eggs. The eggs were left in the terrarium for the required number of days of incubation at ambient temperature. The whole embryo stage 16 dpo (days post-oviposition) and embryonic tail of 22 dpo were harvested into cold PBS and subsequently fixed with RNA*later*® (4 °C).

Organ samples

Three adult tokay geckos were euthanized by an overdose of intramuscular ketamine (150 mg/Kg body weight, intramuscular injection). They were then dissected to harvest organs, namely: pancreas, hearth, muscle, brain, skin, kidney, liver and lung. These eight organ samples were cut up in cold PBS (phosphate-buffered saline) and stored at 4 °C in RNA*later*® stabilization solution (Invitrogen).

RNA extraction and sequencing

RNA isolation and cleaning

The tissues and organs used for RNA sequencing were pancreas, heart, muscle, brain, skin, kidney, liver, lung, whole embryo (16 dpo), embryonic tail (22 dpo), and regenerating tail (4 dpa and 16 dpa). The samples were taken from RNA*later*® solution with sterile forceps and rinsed with sterile PBS. They were then put in 1.5 mL Eppendorf tubes together with 1 mL Trizol and two steel balls (⌀ 4 mm). The tubes were then shaken with a TissueLyser II (Qiagen) at 30/s for 1–2 min. The resulting tissue homogenates were triturated with a hypodermic needle (21G = 0.8mm) attached to a 1 mL syringe until no tissue clumps were visible. The homogenates were then incubated for 30 min and subsequently centrifuged for 10 min at 12,000 RCF (relative centrifugal force).

The supernatants were transferred into new tubes and 0.2 mL of chloroform added to each. The mixture in the tubes were incubated for 2 to 3 min after 11 s of vigorous hand shaking and then centrifuged for 15 min at 16,000 RCF. The colorless aqueous phase of each mixture was transferred to a fresh tube, and 0.7 mL of propan-2-ol (isopropyl alcohol, isopropanol) added to each tube. After leaving for 10 min at room temperature, the samples were centrifuged at 16,000 RCF for 15 min. The supernatants were discarded, and the RNA pellet washed with 1 mL of 75% ethanol. The pellet with 75% ethanol was centrifuged again for 5 min at 7,500 RCF. The supernatants were discarded, and the RNA pellets allowed to dry for 5–10 min. The dry RNA pellets were dissolved in 50 µL of RNase-free water and incubated at 55° C for 10 min. The total RNA of each sample was then cleaned with Qiagen RNeasy Mini Kit (Cat. No. 74104) according to the manufacturer's protocols.

RNA sequencing

Library preparation and transcriptome sequencing were done by the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. The sequencing platform used was Illumina NextSeq 2000 with a read length of 2×100 bp. The RNA sequencing results were used to assist genome annotation.

Genomic DNA extraction and sequencing

We combined two sequencing technologies: Illumina sequencing technology for short reads and Oxford Nanopore technology for long reads. This hybrid approach was in order to get an optimal quality of assembly considering the lower sequencing errors of Illumina, and the longer reads, that facilitate assembly, of Oxford Nanopore (Sović et al., 2016; Utturkar et al., 2014). Each technology requires a different genomic DNA extraction protocol.

Genomic DNA extraction for Illumina sequencing

We used liver and brain tissue samples for Illumina (short read) sequencing. The samples were taken from RNA*later*® solution using sterile forceps and weighed and rinsed in sterile PBS. They were then put in 1.5 mL Eppendorf tubes and genomic DNA extracted and purified using The Wizard® Genomic DNA Purification Kit (Promega, A1120). Chilled nuclei lysis solution (200 µL) was added to the sample tubes and the mixture homogenized using a plastic microtube pestle until tissue clumps were no longer visible. Another 400 µl of chilled nuclei lysis solution was added to the tubes and the samples were homogenized again by trituration with a 23 G needle fitted to a 1 ml syringe. The homogenates were then incubated at 65 °C for 30 min. RNase from the kit was added to the homogenate (3 µL per tube), and the mixture incubated at 37°C for 15–30 min before cooling to room temperature. Next, protein precipitation solution (200 µL) was added, and the solution vortexed and chilled on ice for 5 min. The mixture was then centrifuged at 13,000–16,000 × g for 4 min. The supernatant was transferred to a fresh tube containing 600 µL isopropanol, gently mixed by inversion, and centrifuged again at 13,000–16,000 × g for 1 min. The supernatant was removed, 70% ethanol (600µl) was subsequently added, and the tube was centrifuged again at 13,000–16,000 × g for 1 min. The ethanol was removed, and the pellet was air-dried for 15 min. Finally, the DNA was rehydrated in 100 µL of DNA rehydration solution for 1 h at 65 °C or overnight at 4 °C.

Genomic DNA extraction for Oxford Nanopore sequencing

We used brain, pancreas and heart tissue samples for Oxford Nanopore (long read) sequencing. The samples were taken from RNA*later*® solution with sterile forceps, and weighed, then rinsed with sterile PBS. These samples were then put in 1.5 mL Eppendorf tubes and subsequently proceeded to high molecular weight (HMW) DNA extraction protocols using Monarch® HMW DNA Extraction Kit for tissue (New England Biolabs, cat. NEB T3060). Lysis master mix (300 µl, consisting of 30 parts tissue lysis buffer: 1 part proteinase K) and the sample were homogenized using a clean plastic microtube pestle until all clumps disappeared. Another 300 µl of the lysis master mix was added to the tubes containing the homogenized samples and they were mixed by trituration with wide bore 200 µL micropipette tip with a wide bore to avoid shearing. The mixture was incubated at 56 °C for 45 min with agitation. Then, 10 µL of RNase A from the Monarch® kit was added and mixed by inverting 5–10 ×. Incubation for 10 min at 56 °C with agitation. The heat block in a thermal mixer was changed to accommodate a 2 ml tube, and preheated to 56 °C. Next, 300 µL of protein separation solution was added and mixed by inversion for 1 min. Centrifugation for 10 min at 16,000 × g was conducted. Using a 1000 µL wide-bore micropipette tip (see above), the upper phase containing the DNA was transferred to a labeled Monarch® 2 mL tube.

Two DNA capture beads from the Monarch® kit were added to each sample. Then, 550 µL of isopropanol was added, and the tubes placed on a vertical rotating mixer at 10 rpm for 5 min to allow DNA to attach to the beads. Then, the liquid was removed from the tubes by pipetting, taking care to avoid dislodging any gDNA adhering to the capture beads. To optimize DNA solubility, care was taken not to let the bound DNA dry out on the beads during the following steps. gDNA wash buffer (500 µL) was added, the cap closed, and the tube inverted 2–3 × before removing the wash buffer by pipetting. This wash step was repeated, and the wash buffer was removed by pipetting. A labeled bead retainer was placed into a Monarch® Collection Tube II, and the beads were poured into the bead retainer before closing the cap. A pulse spin (≤ 1 s) in a benchtop minicentrifuge was performed to remove residual wash buffer from the beads. The bead retainer was separated from the collection tube, and the beads were poured into a new, labeled Monarch® 2 mL tube, while the used bead retainer was inserted into a labeled 1.5 mL microfuge tube. Immediately, 100 µL of elution buffer II was added to the glass beads, and the tube with the beads was incubated for 5 min at 56 °C in a thermal mixer with agitation (300 rpm). Ensuring the bead retainer was inserted into the 1.5 mL microfuge tube, the eluate and beads were poured into the bead retainer and the cap closed. Centrifugation for 30 s at 12,000 × g was conducted to separate the eluate from the beads, which were then discarded along with the retainer. The eluate was triturated 5–10 × with a wide bore pipette tip, to ensure that any visible DNA aggregates were dispersed.

Illumina sequencing

Genomic DNA extracted from the tokay gecko liver sample was used for short read library preparation and sequencing, which were done by Genomics Core Facility at EMBL. The sequencing platform used was Illumina NextSeq 400 with 40× coverage.

Oxford Nanopore sequencing

Genomic DNA extracted from the tokay gecko brain sample was used for long read library preparation and sequencing, which were done at Future Genomics Technologies, Leiden, the Netherlands. The ligation sequencing Kit 1D ligation SQK-LSK110 (Oxford Nanopore) was used for library preparation. The genome sequencing was done using FLO-PRO002 (PromethION; Oxford Nanopore Technology) combined with Guppy vs5.0.17 (Oxford Nanopore Technology) at high accuracy (minimum Q-score = 9). The genome was sequenced to give 25× coverage.

Genome assembly and annotation

Genome assembly

De novo contig assembly was performed by Future Genomics Technologies using Flye 2.9-b1768 and were then polished based on Illumina reads using Pilon v. 1.23. The

polished contigs were then scaffolded using RagTag v. 2.1.0 (Alonge et al., 2022) with the genome of Townsend's least gecko *Sphaerodactylus towsendi* (GCF_021028975.2) as a reference.

Genome annotations

Genome annotation was carried out within the computing resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. The genome assembly was soft-masked using RepeatMasker v. 4.1.4. (http://www.repeatmasker.org) before going through the genome annotation pipeline. Genome structural annotation (gene prediction) was done using the BRAKER2 pipeline (Bruna et al., 2021) with gene prediction tools GeneMark-ES/ET/EP/ETP (Brůna et al., 2020) and AUGUSTUS (Stanke et al., 2008), and with RNA-seq data as extrinsic evidence. The structural annotation result from the BRAKER2 pipeline was then functionally annotated using the 'funannotate annotate' command of the Funannotate pipeline (https://funannotate.readthedocs.io/en/latest/install.html). This pipeline utilizes InterProScan5 (Jones et al., 2014), Eggnog-Mapper (Cantalapiedra et al., 2021), SignalP 5.0 (Almagro Armenteros et al., 2019), TMHMM 2.0 (Krogh et al., 2001) to generate gene names and product descriptions from protein-coding models. The genome annotation was then evaluated using BUSCO analysis (Manni et al., 2021).

Ensembl also performed genome annotation of our tokay gecko genome using their in-house genome annotation pipeline Genebuild (https://www.ensembl.org/info/genome/genebuild/index.html).

Duplicate genes classification and Comparative genomics

The protein sequences in the tokay gecko genome predicted by Ensembl genebuild were aligned to eachother using BLAST+ v. 2.14.1 (https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html). The alignment results and the gff (general feature format) file containing gene position in the genome were then further analyzed using the 'Duplicate_gene_classifier' function of MCScanX (Wang et al., 2012) to identify gene duplications.

We then performed all-against-all alignment of coding sequences (CDS) of three gecko genomes: our tokay gecko (GCA_029375565.1), Townsend's least Gecko *Sphaerodactylus townsendi* (Genbank accession number GCA_021028975.2) and the leopard gecko *Eublepharis macularius* (GCA_028583425.1), using BLAST+ v. 2.14.1 (https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html). The alignment results were then concatenated into one file. The gff files were also concatenated. These files were then further analyzed using MCScanX (Wang et al., 2012) to detect gene synteny and collinearity. The results were then explored and visualize using SynVisio (Bandi et al., 2022)

Data Availability

The tokay gecko genome assembly was submitted to GenBank (https://www.ncbi.nlm.nih.gov/genbank/; accession number GCA_029375565.1). All raw sequencing data related to the assembly have been deposited in the NCBI database SRA (sequencing reads archive) under BioProject accession number PRJNA894965. The genome tokay gecko annotation done by Ensembl can be found in the Ensembl rapid-release page (https://rapid.ensembl.org/Gekko_gecko_GCA_029375565.1/Info/Index.)

Results

The Genome assembly of the tokay gecko shows high contiguity and completeness

By implementing a hybrid genome assembly method, using Illumina and Nanopore reads, we produced a complete, high quality, scaffold-level genome assembly of the tokay gecko. The assembly has 4,208 scaffolds with scaffold N50 and L50 of 190.8 MB and 6 respectively (Table 6-1). These scaffolds were constructed from 5,480 contigs with contig N50 and L50 of 10.8 Mbp and 69, respectively (Table 6-1). Furthermore, the genome assembly of the tokay gecko reached 94.8% of complete BUSCOs, of which 92.7% were complete and single-copy, and 2.1% were complete and duplicated (Table 6-1). The complete statistics of the tokay gecko genome assembly can be found in Table S 6-1.

We also examined the interspersed repeats and low complexity DNA sequences contained in the tokay gecko genome assembly (Table 6-2). The total genome consisted of 6.38% interspersed repeats, 0.38% small RNA, 1.18% simple repeats, and 0.25% low complexity repeats (Table 6-2). The total interspersed repeats were largely composed by retroelements (6.34% of the total assembly), especially L2/CR1/Rex (5.14%) from LINEs (Table 6-2). The complete repeat statistics of the tokay gecko genome assembly, including the number and the length occupied, can be found in Table S 6-2.

Table 6-1. **Summary of assembly statistics of the tokay gecko genome.**

Assembly Level	Scaffold
Number of scaffolds	4,208
Total scaffold length	2,741,107,181
Average scaffold length	651,403.80
Scaffold N50	190,862,615
Scaffold L50	6
Largest scaffold	341,137,916
Smallest scaffold	503
Number of contigs	5,480
Total contig length	2,740,979,981
Average contig length	500,178.83

Contig N50	10,858,724
Contig L50	69
Largest contig	78,618,297
Smallest contig	503
Number of gaps in scaffolds	1,272
Total gap length in scaffolds	127,200
Average gap length in scaffolds	100
GC content %	45.37
Complete BUSCOs	7,088 (94.8%)
Complete and single-copy BUSCOs	6,932 (92.7%)
Complete and duplicated BUSCOs	156 (2.1%)
Fragmented BUSCOs	78 (1.0%)
Missing BUSCOs	314 (4.2%)
Total BUSCO groups searched	7480

Table 6-2. **Repeat statistics of the tokay gecko genome**.

Total interspersed repeats (6.38%)	Retroelements	6.34%	SINEs:	0.98%		
			LINEs	5.14%	L2/CR1/Rex	5.14%
			LTR elements	0.22%	Gypsy/DIRS1	0.21%
					Retroviral	0.01%
	DNA transposons	0.04%	hobo-Activator	0.01%		
			Tourist/Harbinger	0.01%		
	Unclassified:	0.01%				
Small RNA	0.38%					
Simple repeats	1.18%					
Low complexity	0.25%					

Notes: Percentages are relative to the total genome sequence. We only used RepeatMasker combined with Dfam-Dfam_3.0 and RepBase-20181026 Databases for repeat identification. *De novo* transposable element (TE) family identification and modeling will be done in the future using Repeat Modeler; we therefore expect the repeat statistics to increase.

The genome of the tokay gecko is well annotated

We used a combination of the BRAKER2 and Funannotate pipelines for genome annotation because BRAKER2 pipeline alone does not support functional annotation, while Funannotate showed a low accuracy in predicting gene structure in gigabyte-sized genome. By combining these two pipelines, we identified 27,350 genes, 172,817 distinct exons and 30,229 predicted transcripts (Table 6-3). Our genome annotation reached 83.9% completed BUSCOs consisting of 77% Complete and single-copy BUSCOs and 11% complete and duplicated BUSCOs. We also asked Ensembl to run our assembly through their in-house genome annotation pipeline, genebuild. This annotation pipeline predicted 22,009 genes, 232,387 distinct exons and 35,820 transcripts. The Ensembl

annotation reached 95.8% completed BUSCOs of which 93.8% were complete and single-copy, and 2.0% were complete and duplicated.

Table 6-3. **Annotation statistics of the tokay gecko genome**.

Criterion	BRAKER2 + Funannotate	Ensembl genebuild
Max number of transcripts per gene	4	10
Mean exon size	226.1	278.5
Mean gene locus size (first to last exon)	31,439.6	53,792.3
Mean number of distinct exons per gene	6.3	10.6
Mean number of transcripts per gene	1.1	1.6
Mean transcript size (UTR, CDS)	1,385	2,667.1
Number of distinct exons	172,817	232,387
Number of genes	27,350	22,009
Number of genes with alternative transcript variants	15,718 (57.5%)	8,011 (36.4%)
Number of multi-exon genes	15,718 (57.5%)	21,171 (96.2%)
Number of predicted transcripts	30,229	35,820
Number of single-exon genes	11,632 (42.5%)	838 (3.8%)
Complete BUSCOs	6,283 (83.9%)	7,161 (95.8%)
Complete and single-copy BUSCOs	5,759 (77%)	7,014 (93.8%)
Complete and duplicated BUSCOs	824 (11%)	147 (2.0%)
Fragmented BUSCOs	89 (1.2%)	106 (1.4%)
Missing BUSCOs	808 (10.8%)	213 (2.8%)
Total BUSCO groups searched	7480	7480

Note: The lineage dataset for BUSCO analysis is sauropsida_odb10 (Creation date: 2024-01-08, number of genomes: 76, number of BUSCOs: 7480). BUSCO was run in euk_genome_met mode utilizing metaeuk gene predictor.

The Ensembl genebuild gave more complete annotations than our combined BRAKER2 and Funannotate analysis. Therefore, we used the Ensembl annotations to perform gene duplication classification in the tokay gecko genome. We found that most of the gene duplications (22,977 genes) are tandem duplications, followed in number by dispersed duplications (8,863 genes), proximal duplications (1,155 genes) and WGD (whole genome duplication) or segmental duplications (1,023 genes). We identified 1,802 singleton genes.

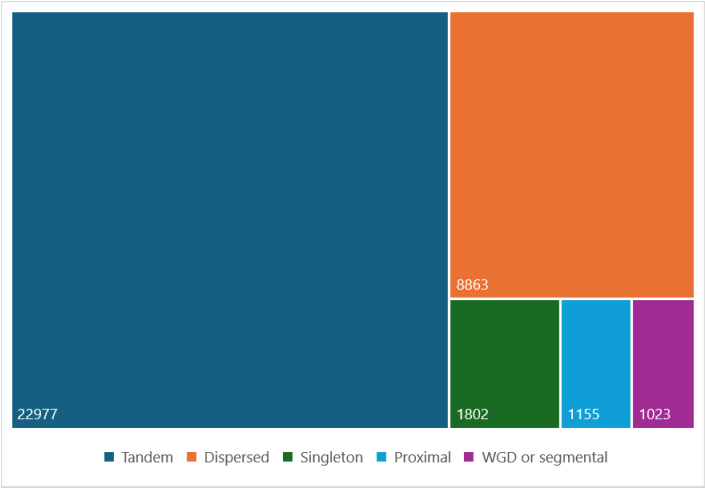

Figure 6-1. **Gene duplication classification in the tokay gecko genome.** Note that a gene can be classified into more than one gene duplication type.

Comparison of the tokay gecko with two other gecko genomes gives an estimate of their phylogenetic distances

We performed synteny analysis by comparing our scaffold-level tokay gecko genome assembly with the chromosome-level assemblies of Townsend's least gecko (Pinto et al., 2022) and the leopard gecko (Pinto et al., 2023). These three geckos belong to the same infraorder (Gekkota; see (Pyron et al., 2013)) but different families (Gekkonidae, Sphaerodactylidae, and Eublepharidae, respectively). We found that the tokay gecko genome assembly is comparable in completeness and contiguity to the those of the Townsend's least gecko and the leopard gecko (Figure 6-2). This is indicated by the extensive syntenic blocks shared between their genomes (Figure 6-2A, B, C). In a comparison of the genomes of the tokay gecko *vs*. the Townsend's least gecko (Figure 6-2B, C), we found that only a few macrosyntenic re-arrangements. By contrast, there were many macrosynteny (chromosomes) and microsynteny (gene clusters) inversions in the comparison of the leopard gecko *vs.* the other two (Figure 6-2A, D). We also found translocations in the comparison of the tokay gecko genome *vs*. the Townsend's least gecko genome (Figure 6-2B, C, D). However, the comparison of the tokay gecko *vs*. the leopard gecko genomes showed more translocations of synteny blocks than in the Townsend's least gecko *vs*. the leopard gecko comparison (Figure 6-2A).

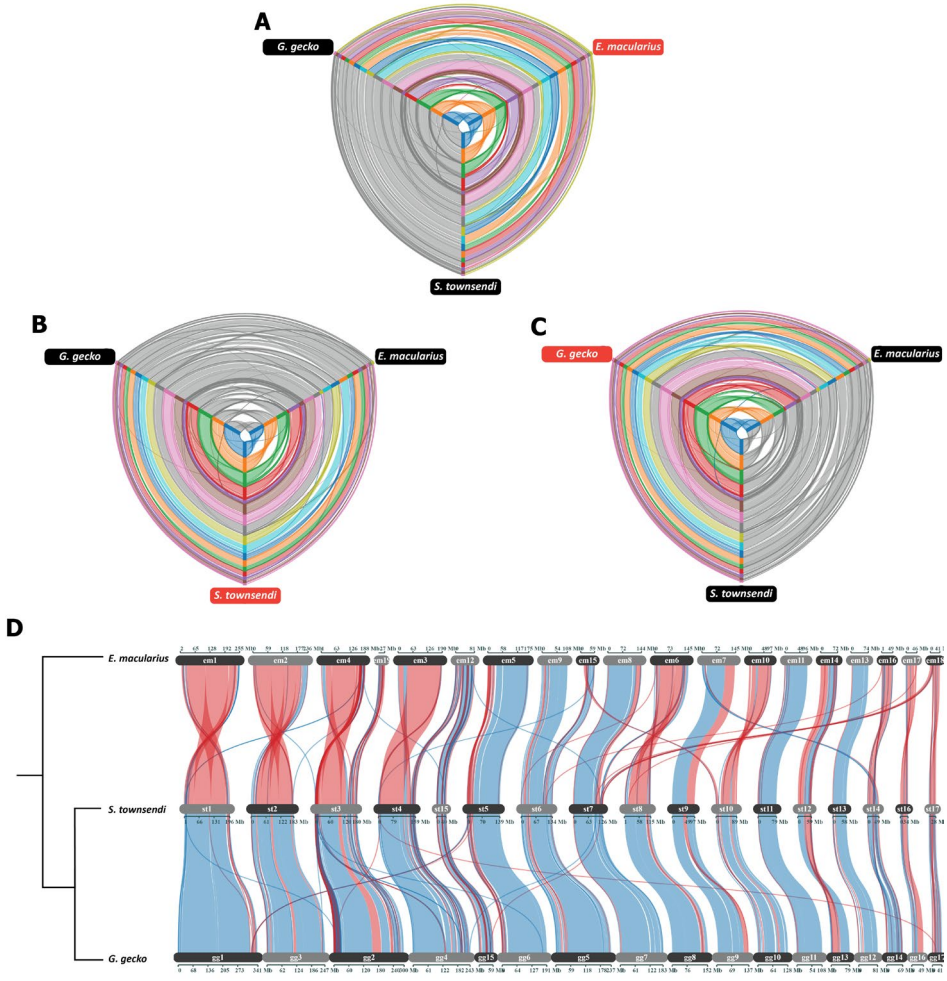

Figure 6-2. **Synteny analysis shows that the genome of the tokay gecko is phylogenetically closer to the Townsend's Least Gecko than the leopard gecko**. (A–C) Overview of synteny comparisons between the tokay gecko (*G. gecko*), the Townsend's least gecko (*S. townsendi*), and the leopard gecko (*E. macularius*). Each color represents one chromosome. (D) Alignment orientation of synteny block between three genomes. Blue color represents regular syntenic block links and red color represents reverse syntenic block links. The chromosomes were reordered for better visualization. The chromosome order in A–C and D is the same. The phylogenetic tree was adapted from (Pyron et al., 2013). *Key* to annotations in D*:* em1–em19, chromosomes of the leopard gecko (chromosome accession NC_072790.1–NC_072808.1); st1–st17, chromosomes of the Townsend's least gecko (chromosome accession NC_059425.1–NC_059441.1); gg1–gg17, chromosomes of the tokay gecko (chromosome accession JAPDNY010000001.1– JAPDNY010000017.1).

Discussion

In this chapter, we sequenced, assembled, and annotated the genome of the tokay gecko. Through the implementation of a hybrid genome assembly approach, that utilized both Illumina and Nanopore sequencing technologies, we were able to produce a high-quality, scaffold-level genome assembly. Our genome annotation was of relatively high-quality, despite specific areas (duplications, fragmentation, missing genes) that we plan to optimize in the future.

Our genome assembly has high contiguity and completeness and has a coverage of 40× (Illumina sequencing) and 24× coverage (Nanopore sequencing). Our assembly result is better than other lizard genomes that were sequenced by Illumina sequencing only but have higher coverage. For instance, the genome assembly of the Schlegel's Japanese gecko (GCA_001447785.1) that used Illumina sequencing and has a coverage of 95× (Liu et al., 2015) has a scaffold N50 of 707.7 kb and scaffold L50 of 963. The other example is the genome of the Madagascar ground gecko (*Paroedura picta*; GCA_003118565.2) which has 77.7× genome coverage, a scaffold N50 of 109 Mb and a scaffold L50 of 6 (Hara et al., 2018; Yamaguchi et al., 2021). These comparisons confirm that combining Illumina and Nanopore sequencing is a good option for providing a more contiguous and complete genome assembly than using Illumina sequencing alone, even if the latter produces a high genome coverage (as has also been argued for example by (Utturkar et al., 2014)).

The tokay gecko genome annotations produced using the BRAKER2 and Funannotate pipelines predicts more genes than the annotations from genebuild. However, the genebuild annotation result is more complete than the combined BRAKER2 + Funannotate annotations. This might be related to the genebuild pipeline utilizing more extensive and comprehensive data integration, including transcriptomic and proteomic data, according to their web notes (https://www.ensembl.org/info/genome/genebuild/index.html). This data integration produces better gene models for genome annotation in general (Prasad et al., 2017; Venturini et al., 2018). Moreover, genebuild has an advanced prediction algorithm, according to their website. Meanwhile, our BRAKER2 + Funannotate genome annotation used only our tokay gecko RNA-seq data (in this thesis) and UniProt (https://www.uniprot.org/) for the protein references. Genebuild, however, uses a larger number of reference databases (https://www.ensembl.org/info/genome/genebuild/annotation_sources.html). Finally, the BRAKER3 pipeline, an update of BRAKER2 pipeline that integrates both RNA-seq data and protein evidence, was not yet stable when we annotated the tokay gecko genome (Bruna et al., 2021; Gabriel et al., 2024). We should consider using BRAKER3 with more

optimal parameters, and re-training gene models, to achieve higher genome annotation completeness in the future.

The tokay gecko genome has a relatively high-quality annotation, especially the annotation done by Ensembl using genebuild. That Ensembl annotation is publicly available (see data availability in Materials and Methods, this chapter). Our tokay gecko genome annotations are even higher in term of the number of genes and BUSCO completeness, compared to the Schlegel's Japanese gecko (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001447785.1) and the Townsend's least gecko genome annotations (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_021028975.2). In general, well-annotated genomes are the most useful for researchers studying molecular mechanisms in model species (Richards, 2015; Worley et al., 2017).

I have shown in this chapter that the tokay gecko genome can be used for chromosomal genome comparison with other genomes having a chromosome-level assembly, namely Townsend's least gecko (Pinto et al., 2022) and the leopard gecko (Pinto et al., 2023). This comparison showed that there are few chromosomal rearrangements, especially translocations, in the three genomes. The fact that I only identified a few chromosomal rearrangements indicates that the three species compared are separated by a relatively small phylogenetic distance (Bhutkar et al., 2008; Sankoff, 2003). Moreover, this comparison showed that the tokay gecko is closer, phylogenetically, to the Towsend least gecko than it is to the leopard gecko. This result is consistent with the squamate phylogeny of (Pyron et al., 2013) showing that the family Gekkonidae, which includes the tokay gecko, is closer to the Sphaerodactylidae, that includes Townsend's least gecko, than to the family Eublepharidae, that includes the leopard gecko.

In summary, this chapter covers the genome assembly and annotation of the tokay gecko. Through the hybrid assembly approach using Illumina and Nanopore sequencing, we produced a high-quality scaffold-level genome assembly for the tokay gecko. Moreover, the genome annotation showed relatively high completeness. Our tokay gecko genome will provide a valuable resource for biological research using the tokay gecko and other lizards as model species, especially in the field regenerative biology. It might also be interesting to do comparative genomic studies of the tokay gecko genome and the genomes of other squamates, since the preliminary genome comparison reported here showed promising results.

References

Alibardi, L. (2019). Review: The Regenerating Tail Blastema of Lizards as a Model to Study Organ Regeneration and Tumor Growth Regulation in Amniotes. *Anat Rec (Hoboken)* **302**, 1469-1490.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**, 420-423.

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C. and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology* **23**, 258.

Bandi, V., Gutwin, C., Siri, J. N., Neufeld, E., Sharpe, A. and Parkin, I. (2022). Visualization Tools for Genomic Conservation. *Methods Mol Biol* **2443**, 285-308.

Bhutkar, A., Schaeffer, S. W., Russo, S. M., Xu, M., Smith, T. F. and Gelbart, W. M. (2008). Chromosomal Rearrangement Inferred From Comparisons of 12 Drosophila Genomes. *Genetics* **179**, 1657-1680.

Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108.

Brůna, T., Lomsadze, A. and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* **2**, lqaa026.

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825-5829.

Cobos, A. J. and Higham, T. E. (2022). Growing up in a rough world: scaling of frictional adhesion and morphology of the Tokay gecko (Gekko gecko). *Beilstein J Nanotechnol* **13**, 1292-1302.

Das, I. (2015). A field guide to the reptiles of South-East Asia. pp. 375-375.

Davalos-Dehullu, E., Baty, S. M., Fisher, R. N., Scott, P. A., Dolby, G. A., Munguia-Vega, A. and Cortez, D. (2023). Chromosome-Level Genome Assembly of the Blacktail Brush Lizard, Urosaurus nigricaudus, Reveals Dosage Compensation in an Endemic Lizard. *Genome Biology and Evolution* **15**, evad210.

Franklin, B. M., Voss, S. R. and Osborn, J. L. (2017). Ion channel signaling influences cellular proliferation and phagocyte activity during axolotl tail regeneration. *Mechanisms of Development* **146**, 42-54.

Gabriel, L., Bruna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M. and Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res* **34**, 769-777.

Hara, Y., Takeuchi, M., Kageyama, Y., Tatsumi, K., Hibi, M., Kiyonari, H. and Kuraku, S. (2018). Madagascar ground gecko genome analysis characterizes asymmetric fates of duplicated genes. *BMC Biology* **16**, 1-19.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240.

Kadafi, A. M., Fauzi, M. A., Ardiantoro, A., Zakky, Q., Nugraha, F. A. D., Priambodo, B., Fahmi, M., Munir, M., , Riyanto, A. and Hamidy, A. (2024). Low genetic diversity of the Indonesian Tokay Gecko (Gekko gecko) (Reptilia: Gekkonidae): Introduce or Native? *Jurnal Biologi Indonesia* **20**, 63-75.

Kongbuntad, W., Tantrawatpan, C., Pilap, W., Jongsomchai, K., Chanaboon, T., Laotongsan, P., Petney, T. N. and Saijuntha, W. (2016). Genetic diversity of the red-spotted tokay gecko (Gekko gecko Linnaeus, 1758) (Squamata: Gekkonidae) in Southeast Asia determined with multilocus enzyme electrophoresis. *Journal of Asia-Pacific Biodiversity* **9**, 63-68.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology* **305**, 567-580.

Li, W., Du, J., Yang, L., Liang, Q., Yang, M., Zhou, X. and Du, W. (2023). Chromosome-level genome assembly and population genomics of Mongolian racerunner (Eremias argus) provide insights into high-altitude adaptation in lizards. *BMC Biol* **21**, 40.

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. *Nature Communications* **6**, 1-11.

Lozito, T. P. and Tuan, R. S. (2017). Lizard tail regeneration as an instructive model of enhanced healing capabilities in an adult amniote. *Connective Tissue Research* **58**, 145-154.

Manni, M., Berkeley, M. R., Seppey, M. and Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols* **1**, e323.

McCusker, C., Bryant, S. V. and Gardiner, D. M. (2015). The axolotl limb blastema: cellular and molecular mechanisms driving blastema formation and limb regeneration in tetrapods. *Regeneration* **2**, 54-71.

McHedlishvili, L., Mazurov, V., Grassme, K. S., Goehler, K., Robl, B., Tazaki, A., Roensch, K., Duemmler, A. and Tanaka, E. M. (2012). Reconstitution of the central and peripheral nervous system during salamander tail regeneration. *Proceedings of the National Academy of Sciences of the United States of America* **109**.

Pinto, B. J., Gamble, T., Smith, C. H., Keating, S. E., Havird, J. C. and Chiari, Y. (2023). The revised reference genome of the leopard gecko (Eublepharis macularius) provides insight into the considerations of genome phasing and assembly. *bioRxiv*.

Pinto, B. J., Keating, S. E., Nielsen, S. V., Scantlebury, D. P., Daza, J. D. and Gamble, T. (2022). Chromosome-Level Genome Assembly Reveals Dynamic Sex Chromosomes in Neotropical Leaf-Litter Geckos (Sphaerodactylidae: Sphaerodactylus). *J Hered* **113**, 272-287.

Polikarpova, A., Ellinghaus, A., Schmidt-Bleek, O., Grosser, L., Bucher, C. H., Duda, G. N., Tanaka, E. M. and Schmidt-Bleek, K. (2022). The specialist in regeneration-the Axolotl-a suitable model to study bone healing? *NPJ Regen Med* **7**, 35.

Prasad, T. S., Mohanty, A. K., Kumar, M., Sreenivasamurthy, S. K., Dey, G., Nirujogi, R. S., Pinto, S. M., Madugundu, A. K., Patil, A. H., Advani, J., et al. (2017). Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Res* **27**, 133-144.

Pyron, R. A., Burbrink, F. T. and Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology* **13**, 93.

Richards, S. (2015). It's more than stamp collecting: how genome sequencing can unify biological research. *Trends Genet* **31**, 411-421.

Rumping, J. M. and Jayne, B. C. (1996). Muscle activity in autotomized tails of a lizard (Gekko gecko): a naturally occurring spinal preparation. *Journal of comparative physiology. A, Sensory, neural, and behavioral physiology* **179**, 525-538.

Sankoff, D. (2003). Rearrangements and chromosomal evolution. *Current Opinion in Genetics & Development* **13**, 583-587.

Sović, I., Križanović, K., Skala, K. and Šikić, M. (2016). Evaluation of hybrid and non-hybrid methods for de novo assembly of nanopore reads. *Bioinformatics* **32**, 2582-2589.

Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637-644.

Utturkar, S. M., Klingeman, D. M., Land, M. L., Schadt, C. W., Doktycz, M. J., Pelletier, D. A. and Brown, S. D. (2014). Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics* **30**, 2709-2716.

Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. and Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience* **7**, giy093.

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49.

Worley, K. C., Richards, S. and Rogers, J. (2017). The value of new genome references. *Experimental Cell Research* **358**, 433-438.

Yamaguchi, K., Kadota, M., Nishimura, O., Ohishi, Y., Naito, Y. and Kuraku, S. (2021). Technical considerations in Hi-C scaffolding and evaluation of chromosome-scale genome assemblies. *Mol Ecol* **30**, 5923-5934.

Supplementary Information

Table S 6-1. **Complete genome assembly statistics of the tokay gecko (*Gekko gecko*)**

# scaffolds	**4,208**
Total scaffold length	2,741,107,181
Average scaffold length	651,403.80
Scaffold N50	190,862,615
Scaffold auN	200,046,474.16
Scaffold L50	6
Scaffold NG50	341,137,916
Scaffold auNG	203,092,158,088.95
Scaffold LG50	1
Largest scaffold	341,137,916
Smallest scaffold	503
# contigs	5,480
Total contig length	2,740,979,981
Average contig length	500,178.83
Contig N50	10,858,724
Contig auN	15,243,932.67
Contig L50	69
Contig NG50	78,618,297
Contig auNG	15,475,301,583.67
Contig LG50	1
Largest contig	78,618,297
Smallest contig	503
# gaps in scaffolds	1,272
Total gap length in scaffolds	127,200
Average gap length in scaffolds	100
Gap N50 in scaffolds	100
Gap auN in scaffolds	100
Gap L50 in scaffolds	636
Largest gap in scaffolds	100
Smallest gap in scaffolds	100
Base composition (A:C:G:T)	A: 748,908,636 C: 621,741,100 G: 621,751,449 T: 748,578,796
GC content %	45.37
# soft-masked bases	1,164,152,571
# segments	5,480
Total segment length	2,740,979,981
Average segment length	500,178.83
# gaps	1,272
# paths	4,208
Scaffold N10	341,137,916
Scaffold N20	299,885,588
Scaffold N30	247,548,350
Scaffold N40	243,323,372
Scaffold N50	190,862,615
Scaffold N60	182,932,204
Scaffold N70	138,791,800
Scaffold N80	107,697,327
Scaffold N90	68,720,646
Scaffold N100	503
Scaffold L10	1
Scaffold L20	2
Scaffold L30	3
Scaffold L40	4
Scaffold L50	6
Scaffold L60	7
Scaffold L70	9
Scaffold L80	11
Scaffold L90	14
Scaffold L100	4,208
Scaffold NG10	341,137,916
Scaffold NG20	341,137,916
Scaffold NG30	341,137,916
Scaffold NG40	341,137,916
Scaffold NG50	341,137,916
Scaffold NG60	341,137,916
Scaffold NG70	341,137,916
Scaffold NG80	341,137,916
Scaffold NG90	341,137,916
Scaffold NG100	341,137,916
Scaffold LG10	1
Scaffold LG20	1
Scaffold LG30	1
Scaffold LG40	1
Scaffold LG50	1
Scaffold LG60	1
Scaffold LG70	1
Scaffold LG80	1
Scaffold LG90	1

Scaffold LG100	1
Contig N10	31,810,334
Contig N20	23,242,895
Contig N30	18,319,056
Contig N40	14,176,216
Contig N50	10,858,724
Contig N60	8,172,830
Contig N70	5,514,941
Contig N80	3,153,579
Contig N90	1,316,638
Contig N100	503
Contig L10	6
Contig L20	17
Contig L30	30
Contig L40	47
Contig L50	69
Contig L60	98
Contig L70	140
Contig L80	206
Contig L90	333
Contig L100	5,480
Contig NG10	78,618,297
Contig NG20	78,618,297
Contig NG30	78,618,297
Contig NG40	78,618,297
Contig NG50	78,618,297
Contig NG60	78,618,297
Contig NG70	78,618,297
Contig NG80	78,618,297
Contig NG90	78,618,297
Contig NG100	78,618,297

Contig LG10	1
Contig LG20	1
Contig LG30	1
Contig LG40	1
Contig LG50	1
Contig LG60	1
Contig LG70	1
Contig LG80	1
Contig LG90	1
Contig LG100	1
Gap N10	100
Gap N20	100
Gap N30	100
Gap N40	100
Gap N50	100
Gap N60	100
Gap N70	100
Gap N80	100
Gap N90	100
Gap N100	100
Gap L10	128
Gap L20	255
Gap L30	382
Gap L40	509
Gap L50	636
Gap L60	764
Gap L70	891
Gap L80	1,018
Gap L90	1,145
Gap L100	1,272

Table S 6-2. **Complete genome repeat statistics of the tokay gecko**.

		number of elements*	length occupied	percentage of sequence
Retroelements		**738,120**	**173,753,694 bp**	**6.34%**
	SINEs:	272,912	26,735,118 bp	0.98%
	Penelope	75	14,042 bp	0.00%
	LINEs:	453,890	141,006,990 bp	5.14%
	CRE/SLACS	0	0 bp	0.00%
	L2/CR1/Rex	453,784	140,986,049 bp	5.14%
	R1/LOA/Jockey	0	0 bp	0.00%
	R2/R4/NeSL	0	0 bp	0.00%
	RTE/Bov-B	0	0 bp	0.00%
	L1/CIN4	31	6,899 bp	0.00%
	LTR elements:	11,318	6,011,586 bp	0.22%
	BEL/Pao	0	0 bp	0.00%
	Ty1/Copia	0	0 bp	0.00%
	Gypsy/DIRS1	9,536	5,663,273 bp	0.21%
	Retroviral	1,720	338,909 bp	0.01%
DNA transposons		**10,622**	**970,491 bp**	**0.04%**
	hobo-Activator	2,535	250,048 bp	0.01%
	Tc1-IS630-Pogo	288	49,187 bp	0.00%
	En-Spm	0	0 bp	0.00%
	MuDR-IS905	0	0 bp	0.00%
	PiggyBac	0	0 bp	0.00%
	Tourist/Harbinger	2,796	190,934 bp	0.01%
	Other (Mirage, P-element, Transib)	0	0 bp	0.00%
Rolling-circles		**15**	**2,208 bp**	**0.00%**
Unclassified:		**1,327**	**205,026 bp**	**0.01%**
Total interspersed repeats:			**174,929,211 bp**	**6.38%**
Small RNA:		**152,557**	**10,439,824 bp**	**0.38%**
Satellites:		**521**	**26,033 bp**	**0.00%**
Simple repeats:		**701,504**	**32,228,786 bp**	**1.18%**
Low complexity:		**94,953**	**6,768,206 bp**	**0.25%**

* as stated in the software report, most repeats fragmented by insertions or deletions have been counted as one element. Runs of >=20 X/Ns in query were excluded in % calcs. The query species was assumed to be *Gallus gallus*. RepeatMasker Combined Database: Dfam-Dfam_3.0, RepBase-20181026. run with rmblastn version 2.9.0+