

Additional file 12

Survey of anuran genome assemblies in NCBI Genome

Methods

All publicly available Anura genome assemblies were retrieved from NCBI Genome [1] on 18 April 2026 using the NCBI Datasets command-line interface (CLI, v16) [2]. For each assembly, accession, organism name, sequencing technology, assembly level, total length, scaffold N50, contig N50, BioSample accession, and assembly name were recorded. Where both a GenBank (GCA) and a RefSeq (GCF) accession existed for the same assembly (identified by shared assembly name), the RefSeq accession was retained and the GenBank duplicate removed, reducing the initial 242 entries to 214 unique assemblies.

Sequencing technology was classified for each assembly using a whitelist approach. All Oxford Nanopore-related terms (Oxford, Nanopore, Technologies, MinION, GridION, PromethION, Flongle) were stripped from the sequencing technology field. Assemblies for which no residual platform terms remained were classified as ONT-only. Assemblies containing other technologies were classified as PacBio + Hi-C, PacBio + Illumina, PacBio only, Illumina only, ONT + other, or Other, based on the platforms present.

Taxonomic classification to family and superfamily level was performed using taxonkit v0.19.0 [3] with the NCBI Taxonomy database (taxdump, April 2026). Terrarana membership was defined as assemblies belonging to Brachycephalidae, Craugastoridae, Strabomantidae, Eleutherodactylidae, or Ceuthomantidae, following Hedges et al. [4] and Heinicke et al. [5].

BioSample tissue metadata were retrieved for each of the 167 unique BioSample accessions via efetch (NCBI Entrez Direct) [6] in batches of 50 and parsed from the XML document summaries. Three fields were extracted: tissue, isolation_source, and sample_type. Each record was classified following the

framework of Taberlet et al. [7], which distinguishes destructive sampling (animal killed: internal organs, whole organism, larvae), nondestructive sampling (animal captured and released: swabs, toe clip), and noninvasive sampling (no capture: shed skin, faeces). Two additional categories were applied: ambiguous (blood, erythrocytes, muscle, tongue, bare skin - tissue types compatible with both live sampling and post-mortem collection, where the animal's fate cannot be determined from metadata alone) and unclassified (records listing only "tissue" or "DNA" without further specification). Records where tissue, isolation_source, and sample_type were all empty were classified as not reported. For BioSamples listing multiple tissue types, the most harmful category present determined the classification. Classification was performed across all three extracted fields (tissue, isolation_source, sample_type) to account for occasional misplacement of tissue information into non-tissue metadata fields.

Results

After deduplication, 214 unique anuran genome assemblies representing 148 species remained. Of these, 82 were at chromosome level, 123 at scaffold level, and nine at contig level. The majority of assemblies (108) were derived from tissues classified as destructive, 33 were classified as ambiguous, and none were classified as nondestructive based on BioSample metadata alone (Table 1). Illumina-only and PacBio + Hi-C were the most common sequencing strategies, together accounting for 153 of 214 assemblies. Only two assemblies were classified as ONT-only (Table 2).

Table 1. Tissue source classification of 214 anuran genome assemblies in NCBI Genome. Accessed 18 April 2026). Tissue metadata were extracted from BioSample records and classified following the framework of Taberlet et al. [7]. No assemblies were classified as nondestructive based on BioSample metadata alone.

Tissue class	Assemblies
Destructive	108
Unclassified	65

Ambiguous	33
Not reported	8
Nondestructive	0

Table 2. Sequencing technology classification of 214 anuran genome assemblies in NCBI Genome.

Technology class	All assemblies	Scaffold or chromosome level
Illumina only	83	81
PacBio + Hi-C	70	69
PacBio + Illumina	31	31
PacBio only	22	16
ONT + other	5	5
ONT only	2	2
Other	1	1

Six Terrarana assemblies from three species were present (Table 3). Five assemblies represented two *Eleutherodactylus* species (*E. coqui*, IUCN Least Concern [8]; *E. marnockii*, IUCN Least Concern [9]), all at chromosome level. The sixth was *P. euphronides* (IUCN Critically Endangered [10]), the assembly reported in this study, and the only *Pristimantis* and the only Strabomantidae (*sensu* NCBI Taxonomy) assembly.

Table 3. All Terrarana genome assemblies in NCBI Genome classified by family. Family classification according to NCBI (April, 2026). Terrarana group defined by Hedges et al. [4] and Heinicke et al. [5]. IUCN status checked manually.

Accession	Species	Family	Tech	Level	IUCN
GCA_019857665.1	<i>Eleutherodactylus coqui</i>	Eleutherodactylidae	Illumina	Chromosome	LC
GCA_035609135.1	<i>Eleutherodactylus coqui</i>	Eleutherodactylidae	PacBio + Hi-C	Chromosome	LC
GCF_035609145.1	<i>Eleutherodactylus coqui</i>	Eleutherodactylidae	PacBio + Hi-C	Chromosome	LC
GCA_055470875.1	<i>Eleutherodactylus marnockii</i>	Eleutherodactylidae	PacBio + Hi-C	Chromosome	LC

GCA_055470885.1	<i>Eleutherodactylus marnockii</i>	Eleutherodactylidae	PacBio + Hi-C	Chromosome	LC
GCA_965278355.2	<i>Pristimantis euphronides</i>	Strabomantidae	ONT only	Scaffold	CR

Two assemblies were classified as ONT-only (Table 4). Both were from IUCN Critically Endangered species and both were at scaffold level. The *Leptodactylus fallax* assembly (GCA_947044405.1; IUCN Critically Endangered [11]) was derived from muscle tissue of a specimen held at the Parc Zoologic de Barcelona and was neither scaffolded nor annotated. The *P. euphronides* assembly (GCA_965278355.2; IUCN Critically Endangered [10]), presented in this study, was derived from blood, toe clip, and oral swab samples collected non-lethally from a wild individual at Grand Etang, Grenada, and was reference-scaffolded and annotated using the HANNO pipeline.

Table 4. ONT-only anuran genome assemblies in NCBI Genome. Tissue classification follows Taberlet et al. [7].

Accession	Species	IUCN	Scaffold N50 (Mb)	Tissue	Tissue class	Scaffolded	Annotated	Sampling location
GCA_947044405.1	<i>Leptodactylus fallax</i>	CR	4.6 (= contig N50)	MUSCLE	Ambiguous	No	No	Parc Zoologic de Barcelona (ex situ)
GCA_965278355.2	<i>Pristimantis euphronides</i>	CR	117.8	blood, toe, oral swab	Ambiguous	Yes	Yes (HANNO)	Grand Etang, Grenada (in situ)

Of the 33 ambiguous assemblies, 30 were at scaffold or chromosome level (Table 5). Among these, 18 listed muscle as the determining tissue term, nine listed blood or erythrocytes, one listed muscle and blood in combination, one listed blood with toe clip and oral swab, and one had no tissue term but contained tissue-related information in the isolation_source metadata field.

Table 5. Ambiguous assemblies at scaffold or chromosome level. Grouped by the tissue term that determined the ambiguous classification. Tissue classification follows Taberlet et al. [7]. Accessions sharing the same BioSample represent haplotype pairs or alternate assemblies of the same data.

Accession	Species	BioSample	Tissue	Level
Muscle as determining tissue term (18 assemblies)				
GCA_009667805.1	<i>Leptobrachium leishanense</i>	SAMN10412277	muscle	Chromosome
GCA_019650415.1	<i>Pipa parva</i>	SAMN12075865	Muscle	Scaffold
GCA_029493135.1	<i>Hyla sarda</i>	SAMN32629340	muscle	Scaffold
GCA_029574335.1	<i>Rana kukunoris</i>	SAMN23768474	muscle	Chromosome
GCA_038048865.1	<i>Mixophyes fleayi</i>	SAMN39657083	muscle	Chromosome
GCA_048565335.1	<i>Mantella aurantiaca</i>	SAMN46714743	muscle	Chromosome
GCA_048565345.1	<i>Mantella aurantiaca</i>	SAMN46714743	muscle	Chromosome
GCA_052815355.1	<i>Rana chensinensis</i>	SAMN49279854	muscle	Chromosome
GCA_054083975.1	<i>Fejervarya limnocharis</i>	SAMN53215033	muscle	Chromosome
GCA_054084005.1	<i>Fejervarya limnocharis</i>	SAMN53215033	muscle	Chromosome
GCA_054643415.1	<i>Pelophylax nigromaculatus</i>	SAMN51991534	muscle	Chromosome
GCA_905336975.1	<i>Bombina variegata variegata</i>	SAMEA6126447	muscle	Scaffold
GCA_933207985.1	<i>Pelobates cultripes</i>	SAMEA10021128	MUSCLE	Chromosome
GCA_947044405.1	<i>Leptodactylus fallax</i>	SAMEA9957016	MUSCLE	Scaffold
GCF_027358695.1	<i>Spea bombifrons</i>	SAMN28408467	muscle	Chromosome
GCF_029499605.1	<i>Hyla sarda</i>	SAMN32629340	muscle	Chromosome
GCF_038048845.1	<i>Mixophyes fleayi</i>	SAMN39657083	muscle	Chromosome
GCF_042186555.1	<i>Aquarana catesbeiana</i>	SAMN37430746	muscle	Chromosome
Blood or erythrocytes as determining tissue term (9 assemblies)				
GCA_009801035.1	<i>Oophaga pumilio</i>	SAMN09702071	blood	Scaffold
GCA_013368275.1	<i>Xenopus tropicalis</i>	SAMN14253885	blood	Chromosome
GCA_019447015.1	<i>Hymenochirus boettgeri</i>	SAMN13066281	Blood	Chromosome
GCA_019512145.1	<i>Engystomops pustulosus</i>	SAMN13066191	Blood	Chromosome
GCA_024363595.1	<i>Xenopus borealis</i>	SAMN27622247	blood	Chromosome
GCA_032062155.1	<i>Leptobrachium ailaonicum</i>	SAMN12163464	Blood	Chromosome
GCF_001663975.1	<i>Xenopus laevis</i>	SAMN04518361	blood	Chromosome
GCF_017654675.1	<i>Xenopus laevis</i>	SAMN18035302	Erythrocytes	Chromosome
GCF_032062135.1	<i>Pyxicephalus adspersus</i>	SAMN08767402	Blood	Chromosome
Muscle and blood in combination (1 assembly)				
GCF_014858855.1	<i>Bufo gargarizans</i>	SAMN14734018	muscle, blood	Chromosome
Blood with nondestructive tissue types (1 assembly)				
GCA_965278355.2	<i>Pristimantis euphronides</i>	SAMEA118241337	blood, toe, oral swab	Scaffold
No tissue term in tissue field; classified via isolation_source field (1 assembly)				
GCF_000935625.1	<i>Nanorana parkeri</i>	SAMN02720813	(isolation_source: muscle sample)	Scaffold

References

1. NCBI Genome. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/datasets/genome>. Accessed 2 Apr 2026.
2. NCBI Datasets CLI. Version 16. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/datasets/docs/v2/download-and-install/>. Accessed 1 Jan 2026.
3. J. W. Ren H. TaxonKit: A practical and efficient NCBI taxonomy toolkit. *J Genet Genomics*. 2021;48(9):844-850.
4. Hedges SB, Duellman WE, Heinicke MP. New World direct-developing frogs (Anura: Terrarana): Molecular phylogeny, classification, biogeography, and conservation. *Zootaxa*. 2008;1737:1-182.
5. Heinicke MP, Duellman WE, Trueb L, Means DB, MacCulloch RD, Hedges SB. A new frog family (Anura: Terrarana) from South America and an expanded direct-developing clade revealed by molecular phylogeny. *Zootaxa*. 2009;2211(1):1-35.
6. Sayers EW, Bolton EE, Brister JR, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2022;50(D1):D13-D25.
7. Taberlet P, Waits LP, Luikart G. Noninvasive genetic sampling: look before you leap. *Trends Ecol Evol*. 1999;14(8):323-327.
8. IUCN SSC Amphibian Specialist Group. 2021. *Eleutherodactylus coqui*. The IUCN Red List of Threatened Species 2021: e.T56522A3041672. <https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T56522A3041672.en>. Accessed on 18 Apr 2026.
9. IUCN SSC Amphibian Specialist Group. 2022. *Eleutherodactylus marnockii*. The IUCN Red List of Threatened Species 2022: e.T56744A53965348. <https://dx.doi.org/10.2305/IUCN.UK.2022-1.RLTS.T56744A53965348.en>. Accessed on 18 Apr 2026.
10. IUCN SSC Amphibian Specialist Group. 2021. *Pristimantis euphronides*. The IUCN Red List of Threatened Species 2021: e.T56593A3043096. <https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T56593A3043096.en>. Accessed on 18 Apr 2026.
11. IUCN SSC Amphibian Specialist Group. 2017. *Leptodactylus fallax*. The IUCN Red List of Threatened Species 2017: e.T57125A3055585. <https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T57125A3055585.en>. Accessed on 18 Apr 2026.