SUPPLEMENTARY DATA

Multi-tissue transcriptomes of caecilian amphibians highlight incomplete knowledge of vertebrate gene families

María Torres-Sánchez, Christopher J. Creevey, Etienne Kornobis, David J. Gower, Mark Wilkinson, Diego San Mauro*

* Author for correspondence (dsanmaur@ucm.es)

This PDF file includes:

Supplementary Tables S1, S2, S3, S4, and S5

Supplementary Figures S1 and S2

Supplementary Table S1. Sample information and quality of RNA extractions for each transcriptome sequenced. Precise sampling locations in French Guiana are unavailable because specimens from different localities were maintained together in captivity.

Species	Specimen voucher	Sampling location	Sex	Tissue	SRA database number	RIN value	Number of reads
Caecilia tentaculata (Caeciliidae)	MW 10281	Kaw Mountains or Nouragues	Male	Spleen	SRR5591446	9.9	49757992
				Foregut	SRR5591452	8.8	54057168
				Heart	SRR5591451	9.8	56478032
				Kidney	SRR5591454	9.6	51051562
				Liver	SRR5591453	9.1	57568844
				Muscle	SRR5591447	9.5	58911078
				Posterior Skin	SRR5591450	8.7	43278838
				Skin	SRR5591449	10	40889140
				Testis	SRR5591445	9.5	56700716
				Lung	SRR5591448	9.5	46260502
Microcaecilia dermatophaga (Siphonopidae)	MW 10280	Angouleme or St Laurent	? (juvenile)	Kidney	SRR5591430	9.5	42750256
				Liver	SRR5591429	9.3	56029426
				Posterior Skin	SRR5591428	9.2	39517710
				Skin	SRR5591427	9.6	43291236
Microcaecilia unicolor (Siphonopidae)	MW 3338	Kaw Mountains or Nouragues	Female	Kidney	SRR5591439	8.9	52616356
				Liver	SRR5591440	8.2	58062338
				Skin	SRR5591437	7.8	32957690
	MW 10282	Kaw Mountains or Nouragues	Female	Foregut	SRR5591426	9.6	47417282
				Liver	SRR5591425	8.9	55237106
				Muscle	SRR5591423	8.6	53570730
				Posterior Skin	SRR5591422	8.2	49216256
				Skin	SRR5591421	8.9	40774934
				Lung	SRR5591424	8.9	48841504
Rhinatrema bivittatum (Rhinatrematidae)	MW 3339	Kaw Mountains or Nouragues or Angouleme	Male	Kidney	SRR5591443	9.2	48775398
				Liver	SRR5591420	8.8	51635654
				Skin	SRR5591419	9.7	39705618
				Testis	SRR5591432	8.9	53962268
	MW 10279	Kaw Mountains or Nouragues or Angouleme	Female	Spleen	SRR5591442	10	56676152
				Foregut	SRR5591438	8.6	47607116
				Liver	SRR5591435	8.8	47561092
				Muscle	SRR5591433	9.4	52900438
				Skin	SRR5591434	8.3	39106530
				Lung	SRR5591436	9.1	51182452
Typhlonectes compressicauda (Typhlonectidae)	MW 10283	Matoury or St George	Male	Foregut	SRR5591431	9.8	53680574
				Heart	SRR5591416	9.4	53159450
				Kidney	SRR5591415	9.3	45034380
				Liver	SRR5591418	9.8	51838608
				Posterior Skin	SRR5591444	10	54023400
				Skin	SRR5591441	9.9	36051312
				Lung	SRR5591417	8.7	44973848

MW, field tag of specimen deposited at The Natural History Museum, London (UK).

Supplementary Table S2. Metrics of the species-specific transcriptomes. Candidate protein-coding genes or contigs with open reading frames (ORFs) found in each caecilian species-specific transcriptome are shown. The alignment percentage represents the amount of pair-end reads that correctly align to each species-specific transcriptome. For the contigs with ORFs, their number (Unique ORFs) and their total number considering all isoforms and post-transcriptional modifications (ORFs' isoforms, note that transcriptomes are built from different tissues) are displayed.

Species	Contigs	N50	Median contig length	Mean contig length	Alignment percentage	ORFs' isoforms	Unique ORFs
Caecilia tentaculata	142,502	1884	429	932.82	96.01	63,540	27,384
Microcaecilia dermatophaga	106,298	1784	426	903.73	97.78	42,510	22,058
Microcaecilia unicolor	146,348	1587	355	850.91	96.93	59,355	26,302
Rhinatrema bivittatum	201,584	1713	398	857.76	96.33	83,643	34,654
Typhlonectes compressicauda	134,394	1263	357	713.87	96.25	59,151	27,603

Supplementary Table S3. Detection of protein-protein interactions (PPIs) and functional enrichment paths in the tissue-specific known vertebrate gene families. The nodes represent proteins and the edges represent PPIs. PPI enrichment p-value is obtained from the STRING enrichment test that compares the obtained number of edges against an expected number of interactions.

		Foregut	**Kidney**	**Liver**	**Spleen**	**Testis**
Known vertebrate gene families		19	21	18	11	80
Number of nodes		17	20	11	11	75
Number of edges		5	5	6	3	19
PPI enrichment p-value		2.76e-06	1.31e-05	0.00258	4.75e-05	4.82e-07
Functional enrichment	Biological Process GO (#)	GO:0007586-Digestion (8)	GO:0098656 - anion transmembrane transport (6); GO:1903825 - organic acid transmembrane transport (5); GO:0003333 - amino acid transmembrane transport (4); GO:0046942 - carboxylic acid transport (5); GO:0015889 - cobalamin transport (2)	GO:0030193 - regulation of blood coagulation (3); GO:0051918- negative regulation of fibrinolysis (2); GO:0072376 - protein activation cascade (3)	GO:0004252 - serine-type endopeptidase activity (4)	-
	KEGG PATHWAY (#)	-	-	-	04972 - Pancreatic secretion (3)	-

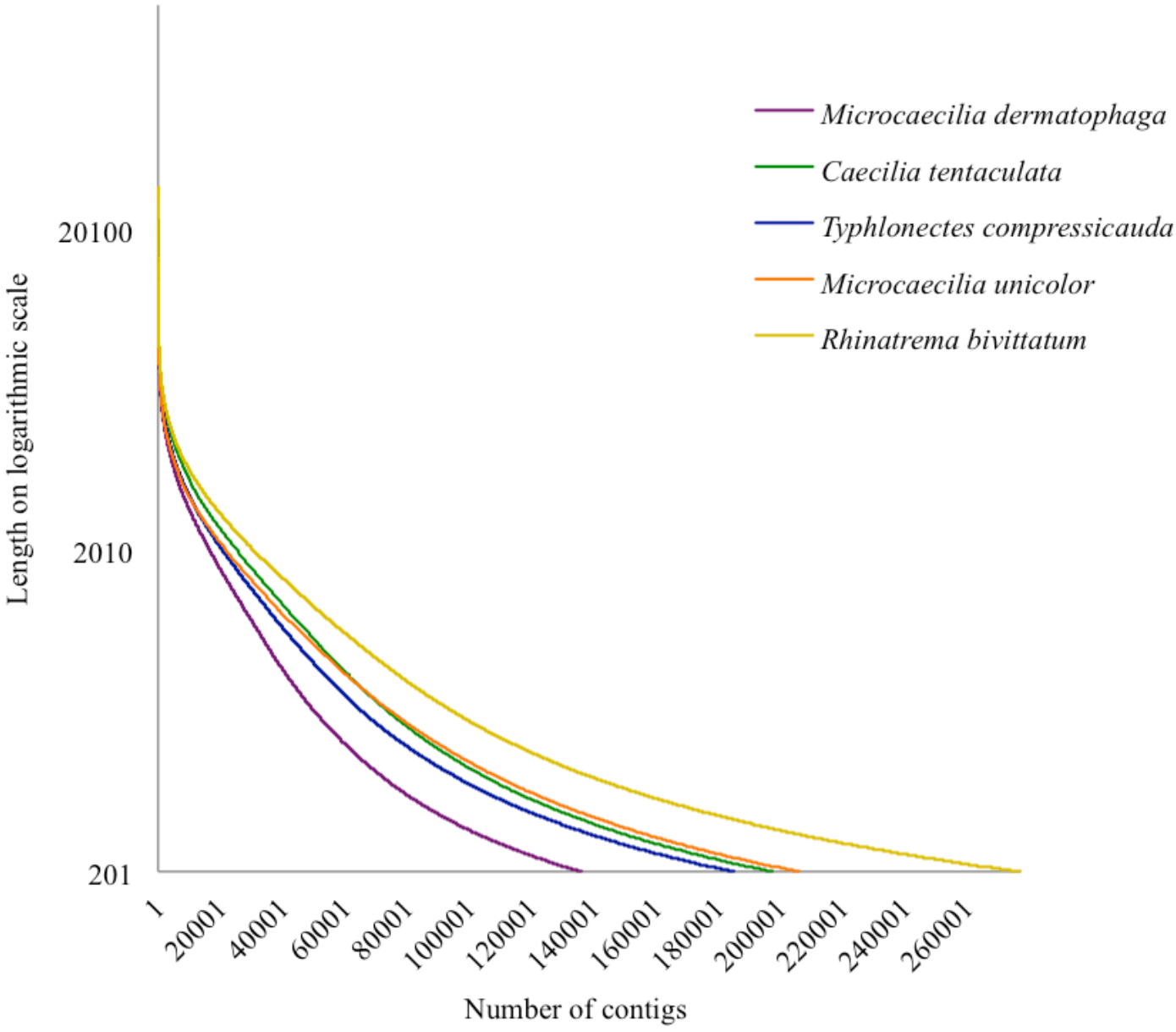
Supplementary Table S4. Results of the Pfam annotation for the candidate novel caecilian gene families with tissue-specific expression.

Tissue	Putative novel caecilian gene families	Annotated novel caecilian gene families	Number of protein family domains	Protein family domain
Foregut	32	3	6	Dimer_Tnp_hAT
				Snapin_Pallidin
				TMF_DNA_bd
				DAP10
				Adeno_E3_CR2
				PEARLI-4
Heart	12	2	4	BORCS8
				ReosigmaC
				TMEM190
				TEX29
Kidney	40	6	14	FISNA
				Tropomyosin
				ATG16
				IncA
				Apolipoprotein
				BMFP
				Tektin
				XhlA
				EzrA
				SlyX
				Cadherin_pro
				IGF
				Exo_endo_phos 2
				LAP2alpha
Liver	44	6	7	NinD
				KRAB
				zf-met
				LAP2alpha
				adh_short
				CENP-P
				Transposase_22
Lung	9	1	1	LAP2alpha
Muscle	27	4	37	IncA
				ATG16
				EMP24_GP25L
				DUF4600
				Fib_alpha
				CLZ
				FlaC_arch
				BRE1
				XhlA
				Nsp1_C
				Reo_sigmaC
				DUF1664
				Spc7
				ABC_tran_CTD
				EzrA
				Laminin_II
				Apolipoprotein
				DUF812
				CENP-F_leu_zip
				YjcZ
				EspB
				Muted
				KASH_CCD
				Spectrin

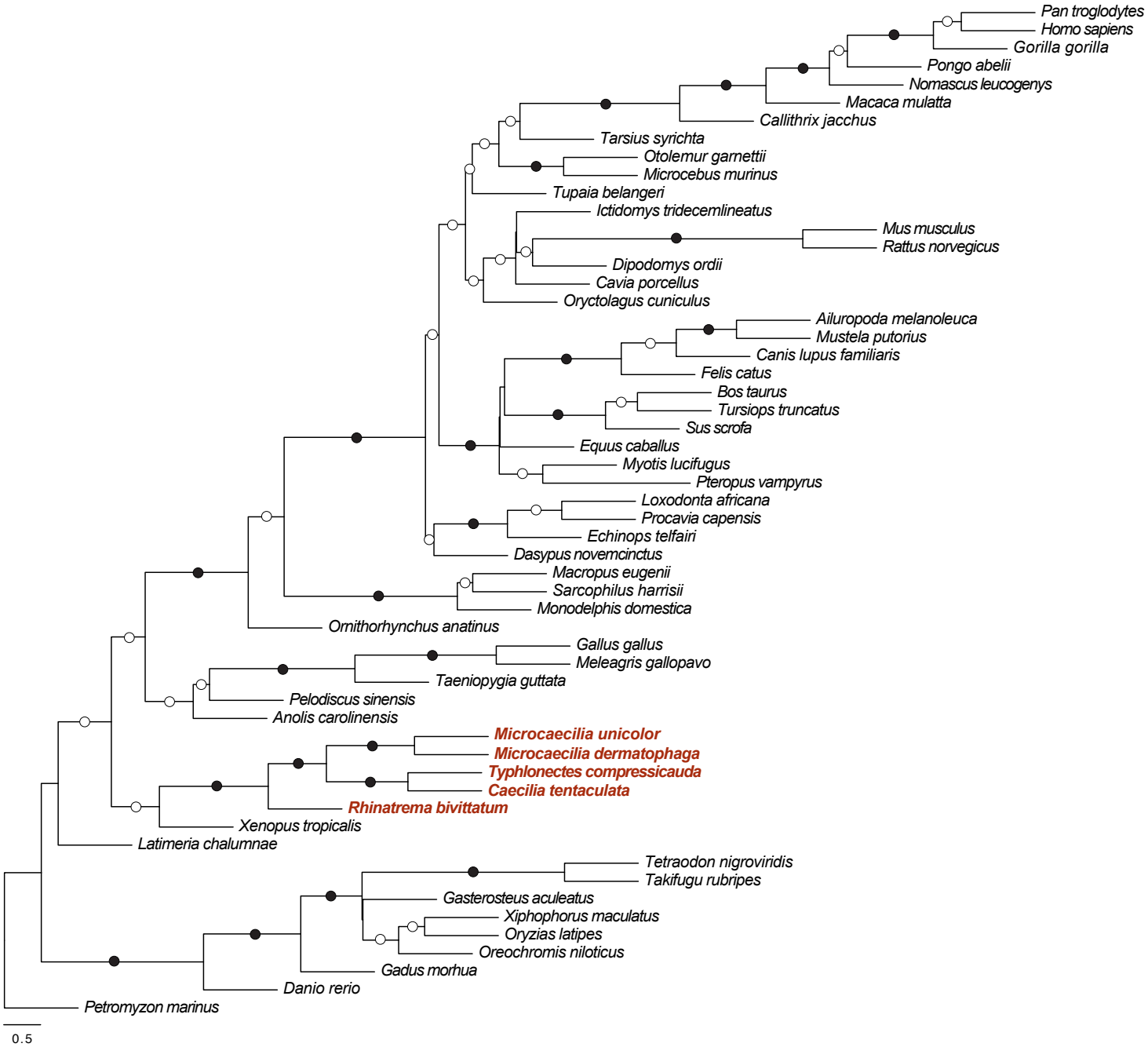
				TMF_DNA_bd
				IFT57
				Prefoldin
				ADIP
				ERM
				Jnk-SapK_ap_N
				MscS_porin
				SlyX
				TOBE_2
				LAP2alpha
				DUF3584
				Kre28
Skin	108	13	20	Ank_5
				KIX_2
				RVT_1
				LAP2alpha
				zf-RVT
				UPAR_LY6
				Toxin_TOLIP
				DUF4061
				DUF724
				DUF4407
				OmpH
				Tup_N
				PhaP_Bmeg
				Asp_protease_2
				gag-asp_proteas
				C1-set
				DUF4381
				DUF4630
				Spectrin
Spleen	8	2	3	Cystatin
				ATP1G1_PLM_MAT8
				SQAPI
Testis	142	12	14	Spc7
				NPV_P10
				DUF1664
				Reo_sigmaC
				CD225
				Myb_DNA-bind_5
				Glyco_transf_11
				RVT_1
				Ferrochelatase
				TIL
				Kazal
				Peptidase_M14
				zf-CCHC
				Serpin

Supplementary Table S5. Number of orthologous genes found in each vertebrate species (out of 1,955). Caecilian species are highlighted in bold type.

	Class	Taxon occupancy
Caecilia tentaculata	**Amphibia**	**333**
Microcaecilia dermatophaga	**Amphibia**	**240**
Microcaecilia unicolor	**Amphibia**	**314**
Rhinatrema bivittatum	**Amphibia**	**357**
Typhlonectes compressicauda	**Amphibia**	**304**
Xenopus tropicalis	Amphibia	134
Gallus gallus	Aves	204
Meleagris gallopavo	Aves	165
Taeniopygia guttata	Aves	215
Petromyzon marinus	Cephalaspidomorphi	83
Latimeria chalumnae	Coelacanthiformes	246
Ailuropoda melanoleuca	Mammalia	536
Bos taurus	Mammalia	477
Callithrix jacchus	Mammalia	701
Canis lupus familiaris	Mammalia	520
Cavia porcellus	Mammalia	426
Dasypus novemcinctus	Mammalia	370
Dipodomys ordii	Mammalia	387
Echinops telfairi	Mammalia	421
Equus caballus	Mammalia	464
Felis catus	Mammalia	493
Gorilla gorilla	Mammalia	966
Homo sapiens	Mammalia	888
Ictidomys tridecemlineatus	Mammalia	430
Loxodonta africana	Mammalia	448
Macaca mulatta	Mammalia	648
Macropus eugenii	Mammalia	248
Microcebus murinus	Mammalia	553
Monodelphis domestica	Mammalia	318
Mus musculus	Mammalia	521
Mustela putorius	Mammalia	422
Myotis lucifugus	Mammalia	389
Nomascus leucogenys	Mammalia	794
Ornithorhynchus anatinus	Mammalia	299
Oryctolagus cuniculus	Mammalia	448
Otolemur garnettii	Mammalia	451
Pan troglodytes	Mammalia	818
Pongo abelii	Mammalia	876
Procavia capensis	Mammalia	433
Pteropus vampyrus	Mammalia	551
Rattus norvegicus	Mammalia	523
Sarcophilus harrisii	Mammalia	339
Sus scrofa	Mammalia	462
Tarsius syrichta	Mammalia	401
Tupaia belangeri	Mammalia	439
Tursiops truncates	Mammalia	545
Anolis carolinensis	Reptilia	185
Pelodiscus sinensis	Reptilia	220
Danio rerio	Teleostei	341
Gadus morhua	Teleostei	302
Gasterosteus aculeatus	Teleostei	327
Oreochromis niloticus	Teleostei	272
Oryzias latipes	Teleostei	280
Takifugu rubripes	Teleostei	204
Tetraodon nigroviridis	Teleostei	255
Xiphophorus maculatus	Teleostei	346



Supplementary Figure S1. Contig lengths for the five studied species-specific caecilian transcriptomes.



Supplementary Figure S2. Supertree of vertebrates reconstructed from 1,955 orthologous gene trees using ASTRAL. Filled bullets on branches denote strong support as measured by both posterior probabilities (≥0.95) and quartet percentages (≥70%) for the respective internal branches. Open bullets denote strong support for posterior probabilities (≥0.95) but not for quartet percentages (<70%). Absence of bullet on a branch denotes weak support as measured by both posterior probabilities (<0.95) and quartet percentages (<70%). Scale bar indicates substitutions per site. The five sampled caecilian species are highlighted.