Genome assembly of the numbat (*Myrmecobius fasciatus*), the only termitivorous marsupial

Emma Peel[1], Luke Silver[1], Parice Brandies[1], Takashi Hayakawa[2,3], Katherine Belov[1] & Carolyn J Hogg [1]*

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia

[2]Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Hokkaido, Japan

[3]Japan Monkey Centre, Inuyama, Aichi, Japan

emma.peel@sydney.edu.au ORCID: 0000-0002-2335-8983

luke.silver@sydney.edu.au ORCID: 0000-0002-1718-5756

parice.brandies@sydney.edu.au ORCID: 0000-0003-1702-2938

hayatak@ees.hokudai.ac.jp ORCID: 0000-0002-1037-0350

kathy.belov@sydney.edu.au ORCID: 0000-0002-9762-5554

*carolyn.hogg@sydney.edu.au ORCID: 0000-0002-6328-398X

Abstract

The numbat (*Myrmecobius fasciatus*) is a critically endangered Australian marsupial, and the last surviving member of the Myrmecobiidae family. The numbat regularly undergoes torpor and is unique amongst marsupials as it is the only diurnal and termitivorous species. Here we sequenced the first draft genome of the numbat using 10x Genomics chromium linked-read technology, resulting in a 3.42 Gbp genome with a scaffold N50 of 223 Kbp. A global transcriptome from liver, lung and tongue was also generated to aid genome annotation with Fgenesh++, identifying 21,465 protein-coding genes and 78.7% complete mammalian BUSCOs. To investigate adaptation to the numbat's termitivorous diet and arid/semi-arid range, we interrogated the most highly expressed transcripts within the tongue and manually annotated taste, vomeronasal and aquaporin gene families. Antimicrobial proteins and proteins involved in digestion were highly expressed in the tongue, as expected. Similarly, umami taste receptors were also expressed in the tongue, however sweet taste receptors were not expressed in this tissue. This pattern of taste receptor expression, combined with the putative contraction of the bitter taste receptor gene repertoire in the numbat genome, may indicate a potential evolutionary adaptation to their specialized termitivorous diet.

Vomeronasal and aquaporin gene repertoires were similar to other marsupials and have not undergone expansion or contraction. The draft numbat genome is a valuable tool for conservation and can be applied to population genetics/genomics studies and to investigate the unique biology of this interesting species.

Research area

Genomics

Classifications

Animal genetics and genetics

Data description

Background and context

The numbat (*Myrmecobius fasciatus*) is a small marsupial (up to 700g), and the only species within the Myrmecobiidae family [1] (Figure 1). Marsupials are one of three lineages of mammals, the others being eutherians (such as humans and mice) and monotremes (platypus, *Ornithorhynchus anatinus* and echidna, *Tachyglossus aculeatus*) [2]. The Myrmecobiidae family is classified under the Dasyuromorphia order, which also contains carnivorous marsupials within the Dasyuridae family such as the Tasmanian devil (*Sarcophilus harrisii*) and the extinct thylacine (*Thylacinus cynocephalus,* the family Thylacinidae) [3]. Numbats are unique amongst marsupials as they are the only diurnal and marsupial. In addition, numbats have a variety of important adaptations to an arid environment including regular torpor.

Figure 1. Adult numbat from the Australian Wildlife Conservancy. Photo credit Wayne Lawler Australian Wildlife Conservancy.

Numbats are the only termitivorous marsupial, consuming up to 20,000 termites per day [4]. Their high visual acuity, powerful front claws and sense of smell enables them to locate and dig out termite mounds and sub-surface structures [4]. Their extremely long tongues allow numbats to then extract termites from within the mounds. This bioturbation is important for ecosystem health, as

55 numbats aerate the soil, facilitate seed germination and remove termite mud from hollows thereby
56 creating habitat for other species [5].

57 The numbat and other marsupials within the Dasyuromorphia, Notoryctemorphia and Diprotodontia
58 orders undergo torpor [6]. Numbats are heterothermic endotherms, and regularly undergo shallow
59 torpor to conserve energy during winter, characterised by a drop in body temperature for up to 15
60 hours [7]. Marsupials typically have a low basal metabolic rate compared to eutherian mammals [8].
61 However, numbat basal metabolic rate is even lower, at 82.5% of other marsupials with an
62 equivalent body mass [9, 10]. During torpor, metabolic rate can drop by up to 60% below the basal
63 rate [6].

64 Historically, numbats inhabited the arid and semi-arid regions of Australia [11]. However,
65 populations have declined by more than 99% due to habitat degradation and predation. It is
66 estimated there are only 1000 individuals remaining in Western Australia [12]. As such, the numbat
67 is currently listed as endangered ("EN") on the IUCN Red List [13] and vulnerable under the
68 Australian Federal Government's Environment Protection and Biodiversity Conservation Act 1999.
69 Numerous reintroductions to wild populations were conducted between 1985 and 2010, however
70 due to predation by introduced cats and foxes many of these were failures, with only four being
71 successful [14, 15]. Due to the ongoing threat of predation, numbats have also been released into
72 several large-scale fenced enclosures to ensure the ongoing persistence of the species [16]. The
73 current numbat recovery plan recommends additional subpopulations be established and the
74 genetic health of all populations be maintained and measured [17].

75 Here we report the first *de novo* reference genome for the numbat using 10X Genomics chromium
76 linked-read sequencing. Assembly resulted in a 3.42 Gb genome with a scaffold N50 of 223Kb and
77 78.7% complete mammalian benchmarking universal single copy gene orthologs (BUSCOs v5.2.2)
78 with 73.2% single-copy and 5.5% duplicated BUSCO genes [18]. A global transcriptome was also
79 generated, consisting of transcripts from the liver, lung, and tongue. This was used to annotate the

genome with Fgenesh++, resulting in 21,465 annotated genes with BLAST hits to the NCBI non-redundant database. Taste and vomeronasal receptors and aquaporin genes were manually annotated within the genome to investigate whether these gene families have expanded or contracted within the numbat compared to other marsupials in response to their unique life history. Annotation revealed a typically marsupial complement of vomeronasal and aquaporin gene families within the numbat which does not reflect adaptation to their arid range. However, a subset of taste receptor genes have contracted in the numbat compared to other marsupials, which may reflect adaptation to a termitivorous diet.

The numbat reference genome is a valuable tool for conservation and will be used alongside population genomic and genetic datasets to measure neutral and functional genetic diversity and health of current and future populations. Obligate termitivorous mammals occur in multiple eutherian lineages including the American anteaters, African aardvark, and the monotreme echidna. The numbat genome and transcriptomes generated in this study provide new insights into the molecular mechanisms which underpin the convergent evolution of this specialised dietary adaptation, and unique physiology of this iconic marsupial.

Methods

Sample collection and sequencing

Numbat liver, lung and tongue were opportunistically sampled from a single female individual housed at Perth Zoo which was euthanized due to medical reasons in 2019. All tissues were flash frozen at -80^{o}C and stored at this temperature until extraction. All samples were collected under Perth Zoo's opportunistic sampling standard operating procedure (export licence EF41000060) and scientific licence number NSW DPIE SL101204.

High molecular weight (HMW) DNA was extracted from 25mg of lung using the MagAttract HMW DNA kit (Qiagen) and quality was assessed using the NanoDrop 6000 with an A260/280 of 1.8 and A260/230 of 1.3. DNA was submitted to the Ramaciotti Centre for Genomics (UNSW) for 10x Genomics chromium library prep, and 150bp paired-end (PE) reads were sequenced on an Illumina

NovaSeq 6000 S1 flowcell. This generated 143GB of raw data, which was quality checked using fastQC v0.11.8 (RRID:SCR_014583) [19].

Total RNA was extracted from 25mg of liver, lung and tongue, using the RNeasy Plus Mini Kit (Qiagen) with on-column DNA digestion using the RNase-free DNase I set (Qiagen). For the tongue, precise isolation of microscopic taste buds was difficult, but these structures are likely to have been included in the section of the tongue surface sampled. RNA purity was assessed using the NanoDrop 6000, with all samples displaying an A260/280 and A260/230 of 1.95 to 2.34. RNA concentration and integrity were measured using a RNA nano 6000 chip (Agilent Technologies), with all samples displaying an RNA integrity number (RIN) from 7 to 8.9. Total RNA was submitted to the Ramaciotti Centre for Genomics (UNSW) for TruSeq mRNA library prep. All tissue libraries were sequenced as 150bp PE reads across one lane of a S1 flowcell on the NovaSeq 6000. This resulted in 22-29GB raw data per sample, which was quality checked using fastQC v0.11.8 (RRID:SCR_014583) [19].

Genome assembly and annotation

De novo genome assembly was performed with Supernova v2.1.1 [20] using default parameters on Amazon Web Services (virtual machine 64 vCPUs; 976 GB RAM; 3 TB storage), obtaining approximately 64x raw coverage and 31x effective coverage. Assembly statistics were generated using BBTools (RRID:SCR_016968) [21], and assembly completeness assessed using BUSCO v5.2.2 and v3.1.0 (RRID:SCR_015008) [18]. The assembly was filtered to remove redundant haplotigs using SLiMSuite v1.8.1 (https://github.com/slimsuite/SLiMSuite) [22]. Read representation was determined by trimming 10x adapters from the raw reads using BBmap (RRID:SCR_016965) [21] which were then mapped back to the assembly using BWA (RRID:SCR_010910) [23]. For annotation, a custom repeat database was generated for the genome using RepeatModeler v2.0.1 (RRID:SCR_015027) [24], then RepeatMasker v4.0.6 (RRID:SCR_012954) [25] used to mask repeats, excluding low complexity regions and simple repeats. Genome annotation was then performed using Fgenesh++ v7.2.2 (RRID:SCR_018928) [26] with general mammalian pipeline parameters, and an optimised gene finding matrix from another species within the Dasyuromorphia order (Tasmanian

devil: *Sarcophilus harrisii*). Transcripts with the longest open reading frame for each predicted gene were extracted from the global transcriptome and used as mRNA-based evidence for gene predictions. Similarly, the non-redundant metazoan protein database was used as protein-based evidence for gene predictions.

Transcriptome assembly and annotation

Raw RNAseq data was quality and length trimmed using Trimmomatic v0.38 (RRID:SCR_011848) [27] with the following flags: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25. Illumina TruSeq sequencing adapters were removed from the dataset (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10) as well as reads shorter than 25bp (MINLEN:25). Reads were quality trimmed and removed where the average quality score fell below 5 within a 4 base pair sliding window (SLIDINGWINDOW:4:5), as well as at the 5' (LEADING:5) and 3' (TRAILING:5) end of the read. Over 99.7% of reads were retained for all datasets post-trimming.

A global transcriptome for the numbat was generated *de novo* using trimmed reads from liver, lung and tongue as input to Trinity v2.8.3 (RRID:SCR_013048) [28] with default parameters and read normalization. The Trinity script TrinityStats.pl was used to generate assembly statistics, representation of full-length protein-coding genes was determined by BLAST to Swiss-Prot, and completeness was assessed using BUSCO v5.2.2 and v3.1.0 (RRID:SCR_015008) [18] against the mammalian database. Functional annotation of the global assembly was performed using Trinotate v3.1.1 (RRID:SCR_018930) [29]. Briefly, TransDecoder v2.0.1 (RRID:SCR_017647) [28] was used to identify coding regions within transcripts, which in addition to the global assembly transcripts, were used to search the Swiss-Prot non-redundant database, the Tasmanian devil reference genome annotations downloaded from NCBI (mSarHar1.11) and the immunome database of marsupials and monotremes [30] using BLAST+ [31] with an e-value cut-off of $1e^{-5}$ and HMMER v3.2.1 (RRID:SCR_005305). To determine the proportion of reads represented in the assembly, trimmed reads were mapped back to the global transcriptome assembly using bowtie2 v2.3.3.1 (RRID:SCR_016368) [32] with the flag -k 20 to indicate a maximum of 20 distinct alignments for each

read. Alignments were used as input to transcript quantification with Salmon v1.4.0 [33] to generate transcript per million (TPM) counts for each tissue. Transdecoder-proteins expressed in the tongue with BLASTp hits to Swiss-prot (e-value of e^{-5}) were used as input to Panther (RRID:SCR_004869) [34] to assign gene ontology (GO) slim terms under the Biological Process and Molecular Function category.

Manual gene annotation

Genes encoding taste and vomeronasal receptors, and aquaporins, were manually annotated in the numbat genome and transcriptomes to investigate gene expansion and/or contraction as a mechanism of evolutionary adaptation to the numbat's unique diet and historically arid range. Briefly, BLAST+ v2.7.1 [31] searches were conducted using known marsupial and eutherian sequences from each gene families as queries, with an e-value cut-off of 10 to ensure any potential hits here not excluded. Putative numbat genes for each family were aligned to other members from eutherians and marsupials using clustalW (RRID:SCR_017277) [35] in BioEdit (RRID:SCR_007361) [36] to confirm expected gene structure and presence of functional protein motifs. For TAS1R and vomeronasal receptor gene families the multiple sequence alignment was then used to construct phylogenetic trees in MEGAX v10.2.4 (RRID:SCR_000667) [37] for each family separately using neighbour-joining method with p-distance and 500 bootstrap replicates, as well as the maximum likelihood method with the James-Thornton-Taylor model. Both methods resulted in the same topology, so only the neighbour-joining trees are presented here. Phylogenetic analysis of TAS2R receptors was conducted in MEGAX v10.2.4 using the maximum likelihood method with the General Time Reversible model and gamma distribution with five categories. Bootstrap analysis was not performed as the topology of marsupial TAS2R sequences has been established previously [38]. Vomeronasal receptor genes were named in order of identification. Taste receptors and aquaporin genes were named according to clustering and bootstrap support within the phylogenetic tree. Extant marsupials have at least 27 orthologous gene groups (OGGs) of bitter taste receptor genes (TAS2Rs) [38]. TAS2R genes identified in the numbat genome were classified into marsupial OGGs

184 based on their clustering with other known marsupial TAS2R genes within the phylogenetic tree in
185 Figure 4. Accession numbers of sequences used as queries in BLAST+ and to generate phylogenetic
186 trees are available in Supplementary table 1

187 ## Results and discussion

188 ### Genome

189 *De novo* assembly and subsequent filtering generated a 3.42Gb genome for the numbat with 30.97x
190 coverage (Table 1). The genome contains 112,299 scaffolds with a scaffold N50 of 223kb (Table 1)
191 and is of a similar quality to the Tasmanian devil genome [39], but less contiguous than the
192 antechinus (*Antechinus stuartii)* [40] and koala (*Phascolarctos cinerus)* genomes [38] (Table 1). The
193 koala genome was generated using PacBio long-reads and multiple scaffolding technologies, so it is
194 not surprising this was more contiguous than the numbat assembly [38]. However, the antechinus
195 genome was generated using 10x Genomics chromium linked-reads and assembled using the same
196 Supernova v2.1.1 pipeline, yet was also more contiguous than the numbat genome [40] (Table 1).
197 This difference in assembly contiguity may arise from molecule length, which represent reads with
198 the same 10x barcode that align to the same region of a contig or scaffold [41]. This metric is an
199 important contributing factor to the quality of 10x linked-read assemblies, with shorter molecules
200 associated with reduced scaffold N50 and mis-assembly [41]. The antechinus genome had a
201 molecule length of 74.08kb [40], compared to only 23.13kb for the numbat genome and below the
202 recommended range of 50-100kb by 10x Genomics [41]. HMW DNA >40kb was used as input to
203 sequencing in both species, although different extraction methods and kits were used in both cases
204 which may have contributed to the difference in molecule length [40]. In addition, the numbat HMW
205 DNA may have degraded during transport, storage, or sequencing, leading to fragmentation.

206 Table 1. Numbat genome assembly statistics compared to koala, antechinus, Tasmanian devil,
207 tammar wallaby and gray short-tailed opossum genomes accessioned with NCBI.

	Numbat	Koala	Antechinus	Tasmanian devil	Tammar wallaby	Gray short-tailed opossum

Reference.	This study	[38]	[40]	[42]	[43]	[44]
Accession no.	-	GCF_002099425.1	GCA_016696395.1	GCF_000189315.1	GCA_000004035.1	GCF_000002295.2
Year	2021	2018	2020	2012	2011	2007
Sequencing technology	10X linked-reads	PacBio, Illumina, BioNano and HiC	10X linked-reads	Illumina	Sanger ABI SOLiD	Whole-genome shotgun (WGS) method
Genome size (Gb)	3.42	3.19	3.31	3.17	3.07	3.59
No. scaffolds	112,299	-	30,876	35,974	277,711	5,223
No contigs	219,447	1,906	106,199	237,291	1,174,000	72,674
Scaffold N50 (Mb)	0.223	-	72.7	1.8	0.0418	59.8
Contig N50 (Mb)	0.037	11.58	0.08	0.02	0.0026	0.108
GC (%)	36.3	39.05	36.20	36.04	38.80	38.00
Gaps (%)	3.52	0.1	2.75	7.66	17.53	2.74

208

209 The repeat content of the numbat genome was similar to other marsupials, with 47.63% of the
210 genome masked as repeats compared to 44.82% in the antechinus [40] and 47.5% in the koala [38].
211 Several different repeat families were identified within the numbat genome (Table 2). Class I long
212 interspersed nuclear elements (LINE1) and mammalian-wide interspersed repeats (MIRs) were the
213 most numerous as identified in other marsupials [38, 40, 42].

214 Table 2. Repeats elements annotated in the numbat genome.

Type of repeat		**Number**	**Repeat sequence**	
			Length (bp)	**Percent**
SINE	ALU	11577	2119505	0.06%
	MIRs	1718103	243123222	7.10%
LINE	LINE1	1402806	733512048	21.42%
	LINE2	807962	176657028	5.16%
	CR1	270927	61050371	1.78%
LTR	ERVL	2082	762885	0.02%
	ERV1	21263	6393956	0.19%
	ERV2	16540	5447399	0.16%
DNA elements	hAT-Charlie	120449	17122026	0.50%

	TcMar-Tigger	20819	5373403	0.16%
Other	Unclassified	1034362	236046652	6.89%
	Small RNA	641	57442	0.00%
	Satellites	56304	14763428	0.43%

Annotation of the numbat genome with Fgenesh++ resulted in 77,806 gene predictions, of which 44,056 were supported by transcript evidence from the global transcriptome and 1,406 by protein evidence. Of these 77,806 genes, 21,465 had BLAST hits to eukaryote genes in the NCBI non-redundant database, which is similar to the number of annotated protein-coding genes in the Tasmanian devil (18,775) [39], antechinus (25,111) [40] and koala (26,558) [38].

Transcriptome

The *de novo* global transcriptome containing transcripts from the liver, lung and tongue contained 2,119,791 transcripts with an average length of 824bp and transcript N50 of 1,393bp. TransDecoder predicted 159,566 coding regions, of which 63% were complete (contained start and stop codon) and 86% had BLAST hits to Swiss-Prot. Given the numbat's specialized diet, we investigated the top 10 transcripts expressed within the tongue with BLAST hits to Swiss-Prot. Antimicrobial proteins from the S100 protein family and L-amino-acid oxidase [45] were highly expressed within the tongue (Figure 2). In addition, transcripts encoding keratin 13 and 4 which form part of the cytoskeleton were also highly expressed, as well as myosin and actin proteins involved in muscle contraction (Figure 2). These transcripts reflect the structure and function of the tongue not only in feeding, but also as an epithelial barrier that forms a first line of defence against infection.

Figure 2. Top 10 transcripts expressed in the tongue with BLAST hits to Swiss-Prot proteins. Papillae were not discriminated at the time of tissue sampling.

To further investigate adaptation to the numbat's unique life history, we manually annotated genes involved in taste, olfaction and water transport in the numbat genome. Duplication or pseudogenization of genes within taste receptor, vomeronasal and aquaporin families have been identified in species with specialized diets. Bitter taste receptor and aquaporin genes have

duplicated in the koala, likely due to their need to detoxify eucalypt leaves and ability to obtain water solely from their diet without drinking [38]. Neofunctionalization and pseudogenization of taste receptor genes has also been linked to highly specialized diets. In large primates which eat a high proportion of leaves, including humans, the umami taste receptor gene has undergone neofunctionalization to recognize L-glutamate contained within leaves, which may promote leaf consumption [46]. In addition, bitter taste receptor genes which recognize leaf-derived toxins have also duplicated in some eutherian lineages [47]. On the other hand, in the giant panda (*Ailuropoda melanoleuca)*, the umami taste receptor is not required for their herbivorous diet of bamboo, hence the umami taste receptor gene is a pseudogene [48]. Our aim was to determine if similar duplication or pseudogenization has occurred in the taste, vomeronasal and aquaporin gene families in the numbat genome in response to their termitivorous diet and semi-arid environment.

Taste receptors

Two types of taste receptors are encoded within the genome and expressed within the mammalian oral cavity; type 1 (TAS1R) umami and sweet taste receptors, and type 2 (TAS2R) bitter taste receptors. Type 1 taste receptors are encoded by three genes in most mammals including marsupials and monotremes, TAS1R1, TAS1R2 and TAS1R3 [38, 49]. The TAS1R1/TAS1R3 heterodimer and TAS1R2/TAS1R2 homodimer generate functioning umami and sweet taste receptors respectively [50].

Orthologs of all three mammalian TAS1R genes were identified in the numbat genome (Figure 3, Supplementary table 2). Only partial sequences were characterised for TAS1R1 and TAS1R3 as short internal exons could not be identified. The TAS1R gene cluster was fragmented within the genome, as all three genes were encoded on different scaffolds compared to the single gene cluster in human and mouse [38, 50]. All TAS1R genes are likely functional in the numbat, owing to the lack of premature stop codons, insertions or deletions within the gene sequence, and identification of transcripts within the global transcriptome, pseudogenization has not occurred in response to their specialized diet as observed in the giant panda [48]. However, the expression of TAS1R1 and TAS1R3

in the numbat tongue transcriptome (TAS1R1 0.073573 TPM and TAS1R3 0.057388 TPM) may reflect their dietary preference for termites, as the TAS1R1/TAS1R3 heterodimer which forms the umami taste receptor recognizes free nucleotides that are abundant in insects [46]. TAS1R2 which forms the sweet taste receptor was not expressed in the tongue transcriptome, which may relate to the low level of free sugars in termites. While it is possible that TAS1R2 may be expressed in other areas of the tongue not sampled, functional TAS1R2 genes are known to be expressed in non-taste organs for carbohydrate metabolism [51].

Figure 3. Phylogenetic relationship amongst numbat (Myfa; yellow), Tasmanian devil (Saha; red), koala (Phci; green), gray short-tailed opossum (*Monodelphis domestica)* (Modo; purple) and human (Hosa; black) TAS1R genes. The neighbor-joining phylogenetic tree was constructed using the p-distance method and 1000 bootstrap values based on the amino acid sequence alignment. Only bootstrap values greater than 50% are shown.

Type 2 taste receptors (TAS2R) are G-protein coupled receptors which detect bitter substances [50]. The number of TAS2R genes varies between species, with 26 and 40 genes in humans and mice respectively [47]. Marsupials have several TAS2R genes which are orthologous to eutherians (TAS2R1, 2, 4, 38 and 60). However, some marsupials such as the koala have undergone large duplications within this gene family, with 66 TAS2R genes identified (24 intact and 42 pseudogenised) [38]. This expansion is thought to reflect adaptation to the koala's eucalypt diet, as duplications have occurred within TAS2R genes that presumably detect β-glucosides such as cyanogenic glycosides (TAS2R41 and 705) which are a component of eucalypt leaves [38]. Similarly, a large expansion of a sister OGG to eutherian and marsupial TAS2R41, TAS2R705 and TAS2R60 has also been identified in the echidna and platypus genome, indicating this TAS2R gene cluster is highly conserved across mammals [49].

22 TASR2 genes were identified in the numbat genome, of which 11 were putative pseudogenes owing to the presence of premature stop codons within the open reading frame (Figure 4, 5,

289 Supplementary table 2). While only two of the the 11 numbat TAS2R genes with complete coding
290 sequences were identified in the global transcriptome, this is likely due to the fact that TAS2Rs are
291 mainly expressed within papillae which were not discriminated at the time of tongue tissue sampling
292 [50]. The number of TAS2R genes in the numbat was the smallest amongst marsupials studied to
293 date, and more similar in size to monotremes [38] (Figure 4). Despite this, numbat TAS2R genes
294 cluster within the OGG clade containing other marsupial and eutherian TAS2R genes which may
295 detect harmful β-glucosides (TAS2R41, 60 and 705) contained within arthropods (Figure 5). The
296 insectivorous echidna (*Tachyglossus aculeatus)* has a similarly reduced TAS2R gene repertoire
297 compared to the platypus. However, the OGG of TAS2R41, 60 and 705 have also been retained in the
298 echidna, as observed in the numbat [49]. The presence of this OGG and overall reduction of the
299 TAS2R gene repertoire in these two species may reflect adaptation to an insectivorous diet.
300 However, TAS2R gene contraction in the numbat may also be due to assembly error. Numbat TAS2R
301 genes were encoded across six scaffolds compared to two main gene clusters in the human and
302 mouse genome [50]. Future improvements to numbat genome contiguity may uncover additional
303 taste receptor genes in the numbat, enabling reconstruction of gene clusters and investigation of
304 synteny within this genomic region.

305 Figure 4. Number of type 2 taste receptor genes in the numbat compared to other marsupials and
306 monotremes with intact coding sequences (CDS), truncated CDS or disrupted CDS which likely
307 represent pseudogenes. TAS2Rs with truncated CDS may represent incomplete sequences due to
308 short contigs and/or scaffolds, or pseudogenes.

309 Figure 5. Phylogenetic relationship amongst marsupial TAS2R genes. Numbat TAS2R genes are
310 indicated by the yellow circle and the other circle colours mirror Figure 3 with the addition of
311 tammar wallaby (*Notamacropus eugenii*) indicated by the blue circle. TAS2R sequences of non-
312 numbat marsupials were annotated in [38]. The phylogenetic tree was constructed using the
313 Maximum Likelihood method (GTL+G, 5 categories) based on the nucleotide sequence alignment.

Only topology is shown. The clades of 27 marsupial OGGs are indicated by the black circles around each respective node with the corresponding gene name. TAS2R gene names with a strikethrough (e.g. TAS2R701) indicate the corresponding numbat TAS2R gene is pseudogenised, and grey gene names indicate the corresponding numbat TAS2R gene is absent. The marsupial OGGs which may recognize β-glucoside are shaded in orange, note the numbat TAS2R genes within this clade (TAS2R60, 705 and 41) are intact within the genome.

Vomeronasal receptors

Vomeronasal receptors (VR) are a class of olfactory receptor primarily expressed in the vomeronasal organ (VNO) within the nose and involved in the detection of pheromones. There are two types of VR (V1R and V2R) encoded by separate gene families which differ in their expression pattern and gene structure [52]. Binding of pheromones or odorants to V1R and V2R initiates chemical sensing that has important roles in many behaviours such as mating and aggression. The number of genes encoding V1R and V2R differs amongst species, with many mammals displaying a discrepancy in the ratio of V1R to V2R genes [52].

Type 1 vomeronasal receptors (V1R) are involved in the detection of small pheromones within the air, such as those involved in sex and mating [52, 53]. V1Rs are encoded by an intronless gene and are primarily expressed within the apical layer of the epithelium within the VNO [52]. The majority of V1Rs in humans are pseudogenes [54], while rodents have a large expansion of more than 200 functional genes [55]. 162 V1R genes were identified in the numbat genome, of which 112 contained intact coding sequences and five were expressed in the global transcriptome (Supplementary table 2). The lack of expression for the majority of numbat V1Rs is not surprising given they are solely expressed within the VNO in other mammals [52]. The 112 putative functional numbat V1R genes were encoded across 68 different scaffolds. This is compared to large clusters of duplicated genes in the mouse genome, indicating the VR gene family was highly fragmented in the numbat genome similar to taste receptors.

The numbat V1R gene repertoire is similar to other marsupials and monotremes, with more than 90 genes identified in the opossum [53] and tammar wallaby [53, 55] and 280 genes in the platypus [55, 56]. V1Rs from the numbat and other marsupials form both marsupial-specific and species-specific clades in the phylogenetic tree, which may reflect marsupial-specific adaptations (Figure 6). For example, VRs are thought to be involved in the unaided movement of altricial marsupial young from the birth canal to the mother's pouch and teat [57].

Figure 6. Phylogenetic relationship amongst numbat (yellow), opossum (purple), human (black), mouse (blue), platypus (orange) and echidna (pink) type 1 vomeronasal receptor genes. Marsupial-specific clades are denoted by the red branches. The neighbor-joining phylogenetic tree was constructed using the p-distance method and 1000 bootstrap values based on the amino acid sequence alignment. Only bootstrap values greater than 50% are shown.

Type 2 vomeronasal receptors (VR2) are expressed in the basal layer of the VNO epithelium and detect water-soluble peptides and pheromones [52, 53]. Similar to V1Rs, V2R gene number varies significantly amongst species [52]. V2R genes have expanded in rodents, with more than 100 and functional genes and 150 pseudogenes [58]. In comparison, the platypus genome contains 15 V2Rs [58, 59], while humans and primates do not encode functional V2R genes [53, 54, 58].

29 V2R sequences were identified in the numbat genome, of which 22 likely represent functional genes (Supplementary table 2). The number of V2R genes in the numbat is low compared to the 86 functional (70 pseudogene) V2R genes in opossum [53]. The low number of V2R genes identified in the numbat may be due to assembly fragmentation, as V2R genes were encoded across 17 scaffolds many of which were short and only contained partial V2R sequences. However, V2Rs have not been manually annotated in other marsupial genomes which hinders our interpretation of these results.

Overall, numbats encode a number of functional type 1 and 2 vomeronasal receptors unlike many eutherian mammals. While V1R genes have expanded in numbats as in other marsupials, the lack of

363 annotated marsupial V2R genes limits our ability to identify if the low gene number in numbat
364 results from incomplete gene annotations or gene family contraction.

365 ## Aquaporins

366 Numbats historically inhabited the arid and semi-arid areas of southern and central Australia [12].
367 Unlike many arid marsupials, the numbat's renal morphology and urinary concentration does not
368 reflect this environment [60]. However, gene families involved in water metabolism such as
369 aquaporins have not been explored. Aquaporins are plasma membrane channels that are involved in
370 the transport of water and other small molecules, and are essential for water balance [61].
371 Aquaporin genes have undergone duplications in the koala, an adaptation to their highly specialized
372 diet of eucalypt leaves and ability to "taste water" [38].

373 12 aquaporin genes were identified in the numbat genome, including the water selective aquaporins
374 (AQP1, 2, 4, 5, 6 and 8), aquaglyceroporins (AQP3, 7, 9 and 10) and superaquaporins (AQP11 and 12)
375 [61] (Supplementary table 2). This gene content is identical to most marsupials [38] and similar to
376 the 13 aquaporin genes in humans [61]. All 12 numbat aquaporins were expressed in the global
377 transcriptome, and all but AQP2 were expressed in the tongue. AQP5 is highly expressed in the
378 tongue of other mammals as it is central to sensing water concentration [61]. This was not the case
379 for the numbat, as AQP5 was expressed in the tongue at low levels (0.013 TPM) whereas AQP3 was
380 the most highly expressed aquaporin in this tissue (62.9 TPM). The numbat aquaporin gene
381 repertoire is typical of other mammals and has not undergone gene duplications as observed in the
382 koala [38].

383 ## Data validation and quality control

384 Functional completeness of the genome and global transcriptome was assessed by searching for the
385 presence of single copy gene orthologs using BUSCO [18]. 78.7% and 76.4% of complete mammalian
386 BUSCO v5.2.2 genes, and 82.7% and 72% of complete mammalian BUSCO v3.1.0 genes were
387 identified in the genome and global transcriptome respectively. These BUSCO scores are lower than
388 reported for the koala [38], Tasmanian devil [39] and antechinus genomes [40], indicating the draft

389 status of the numbat genome. Despite this, 93.88% of input reads mapped to the genome assembly,
390 and more than 96% of trimmed RNAseq reads from each of the three tissues mapped to the global
391 transcriptome. This indicates that the genome and transcriptome assemblies are an accurate
392 representation of the input sequencing reads.

393 ## Re-use potential

394 The *de novo* assembly of the numbat genome using 10X Genomics chromium linked-reads resulted
395 in a 3.42Gb draft-quality genome. As the numbat is the sole member of the Myrmecobiidae family
396 and the only diurnal and termitivorous marsupial, this genome provides an opportunity to study the
397 genetic basis of these unique traits. The numbat genome is one of few arid marsupial genomes that
398 have been sequenced and represents an important contribution to studying adaptation to aridity,
399 particularly given climate change.

400 The numbat genome can immediately be used for conservation management through alignment of
401 population genetics datasets such as reduced representation sequencing, which will enable
402 monitoring of both genome-wide and functional genetic health and diversity of numbat populations.
403 Despite the fragmented nature of the genome, the draft numbat assembly enabled investigation of
404 taste, vomeronasal and aquaporin gene families in this unique marsupial and provides a basis for
405 future sequencing projects.

406 ## Conclusion

407 We have generated a draft genome assembly and global transcriptome assembly for the numbat,
408 the only member of the Myrmecobiidae family and only termitivorous marsupial. Given the
409 numbat's specialized diet, we investigated highly expressed transcripts within the tongue, and
410 manually annotated taste and vomeronasal receptors in the genome. The tongue contains
411 numerous transcripts involved in feeding and immunity, highlighting its role as a first line of defence
412 against foreign agents. The pattern of taste receptor expression in the tongue and putative
413 contraction of the bitter taste receptor gene repertoire in the numbat genome may reflect their
414 specialized termitivorous diet. Vomeronasal receptor gene families in the numbat did not show

evidence of gene expansion or contraction, as observed in other mammals with specialized diets. Similarly, numbat aquaporin genes were similar to other mammals and did not reflect adaptation to an arid environment. However, genome fragmentation influenced the quality of manual gene annotation and further work is required for confirmation. The numbat genome is an important resource for conserving this distinctive marsupial and understanding its unique life history and termitivorous diet.

Data availability statement

The numbat genome and global transcriptome assembly supporting the results of this article are available through Amazon Web Services open datasets program https://registry.opendata.aws/australasian-genomics/. The genome assembly and all raw sequencing reads including the 10x linked-reads and RNAseq reads are available through NCBI under BioProject number PRJNA786364.

Declarations

List of abbreviations

Aquaporin (AQP), basic local alignment search tool (BLAST), base pair (bp), benchmarking universal single copy orthologs (BUSCO), class 1 long interspersed nuclear elements (LINE1), high molecular weight (HMW), kilobase pair (kbb), mammalian-wide interspersed repeats (MIRs), megabase pair (Mbp), National Centre for Biotechnology Information (NCBI), transcripts per million (TPM), type 1 taste receptor (TAS1R), type 2 taste receptor (TAS2R), type 1 vomeronasal receptor (V1R), type 2 vomeronasal receptor (V2R), vomeronasal organ (VNO).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been funded by the Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and Discovery Project (DP180102465), and the Bioplatforms Australia Oz Mammals Genomics consortia. LS and PB are supported by an Australian postgraduate award. TH is supported by a KAKENHI grant from the Japan Society for the Promotion of Science (JSPS) (19K16241, 21H04919, and 21KK0106) and the JSPS Bilateral Joint Research Project "Earth-wide comparative genomics of endangered mammals in Japan and Australia" (JPJSBP 120219902).

Authors' contributions

PB and LS assembled and annotated the genome. EP assembled and annotated the global transcriptome, conducted transcript counts and manually annotated taste and vomeronasal receptors, and aquaporin genes. TH analysed molecular evolution of taste receptors. KB and CJH designed the study. All authors viewed, commented on and agreed to publication of the manuscript.

Acknowledgements

The authors would like to acknowledge Perth Zoo for providing samples.

References

1. Gill T. Arrangement of the families of mammals with analytical tables. The Smithsonian Institution; 1874.
2. Tyndale-Biscoe CH. Life of marsupials. Collingwood: CSIRO Publishing; 2005.
3. Feigin CY, Newton AH, Doronina L, Schmitz J, Hipsley CA, Mitchell KJ, et al. Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. Nature Ecology & Evolution. 2018;2 1:182-92. doi:10.1038/s41559-017-0417-y.
4. Richardson B and Walton D. Fauna of Australia. Australian Government Publishing Service, Canberra; 1989.
5. Christensen P, Maisey K and Perry DH. RADIOTRACKING THE NUMBAT, MYRMECOBIUS-FASCIATUS, IN THE PERUP FOREST OF WESTERN-AUSTRALIA. Aust Wildlife Res. 1984;11 2:275-88.
6. Geiser F and Körtner G. Hibernation and daily torpor in Australian mammals. Australian Zoologist. 2011;35 2:204-15. doi:10.7882/AZ.2010.009.
7. Cooper CE and Withers PC. Patterns of body temperature variation and torpor in the numbat, *Myrmecobius fasciatus* (Marsupialia: Myrmecobiidae). Journal of Thermal Biology. 2004;29:277-84.
8. Dawson TJ and Hulbert AJ. Standard Energy Metabolism of Marsupials. Nature. 1969;221 5178:383-. doi:10.1038/221383a0.
9. Cooper CE and Withers PC. Metabolic physiology of the numbat (*Myrmecobius fasciatus*). Journal of Comparative Physiology B. 2002;172:669-75.

10. Cooper C, Withers P and Bradshaw S. Field metabolic rate and water turnover of the numbat (Myrmecobius fasciatus). Journal of Comparative Physiology B. 2003;173 8:687-93.
11. Friend JA. The numbat *Myrmecobius fasciatus* (Myrmecobiidae): history of decline and potential for recovery. Proceedings of the Ecological Society of Australia. 1990;16:329-77.
12. Friend J. The numbat Myrmecobius fasciatus (Myrmecobiidae): History of decline and potential for recovery. Proceedings of the Ecological Society of Australia. 1990;16:369-77.
13. Woinarski J and Burbidge AA: *Bettongia penicillata* The IUCN Red List of Threatened Species 2016: e. T27858A21961347. https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T2785A21961347.en (2016). Accessed 25th November 2021.
14. Friend J and Thomas N. Conservation of the numbat (Myrmecobius fasciatus). 2003. p. 452-63.
15. Bester AJ and Rusten K. Trial translocation of the numbat (Myrmecobius fasciatus) into arid Australia. Aust Mammal. 2009;31 1:9-16. doi:https://doi.org/10.1071/AM08104.
16. Legge S, Woinarski J, Burbidge AA, Palmer R, Ringma J, Radford JQ, et al. Havens for threatened Australian mammals: the contributions of fenced areas and offshore islands to the protectoin of mammal species susceptible to introduced predators. Wildlife Research. 2018;45 7:627-44.
17. Friend JA and Page MJ. *Numbat (Myrmecobius fasciatus) recovery plan. Wildlife Management Porgram No. 60*. 2017. Department of Parks and Wildlife, Perth, WA.
18. Seppey M, Manni M and Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M, editor. Gene Prediction: Methods and Protocols. New York, NY: Springer New York; 2019. p. 227-45.
19. Andrews S: FastQC. A quality control analysis tool for high throughput sequencing data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
20. Weisenfeld NI, Kumar V, Shah P, Church DM and Jaffe DB. Direct determination of diploid genome sequences. Genome Research. 2017;27 5:757-67. doi:10.1101/gr.214874.116.
21. Bushnell B: BBMap. https://sourceforge.net/projects/bbmap/ (2014).
22. Edwards R: SLiMSuite. https://github.com/slimsuite/SLiMSuite/ (2019).
23. Li H and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2010;26 5:589-95. doi:10.1093/bioinformatics/btp698.
24. Smit A, Hubley R and Green P: RepeatModeler Open-1.0. http://www.repeatmasker.org (2008-2015).
25. Smit A, Hubley R and Green P: RepeatMasker Open-4.0. http://www.repeatmasker.org (2013-2015).
26. Solovyev V, Kosarev P, Seledsov I and Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biology. 2006;7 Suppl 1:S10. doi:10.1186/gb-2006-7-s1-s10.
27. Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30 15:2114-20.
28. Haas BJ, Papanicolaou A, Yassour M, Gradherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols. 2013;8 8:1494-512.
29. Bryant DM, Johnson KM, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Reports. 2017;18:762-76.
30. Wong ESW, Papenfuss AT and Belov K. Immunome database for marsupials and monotremes. BMC Immunology. 2011;12 48:1-6.
31. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC bioinformatics. 2009;10 421.
32. Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie2. Nature Methods. 2012;9 4:357-9.

33. Patro R, Duggal G, Love MI, Irizarry RA and Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14 4:417-9. doi:10.1038/nmeth.4197.
34. Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Research. 2021;49 D1:D394-D403. doi:10.1093/nar/gkaa1106.
35. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, et al. Clustal W and clustal X version 2.0. Bioinformatics. 2007;23:2947-8.
36. Hall T. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series. 1999;41:95-8.
37. Kumar S, Stecher G, Li M, Knyaz C and Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular biology and evolution. 2018;35 6:1547-9. doi:10.1093/molbev/msy096.
38. Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, et al. Adaptation and conservation insights from the koala genome. Nature Genetics. 2018;50 8:1102-11. doi:10.1038/s41588-018-0153-5.
39. Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandriv LB and al e. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell. 2012;148:780-91.
40. Brandies PA, Tang S, Johnson RSP, Hogg CJ and Belov K. The first *Antechinus* reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies. Gigabyte. 2020;2020:0. doi:10.46471/gigabyte.7.
41. 10x Genomics: Molecule length calculation. https://support.10xgenomics.com/de-novo-assembly/software/pipelines/latest/output/moleculelen (2020). Accessed 22nd July 2021.
42. Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrow LB, Bauer MJ, Fu B, et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell. 2012;148:780-91.
43. Renfree MB, Papenfuss AT, Deakin JE, Lindsay J, Heider T, Belov K, et al. Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. Genome Biology. 2011;12:2-25.
44. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature. 2007;447:168-78.
45. Castellano F and Molinier-Frenkel V. An Overview of l-Amino Acid Oxidase Functions from Bacteria to Mammals: Focus on the Immunoregulatory Phenylalanine Oxidase IL4I1. Molecules. 2017;22 12 doi:10.3390/molecules22122151.
46. Toda Y, Hayakawa T, Itoigawa A, Kurihara Y, Nakagita T, Hayashi M, et al. Evolution of the primate glutamate taste sensor from a nucleotide sensor. Current Biology. 2021;31 20:4641-9.e5. doi:https://doi.org/10.1016/j.cub.2021.08.002.
47. Hayakawa T, Suzuki-Hashido N, Matsui A and Go Y. Frequent Expansions of the Bitter Taste Receptor Gene Repertoire during Evolution of Mammals in the Euarchontoglires Clade. Molecular Biology and Evolution. 2014;31 8:2018-31. doi:10.1093/molbev/msu144.
48. Zhao H, Yang J-R, Xu H and Zhang J. Pseudogenization of the Umami Taste Receptor Gene Tas1r1 in the Giant Panda Coincided with its Dietary Switch to Bamboo. Molecular Biology and Evolution. 2010;27 12:2669-73. doi:10.1093/molbev/msq153.
49. Zhou Y, Shearwin-Whyatt L, Li J, Song Z, Hayakawa T, Stevens D, et al. Platypus and echidna genomes reveal mammalian biology and evolution. Nature. 2021; doi:10.1038/s41586-020-03039-0.
50. Bachmanov AA and Beauchamp GK. Taste Receptor Genes. Annual Review of Nutrition. 2007;27 1:389-414. doi:10.1146/annurev.nutr.26.061505.111329.
51. Margolskee RF, Dyer J, Kokrashvili Z, Salmon KSH, Ilegems E, Daly K, et al. T1R3 and gustducin in gut sense sugars to regulate expression of Na⁺-glucose

cotransporter 1. Proceedings of the National Academy of Sciences. 2007;104 38:15075. doi:10.1073/pnas.0706678104.

52. Silva L and Antunes A. Vomeronasal Receptors in Vertebrates and the Evolution of Pheromone Detection. Annual Review of Animal Biosciences. 2017;5 1:353-70. doi:10.1146/annurev-animal-022516-022801.

53. Shi P and Zhang J. Comparative genomic analysis identified an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. Genome Research. 2007;17:166-74.

54. Young JM, Kambere M, Trask BJ and Lane RP. Divergent V1R repertoires in five species: amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs. Genome Research. 2005;15:231-40.

55. Young JM, Massa HF, Hsu L and Trask BJ. Extreme variability among mammalian V1R gene families. Genome Research. 2010;20:10-8.

56. Grus WE, Shi P and Zhang J. Largest Vertebrate Vomeronasal Type 1 Receptor Gene Repertoire in the Semiaquatic Platypus. Molecular Biology and Evolution. 2007;24 10:2153-7. doi:10.1093/molbev/msm157.

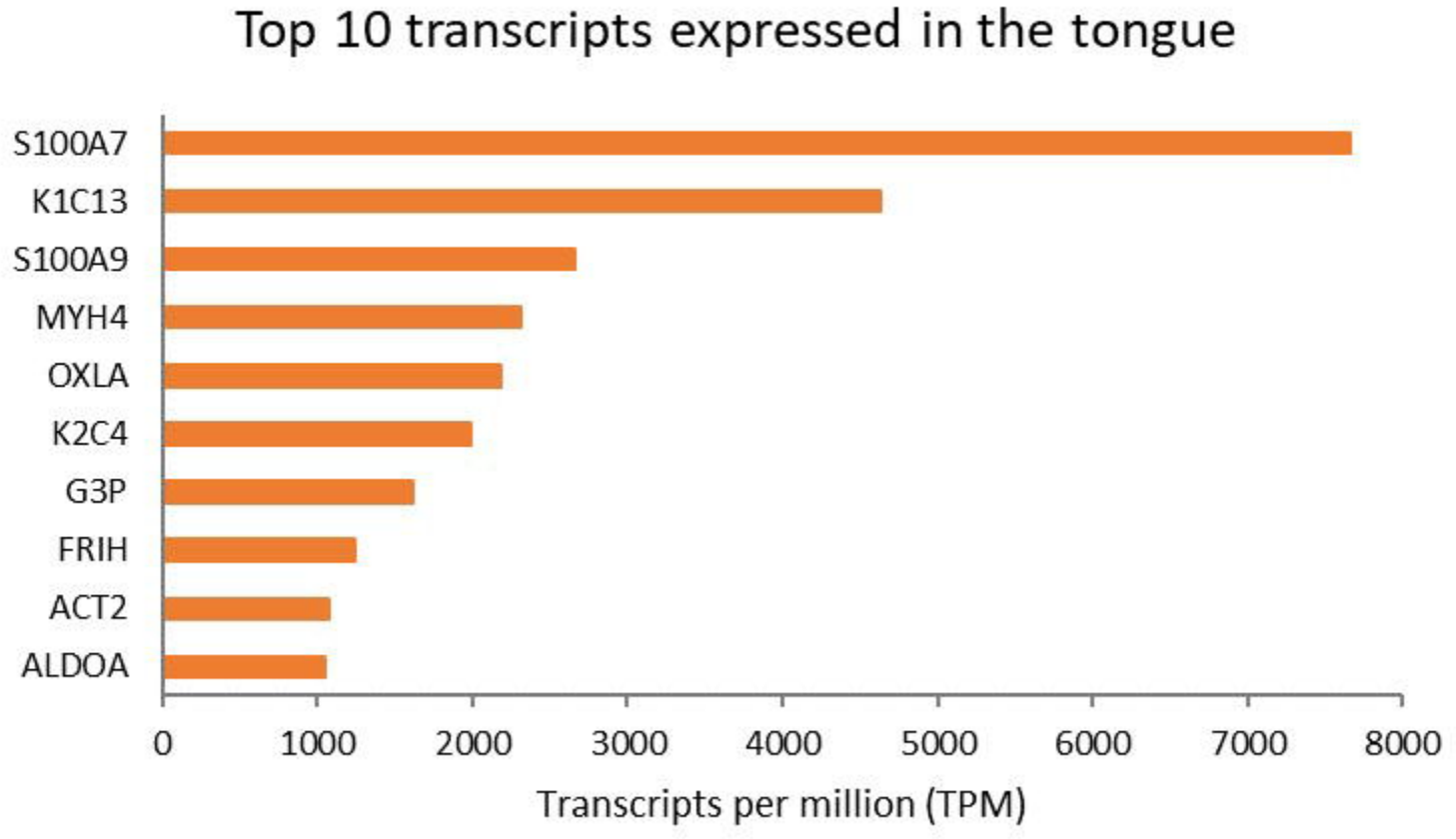
57. Schneider NY. The development of the olfactory organs in newly hatched monotremes and neonate marsupials. Journal of Anatomy. 2011;219 2:229-42. doi:https://doi.org/10.1111/j.1469-7580.2011.01393.x.

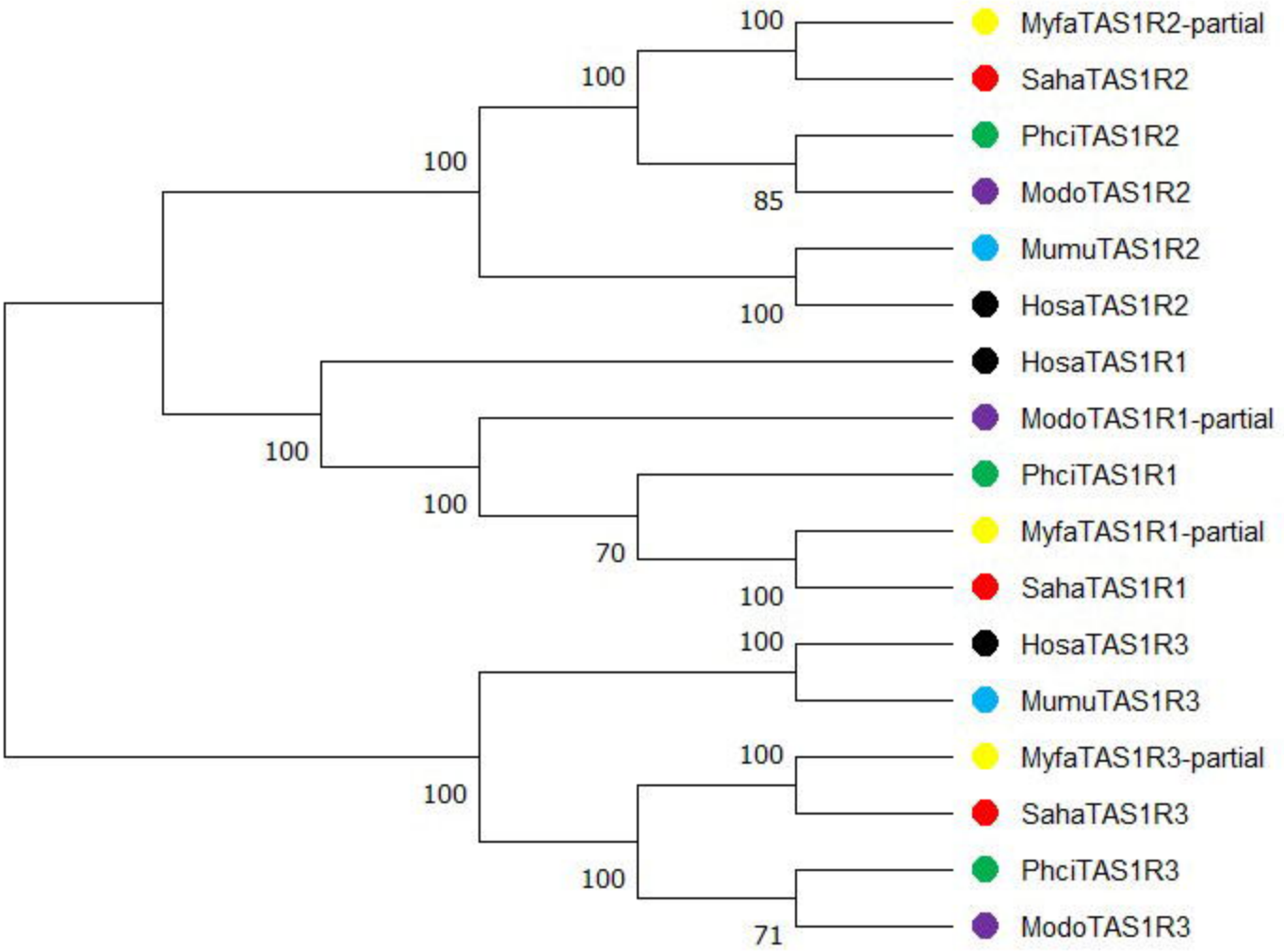
58. Dong D, Jin K, Wu X and Zhong Y. CRDB: Database of Chemosensory Receptor Gene Families in Vertebrate. PLOS ONE. 2012;7 2:e31540. doi:10.1371/journal.pone.0031540.

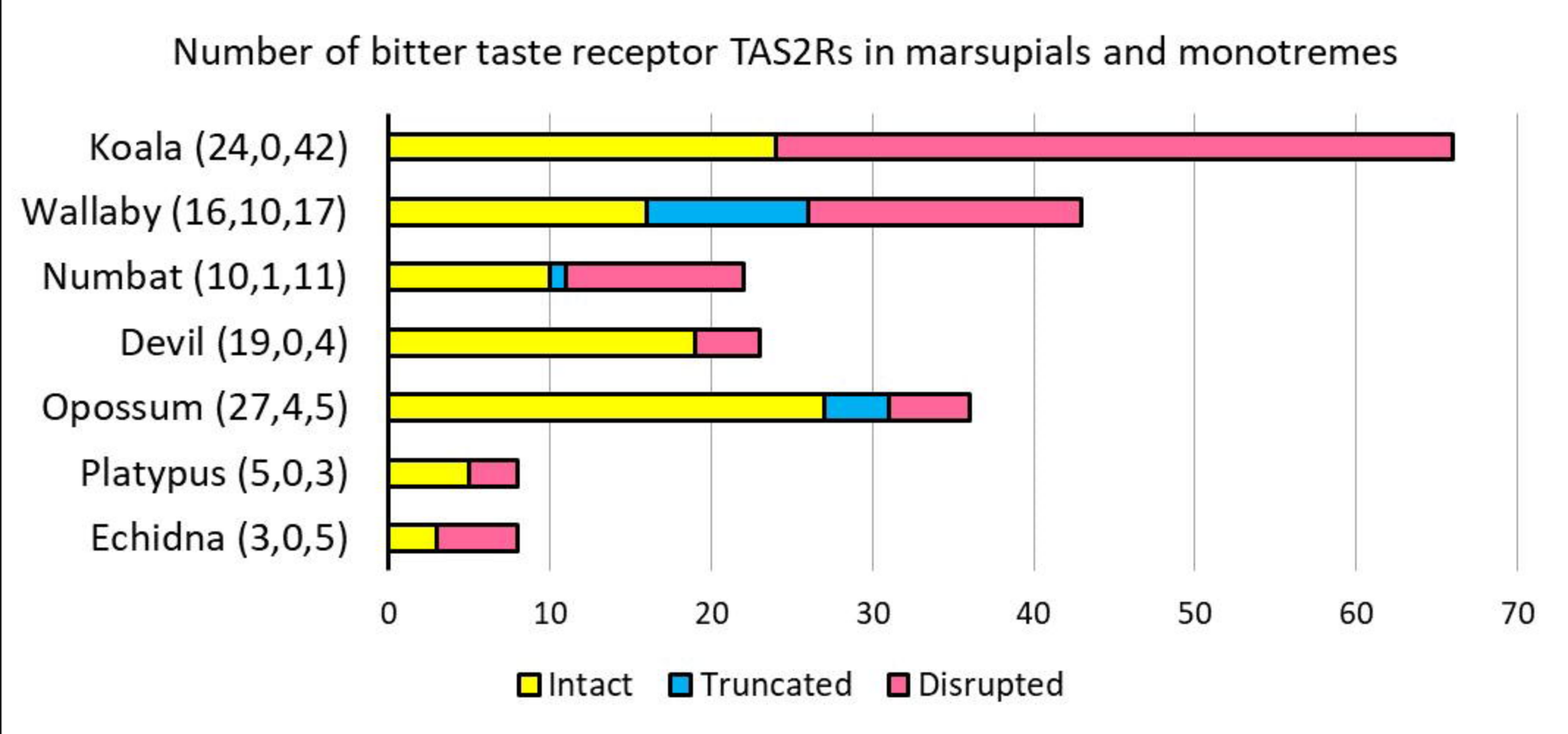
59. Brykczynska U, Tzika AC, Rodriguez I and Milinkovitch MC. Contrasted Evolution of the Vomeronasal Receptor Repertoires in Mammals and Squamate Reptiles. Genome Biology and Evolution. 2013;5 2:389-401. doi:10.1093/gbe/evt013.

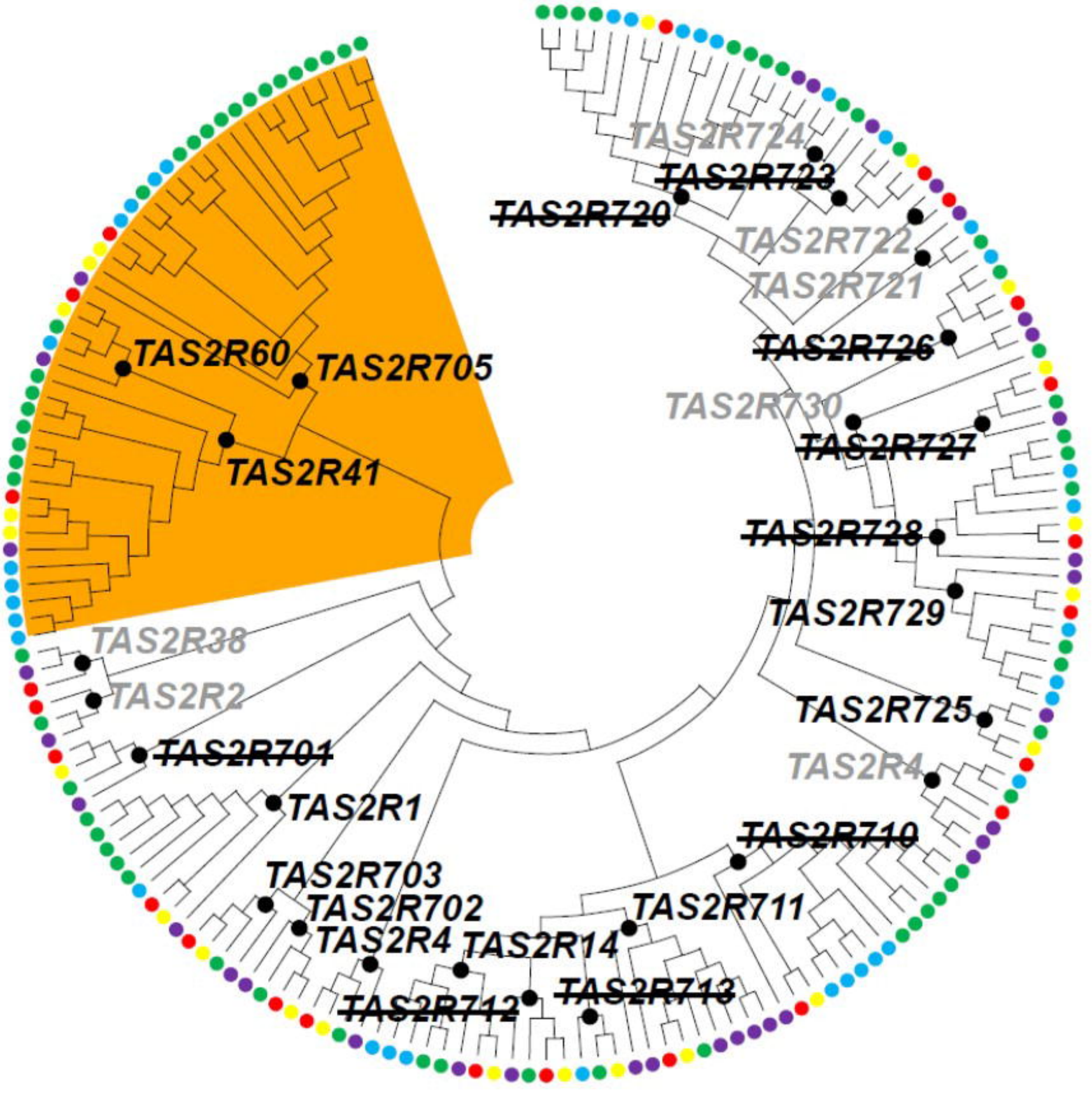
60. Cooper CE and Withers PC. Gross renal morphology of the numbat (Myrmecobius fasciatus) (Marsupialia:Myrmecobiidae). Australian Mammalogy. 2010;32 2:95-7. doi:https://doi.org/10.1071/AM10005.

61. Ishibashi K, Hara S and Kondo S. Aquaporin water channels in mammals. Clinical and Experimental Nephrology. 2009;13 2:107-17. doi:10.1007/s10157-008-0118-6.


Top 10 transcripts expressed in the tongue
S100A7
K1C13
S100A9
MYH4
OXLA
K2C4
G3P
FRIH
ACT2
ALDOA
0
1000
2000
3000
4000
5000
6000
7000
8000
Transcripts per million (TPM)


100
100
100
85
100
100
100
100
70
100
100
100
100
100
100
71
MyfaTAS1R2-partial
SahaTAS1R2
PhciTAS1R2
ModoTAS1R2
MumuTAS1R2
HosaTAS1R2
HosaTAS1R1
ModoTAS1R1-partial
PhciTAS1R1
MyfaTAS1R1-partial
SahaTAS1R1
HosaTAS1R3
MumuTAS1R3
MyfaTAS1R3-partial
SahaTAS1R3
PhciTAS1R3
ModoTAS1R3


Number of bitter taste receptor TAS2Rs in marsupials and monotremes
Koala (24,0,42)
Wallaby (16,10,17)
Numbat (10,1,11)
Devil (19,0,4)
Opossum (27,4,5)
Platypus (5,0,3)
Echidna (3,0,5)
0
10
20
30
40
50
60
70
Intact
Truncated
Disrupted


TAS2R724
TAS2R723
TAS2R720
TAS2R722
TAS2R721
TAS2R726
TAS2R730
TAS2R727
TAS2R728
TAS2R729
TAS2R725
TAS2R4
TAS2R710
TAS2R711
TAS2R713
TAS2R14
TAS2R712
TAS2R4
TAS2R702
TAS2R703
TAS2R1
TAS2R701
TAS2R2
TAS2R38
TAS2R41
TAS2R705
TAS2R60