



ARTICLE

<https://doi.org/10.1038/s41467-019-13531-5>

OPEN

# Genomic and transcriptomic insights into molecular basis of sexually dimorphic nuptial spines in *Leptobrachium leishanense*

Jun Li <sup>1</sup>, Haiyan Yu<sup>2</sup>, Wenxia Wang<sup>1</sup>, Chao Fu<sup>1</sup>, Wei Zhang<sup>1</sup>, Fengming Han<sup>2</sup> & Hua Wu <sup>1\*</sup>

Sexually dimorphic (SD) traits are important in sexual selection and species survival, yet the molecular basis remains elusive, especially in amphibians where SD traits have evolved repeatedly. We focus on the Leishan moustache toad (*Leptobrachium leishanense*), in which males develop nuptial spines on their maxillary skin. Here we report a 3.5 Gb genome assembly with a contig N50 of 1.93 Mb. We find a specific expansion of the intermediate filament gene family including numerous keratin genes. Within these genes, a cluster of duplicated hair keratin genes exhibits male-biased and maxillary skin-specific expression, suggesting a role in developing nuptial spines. We identify a module of coexpressed genes significantly associated with spine formation. In addition, we find several hormones likely to be involved in regulating spine development. This study not only presents a high-quality anuran genome but also provides a reference for studying skin-derived SD traits in amphibians.

<sup>1</sup>Institute of Evolution and Ecology, School of Life Sciences, Central China Normal University, 152 Luoyulu, Hongshan District, Wuhan 430079, China.

<sup>2</sup>Biomarker Technologies Corporation, Beijing 101300, China. \*email: [wuhua@mail.ccnucnu.edu.cn](mailto:wuhua@mail.ccnucnu.edu.cn)

Males and females of a species often exhibit different strategies to maximize fitness. This difference in optimal strategies can select for phenotypic differences between two sexes<sup>1</sup>. Sexually dimorphic (SD) modifications include male-specific morphology, such as the lion's mane<sup>2</sup>, the sword-like tail of swordtail fishes<sup>3</sup>, and the greatly enlarged tooth of male narwhals (*Monodon monoceros*)<sup>4</sup>. As most of the genome is shared between sexes with the exception of genes in sex chromosomes, sexual dimorphism is primarily caused by differential, sex-biased gene expression<sup>1</sup>. To achieve sex-biased expression, gene duplication at the genomic level is a potential solution. Redundant paralogs generated by duplication events can allow one or more genes to diverge in expression levels and/or protein structure<sup>5</sup>. A wealth of evidence has proven that duplicated genes often acquire male-biased and tissue-specific expression<sup>6</sup>. Recent studies exploring genome-wide sex-biased expression patterns in different species revealed a broad variation in the percentage of sex-biased genes (ranging from 2% of transcripts in *Littorina saxatilis* to 90% in *Drosophila melanogaster*)<sup>7</sup>. Although these studies provide insights into sex-biased expression patterns, very few of them link sex-biased genes to particular traits. Furthermore, most of conclusions regarding the molecular regulation of SD traits come from studies on model species<sup>8</sup>. The main obstacle linking interesting traits to sex-biased genes outside of model organisms is that we often lack genomic information for nonmodel species, including genomic resource and knowledge of genetic networks governing trait development.

Amphibians are a group of vertebrates with abundant SD traits. In anurans (toads and frogs), >90% of the species exhibit larger body sizes in females than in males<sup>9</sup>. Another type of SD trait in anurans is skin-derived excrescences, such as the nuptial pads on the digits of hands or on the ventral surfaces of forelimbs<sup>10</sup> and the nuptial spines on the upper jaw<sup>11</sup>. These traits are present mainly in adult males during the breeding period and exhibit a seasonal cycle. A morphological survey on nuptial pads from 26 species in Phyllomedusinae (Hylidae) found that the pads consisted of dark epidermal projections (EPs). The shape and density of EPs and the separation level between adjacent EPs differ among species<sup>12</sup>. Hormone implant experiments revealed that the pads could be induced by androgens in adult males, adult females, and even tadpoles<sup>13</sup>. However, in contrast to the results of morphological examination and hormone manipulation, the genetic basis of shaping nuptial excrescences remains elusive, which can be largely attributed to the lack of genomic information. For instance, a recent study on *Leptobranchium boringii* using nonreference transcriptomic analyses found several processes (such as cytosolic processes and peptidase inhibitor activity) and a list of potential genes (such as insulin-like growth factor genes and sex steroid hormone-related genes) that may be associated with the seasonal development of nuptial spines<sup>14</sup>. However, owing to the absence of a reference genome, the annotation rate of unigenes was very low (30.98%), indicating that most of the unigenes were unannotated<sup>14</sup>. This limitation will hinder our understanding of important biological processes and genes associated with SD traits.

The genomes of amphibians are exceptionally large (up to 120 Gb in salamanders) and feature high levels of repetitive sequences<sup>15</sup>, which makes both sequencing and assembly challenging. To date, six anuran genomes have been sequenced and annotated<sup>16–22</sup>, among which only the genomes of *Xenopus laevis* and *Xenopus tropicalis* have been assembled to chromosome level<sup>17,19</sup>. Among the sequenced anurans, *X. laevis* and *X. tropicalis* belong to the Archaeobatrachia, while the other four species (*Nanorana parkeri*, *Rana catesbeiana*, *Rhinella marina*, and *Oophaga pumilio*) belong to Neobatrachia<sup>23</sup>. However, to the best of our knowledge, no genomic data for the spadefoot toads

(Pelobatoidea), a monophyletic clade sister to Neobatrachia relative to the paraphyletic Archaeobatrachia<sup>23</sup>, has been published. Within the Pelobatoidea, Megophryidae that is the most widely diversified family provides an appropriate system for studying skin-derived SD traits in amphibians. For example, in the Leishan moustache toad (*Leptobranchium leishanense*), males have skin excrescences that are similar to skin-derived SD traits in other anurans in terms of tissue structure and developmental cycles. During the breeding period, adult males develop four sharp and conical black spines on the maxillary skin (MS) (two on each side), whereas such a structure is absent in females. During the post-breeding period, adult males lose the conical outer casing<sup>24</sup>. Immature males have only red spots in the same area. It has been proposed that such nuptial spines may be used for male–male combat, stimulation of females, or nest construction/maintenance during the breeding season<sup>25</sup>.

Herein, to build a representative genome of Pelobatoidea and to understand the molecular basis of skin-derived SD traits in amphibians, we focus on the Leishan moustache toad. We first assemble a 3.5 Gb of chromosome-scale genome. Comparative genomic analysis reveals expansion of intermediate filament (IF) gene families (GFs) in *L. leishanense*, which include numerous keratin genes (*krts*). To identify biological processes and genes associated with the production of nuptial spines, we compare the transcriptomes of multiple tissues, including dorsal skin (DS), MS, brain, and gonad, between males and females at three developmental stages. Stage A is the subadult stage, when males have not developed spines and females have no eggs; stage B is the breeding stage, when males have developed spines and females have eggs; stage C is the postbreeding period, when males' spines fall off and females have laid eggs. We identify a module of coexpressed genes to be significantly associated with spine formation. In addition, hormones such as androgen, thyroid hormone (TH), prolactin (PRL), and relaxin (RLN) are likely involved in the regulation of spine development. In summary, we obtain a high-quality reference genome and reveal a series of candidate genes underlying the production and regulation of nuptial spines in *L. leishanense*. Similar regulatory patterns can be used to guide future studies on skin-derived SD traits in amphibians.

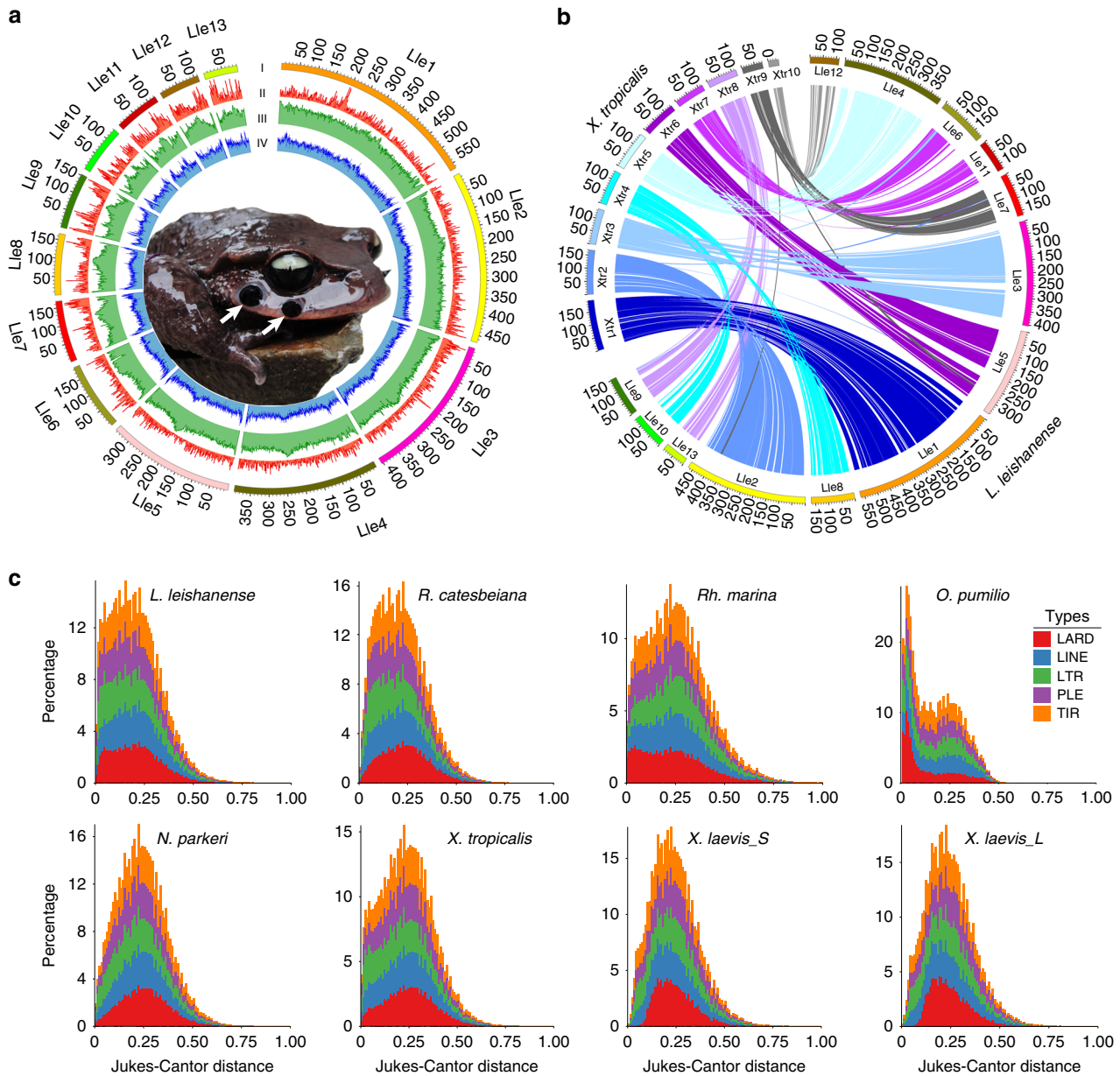
## Results

**Genome assembly and characterization.** A male *L. leishanense* toad was selected for genome sequencing and assembly. The genome size was estimated to be 3.56 Gb based on the *k*-mer distribution (Supplementary Fig. 1). A total of 285.81 Gb (~80×) of PacBio long reads were assembled using Canu v1.5<sup>26</sup> combined with WTDBG v1.1.006 (<https://github.com/ruanjue/wtdbg>). This assembly was polished based on 174.86 Gb (~50×) of Illumina paired-end reads using Pilon v1.22<sup>27</sup>. Subsequently, we used LACHESIS (<http://shendurelab.github.io/LACHESIS/>) to assemble contigs on 13 pseudochromosomes<sup>28</sup> based on 155.16 Gb (~44×) of Hi-C data. A total of 3296 contigs with a length of 3.31 Gb were assembled to chromosome-level scaffolds (Supplementary Fig. 2). We finally generated a 3.54 Gb of *L. leishanense* genome, with a contig N50 of 1.93 Mb and scaffold N50 of 394.69 Mb, providing the first chromosome-anchored genome among Pelobatoidea species (Table 1, Fig. 1a, Supplementary Tables 1 and 2).

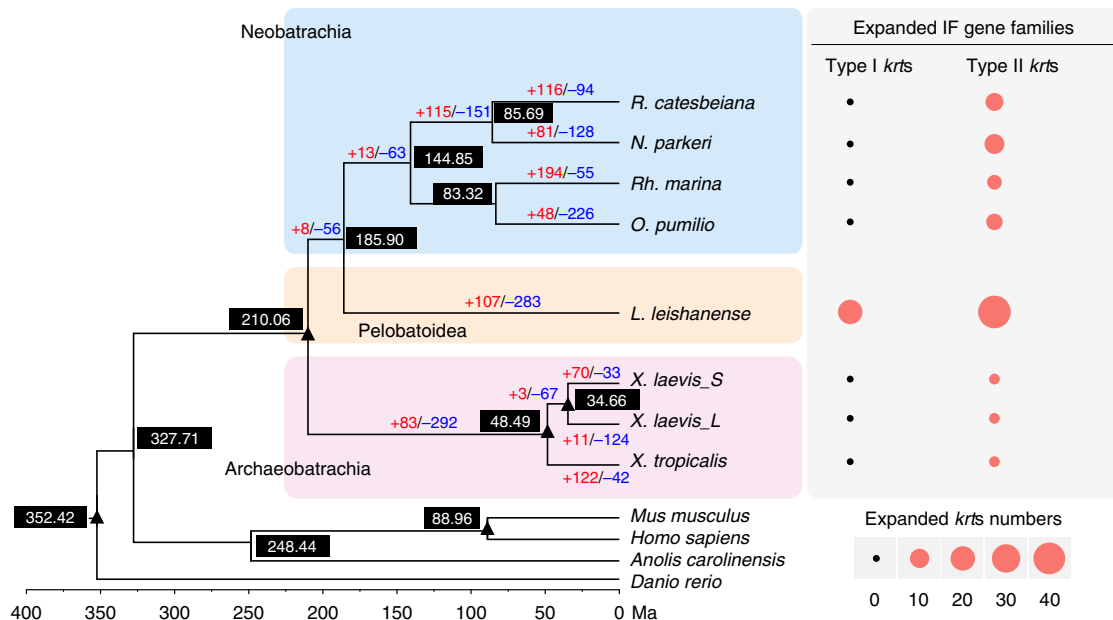
To evaluate the assembly quality, we mapped the Illumina reads on the reference genome. Approximately 95% of the paired-end reads were mapped properly. Then we aligned 258,442 transcriptomic unigenes from nine *L. leishanense* tissues to the assembly, with >99.6% of unigenes being aligned, indicating excellent coverage of the expressed genes (Supplementary

**Table 1 Statistics of assembled genomes among different anurans.**

	<i>Leptobrachium leishanense</i>	<i>Nanorana parkeri</i>	<i>Rana catesbeiana</i>	<i>Rhinella marina</i>	<i>Oophaga pumilio</i>	<i>Xenopus tropicalis</i>	<i>Xenopus laevis</i>
Assembled genome size (Gb)	3.5	2.1	5.8	2.6	5.5	1.5	2.7
Contig N50 (bp)	1,931,129	32,920	5239	167,489	3942	71,041	19,713
Scaffold N50 (bp)	394,693,044	1,075,367	68,964	167,489	59,503	135,134,832	136,570,856
GC content (%)	43.7	42.5	43.0	43.2	43.5	40.0	40.0
Repetitive sequences (%)	77.1	56.7	66.9	86.6	62.9	51.9	38.3
Protein coding genes	23,420	21,477	22,000	58,302	17,051	27,047	37,385



**Fig. 1** Genomic landscape of *Leptobrachium leishanense* and comparison with other anurans. **a** From outer to inner: (I) sizes of 13 pseudo-chromosomes; (II) gene density (the percentage of genes in per 200-kb window); (III) repeat sequence distribution; (IV) GC content (%). The photo in the circle shows an adult male with nuptial spines on the upper jaw as marked with two white arrows (two spines on each side). The toad photo was taken by Wei Zhang. **b** Chromosome synteny between *L. leishanense* and *X. tropicalis*. Chromosomes of *X. tropicalis* are marked as “Xtr1-10” and chromosomes of *L. leishanense* are marked as “Lle1-13”. Collinear blocks between two species are linked by lines with the same colors. **c** Divergence distribution of five dominant transposable elements (TEs) in different genomes based on the same analysis procedures (see “Methods”). The x-axis is the Jukes-Cantor distance from the consensus in the TE library. The y-axis shows percentages of TE occupancy in genomes. LARD large retrotransposon derivatives, PLE Penelope-like elements, TIR terminal inverted repeat.



**Fig. 2** Comparative genomic analyses across anurans. Values in black boxes are divergence times in units of million years ago (Ma, see “Methods”). The number near each branch indicates the number of significantly expanded (red) and contracted (blue) gene families for each clade. Nodes marked with black triangles are used as time calibration nodes. Two intermediate filament (IF) gene families were detected under significant expansion, which include type I and type II keratin genes, respectively. The numbers of expanded keratin genes in different anurans compared to their most recent common ancestor are displayed in circle sizes.

Table 3). Furthermore, we examined the completeness of the conserved core eukaryotic genes (CEGs) and universal single-copy orthologs using CEGMA v2.5<sup>29</sup> and BUSCO v3<sup>30</sup> (tetrapoda\_odp9 database), respectively. The reference genome of *L. leishanense* includes 241 of the 248 (97.2%) complete CEGs and 3840 of the 3950 (97.2%) complete and partial BUSCO genes, indicating high completeness of the assembled genome. These metrics are higher than those of sequenced anuran genomes as evaluated by the same method (Supplementary Figs. 3 and 4 and Supplementary Tables 4 and 5).

**Genome annotation and chromosome synteny.** We annotated the repetitive sequences based on the de novo repeat sequence database of *L. leishanense* combined with Repbase 20.01<sup>31</sup>. We found that 77.1% (2.73 Gb) of the *L. leishanense* genome was repetitive sequences, which is higher than the values for most anurans with sequenced genomes (Table 1). Similar to other species, the most abundant transposable elements (TEs) are terminal inverted repeats (Supplementary Table 6). Notably, two major retrotransposons, long interspersed nuclear element and long terminal repeat (LTR), constitute a higher proportion of the *L. leishanense* genome (24.8% and 17.5%, respectively) than other anurans, suggesting the high accumulation of these TEs in *L. leishanense*. Such accumulation of retrotransposons provides the potential for genome size expansion via duplicative transposition. Moreover, the TE expansion pattern of *L. leishanense* is similar to that of *R. catesbeiana* and *R. marina*, which exhibits one major expansion wave followed by a slight burst that occurred recently (Fig. 1c). After masking repetitive sequences, we annotated a total of 23,420 genes in the reference genome with an average gene length of 30 kb (Supplementary Fig. 5), and 96.4% of these genes were annotated in public databases.

To compare structural characteristics of the genomes between *L. leishanense* and *X. tropicalis*, we analyzed chromosomal synteny based on genome-scale ortholog alignment. We found extensive chromosome synteny between two species, with a vast majority of orthologs located on one chromosome in *X. tropicalis*

also being located on a single chromosome in *L. leishanense* (Fig. 1b). In *L. leishanense*, a total of 346 blocks with an average size of 7.68 Mb were collinear with the *X. tropicalis* blocks. The chromosome number in *L. leishanense* is 13, whereas the number in *X. tropicalis* is 10. This difference is reflected in the corresponding chromosomal fissions in *L. leishanense*. Specifically, the blocks on the *X. tropicalis* chromosomes (Xtr4, 7, and 8) are distributed on two separate chromosomes in the *L. leishanense* genome (Lle8+10, 6+11, and 9+13, respectively). In addition, chromosomal rearrangements such as inversion and translocation were detected. In the *L. leishanense* genome, a total of 179 out of 346 blocks were inverted (Supplementary Data 1). Seven regions in *L. leishanense* with sizes of 502 kb–5.93 Mb corresponding to Xtr2, 3, 6, 9, and 10 of *X. tropicalis* were translocated into other non-homologous chromosomes (Fig. 1b, Supplementary Figs. 6 and 7).

**Expanded GFs in *L. leishanense*.** We compared the *L. leishanense* genome with six other anurans (*R. catesbeiana*, *N. parkeri*, *R. marina*, *O. pumilio*, *X. tropicalis*, and *X. laevis*) and four vertebrates (*Mus musculus*, *Homo sapiens*, *Anolis carolinensis*, and *Danio rerio*) to analyze anuran divergence. A phylogenetic tree based on single-copy orthologs revealed the middle position of *L. leishanense* between Neobatrachia and Archaeobatrachia (Fig. 2). The most recent common ancestor (MRCA) among the seven anurans was estimated to occur at 210.06 million years ago (Ma) (176.22–248.83 Ma). *L. leishanense* shared an MRCA with Neobatrachia species at 185.90 Ma (154.10–221.31 Ma, Fig. 2).

It has been proposed that changes in gene copy number would support adaptive evolution<sup>32</sup>. We thus estimated the expanded GFs in *L. leishanense* to explore targets of adaptive evolution. A total of 107 GFs were significantly expanded in *L. leishanense* compared with its MRCA (Fig. 2). Genes from expanded families were mainly enriched in signal response categories, including sensory perception of smell, response to stimulus, and signal transduction (Supplementary Data 2). These categories were also significantly expanded in other anurans such as *N. parkeri*,

*R. catesbeiana*, and *R. marina* (Supplementary Data 2). The PI3K-Akt signaling pathway, which plays important roles in responding to extracellular stimuli and regulating multiple cellular functions, such as cell proliferation, apoptosis, and survival<sup>33</sup>, was significantly overrepresented in *L. leishanense* (Fisher's exact test, Benjamini-Hochberg (BH) corrected  $p < 0.001$ ). In addition, biological processes associated with immune responses, including the immune system process, antigen processing and presentation, and immune response, were significantly and specifically expanded in *L. leishanense* (Supplementary Data 2). Further analyses and empirical data are needed to explore the relationship between habitat adaptation and the expansion of immune response-related genes in *L. leishanense*.

**Expanded keratin genes support nuptial spine formation.** We found two GFs encoding IF proteins were significantly expanded in *L. leishanense* (Fig. 2). These GFs include genes encoding keratins that are filament-forming proteins of epithelial cells and are involved in shaping keratinized tissues in vertebrates, such as human hairs and nails<sup>34</sup>. Considering the hardness of nuptial spines, we hypothesize that the expansion of the IF GFs may be associated with the occurrence of nuptial spines in *L. leishanense*. We therefore conducted a comprehensive analysis of *krts* in *L. leishanense* and compared to other vertebrates.

In tetrapod, an important component of integument rigidity is formed by two classes of  $\alpha$ -keratins (termed type I and type II  $\alpha$ -keratins)<sup>34</sup>. Genomic analyses have demonstrated that humans possess a total of 54 functional *krts* (28 type I and 26 type II), separately clustered on two chromosomes<sup>35</sup>. In *L. leishanense*, we identified a total of 101 complete  $\alpha$ -*krts*, including 53 type I and 48 type II genes (Supplementary Data 3), which is much higher than the number observed in other species<sup>34</sup> (Supplementary Table 7). By comparing the genomic distribution of  $\alpha$ -*krts*, we found that type I and II  $\alpha$ -*krts* were separately clustered on two chromosomes (Lle 12 and Lle 2) in *L. leishanense* and were flanked by the same genes as those in other vertebrates, suggesting conserved arrangement order of *krts*. There are several clusters of lineage-specific paralogous *krts* in mammals (human and mouse), *L. leishanense*, and other anurans (Fig. 3a; Supplementary Fig. 8). The most diversified and duplicated *krts* within *L. leishanense* are homologous to mammalian hair keratins (HKs). Notably, the copy number of HK genes in *L. leishanense* is almost double than in other anurans (Fig. 3a).

To further explore the correlation of duplicated HK genes with nuptial spines in *L. leishanense*, we sequenced the expressed RNAs and compared the expression patterns of HK and non-HK genes in multiple samples (Supplementary Fig. 9). We found that the duplicated HK genes (types I and II) in *L. leishanense* showed male-biased and MS-specific expression during the breeding stage (B), whereas non-HK genes expressed similarly in both the MS and DS in two sexes (Fig. 3b, Supplementary Data 4). These results support that the expansion of the IF GFs, especially the highly duplicated HK genes, is associated with the occurrence of nuptial spines in *L. leishanense*.

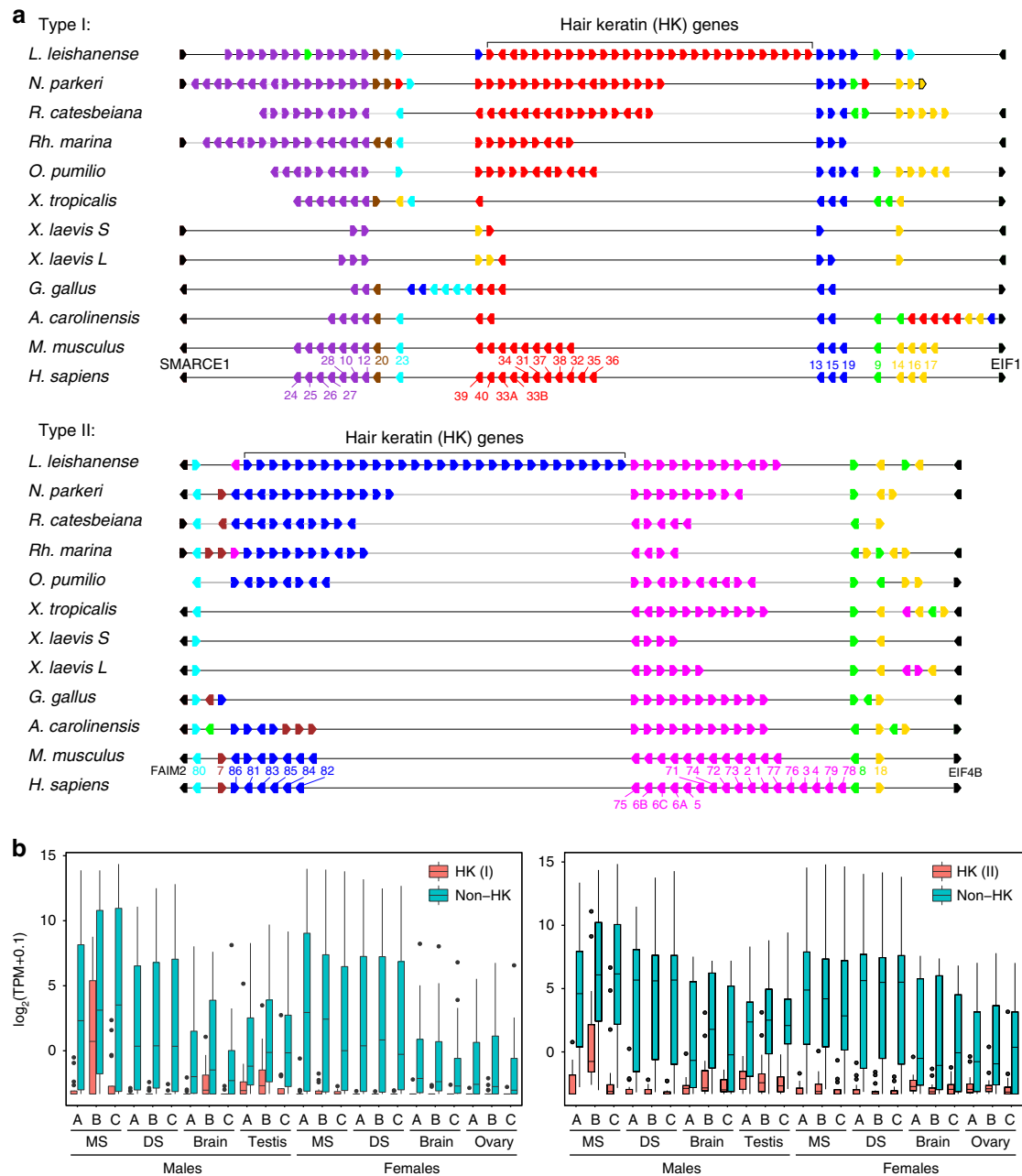
**Biological processes and pathways in producing spines.** To investigate biological processes associated with nuptial spines, we sequenced transcriptomes from the MS in two sexes and three developmental stages (Table 2). We also sequenced transcriptomes from the DS as controls for skin. For each stage, we performed three groups of independent comparisons: (1) comparison of MS between males and females; (2) comparison of MS against DS in males; and (3) comparison of MS against DS in females. We identified differentially expressed genes (DEGs) for

each comparison and conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichments.

Stage A is the subadult period, when males and females show similar red spots on the MS. In adult males, these spot regions develop black spines during breeding seasons. Thus DEGs in males' MS at stage A would be associated with early preparation for spine development. Differential expression analyses revealed that only 19 out of the 15,135 genes were expressed differentially between males' and females' MS ( $|\log_2(\text{fold change})| > 1$  and BH corrected  $p < 0.01$ ) and no GO term was enriched (Fig. 4, Supplementary Table 8), which is consistent with the similar phenotypes between two sexes. DEGs between the MS and DS in males were mainly involved in transcriptional regulation processes (including positive and negative regulation of transcription from the RNA polymerase II promoter) and skeletal development-related processes (including embryonic skeletal system morphogenesis and cartilage development). Similar processes were not enriched in the comparison of females (Fig. 4). Stage B is the adult breeding period, when males have developed spines and females retain red spots in the MS. Thus DEGs in males' MS at stage B would be directly associated with the formation of nuptial spine. We found that 83 out of 14,409 genes were differentially expressed in the MS between two sexes. These DEGs were enriched in hormone activities (such as TH transport and steroid biosynthetic process) and epithelial cell differentiation (Fig. 4). Similar processes were also enriched for DEGs in males' MS vs DS but were absent in the same comparison for females (Fig. 4). At the postbreeding stage (C) when males' spines fall off, we found that the DEGs of males' MS vs DS were significantly enriched in the proteolytic process that contains multiple genes encoding trypsin and cathepsin (Fig. 4). Proteolysis can break peptide bonds and hydrolyze proteins to small polypeptides (<https://www.ebi.ac.uk/QuickGO/term/GO:0006508>). It was reported that trypsin could degrade human epidermal keratins in vitro<sup>36</sup>. We thus consider that the proteolytic process may be involved in triggering the falling off of the spines, probably by degrading the keratin protein complex.

Findings from amphibian endocrine research have revealed that male-biased structures such as enlarged flexor muscles and nuptial thumb pads are androgen dependent<sup>13</sup>. In amphibians, as in other vertebrates, androgens (testosterone and 5 $\alpha$ -dihydrotestosterone) are mainly produced by testes and regulated by gonadotropins released from the pituitary gland<sup>37</sup>. Therefore, to explore the hormonal regulation of nuptial spines in *L. leishanense*, we sequenced the transcriptomes from brain and gonads (Table 2). DEGs from brains were significantly enriched in the neuroactive ligand-receptor interaction pathway (BM3 vs BF3 and AM3 vs BM3, Supplementary Fig. 10), which includes numerous signaling molecules, such as hormones and the associated receptors. Within this pathway, five genes encoding hormone subunits were specifically highly expressed in males' brains at stage B (BM3 in Fig. 5a). Among these genes, the *lh* (luteinizing hormone) and *fsh* (follicle-stimulating hormone) encode peptide hormones that stimulate the synthesis of steroid hormones by traveling via the blood to the gonads<sup>38</sup>. High expression of these genes indicates the activation of the hypothalamus-pituitary-gonad axis in males during the breeding season. Consistently, the steroid hormone biosynthesis pathway was significantly enriched in males' testes between different stages (AM4 vs BM4 and BM4 vs CM4, Supplementary Figs. 10 and 11). Within this pathway, genes encoding enzymes involved in sex steroid biosynthesis were highly expressed in male testes at stage B (Fig. 5b), suggesting a high level of androgen biosynthesis.

In addition to gonads, the skin is a major nonclassic tissue for steroidogenesis via the expression of functionally active enzymes<sup>39</sup>. In *L. leishanense*, DEGs between the male and female



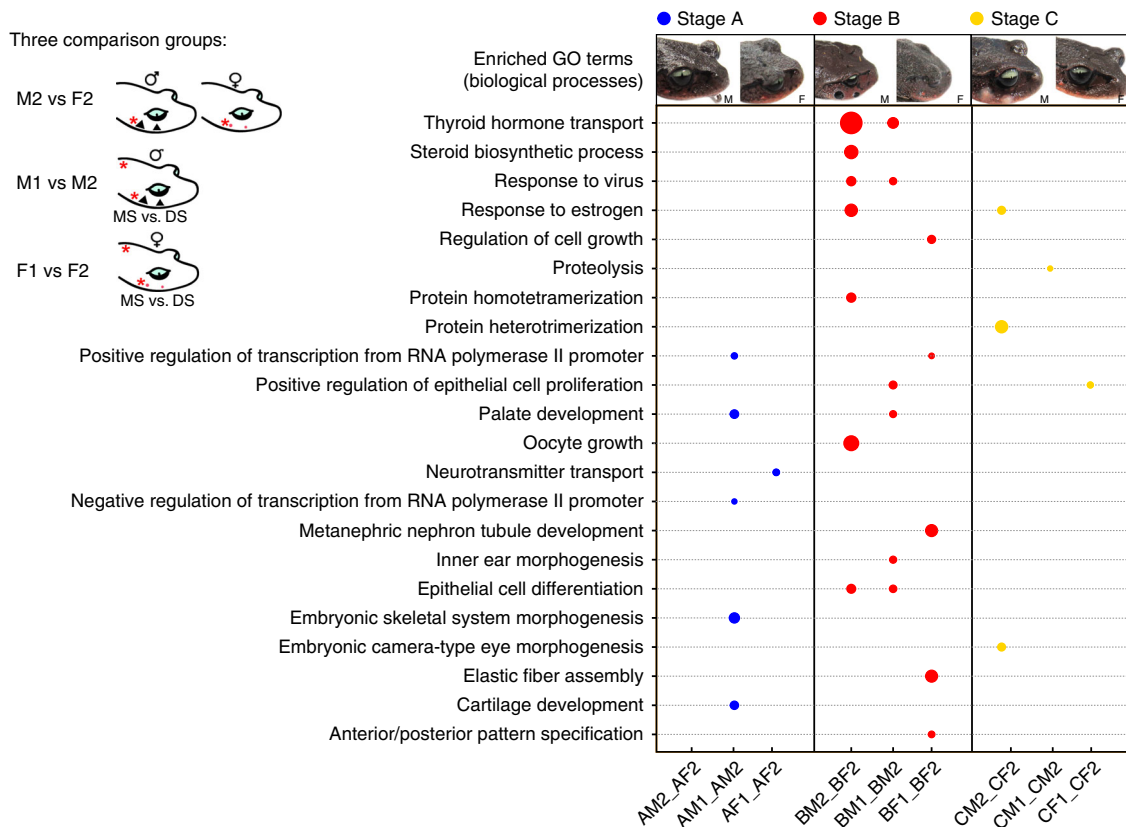
**Fig. 3** Genomic distribution and expression pattern of the keratin gene family. **a** Genomic organization of type I and type II  $\alpha$ -keratin genes in different species. Polygon arrows indicate the direction of keratin genes in the genome. Gene colors are marked based on homologous relationships with human keratin genes. Protein names in human are listed. **b** Box plots of the expression levels for HK genes and non-HK genes in different samples. The expression level was evaluated by  $\log_2(\text{TPM} + 0.1)$ . For each boxplot of type I HK,  $n = 17$  genes' mean expression values based on three biological replicates; for each boxplot of type I non-HK,  $n = 22$  genes; for each boxplot of type II HK,  $n = 17$  genes; for each boxplot of type II non-HK,  $n = 15$  genes (Supplementary Data 4). The middle line of a box indicates the median (50 percentile) of the expression values for HK (red boxes) and non-HK (blue boxes) genes. The upper and lower terminal lines of a box represent the 25th and 75th percentiles, respectively. Lines extending vertically from the boxes (whiskers) indicate variability outside the upper and lower percentiles.

**Table 2** Sampling codes for transcriptomic analyses.

Stages	Dorsal skin	Maxillary skin	Brain	Gonad
A	AM1, AF1	AM2, AF2	AM3, AF3	AM4, AF4
B	BM1, BF1	BM2, BF2	BM3, BF3	BM4, BF4
C	CM1, CF1	CM2, CF2	CM3, CF3	CM4, CF4

M male, F female

MS were also significantly enriched in the steroid biosynthesis process, which includes genes encoding 3-beta-hydroxysteroid dehydrogenase (*hsd3b*), 17 $\alpha$ -hydroxylase (*cyp17a1*), and steroidogenic acute regulatory protein (*star*)<sup>40</sup>. High expression of these enzymes indicates the accumulation of C19 steroids of androgens in males' MS. Although the level of steroidogenesis in nonclassic tissues is quite modest compared with that in gonads, this process could be very important in locally autocrine or

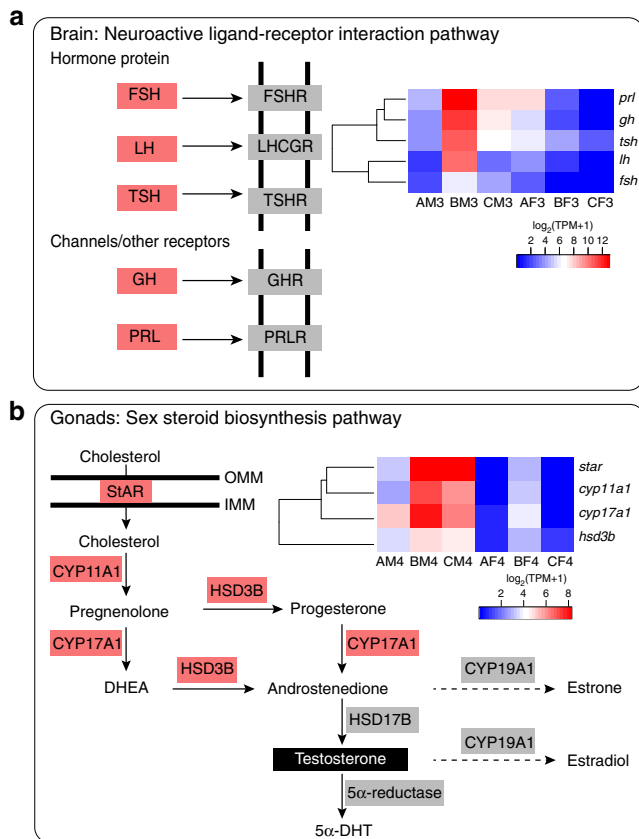


**Fig. 4** Biological processes associated with producing nuptial spines in *Leptobrachium leishanense*. The legend in the left corner shows three groups of comparisons conducted in each stage: M2 vs F2 indicates the comparison of maxillary skin between males and females (AM2\_AF2, BM2\_BF2, and CM2\_CF2 in the right dot plot panel); M1 vs M2 indicates the comparison of male dorsal skin and maxillary skin (AM1\_AM2, BM1\_BM2, CM1\_CM2); F1 vs F2 indicates the comparison of female dorsal skin and maxillary skin (AF1\_AF2, BF1\_BF2, CF1\_CF2). Red stars on the cartoon toads represent tissues used for comparison in each group. Jun Li made the hand-drawn cartoon toads in Adobe Illustrator CC 2018 with the help of Graphics Tablet (Huion, Model: H420, made in Shenzhen, China). The head figures on the top show the phenotypes of male (M) and female (F) *L. leishanense* toads at three stages. These toad photos were taken by Jun Li. The colored circles in the dot plot represent three stages, and the circle sizes represent enrichment factors with large size indicating a high degree of enrichment.

paracrine regulation<sup>39</sup>. However, androgens alone are insufficient to support male SD traits, and other hormones play critical roles as well<sup>13</sup>. Here we found that TH, PRL, and RLN may be involved in the regulatory process. The “thyroid hormone transport” term was highly enriched in DEGs in BM2 vs BF2 and BM1 vs BM2 (Fig. 4). Furthermore, genes encoding PRL and RLN were more highly expressed in the male MS than in the female MS based on both RNA-seq and quantitative PCR (qPCR) data (false discovery rate (FDR)<sub>pri</sub> < 0.001, log<sub>2</sub>(FC)<sub>pri</sub> = -8.41; FDR<sub>rln</sub> < 0.001, log<sub>2</sub>(FC)<sub>rln</sub> = -10.59, Supplementary Figs. 12 and 13).

**Gene coexpression analyses reveal the genetic networks underlying the spine production.** To further identify the key genes controlling the production of nuptial spines, we analyzed the gene coexpression networks and identified genetic modules using weighted gene correlation network analysis (WGCNA)<sup>41</sup>. We sought to identify modules (groups of highly correlated genes that may exhibit the same biological activity) that were associated with specific traits. Based on the correlation of expression profiles among genes, we identified 12 modules in total, with the 13th module including genes that were not assigned to any module (Fig. 6a; Supplementary Fig. 14; Supplementary Data 5). Then we characterized the gene expression pattern of each module based on the module eigengene (ME), which represents the first principal component of the scaled module expression profiles. MEs of

each module were correlated to external sample traits, with significantly correlated modules suggesting specifically high expression patterns for a particular trait. We found that ME03, representing 170 genes, was significantly correlated with MS at stage A in both males and females ( $r_{\text{male}} = 0.53$ ,  $p < 0.001$ ;  $r_{\text{female}} = 0.32$ ,  $p = 0.007$ , Fig. 6a), suggesting that genes in this module are mainly involved in MS specialization prior to spine development. Within this module, positive regulation of the canonical Wnt signaling pathway was significantly enriched (corrected  $p = 0.012$ , Supplementary Data 6). In addition, the Wnt signaling pathway (ko04310) was enriched from KEGG analysis ( $p = 0.02$ , corrected  $p = 0.27$ , Supplementary Data 6). In the canonical Wnt signaling pathway, genes encoding the Wnt protein, two leucine-rich repeat-containing G-protein coupled receptors (LGRs), and Wnt inhibitory factor 1 (Wif1) were included in ME03 (Fig. 6b). The coexpression pattern between the Wnt protein and LGR receptors suggests the activation of Wnt signaling in the MS prior to the formation of nuptial spines. In addition, we analyzed the positively selected genes in *L. leishanense* compared with six other anurans using the branch-site model of CODEML in PAML v4.9<sup>42</sup>. The gene encoding adenomatous polyposis coli, which facilitates  $\beta$ -catenin degradation<sup>43</sup>, had experienced significantly positive selection in *L. leishanense* ( $p = 0.004$ ; Supplementary Data 7). Signaling by the Wnt proteins is one of the fundamental mechanisms that directs cell proliferation during embryogenesis<sup>43</sup>. Moreover, the



**Fig. 5** Hormone-related pathways enriched in DEGs of the brain and gonads. **a** An illustration of the neuroactive ligand-receptor interaction pathway. The red boxes are ortholog genes that are highly expressed in male brains at stage B. The heatmap shows the expression levels of five genes across six brain samples. *fsf* follicle-stimulating hormone, *lh* luteinizing hormone, *tsh* thyroid-stimulating hormone, *gh* growth hormone. **b** An illustration of the sex steroid biosynthesis pathway revised from the map00140 in the KEGG database. The red boxes are enzymes that are highly expressed in male testes at stage B. The heatmap shows the expression levels of four genes across six gonad samples. *star* steroidogenic acute regulatory protein, *cyp11a1* cytochrome P450 side-chain cleavage enzyme (P450<sub>sc</sub>), *cyp17a1* cytochrome P450 17 $\alpha$ -hydroxylase/17,20-lyase (P450<sub>c17</sub>), *hsd3b* 3 $\beta$ -hydroxysteroid dehydrogenase, DHEA dehydroepiandrosterone, 5 $\alpha$ -DHT 5 $\alpha$ -dihydrotestosterone, OMM outer mitochondrial membrane, IMM inner mitochondrial membrane.

Wnt signaling pathway has been reported to facilitate the formation of skin appendages such as teeth, hair follicles, and deer antlers by mediating epithelial–mesenchymal interactions and determining cell fate<sup>44,45</sup>. Therefore, we propose that activation of the Wnt signaling pathway is involved in triggering the production of nuptial spines in *L. leishanense*. This result reveals the co-option of the pre-existing Wnt signaling pathway for generating novel traits.

We also found that module ME05 showed a significant correlation with the male MS at stage B ( $r = 0.86$ ,  $p < 0.001$ , Fig. 6a). This module contains 43 genes, 16 of which are *krt* genes (Fig. 6c). Within this module, the coexpression network reveals that *krt* genes showed strong correlation with other genes, supporting the core roles of *krt* genes in the formation of SD spines. This module also contains three *tgf2* genes (encoding protein-glutamine gamma-glutamyltransferase 2) that are associated with cross-linking in the keratin assembly process<sup>46</sup>. In addition, the *hoxc13* gene showed a distinct correlation with other

genes, although the edge weight was not as high as that of the *krt* genes (Fig. 6c). It has been shown that *hoxc13* is involved in the regulation of human HK expression<sup>47</sup>. The gene encoding tyrosinase (*tyr*), which is the essential enzyme in melanin synthesis<sup>48</sup>, was also included in this module, which suggests the role of this gene in the maintenance of the black color of the spines. Thus this module provides the most potential candidate genes that are directly associated with the occurrence of nuptial spines in male *L. leishanense* toads.

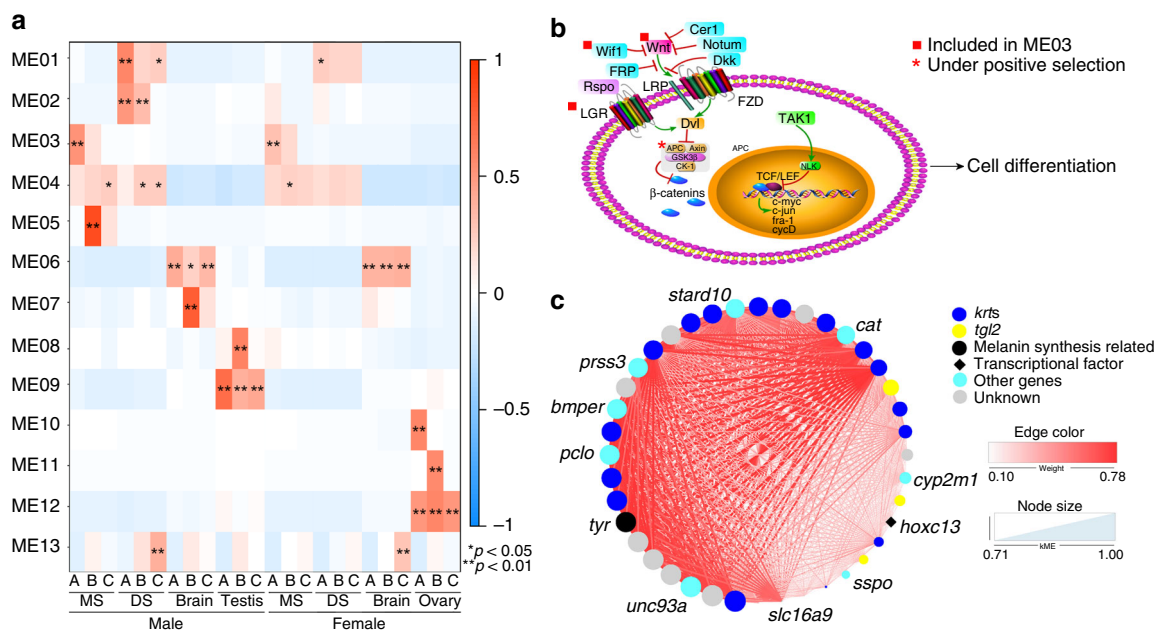
## Discussion

In this study, we generated a Pelobatoidea reference genome by combining PacBio long-read sequencing and chromatin interaction scaffolding methods. We detected two recent burst events of TEs in *L. leishanense* genome, which is similar to the genomes of *R. catesbeiana* (5.8 Gb), *R. marina* (2.6 Gb), and *O. pumilio* (5.5 Gb) from Neobatrachia. In contrast, *Xenopus* species (sizes in 1.5–2.7 Gb) from Archaeobatrachia only presented one peak. However, the number of protein-coding genes is similar across these species, indicating that different expansion and accumulation of TEs is one of the major factors contributing to the variation in genome size in amphibians. Diverse genome sizes in amphibians may be driven by adaptive and/or nonadaptive forces. On the one hand, the genome size is strongly associated with developmental rates in Anura. Natural selection will reduce the genome size in environments that foster rapid embryonic and larval development<sup>49</sup>. On the other hand, the evolution of genome size can be affected by neutral factors, such as mutation rates and effective population size<sup>49,50</sup>. For example, the substitution rates of both mitochondrial and nuclear datasets are positively correlated with genome size in Anura<sup>49</sup>. Exploring the relative roles of adaptive and neutral factors in driving changes in genome size would be interesting when an increasing number of amphibian genomes have been sequenced.

Although vertebrate genomes differ greatly in DNA content and chromosome karyotype, extensive synteny exists across species<sup>51</sup>. Here we detected a highly conserved genomic arrangement pattern between *L. leishanense* and *X. tropicalis*, despite the different numbers of chromosomes and the long-term divergence (~210.1 Ma) between two species. Similar genomic synteny analysis has been conducted between *N. parkeri* and *X. tropicalis*<sup>18</sup>. Because of the relatively fragmented assembly of the *N. parkeri* genome (scaffold N50 of 1.1 Mb), most of the collinear blocks that belong to one chromosome in *X. tropicalis* were located on separate scaffolds in *N. parkeri*. For example, the orthologs of chromosome Xtr1 were mapped on scaffolds13 and 18 in *N. parkeri*. Thus fragmented assembly will underestimate the number of collinear blocks identified between two genomes. Despite of the fragmented assembly, whole-genome alignments between *N. parkeri* and *X. tropicalis* still revealed a large amount of synteny<sup>18</sup>. Nevertheless, collinear blocks may vary in orientation or position between species. For example, we detected inversions and translocations in *L. leishanense* compared with the chromosomes of *X. tropicalis*, suggesting the presence of structural variation in amphibian genomes.

In this study, we focused on examining the molecular basis underlying the skin-derived nuptial spines in *L. leishanense* by combining genomic and transcriptomic analyses. A previous study has been conducted on *L. boringii* by analyzing the non-reference transcriptomes of males' MS, brain, and testes between the breeding (B) and postbreeding (C) stages<sup>14</sup>. Although a total of 1181 DEGs were identified between stages B and C in males' MS, no nuptial spine-specific GO term was enriched, which may be largely due to the lack of a reference genome. Here we designed a comprehensive sampling strategy by collecting more





**Fig. 6** Gene modules correlated with forming nuptial spines in *Leptobrachium leishanense*. **a** Correlation between coexpressed modules and sampling traits. Heatmap colors represent correlation coefficients (\*Fisher’s asymptotic  $p < 0.05$ ; \*\* $p < 0.01$ ). Rows are different modules, and columns are different sampling traits. **b** An illustration of the canonical Wnt pathway. Genes marked with red squares are included in the ME03 module; the gene marked with a red star is under significant positive selection in *L. leishanense* compared to other anurans. **c** Weighted coexpression network for genes in the ME05 module. Dark-blue circles are keratin genes; yellow circles are genes encoding protein-glutamine gamma-glutamyltransferase 2 (*tgf2*); the black circle is *tyr* gene that is annotated as a melanin synthesis-related gene; the black diamond is a transcriptional factor (*hoxc13*); light-blue circles are other annotated genes involved in diverse metabolic processes; gray circles are unknown genes. Colored lines represent the weight of connectivity between genes, with white to red indicating weights from low (0.10) to high (0.78). The node sizes are associated with the kME values ranging from 0.71 to 1.00. *stard10* StAR-related lipid transfer (START) domain containing 10, *cat* catalase, *tgf2* protein-glutamine gamma-glutamyltransferase 2, *cyp2m1* cytochrome P450 2M1, *hoxc13* homeobox protein Hox-C13, *sspo* SCO-spondin (precursor), *slc16a9* monocarboxylate transporter 9, *unc93a* Protein unc-93 homolog, *tyr* tyrosinase, *pcl* protein piccolo-like, *bmp* BMP-binding endothelial regulator, *prss3* trypsin-3.

tissues (added DS as a control) at three developmental stages (added subadult stage A) from two sexes (added the same samples from females). Moreover, the high-quality *L. leishanense* genome provided a solid reference for identifying DEGs and for GO enrichments. Transcriptomic analyses revealed meaningful biological processes associated with shaping nuptial spines, such as epithelial cell differentiation, steroid biosynthesis, and TH transport, which have not been identified previously. In addition, genomic-level examination allowed us to analyze the species-specific duplication of *krts* and the homologous relationship with HK genes in mammals. Based on the expression pattern of *krts*, we conclude that duplicated HK genes are closely associated with shaping nuptial spines. The study on *L. boringii* also obtained one *krt* (comp28513\_c0) expressed differentially; however, it is difficult to extensively annotate *krts* based solely on transcriptomic data. Therefore, the reference genome of *L. leishanense* contributes greatly to the exploration of the molecular bases underlying nuptial spine formation.

Notably, the duplication of HK genes also occurs in the other four Neobatrachia anurans (*N. parkeri*, *R. catesbeiana*, *R. marina*, and *O. pumilio*). Most of these terrestrial anurans present skin-derived SD structures, such as tiny nuptial pads on the male chest (*N. parkeri*)<sup>52</sup> and fingers (*R. catesbeiana* and *R. marina*)<sup>53,54</sup>. The skin of male *O. pumilio* exhibits specialized dorsal color<sup>55</sup>. In contrast, the aquatic *Xenopus* species (*X. laevis* and *X. tropicalis*) only have one type I HK gene and do not have any type II HK gene (Fig. 3a). Such differences in the number and duplication of HK genes between terrestrial and aquatic anurans suggest that HK genes may experience species-specific divergence and support the formation of diverse structures in the adaptation to terrestrial

habitats in amphibians. Therefore, extensive studies on keratin GFs across amphibians are required.

However, there remain several unanswered questions. First, sex chromosomes that differ between males and females are important to understand the molecular mechanisms underlying sexual dimorphism. Insights obtained from model species, particularly fruit flies, proved that the central mechanism underlying the generation of SD traits is the integration of sex-determination system and the genetic networks governing trait development<sup>8</sup>. Sex chromosomes are usually associated with sex determination. Thus identifying sex chromosomes is crucial for screening genes that control the development of sexual dimorphism. Unfortunately, based on available data, we cannot identify sex chromosomes due to the lack of morphologically distinguishable sex chromosomes in *L. leishanense*<sup>28</sup>. In fact, the sex chromosome system varies considerably in amphibians, and the heterogametic sex (either XX/XY or ZZ/ZW) can vary even between populations of a single species<sup>56</sup>. Pool sequencing based on confidently sexed males and females from multiple populations and assembly to the reference genome may be helpful for identifying sex-specific molecular markers and localizing sex chromosomes. However, the success of pool sequencing depends on the size of detectable differences between sex chromosomes. Very small genetic differences may be difficult to accurately screen.

In addition to the differences in gene content on sex chromosomes, the secretion of sex steroids (androgens and estrogens) differs between males and females in vertebrates, which is another factor affecting sexual dimorphism. Here we found two pathways involved in the hypothalamus–pituitary–gonad regulation axis were significantly enriched in the brain and testes, suggesting a

high level of androgen biosynthesis in male testes at stage B. In addition, three genes involved in testosterone synthesis were present at significantly high levels in the male MS during spine development, suggesting local accumulation of androgens in MS. These findings reveal the essential role of androgens in the formation of nuptial spines. However, it is difficult to directly identify the regulatory processes of androgens in triggering the expression of nuptial spine-related genes.

We also observed the involvement of other hormones, such as PRL and RLN, supported by sex-related DEGs in the MS. PRL, which was named for its role in the promotion of lactation in mammals, has been found to serve multiple roles in reproduction<sup>57</sup>. Interestingly, PRL was reported to induce the formation of nuptial pads in combination with TH in male red-spotted newts<sup>58</sup>. Thus we propose that PRL plays an essential role in regulating the production of nuptial spine in *L. leishanense*. RLN is another small peptide hormone that was broadly implicated in regulating the female reproductive process and improving spermatogenesis in vertebrates<sup>59,60</sup>. However, its role in the regulation of SD trait development has rarely been reported<sup>14</sup>. We found that the gene encoding RLN is highly and specifically expressed in the male MS, suggesting the crucial role in nuptial spine production in *L. leishanense*. Further studies on hormone regulation are needed to examine the roles of different hormones in regulating nuptial spine development in *L. leishanense*.

## Methods

**Samples used for different sequencing methods.** The Leishan moustache toad samples used in this study were collected from the Leishan county, Guizhou Province, China. For genomic sequencing, we collected muscle from one adult male. Muscle and other eight tissues from the same individual (including DS, MS, brain, testis, liver, heart, kidney, and spleen) were collected for transcriptomic sequencing. For the Hi-C library preparation, we collected 4 ml of heart blood from four adult males to achieve enough sample volume. For comparative transcriptomic sequencing, RNA was extracted from four types of tissues (MS, DS, brain, and gonads) in two sexes at three developmental stages (Supplementary Fig. 9). Each sample includes three biological replicates. For the qPCR validation, we used MS and DS from adult males and adult females during the breeding season. For detailed information on sampling strategies, please see Supplementary Table 9. All experiments involving animals in this study were approved by the Animal Ethics Committee of the School of Life Sciences, Central China Normal University (CCNU-IACUC-2019-008). We have complied with all relevant ethical regulations for animal testing and research.

**Illumina sequencing and genome survey.** Muscle tissue from an adult male was flash frozen in liquid nitrogen. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA). Eight paired-end libraries and 12 mate-pair libraries were prepared following Illumina protocols. Library sequencing was performed on the Illumina HiSeq 4000 system (Illumina, San Diego, CA, USA). After removal of sequencing adapters, contaminant reads (mitochondrial, bacterial, and viral sequences), and low-quality reads, we finally obtained 572.38 Gb (~163× coverage) of clean reads (Supplementary Table 10). We estimated the genome size by *k*-mer distribution (*k* = 21, Supplementary Fig. 1).

**Sequencing of PacBio long reads and Hi-C libraries.** Genomic DNA was sheared by a g-TUBE device (Covaris, Woburn, MA, USA) setting with 20 kb. The sheared DNA was purified and concentrated with AmpureXP beads (Agencourt, Beverly, MA, USA) and further used for Single Molecular Real Time (SMRTbell) library preparation according to the PacBio 20-kb template preparation protocol. The isolated SMRTbell fractions were purified and used for primer and polymerase (P6) binding according to the manufacturer's binding calculator (Pacific Biosciences, Menlo Park, CA, USA). Single-molecule sequencing was conducted on a PacBio RS-II platform with C4 chemistry. After adapter removal, we totally obtained 285.81 Gb of subreads (80.3× coverage), with an average length of 10.03 kb.

Hi-C libraries were created from the whole blood cells (WBCs) from adult males. According to the protocol, nuclear DNA from WBCs was cross-linked and enzymatically digested with Hind III, leaving pairs of distally located but physically interacting DNA molecules attached to each other. The sticky ends of the digested fragments were biotinylated and ligated to each other to form chimeric circles. Biotinylated circles, which are chimeras of physically associated DNA molecules from the original cross-linking, were enriched, sheared, and sequenced with the Illumina HiSeq 2500 platform. A total of 517.9 million clean Hi-C reads pairs (155.16 Gb, 44× coverage) were obtained.

**Genome assembling processes.** We first used Illumina reads (163×) and PacBio long reads (10×) to obtain the initial assembly (this version is referred to as Illumina-and-PacBio Assembly, IPA, Supplementary Table 1). The paired-end reads were assembled to contigs using Platanus<sup>61</sup>. The resulting contigs and PacBio reads were assembled to hybrid contigs with DBG2LOC<sup>62</sup>. Then we used SSPACE<sup>63</sup> to assemble contigs to scaffolds based on the mate-pair reads, which resulted in the IPA version of the reference genome with a contig N50 of 408.18 kb and scaffold N50 of 754.32 kb (Supplementary Table 1). Considering the poor quality of the IPA version and the high proportion of repeat sequences estimated from the genome survey, we chose to use pure PacBio reads to assemble the genome of *L. leishanense*.

We used the 285.81 Gb of PacBio reads to assemble the second version of genome. Prior to assembly, we used the error correction module in Canu v1.5 to correct long subreads. Sequencing errors were corrected with a corrected error rate of 0.025. The corrected subreads were used for genome assembly using WTDBG v1.1.006. The draft genome was polished using Pilon v1.22 with 174.86 Gb of paired-end reads, which produced a *L. leishanense* reference genome with a contig N50 of 2.29 Mb (this version is referred to as PacBio-and-Illumina Polishing, version PIP, Supplementary Table 1).

Then we used the Hi-C data to correct the PIP version and assembled contigs to chromosome-level scaffolds. We used the Hi-C interaction signals to correct assembling errors. Contigs within the assembled genome were broken into fragments with a length of 50 kb. These fragments were used for correcting errors by reassembling based on Hi-C interaction signals. We regarded those positions that cannot be restored to the original positions on the assembled genome as candidate error regions. Within these regions, the positions with low Hi-C coverage were identified as error points. Based on the Hi-C correction, the contig N50 of the PIP version decreased slightly from 2.29 to 1.93 Mb. Then we assembled 3296 (out of 8601) contigs to 13 pseudochromosomes. We finally generated a 3.54 Gb of the *L. leishanense* reference genome, with a contig N50 of 1.93 Mb and scaffold N50 of 394.69 Mb (this version is referred to as PacBio-Illumina polishing and Hi-C assembly, version PIH, Supplementary Table 1), which was used for subsequent analyses.

**Evaluation of genome assembly.** To evaluate the quality of the *L. leishanense* reference genome, we aligned the Illumina reads onto the assembly. We also aligned 258,442 unigenes (≥100 bp) from 9 tissues on the assembly to assess the completeness of gene regions. In addition, we identified conserved eukaryotic core genes and single-copy orthologs in tetrapods based on the CEGMA v2.5 and BUSCO v3 databases, respectively. Based on the standard evaluation procedures, we searched 208 (out of 248, 83.9%) CEGs and 3142 (out of 3950, 79.5%) BUSCO genes in the assembly. To further check whether the missing genes exist in the assembled genome, we aligned the predicted genes in our genome against the hmmer profiles of missing genes in the CEGMA and BUSCO databases, respectively. For the CEGMA evaluation, we defined the genes with >70% of coverage as the complete CEGs based on the standard procedure. We finally found 33 complete CEGs presented in our assembly, which were missed in the standard evaluation procedure. To further explore why these CEGs cannot be detected based on the standard procedure, we examined the features of missing genes and found that these genes usually have longer introns (Supplementary Fig. 3). For the BUSCO evaluation, we also searched the missing genes based on the hmmer model. We finally identified 421 previously missed BUSCO genes in *L. leishanense* genome, which resulted in 97.22% of complete and fragmented BUSCOs. Similar with the missing CEGs, the missing genes that omitted by original BUSCO evaluation procedure are featured by long introns compared with the complete BUSCOs. Similar phenomenon also exists in other anuran genomes (Supplementary Fig. 4).

**Genome annotation and chromosome synteny analysis.** We constructed the de novo repetitive sequence database of *L. leishanense* and combined with the Repbase database 20.01 to create the final repeat library (for detailed software list, please see Supplementary Table 6). Repeat sequences in the *L. leishanense* genome were identified and classified using RepeatMasker 4.0.6<sup>64</sup>. The LTR family classification criterion was defined by the 5'LTR sequences of the same family that shared at least 80% identity over at least 80% of their lengths. To compare repeat sequences across different anurans, we applied the same methods for six other genomes (*N. parkeri*, *R. catesbeiana*, *X. tropicalis*, *X. laevis*, *R. marina*, and *O. pumilio*).

We integrated three approaches, namely, de novo prediction, homology search, and transcript-based assembly, to annotate protein-coding genes in a repeat-masked genome (for detailed software list, please see Supplementary Fig. 5). Consensus gene models were generated by integrating the de novo prediction and protein and transcript alignments using EVIDENCEModeler v1.1.1<sup>65</sup>. To assign gene functions, the predicted gene sequences were searched against seven databases: NR, GO, KEGG, KOG, Pfam, SwissProt, and TrEMBL.

We analyzed the chromosome synteny between *L. leishanense* and *X. tropicalis* via all-to-all BLASTP searches of protein sequences (with an *E*-value cut-off of 1e−5). Collinear blocks containing at least 10 genes (−s 10) and a maximum of 15 gaps (genes) between two proximal orthologs within a block (−m 25) were identified using MCScanX<sup>66</sup>. The serial numbers of the chromosomes were manually adjusted to reflect the descending order of chromosome length (Lle1 is the longest chromosome of *L. leishanense*, and Lle13 is the shortest chromosome).



8. Williams, T. M. & Carroll, S. B. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nat. Rev. Genet.* **10**, 797–804 (2009).
9. Shine, R. Sexual selection and sexual dimorphism in the Amphibia. *Copeia* **1979**, 297 (1979).
10. Kurabuchi, S. Fine structures on the surface of nuptial pads of male hylid and rhacophorid frogs. *J. Morphol.* **219**, 173–182 (1994).
11. Zheng, Y., Li, S. & Fu, J. A phylogenetic analysis of the frog genera *Vibrissaphora* and *Leptobranchium*, and the correlated evolution of nuptial spine and reversed sexual size dimorphism. *Mol. Phylogenet. Evol.* **46**, 695–707 (2008).
12. Luna, M. C., Taboada, C., Baëta, D. & Faivovich, J. Structural diversity of nuptial pads in Phyllomedusinae (Amphibia: Anura: Hylidae). *J. Morphol.* **273**, 712–724 (2012).
13. Sever, D. M. & Staub, N. L. Hormones, sex accessory structures, and secondary sexual characteristics in amphibians. in *Hormones and Reproduction of Vertebrates* (eds Norris, D. O. & Lopez, K. H.) Ch. 5 (Elsevier, 2011).
14. Zhang, W. et al. Transcriptome analysis reveals the genetic basis underlying the seasonal development of keratinized nuptial spines in *Leptobranchium boringii*. *BMC Genomics* **17**, 978 (2016).
15. Sun, C. et al. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **4**, 168–183 (2012).
16. Hammond, S. A. et al. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nat. Commun.* **8**, 1433 (2017).
17. Session, A. M. et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **538**, 336–343 (2016).
18. Sun, Y. et al. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc. Natl Acad. Sci. USA* **112**, E1257–E1262 (2015).
19. Hellsten, U. et al. The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636 (2010).
20. Nowoshilow, S. et al. The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018).
21. Edwards, R. J. et al. Draft genome assembly of the invasive cane toad, *Rhinella marina*. *Gigascience* <https://doi.org/10.1093/gigascience/giy095> (2018).
22. Rogers, R. L. et al. Genomic takeover by transposable elements in the strawberry poison frog. *Mol. Biol. Evol.* **35**, 2913–2927 (2018).
23. Zhang, P. et al. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. *Proc. Natl Acad. Sci. USA* **114**, E5864–E5870 (2017).
24. Qi, Y. et al. Significant male biased sexual size dimorphism in *Leptobranchium leishanensis*. *Asian Herpetol. Res.* **6**, 298–304 (2015).
25. Hudson, C. M., He, X. & Fu, J. Keratinized nuptial spines are used for male combat in the Emei moustache toad (*Leptobranchium boringii*). *Asian Herpetol. Res.* **2**, 142–148 (2011).
26. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
27. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
28. Zhao, E., Wu, G. & Yang, M. A comparative study of the karyotypes of the genus *Vibrissaphora*. *Acta Herpetol. Sin.* **2**, 15–20 (1983).
29. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
30. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
31. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
32. Li, J. et al. Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proc. Natl Acad. Sci. USA* **115**, 8406–8411 (2018).
33. Song, G., Ouyang, G. & Bao, S. The activation of Akt/PKB signaling pathway and cell survival. *J. Cell. Mol. Med.* **9**, 59–71 (2005).
34. Vandeborgh, W. & Bossuyt, F. Radiation and functional diversification of alpha keratins during early vertebrate evolution. *Mol. Biol. Evol.* **29**, 995–1004 (2012).
35. Schweizer, J. et al. New consensus nomenclature for mammalian keratins. *J. Cell Biol.* **174**, 169–174 (2006).
36. Bjelland, S., Hjelmeland, K. & Volden, G. Degradation of human epidermal keratin by cod trypsin and extracts of fish intestines. *Arch. Dermatol. Res.* **280**, 469–473 (1989).
37. Canosa, L. F., Pozzi, A. G., Rosembli, C. & Ceballos, N. R. Steroid production in toads. *J. Steroid Biochem. Mol. Biol.* **85**, 227–233 (2003).
38. Moore, F. L. Reproductive endocrinology of amphibians. in *Fundamentals of Comparative Vertebrate Endocrinology* (eds Chester-Jones, I., Ingleton, P.M. & Phillips J.G.) 207–221 (Springer, Boston, MA, 1987).
39. Slominski, A. et al. Steroidogenesis in the skin: implications for local immune functions. *J. Steroid Biochem. Mol. Biol.* **137**, 107–123 (2013).
40. Schiffer, L., Arlt, W. & Storbeck, K. H. Intracrine androgen biosynthesis, metabolism and action revisited. *Mol. Cell. Endocrinol.* **465**, 4–26 (2018).
41. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).
42. Yang, Z. & Bielawski, J. PAML4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
43. MacDonald, B. T., Tamai, K. & He, X. Wnt/β-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17**, 9–26 (2009).
44. Ahn, Y. Signaling in tooth, hair, and mammary placodes. *Curr. Top. Dev. Biol.* **111**, 421–459 (2015).
45. Mount, J. G. et al. Evidence that the canonical Wnt signalling pathway regulates deer antler regeneration. *Dev. Dyn.* **235**, 1390–1399 (2006).
46. Buxman, M. M. & Wuepper, K. D. Keratin cross linking and epidermal transglutaminase: a review with observations on the histochemical and immunochemical localization of the enzyme. *J. Invest. Dermatol.* **65**, 107–112 (1975).
47. Jave-Suarez, L. F., Winter, H., Langbein, L., Rogers, M. A. & Schweizer, J. HOXC13 is involved in the regulation of human hair keratin gene expression. *J. Biol. Chem.* **277**, 3718–3726 (2002).
48. Slominski, A. Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiol. Rev.* **84**, 1155–1228 (2004).
49. Liedtke, H. C., Gower, D. J., Wilkinson, M. & Gomez-Mestre, I. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat. Ecol. Evol.* **2**, 1792–1799 (2018).
50. Conery, J. S. & Lynch, M. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
51. Voss, S. R. et al. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res.* **21**, 1306–1312 (2011).
52. Ma, X. & Lu, X. Annual cycle of reproductive organs in a Tibetan frog, *Nanorana parkeri*. *Anim. Biol.* **60**, 259–271 (2010).
53. Beaty, L. E., Emmering, Q. C. & Bernal, X. E. Mixed sex effects on the second-to-fourth digit ratio of Túngara Frogs (*Engystomops pustulosus*) and Cane Toads (*Rhinella marina*). *Anat. Rec.* **299**, 421–427 (2016).
54. Liu, X. et al. Diet and prey selection of the invasive american bullfrog (*Lithobates catesbeianus*) in Southwestern China. *Asian Herpetol. Res.* **6**, 34–44 (2015).
55. Maan, M. E. & Cummings, M. E. Sexual dimorphism and directional sexual selection on aposematic signals in a poison frog. *Proc. Natl Acad. Sci. USA* **106**, 19072–19077 (2009).
56. Nakamura, M. Sex determination in amphibians. *Semin. Cell Dev. Biol.* **20**, 271–282 (2009).
57. Freeman, M. E., Kanyicska, B., Lerant, A. & Nagy, G. Prolactin: structure, function, and regulation of secretion. *Physiol. Rev.* **80**, 1523–1631 (2000).
58. Singhas, C. A. & Dent, J. N. Hormonal control of the tail fin and of the nuptial pads in the male red-spotted newt. *Gen. Comp. Endocrinol.* **26**, 382–393 (1975).
59. Sherwood, O. D. Relaxin's physiological roles and other diverse actions. *Endocr. Rev.* **25**, 205–234 (2004).
60. de Rienzo, G., Aniello, F., Branno, M. & Minucci, S. Isolation and characterization of a novel member of the relaxin/insulin family from the testis of the frog *Rana esculenta*. *Endocrinology* **142**, 3231–3238 (2001).
61. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
62. Ye, C., Hill, C. M., Wu, S., Ruan, J. & Ma, Z. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci. Rep.* **6**, 1–9 (2016).
63. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
64. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. in *Curr. Protoc. Bioinformatics* **Chapter 4**, Unit 4.10 (2009).
65. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, 1–22 (2008).
66. Wang, Y. et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, 1–14 (2012).
67. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
68. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
69. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
70. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

## Acknowledgements

We are thankful to Professor Patricia Wittkopp and all members of the Wittkopp laboratory from the University of Michigan, Ann Arbor (USA) for their helpful comments and suggestions for this paper. We thank Dr. Jinzhong Fu from the University of Guelph for his kind help with explaining the results and writing the manuscript. We are thankful to Dr. Rebekah Rogers and Dr. Rasmus Nielsen for sharing the genomic data of the strawberry poison frog. This work was supported by the National Natural Science Foundation of China (No.31770405) and the Biodiversity Survey, Monitoring and Assessment Project of Ministry of ecology and environment, China (2019–2023).

## Author contributions

J.L. and H.W. designed the original concept and scientific objectives. J.L., W.W., C.F., and W.Z. acquired samples for sequencing. F.H. performed library preparation, genome sequencing, assembly, and annotation. H.Y. and F.H. characterized repetitive sequences and analyzed transcriptome data. J.L., H.Y., and H.W. investigated important genes related to developing nuptial spines. J.L. and H.Y. conducted WGCNA and analyzed keratin gene family. J.L., W.W., and C.F. conducted quantitative PCR test. H.W. obtained funding and other resources. J.L. and H.W. wrote the manuscript with input from other authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-019-13531-5>.

**Correspondence** and requests for materials should be addressed to H.W.

**Peer review information** *Nature Communications* thanks Andrew Crawford and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019