Coloring inside the lines: genomic architecture and evolution of a widespread color pattern in frogs

Authors

Sandra Goutte[a*], Imtiyaz Hariyani[a], Kole Deroy Utzinger[a], Yann Bourgeois[b] & Stéphane Boissinot[a*]

Affiliations

[a] New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE

[b] School of Biological Sciences, University of Portsmouth, Portsmouth, United Kingdom

[*]**Corresponding authors email addresses:** sb5272@nyu.edu (SB), sg5533@nyu.edu (SG)

Summary

Traits shared among distantly related lineages are indicators of common evolutionary constraints, at the ecological, physiological or molecular level. The vertebral stripe is a color pattern that is widespread across the anuran phylogeny. Despite its prevalence in the order, surprisingly little is known about the genetic basis and evolutionary dynamic of this color pattern. Here we combine histology, genome- and transcriptome-wide analyses with order-scale phylogenetic comparative analyses to investigate this common phenotype. We show that the vertebral stripe has evolved hundreds of times in the evolutionary history of anurans and is selected for in terrestrial habitats. Using the Ethiopian *Ptychadena* radiation as a model system, we demonstrate that variation at the *ASIP* gene is responsible for the different vertebral stripe phenotypes. Alleles associated to these phenotypes are younger than the split between closely related *Ptychadena* species, thus indicating that the vertebral stripe results from parallel evolution within the group. Our findings demonstrate that this widespread color pattern evolves rapidly and recurrently in terrestrial anurans, and therefore constitute an ideal system to study repeated evolution.

Keywords

Repeated evolution, adaptation, Agouti signaling protein, color polymorphism, Anura

Introduction

Animal color patterns are conspicuous hallmarks of selection. Color patterns may evolve because they are linked to a beneficial physiological trait[1,2] or because they serve as

sexual[3,4] or warning signals[5]. Alternatively, color pattern can help avoid detection from visually-oriented predators or prey by disrupting body shape recognition[6–9], masquerading as an object or animal[10], countershading[11], or substrate-matching[12–14]. In many species, multiple color patterns coexist within or between populations. These polymorphisms can be maintained by divergent mating strategies[15,16], apostatic selection (preference for the most common morph by predators), temporal or spatial habitat heterogeneity[5], or heterozygote advantage on correlated traits[2]. This diversity of selective regimes makes color patterns an ideal system to investigate the evolutionary mechanisms underlying phenotypic evolution.

The vertebral stripe is a common color pattern found across numerous distantly related anuran amphibians around the world (Fig. 1a). Phenotypes shared across highly divergent lineages can result from ancestral alleles conserved over millions of years of evolution, or have evolved independently multiple times, perhaps driven by similar selective forces. While predator-mediated selection is a widely assumed mechanism for the evolution and maintenance of most color patterns in anurans[17], the link between the anuran vertebral stripe and survival is only empirically supported in a few species[13,18,19] (but see[20]). In order to understand the evolutionary history of the vertebral stripe in anurans, a broad-scale comparative analysis is necessary. Understanding the genomic architecture underlying phenotypes can also inform on the evolutionary mechanisms at play. Despite the commonness of the pattern, the genetic basis of the anuran vertebral stripe is largely unknown. In the few species investigated, the stripe is determined by a

dominant allele at a single locus[17,21]. However, as these inferences were made solely on crossings, the identity of the locus or loci involved remains to be determined.

Here we combine macro- and microevolutionary analyses with transcriptomic and histological data to investigate the evolutionary history and genomic architecture underlying the vertebral stripe in anurans. By integrating results at three different scales (order, species group, and species), our study exemplifies how natural selection combined with rapidly evolving genomic regions may result in recurrent phenotypic evolution.

Results

Evolutionary history of the vertebral stripe in anurans

To retrace the evolutionary history of the vertebral stripe in anurans, we examined the dorsal color pattern of 2,785 anuran species for which phylogenetic data was available[22], representing 37.6% of species and 96.5% of families currently recognized in Anura[23]. A vertebral stripe morph was present in 15.7% of the 2,785 species included, and of those, 78% were polymorphic for the trait (Fig. 1b). Our analysis estimated that the vertebral stripe pattern evolved independently 330 ± 18 times across the phylogeny and was lost 515 ± 42 times (Fig. 1c). This result strongly supports the hypothesis of multiple origins of the anuran dorsal stripe and rejects the hypothesis of ancestral alleles conserved across anurans' evolutionary history.

Once we established that the vertebral stripe evolved independently multiple times, we investigated the role of the environment in the evolution of this trait. We hypothesized

that recurrent evolution of the vertebral stripe across anurans may be due to similar selective pressures in shared habitat types. To test this hypothesis, we carried a comparative analysis on 2,620 anuran species, and compared habitat-dependent and habitat-independent models of evolution for the vertebral stripe. Our analysis revealed that the rate of transition between unstriped and striped morphs is correlated with habitat (Supp. Fig. S1). The vertebral stripe evolved significantly more often in terrestrial clades compared to terrestrial-aquatic, arboreal, and terrestrial-arboreal linages (Fig 1d and 1e; Table 1). Arboreal lineages also showed the lowest gain and highest loss rates for the color pattern. The vertebral stripe may thus be selected against in arboreal habitats and selected for in terrestrial habitats.

Cellular organization of the vertebral stripe

To investigate the molecular and cellular mechanisms underlying the vertebral stripe in anurans, we focused on a radiation of terrestrial frogs, the *Ptychadena neumanni* species complex. This monophyletic group encompasses 12 species[24], which all present vertebral stripe polymorphism (the stripe could be thin, wide, or absent), except for two species: *P. harenna*, in which the vertebral stripe morph is always absent and *P. cooperi*, for which all individuals present a thin vertebral stripe (Fig. 2a and 2b). We examined the pigment cells organization in the dorsal skin of ten individuals of the *P. neumanni* species complex (*P. robeensis, P. nana, P. erlangeri* and *P. amharensis*) presenting different vertebral stripe phenotypes (thin or wide striped, or unstriped; Fig. 2b). Outside the stripe, numerous melanophores with dispersed melanosomes covering other chromatophores, and melanosomes in the epidermal layer create a dark coloration (Fig. 2c). Within the

stripe, melanophores with aggregated melanosomes (when present) and no or few epidermal melanosomes result in a lighter shade (Fig. 2c). The number and state (aggregated vs. dispersed) of the melanophores, as well as the concentration of epidermal melanosomes thus seem to be the major determinants of the vertebral stripe pattern in the species examined.

Genomic architecture of the vertebral stripe in Ptychadena robeensis

To identify the genomic region(s) involved in the dorsal stripe pattern, we conducted a genome-wide association study (GWAS) on one species of the *P. neumanni* complex, *Ptychadena robeensis,* which is polymorphic for the trait. We produced whole-genome resequencing data (4.84X average coverage) for 52 individuals with either a wide (n = 25) or a thin (n = 27) vertebral stripe and aligned the reads on the recently assembled chromosome-level genome of the species[25], resulting in a total of a 17,797,568 single-nucleotide polymorphisms (SNPs) dataset. The number of unstriped individuals collected being low (n=5), we excluded this phenotype from the analysis. We identified a single genomic region associated with the color pattern, which included multiple significant SNPs (Fig. 3a).

The identified SNPs are located on chromosome 11 downstream of a region containing two predicted[25] copies of the *Agouti signaling protein* gene (*ASIPa* and *ASIPb*; see Methods and Supp. Fig. S3). The majority of the SNPs are in the region overlapping with *ASIPb* and a non-coding RNA (ncRNA) sequence (Fig. 3a) located downstream of this gene, and all of them fall outside coding sequences. *ASIP* is known to be involved in melanophore

differentiation and melanin production in vertebrates[26,27]. By examining the phenotypes of homozygote and heterozygote individuals, we determined that the *wide* allele is dominant over the *thin* allele. Because all of these SNPs are located outside the predicted protein coding regions, we hypothesize that they affect the expression of the gene. An up-regulation of *ASIP* is linked to lighter phenotypes in mammals[28,29] and fishes[27], we could thus expect an up-regulation of *ASIP* in individuals with a wide vertebral stripe.

Differential gene expression associated with the vertebral stripe

We explored transcriptome-wide patterns of gene expression in the skin of adult *P. robeensis* presenting a thin, wide, or no vertebral stripe (Fig. 3b and 3c). Surprisingly, the expression levels of *ASIPa*, *ASIPb* and the ncRNA were very low in dorsal skin, and no significant differential expression could be detected between phenotypes or between skin samples collected within versus outside the stripe (Supp. Fig. S4). The transcripts were at considerably greater concentration in the ventral skin, which is white and lacks any melanization (Supp. Fig. S4a). To quantify more precisely *ASIP* expression in the dorsal skin, we conducted a quantitative real-time PCR experiment (qPCR), which confirmed comparable expression levels of *ASIPb* within and outside the vertebral stripe (Supp. Fig. S5).

Thirty two transcripts were found at significantly different abundance between the *thin* and *wide* morphs (Fig. 3b; Supp. Table S2), ten of which were associated with *per3*, a gene involved in the circadian rhythm of vertebrates[30,31]. Six transcripts were at lower concentration within the stripe compared to dorsal skin outside the stripe, and none

showed a higher abundance (Fig. 3c). Three of these were transcripts of *aldoa,* a gene expressed during melanogenesis in mammals[32] and up-regulated in yellow compared to black carp[33].

Recent evolution of the thin and wide alleles

To estimate the age of the *wide* and *thin* alleles and look for signatures of selection, we used Ancestral Recombination Graphs (ARG) analyses on four individuals of *P. robeensis* representative of the different phenotypes which were sequenced at a higher coverage (12.97X on average). Times to the Most Recent Common Ancestor (TMRCA) for the *thin* and *wide* alleles were more recent than the surrounding genomic regions or the total population TMRCAs in region directly surrounding the most significant SNPs outputted by the GWAS (Fig. 4a). This result is consistent with a partial selective sweep and excludes the possibility of an ancient polymorphism maintained by balancing selection. Both the *wide* and *thin* alleles have evolved recently, 100,000-300,000 years ago. These alleles thus arose long after the divergence between *Ptychadena robeensis* and its closest relatives, estimated at 3.8-8.3 million years ago[34].

When comparing the genomic region of interest between all 12 species of the *P. neumanni* complex (8.29X average coverage), we found that the *thin* and *wide* alleles were indeed private to *Ptychadena robeensis* and not shared with the other *Ptychadena* species (Fig. 4.b). This result is further supported by a phylogeny of the region of interest where haplotypes are grouped according to clades within the radiation, and not phenotypes (Fig. 4.b). Additional GWAS analyses on two other members of the *P. neumanni* complex, *P.*

amharensis (n = 32; 1.95X average coverage; 82,580,376 SNPs) *and P. beka* (n = 42; 4.58X average coverage; 33,430,567 SNPs), failed to detect a single locus responsible for the dorsal pattern for either species. These results indicate that the alleles found in *P. robeensis* have evolved recently and other alleles are responsible for similar dorsal patterns in closely related species.

Discussion

In this study, we show that the anuran vertebral stripe evolved multiple times, and significantly more often in terrestrial lineages compared to terrestrial-aquatic, arboreal, and terrestrial-arboreal linages. The vertebral stripe might increase concealment from visually oriented predators, such as birds or mammals, which are more prevalent in terrestrial habitats, and thus confer a fitness advantage. The widespread polymorphism for the trait may be maintained because some predators attack preferentially the most common morph and/or because the cryptic advantage of the dorsal pattern changes across the environment or through seasons. The vertebral stripe polymorphism could also be the result of recently evolved alleles on their way to fixation. This process could be particularly slow if alleles causing the vertebral stripe are dominant[17,21] and the stripe adaptive advantage is only moderate. Interestingly, the vertebral stripe was lost significantly more often in arboreal lineages than in other groups, indicating a potential fitness cost of the pattern in this habitat. Other color patterns could also be selected for in arboreal habitats and the vertebral stripe could be lost as a side-effect. Although the molecular and evolutionary mechanisms may vary across lineages, a shared selective pressure favored the presence of striped morphs in terrestrial anurans.

196

197 In the grass frog *Ptychadena robeensis,* we identified two *ASIP* alleles responsible for
198 distinct vertebral stripe morphs. In doing so, we establish the first evidence of a causal link
199 between *ASIP* and a color pattern in amphibians. The involvement of *ASIP* in melanophore
200 differentiation and melanin production has been extensively studied in mammals[28,35–37],
201 and in rodents in particular[38–40]. Vertebral stripe pattern differentiation is likely governed
202 by differential expression of *ASIP* as alleles differ in non-protein coding regions. However,
203 the expression level of *ASIP* is extremely low and does not differ between morphs in adult
204 dorsal skin. Morph-dependent differential expression of *ASIP* may thus occur at an earlier
205 stage of the animal's development[41], or punctually during chromatophore differentiation
206 resulting in an overall low level of expression in the dorsal skin. In the *Ptychadena*
207 *neumanni* species complex, the vertebral stripe first appears at the final stages or after
208 metamorphosis (personal observation based on 92 individuals identified through
209 barcoding at different developmental stages). Investigating *ASIP's* expression levels in the
210 dorsal skin of metamorphic and juvenile individuals will be necessary to determine the
211 gene's role in the establishment of the color pattern during development.

212

213 Other genes were found to be differentially expressed (DE genes) between morphs
214 and between skin sections (within and outside the vertebral stripe). While morph-
215 dependent DE genes are likely to be involved in the same pathways as *ASIP*, the genes
216 differentially expressed between sections of the dorsal skin could be produced by the
217 different mature chromatophores. As opposed to the dorsal skin, *ASIP* expression levels
218 were high in the ventral skin, which is, as in many other anurans, uniformly white. This

dorso-ventral differential expression is comparable to expression patterns found in several species of fish which also present a white ventrum[27,41]. Interestingly, the difference in expression level between dorsal and ventral skin was most marked for the ncRNA located downstream of *ASIPb* (Supp. Fig. S4). Together with the proximity of significant SNPs outputted by the GWAS (Fig. 3), this indicates that this ncRNA is a major actor in color pattern determination in this species, likely by having a regulatory role on *ASIPa* and/or *ASIPb*. *ASIP* thus seems to play a determining role in both the dorsal color pattern and the lack of melanization in ventral skin.

In multiple organisms, *ASIP* alleles have evolved rapidly leading to parallel evolution of similar phenotypes within species[42] or groups of closely-related species[28]. In the *P. neumanni* species complex, species presenting the same color patterns did not share the *ASIP* alleles with *P. robeensis*. The lack of signal in the GWAS conducted for *P. amharensis* and *P. beka* could also be indicative of multiple haplotypes leading to the same phenotypes within these species, which occupies much larger distribution ranges than *P. robeensis* and presents greater intraspecific genetic variances. We estimated *P. robeensis' ASIP* alleles to have evolved less than 300,000 years ago, much after diverging from its closest known relatives. *Ptychadena robeensis* occupies grasslands and cultivated fields where opportunities for hiding from predator under the vegetation are rare and substrate varies from grass to bare soil. A vertebral stripe (wide or thin) may thus provide a significant fitness advantage by enhancing crypsis or disrupting body shape recognition.

241 Recurrent evolution in the regulatory region of *ASIP* could have led to the same
242 color patterns in this group of closely-related species, similarly to the horizontal stripes in
243 African cichlids caused by repeated evolution at *agrp2* regulatory region[43]. However,
244 mutations impacting the expression of genes interacting with *ASIP* (such as *MC1R* for
245 example[27]) may also be responsible for the vertebral stripe in *Ptychadena* spp. and other
246 terrestrial anurans. By demonstrating the involvement of *ASIP* in a widespread trait, our
247 study opens new research avenues on color patterns in anurans. Mutations in the
248 regulatory regions of *ASIP* or interacting genes causing the appearance or loss of the
249 vertebral stripe are likely occurring at a high rate in anurans, making this trait an ideal
250 system to study parallel evolution.
251

252 **Methods**

253 *Comparative analysis of vertebral stripe evolution in anurans*

254 We conducted a comparative analysis across the Anura using the largest dated
255 molecular phylogeny of amphibians published to date[22]. This phylogeny comprises 3,309
256 species (=45% of currently recognized species[23]), representing most families, subfamilies
257 and genera[22]. We collected data on dorsal color pattern for 2,785 of these species by
258 examining all photographs available for each species on *Amphibiaweb*
259 (https://amphibiaweb.org). When no or few photos were available, we searched additional
260 sources such as original species descriptions and the number of photographs examined for
261 each species was systematically recorded (Supp. Table S1). Dorsal color patterns were
262 classified in the following categories: thin, medium or wide stripe, and unstriped. If a

263 species had at least one individual counted in the *unstriped* and one of the *striped*
264 categories, the species was considered polymorphic for the trait.
265
266 Habitat use data was collected independently for 2,620 species based on multiple
267 large studies (Supp. Table S1). Anuran habitats were categorized as arboreal, aquatic,
268 terrestrial, arboreal/terrestrial and terrestrial/aquatic, based on the main habitat occupied
269 by adult individuals (when reproductive and general habitats were available).
270
271 *Ancestral state reconstruction of dorsal color patterns*
272 To reconstruct the evolutionary history of the vertebral stripe, we created 1,000
273 stochastic maps of the trait onto the phylogeny using the function *make.simmap* in the R
274 package *phytools*[44]. Because most of the *striped* species were polymorphic for the trait
275 (78%), and as erroneous categorization was more likely for the fixed than for the
276 polymorphic categories (if only few photos were available), we recategorized species as
277 either having at least some individuals presenting a vertebral stripe, or without any striped
278 individuals. We thus used two categories, *striped* (including polymorphic species) and
279 *unstriped* with a model allowing transition rates between the two morphs to be different
280 (ARD), and estimated the number of transitions during the evolutionary history of anurans.
281
282 *Dorsal pattern evolution in different habitats*
283 To test the hypothesis that the dorsal stripe might be selected for in particular
284 habitats, we first built 100 stochastic maps of habitat data for the 2,620 species categorized
285 on the phylogeny. For each of the stochastic trees, we fitted a model for which transition

rates between dorsal color patterns was independent of habitat and a model for which transition rates differed for each habitat using the *fitmultiMk* function from the R package *phytools*, and compared them using a likelihood ratio test (Supp. Figure S1). As the habitat-dependent model of evolution for the vertebral stripe was systematically better fitted than the independent model, we extracted the estimated transition rates between *striped* and *unstriped* phenotypes for each habitat and each of the 100 stochastic maps. Because few species were categorized as aquatic (1.87 % of included taxa, i.e. 49 species), the variance in transition rate estimates was much greater than for the other habitats (Supp. Figure S2), so we excluded aquatic lineages from further analyses. We compared the transition rates between pairs of habitats using a Tukey honest significant differences test (Table 1).

Sampling of Ethiopian Ptychadena

Individuals of the *Ptychadena neumanni* species complex were collected in the highlands of Ethiopia between 2011 and 2019. Our study was approved by the relevant Institutional Animal Care and Use Committee at Queens College and New York University School of Medicine (IACUC; Animal Welfare Assurance Number A32721–01 and laboratory animal protocol 19–0003). Frogs were sampled according to permits DA31/305/05, DA5/442/13, DA31/454/07, DA31/192/2010, DA31/230/2010, DA31/7/2011 and DA31/02/11 provided by the Ethiopian Wildlife Conservation Authority. We photographed individuals in life and euthanized them by ventral application of 20% benzocaine gel. We extracted tissue samples and stored them in RNAlater or 95% ethanol. Adult individuals were fixed in 10% formalin for 24 to 48 hours, and then transferred to 70% ethanol. After preservation, we took additional photographs of all individuals. All specimens were

309 deposited at the Natural History Collection of the University of Addis Ababa, Ethiopia.
310 Tissue samples are deposited at the Vertebrate Tissue Collection, New York University Abu
311 Dhabi (NYUAD).
312
313 *Histological skin sections*
314 Dorsal and ventral skin sections were extracted from ten preserved adult
315 specimens: two thin striped (SB81, SB82) and one unstriped (SB69) *Ptychadena robeensis*
316 specimens, two thin striped (SB493, SB510) and one wide striped (SB494) *P. nana*, two
317 wide striped (SB552, SB548) *P. erlangeri*, and two wide striped (SB584, SB593) *P.*
318 *amharensis.* The skin samples were embedded in paraffin blocks and sections of 4 µm
319 thickness were produced. The sections were stained with hematoxylin-eosin (HE) and
320 chromatophores were examined using a Leica DMI 6000 B microscope.
321
322 *DNA and RNA extractions and sequencing*
323 Genomic DNA of 61 *Ptychadena robeensis* individuals was extracted from liver tissue
324 using the DNeasy blood and tissue kit (Qiagen, Valencia, CA). RNA was extracted from the
325 skin of 13 individuals using a RNeasy mini kit (Qiagen, Valencia, CA). For eight individuals,
326 RNA was extracted within and outside the vertebral stripe separately, for three individuals
327 lacking any dorsal pattern, a single sample of dorsal skin was used, and for two individuals,
328 RNA from ventral skin was extracted. We quantified extracted DNA and RNA using a Qubit
329 fluorometer (Life Technologies). Libraries were prepared using a NEB library prep kit and
330 sequenced on Illumina NextSeq 550 flow cells at the Genome Core Facility of New York
331 University Abu Dhabi. After quality filtering, reads were aligned to the *de novo* assembly

Ptychadena robeensis reference genome[25]. The average coverage of the genomic data was 4.84X, except for four individuals which we sequenced at an average of 12.97X. Variants were called using the function *HaplotypeCaller* from *gatk v3.5*[45]. The low-coverage and higher-coverage samples were then combined and genotyped in two separate datasets using *CombineGVCF* and *GenotypeGVCFs* functions from *gatk*.

Genome-wide association study on P. robeensis

After examination of the low-coverage genomic dataset (n=61 individuals; SNPs PCA and visual examination in IGV), we realized that five individuals were hybrids , likely resulting from the crossing of *P. robeensis* and *P. levenorum*, a closely related species with a partially overlapping distribution range[24]. After removing the hybrids, the dataset comprised 17,797,568 SNPs. We excluded four individuals, which presented no dorsal melanization and could not be categorized as thin or wide striped, and our final dataset contained 52 individuals (25 wide striped and 27 thin striped. Stripe phenotype (thin, wide or unstriped) and coloration (brown or green) were not correlated. Quality checks and the genome-wide association study (GWAS) were done using *PLINK 1.9*[40,41]. We checked for individual relatedness as well as major discrepancies between samples in data missingness and minor allele frequency. However, because of the low-coverage nature of our data, we did not apply any stringent quality filtering. We extracted and visualized the result of the GWAS using the R package *qqman*[48].

Test for selection and alleles ages in Ptychadena robeensis

We searched for signatures of selection in the regions linked to dorsal stripe pattern and determined the age of the alleles using *ARGweaver*. In short, *ARGweaver* reconstructs a set of Ancestral Recombination Graphs (ARGs) for every non-recombining interval in the genomic region of interest. The program then samples from the posterior distribution of ARGs given an evolutionary model. Regions under positive selection should display a reduced coalescence time whereas regions under ancient balancing selection should have older coalescence time compared to neutral regions.

We ran *ARGweaver* on the high-coverage *Ptychadena robeensis* dataset (12.97X average coverage) comprising four individuals. We used the mutation rate estimated for the species group[34] 6.98e-10/bp/generation (with a 2-year generation time) and the average recombination rate estimated for *Xenopus tropicalis*[49], 9.73e-9/bp/generation as priors. The effective population size was estimated as a function of time in SMC++[50] elsewhere[51], and we used a maximum coalescence time of 4 million generations, around twenty times the harmonic mean of the estimated effective population size over time. *ARGweaver* was run for 5,000 iterations and ARGs were sampled every 100 iteration. Convergence of the chain was monitored visually by plotting multiple ARG statistics (priors, likelihood, number of recombination events, total branch length, number of variant sites not explained by a single mutation) against iteration number. All statistics were stationary after the first 2,000 iterations, so we discarded these 2,000 first iterations as burn-in for further analyses.

We inspected the phased haplotypes and categorized them in *wide* or *thin* haplotypes, thereby considering haplotypes rather than individuals in the subsequent analyses. We extracted times to the most recent common ancestor (TMRCA) for the *wide* and *thin* haplotypes, as well as for the whole population and compared the three curves across our region of interest. Ancient balancing selection should show TMRCA older than neutral genomic regions and an equal TMRCA between the haplotypes and the overall population. Positive selection or recent balancing selection, on the other hand, should have and overall TMRCA similar to neutral regions but haplotype TMRCAs more recent than the overall TMRCA.

Phylogeny of haplotypes

In order to compare the region of interest across the *Ptychadena neumanni* species complex, we reconstructed a haplotype phylogeny. The genomes of the 12 species (one individual per species, except for *P. robeensis* for which the genomes of five individuals were included) were phased using Beagle 5.1[52]. We then built a phylogeny of haplotypes in the region of interest (40kb region; 6,677 SNPs) using the R package *SNPRelate*[53] (Fig. 4b).

Gene expression analysis

Reads were aligned to the annotated reference genome using HISAT2[54] and StringTie2[55]. A transcriptome-wide gene count matrix was then created using the script *prepDE.py3* provided on the StringTie website. Subsequent analyses were conducted in the *R* environment[56]. We used the *R* package *edgeR*[57] to filter and normalized our data prior analysis. We filtered out genes which had a count inferior to 1 count-per-million (cpm) in at

400 least 16 samples (>50% of the 21 samples in total) and applied a "Trimmed Mean of M-
401 values" (TMM) normalization of the data using the *R* package *DESeq2*[58]. We then identified
402 differentially expressed (DE) genes between wide and thin striped individuals as well as
403 between sections of dorsal skin within and outside the vertebral stripe using the *DESeq*
404 function of the same package.
405
406 *Annotation of the region of interest*
407 In order to determine the type of variant responsible for the vertebral stripe pattern
408 in *Ptychadena robeensis*, we annotated the region of chromosome 11 containing significant
409 SNPs in the GWAS. We visually examined the transcripts obtained from our RNAseq data
410 against the annotation for protein coding sequences predicted by Augustus[25,59] in the
411 region using IGV[60]. Significant SNPs outputted by the GWAS were all located outside
412 protein-coding regions, in between three genes: two predicted genes 40 kb apart were
413 identified as *ASIP* and a third, 38kb downstream, was identified as *AHCY* using MegaBlast[61].
414
415 While a single *ASIP* gene is known in birds and mammals, two genes, *ASIP1* and
416 *ASIP2* are present in teleost fish[27,62]. In *Xenopus tropicalis*, a single *ASIP* gene is predicted,
417 but no focal study in amphibians has been conducted to date. In order to determine
418 whether the two genes in *P. robeensis* correspond to *ASIP*, *ASIP1,* or *ASIP2*, we translated
419 the genomic sequences and produced a protein alignment and maximum likelihood
420 phylogeny with the *ASIP* gene family in vertebrates using *seaview*[63] (Supp. Fig. S3). Both
421 genes grouped together and with other amphibians' *ASIP*. We thus named them *ASIPa* and
422 *ASIPb* to avoid any confusion with *ASIP1* and *ASIP2* from teleost fish. *ASIPa* is composed of

six exons, two of which are copies of *ASIPb*'s only two exons. Additionally, we detected a non-coding RNA (ncRNA) downstream *ASIPb*. A protein alignment revealed that this ncRNA contains a region aligning with the third exon of *ASIP* in other vertebrates, which is absent from *ASIPa* and *ASIPb*.

Quantitative real-time PCR experiment

Quantitative real time PCR (qPCR) was conducted on RNA extracted from dorsal skin within (n=2) and outside (n=4) the vertebral stripe, and ventral skin (n=1). Each reaction was triplicated to minimize the impact of experimental error. Two reference genes were selected using our RNAseq dataset with the following criteria: a minimum of 50 count per million in all samples, the lowest possible variance in expression level among samples and a minimum of two exons. Candidate reference genes were then checked for functional independence and compared to genes typically used for qPCR in *Xenopus laevis*. We retained *rpl27* and *abce1* as candidate reference genes.

Primers for *ASIPb, rpl27* and *abce1* were designed based on the *Ptychadena robeensis* reference genome and annotation using Primer3Plus[64]. Primers were designed to span an exon-intron junction to avoid amplification of genomic DNA during the qPCR. The experiment was run using a StepOnePlus real-time PCR system and a Power SYBR Green RNA-to-CT 1-step kit (Applied Biosystems) on a volume of 20µl. Results were analyzed using the R package *pcr*[65]. We compared the expression levels of *rpl27* and *abce1* across samples and retained *rpl27*. Relative expression levels of *ASIPb* between our samples were

445 calculated using *rpl27* as reference gene and dorsal skin outside the stripe as reference
446 group as it should have the lowest level of *ASIPb*.
447

448 **Acknowledgments**

449 We would like to thank the Ethiopian Wildlife Conservation Authority and the
450 Ethiopian Biodiversity Institute for providing us with collecting and export permits for the
451 samples. Fieldwork in Ethiopia would not have been possible if not for the invaluable
452 assistance of Megersa Kelbessa, Itbarek, and Samuel Woldeyes of Rock Hewn Tours. We
453 also thank the important number of students and postdocs who collected *Ptychadena*
454 specimens and samples over the years, and in particular Xenia Freilich, Jacobo Reyes-
455 Velasco, Justin Wilcox, Sebastian Kirchhof and Marcin Falis. We are very thankful for the
456 help from Marc Arnoux and Nizar Drou, from the Genome Core Facility and the
457 Bioinformatics group at NYUAD. This research was carried out on the High-Performance
458 Computing resources at New York University Abu Dhabi. We also thank David Howse and
459 Sayel Daoud of The National Reference Laboratory and Rachid Rezgui from the Microscopy
460 Core Facility at NYUAD for their help in producing and visualizing histological sections.
461

462 **Funding**. This project was funded by NYUAD Grant AD180 to SB. The NYUAD Sequencing
463 Core is supported by NYUAD Research Institute grant G1205A to the NYUAD Center for
464 Genomics and Systems Biology.
465

Author contributions

SG and SB designed the study. KDU and SG collected color pattern and habitat data. SG, YB, and SB collected *Ptychadena* spp. specimens and samples. SG extracted DNA and RNA from *Ptychadena* spp. samples and ran the and qPCR. SG produced and interpreted histological section photographs and conducted comparative, genomic and gene expression analyses. IH and YB provided help and advice on genomic and transcriptomic analyses. All authors read and contributed to the manuscript.

Competing interests

The authors declare no competing interest.

References

1. Bull, C. M. Black pattern polymorphism and tadpole growth rate in two Western Australian frogs. *Aust. J. Zool.* **25**, 243–248 (1977).
2. Otaki, J. M. Physiological side-effect model for diversification of non-functional or neutral traits : a possible evolutionary history of Vanessa butterflies (*Lepidoptera*, Nymphalidae). *Lepidoptera Sci.* **59**, 87–102 (2008).
3. Chen, I.-P., Stuart-Fox, D., Hugall, A. F. & Symonds, M. R. E. Sexual Selection and the Evolution of Complex Color Patterns in Dragon Lizards. *Evolution* **66**, 3605–3614 (2012).
4. Zhou, W. *et al.* Sexual selection on jumping spider color pattern: investigation with a new quantitative approach. *Behav. Ecol.* (2021) doi:10.1093/beheco/arab008.
5. Briolat, E. S. *et al.* Diversity in warning coloration: selective paradox or the norm? *Biol. Rev.* **94**, 388–414 (2019).

6. Osorio, D., Srinivasan, M. V. & Pettigrew, J. D. Camouflage by edge enhancement in animal coloration patterns and its implications for visual mechanisms. *Proc. R. Soc. Lond. B Biol. Sci.* **244**, 81–85 (1991).
7. Stevens, M. & Cuthill, I. C. Disruptive coloration, crypsis and edge detection in early visual processing. *Proc. R. Soc. B Biol. Sci.* **273**, 2141–2147 (2006).
8. Cuthill, I. C. & Székely, A. Coincident disruptive coloration. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 489–496 (2009).
9. Ruxton, G. D., Allen, W. L., Sherratt, T. N. & Speed, M. P. *Avoiding Attack: The Evolutionary Ecology of Crypsis, Aposematism, and Mimicry*. *Avoiding Attack* (Oxford University Press, 2018).
10. Skelhorn, J., Rowland, H. M., Speed, M. P. & Ruxton, G. D. Masquerade: Camouflage Without Crypsis. *Science* **327**, 51–51 (2010).
11. Rowland, H. M., Cuthill, I. C., Harvey, I. F., Speed, M. P. & Ruxton, G. D. Can't tell the caterpillars from the trees: countershading enhances survival in a woodland. *Proc. R. Soc. B Biol. Sci.* **275**, 2539–2545 (2008).
12. Godfrey, D., Lythgoe, J. N. & Rumball, D. A. Zebra stripes and tiger stripes: the spatial frequency distribution of the pattern compared to that of the background is significant in display and crypsis. *Biol. J. Linn. Soc.* **32**, 427–433 (1987).
13. Woolbright, L. L. & Stewart, M. M. Spatial and Temporal Variation in Color Pattern Morphology in the Tropical Frog, *Eleutherodactylus coqui*. *Copeia* **2008**, 431–437 (2008).
14. Rojas, B. Behavioural, ecological, and evolutionary aspects of diversity in frog colour patterns. *Biol. Rev.* **92**, 1059–1080 (2017).

15. Hurtado-Gonzales, J. L. & Uy, J. A. C. Alternative mating strategies may favour the persistence of a genetically based colour polymorphism in a pentamorphic fish. *Anim. Behav.* **77**, 1187–1194 (2009).

16. Pérez i de Lanuza, G., Font, E. & Carazo, P. Color-assortative mating in a color-polymorphic lacertid lizard. *Behav. Ecol.* **24**, 273–279 (2013).

17. Hoffman, E. A. & Blouin, M. S. A review of colour and pattern polymorphisms in anurans. *Biol. J. Linn. Soc.* **70**, 633–665 (2000).

18. Stewart, M. M. Parallel Pattern Polymorphism in the Genus *Phrynobatrachus* (Amphibia: Ranidae). *Copeia* **1974**, 823–832 (1974).

19. McElroy, M. T. Teasing apart crypsis and aposematism – evidence that disruptive coloration reduces predation on a noxious toad. *Biol. J. Linn. Soc.* **117**, 285–294 (2016).

20. Anderson, N. K., Gutierrez, S. O. & Bernal, X. E. From forest to city: urbanization modulates relative abundance of anti-predator coloration. *J. Urban Ecol.* **5**, (2019).

21. O'Neill, E. M. & Beard, K. H. Genetic Basis of a Color Pattern Polymorphism in the Coqui Frog *Eleutherodactylus coqui*. *J. Hered.* **101**, 703–709 (2010).

22. Pyron, R. A. Biogeographic Analysis Reveals Ancient Continental Vicariance and Recent Oceanic Dispersal in Amphibians. *Syst. Biol.* **63**, 779–797 (2014).

23. Frost, D. R. Amphibian species of the world: an online reference. Version 6.0. http://research.amnh.org/herpetology/amphibia/index.html (2018).

24. Goutte, S., Reyes-Velasco, J., Freilich, X., Kassie, A. & Boissinot, S. Taxonomic revision of grass frogs (Ptychadenidae, *Ptychadena*) endemic to the Ethiopian highlands. *ZooKeys* **1016**, 77–141 (2021).

25. Hariyani, I., Reyes-Velasco, J. & Boissinot, S. A chromosome-scale genome assembly for the Ethiopian Highland frog *Ptychadena robeensis*. (*in preparation*).

26. Sviderskaya, E. V., Hill, S. P., Balachandar, D., Barsh, G. S. & Bennett, D. C. Agouti signaling protein and other factors modulating differentiation and proliferation of immortal melanoblasts. *Dev. Dyn.* **221**, 373–379 (2001).

27. Cal, L., Suarez-Bregua, P., Cerdá-Reverter, J. M., Braasch, I. & Rotllant, J. Fish pigmentation and the melanocortin system. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **211**, 26–33 (2017).

28. Liang, D. *et al.* Genomic Analysis Revealed a Convergent Evolution of LINE-1 in Coat Color: A Case Study in Water Buffaloes (*Bubalus bubalis*). *Mol. Biol. Evol.* **38**, 1122–1136 (2020).

29. Jones, M. R. *et al.* Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* (2018).

30. Nadkarni, N. A., Weale, M. E., von Schantz, M. & Thomas, M. G. Evolution of a Length Polymorphism in the Human PER3 Gene, a Component of the Circadian System. *J. Biol. Rhythms* **20**, 490–499 (2005).

31. Li, Y., Li, G., Wang, H., Du, J. & Yan, J. Analysis of a Gene Regulatory Cascade Mediating Circadian Rhythm in Zebrafish. *PLOS Comput. Biol.* **9**, e1002940 (2013).

32. Slominski, A. *et al.* The role of melanogenesis in regulation of melanoma behavior: Melanogenesis leads to stimulation of HIF-1α expression and HIF-dependent attendant pathways. *Arch. Biochem. Biophys.* **563**, 79–93 (2014).

33. Ye, X. *et al.* ITRAQ Proteomic Analysis of Yellow and Black Skin in Jinbian Carp (Cyprinus carpio). *Life* **10**, (2020).

555 34. Freilich, X., Tollis, M. & Boissinot, S. Hiding in the highlands: Evolution of a frog species
556 complex of the genus *Ptychadena* in the Ethiopian highlands. *Mol. Phylogenet. Evol.* **71**,
557 157–169 (2014).
558 35. Girardot, M. *et al.* The insertion of a full-length Bos taurus LINE element is responsible for a
559 transcriptional deregulation of the Normande Agouti gene. *Pigment Cell Res.* **19**, 346–355
560 (2006).
561 36. Norris, B. J. & Whan, V. A. A gene duplication affecting expression of the ovine ASIP gene is
562 responsible for white and black sheep. *Genome Res.* **18**, 1282–1293 (2008).
563 37. Almathen, F., Elbir, H., Bahbahani, H., Mwacharo, J. & Hanotte, O. Polymorphisms in MC1R
564 and ASIP Genes are Associated with Coat Color Variation in the Arabian Camel. *J. Hered.*
565 **109**, 700–706 (2018).
566 38. Galbraith, D. B. The agouti pigment pattern of the mouse: A quantitative and experimental
567 study. *J. Exp. Zool.* **155**, 71–89 (1964).
568 39. Siracusa, L. D. The agouti gene: turned on to yellow. *Trends Genet.* **10**, 423–428 (1994).
569 40. Dolinoy, D. C. The agouti mouse model: an epigenetic biosensor for nutritional and
570 environmental alterations on the fetal epigenome. *Nutr. Rev.* **66**, S7-11 (2008).
571 41. Ceinos, R. M., Guillot, R., Kelsh, R. N., Cerdá-Reverter, J. M. & Rotllant, J. Pigment patterns
572 in adult fish result from superimposition of two largely independent pigmentation
573 mechanisms. *Pigment Cell Melanoma Res.* **28**, 196–209 (2015).
574 42. Jones, M. R., Mills, L. S., Jensen, J. D. & Good, J. M. Convergent evolution of seasonal
575 camouflage in response to reduced snow cover across the snowshoe hare range. *Evolution*
576 **74**, 2033–2045 (2020).

43. Kratochwil, C. F. *et al.* Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science* **362**, 457–460 (2018).

44. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

45. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

46. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).

47. Purcell, S. & Chang, C. *PLINK*. (2017).

48. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv* 005165 (2014) doi:10.1101/005165.

49. Wells, D. E. *et al.* A genetic map of *Xenopus tropicalis*. *Dev. Biol.* **354**, 1–8 (2011).

50. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).

51. Hariyani, I., Kirchhof, S., Bourgeois, Y., Reyes-Velasco, J. & Boissinot, S. Niche divergence and genomic adaptation shape an adaptive radiation of Ethiopian Highland anurans. *In prep*.

52. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet* **108**, 1880–1890.

53. Zheng, X., Gogarten, S., Laurie, C. & Weir, B. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* **28**, 3326–3328 (2012).

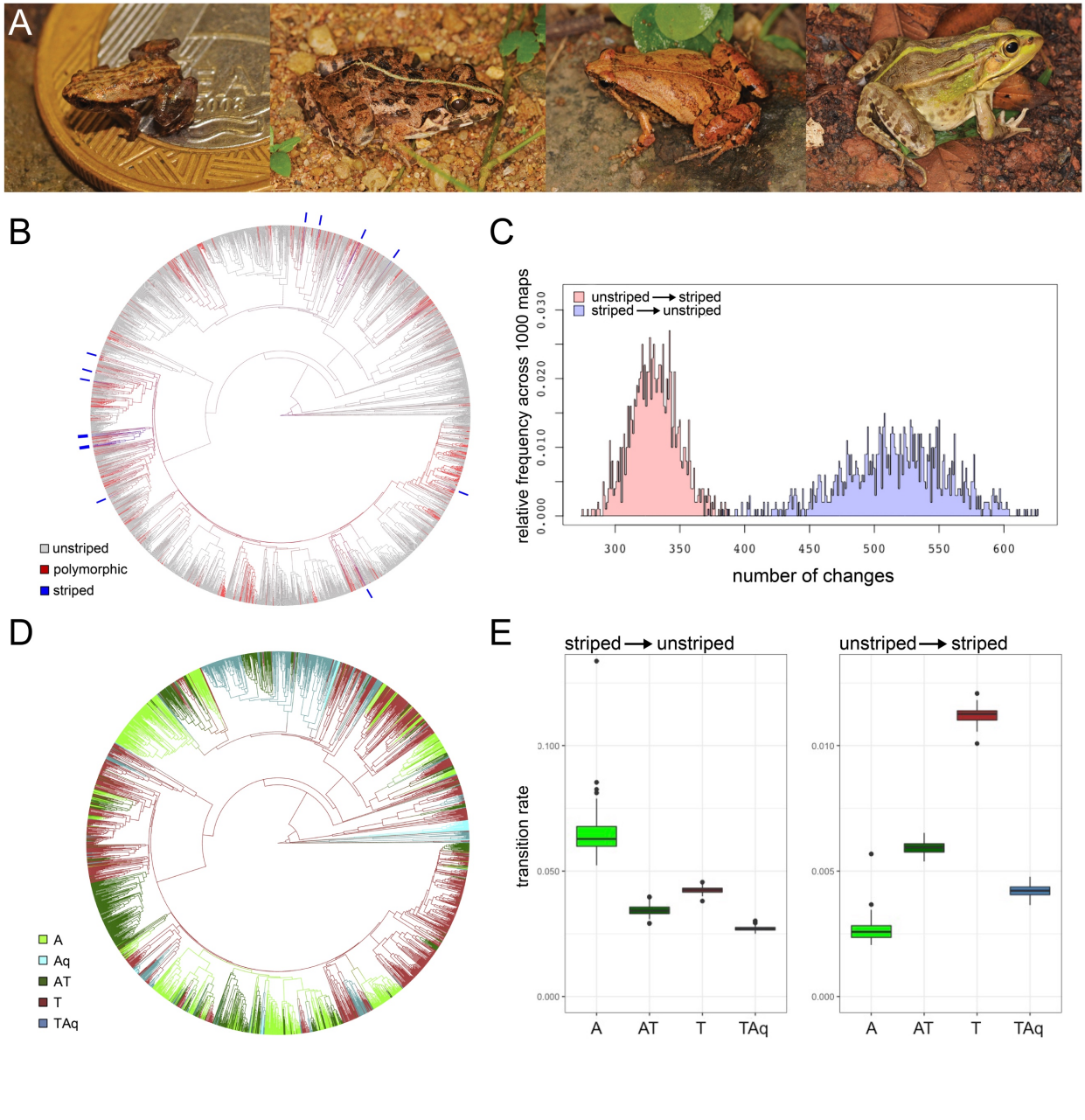
599 54. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment
600 and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
601 55. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with
602 StringTie2. *Genome Biol.* **20**, 278 (2019).
603 56. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for
604 Statistical Computing, Vienna, Austria, 2020).
605 57. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
606 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140
607 (2010).
608 58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
609 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
610 59. Keller, O., Kollmar, M., Stanke, M. & Waack, S. A novel hybrid gene prediction method
611 employing protein multiple sequence alignments. *Bioinformatics* **27**, 757–763 (2011).
612 60. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
613 61. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5-9 (2008).
614 62. Cortés, R. *et al.* Evolution of the melanocortin system. *Gen. Comp. Endocrinol.* **209**, 3–10
615 (2014).
616 63. Galtier, N., Gouy, M. & Gautier, C. SEAVIEW and PHYLO_WIN: two graphic tools for
617 sequence alignment and molecular phylogeny. *Comput. Appl. Biosci. CABIOS* **12**, 543–548
618 (1996).
619 64. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.*
620 **35**, W71–W74 (2007).

621 65. Ahmed, M. & Kim, D. R. pcr: an R package for quality assessment, analysis and testing of
622 qPCR data. *PeerJ* **6**, e4473 (2018).
623 66. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
624 *Nucleic Acids Res.* **32**, 1792–1797 (2004).

625
626

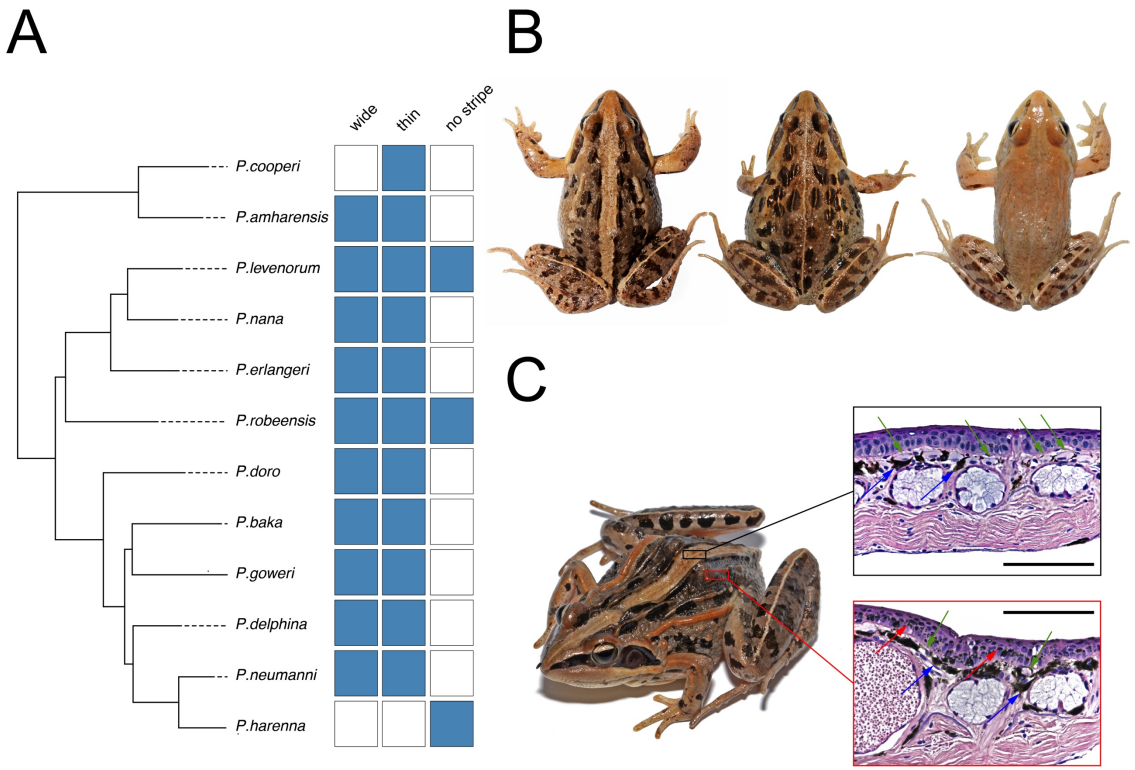
627 **Figure titles and legends**

628 **Figure 1. The evolution of the vertebral stripe in anurans. A** Examples of the vertebral
629 stripe in distantly related species: from left to right, *Brachycephalus hermogenesi* (family:
630 Brachycephalidae), *Fejervarya limnocharis* (family: Dicroglossidae), *Microhyla ornata*
631 (family: Microhylidae), *Pelophylax nigromaculatus* (family: Ranidae), **B** vertebral stripe
632 morphs (unstriped = grey, striped = blue, polymorphic = red; monomorphic striped taxa
633 are further indicated by blue bars for readability) mapped on the phylogeny of Anura
634 (n=2,785 species; 1,000 stochastic maps), **C** estimated number of transitions between
635 striped and unstriped morphs in the evolution of anurans based on 1,000 stochastic maps,
636 **D** habitat use mapped on the phylogeny of Anura (n= 2,620 species; 100 stochastic maps),
637 **E** transition rates between striped and unstriped phenotypes for each habitat, estimated
638 for 100 stochastic maps. Habitat categories: A = arboreal, Aq = aquatic, AT = arboreal-
639 terrestrial, T = terrestrial, TAq = terrestrial-aquatic.



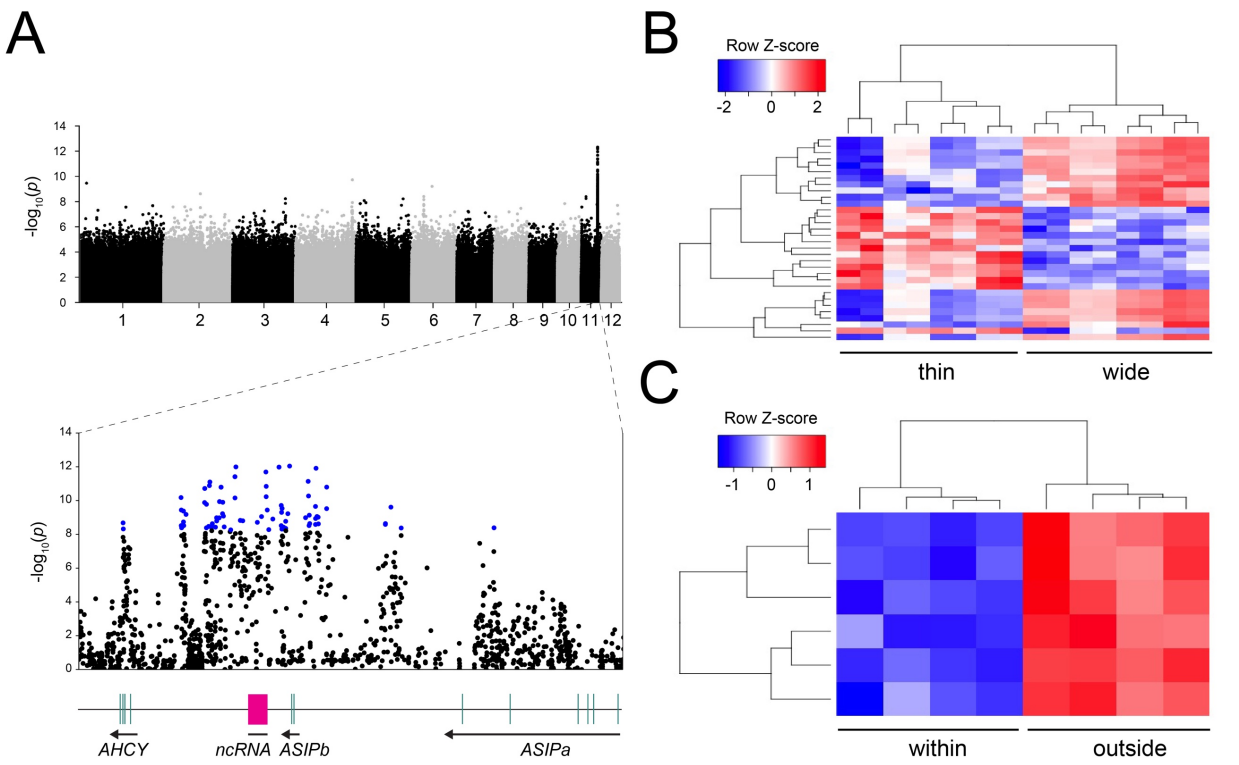
640
641
642

643 **Figure 2. The vertebral stripe in Ethiopian *Ptychadena.* A** Polymorphism of the
644 vertebral stripe (wide or thin striped, or unstriped) in the *Ptychadena neumanni* species
645 complex (phylogeny based on 500,000 genome-wide distributed SNPs, reproduced from[51]).
646 Presence of the morph is indicated in blue, absence in white. **B** Adult *Ptychadena robeensis*
647 presenting the three possible vertebral stripe morphs. From left to right: wide striped, thin
648 striped, unstriped. **C** Histological sections of the dorsal skin within (top) and outside
649 (bottom) the vertebral stripe in a female *Ptychadena erlangeri* (SB548; live photograph on
650 the left). Scale bar = 200µm. Within the stripe (top), the few melanophores (blue arrows)
651 are in a contracted state and do not entirely cover the xanthophores (green arrows), in
652 contrast with the skin outside the stripe (bottom). Outside the stripe, numerous
653 melanosomes (red arrows) are also present in the epidermal layer, creating a very dark
654 coloration.



655

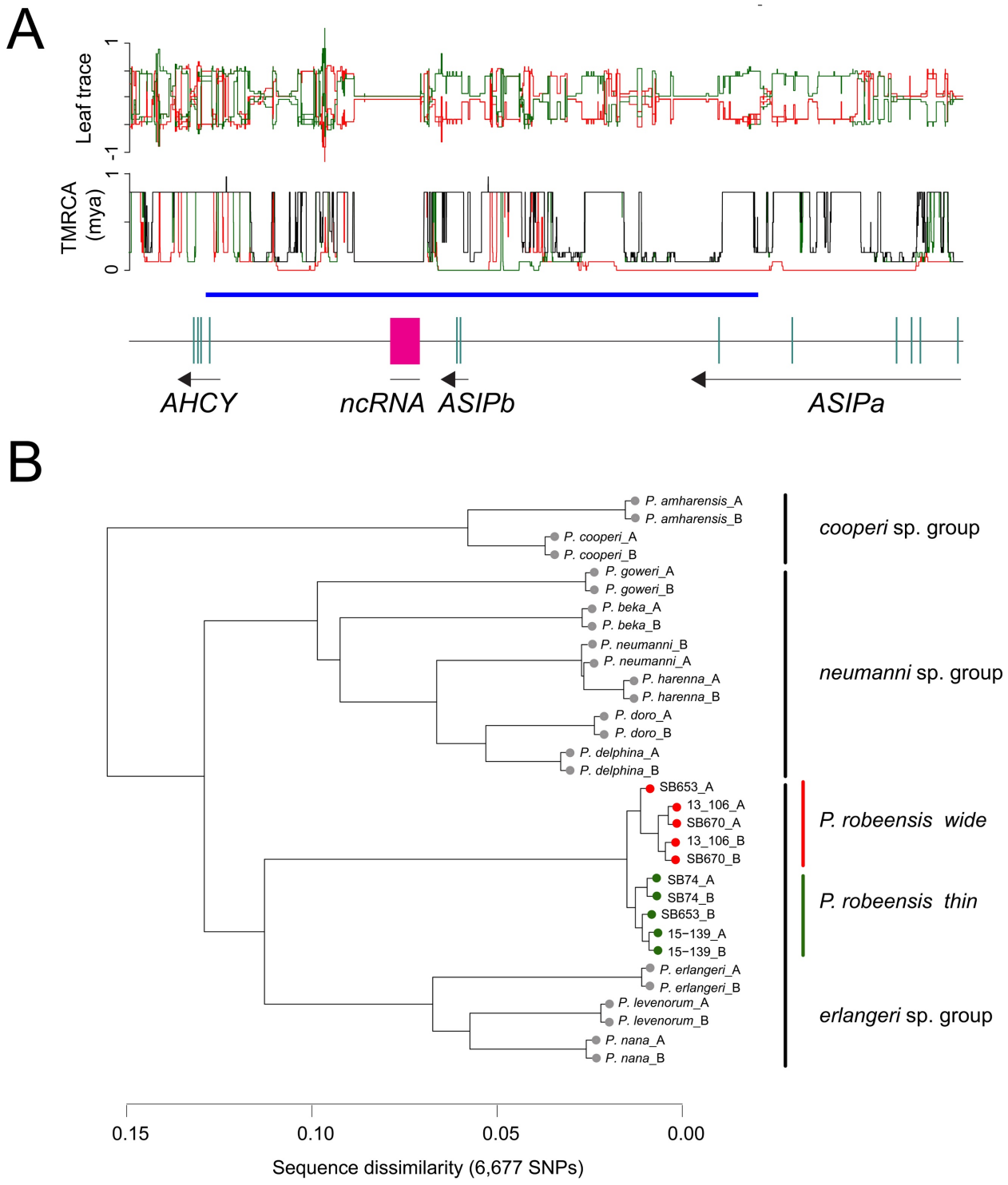
656 **Figure 3. Genes associated with the stripe phenotype in *Ptychadena robeensis.* A**
657 Genome-wide association study (GWAS) reveals a single locus governing the vertebral
658 stripe phenotype on chromosome 11 of *Ptychadena robeensis* (top panel). Significant SNPs
659 (indicated in blue in the bottom panel) are located in between and downstream *ASIP* exons
660 (indicated in light green below) and a non-coding RNA (pink). **B** and **C** Differential gene
661 expression analysis in the skin of adult *Ptychadena robeensis.* 32 genes are differentially
662 expressed between vertebral stripe phenotypes (**B**) and six within versus outside the
663 vertebral stripe in the same individuals (**C**).



664

Figure 4. Recent evolution of the *thin* and *wide* alleles. A Leaf trace plot and TMRCA across the region of interest computed using ancestral recombination graphs analysis. Red and green solid lines indicate *wide* and *thin* haplotypes, respectively. Black solid lines indicate the overall population. The positions of *ASIPa, ASIPb*, *AHCY* and the ncRNA are indicated below and the region containing significant SNPs in the GWAS analysis is indicated by a horizontal blue bar. **B** Dendrograms based on dissimilarity of sequences in the region of interest (40kb region; 6,677 SNPs) in the *Ptychadena neumanni* species complex (8.29X average coverage). Haplotypes are denoted by A or B after the species or sample name. For *P. robeensis*, leaves are color-coded by haplotypes (green for *thin* and red for *wide*). Within *P. robeensis*, haplotypes are grouped by color pattern, while across the *Ptychadena neumanni* complex, they are grouped according to species relatedness and form the three clades previously described[24] (Fig. 2a).

678



679

680

681

682 **Tables**

683

684 **Table 1. Tukey honest significant differences test of morph transition rates between**

685 **habitat pairs.** Values were multiplied by 10^3 for readability. Average (Δ), lower and upper

686 values of the difference between rates based on 100 stochastic maps are given. Habitat

687 categories: A = arboreal, Aq = aquatic, AT = arboreal-terrestrial, T = terrestrial, TAq =

688 terrestrial-aquatic.

	striped → unstriped				unstriped → striped			
	Δ	lower	upper	adj. p-value	**Δ**	lower	upper	adj. p-value
AT - A	-30.92	-32.76	-29.09	< 0.001	3.30	3.19	3.42	< 0.001
T - A	-22.84	-24.68	-21.00	< 0.001	8.59	8.48	8.71	< 0.001
Taq - A	-38.19	-40.03	-36.35	< 0.001	1.58	1.46	1.69	< 0.001
T - AT	8.08	6.24	9.92	< 0.001	5.29	5.17	5.41	< 0.001
TAq - AT	-7.27	-9.11	-5.43	< 0.001	-1.73	-1.84	-1.61	< 0.001
TAq - T	-15.35	-17.19	-13.51	< 0.001	-7.02	-7.13	-6.90	< 0.001

689

690

691

692

693

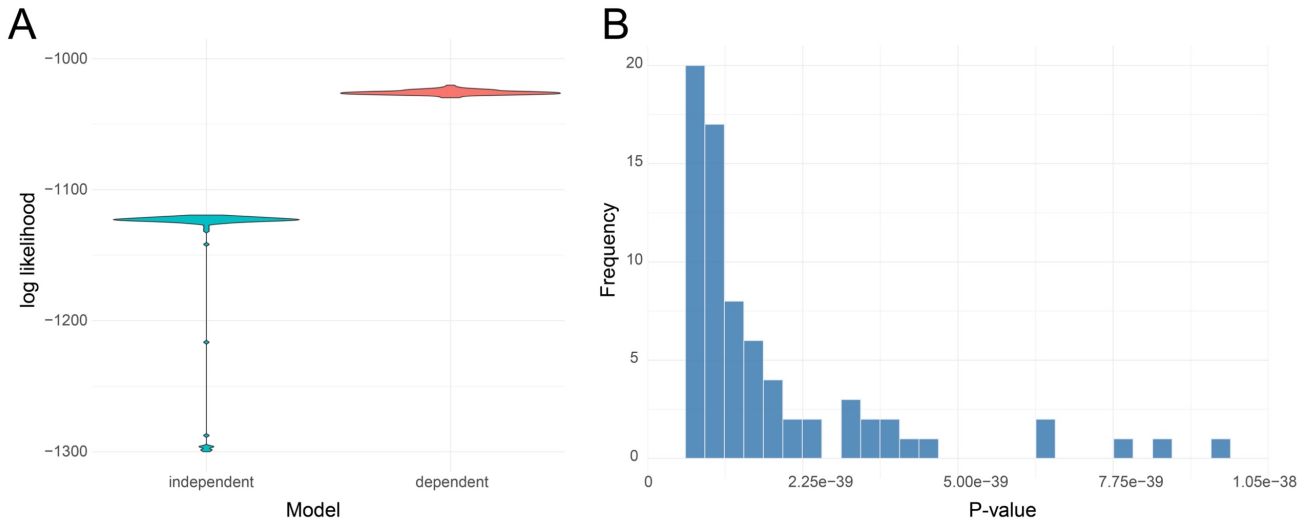
694 **Supplemental information**

695

696 **Table S1. Phenotype and habitat data and references.**

697

698 **Figure S1. Gain and loss rates of the anuran vertebral stripe are habitat-dependent. A**
699 Likelihood of the habitat-dependent and independent models estimated for 100 stochastic
700 maps. **B** P-values of the 100 likelihood ratio tests between dependent and independent. For
701 all 100 stochastic maps, the model with transition rates dependent of the habitat is favored.



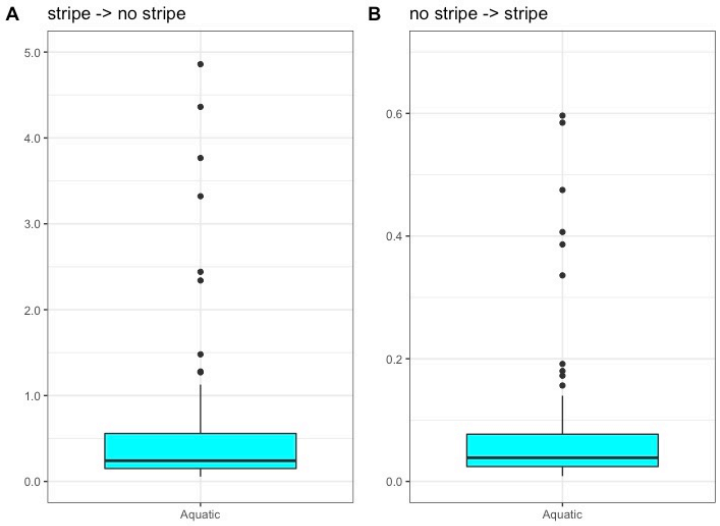
702

703

704

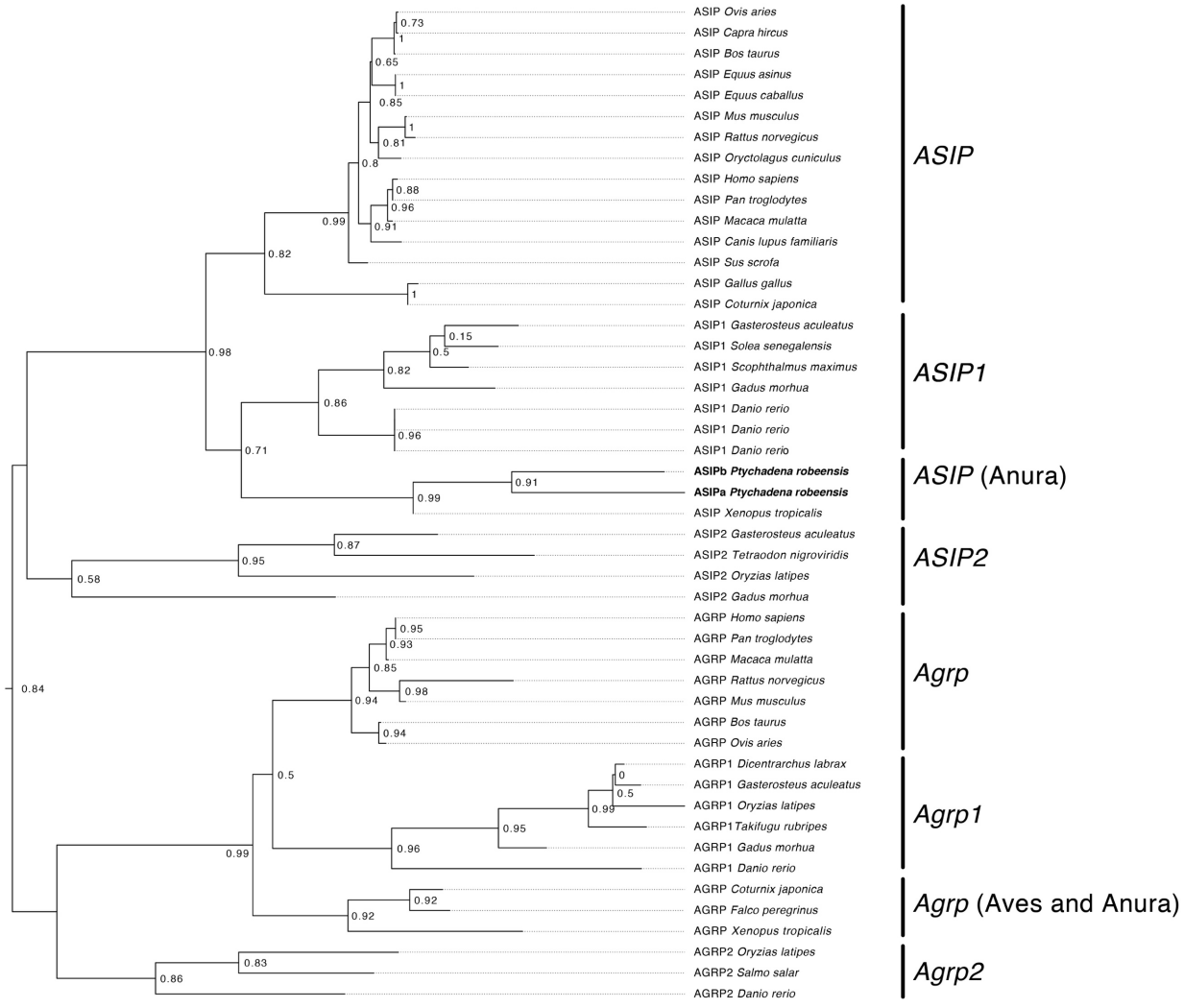
705

706 **Figure S2. Transition rates between striped and unstriped phenotypes for aquatic**
707 **lineages estimated for 100 stochastic maps.** The low number of species in the aquatic
708 category (1.87 % of included taxa, i.e. 49 species) results in important variation in rates
709 among the stochastic maps.
710



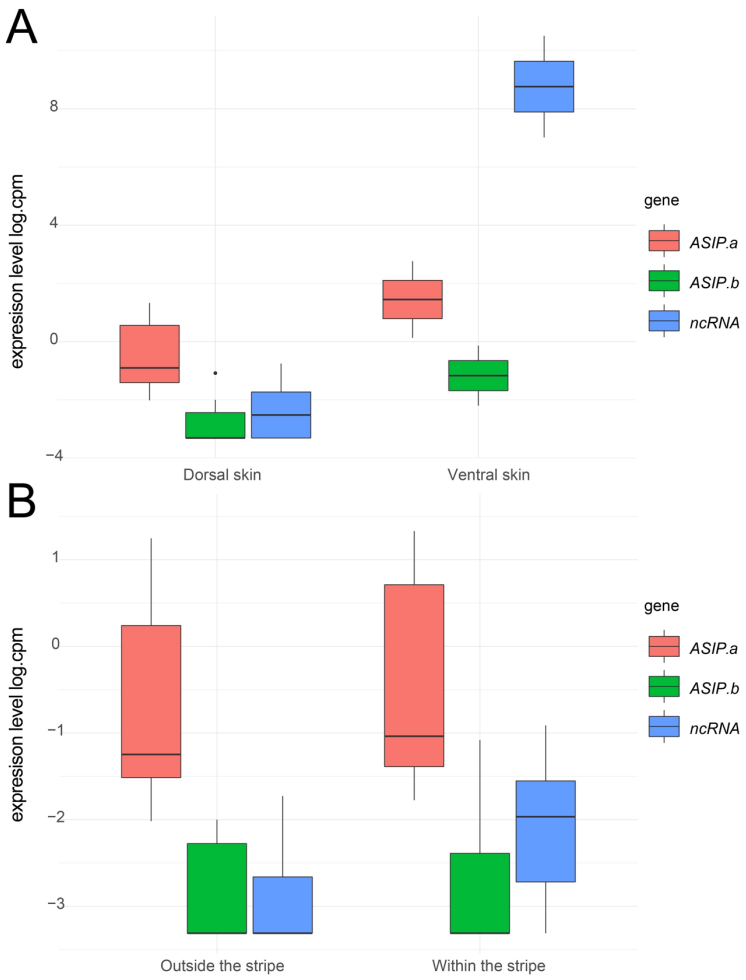
711
712
713

714 **Figure S3. Position of *ASIPa* and *ASIPb* in the evolution of *ASIP*.** Maximum likelihood
715 tree based on a protein alignment using MUSCLE[66]. *ASIPa* and *ASIPb* are grouped with *ASIP*
716 of *Xenopus tropicalis,* excluding the possibility for the two genes to correspond to fish's
717 *ASIP1* and *ASIP2*. Examination of the protein alignment showed that *ASIPa* and *ASIPb* share
718 two exons, resulting from a gene duplication event.
719



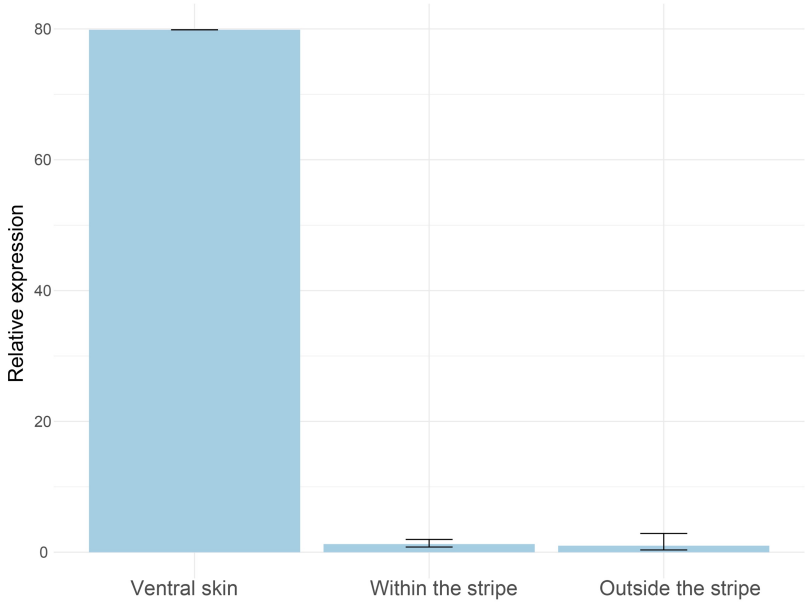
720

721 **Figure S4. Expression levels of ASIP in the skin of *Ptychadena robeensis*.** Normalized
722 expression levels of *ASIPa*, *ASIPb* and the ncRNA from RNAseq data is given in log count per
723 million. **A** ventral skin (n=2) shows a greater number of all three transcripts than dorsal
724 skin (n=19), **B** skin within (n=4) and outside (n=4) the vertebral stripe do not show
725 significant differences in expression level of *ASIPa*, *ASIPb* or the ncRNA.
726



727
728
729
730

731 **Figure S5. Quantitative real-time PCR of *ASIPb* in *Ptychadena robeensis***. Relative
732 expression levels of *ASIPb* in ventral skin (n=1) and dorsal skin within (n=2) and outside
733 (n=4) the vertebral stripe measured by qPCR. Each reaction was triplicated and average CT
734 value for each individual was used. *ASIPb* expression level in ventral skin is 80 times
735 greater than in dorsal skin, while no significant differential expression is detected within
736 versus outside the vertebral stripe.



737
738
739
740