A novel amphibian sex determination candidate gene, evolved by structural variation in the regulatory region between X and Y

Heiner Kuhl[1,*], Wen Hui Tan [2,*], Christophe Klopp[3], Wibke Kleiner[1], Baturalp Koyun[1,4], Mitica Ciorpac[5,6], Romain Feron[7,8], Martin Knytl[9,10], Werner Kloas[1], Manfred Schartl[11,12], Christoph Winkler[2,§], Matthias Stöck[1,§]

[1] Leibniz-Institute of Freshwater Ecology and Inland Fisheries, IGB, Müggelseedamm 301 & 310, 12587 Berlin, Germany,

[2] Department of Biological Sciences and Centre for Bioimaging Sciences, National University of Singapore, 14 Science Drive 4, Block S1A, Level 6, Singapore 117543,

[3] SIGENAE, Plate-forme Bio-informatique Genotoul, Mathématiques et Informatique Appliquées de Toulouse, INRAe, 31326 Castanet-Tolosan, France

[4] Bilkent University, Department of Molecular Biology and Genetics, Faculty of Science, SB Building, Ankara, Turkey 06800,

[5] Danube Delta National Institute for Research and Development, Tulcea 820112, Romania,

[6] Advanced Research and Development Center for Experimental Medicine – CEMEX, "Grigore T. Popa", University of Medicine and Pharmacy, Mihail Kogălniceanu Street 9-13, Iasi 700259, Romania,

[7] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland,

[8] Swiss Institute of Bioinformatics, Lausanne, Switzerland,

[9] Department of Cell Biology, Faculty of Science, Charles University, Viničná 7, Prague, 12843, Czech Republic

[10] Department of Biology, McMaster University, 1280 Main Street West, Hamilton L8S 4K1, Ontario, Canada

[11] Developmental Biochemistry, Biocenter, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany,

[12] The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas, TX, 78666, USA,

[*] These authors contributed equally to the paper.

[§] Co-corresponding authors.

Abstract

Most vertebrates develop distinct females and males, where sex is determined by repeatedly-evolved environmental or genetic triggers. Undifferentiated sex chromosomes and large genomes have caused major knowledge gaps in amphibians. Only a single master gene is known in >8650 species, the *dmrt1*-paralogue (*dm-w*) of female-heterogametic clawed frogs (*Xenopus*; ZW♀/ZZ♂). Combining a chromosome-scale genome of a non-model amphibian, the European green toad, *Bufo*(*tes*) *viridis*, with ddRAD- and whole genome pool-sequencing revealed a novel candidate master gene (*bod1l*), governing a male-heterogametic system (XX♀/XY♂). Targeted sequencing across multiple taxa and a male long-read assembly uncovered structural X/Y-variation in the 5'-regulatory region, where a Y-specific non-coding RNA, only expressed in males, suggests regional enhancer-properties. Developmental transcriptomes and RNA *in-situ* hybridization provide evidence for timely and spatially relevant, sex-specific *bod1l*-gene expression in primordial gonads with coinciding differential *H3K4me*-methylation in pre-granulosa/pre-Sertoli cells, pointing to a novel mechanism of amphibian sex determination.

Introduction

Most vertebrates occur as females and males. Vertebrate sex determination comprises environmental triggers, or simple to complex genetic (genotypic) sex determination – and perhaps combinations thereof – that evolved several times[1,2]. Likewise, the genetic and cellular developmental processes directing undifferentiated gonads towards testes or ovaries are diverse[3].

Genetic sex determination is governed by sex chromosomes that are typically differentiated and often microscopically distinguishable in mammals, birds, and several reptile groups[4]. With several exceptions, some reptilians and most other cold-blooded (poikilothermic) vertebrates, like fish and amphibians, exhibit poorly or undifferentiated (homomorphic) sex chromosomes. Three main evolutionary concepts explain this prevalence: (i) frequent turnovers, overthrowing sex-determining genes by novel autosomal mutations, or even the sex determination system (e.g. XY to WZ)[5], (ii) recurrent X-Y recombination, e.g. by occasional sex reversal to XY-females[6], and (iii) "jumping master genes" with conserved genetic triggers translocating to other chromosomes[7] (incl. refs.). Conserved poorly or non-differentiated sex chromosomes

are known in several clades of poikilothermic vertebrates, e.g. in sturgeons, pikes, and skinks[8-10].

More than 95% of amphibians possess undifferentiated sex chromosomes with either female (ZZ/ZW) or male (XY/XX) heterogamety[11,12]. Few examples of cytogenetically differentiated sex chromosomes[12], a female W0/00 male sex determination system[13], and several amphibians with multiple sex chromosomes have been identified in both laboratory strains[14] and natural populations[15]. Notably, to date, in >8,650 species of amphibians[16], the widespread combination of poorly or undifferentiated sex chromosomes with large genome-sizes (see below) have caused a basic knowledge gap with only a single though very well-characterized amphibian master sex-determination gene known in a model species, the African clawed frog (*Xenopus laevis*[11,17]) and some related species[18]. *Dm-w* arose after (and perhaps in response to) tetraploidization, and has been lost secondarily in some tetraploid species[19].

True toads, Bufonidae, share large genomes, conserved karyotypes[20] and undifferentiated sex chromosomes[21]. Typical of most bufonid toads is also Bidder's organ that develops in males and - in many species - also in females during early gonadogenesis from the anterior parts of the primordial gonads, considered to possibly present a rudimentary ovary[22]. Using sex-reversal experiments, castrating males and regrowing functional ovaries from Bidder's organs, early on, Ponse[23] inferred female heterogamety (ZW) in common toads (*B. bufo*). This was also cytogenetically deduced in related *B. gargarizans*[24], and cane toads (*Rhinella marina*[25]). With few exceptions, a meta-analysis of sex-differential F1-hybrid fitness of interspecific crosses in bufonids likewise supported widespread female heterogamety[26].

This situation contrasts with that in our target species, Palearctic green toads of the *Bufo*(*tes*) *viridis* sub-group. While their sex chromosomes are also undifferentiated[21,27] (refs. therein), sex-reversal experiments in '*B. viridis*' from Asia Minor[28] suggested an XY-system. Molecular evidence for male heterogamety resulted from sibship analyses using microsatellites and nuclear sequence markers[29-31], identifying the largest linkage group (LG1), homologous to autosomal LG1 of *X. tropicalis* and harboring *dmrt1*, as sex-linked in several diploid green toad species. All feature strongly-reduced X/Y-recombination but occasional X/Y-exchange is suggested at evolutionary time scales[29]. However, phylogenetic analysis revealed X- and Y-alleles of *dmrt1* to cluster by species and not by gametologue, suggesting X/Y-recombination involving *dmrt1*,

100 which rejected its role as master sex determiner[31]. Details of sex determination have
101 not been examined in allopolyploids green toads[32].
102 In mammals (6,649 species[33]), genome assemblies are available for about one third
103 (2740[34]). Although there are more species described in amphibians (>8650[16]), owing
104 to large genome sizes and excessive repetitive regions, to date only >80 amphibian
105 genomes have been assembled, including 56 out of 7627 Anura, 19 out of 805 Urodela
106 (Caudata) and 9 out of 221 Gymnophiona[16,34]. However, profiting from long-read- and
107 3D-sequencing technologies, amphibian genomics is on the rise.
108 Here, we provide a chromosome-scale genome of a non-model amphibian, the
109 European green toad, *Bufo*(*tes*) *viridis*, and identify a novel candidate master sex
110 determining gene - present in multiple related green toad taxa - that evolved by
111 structural variation, involving a Y-specific ncRNA, in the regulatory region between X
112 and Y chromosome.
113
114 **Results**
115 **Chromosome-scale genome assembly of *B. viridis***
116 We obtained a 3.8 Gbp chromosomal-level genome assembly using Oxford nanopore
117 (ONT) long-read (N50 = 11.6 kbp; sequencing coverage 21x) and Illumina short-read
118 data (2x150 bp; sequence coverage 47x) for the European green toad (Fig.1a,b). N50
119 contig size of the final assembly was 1.9 Mbp. Eleven chromosomes (scaffold N50 =
120 470 Mbp; consisting of 94% of the contigs total length), consistent with the karyotype
121 of *B. viridis* (Fig.1c), were constructed using chromosome conformation capture (Hi-
122 C) short reads (318 Gbp). Many unplaced contigs contain a known long satellite DNA
123 family[35]. Comparing the chromosomes to the *B. bufo* assembly revealed genome-wide
124 synteny and collinearity with few inversions, despite significant differences in assembly
125 size (VGP's aBufBuf1.1 = 5.0 Gbp; Fig.1b). *De novo* analysis revealed 68.5% repeat
126 content in *B. viridis*. Using published data from related species and RNAseq data, we
127 annotated 21,605 protein coding genes in the genome. BUSCO (Geneset
128 vertebrata_ODB9)-scoring of protein predictions revealed 96.4% complete coding
129 sequences and a low missing gene fraction of 1.7% (C:96.4% [S:94.2%, D:2.2%],
130 F:1.9%, M:1.7%, n:2586).
131

132 **ddRADseq in *B. viridis* genetic families**

133 We first attempted to identify the sex determination region by double digest restriction-
134 site associated DNA sequencing (ddRADseq). In three genetic families of European
135 green toads, we analyzed a total of 443,378 markers, present in at least one individual
136 with a depth higher than 4. Of these, 96 were significantly associated with male sex (p
137 < 0.05; Bonferroni correction), but none was found in all males (Supplementary Figure
138 1), probably due to high stochasticity of random sequencing of large genome fractions.
139 When mapped to the *B. viridis* genome, 64 of these 96 markers were located along
140 the entire chromosome 1 (scf1) (Supplementary Figure 2), one on chromosome 7
141 (scf7), five on unplaced scaffolds, and 26 were not aligned uniquely or with low quality
142 (< 20). Overall, this strongly supports a male (XY) genetic sex determination system
143 on chromosome 1, however, with a very small sex-specific region that is not fully
144 covered in each individual. While this approach powerfully corroborated previous
145 microsatellite-based results[29], it did not provide a sex-diagnostic marker.

146

147 **Pool-seq whole genome sequencing yields a cross-amplifying marker in**
148 **multiple taxa**

149 To narrow down the sex determining region on scaffold 1, we used whole genome
150 pool sequencing (pool-sex), involving 22 female and 22 male phenotypically-sexed
151 adult green toads from multiple European green toad populations (Supplementary
152 Table 1). Mapping pool-sex short reads to the reference, genome-wide revealed a
153 single peak in a region on scaffold 1 (scf1), where the number of clustered
154 heterozygous SNPs (N = 105) in males surpassed about nine times the maximum
155 number of clustered heterozygous SNPs (N = 12) in females (Fig.1d, Supplementary
156 Table 2; Supplementary Figure 3).

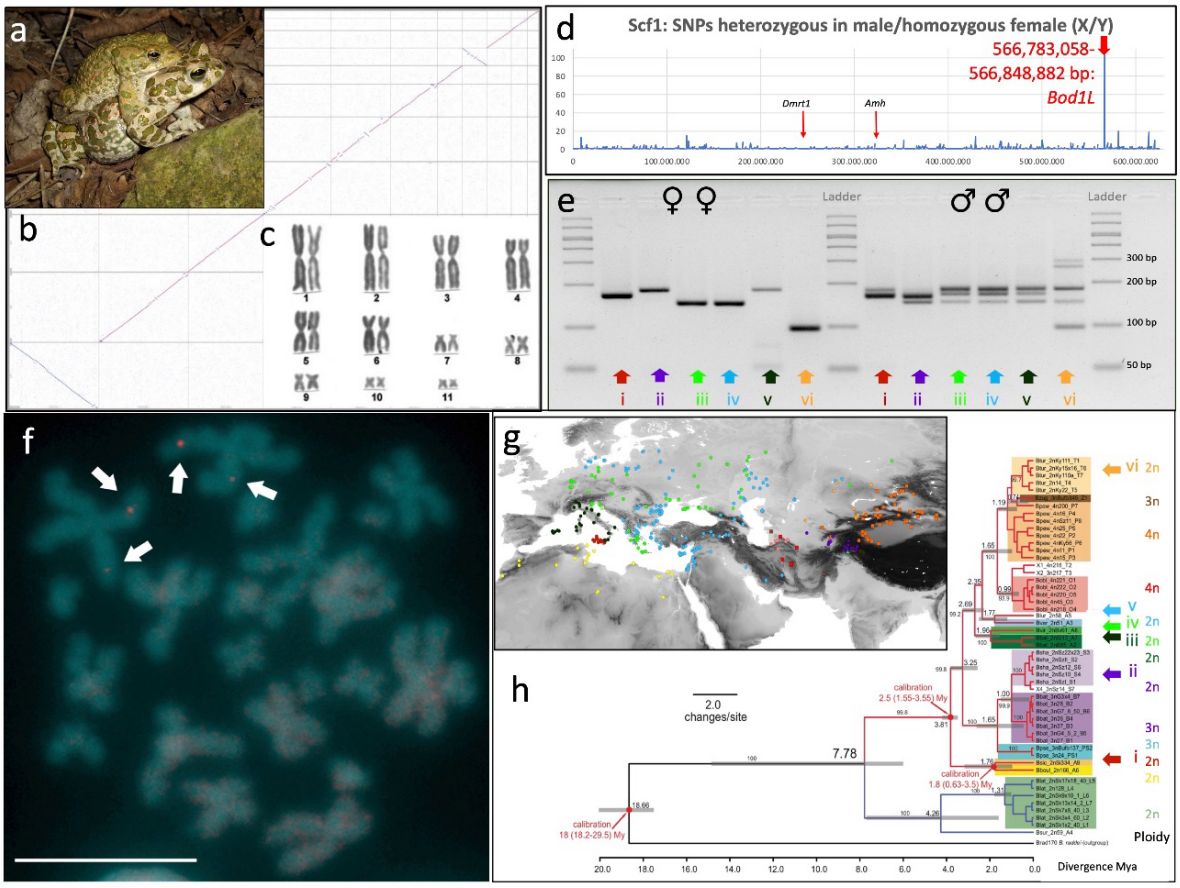


Fig. 1| The chromosome-scale reference-genome of *Bufo*(*tes*) *viridis* and the new sex-diagnostic marker. a, Breeding pair of *B. viridis* in amplexus (© M. Stöck). **b,** *B. viridis*-assembly compared to *B. bufo* assembly by the Vertebrate Genome Project (VGP, Genbank accession number GCA_905171765.1) showing high synteny of both bufonid genomes across all 11 chromosomes. **c,** Giemsa-stained karyotype of *B. viridis*, changed from[27]. **d,** Mapping the pool-seq short reads to the *B. viridis* reference: genome-wide a single region (scf:566,783,058-566,848,882), where the number of heterozygous SNPs (y-axis) on scaffold 1 (N = 105) in males surpassed about nine times the maximum number of genomic positions (N = 12) where females exhibit heterozygous SNPs, provides strong evidence for male heterogamety (XX/XY); for comparisons with a ZZ/ZW system see Supplementary Fig. 2. **e,** Overview of sex-specific PCR-products of primers BvXY1 in females (♀♀, left) and males (♂♂, right) of six (i-vi) nominal diploid green toad taxa, electrophoresis result using 4 µL of PCR-products on a 4% high-resolution Sieve-agarose gel run for 60 min on 70 V; Palearctic green toad taxa shown are: (i) *siculus*, (ii) *shaartusiensis*, (iii) *viridis*, (iv) *variabiilis*, (v) *balearicus*, (vi) *turanensis*. **f,** FISH-TSA of a metaphase spread of a *B. viridis* male with red *bod1l*-signals on chromosomes 1 on their short (p) arm (arrows), chromosomes were counterstained with DAPI (blue-green), scale bar in lower left corner represents 10 µm. **g,** Range map of Palearctic green toad taxa, numbers (i-vi) and color-coding as in (e) and (h), diploid (2n) taxa are indicated by circles, triploids (3n) by triangles and tetraploids (4n) by squares. **h,** Dated mitochondrial phylogeny of Palearctic green toads (adapted from[32], with the phylogenetic positions of the diploid taxa, in which BvXY1 is sex-diagnostic, indicated

178 by color-coded arrows as in (**e**); the three ploidy levels are provided to the right of the leaves, note that
179 the sex marker is seemingly conserved only in the diploid green toads; all polyploids are of hybrid
180 origin[32].
181
182 This strongly supports a male-heterogametic (XY) system, involving a short 66 kb
183 region (scf1:566,783,058-566,848,882 bp). It encompasses one third of the 5'-
184 upstream-region and two thirds of the gene *bod1l* (bufVir1.4895.2, biorientation of
185 chromosomes in cell division 1 like 1). Transcriptomes and cDNA-based PCR-
186 products of males of *B. viridis* exhibited consistent heterozygosity of exons 7 to 10
187 (scf1:566,831,904-566,833,300), while females were always homozygous.
188 The heterozygous 5'-region exhibited several INDELs between X and Y haplotypes,
189 enabling primer design. In females, BvXY1 amplified a single PCR-product - visible as
190 one band on the gel, and two to three bands in males, widely consistent with a an
191 XX♀/XY♂-system, but suggesting multiple Y-specific signals in males. Of note,
192 BvXY1-primers amplified not only sex-specific products in individual *B. viridis* (sensu
193 lato) males and females of the pool-sex-approach but also in multiple phenotypically-
194 sexed males and females of this and four additional taxa (*B. balearicus*, *B. siculus, B.*
195 *turanensis, B. shaartusiensis*), comprising a monophyletic radiation, probably >3 Mio.
196 years old[32] (Fig. 1e-h). The sex-diagnostic PCR-marker BvXY1 amplified a single
197 specific 100-200 bp band in females of six green toad taxa, but 2-3 bands in their
198 males, one in females' size, presenting the X-copy. The additional male bands with
199 the initial BvXY1-primers yielded identical or similar sequences indicating duplicates
200 on the Y-homologue.
201
202 **Position of the sex-specific marker and FISH on the karyotype dismiss**
203 **duplication with translocation**
204 New sex determining genes evolve by allelic diversification or from gene
205 duplications[36]. To elucidate the situation in *B. viridis*, we examined synteny and
206 localization of the candidate gene. *Bod1l* lies between the genes *Cpeb2* and *nkx3-2*
207 (NK3 homeobox 2), on the long arm of scaffold 1, an evolutionarily conserved gene
208 order likewise found in *B. bufo*, *Rana temporaria* and *Xenopus tropicalis*, suggesting
209 synteny throughout anurans.
210 Fluorescence *in-situ* hybridization with tyramide signal amplification (FISH-TSA) on
211 metaphase chromosomes of a male *B. viridis* revealed two signals, one on each

212 homologue, for a *bod1l*-specific probe at a position on the distal part of the short (p)
213 arm of chromosome 1 (Fig. 1f). Two (and not three) signals suggest that this gene
214 exists in two copies in XY-males, implying that both heterozygous sequences evolved
215 by structural diversification[36] rather than translocation of a male-specific duplication,
216 as shown for *dm-w* of female *Xenopus laevis*[17]. This result is further supported by the
217 pool-seq and long-read male coverage (see below).
218

219 **Targeted sequencing analyses in multiple related species sharing the genetic**
220 **sex marker**

221 If the hypothetically sex-determining SNPs in coding *bod1l*-regions arose in a common
222 ancestor, we expected the five diploid taxa, sharing BvXY1 in 5'-*bod1l*-regions (Fig.
223 1d), to share at least some exonic, possibly functional SNPs. To test this, we applied
224 targeted enrichment to multiple females and males of five taxa, using an AmpliSeq
225 custom DNA panel on 80 kb (93%) of the *bod1l*-X-copy, including its 5'-region with
226 BvXY1, and analyzed 8183 SNPs. Of note, none of the *bod1l* coding SNPs, present
227 in *B. viridis* (Supplementary File 2), was conserved across all five toad taxa
228 (Supplementary Files 3, 4, Supplementary Figure 4a), suggesting that this coding
229 region (cds) does not contain the mechanistic key to green toad sex determination.
230 Instead, the targeted enrichment approach narrowed down the sex-specific region to
231 the 5'-region of *bod1l*, containing BvXY1 (Supplementary Figure 4b).
232

233 **Y-chromosomal structure of *B. viridis* from long-read ONT whole genome**
234 **sequencing**

235 After first working with the female reference genome, X-haplotype and *bod1l*-coding
236 region, we then moved the focus back to the 5'-end of *bod1l* and the BvXY1-vicinity
237 on the Y-haplotype. To elucidate this Y-chromosomal structure, we reconstructed this
238 region from ca. 30x whole genome long-reads of a single male *B. viridis* and aligned
239 it to the X-copy of the *bod1l*-region. Female-to-male pool-seq-coverage of X- and Y-
240 haplotypes yielded the expected diploid female and haploid male coverage
241 (Supplementary Figure 5 and 6).
242 In the *bod1l*-region, the confidently assembled Y-haplotype specific sequence
243 comprises about 55 kb with some greater-than-zero female pool-seq-coverage,
244 explicable by repeat-enriched fragments, not assignable to X or Y (Supplementary
245 Figure 7). The male long-read assembly revealed several structural changes between

246 X- and Y-haplotypes in the 5'-*bod1l*-region. The *B. viridis*-Y contains large insertions,
247 mainly stemming from repetitive elements, and a non-coding RNA (ncRNA, see
248 below). None of the so-called 'usual suspects', genes that often acquire sex-
249 determination function in vertebrates[1], was detected in this region.
250

251 **Gene expression during larval development in sibling *B. viridis* tadpoles from**
252 **multiple European localities**

253 Pool-sequencing of genomic DNA in *B. viridis* had revealed consistent heterozygosity
254 in males for *bod1l*-exons 7 to 10. Accordingly, RNAseq revealed sex-differential
255 expression (scf1:566,831,904-566,833,300), where heterozygous *B. viridis* males
256 showed 1 SNP in exon 7, and 14 SNPs in exon 10, while females were always
257 homozygous (confirmed by cDNA-based PCR-products). To test whether *bod1l*-
258 products are differentially composed after transcription of X- and Y-copy, we analyzed
259 splice sites in developmental transcriptomes, including the sex determination period.
260 The absence of significant sex-specific differences excluded alternative splicing.
261 However, this analysis allowed manual improvement of the *bod1l*-gene model.
262 To study the expression of the candidate master sex determination gene (*bod1l*) and
263 ca. 80 other genes (Supplementary File 5) playing roles in sexual development and/or
264 sex determination in other vertebrates ('usual suspects'[1]), we then examined larval
265 transcriptomes of six sibling groups, in total 54 individual toads from six locations,
266 ranging from day 10 after fertilization until Gosner[37] stage 46 (the last about 6 weeks
267 after metamorphosis) as well as from two adult gonads.
268 Since gonadal differentiation is histologically detectable only at or after metamorphosis
269 completion (>Gosner 46[31,37]), the critical period of genetic sex determination lies
270 earlier, within our sampling period. Until stage 5a, overall *bod1l*-expression varies in
271 *B. viridis* between males and females (Gosner 34-38; Fig. 2). From stages 4 to 5b,
272 other genes, specific for male sexual development, like *sox9*, *sox10*, *sox17*, and *sf1*,
273 show higher expression in males than in females. Other such genes change their
274 expression distinctly after stage 4, like *ck18* (bufVir1.18618.1) or *11ßHD*
275 (bufVir1.16559.1) or - in our series - were detected only in Gosner 46 and subadult
276 testes (*dmrt1, amh*). Female-specific genes also changed their expression around
277 stage 4; *wnt4*, *ihh*, *foxL2*, *ptch1*, *gli3*, *fshr* from stage 5a or 5b; however, their
278 expression was stronger in males than females. Expression of *cyp19a1a*, *cyp17a1*
279 and *gdf9* was detectable only in stage 5b (Gosner 46) and adult ovaries. *Bod1l*-

expression differences between females and males (Fig. 2) lack statistical significance but occur at the right developmental period (2 to 5a), as required for a sex-determining gene. Early male transcriptomes (stages 1 to 5a) contain also the bufonid-specific Bidder's organ, considered similar to a rudimentary ovary, that develops adjacent to the anterior primordial gonads[22]. Published qPCR-data suggested that gene expression in adult Bidder's organs differs from that of testes and ovaries but resembles the latter, consistent with classical, mid-20th-century male-to-female sex-reversal experiments, when Bidder's organs grew and resumed oogenesis after the removal of testes[23] (reviewed[22]). In green toads, Bidder's organ develops in males but barely in females[22]. Given that the "ovary-like" primordial Bidder's organ is contributing tissue to our RNAseq-analysis of male tadpoles until stage 5a, we assume a "female-bias" in males' transcriptomes (stages 1 to 5a). In conclusion and in line with the AmpliSeq-data from multiple species, only sharing the BvXY1-marker across taxa, we consider these slight quantitative differences between males and females as sufficient to trigger green toad sex determination. Of note, there is a trend of greater Y-haplotypic expression of *bod1l* between stages 4 to 5a (coinciding with an increase of the ncRNA-Y expression, next paragraph) and a distinctly greater additive XY-expression in males than XX-expression in females. *Bod1l*-Y also shows the highest expression in testis compared to other male organs (heart, brain, liver).

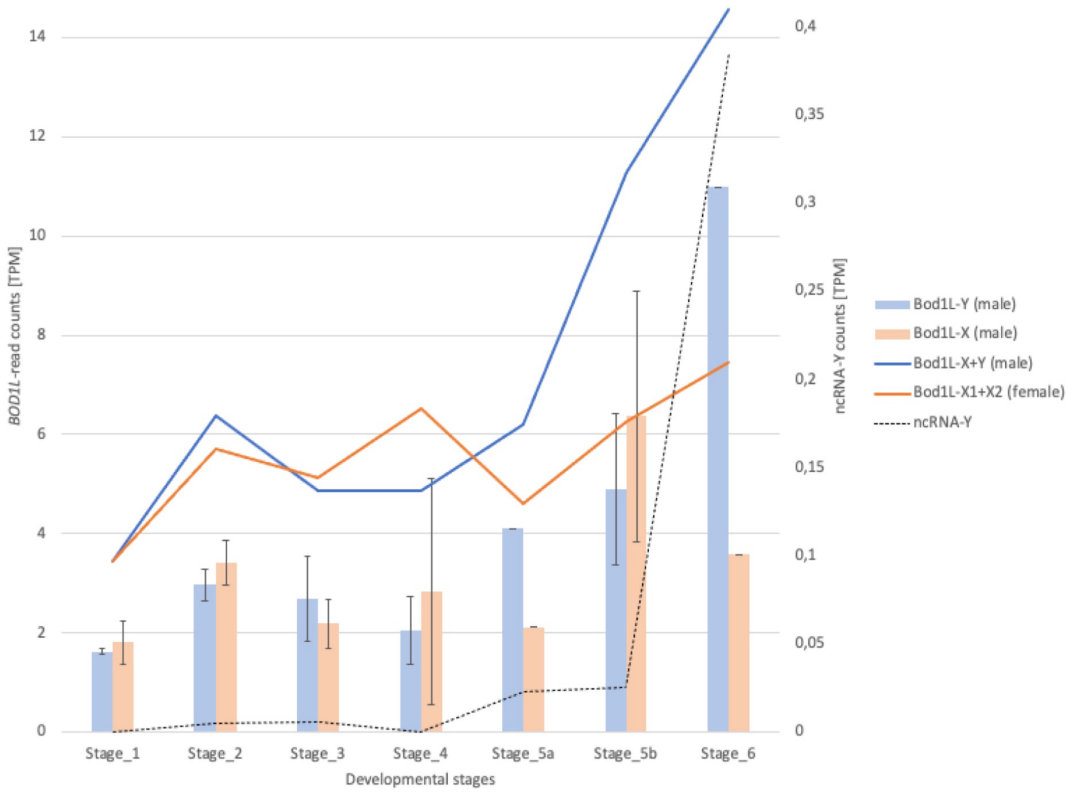
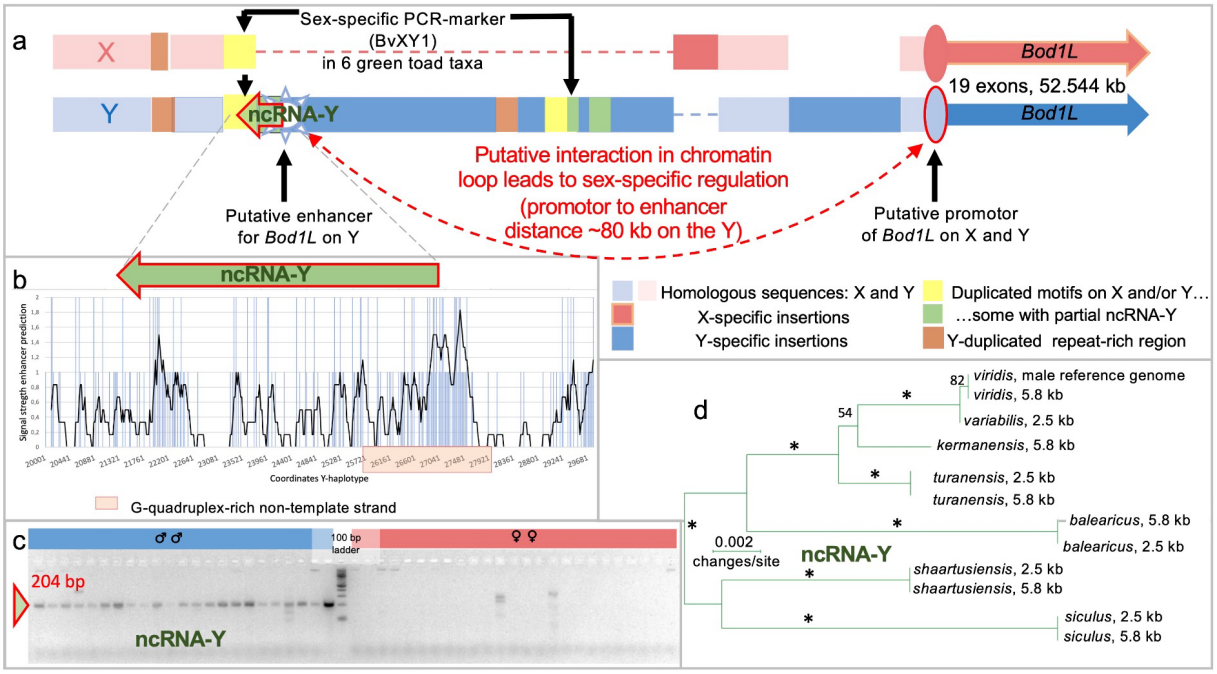


Fig. 2 | Haplotype-specific and trend of XX- or XY- additive expression of *bod1l* and the ncRNA-Y in females and males in seven developmental stages. Expression in females *vs*. males. Stages: 1: 10 days after fertilization, 2: 15 days after fertilization, 3: 18 days after fertilization; 4: 34 days after fertilization/Gosner[37] stages 34-38, 5a: Gosner stages 43-44, beginning of metamorphosis, 5b: Gosner stage 46, six weeks after metamorphosis, 6: subadult male and adult female. Numbers of available tadpole transcriptomes varied by stage; TPM: Transcript reads Per Million RNAseq reads.

A non-coding RNA (ncRNA) is Y-specifically expressed upstream of *bod1l*

When further comparing RNAseq (mapped to Y- or X-haplotypes) from ovaries *vs.* testes (in tadpoles plus adjacent Bidder's organ), we discovered a long non-coding RNA ("ncRNA-Y", ca. 1.2 kb, 4 exons, Fig. 3) with its exons 1 and 2 being Y-specific, and expressed exclusively in males, as confirmed by PCR from cDNAs (Fig. 3c, Supplementary Fig. 8). The ncRNA-Y harbors the initially detected gDNA-marker (BvXY1) in intron 3. We hypothesize that this male-specifically expressed ncRNA-Y hints at an enhancer element, missing on the X, which regulates male *bod1l*-expression and simultaneously transcribes the ncRNA-Y in a one-dimensional (1D) manner from the opposite strand[38]. With ca. 80 kb, the putative enhancer sits in a reasonable distance to the *bod1l*-promoter (Fig. 3a,b)[39]. We also tested 200 bp-"sliding-windows" of the ncRNA-region for their *in-silico* potential as enhancers of

319 *bod1l* (Supplementary Table 3; Supplementary Fig. 9). The most stringent threshold
320 of 0.9 still predicted larger parts of the Y-specific region to function as strong
321 enhancers, intensified by numerous G-quadruplexes (>25 alone in intron 1 of ncRNA-
322 Y), known to accumulate at promoters and enhancers[40]. As shown by PCR using
323 cDNA, this long ncRNA-Y is also expressed in skin, liver, brain and heart of males
324 after metamorphosis as shown using cDNA, and exhibits simultaneous expression of
325 the Y-haplotype of *bod1l* with the strongest signals in testis (Fig. 2). Due to its low
326 expression level, it is only detectable in high-coverage transcriptomes of the primordial
327 male gonads and adult testes but not in heart, brain and liver high-coverage
328 transcriptomes. It comprises 5.8 kb on the gDNA level and can be male-specifically
329 PCR-amplified in all those six taxa containing the BvXY1-marker and suggests a
330 common ancestral feature (Fig. 3d).



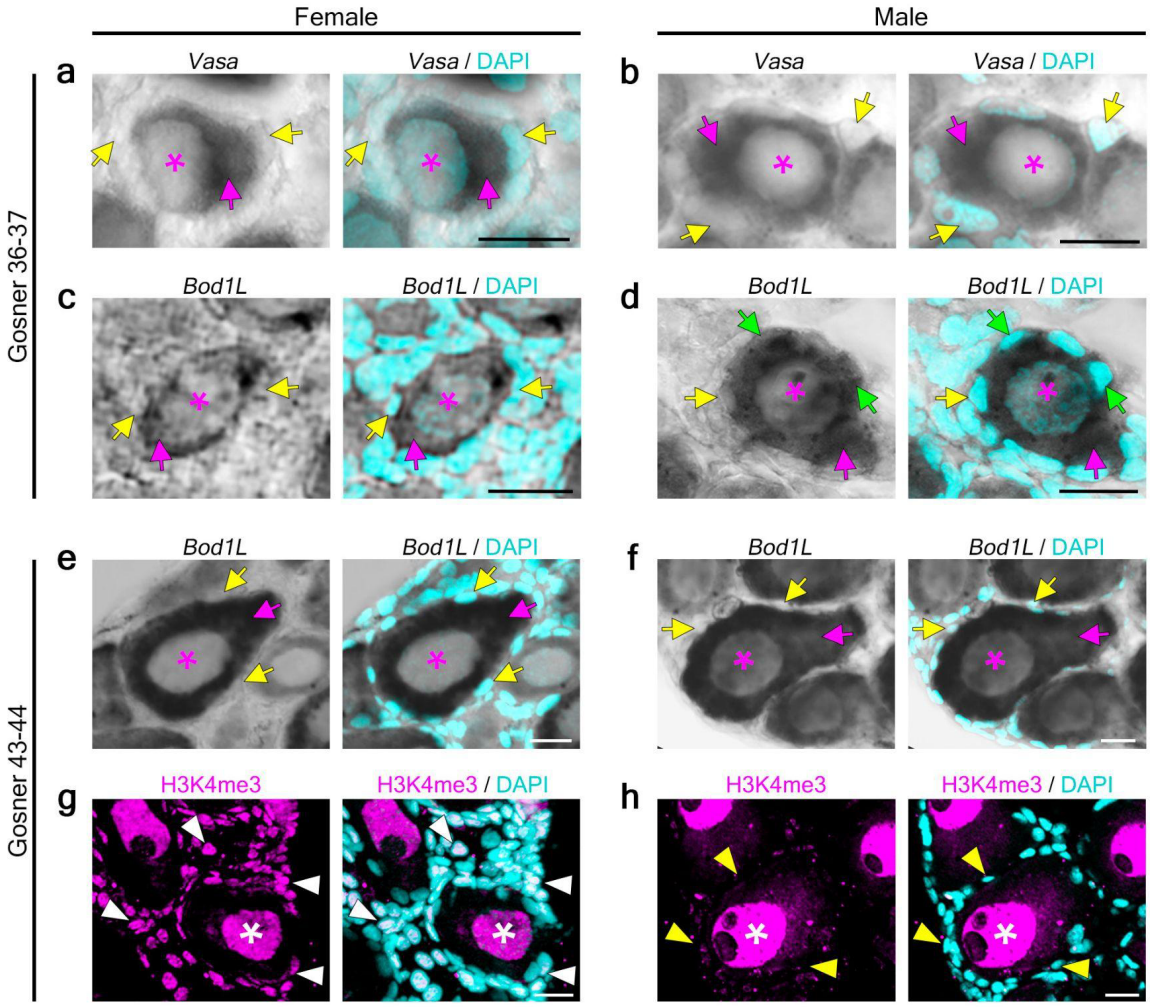
331

332 **Fig. 3 | Structure of the X and Y-haplotype of the *bod1l*-region of *Bufo*(*tes*) *viridis* with focus on**
333 **the 5'-end of the candidate master sex-determining gene and the Y-specific ncRNA. a,** X-
334 chromosomal (above, pink) and Y-chromosomal haplotype (below, blue) with color-coded homologous
335 or X- and Y-specific INDELs, and the Y-specific long (ca. 5.8 kb) non-coding RNA (ncRNA-Y), flanking
336 a putative enhancer at the 5'-end of the gene, ca. 80 kilobases before its promoter. **b,** Graph of the
337 signal of the AI-based software *i-enhancer*, reaching from 0 to 2, predicting strong enhancer properties
338 of the Y-specific 5'-end of *bod1l*, including a G-quadruplex-rich region on the non-template strand,
339 overlapping with the ncRNA-Y. **c,** Y- and thus male-specific expression of the ncRNA-Y (204 bp PCR-
340 product from cDNA) in larval green toads, from 10 days after hatching until early metamorphosis; in
341 females, only non-specific products occurred, since X-copies of *bod1l* do not contain this ncRNA (for
342 details and positive controls of females: Supplementary Figure 8). **d,** Unrooted Maximum-Likelihood-

343 based phylogenetic tree, showing that the ncRNA-Y evolved in a common ancestor of diploid Palearctic
344 green toads sharing this XY-system, with *siculus* and *viridis* spanning the most divergent lineages; two
345 PCR-products (2.5 and 5.8 kb), directly amplified by independent PCRs from six related diploid green
346 toad taxa, clustered by Y-gametologue and thus species (* symbolized 100% bootstrap support; only
347 support values >50% are shown).
348

349 **Spatially different gene expression in the developing gonads**

350 For a master sex-determination gene, not only the accurate timing of its expression
351 but also its organ- and cell-specific spatial expression is essential. We provide
352 evidence by RNA *in-situ* hybridization in histological sections of gonads in Gosner
353 stage 36-37, where *bod1l* shows sex-specific differential expression. In female
354 gonads, *bod1l*-expression was evident in *vasa*-positive primordial germ cells
355 (developing oogonia) but absent from surrounding somatic pre-granulosa cells (Fig.
356 4a,c, Supplementary Figure 10). In contrast, *vasa*-positive male primordial germ cells
357 (developing spermatogonia) as well as a majority of adjacent somatic pre-Sertoli cells
358 exhibited *bod1l*-expression (Fig. 4b,d, green arrows). In Gosner stages 43-44, at the
359 beginning of metamorphosis, *bod1l*-expression in females remained similar to Gosner
360 36-37. In males, however, *bod1l*-expression was no longer detectable in somatic pre-
361 Sertoli cells (Fig. 4e,f; Supplementary Figure 10). This suggests a dynamic regulation
362 of *bod1l* in early male and female gonads, with expression in somatic cells exclusively
363 in male but not female gonads at the relevant stages (prior stage 4), when sex
364 determination occurs.
365 It has been shown that *Bod1l* interacts with *Setd1a* to facilitate tri-methylation of
366 histone H3 at lysine 4 (H3K4me3[41]), with similarities in eukaryotes from yeast to
367 mammals[42]. To examine its status in male and female gonads, we performed
368 H3K4me3-immunostaining on Gosner 43-44 gonads (Fig. 4g,h). In females, gonads
369 exhibited distinct H3K4me3-signals in nuclei of both germ and somatic pre-granulosa
370 cells (Fig. 4g, white asterisks and arrowheads, respectively; Supplementary Figure
371 10). In contrast, H3K4me3 was strongly detected in male germ cell nuclei (Fig. 4h,
372 white asterisks) but hardly detectable or absent in nuclei of somatic pre-Sertoli cells
373 (Fig. 4h, yellow arrowheads). In Gosner stage 46, six weeks after metamorphosis,
374 *bod1l* was expressed (RNAseq, PCR of cDNA) in both sexes in several somatic
375 tissues (brain, heart, liver, intestine) as well as in gonads.
376



i

		Expression in somatic cells		Expression in germ cells	
		Male	Female	Male	Female
Gosner 36-37	*Bod1L*	+	-	+	+
Gosner 43-44	*Bod1L*	-	-	+	+
	H3K4me3	low	high	high	high

Fig. 4 | *Vasa*- and *bod1l*-expression and H3K4me3-methylation in primordial gonads of *B. viridis*, i.e. in the female early ovary, and in male Bidder's organ with testes (Gosner stages 36-37) as well as in early metamorphosis (Gosner stages 43-44). a-f, *Vasa* and *bod1l* expression shown by RNA *in-situ* hybridization (ISH) on 20 µm transverse cryosections of female (a, c, e, g) and male (b, d, f, h) gonads at developmental stages Gosner 36-37 (a-d) and 43-44 (e, f), the latter presenting early metamorphic stages. DAPI was used to stain nuclei. Magenta asterisks indicate nuclei of germ cells; magenta arrows: cytoplasm of germ cells; yellow arrows: nuclei of somatic cells not stained by RNA ISH; green arrows: nuclei of somatic cells stained by RNA ISH. **g,h,** H3K4me3 and DAPI staining on 20 µm transverse cryosections of female (g) and male (h) green toad Bidder's organ and gonads, Gosner stage 43-44. White asterisks indicate H3K4me3-positive germ cell nuclei; white arrowheads, H3K4me3-positive somatic cell nuclei; yellow arrowheads, H3K4me3-negative somatic cell nuclei.

389 Scale bars: 20 µm. **i,** Summary of *bod1l*-expression and H3K4me3-methylation in somatic cells and
390 germ cells in Gosner 36-37 and Gosner 43-44 male and female gonads.
391

392 **XY-phylogeny in six Eurasian diploid taxa from targeted sequencing**

393 For a master sex-determining locus, evolved from a common ancestor, non-
394 recombining loci are expected to cluster by gametologue (i.e. X- or Y-specific
395 haplotypes) and not by species. Analyzing X- and Y-specific sequences obtained from
396 the AmpliSeq custom DNA panel and amplicon sequencing of ncRNA-Y of six diploid
397 green toad taxa, we show that in the 2,671 bp-region (scf1:566,788,797-566,791,468),
398 containing the sex-specific BvXY1-marker, sequences cluster by gametologue (Xs *vs*.
399 Ys, Supplementary Figure 4c), while in the similar-sized adjacent upstream region
400 (scf1:566,786,116-566,788,796), sequences clustered by taxon/species, suggesting
401 occasional recombination (Supplementary Figure 4d). This provided evidence of both
402 the stop of recombination since the sequence arose in a common ancestor and the
403 apparent functional relevance of this region, since these species also do not share
404 coding sequences.
405

406 **Discussion**

407 Our data in *B. viridis* and five related diploid green toad taxa are in accordance with a
408 gene that shows structural changes in its 5'-regulatory region on the genomic level,
409 and sex-specific gene expression in heterozygous (XY) males *vs.* homozygous (XX)
410 females, including a strong phylogenetic and poolseq-supported signal of local
411 recombination arrest. Its sex-specific expression occurs at the right time, coinciding
412 with that of an ncRNA on the Y, much earlier than histologically detectable
413 differentiation, in the target organs (primordial gonads), and therein in the germ and
414 surrounding somatic cell populations (in male pre-Sertoli but not in female pre-
415 granulosa cells). These properties qualify *bod1l* as a candidate master sex
416 determination gene. While future evidence (e.g. using CRISPR/Cas9) is desirable for
417 its better mechanistic understanding, a single genome-wide peak of heterozygosity
418 revealed by pool-seq and evolutionary conservation of the structural XY-differentiation
419 in several Eurasian taxa along with the shared Y-specific ncRNA, support its
420 candidacy as the Palearctic green toad master gene.
421 The only known amphibian master sex determining gene, *dm-w* of clawed frogs
422 (*Xenopus*), governs female-heterogamety (ZW♀/ZZ♂) and evolved by gene

423 duplication as a *dmrt1*-paralog, involving neofunctionalization of a non-coding part of
424 a DNA-transposon[11,17-19,43]. Here, we provide evidence for a 5'-upstream-region of a
425 gene in a male heterogametic bufonid toad system (XX♀/XY♂), evolved by structural
426 diversification in a common ancestor, as shown by the associated ncRNA-Y-
427 phylogeny (Fig. 3c). To our knowledge, this reveals only the 2nd master sex
428 determining gene in amphibians. *Bod1l* on the Y involves (on the opposite strand) an
429 ancestral insertion of a ncRNA-Y, harboring the partially duplicated BvXY1, and a
430 neighboring signal of a G-quadruples-rich enhancer towards its 3'-end. This is a
431 mechanistic hint that the enhancer modifies *bod1l*-expression in XY-males (Fig. 3), as
432 shown by timely and spatially differential expression in germ and pre-Sertoli but not in
433 pre-granulosa cells (Fig. 4).
434 There are several, partly interlinked pathways by which spatially and/or timely
435 differential *bod1l*-expression may determine sex in green toads.
436 (i) Sex determination of vertebrates differs regarding the timing of primordial germ cells
437 to enter meiosis[44], with females entering meiosis earlier than males, exhibiting an
438 arrest. *Bod1l* is well-known to maintain genomic integrity at sites of DNA-replication
439 damage[45]. *Bod1l* interacts towards its N-terminus with the histone methyltransferase
440 *Setd1a*, which triggers the histone chaperone function of *Fancd2*[46], part of the Fanconi
441 anemia group (*Fanc*), that plays roles in zebrafish sex reversal[47]. Thus, differential
442 *bod1l*-expression might cause a compromised DNA repair system by shifting sex
443 determination towards males.
444 (ii) Interaction of *Bod1l* with *Setd1a* and H3K4me3 refers to tri-methylation on lysine
445 residue 4 in the amino-terminal tail of histone H3, is associated with potentially active
446 promoter regions[48] and linked to double-strand breaks, essential for meiosis 1
447 crossing-over, ensuring chromosome segregation[49,50]. Thus, the differential *bod1l*-
448 expression might enable double-strand breaks as a precondition for the earlier start of
449 recombination in females but not in males. This suggests that *bod1l*-expression in
450 males at the SD stage is, though not significantly, lower than in females (Fig. 2: stage
451 4), presumably by decreased expression of both functional copies, since we have no
452 evidence for a compromised Y-allele.
453 (iii) H3K4-methylation in mammals[48] is associated with *de novo* DNA-methylation in
454 germ cells, and essential for maintaining transcription in the growing oocyte and for
455 triggering zygotic genome activation[51], with XY-mice deficient for the H3K9-
456 demethylating enzyme *Jmjd1a* showing increased H3K9-dimethylation and decreased

457 activating marks of H3K4-trimethylation across the *Sry*-locus, leading to a high
458 frequency of sex reversal[2].
459 According to Dupont and Capel[52] "prior to sex determination" in placental mammals,
460 bipotential supporting gonadal cells of both genetic sexes maintain identical
461 transcriptional programs, allowing commitment to either Sertoli (testes) or granulosa
462 (ovary) cell fate. In pre-Sertoli and pre-granulosa cells, bivalent promoters of key sex
463 determining genes (e.g. *Sox9*, *Fgf9*, *Wnt4*, *Foxl2*) have both, H3K27me3 and
464 H3K4me3 marks[52]. After sex determination by the master sex determination gene *Sry*,
465 promoters to be activated in that sex lost the repressive H3K27me3 mark, while genes
466 to be silenced maintained their bivalency[53]. This suggested that (a) sex-determining
467 genes are poised for activation (or repression) at the bipotential stage, (b) gene
468 activation involves removal of H3K27me3 marks, while (c) repression of the alternative
469 supporting cell fate requires maintenance of existing H3K27me3 repressive marks
470 rather than *de novo* deposition of these marks after commitment to the Sertoli or
471 granulosa program[52].
472 Our preferred hypothesis is that differential *bod1l*-expression (reminiscent of *Sry*)
473 might determine sex by changing H3K4-methylation in green toads during early
474 gonadal development, prior to metamorphosis.
475 To test the hypothesis that *bod1l-expression* coincides in time and space with changes
476 of H3K4-methylation in developing green toad gonads, we used H3K4me-antibodies
477 together with RNA *in-situ* hybridization (Fig. 4). In Gosner stage 43-44, slightly prior to
478 morphological differentiation, we found a strong difference between male and female
479 gonadal somatic cell H3K4-methylation. This suggests that spatially different *bod1l*-
480 expression results in sex-specific methylation patterns that initiate sex-specific
481 gonadal differentiation.
482 Given the multiple potential links between *bod1l*-expression and sex-specific meiotic
483 stages shown above, we hypothesize that this gene functions as a suppressor of the
484 female sex determination pathway – a scenario found in many vertebrates[54]. Following
485 differential *bod1l*-expression, in the bipotential gonad, ovarian pre-granulosa cells
486 transdifferentiate to testicular pre-Sertoli cells. In contrast to the only known amphibian
487 master gene in clawed frogs (*Xenopus*), green toad *bod1l* did not evolve by gene
488 duplication but structural diversification in the 5'-region of the gene, namely by Y-
489 specific sequence insertion with emergence of an enhancer, which is transcribing
490 ncRNA-Y. This evolution of sex determination seems broadly similar to allelic

diversification, so far unknown in amphibians, but known from many examples of master genes in fish[36].

References:

1. Herpin, A. & Schartl, M. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep* **16**, 1260-74 (2015).
2. Capel, B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat Rev Genet* **18**, 675-689 (2017).
3. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol* **12**, e1001899 (2014).
4. Rovatsos, M. *et al.* Conservation of sex chromosomes in lacertid lizards. *Mol Ecol* **25**, 3120-6 (2016).
5. Volff, J.N., Nanda, I., Schmid, M. & Schartl, M. Governing sex determination in fish: regulatory putsches and ephemeral dictators. *Sex Dev* **1**, 85-99 (2007).
6. Perrin, N. Sex reversal: a fountain of youth for sex chromosomes? *Evolution* **63**, 3043-9 (2009).
7. Bertho, S., Herpin, A., Schartl, M. & Guiguen, Y. Lessons from an unusual vertebrate sex-determining gene. *Philos Trans R Soc Lond B Biol Sci* **376**, 20200092 (2021).
8. Kuhl, H. *et al.* A 180 Myr-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes. *Philos Trans R Soc Lond B Biol Sci* **376**, 20200089 (2021).
9. Kostmann, A., Kratochvíl, L. & Rovatsos, M. Poorly differentiated XX/XY sex chromosomes are widely shared across skink radiation. *Proc Biol Sci* **288**, 20202139 (2021).
10. Pan, Q. *et al.* The rise and fall of the ancient northern pike master sex-determining gene. *Elife* **10** (2021).
11. Ito, M. Sex Determination and Differentiation in Frogs. in *Reproductive and Developmental Strategies: The Continuity of Life* (eds. Kobayashi, K., Kitano, T., Iwao, Y. & Kondo, M.) 349-366 (Springer Japan, Tokyo, 2018).
12. Schmid, M. & Steinlein, C. Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class Amphibia. in *Genes and Mechanisms in Vertebrate Sex Determination* (eds. Scherer, G. & Schmid, M.) 143-176 (Birkhäuser Basel, Basel, 2001).
13. Green, D.M. Cytogenetics of the endemic New Zealand frog, Leiopelma hochstetteri: extraordinary supernumerary chromosome variation and a unique sex-chromosome system. *Chromosoma* **97**, 55-70 (1988).
14. Roco Á, S. *et al.* Coexistence of Y, W, and Z sex chromosomes in Xenopus tropicalis. *Proc Natl Acad Sci U S A* **112**, E4752-61 (2015).
15. Furman, B.L.S. *et al.* A frog with three sex chromosomes that co-mingle together in nature: Xenopus tropicalis has a degenerate W and a Y that evolved from a Z chromosome. *PLoS Genet* **16**, e1009121 (2020).
16. AmphibiaWeb. See www. amphibiaweb.org (accessed June 2023).
17. Yoshimoto, S. *et al.* A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis. *Proc Natl Acad Sci U S A* **105**, 2469-74 (2008).
18. Cauret, C.M.S. *et al.* Developmental Systems Drift and the Drivers of Sex Chromosome Evolution. *Mol Biol Evol* **37**, 799-810 (2020).

19. Evans, B.J. *et al.* New insights into Xenopus sex chromosome genomics from the Marsabit clawed frog X. borealis. *J Evol Biol* **35**, 1777-1790 (2022).
20. Bogart, J.P. Karyotypes. in *Evolution of the genus Bufo* (ed. Blair, W.F.) 171-232 (Austin, Texas: University of Texas Press 1972).
21. Schmid, M. Chromosome banding in Amphibia I. Constitutive heterochromatin and nucleolus organizer regions in Bufo and Hyla. *Chromosoma (Berl.)* **66**, 361-388 (1978).
22. Piprek, R.P., Kloc, M. & Kubiak, J.Z. Bidder's organ--structure, development and function. *Int J Dev Biol* **58**, 819-27 (2014).
23. Ponse, K. La proportion sexuelle dans la descendance issue des oeufs produits par l'organe de Bidder des crapauds femelles. *Rev. Suisse Zool.* **48**, 541–544 (1941).
24. Changxiang, W., Qun, L. & Wei, X. Studies of chromosome banding and sister chromatid exchange in Bufo bufo gargarizans. *Acta Gen Sinica* **10**, 291–297 (1983).
25. Abramyan, J., Ezaz, T., Graves, J.A. & Koopman, P. Z and W sex chromosomes in the cane toad (Bufo marinus). *Chromosome Res* **17**, 1015-24 (2009).
26. Malone, J.H. & Fontenot, B.E. Patterns of reproductive isolation in toads. *PLoS One* **3**, e3900 (2008).
27. Stöck, M., Steinlein, C., Lamatsch, D.K., Schartl, M. & Schmid, M. Multiple origins of tetraploid taxa in the Eurasian Bufo viridis subgroup. *Genetica* **124**, 255-72 (2005).
28. Ueda, H. Offspring of sex-reversed males in Bufo viridis. *Sci. Rep. Lab. Amphib. Biol. Hiroshima. Univ.* **10**(1990).
29. Stöck, M. *et al.* Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (Bufo viridis subgroup). *J Evol Biol* **26**, 674-82 (2013).
30. Brelsford, A. *et al.* Homologous sex chromosomes in three deeply divergent anuran species. *Evolution* **67**, 2434-40 (2013).
31. Tamschick, S. *et al.* Sex chromosome conservation, DMRT1 phylogeny and gonad morphology in diploid Palearctic green toads (Bufo viridis subgroup). *Cytogenet Genome Res* **144**, 315-24 (2014).
32. Betto-Colliard, C., Hofmann, S., Sermier, R., Perrin, N. & Stöck, M. Profound genetic divergence and asymmetric parental genome contributions as hallmarks of hybrid speciation in polyploid toads. *Proc Biol Sci* **285**(2018).
33. Burgin, C. *et al.* Mammal Diversity Database (Version 1.11) dataset. Zenodo http://dx.doi.org/10.5281/zenodo.7830771 (2023).
34. NCBI. See www.ncbi.nlm.nih.gov (accessed June 2023).
35. Guzmán, K. *et al.* Identification and characterization of a new family of long satellite DNA, specific of true toads (Anura, Amphibia, Bufonidae). *Scientific Reports* **12**, 13960 (2022).
36. Schartl, M. & Herpin, A. Sex Determination in Vertebrates. in *Encyclopedia of Reproduction (Second Edition)* (ed. Skinner, M.K.) 159-167 (Academic Press, Oxford, 2018).
37. Gosner, K.L. A simplified table for staging anuran embryos and larvae with notes on identification. *Herpetologica*, 183–190 (1960).
38. Burren, O.S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol* **18**, 165 (2017).
39. Jackson, D.A., Dickinson, P. & Cook, P.R. The size of chromatin loops in HeLa cells. *Embo j* **9**, 567-71 (1990).
40. Lyu, J., Shao, R., Kwong Yung, P.Y. & Elsässer, S.J. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res* **50**, e13 (2022).
41. Bayley, R. *et al.* H3K4 methylation by SETD1A/BOD1L facilitates RIF1-dependent NHEJ. *Mol Cell* **82**, 1924-1939.e10 (2022).

589 42. Giovanni, C. *et al.* The BOD1L subunit of the mammalian SETD1A complex sustains
590 the expression of DNA damage repair genes despite restraining H3K4 trimethylation.
591 Preprint at https://www.biorxiv.org/content/10.1101/2023.04.06.535882v2 (2023).
592 43. Hayashi, S. *et al.* Neofunctionalization of a Noncoding Portion of a DNA Transposon
593 in the Coding Region of the Chimerical Sex-Determining Gene dm-W in Xenopus
594 Frogs. *Mol Biol Evol* **39**(2022).
595 44. Bowles, J. & Koopman, P. Retinoic acid, meiosis and germ cell fate in mammals.
596 *Development* **134**, 3401-11 (2007).
597 45. Higgs, M.R. *et al.* BOD1L Is Required to Suppress Deleterious Resection of Stressed
598 Replication Forks. *Mol Cell* **59**, 462-77 (2015).
599 46. Higgs, M.R. *et al.* Histone Methylation by SETD1A Protects Nascent DNA through the
600 Nucleosome Chaperone Activity of FANCD2. *Mol Cell* **71**, 25-41.e6 (2018).
601 47. Rodríguez-Marí, A. & Postlethwait, J.H. The role of Fanconi anemia/BRCA genes in
602 zebrafish sex determination. *Methods Cell Biol* **105**, 461-90 (2011).
603 48. Smallwood, S.A. & Kelsey, G. De novo DNA methylation: a germ cell perspective.
604 *Trends Genet* **28**, 33-42 (2012).
605 49. Youds, J.L. & Boulton, S.J. The choice in meiosis – defining the factors that influence
606 crossover or non-crossover formation. *Journal of Cell Science* **124**, 501-513 (2011).
607 50. Ki, B.S. *et al.* Epigenetic regulator Cfp1 safeguards male meiotic progression by
608 regulating meiotic gene expression. *Experimental & Molecular Medicine* **54**, 1098-
609 1108 (2022).
610 51. Fan, H.-Y. & Sun, Q.-Y. Chapter 12 - Oocyte Meiotic Maturation. in *The Ovary (Third*
611 *Edition)* (eds. Leung, P.C.K. & Adashi, E.Y.) 181-203 (Academic Press, 2019).
612 52. Dupont, S. & Capel, B. The Chromatin State during Gonadal Sex Determination. *Sex*
613 *Dev* **15**, 308-316 (2021).
614 53. Garcia-Moreno, S.A. *et al.* CBX2 is required to stabilize the testis pathway by
615 repressing Wnt signaling. *PLoS Genet* **15**, e1007895 (2019).
616 54. Nakamura, M. Sex determination in amphibians. *Semin Cell Dev Biol* **20**, 271-82
617 (2009).
618

619 **Online content**

620 **Methods**

621 ***Ethics and animal welfare statement***

622 Our research complies with all relevant ethical regulations; the relevant animal
623 protection committee of the IGB and the city of Berlin (LAGESO) approved the study
624 protocol where relevant. Most adult toads were sampled more than 20 years ago by
625 either minimal invasive sampling methods using tail tips of larvae or saliva of adults,
626 obtained by buccal cotton swabs. We also used tissues of road kill toads and/or
627 collected a few adult animals in Greece. For sampling and export of toads and
628 tadpoles, permits were kindly provided by the Ministry of the Environment of Greece
629 (115790/229) 2014-2017. Developmental stages of tadpoles were obtained from
630 control groups of an animal experiment approved by the German State Office of Health
631 and Social Affairs (LaGeSo, Berlin, Germany; G0359/12) and a commercial breeder.

632 The tadpoles were kept according to approved practice (permit ZH 114 to IGB),
633 including the provision of appropriate tank size, sufficient rate of waterflow, natural
634 photoperiod, *ad libitum* food supply, as well as temperatures within the species'
635 thermal tolerance range. This ensured that no pain, suffering, distress or lasting harm
636 was inflicted on the animals. According to the permit G0359/12, tadpoles and juvenile
637 toads were euthanized by an overdose of Tricaine PHARMAQ 1000mg/g (MS222)
638 since this does not incur pain, suffering or distress.
639
640 ***Chromosome-scale genome assembly***
641 The female *B. viridis* reference genome was sequenced in-house at IGB using ONT
642 long reads (MinIon device, SQK-LSK109 kits and R9.4.1 flow cells) and Illumina short
643 reads (Novogene, Europe). The long reads and pre-assembled short reads IDBA-UD
644 assembly[55] were assembled using the WTDBG2.2 assembler[56]. Sequences were
645 error-corrected using two rounds of Flye polish with long reads[57] and four rounds of
646 VCFconsensus using short reads[58]. The chromosome-scale assembly was achieved
647 using Arima HiC-data and the JUICER/3DDNA package and manual curation by
648 juicebox[59-61]. Protein coding genes were annotated using RNAseq data of *Bufonidae*
649 as well as Hyloidea transcriptomes and more distantly related amphibian proteins,
650 available at NCBI. A reference genome browser was created http://genomes.igb-
651 berlin.de/Greentoad/.[62]
652
653 ***Double-digest RAD sequencing (ddRADseq)***
654 DNA of three genetic families consisting for one data set of both parents (buccal
655 swabs) and 40 sexed mostly phenotypically offspring and for two other sample sets of
656 25 or 15 siblings, was extracted using the DNeasy Tissue Kit or the BioSprint robotic
657 workstation (QIAGEN), following the manufacturer's protocols. Double-digest RADseq
658 libraries were prepared according to this study[30]. In brief, genomic DNA digestion by
659 two restriction enzymes (*Sbf*I, *Mse*I) and subsequent ligation of fragments to adapters
660 containing unique barcodes (4–8 bp) was followed by library-PCR for 20 cycles for
661 only different cut-sites containing fragments, and a gel-based size selection, isolating
662 the 400–500 bp fraction. RADseq libraries were single-end sequenced on an Illumina
663 HiSeq 2500, at a 20x coverage.
664 Raw ddRAD reads were demultiplexed with STACKS *process_radtags* version 2.55
665 (cite https://onlinelibrary.wiley.com/doi/full/10.1111/mec.15253), using the enzymes

666 sbfI and mseI (*--renz_1 sbfI --renz_2 mseI*) and the parameters *-c* (clean data), -r
667 (rescue barcodes and RAD-Tag cut sites), and -q (discard reads with low quality
668 scores).
669 Demultiplexed reads were used as input for the RADSEX workflow[63]. First, a table
670 containing the number of occurrences of each unique sequence in the reads file in
671 each individual was generated with *radsex process*, using default parameter values.
672 Then, the distribution of markers between males and females was computed with
673 *radsex distrib* using a minimum depth to consider a marker present in an individual of
674 5 (--min-depth 5). Markers were aligned to the genome of *B. viridis* (bufVir1) using
675 *radsex map* using a minimum depth of 5 (--min-depth 5) and other parameters set to
676 default. Results were plotted with the *SGTR* package, with the entire process provided
677 as a Snakemake workflow[64].
678
679 ***Pool-sex approach***
680 Of 22 safely sexed adult toads of each sex from multiple populations in the range of
681 the European green toad taxa (*B. viridis viridis* and *B. viridis variabilis*), using buccal
682 swabs or tissues genomic DNA was extracted using the DNeasy Tissue Kit or the
683 BioSprint robotic workstation (QIAGEN), following the manufacturer's protocols.
684 Individual DNAs were quantified using the QuantiFluor® dsDNA System kit
685 (Promega), adjusted and equimolarly mixed to yield a female and a male pool. Whole
686 genome library preparation for 300 bp fragments and whole genome sequencing was
687 performed at NOVOGENE (UK), on a NovaSeq PE150 to yield 131 and 140 Gbp of
688 data for male and female pool, respectively. Reads were mapped to the reference
689 genome using the short read mode of MINIMAP2 (parameters -K 2000M -I 100G -a -
690 x sr -t 80[65], and converted to sorted bam-files by SAMTOOLS[66]. Variants (SNPs,
691 MNPs, INDELs) were called from both bam files using the PLATYPUS variant caller[67]
692 with integrated duplicate fragment reads removal (–filterDuplicates = 1) and
693 reassembly of reads (–assembleAll = 1). The resulting vcf-file was screened for XY-
694 or XX-specific variants using AWK-scripting and different variant read coverage cutoffs
695 (i.e. SNP read coverage in females and males >=15; variant read count in females
696 <=7% or >=93% of total SNP read coverage as a condition for homozygosity; variant
697 frequency in males between 45 - 55% of total SNP read cov. as a condition for
698 heterozygosity; for screening ZW SNPs conditions for males and females were
699 interchanged). Variants were clustered according to their distance in the genome

700 (maximal distance of variants in a cluster: 9000 bp), counted using BEDTOOLS
701 (merge, annotate[68]) and sorted by variant count. Regions with highest variant counts
702 or densities and neighboring regions were inspected for genes related to sex
703 determination.
704
705 ***Sex-diagnostic PCR marker design***
706 The region 566,783,058-566,848,882 bp on scaffold 1 includes several INDELs
707 apparent between the X and the Y haplotype, allowing design of a number of primers
708 that were primarily tested in gradient-PCRs on phenotypically sexed *B. viridis*. While
709 several primers showed sex-specific amplification in this species, one (BvXY1) could
710 be optimized to work across multiple species.
711 A collection of DNAs (extraction: see above) comprising samples of phenotypically
712 sexed green toads of the entire radiation[32] served to test sex-specificity of all PCR
713 markers. PCR-amplifications were carried out in 25 µL reactions containing 1.25 µL of
714 each primer (BvXY1_F: 5'_CTGAATAGCAATTGGACACAGC_3; BvXY1_R:
715 5'_ATGTGAGAGCTGTTACAATGGA_3'), 14.5 µL of water, 5 µL of AllTaq buffer (5x),
716 0.5 µL of AllTaq polymerase (Qiagen), and 2 µL of DNA (concentration 4-20 ng/µL).
717 Temperature regime comprised 2 min at 95°C (initial denaturation), and 38 cycles of
718 (1 min at 95°C denaturation, 45 s at 59°C annealing, 45 s at 72°C extension), and 5
719 min at 72°C final extension. PCR products (4 µL) were visualized by electrophoresis
720 on 4% high-resolution Sieve-agarose (Biozym) gels, run for 60 min at 70 V.
721 To PCR-amplify either a 2.5 kb fragment or nearly the entire ncRNA-Y, comprising 5.8
722 kb based on gDNA on the Y-chromosome in all diploid green toad taxa with sex-
723 specific BvXY1-markers, we used the same ingredients and volumes as above either
724 with primers nRNA_Y_ex1F: 5'_GGGCAGTGAAGAGAAGTCCT_3', nRNA_Y_ex2R
725 5'_TTCACCATTCTCCAGCCTGT_3' for the 2.5 kb-fragment or
726 5.8kb_ncRNA_Y_F: 5'_GTGGTAGTTGCTGTGCCTTT_3', 5.8kb_ncRNA_Y_R:
727 5'_CCGGTGTGGTGTAAAATATCCA_3' for the 5.8 kb-piece. Temperature regimes
728 involved 2 min at 95°C, and 37 cycles of (1 min at 95°C denaturation; 2.5 min at 63°C
729 (for the 2.5 kb) or 5 min at 59.4°C (for 5.8 kb) annealing; 1 min at 72°C extension and
730 5 min at 72°C final extension (note extended annealing times). PCR products (4 µL)
731 were visualized by electrophoresis on 1.5% agarose (Biozym) gels, run for 60 min at
732 80 V. PCR-products (21 µL) were barcoded using the NBD196 and LSK110 ONT-kits,
733 and sequenced using the ONT-MinIon device, as described. Basecalling was

734 performed using ONTs GUPPY version 6 and the super accuracy base-calling model
735 (SUP-model).
736
737 **Phylogenetic tree of a 2.7 kb-region at the Y-specific region boundary**
738 Consensus sequences of 100 ONT reads of Y-specific amplicons were calculated
739 using NGSPECIESID[69]. Homologous X-sequences were reconstructed from targeted
740 enrichment sequencing data of females using BCFTOOLS consensus[70] for the region
741 scf1:566,788,797-566,791,468. We also excluded regions of the sequence not
742 covered by enrichment data by supplying a masking file through parameter "-m" of
743 bcftools consensus. All resulting sequences were multiple aligned using MAFFT
744 (automatic alignment parameters). The alignment was filtered by GBLOCKS[71] to
745 remove columns with excessive missing data. IQTREE2[72] was used with parameter "-
746 alrt 1000 -bb 1000" to calculate a phylogenetic tree. To calculate a phylogenetic tree
747 of the upstream neighboring region (scf1:566,786,116-566,788,796) data from target
748 enrichment for males and females was used as described above.
749
750 ***Phenotypic sexing in multiple species***
751 Phenotypic sex of adult toads of multiple diploid Eurasian green toad taxa had been
752 established during the breeding period when males exhibited secondary sexual
753 characters like nuptial excrescences and thorny glandular tips on the back, or were
754 caught producing male advertisement calls and/or male release calls. Females were
755 identified as filled with eggs, caught in amplexus, and/or while spawning, producing
756 female release calls and/or lacking the male secondary sexual characters. During the
757 spawning migrations, several road kill toads also allowed for anatomical identification
758 (possession of ovaries or testes) of phenotypic sex.
759
760 ***Assembly of Y-haplotype***
761 Several iterative approaches were used to extract Y-haplotypic sequencing reads of
762 *B. viridis* GR57 male[73] long read data (Oxford Nanopore Promethion platform;
763 Novogene). First, we mapped the reads to the BvXY1-marker sequences to separate
764 Y reads of the region. Second, we performed a FLYE[57] assembly of the male data,
765 identified several contigs with variation compared to the X haplotype (*bod1l*-region),
766 and extracted the corresponding reads. The Y-specific reads were assembled, then
767 added to the female assembly and all male reads were mapped to the combined

768 sequences, which allowed finding more reads matching better to Y than to X contigs.
769 This procedure was iteratively repeated several times until the Y-haplotype and its
770 structure could be assembled. Pool-seq data was again mapped to the female
771 assembly with added Y-haplotype sequences by MINIMAP2[65] showing the expected
772 coverage differences between female and male pools. A dot plot comparison of X
773 versus Y haplotype involved the corresponding function of the NCBI Blast website.
774
775 **Testing enhancer activity of the Y-haplotype**
776 We used the AI-based predictor "iEnhancer-Deep" in a 200 bp-"sliding-windows"
777 approach in the ncRNA region to test its *in-silico* potential as enhancers of *bod1l-*
778 *expression*[74].
779
780 ***Gene expression using RNAseq during development in sibling tadpoles from***
781 ***multiple European populations***
782 In total 54 transcriptomes of sibling tadpoles comprising brothers and sisters (plus
783 three unsexed larvae) of four genetic families from four different geographic locations
784 as well as gonadal tissue of a subadult male and adult female from two additional
785 localities were examined by RNAseq (transcriptomics). Tadpoles presented five
786 developmental stages from control groups of an ecotoxicological experiment[31]. They
787 were euthanized in MS222 and flash-frozen in liquid nitrogen. In tadpoles older than
788 Gosner stage 37, the intestine was removed prior to flash-freezing. In tadpoles at
789 Gosner stage 43-44, only the body segment that contained the gonads was used for
790 RNA extraction. In toadlets of Gosner stage 46 (specifically 6 weeks after
791 metamorphosis), ovaries, testes (separated from Bidder's organ), and several other
792 organs (brain, heart, liver, intestine) were dissected and stored in RNAlater prior to
793 flash-freezing. DNA from tailtips, or skin samples of each tadpole or buccal swabs
794 (juvenile toads) served for genetic sexing using BvXY1-primers. RNAs were extracted
795 using the TRIzol Reagent (Thermo Fisher Scientific, Waltham, USA) according to the
796 supplier's recommendation and cleaned using the RNeasy Mini Kit (Qiagen). All RNAs
797 included in this study exhibited RIN values >9.0. Library processing and RNAseq were
798 carried out by NOVOGENE (Cambridge, UK) on a NovaSeq 6000 PE 150, generating
799 9 Gb of raw data per sample.
800 Transcriptome sequences of 54 male and female juvenile developmental stages and
801 two adult gonads were normalized and mapped to the genome using the nf-core RNA-

802 Seq pipeline in STAR aligner (https://github.com/alexdobin/STAR/releases,—
803 quantMode GeneCounts). A list comprising genes that play roles in sexual
804 development and/or sex determination in other vertebrates (also known as 'usual
805 suspects'[1]) is shown as Supplementary File 5). To plot *bod1l* and *ncRNA-Y* expression
806 trends over the developmental stages and separate expression analysis of *bod1l-X*
807 and *bod1l-Y,* we mapped all reads, including the male haplotype, to the female
808 assembly. As the *bod1l-Y* gene model is 3'-partial, we used the corresponding parts
809 of the *bod1l-X* gene model for quantification. Counts of reads matching the different
810 haplotypes' gene models were extracted using FEATURECOUNTS (from SUBREAD-
811 2.0.3 package[75] with mapping quality (MQ) higher than 19, to increase haplotype
812 specificity. Read counts were normalized by total mapped reads of each sample (TPM
813 normalization). Average expression and standard deviation was calculated and plotted
814 combining samples of the different stages using spreadsheet software.
815
816 ***cDNA of developmental stages***
817 Using the iScript RT Supermix for RT-qPCR kit (Biorad), 10 µL of RNA (concentration
818 of 80 ng/µL) from all developmental stages (as used for RNAseq) were randomly
819 reverse-transcribed into cDNA according to the manufacturer's recommendations. To
820 test allele-specific *bod1l*-gene expression, we designed primers (bod1l_RNA1F:
821 5'_AACTGGAAGAAAAGCGCAAA_3'; bod1l_RNA1R:
822 5'_GTGCTCTTGGCATTCTGTGA_3') and amplified 577 bp from cDNA under the
823 same conditions as for BvXY1 that were Sanger-sequenced in both directions. To test
824 expression of the ncRNA-Y, we used primers designed to match its inter-exonic
825 regions NcRNAY_Ex1to2_F: 5'_CATGAGGAGAGCAGTAATGAACC_3',
826 NcRNAY_Ex3to2_R: 5'_GCTGTCATCAACCTGCTCAA_3' and amplified and
827 visualize 204 bp male-specific products from cDNA under the same conditions as for
828 BvXY1 but at 55°C annealing temperature.
829
830 ***Probe preparation for RNA* in-situ *hybridization***
831 The sequence marker obtained by cDNA amplification using bod1l_RNA1F/R primers
832 was also used as a probe for RNA *in-situ* hybridization on histological gonad sections
833 of various developmental stages. As a probe of a control gene, expressed in gonads
834 of both sexes, we used a 501 bp cDNA fragment of *Vasa* (DDX4-DEAD-Box-Helicase-
835 4, VASA-homolog), amplified with the primers Vasa1F:

836 5'_AACCAAAAAGAAGGCGGATT_3'; Vasa1R:
837 5'_ATGCCAGGATTGCTGTAACC_3' under the same conditions as for BvXY1 but
838 with annealing temperature of 57°C.
839
840 ***RNA* in-situ *hybridization and H3K4me antibody staining***
841 We opened the body cavity of freshly euthanized tadpoles of three developmental
842 stages and fixed them in 4% paraformaldehyde (PFA) buffered in 1x PBS
843 (Thermofisher), pH = 7.0, at 4°C overnight. Samples were washed in filtered 1×
844 phosphate-buffered saline with tween 20 (PBST, ROTI®Stock) twice for 5 min, then
845 dehydrated in a series of methanol (MetOH) with shaking at room temperature (25%
846 MetOH in PBST for 10 min to 1 h, 50% MetOH in PBST for 10 min to 1 h, 75% MetOH
847 in PBST for 10 min to 1 h, 100% MetOH for 10 min to 1h, timing dependent on tissue
848 sample size). Samples were stored in 100% MetOH at -20°C until histological
849 sectioning.
850 Cryosectioning, RNA *in-situ* hybridization and immunohistochemistry were performed
851 as described[76] with slight modifications. Transverse cryosections (20 µm) were
852 collected on SuperFrost Plus glass slides (Menzel Gläser J1800AMNZ) and stored at
853 -20°C for up to 2 months. To generate *bod1l* and *vasa* riboprobes, primers
854 Bod1_RNA_Bv_1F (5'_AACTGGAAGAAAAGCGCAAA_3') and Bod1_RNA_Bv_1R
855 (5'_GTGCTCTTGGCATTCTGTGA_3') were used to amplify a 577 bp *bod1l* sequence
856 and primers vasa_1F (5'_AACCAAAAAGAAGGCGGATT_3') and vasa_1R
857 (5'_ATGCCAGGATTGCTGTAACC_3') were used to amplify a 501 bp *Vasa* sequence
858 from Gosner 36-37 *B. viridis* cDNA. These sequences were then ligated to a pCR-
859 bluntII-TOPO vector (Invitrogen) and plasmids were purified using a plasmid DNA
860 miniprep kit according to manufacturer's instructions (Promega A1460). For synthesis
861 of sense and antisense riboprobes, plasmids were linearized using *BamHI* or *NotI*
862 (New England Biolabs) and *in vitro*-transcribed using T7 or SP6 polymerase (DIG RNA
863 Labeling Mix, Roche), respectively. Rabbit polyclonal anti-histone H3 (tri methyl K4)
864 primary antibody (Abcam ab8580) and secondary goat anti-rabbit Alexa Fluor 568
865 (1:500, Invitrogen A-11011) were used. Sections were counterstained with 0.25 µg/mL
866 DAPI and mounted in Mowiol 4-88 (Calbiochem). Imaging was performed using the
867 Nikon SMZ18 stereomicroscope and Olympus FluoView FV3000 confocal
868 microscope. To process images, Bitplane Imaris and Adobe Photoshop were used.
869

870 ***FISH-TSA***

871 As a probe for fluorescence *in-situ* hybridization with tyramide signal amplification
872 (FISH-TSA), we designed novel primers to amplify a longer portion of *bod1l* from cDNA
873 (*bod1l*_Exon7To10_F: 5′_AACTGGAAGAAAAGCGCAAA_3′, *bod1l*_Exon7To10_R:
874 5′_TGTGTTGTCTGGAATCACTGG_3′; same conditions as for BvXY1 but with
875 annealing Temperature of 60°C.), exhibiting the highest heterozygosity in the coding
876 segment of the gene in *B. viridis* males. A PCR-product was produced using the PPP
877 Master Mix (Top-Bio, Czech Republic) according to manufacturer's instructions
878 (annealing temperature 60°C) and purified[77]. The cDNA amplicons were cloned using
879 the TOPO-TA cloning kit and DH5α-T1R Competent Cells (Thermo Fisher Scientific).
880 Plasmid DNA was isolated from bacterial colonies using the EZNA Plasmid DNA Mini
881 Kit I (Omega Bio-Tek Inc.), sequenced, and tested to contain the correct insert by
882 BLAT search on the green toad genome browser (see Data availability). The sequence
883 of the *bod1l* clone was compared to the reference genome. Subsequently, the *bod1l*
884 locus was re-amplified and purified[77]. A 20 years old tissue suspension (bone marrow),
885 obtained from a male European green toad from Germany, treated with colchicine and
886 fixed in methanol-acetic acid[27] was dropped onto a microscope slide[78], followed by
887 FISH. The probe was labelled with digoxigenin-11-dUTP (Jena Bioscience) using
888 DecaLabel DNA Labelling Kit (Thermo Fisher Scientific). The FISH-TSA followed[79]
889 with minor modifications[80]. Mitotic metaphase spreads were counterstained with
890 ProLongTM Diamond Antifade Mountant with the fluorescent 4',6-diamidino-2-
891 phenylindole, DAPI stain (ThermoFisher Scientific). Microscopy using Leica
892 Microsystem (Wetzlar, Germany) and processing of metaphase images followed[81].
893

894 ***Analysis of potential alternative splicing of bod1l***

895 For the entire *bod1l* region, including the 3'- and 5-' UTR-regions (scf1:566495375-
896 567417162), aligned RNAseq-reads were extracted from each RNA-seq alignment
897 file, the splice sites locations were retrieved and counted. Using all the splice sites, a
898 reference file of them was built. The splice event counts were merged for all samples
899 using the reference and stored as csv text file (Supplementary file 6).
900

901 ***Structural evaluation of the* bod1l-*X and Y protein copies***

902 We used Alphafold2[82] via its cloud-based application and its local installation
903 ColabFold[82] and analyzed the structure predictions by RaptorX[83] for structural and

904 potential functional amino-acid differences between the X and the Y-copy of *B. viridis*.
905 The molecular data viewer iCn3D[84] was used to inspect the structures and highlight
906 specific amino acid residues (Supplementary text 1).
907
908 ***AmpliSeq custom DNA panel***
909 After providing pre-publication access to the *B. viridis* genome-browser, the Illumina
910 concierge design service team developed two primer pools covering ca. 80 kb of the
911 *bod1l*-region (scf1:566,785,000-566,865,000) and 10 kb for a control gene (*amh*),
912 known to play roles in sexual development and sex determination in other vertebrates
913 and lying on the same scaffold (scf1:324181812-324192250). Overall, the theoretical
914 coverage was 92.3% for *bod1l*, and 98.2% for *Amh*.
915 Wet lab approaches followed the "AmpliSeq for Illumina On-Demand, Custom, and
916 Community Panels" (Document #1000000036408 v09; pp. 5-23[85]) for two primer pools
917 with 193-384 primer pairs in the standard workflow. Briefly, this included ten steps, (i)
918 quantification of DNA samples of multiple sexed green toads (QuantiFluor® dsDNA
919 System kit, Promega), (ii) target amplification in two separate wells for the two primer
920 pools, (iii) partial digestion of amplicons, (iv) ligation of individual indices/sample, (v)
921 library clean up and (vi) amplification, (vii) 2nd clean up, (viii) checking libraries on the
922 Bioanalyser for expected insert size and concentration (QuantiFluor® dsDNA System
923 kit, Promega), (ix) dilution and mixing, and (x) sequencing, performed by Novogene
924 on a NovaSeq PE250 for individual libraries obtained for 54 diploid green toads (ca.
925 100 Mio bp raw data per library).
926 We included Palearctic green toads by the following numbers, taxa and sexes: 8 *B.*
927 *siculus* (4 ♂♂, 3 ♀♀), 9 *B. viridis* (5 ♂♂, 4 ♀♀), 8 *B. balearicus* (5 ♂♂, 3 ♀♀), 9 *B.*
928 *shaartusiensis* (4 ♂♂, 5 ♀♀), 10 *B. turanensis* (4 ♂♂, 5 ♀♀). Species' nomenclature
929 follows[32].
930 The sequence reads from target enrichment were mapped against the reference
931 genome using MINIMAP22 (-I 100G -a -x sr -t 4)[65]. SAMTOOLS[66] was used to add
932 read group and sample names, to fix mates and sort the bam file output. All BAM files
933 were merged and variant calling for the target regions was performed by SAMTOOLS
934 and BCFTOOLS[66] with parameters to cope for extreme sequencing depth (samtools
935 view -bu input.bam scf1:324181812-324192250 scf1:566783149-566866298 |
936 bcftools mpileup --annotate DP4 --max-depth 1000000000 -f bufVir1d.fa - | bcftools
937 call -mv -Ob -o $1.calls.bcf). The output was converted to vcf format and summarized

938 by an awk script that filtered SNP position and calculated allele frequency from DP4
939 annotation for each sample. The resulting table was further inspected in Excel and
940 SNPs that were heterozygous in males while homozygous in females were filtered. To
941 fine map the region of the sex marker male specific SNPs of the the different species
942 were intersected (resulting in male specific SNPs present in 2, 3, 4, 5 species) and
943 plotted along the target regions in our genome browser.
944

945 **References:**
946

947 55. Peng, Y., Leung, H.C., Yiu, S.M. & Chin, F.Y. IDBA-UD: a de novo assembler for
948 single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*
949 **28**, 1420-8 (2012).
950 56. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**,
951 155-158 (2020).
952 57. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P.A. Assembly of long, error-prone
953 reads using repeat graphs. *Nat Biotechnol* **37**, 540-546 (2019).
954 58. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8
955 (2011).
956 59. Durand, N.C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution
957 Hi-C Experiments. *Cell Syst* **3**, 95-8 (2016).
958 60. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields
959 chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
960 61. Durand, N.C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps
961 with Unlimited Zoom. *Cell Syst* **3**, 99-101 (2016).
962 62. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006
963 (2002).
964 63. Feron, R. *et al.* RADSex: A computational workflow to study sex determination using
965 restriction site-associated DNA sequencing data. *Mol Ecol Resour* **21**, 1715-1731
966 (2021).
967 64. https://github.com/RomainFeron/paper-sexdetermination-greentoad-2021.
968 65. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,
969 3094-3100 (2018).
970 66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
971 2078-9 (2009).
972 67. Rimmer, A. *et al.* Integrating mapping-, assembly- and haplotype-based approaches for
973 calling variants in clinical sequencing applications. *Nat Genet* **46**, 912-918 (2014).
974 68. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing
975 genomic features. *Bioinformatics* **26**, 841-2 (2010).
976 69. Sahlin, K., Lim, M.C.W. & Prost, S. NGSpeciesID: DNA barcode and amplicon
977 consensus generation from long-read sequencing data. *Ecol Evol* **11**, 1392-1398 (2021).
978 70. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2021).
979 71. Castresana, J. Selection of conserved blocks from multiple alignments for their use in
980 phylogenetic analysis. *Mol Biol Evol* **17**, 540-52 (2000).
981 72. Minh, B.Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
982 Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534 (2020).

73. Gerchen, J.F. *et al.* A Single Transcriptome of a Green Toad (Bufo viridis) Yields Candidate Genes for Sex Determination and -Differentiation and Non-Anonymous Population Genetic Markers. *PLoS One* **11**, e0156419 (2016).
74. Kamran, H., Tahir, M., Tayara, H. & Chong, K.T. iEnhancer-Deep: A Computational Predictor for Enhancer Sites and Their Strength Using Deep Learning. *Applied Sciences* **12**, 2120 (2022).
75. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
76. Dasyani, M. *et al.* Lineage tracing of col10a1 cells identifies distinct progenitor populations for osteoblasts and joint cells in the regenerating fin of medaka (Oryzias latipes). *Dev Biol* **455**, 85-99 (2019).
77. Knytl, M. & Fornaini, N.R. Measurement of Chromosomal Arms and FISH Reveal Complex Genome Architecture and Standardized Karyotype of Model Fish, Genus Carassius. *Cells* **10**(2021).
78. Courtet, M., Flajnik, M. & Du Pasquier, L. Major histocompatibility complex and immunoglobulin loci visualized by in situ hybridization on Xenopus chromosomes. *Dev Comp Immunol* **25**, 149-57 (2001).
79. Krylov, V., Tlapakova, T. & Macha, J. Localization of the single copy gene Mdh2 on Xenopus tropicalis chromosomes by FISH-TSA. *Cytogenet Genome Res* **116**, 110-2 (2007).
80. Knytl, M., Tlapakova, T., Vankova, T. & Krylov, V. Silurana Chromosomal Evolution: A New Piece to the Puzzle. *Cytogenet Genome Res* **156**, 223-228 (2018).
81. Seroussi, E. *et al.* Avian Expression Patterns and Genomic Mapping Implicate Leptin in Digestion and TNF in Immunity, Suggesting That Their Interacting Adipokine Role Has Been Acquired Only in Mammals. *Int J Mol Sci* **20**(2019).
82. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).
83. Peng, J. & Xu, J. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins* **79 Suppl 10**, 161-71 (2011).
84. Wang, J. *et al.* iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* **36**, 131-135 (2020).
85. https://support.illumina.com/downloads/ampliseq-for-illumina-custom-and-community-panels-reference-guide-1000000036408.html.

Acknowledgements

Funding has been in part provided by the German Research Foundation (DFG) to MStöck and HK (STO493/8), and the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB). WKloas and MStöck also received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement No. 825753 (ERGO). FISH-TSA experiments were partly covered by the PJAC-project CZ.02.01.01/00/22_010/0002902, MSCA Fellowships CZ–UK (MK). We thank Petros Lymberakis, Natural History Museum of Crete, University of Crete, for help with the sampling and export permits of green toads by the Ministry of the Environment of Greece (115790/229) 2014-2017. The tissue and DNA collection of MStöck profited

from donations of multiple international colleagues over 25 years; he thanks everyone. We are grateful to Sandra Bittmann (IGB) for HMW-DNA extraction, Eva Kreuz (IGB) for initial efforts to clone the SD-region, Julia Bindl (IGB) for help with the DNA and RNA work, Hughes Parinello (Plateforme MGX) for preparing the HiC-libraries and sequencing, the BeGenDiv (Berlin, Germany) for providing lab space and Illumina-sequencing support for the ddRAD. We thank Sebastian Petri (Biozyme) for discussions on high-resolution agarose gel-electrophoresis.

Authors' contributions:

HK: Contributed to study design, sequenced long reads, assembled and annotated the genome, set up the genome browser, mapped and analyzed pool-seq sequences and SNPs of the AmpliSeq custom DNA panel, performed phylogenetic analyses, contributed to 1st manuscript draft

TWH: Made histological sections and performed all RNA *in-situ* hybridizations and H3K4me immunostainings, contributed to 1st manuscript draft

CK: Analyzed all transcriptomes, and gene expression data, analyzed splice sites of *bod1l*

WKleiner: Performed RNA extraction, bioanalyzer quantification and evaluation, did most marker testing by PCRs, co-worked in the wet lab for the AmpliSeq custom DNA panel

BK: Performed molecular work to characterize the ncRNA of multiple green toad taxa, prepared RNA for transcriptomes and cDNA-work at Gosner stage 46

MC: Contributed to ddRAD wet lab

RF: Analyzed the ddRADseq data

MK: Performed FISH-TSA

WKloas: Contributed to study design, discussed developmental histology and data.

MSchartl: Advised in many respects, initiated some collaborations, discussed data and study progress and results.

CW: Contributed to study design, supervised and analyzed RNA *in-situ* hybridization and H3K4me immunostainings.

MStöck: Conceived the study, did all field work/sampling, raised tadpoles, prepared and fixed tissues for chromosome and RNA *in situ* work, collected buccal swabs and tissues over 25 years, sexed toads, extracted all DNAs, performed ddRAD wet lab,

1061 developed all primers, did some marker testing by PCRs, co-worked in the wet lab for
1062 the AmpliSeq custom DNA panel, wrote the 1st manuscript draft.
1063 All authors contributed text to the manuscript, and read and improved the final version.
1064
1065 **Competing interests**
1066 The authors declare to have no competing interests.
1067
1068 **Data availability**
1069 All data generated in this study have been deposited in publicly accessible databases.
1070 Genome assembly, whole genome sequencing (WGS) data and RNA-seq reads have
1071 been submitted to GenBank on 25 February 2022 (submission ID: SUB11111520,
1072 Bioproject: PRJNA292872, Biosample: SAMN26135295). Alternatively, the IGB-
1073 Green Toad genome browser http://genomes.igb-berlin.de/Greentoad/ provides easy
1074 access to the assembly and its annotation.
1075
1076 **Additional information**
1077
1078 **Supplementary File 1: Description of Supplementary Data.**
1079 **Supplementary File 2: Analyses of SNPs causing 6 non-synonymous changes**
1080 **in the coding region of *bod1l* in 49 male and female transcriptomes as compared**
1081 **to the female reference genome.** None of these changes was 100% associated with
1082 genetic sex as in the 5'-region of *bod1l*.
1083 **Supplementary File 3: Percentages of covered parts of ca. 80 kb of the *bod1l*-**
1084 **region by targeted sequencing in six diploid green toad taxa, using a panel**
1085 **based on the female *B. viridis* reference genome, compared to ca. 20 kb of the**
1086 **gene for *anti-Muellerian hormone* (*amh*).** Taxa abbreviations: Bbal - *balearicus*,
1087 Bsha - *shaartusiensis*, Bsic - *siculus*; Btur - *turanensis*, Bvir - *viridis* and *variabilis*.
1088 **Supplementary File 4: SNP-comparisons in ca. 80 kb of the *bod1l*- (scf1:**
1089 **566,783,513 - 566,865,063) and ca. 10 kb of the *amh*-regions (scf1: 324,181,734 –**
1090 **324,192,327) by targeted sequencing in six diploid green toad taxa in a panel**
1091 **based on the female *B. viridis* reference genome.** The coordinates of the cluster of
1092 male specific SNPs present in all five species are colored in red (scf1: 566,790,085 -
1093 566,790,204). Values for color-coded SNP-comparisons with the reference genome:
1094 >=0.9: homozygous (green/yellow) = identical to reference; between 0.9 and 0.1:

heterozygous (orange) = one allele identical to reference; <=0.1: homozygous(red) = completely different from reference; NA: SNP not covered by sequencing data in the relevant individual / taxon; M - male, F - female; position: SNP position on the female *B. viridis* reference genome; taxa abbreviations as in Supplementary File S3.

Supplementary File 5: Boxplots showing expression of 80 genes that play roles in sexual development and/or sex determination in other vertebrates (also known as 'usual suspects'[1], in developmental stages of *B. viridis*. X-axis showing the normalized intensity of expression; Y-axis marks six developmental stages (details as in Fig. 2).

Supplementary File 6: File of splice site read counts in *bod1l*-region.

Supplementary Figures

Supplementary Figure 1: **Analyses of ddRAD markers in three genetic families of *B. viridis* using RADsex[63].** Of the 443,378 markers present in at least one individual with a coverage >5; 96 markers were significantly associated with male sex ($p < 0.05$ after Bonferroni correction), but none of these markers were found in all males.

Supplementary Figure 2: **Alignment of ddRAD markers shown in Supplementary Figure 1 to the *B. viridis* reference genome.** 64 of the 96 markers significantly associated with sex were located on chromosome 1 (scaffold 1, scf1), one marker was aligned to chromosome 7 (scf7), five markers were aligned to unplaced scaffolds, and 26 markers were not aligned uniquely or with a mapping quality lower than 20.

Supplementary Figure 3: **Mapping pool-sex short reads to the *B. viridis* reference genome with SNP-heterozygosity in males (a) *vs*. females (b) on scaffold (scf) 1.**

Supplementary Figure 4: Male-specific SNPs of 6 species over *bod1l*- region. a SNPs show different blocks of male-specific heterozygosity between species, which exclude coding changes of *bod1l* as sex-determination mechanism. **b** Highest congruence of male-specific SNPs among all species is in the PCR marker/ncRNA-region (red ellipsis). **c** Phylogenetic tree of haplotypes from the marked region shows clustering by gametologue. **d** sequences from the upstream region were clustered by species.

Supplementary Figure 5: Coverage distributions of X/Y haplotypes depending on female/male pool-seq in *B. viridis*. a Y-haplotype, **b** X-haplotype.

1129 **Supplementary Figure 6: Female/male pool-seq coverage over haplotypes in *B.***
1130 ***viridis*. a** female pool-sequencing, **b** male pool-sequencing.
1131 **Supplementary Figure 7: X- and Y-haplotype comparisons by dot plots in the**
1132 ***bod1l*-region. a** X-*versus*-Y dot plot, **b** Y-*versus*-Y, **c** X-*versus*-X.
1133 **Supplementary Figure 8: a, Y- and thus male-specific expression of the ncRNA-**
1134 **Y (204 bp PCR-product from cDNA) in larval green toads.** PCR-products from
1135 cDNA (204 bp), obtained with primers NcRNAY_Ex1to2_F/_R; to the left males, to the
1136 right females; each sex is shown from 10 days after hatching until early
1137 metamorphosis (Gonser 43-44) and one adult gonad for each sex; in females, only
1138 non-specific products occurred, since X-copies of *bod1l* do not contain and thus do
1139 not express this ncRNA. **b,** Gel image of a control-PCR using the females from (a)
1140 **Supplementary Figure 9: Graph of the signal of the AI-based software i-**
1141 **enhancer**. Sinal bars reaching from 0 to 2, predicting strong enhancer properties of
1142 the Y-specific 5'-end of *bod1l*, including a G-quadruplex-rich region on the non-
1143 template strand, overlapping with the ncRNA-Y.
1144 **Supplementary Figure 10: *Vasa*, *bod1l* expression and H3K4me3-methylation in**
1145 ***B. viridis* gonads, overview images and additional representative examples.**
1146 **a-k,** *Vasa* and *bod1L* expression shown by RNA *in-situ* hybridization (ISH) on 20 µm
1147 transverse cryosections of female (**a, b, e, f**) and male (**c, d, g-k**) gonads, Gosner 36-
1148 37. DAPI was used to stain nuclei. Sense controls for RNA ISH are shown in **a, c, e**,
1149 **g**, right panels. Red boxes in **a, c, e, g** (left panel) indicate zoomed images shown in
1150 **b, d, f, h**, respectively. Yellow boxes in **b, d, f, h** (right panel) indicate zoomed images
1151 shown in **Fig. 4a-d,** respectively. Red boxes in **i, k** (left panel) indicate zoomed images
1152 shown in right panels of **i, k**. *Bod1L* staining on three different cryosections of male
1153 gonads are shown in **g, i, k**. Magenta asterisks indicate nuclei of germ cells; magenta
1154 arrows: cytoplasm of germ cells; yellow arrows: somatic cells not stained by RNA ISH;
1155 green arrows: somatic cells stained by RNA ISH. **l,m,** H3K4me3 and DAPI staining on
1156 20 µm transverse cryosections of female (**l**) and male (**m**) *B. viridis* Bidder's organ and
1157 gonads, Gosner stage 43-44. Yellow boxes in **l, m** indicate zoomed images shown in
1158 **Fig. 4g** and **Fig. 4h,** respectively. White asterisks indicate H3K4me3-positive germ
1159 cell nuclei; white arrowheads, H3K4me3-positive somatic cell nuclei; yellow
1160 arrowheads, H3K4me3-negative somatic cell nuclei. Scale bars: 50 µm.

1161

1162 **Supplementary Tables**

1163 **Supplementary Table 1**: Localities/populations of European green toads that
1164 contributed to the pool-seq experiment.

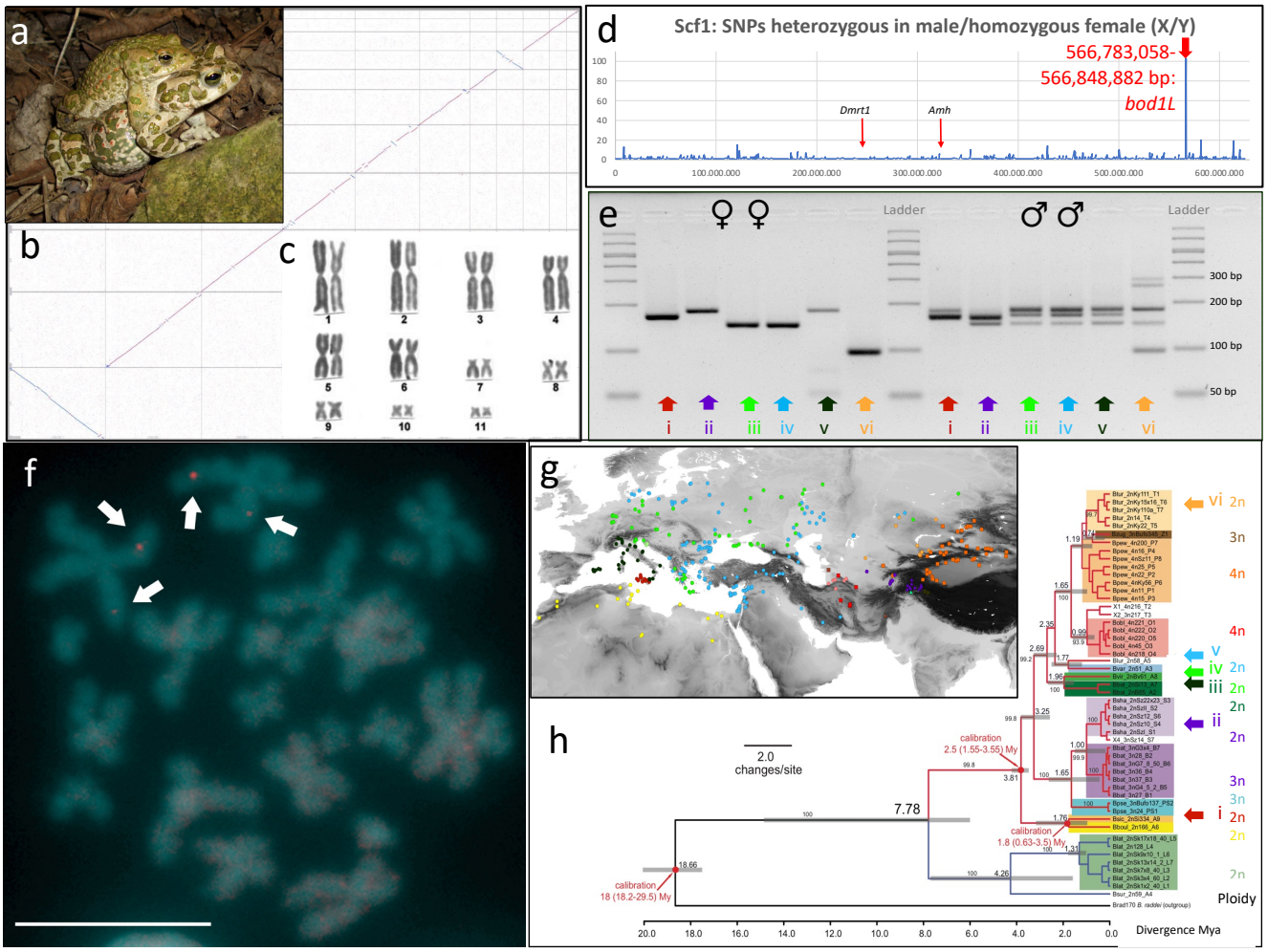
1165 **Supplementary Table 2**: Summary of the pool sequencing approach for scaffold 1
1166 (scf1) for SNPs homozygous in males/heterozygous females (Z/W) as well as SNPs
1167 heterozygous in males/homozygous females (X/Y).

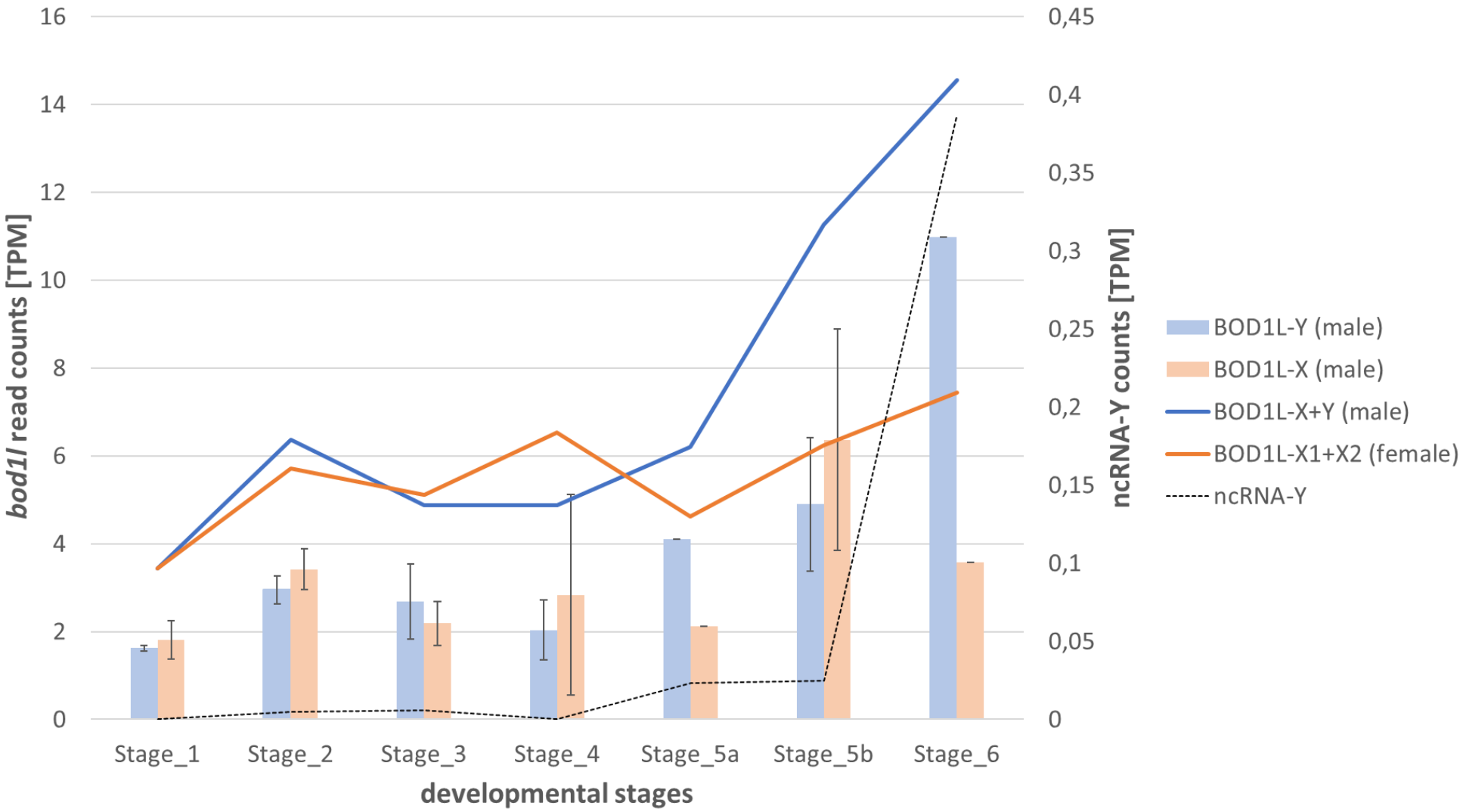
1168 **Supplementary Table 3:** Signal of the AI-based software i-enhancer, reaching from
1169 0 to 2, predicting strong enhancer properties of the Y-specific 5'-end of *bod1l*, including
1170 a G-quadruplex-rich region on the non-template strand, overlapping with the ncRNA-
1171 Y.

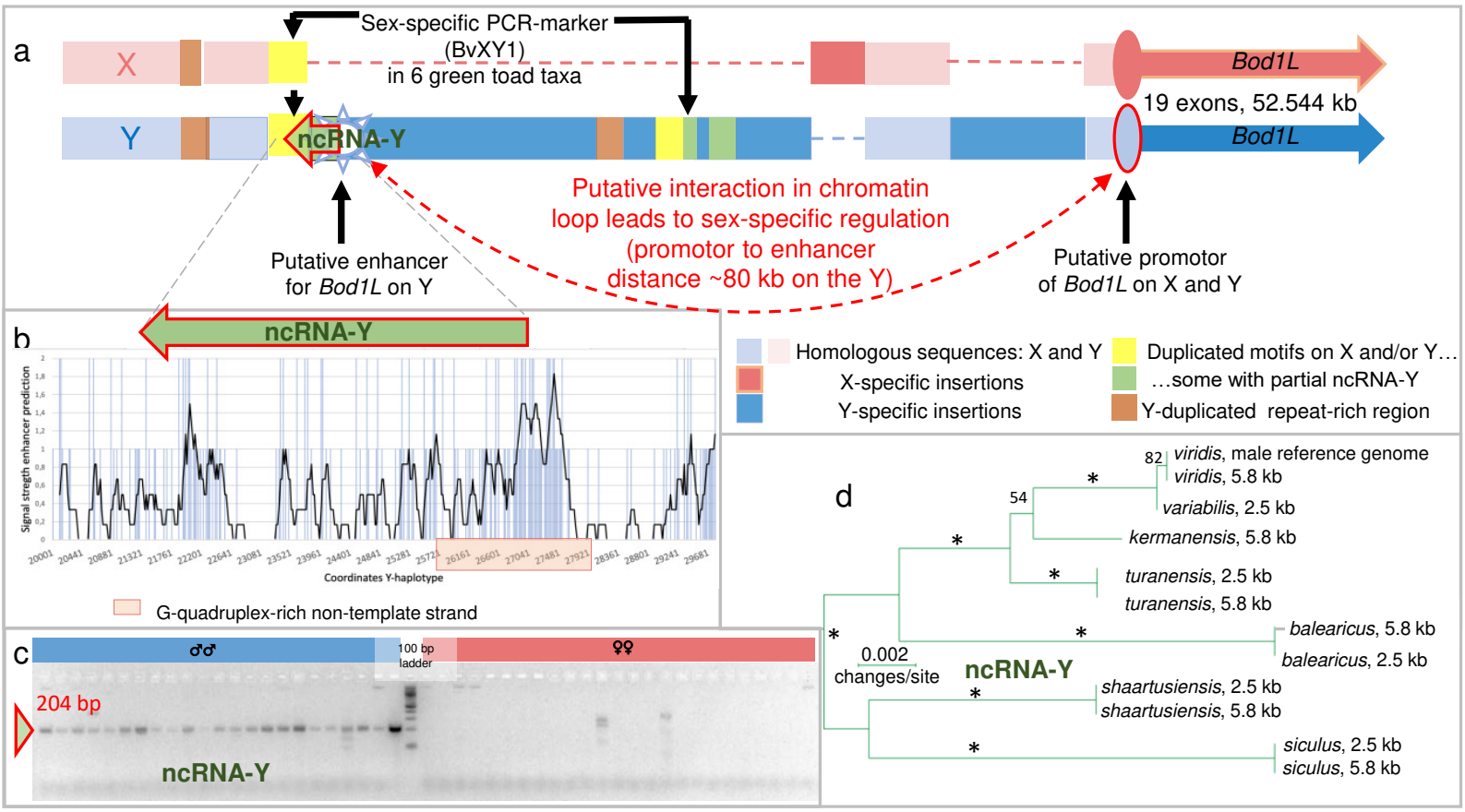
1172

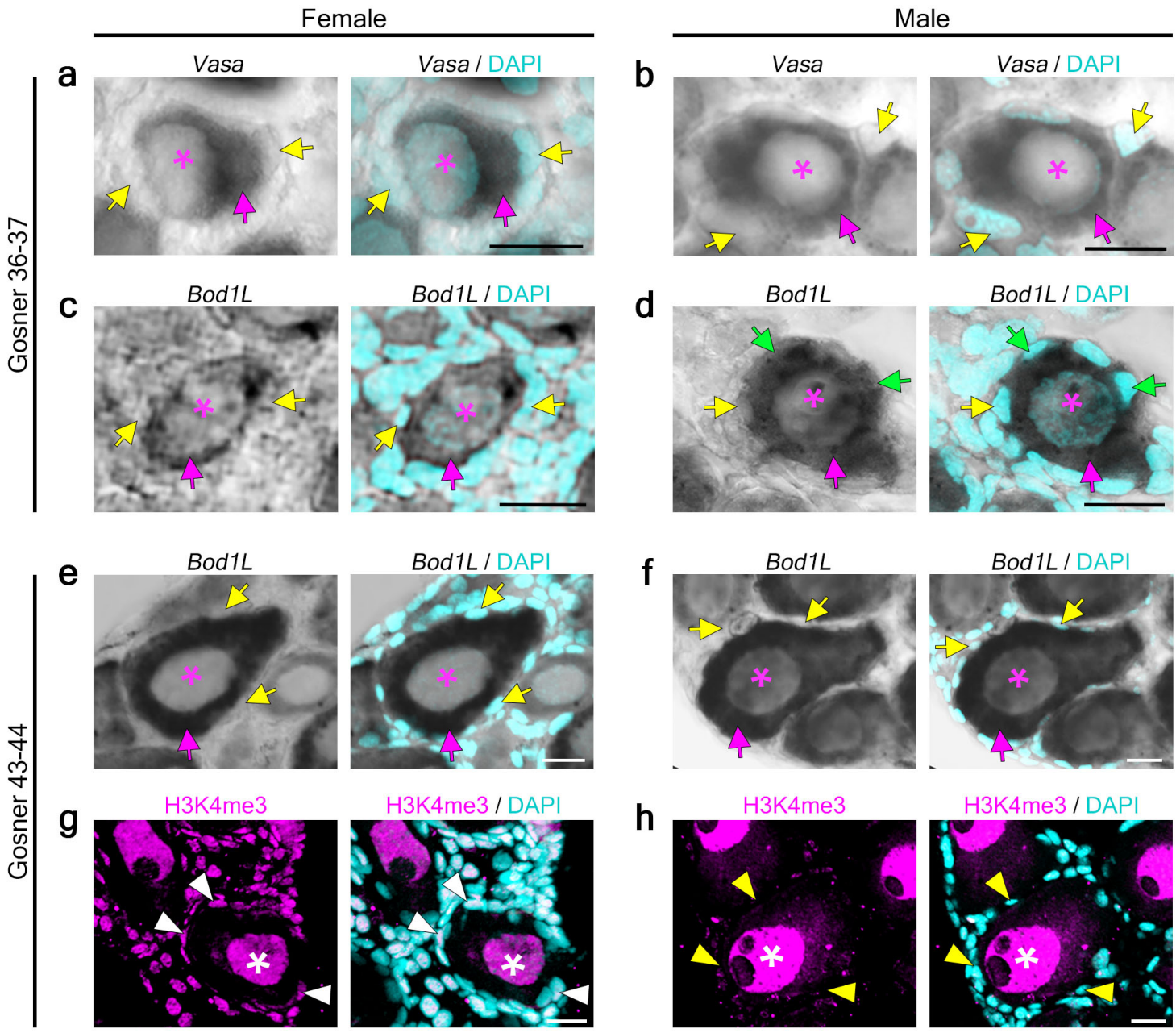
1173 **Supplementary Texts**

1174 **Supplementary Text 1:** Potential impact of X- *vs*. Y-*bod1l* coding-sequences in *B.*
1175 *viridis*.


a
b
c
d
Scf1: SNPs heterozygous in male/homozygous female (X/Y)
566,783,058-
566,848,882 bp:
bod1L
Dmrt1
Amh
100
80
60
40
20
0
0
100.000.000
200.000.000
300.000.000
400.000.000
500.000.000
600.000.000
e
♀♀
Ladder
♂♂
Ladder
300 bp
200 bp
100 bp
50 bp
i
ii
iii
iv
v
vi
i
ii
iii
iv
v
vi
f
g
h
2.0
changes/site
calibration
2.5 (1.55-3.55) My
calibration
1.8 (0.63-3.5) My
calibration
18 (18.2-29.5) My
7.78
18.66
3.81
4.26
vi 2n
3n
4n
4n
v
iv 2n
iii 2n
2n
ii
2n
3n
3n
i 2n
2n
2n
Ploidy
20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
Divergence Mya


16
14
12
10
8
6
4
2
0
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
bod1l read counts [TPM]
ncRNA-Y counts [TPM]
Stage_1
Stage_2
Stage_3
Stage_4
Stage_5a
Stage_5b
Stage_6
developmental stages
BOD1L-Y (male)
BOD1L-X (male)
BOD1L-X+Y (male)
BOD1L-X1+X2 (female)
ncRNA-Y


a
Sex-specific PCR-marker
(BvXY1)
in 6 green toad taxa
X
Y
ncRNA-Y
Bod1L
19 exons, 52.544 kb
Bod1L
Putative interaction in chromatin
loop leads to sex-specific regulation
(promotor to enhancer
distance ~80 kb on the Y)
Putative enhancer
for Bod1L on Y
Putative promotor
of Bod1L on X and Y
b
ncRNA-Y
Signal strength enhancer prediction
Coordinates Y-haplotype
G-quadruplex-rich non-template strand
Homologous sequences: X and Y
X-specific insertions
Y-specific insertions
Duplicated motifs on X and/or Y…
…some with partial ncRNA-Y
Y-duplicated repeat-rich region
c
♂♂
100 bp
ladder
♀♀
204 bp
ncRNA-Y
d
82
viridis, male reference genome
viridis, 5.8 kb
54
variabilis, 2.5 kb
kermanensis, 5.8 kb
turanensis, 2.5 kb
turanensis, 5.8 kb
balearicus, 5.8 kb
balearicus, 2.5 kb
0.002
changes/site
ncRNA-Y
shaartusiensis, 2.5 kb
shaartusiensis, 5.8 kb
siculus, 2.5 kb
siculus, 5.8 kb



i

		Expression in somatic cells		Expression in germ cells	
		Male	Female	Male	Female
Gosner 36-37	*Bod1L*	+	-	+	+
Gosner 43-44	*Bod1L*	-	-	+	+
	H3K4me3	low	high	high	high