A candidate sex determination locus in amphibians, evolved by structural variation between X- and Y-chromosomes

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 sex-determining gene is known in >8650 species, the *dmrt1*-paralogue (*dm-w*) of female-heterogametic clawed frogs (*Xenopus*; ZW♀/ZZ♂). Combining chromosome-scale female and male genomes of a non-model amphibian, the European green toad, *Bufo*(*tes*) *viridis*, with ddRAD- and whole genome pool-sequencing revealed a candidate master locus, governing a male-heterogametic system (XX♀/XY♂). Targeted sequencing across multiple taxa uncovered structural X/Y-variation in the 5'-regulatory region of the gene *bod1l*, where a Y-specific non-coding RNA, only expressed in males, suggests that this locus may initiate sex-specific differentiation. Developmental transcriptomes and RNA *in-situ* hybridization show timely and spatially relevant, sex-specific ncRNA-Y and *bod1l*-gene expression in primordial gonads with coinciding differential *H3K4me*-methylation in pre-granulosa/pre-Sertoli cells, pointing to a specific 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, which are prevailing in the tree of life[5]. In vertebrates, 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)[6], (ii) recurrent X-Y recombination, e.g. by occasional sex reversal to XY-females[7], and (iii) "jumping master genes" with conserved genetic triggers translocating to other chromosomes[8].

Conserved poorly or non-differentiated sex chromosomes are known in several clades of poikilothermic vertebrates, e.g., in sturgeons, pikes, and skinks[9-11].

The majority of amphibians possess undifferentiated sex chromosomes with either female (ZZ/ZW) or male (XY/XX) heterogamety[12,13]. Several examples of cytogenetically differentiated sex chromosomes[13,14.], a female W0/00 male sex determination system[15], and several amphibians with multiple sex chromosomes have been identified in both laboratory strains[16] and natural populations[17]. Notably, to date, in >8,650 species of amphibians[18], 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*[12,19]) and some related species[20]. The *Dm-w* gene arose after (and perhaps in response to) tetraploidization, is essential as recently shown in knock-out females[21], and has been lost secondarily in some tetraploid species[22].

True toads, Bufonidae, share large genomes, conserved karyotypes[23] and undifferentiated sex chromosomes[24]. 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[25]. Using sex-reversal experiments, castrating adult males and regrowing functional ovaries from Bidder's organs by Harms[26], early on, Ponse[27] inferred female heterogamety (ZW) in common toads (*B. bufo*). This was also cytogenetically deduced in related *B. gargarizans*[28], and cane toads (*Rhinella marina*[29]). With few exceptions, a meta-analysis of sex-differential F1-hybrid fitness of interspecific crosses in bufonids likewise supported widespread female heterogamety[30].

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[24,31], sex-reversal experiments in '*B. viridis*' from Asia Minor[32] suggested an XY-system. Molecular evidence for male heterogamety resulted from sibship analyses using microsatellites and nuclear sequence markers[33-35], 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 X/Y-exchange, in the pseudo-autosomal region, is suggested at evolutionary time scales[33]. However, phylogenetic analysis

revealed X- and Y-alleles of *dmrt1* to cluster by species and not by gametologue, suggesting X/Y-recombination involving *dmrt1*, which rejected its role as master sex determiner[35]. Details of sex determination have not been examined in allopolyploids green toads[36].

In mammals (6,649 species[37]), genome assemblies are available for about one third (2740[38]). Although there are more species described in amphibians (>8650[18]), owing to large genome sizes and excessive repetitive regions, to date only >80 amphibian genomes have been assembled, including 56 out of 7627 Anura, 19 out of 805 Urodela (Caudata) and 9 out of 221 Gymnophiona[18,38]. However, profiting from long-read- and 3D-sequencing technologies, amphibian genomics is on the rise.

Here, we provide a chromosome-scale genome of a non-model amphibian, the European green toad, *Bufo*(*tes*) *viridis*, and identify a specific candidate master sex determining locus - present in multiple related green toad taxa - that evolved by structural variation, involving a Y-specific ncRNA, in the regulatory region between X and Y chromosome.

Results

Chromosome-scale genome assembly of *B. viridis*

We obtained a 3.8 Gbp chromosomal-level genome assembly using Oxford nanopore (ONT) long-read (N50 = 11.6 kbp; sequencing coverage 21x) and Illumina short-read data (2x150 bp; sequence coverage 47x) for the European green toad (Fig.1a,b). N50 contig size of the final assembly was 1.9 Mbp. Eleven chromosomes (scaffold N50 = 470 Mbp; consisting of 94% of the contigs total length), consistent with the karyotype of *B. viridis* (Fig.1c), were constructed using chromosome conformation capture (Hi-C) short reads (318 Gbp). Many unplaced contigs contain a known long satellite DNA family[39]. Comparing the chromosomes to the *B. bufo* assembly revealed genome-wide synteny and collinearity with few inversions, despite significant differences in assembly size (VGP's aBufBuf1.1 = 5.0 Gbp; Fig.1b). *De novo* analysis revealed 68.5% repeat content in *B. viridis*. Using published data from related species and RNAseq data, we annotated 21,605 protein coding genes in the genome. BUSCO (Geneset vertebrata_ODB9)-scoring of protein predictions revealed 96.4% complete coding sequences and a low missing gene fraction of 1.7% (C:96.4% [S:94.2%, D:2.2%], F:1.9%, M:1.7%, n:2586).

ddRADseq in *B. viridis* genetic families

We first attempted to identify the sex determination region by double digest restriction-site associated DNA sequencing (ddRADseq). In three genetic families of European green toads, we analyzed a total of 443,378 markers, present in at least one individual with a depth higher than 4. Of these, 96 were significantly associated with male sex ($p < 0.05$; Bonferroni correction), but none was found in all males (Supplementary Fig. 1), probably due to high stochasticity of random sequencing of large genome fractions. When mapped to the *B. viridis* genome, 64 of these 96 markers were located along the entire chromosome 1 (scf1) (Supplementary Fig. 2), one on chromosome 7 (scf7), five on unplaced scaffolds, and 26 were not aligned uniquely or with low quality (mapping quality, MQ < 20). Overall, this strongly supports a male (XY) genetic sex determination system on chromosome 1, however, with a very small sex-specific region that is not fully covered in each individual, and sex-linkage of most ddRAD markers is not evolutionarily conserved due to X-Y recombination, in the huge pseudo-autosomal region (PAR) of this chromosome 1. While this approach powerfully corroborated previous microsatellite-based results[33], it did not provide a sex-diagnostic marker.

Pool-seq yields a cross-amplifying marker in multiple taxa

To narrow down the sex determining region on scaffold 1, we used whole genome pool sequencing (from here: “pool-sex”), involving 22 female and 22 male phenotypically-sexed adult green toads from multiple European green toad populations (Supplementary Table 1), chosen to be closely related according to phylogeographic information[40,41] and presumably sharing the same sex chromosome, based on microsatellites[33-35]. Mapping pool-sex short reads to the reference, genome-wide revealed a single peak in a region on scaffold 1 (scf1), where the number of clustered heterozygous SNPs (N = 103) in males surpassed about nine times the maximum number of clustered heterozygous SNPs (N = 12) in females (Fig.1c, Supplementary Table 2; Supplementary Fig. 3). This result was later confirmed quantifying male-heterozygous SNPs on scaffold 1 of the male genome assembly as well as male-specific pool-seq coverage on the male genome (see below and Fig. 1c).

This supports a male-heterogametic (XY) system, involving a short 66 kb region (scf1:566,783,058-566,848,882 bp). This locus encompasses one third of the 5’-

upstream-region and two thirds of the gene *bod1l* (bufVir1.4895.2, biorientation of chromosomes in cell division 1 like 1). Transcriptomes and cDNA-based PCR-products of males of *B. viridis* exhibited consistent heterozygosity of exons 7 to 10 (scf1:566,831,904-566,833,300), while females were always homozygous.

The heterozygous 5'-region exhibited several INDELs between X and Y haplotypes, enabling primer design. In females, BvXY1 amplified a single PCR-product - visible as one band on the gel, and two to three bands in males, widely consistent with an XX♀/XY♂-system, but suggesting multiple Y-specific signals in males, which might be sex-chromosomal duplications, nonspecific PCR-amplicons or even PCR-recombinants. Of note, BvXY1-primers amplified not only sex-specific products in individual *B. viridis* (sensu lato) males and females of the pool-sex-approach but also in multiple phenotypically-sexed males and females of this and four additional taxa (*B. balearicus*, *B. siculus, B. turanensis, B. shaartusiensis*), comprising a monophyletic radiation, probably >3 Mio. years old[36] (Fig. 1e-h). The sex-diagnostic PCR-marker BvXY1 amplified a single specific 100-200 bp band in females of six green toad taxa, but 2-3 bands in their males, one in females' size, presenting the X-copy. The additional male bands with the initial BvXY1-primers yielded identical or similar sequences indicating duplicates on the Y-homologue.

Synteny and FISH-TSA of *bod1l* dismiss duplication with translocation

New sex determining genes evolve by allelic diversification or from gene duplications[42]. To elucidate the situation in *B. viridis*, we examined synteny and localization of the candidate locus. *Bod1l* lies between the genes *Cpeb2* and *nkx3-2* (NK3 homeobox 2), on the long arm of scaffold 1, an evolutionarily conserved gene order likewise found in *B. bufo*, *Rana temporaria* and *Xenopus tropicalis*, suggesting synteny throughout anurans.

Fluorescence *in-situ* hybridization with tyramide signal amplification (FISH-TSA) on metaphase chromosomes of a male *B. viridis* revealed two signals, one on each homologue, for a *bod1l*-specific probe at a position on the distal part of the short (p) arm of chromosome 1 (Fig. 1f). Two (and not three) signals suggest that this gene exists in two copies in XY-males of *B. viridis*, implying that both heterozygous sequences evolved by structural diversification[42] rather than translocation of a male-specific duplication, as shown for *dm-w* of female *Xenopus laevis*[19]. This result is further supported by the pool-seq and long-read male coverage (see below).

Targeted sequencing in multiple related species sharing the sex marker

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

Y-chromosomal structure of *B. viridis* from long-read WGS

After first working with the female reference genome, X-haplotype and *bod1l*-coding region, we then moved the focus back to the 5'-end of *bod1l* and the BvXY1-vicinity on the Y-haplotype. To elucidate this Y-chromosomal structure, we reconstructed this region, which showed about 15x coverage (haploid), from ca. 30x whole genome long-reads of a single male *B. viridis* and aligned it to the X-copy of the *bod1l*-region. Female-to-male pool-seq-coverage of X- and Y-haplotypes yielded the expected diploid female and haploid male coverage (Supplementary Fig. 5 and 6). Additionally, improved base-calling and haplotype-aware correction algorithms became available and enabled us to also assemble a whole male genome from that data, which supported the results of the manual approach (Supplementary File 2).

In the *bod1l*-region, the confidently assembled Y-haplotype-specific sequence comprises about 55 kb with some greater-than-zero female pool-seq-coverage, explicable by repeat-enriched fragments, not assignable to X or Y (Supplementary Fig. 7). The male long-read whole genome assembly revealed several structural changes between X- and Y-haplotypes in the 5'-*bod1l*-region. The *B. viridis*-Y contains large insertions, mainly stemming from repetitive elements, and a non-coding RNA (ncRNA, see below). None of the so-called 'usual suspects', genes that often acquire sex-determination function in vertebrates[1], was detected in this region.

Larval gene expression in sibling *B. viridis* tadpoles from multiple localities

Pool-sequencing of genomic DNA in *B. viridis* had revealed consistent heterozygosity in males for *bod1l*-exons 7 to 10. Accordingly, RNAseq revealed sex-differential expression (scf1:566,831,904-566,833,300), where heterozygous *B. viridis* males showed 1 SNP in exon 7, and 14 SNPs in exon 10, while females were always homozygous (confirmed by cDNA-based PCR-products). To test whether *bod1l*-products are differentially composed after transcription of X- and Y-copy, we analyzed splice sites in developmental transcriptomes, including the sex determination period. The absence of significant sex-specific differences excluded alternative splicing. However, this analysis allowed manual improvement of the *bod1l*-gene model.

To study the expression of the candidate sex determination gene (*bod1l*),the adjacent ncRNA-Y and ca. 80 other genes (Supplementary File 6) playing roles in sexual development and/or sex determination in other vertebrates ('usual suspects'[1]), we then examined larval transcriptomes of six sibling groups, in total 54 individual toads from six locations, ranging from Gosner[43] stage 23 (day 10 after fertilization) until Gosner stage 46 (the last about 6 weeks after metamorphosis) as well as from two adult gonads.

Since gonadal differentiation is histologically detectable only at or after metamorphosis completion (>Gosner 46[25,35,43]), the critical period of genetic sex determination lies earlier, within our sampling period. In Gosner stage 46, 6 weeks after metamorphosis, we could anatomically distinguish between Bidder's organ and early testis, proofing sexual differentiation. Until Gosner stages 34-38 (Fig. 2), overall *bod1l*-expression varies in *B. viridis* between males and females (. From Gosner-stages 38 to 43, other genes, specific for male sexual development, like *sox9*, *sox10*, *sox17*, and *sf1*, show higher expression in males than in females. Other such genes change their expression distinctly after Gosner stage 38, like *ck18* (bufVir1.18618.1) or *11ßHD* (bufVir1.16559.1) or - in our series - were detected only in Gosner stage 46 and subadult testes (*dmrt1, amh*). Female-specific genes also changed their expression around Gosner stages 34-38; *wnt4*, *ihh*, *foxL2*, *ptch1*, *gli3*, *fshr* from Gosner stage 43 or 45; however, their expression was stronger in males than females. Expression of *gdf9*, *cyp19a1a*, and *cyp17a1* was detectable only in Gosner 46 and adult ovaries, with *cyp19a1a* and *cyp17a1* presenting good markers for the

timing of gonadal sexual differentiation in anurans. *Bod1l*-expression differences between females and males (Fig. 2) lack statistical significance but occur at the right developmental period (Gosner stages 27-44), as required for a sex-determining gene. Early male transcriptomes (Gosner stages 23-44) contain also the bufonid-specific Bidder's organ, considered similar to a rudimentary ovary, that develops adjacent to the anterior primordial gonads[25]. 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[27] (reviewed[25]). In green toads, Bidder's organ develops in males but barely in females[25]. Given that the "ovary-like" primordial Bidder's organ is contributing tissue to our RNAseq-analysis of male tadpoles until Gosner stage 44, we assume a "female-bias" in males' transcriptomes (Gosner stages 23 to 44). In conclusion and in line with the AmpliSeq-data from multiple species, only sharing the BvXY1-marker across taxa, we consider the male-only expression of ncRNA-Y and the quantitative differences in additive *bod1l*-expression between males and females (Fig. 2) as sufficient to trigger green toad sex determination. Of note, there is a trend of greater Y-haplotypic expression of *bod1l* between Gosner-stages 38 and 43 (coinciding with an increase of the ncRNA-Y expression, next paragraph) and a distinctly greater additive *bod1l*-XY-expression in males than XX-expression in females (Fig. 2, Supplementary Fig. 8). *Bod1l*-Y also shows the highest expression in testis compared to other male organs (heart, brain, liver).

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. 9). 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[44]. With ca. 80 kb, the putative enhancer sits in a reasonable distance to the *bod1l*-promoter (Fig. 3a,b)[45]. We also tested 200 bp-

"sliding-windows" of the ncRNA-region for their *in-silico* potential as enhancers of *bod1l* (Supplementary Table 3; Supplementary Fig. 10). The most stringent threshold of 0.9 still predicted larger parts of the Y-specific region to function as strong enhancers, intensified by numerous G-quadruplexes (>25 alone in intron 1 of ncRNA-Y), known to accumulate at promoters and enhancers[46]. As shown by PCR using cDNA, this long ncRNA-Y is also expressed in skin, liver, brain and heart of males after metamorphosis as shown using cDNA, and exhibits simultaneous expression of the Y-haplotype of *bod1l* with the strongest signals in testis (Fig. 2). Due to its low expression level, it is only detectable in high-coverage transcriptomes of the primordial male gonads and adult testes but not in heart, brain and liver high-coverage transcriptomes. It comprises 5.8 kb on the gDNA level and can be male-specifically PCR-amplified in all those six taxa containing the BvXY1-marker and suggests a common ancestral feature (Fig. 3d).

Spatially different gene expression in the developing gonads

For a sex-determination gene, not only the accurate timing of its expression but also its organ- and cell-specific spatial expression is essential. We provide evidence by RNA *in-situ* hybridization in histological sections of gonads in Gosner stages 36-37, where *bod1l* shows sex-specific differential expression. In female gonads, *bod1l*-expression was evident in *vasa*-positive primordial germ cells (developing oogonia) but absent from surrounding somatic pre-granulosa cells (Fig. 4a,c, Supplementary Fig. 11). In contrast, *vasa*-positive male primordial germ cells (developing spermatogonia) as well as a majority of adjacent somatic pre-Sertoli cells exhibited *bod1l*-expression (Fig. 4b,d, green arrows). In Gosner stages 43-44, at the beginning of metamorphosis, *bod1l*-expression in females remained similar to Gosner 36-37. In males, however, *bod1l*-expression was no longer detectable in somatic pre-Sertoli cells (Fig. 4e,f; Supplementary Fig. 11). This suggests a dynamic regulation of *bod1l* in early male and female gonads, with expression in somatic cells exclusively in male but not female gonads. This precedes the differences observed by RNAseq from trunk sections in Gosner stage 43-44, when quantitative differences of sex-specific *bod1l*-expression occur (Fig. 2).

It has been shown that *Bod1l* interacts with *Setd1a* to facilitate tri-methylation of histone H3 at lysine 4 (H3K4me3[47]), with similarities in eukaryotes from yeast to mammals[48]. To examine its status in male and female gonads, we performed

H3K4me3-immunostaining on Gosner 43-44 gonads (Fig. 4g,h). In females, gonads exhibited distinct H3K4me3-signals in nuclei of both germ and somatic pre-granulosa cells (Fig. 4g, white asterisks and arrowheads, respectively; Supplementary Fig. 11). In contrast, H3K4me3 was strongly detected in male germ cell nuclei (Fig. 4h, white asterisks) but hardly detectable or absent in nuclei of somatic pre-Sertoli cells (Fig. 4h, yellow arrowheads). In Gosner stage 46, six weeks after metamorphosis, *bod1l* was expressed (RNAseq, PCR of cDNA) in both sexes in several somatic tissues (brain, heart, liver, intestine) as well as in gonads.

XY-phylogeny in six Eurasian diploid taxa from targeted sequencing

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

Discussion

Our data in *B. viridis* and five related diploid green toad taxa are in accordance with a locus that comprises a ncRNA and structural changes in the 5'-regulatory region of a gene (*bod1l*), as well as sex-specific ncRNA-expression in heterozygous (XY) males *vs.* homozygous (XX) females, accompanied by strong phylogenetic and poolseq-supported signal of local recombination arrest. Sex-specific expression occurs at the right time, with male-restricted expression of the ncRNA from the Y, much earlier than histologically detectable differentiation, and *bod1l*-expression in the target organs (primordial gonads), and therein in the germ and surrounding somatic cell populations (in male pre-Sertoli but not in female pre-granulosa cells). These properties suggest that the region containing ncRNA-Y and *bod1l* presents a

candidate sex determination locus. Future evidence (e.g., using CRISPR/Cas9) is desirable for its better mechanistic understanding. A single genome-wide peak of heterozygosity between female and male revealed by pool-seq and coverage differences on whole genome assemblies of both sexes (Fig. 1c), along with evolutionary conservation of the structural XY-differentiation in several Eurasian taxa sharing the Y-specific ncRNA, support its candidacy as the Palearctic green toad sex determining locus. The expression of the ncRNA-Y and *bod1l* in Bidder's organ of the differentiating male gonad (Gosner stage 46) may contradict the view that Bidder's organ in this XX/XY system can be simply interpreted as a 'proto-ovary'.

The only known amphibian master sex determining gene, *dm-w* of clawed frogs (*Xenopus*), governs female-heterogamety (ZW♀/ZZ♂) and evolved by gene duplication as a *dmrt1*-paralog, involving neofunctionalization of a non-coding part of a DNA-transposon[12,19-22,49]. Here, we provide evidence for a sex-locus in a male heterogametic bufonid toad system (XX♀/XY♂), evolved by structural diversification in a common ancestor, as shown by the associated ncRNA-Y-phylogeny (Fig. 3c). To our knowledge, this reveals only the 2nd master sex determining locus in amphibians.

Our data suggest either a sex-determining function of the ncRNA-Y itself or its functional interaction with *bod1l*, e.g., as an enhancer[50,51].

Long non-coding RNAs (lncRNAs) play important roles in gonadal development and sex differentiation[50,52], and thus the ncRNA-Y might suffice as a master sex determinator, although, so far, we have not detected homology with genes known to play roles in sexual development. We rather consider a regulatory effect on the neighboring *bod1l*. The ncRNA-Y, harboring the partially duplicated BvXY1, contains a neighboring signal for a G-quadruples-rich enhancer towards its 3'-end. This is a mechanistic hint that the enhancer modifies *bod1l*-expression in XY-males (Fig. 3), as shown by timely and spatially differential expression in germ and pre-Sertoli but not in pre-granulosa cells (Fig. 4).

There are also several, partly interlinked pathways by which spatially and/or timely differential *bod1l*-expression may determine sex in green toads.

(i) Sex determination of vertebrates differs regarding the timing of primordial germ cells to enter meiosis[53], with females entering meiosis earlier than males, exhibiting an arrest. *Bod1l* is well-known to maintain genomic integrity at sites of DNA-replication damage[54]. *Bod1l* interacts towards its N-terminus with the histone

methyltransferase *Setd1a*, which triggers the histone chaperone function of *Fancd2*[55], part of the Fanconi anemia group (*Fanc*), that plays roles in zebrafish sex reversal[56]. Thus, differential *bod1l*-expression might cause a compromised DNA repair system by shifting sex determination towards males.

(ii) Interaction of *Bod1l* with *Setd1a* and H3K4me3 refers to tri-methylation on lysine residue 4 in the amino-terminal tail of histone H3, is associated with potentially active promoter regions[57] and linked to double-strand breaks, essential for meiosis 1 crossing-over, ensuring chromosome segregation[58,59]. Thus, the differential *bod1l*-expression might enable double-strand breaks as a precondition for the earlier start of recombination in females but not in males. (iii) H3K4-methylation in mammals[57] is associated with *de novo* DNA-methylation in germ cells, and essential for maintaining transcription in the growing oocyte and for triggering zygotic genome activation[60], with XY-mice deficient for the H3K9-demethylating enzyme *Jmjd1a* showing increased H3K9-dimethylation and decreased activating marks of H3K4-trimethylation across the *Sry*-locus, leading to a high frequency of sex reversal[2].

According to Dupont and Capel[61] “prior to sex determination” in placental mammals, bipotential supporting gonadal cells of both genetic sexes maintain identical transcriptional programs, allowing commitment to either Sertoli (testes) or granulosa (ovary) cell fate. In pre-Sertoli and pre-granulosa cells, bivalent promoters of key sex determining genes (e.g., *Sox9*, *Fgf9*, *Wnt4*, *Foxl2*) have both, H3K27me3 and H3K4me3 marks[61]. After sex determination by the master sex determination gene *Sry*, promoters to be activated in that sex lost the repressive H3K27me3 mark, while genes to be silenced maintained their bivalency[62]. This suggested that (a) sex-determining genes are poised for activation (or repression) at the bipotential stage, (b) gene activation involves removal of H3K27me3 marks, while (c) repression of the alternative supporting cell fate requires maintenance of existing H3K27me3 repressive marks rather than *de novo* deposition of these marks after commitment to the Sertoli or granulosa program[61].

One hypothesis is that differential *bod1l*-expression (reminiscent of *Sry*) might determine sex by changing H3K4-methylation in green toads during early gonadal development, prior to metamorphosis.

To test the hypothesis that *bod1l-expression* coincides in time and space with changes of H3K4-methylation in developing green toad gonads, we used H3K4me-antibodies together with RNA *in-situ* hybridization (Fig. 4). In Gosner stage 43-44,

slightly prior to morphological differentiation, we found a strong difference between male and female gonadal somatic cell H3K4-methylation. This suggests that spatially different *bod1l*-expression results in sex-specific methylation patterns that initiate sex-specific gonadal differentiation. However, opposite to the situation in mammals, H3K4me3 is weaker in somatic cells of green toad males than those of females (Fig. 4), while increasing H3K4-methylation allows *Sry*-expression activation (male determination). Another major difference is that the primordial gonad of bufonid toads, specifically in green toad males, includes the “ovary-like” Bidder’s organ.

Given the multiple potential links between *bod1l*-expression and sex-specific meiotic stages shown above, we hypothesize that this gene may function as a suppressor of the female sex determination pathway – a scenario found in many vertebrates[63]. Following differential *bod1l*-expression, in the bipotential gonad, ovarian pre-granulosa cells transdifferentiate to testicular pre-Sertoli cells.

In contrast to the only known amphibian master sex determining gene in clawed frogs (*Xenopus*), the green toad sex-locus with ncRNA-Y and *bod1l* did not evolve by gene duplication but structural diversification in the 5’-region of the gene, namely by Y-specific sequence insertion. 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[42].

Methods

Ethics and animal welfare statement

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

developmental stages (table) by Gosner[43]. Tadpoles were reared in a 12/12 h light/dark cycle at constantly 22 ± 1 °C in sufficiently aerated and regularly cleaned tanks. Weekly monitored water parameters comprised: dissolved oxygen, nitrate, ammonium, pH, conductivity, and hardness and kept as described[64]. Tadpoles were fed SeraMicron (Sera, Germany), later supplied with TetraMin (Tetra, Germany). To imitate natural conditions at metamorphosis, animals were transferred to glass terraria at Gosner stage 46. Thus, the tadpoles were kept according to approved practice (permit ZH 114 to IGB), including the provision of appropriate tank size, sufficient rate of waterflow, natural photoperiod, *ad libitum* food supply, as well as temperatures within the species' thermal tolerance range. This ensured that no pain, suffering, distress or lasting harm was inflicted on the animals. According to the permit G0359/12, tadpoles and juvenile (and two adult) toads were euthanized by an overdose of Tricaine PHARMAQ 1000 MG/G (MS222) (concentration: 500 mg/L)) since this does not incur pain, suffering or distress.

Chromosome-scale female genome assembly

The female *B. viridis* reference genome was sequenced in-house at IGB using ONT long reads (MinIon device, SQK-LSK109 kits and R9.4.1 flow cells) and Illumina short reads (Novogene, Europe). The long reads and pre-assembled short reads IDBA-UD assembly[65] were assembled using the WTDBG2.2 assembler[66]. Sequences were error-corrected using two rounds of Flye polish with long reads[67] and four rounds of VCFconsensus using short reads[68]. The chromosome-scale assembly was achieved using Arima HiC-data and the JUICER/3DDNA package and manual curation by juicebox[69-71]. *De novo* repeat analysis of the assembly was performed by the RepeatModeler / RepeatMasker package. Protein coding genes were annotated using RNAseq data of *Bufonidae* as well as Hyloidea transcriptomes and more distantly related amphibian proteins, available at NCBI. A reference genome browser was created http://genomes.igb-berlin.de/Greentoad/.[72]

Assembly of Y-haplotype

Several iterative approaches were used to extract Y-haplotypic sequencing reads of *B. viridis* GR57 male[73] long read data (Oxford Nanopore Promethion platform; Novogene). First, we mapped the reads to the BvXY1-marker sequences to separate Y reads of the region. Second, we performed a FLYE[67] assembly of the male data,

identified several contigs with variation compared to the X haplotype (*bod1l*-region), and extracted the corresponding reads. The Y-specific reads were assembled, then added to the female assembly and all male reads were mapped to the combined sequences, which allowed finding more reads matching better to Y than to X contigs. This procedure was iteratively repeated several times until the Y-haplotype and its structure could be assembled. Pool-seq data was again mapped to the female assembly with added Y-haplotype sequences by MINIMAP2[74] showing the expected coverage differences between female and male pools. A dot plot comparison of X versus Y haplotype involved the corresponding function of the NCBI Blast website.

Assembly of a male genome

Data that was used for the manual assembly of the Y-haplotype (see above), served also to perform a complete de novo assembly of a male B. viridis individual. To improve assembly results RAW data was first base-called by the most up-to-date basecaller for ONT data (dorado-0.5.1-linux-x64 with dna_r9.4.1_e8_sup@v3.6 model). Then the reads were corrected using the haplotype-aware read correction method, implemented in PECAT[75]. The corrected reads and the prior flye assembly contigs (see above) were assembled using wtdbg2. Contigs were corrected by one round of Flye polish (with corrected long-reads) and two rounds of VCFconsensus (with male poolseq data). Male Hi-C data was mapped to the contigs using CHROMAP[76] and chromosomal scaffolds were computed by YAHS[77]. Some unplaced contigs could be placed into the YAHS scaffolds using RAGOUT2[78] and synteny analysis with the female genome. The assembly was scored using BUSCO analysis and had highly similar scores as the female assembly.

Double-digest RAD sequencing (ddRADseq)

DNA of three genetic families consisting for one data set of both parents (buccal swabs) and 40 sexed mostly phenotypically offspring and for two other sample sets of 25 or 15 siblings, was extracted using the DNeasy Tissue Kit or the BioSprint robotic workstation (QIAGEN), following the manufacturer's protocols. Double-digest RADseq libraries were prepared according to this study[79]. In brief, genomic DNA digestion by two restriction enzymes (*Sbf*I, *Mse*I) and subsequent ligation of fragments to adapters containing unique barcodes (4–8 bp) was followed by library-PCR for 20 cycles for only different cut-sites containing fragments, and a gel-based

size selection, isolating the 400–500 bp fraction. RADseq libraries were single-end sequenced on an Illumina HiSeq 2500, at a 20x coverage.

Raw ddRAD reads were demultiplexed with STACKS *process_radtags* version 2.55[80], using the enzymes sbfI and mseI (*--renz_1 sbfI --renz_2 mseI*) and the parameters *-c* (clean data), -r (rescue barcodes and RAD-Tag cut sites), and -q (discard reads with low quality scores).

Demultiplexed reads were used as input for the RADSEX workflow[81]. First, a table containing the number of occurrences of each unique sequence in the reads file in each individual was generated with *radsex process*, using default parameter values. Then, the distribution of markers between males and females was computed with *radsex distrib* using a minimum depth to consider a marker present in an individual of 5 (--min-depth 5). Markers were aligned to the genome of *B. viridis* (bufVir1) using *radsex map* using a minimum depth of 5 (--min-depth 5) and other parameters set to default. Results were plotted with the *SGTR* package, with the entire process provided as a Snakemake workflow[82].

Poolseq approach using phenotypically sexed toads: 'poolsex'

Of 22 safely sexed adult toads of each sex from multiple populations in the range of the European green toad taxa (*B. viridis viridis* and *B. viridis variabilis*), using buccal swabs or tissues genomic DNA was extracted using the DNeasy Tissue Kit or the BioSprint robotic workstation (QIAGEN), following the manufacturer's protocols. Individual DNAs were quantified using the QuantiFluor® dsDNA System kit (Promega), adjusted and equimolarly mixed to yield a female and a male pool. Whole genome library preparation for 300 bp fragments and whole genome sequencing was performed at NOVOGENE (UK), on a NovaSeq PE150 to yield 131 and 140 Gbp of data for male and female pool, respectively. Reads were mapped to the reference genome using the short-read mode of MINIMAP2[74] (parameters -K 2000M -I 100G -a -x sr -t 80, and converted to sorted bam-files by SAMTOOLS[83]. Variants (SNPs, MNPs, INDELs) were called from both bam files using the PLATYPUS variant caller[84] with integrated duplicate fragment reads removal (–filterDuplicates = 1) and reassembly of reads (–assembleAll = 1). The resulting vcf-file was screened for XY- or XX-specific variants using AWK-scripting and different variant read coverage cutoffs (i.e. SNP read coverage in females and males >=15; variant read count in females <=7% or >=93% of total SNP read coverage as a

condition for homozygosity; variant frequency in males between 45 - 55% of total SNP read cov. as a condition for heterozygosity; for screening ZW SNPs conditions for males and females were interchanged). Variants were clustered according to their distance in the genome (maximal distance of variants in a cluster: 9000 bp), counted using BEDTOOLS (merge, annotate[85]) and sorted by variant count. Regions with highest variant counts or densities and neighboring regions were inspected for genes related to sex determination. The same heterozygous SNP analysis was repeated using a male genome assembly (see above). Additionally, an analysis of male-specific poolsex coverage of the male genome was performed. Here, we used the bam files (MQ60 reads only) to analyze bp-level sequencing coverage using BEDTOOLS genomecov. The resulting bedgraph files for female poolseq and male poolsex were combined by BEDTOOLS unionbedg. Coverage of both poolsex datasets was normalized by their modal coverage peak (setting diploid coverage => 1.0). Then bp positions were extracted, if they had values of 0.25 - 0.75 for male- and < 0.05 for female-poolsex coverage. These genomic positions were clustered by distance and the number of male-specific covered bases in each cluster was counted BEDTOOLS(merge, annotate).

Sex-diagnostic PCR marker design

The region 566,783,058-566,848,882 bp on scaffold 1 includes several INDELs apparent between the X and the Y haplotype, allowing design of a number of primers that were primarily tested in gradient-PCRs on phenotypically sexed *B. viridis*. While several primers showed sex-specific amplification in this species, one (BvXY1) could be optimized to work across multiple species.

A collection of DNAs (extraction: see above) comprising samples of phenotypically sexed green toads of the entire radiation[36] served to test sex-specificity of all PCR markers. PCR-amplifications were carried out in 25 µL reactions containing 1.25 µL of each primer (BvXY1_F: 5'_CTGAATAGCAATTGGACACAGC_3; BvXY1_R: 5'_ATGTGAGAGCTGTTACAATGGA_3'), 14.5 µL of water, 5 µL of AllTaq buffer (5x), 0.5 µL of AllTaq polymerase (Qiagen), and 2 µL of DNA (concentration 4-20 ng/µL). Temperature regime comprised 2 min at 95°C (initial denaturation), and 38 cycles of (1 min at 95°C denaturation, 45 s at 59°C annealing, 45 s at 72°C extension), and 5 min at 72°C final extension. PCR products (4 µL) were visualized

by electrophoresis on 4% high-resolution Sieve-agarose (Biozym) gels, run for 60 min at 70 V.

To PCR-amplify either a 2.5 kb fragment or nearly the entire ncRNA-Y, comprising 5.8 kb based on gDNA on the Y-chromosome in all diploid green toad taxa with sex-specific BvXY1-markers, we used the same ingredients and volumes as above either with primers ncRNA_Y_ex1F: 5'_GGGCAGTGAAGAGAAGTCCT_3', ncRNA_Y_ex2R 5'_TTCACCATTCTCCAGCCTGT_3' for the 2.5 kb-fragment or 5.8kb_ncRNA_Y_F: 5'_GTGGTAGTTGCTGTGCCTTT_3', 5.8kb_ncRNA_Y_R: 5'_CCGGTGTGGTGTAAAATATCCA_3' for the 5.8 kb-piece. Temperature regimes involved 2 min at 95°C, and 37 cycles of (1 min at 95°C denaturation; 2.5 min at 63°C (for the 2.5 kb) or 5 min at 59.4°C (for 5.8 kb) annealing; 1 min at 72°C extension and 5 min at 72°C final extension (note extended annealing times). PCR products (4 µL) were visualized by electrophoresis on 1.5% agarose (Biozym) gels, run for 60 min at 80 V. PCR-products (21 µL) were barcoded using the NBD196 and LSK110 ONT-kits, and sequenced using the ONT-MinIon device, as described. Basecalling was performed using ONTs GUPPY version 6 and the super accuracy base-calling model (SUP-model).

Phylogenetic tree of a 2.7 kb-region at the Y-specific region boundary

Consensus sequences of 100 ONT reads of Y-specific amplicons were calculated using NGSPECIESID[86]. Homologous X-sequences were reconstructed from targeted enrichment sequencing data of females using BCFTOOLS consensus[87] for the region scf1:566,788,797-566,791,468. We also excluded regions of the sequence not covered by enrichment data by supplying a masking file through parameter "-m" of bcftools consensus. All resulting sequences were multiple aligned using MAFFT (automatic alignment parameters). The alignment was filtered by GBLOCKS[88] to remove columns with excessive missing data. IQTREE2[89] was used with parameter "-alrt 1000 -bb 1000" to calculate a phylogenetic tree. To calculate a phylogenetic tree of the upstream neighboring region (scf1:566,786,116-566,788,796) data from target enrichment for males and females was used as described above.

Phenotypic sexing in multiple species

Phenotypic sex of adult toads of multiple diploid Eurasian green toad taxa had been established during the breeding period when males exhibited secondary sexual

characters like nuptial excrescences and thorny glandular tips on the back, or were caught producing male advertisement calls and/or male release calls. Females were identified as filled with eggs, caught in amplexus, and/or while spawning, producing female release calls and/or lacking the male secondary sexual characters. During the spawning migrations, several road kill toads also allowed for anatomical identification (possession of ovaries or testes) of phenotypic sex.

Testing enhancer activity of the Y-haplotype

We used the AI-based predictor "iEnhancer-Deep" in a 200 bp "sliding-window" approach in the ncRNA region to test its *in-silico* potential as enhancers of *bod1l-expression*[90].

Development gene expression in sibling tadpoles from multiple locations

In total 54 transcriptomes of sibling tadpoles comprising brothers and sisters (plus three unsexed larvae) of four genetic families from four different geographic locations as well as gonadal tissue of a subadult male and adult female from two additional localities were examined by RNAseq (transcriptomics). Tadpoles presented five developmental stages from untreated control groups of a previous experiment[91]. They were euthanized in MS222 and flash-frozen in liquid nitrogen. In tadpoles older than Gosner stage 37, the intestine was removed prior to flash-freezing. In tadpoles at Gosner stage 43-44, only the body segment that contained the gonads was used for RNA extraction. In toadlets of Gosner stage 46 (specifically 6 weeks after metamorphosis), ovaries, testes (separated from Bidder's organ), and several other organs (brain, heart, liver, intestine) were dissected and stored in RNAlater prior to flash-freezing. DNA from tailtips, or skin samples of each tadpole or buccal swabs (juvenile toads) served for genetic sexing using BvXY1-primers. RNAs were extracted using the TRIzol Reagent (Thermo Fisher Scientific, Waltham, USA) according to the supplier's recommendation and cleaned using the RNeasy Mini Kit (Qiagen). All RNAs included in this study exhibited RIN values >9.0. Library processing and RNAseq were carried out by NOVOGENE (Cambridge, UK) on a NovaSeq 6000 PE 150, generating 9 Gb of raw data per sample.

Transcriptome sequences of 54 male and female juvenile developmental stages and two adult gonads were normalized and mapped to the genome using the nf-core RNA-Seq pipeline in STAR aligner (https://github.com/alexdobin/STAR/releases,—

quantMode GeneCounts). A list comprising genes that play roles in sexual development and/or sex determination in other vertebrates (also known as 'usual suspects'[1]) is shown as Supplementary File 6). To plot *bod1l* and *ncRNA-Y* expression trends over the developmental stages and separate expression analysis of *bod1l-X* and *bod1l-Y,* we mapped all reads, including the male haplotype, to the female assembly. As the *bod1l-Y* gene model is 3'-partial, we used the corresponding parts of the *bod1l-X* gene model for quantification. Counts of reads matching the different haplotypes' gene models were extracted using FEATURECOUNTS (from SUBREAD-2.0.3 package[92] with mapping quality (MQ) higher than 19, to increase haplotype specificity. Read counts were normalized by total mapped reads of each sample (TPM normalization). Average expression and standard deviation were calculated and plotted combining samples of the different stages using Microsoft Excel software.

cDNA of developmental stages

Using the iScript RT Supermix for RT-qPCR kit (Biorad), 10 µL of RNA (concentration of 80 ng/µL) from all developmental stages (as used for RNAseq) were randomly reverse-transcribed into cDNA according to the manufacturer's recommendations. To test allele-specific *bod1l*-gene expression, we designed primers (bod1l_RNA1F: 5'_AACTGGAAGAAAAGCGCAAA_3'; bod1l_RNA1R: 5'_GTGCTCTTGGCATTCTGTGA_3') and amplified 577 bp from cDNA under the same conditions as for BvXY1 that were Sanger-sequenced in both directions. To test expression of the ncRNA-Y, we used primers designed to match its inter-exonic regions NcRNAY_Ex1to2_F: 5'_CATGAGGAGAGCAGTAATGAACC_3', NcRNAY_Ex3to2_R: 5'_GCTGTCATCAACCTGCTCAA_3' and amplified and visualize 204 bp male-specific products from cDNA under the same conditions as for BvXY1 but at 55°C annealing temperature.

Probe preparation for RNA* in-situ *hybridization

The sequence marker obtained by cDNA amplification using bod1l_RNA1F/R primers was also used as a probe for RNA *in-situ* hybridization on histological gonad sections of various developmental stages. As a probe of a control gene, expressed in gonads of both sexes, we used a 501 bp cDNA fragment of *Vasa* (DDX4-DEAD-Box-Helicase-4, VASA-homolog), amplified with the primers Vasa1F:

5’_AACCAAAAAGAAGGCGGATT_3’; Vasa1R: 5’_ATGCCAGGATTGCTGTAACC_3’ under the same conditions as for BvXY1 but with annealing temperature of 57°C.

RNA* in-situ *hybridization and H3K4me antibody staining

We opened the body cavity of freshly euthanized tadpoles of three developmental stages and fixed them in 4% paraformaldehyde (PFA) buffered in 1x PBS (Thermofisher), pH = 7.0, at 4°C overnight. Samples were washed in filtered 1× phosphate-buffered saline with tween 20 (PBST, ROTI®Stock) twice for 5 min, then dehydrated in a series of methanol (MetOH) with shaking at room temperature (25% MetOH in PBST for 10 min to 1 h, 50% MetOH in PBST for 10 min to 1 h, 75% MetOH in PBST for 10 min to 1 h, 100% MetOH for 10 min to 1h, timing dependent on tissue sample size). Samples were stored in 100% MetOH at -20°C until histological sectioning.

Cryosectioning, RNA *in-situ* hybridization and immunohistochemistry were performed as described[93] with slight modifications. Transverse cryosections (20 µm) were collected on SuperFrost Plus glass slides (Menzel Gläser J1800AMNZ) and stored at -20°C for up to 2 months. To generate *bod1l* and *vasa* riboprobes, primers Bod1_RNA_Bv_1F (5’_AACTGGAAGAAAAGCGCAAA_3') and Bod1_RNA_Bv_1R (5’_GTGCTCTTGGCATTCTGTGA_3') were used to amplify a 577 bp *bod1l* sequence and primers vasa_1F (5’_AACCAAAAAGAAGGCGGATT_3') and vasa_1R (5’_ATGCCAGGATTGCTGTAACC_3') were used to amplify a 501 bp *Vasa* sequence from Gosner 36-37 *B. viridis* cDNA. These sequences were then ligated to a pCR-bluntII-TOPO vector (Invitrogen) and plasmids were purified using a plasmid DNA miniprep kit according to manufacturer’s instructions (Promega A1460). For synthesis of sense and antisense riboprobes, plasmids were linearized using *BamHI* or *NotI* (New England Biolabs) and *in vitro*-transcribed using T7 or SP6 polymerase (DIG RNA Labeling Mix, Roche), respectively. Rabbit polyclonal anti-histone H3 (tri methyl K4) primary antibody (Abcam ab8580) and secondary goat anti-rabbit Alexa Fluor 568 (1:500, Invitrogen A-11011) were used. Sections were counterstained with 0.25 µg/mL DAPI and mounted in Mowiol 4-88 (Calbiochem). Imaging was performed using the Nikon SMZ18 stereomicroscope and Olympus FluoView FV3000 confocal microscope. To process images, Bitplane Imaris and Adobe Photoshop were used.

FISH-TSA

As a probe for fluorescence *in-situ* hybridization with tyramide signal amplification (FISH-TSA), we designed novel primers to amplify a longer portion of *bod1l* from cDNA (*bod1l*_Exon7To10_F: 5′_AACTGGAAGAAAAGCGCAAA_3′, *bod1l*_Exon7To10_R: 5′_TGTGTTGTCTGGAATCACTGG_3′; same conditions as for BvXY1 but with annealing Temperature of 60°C.), exhibiting the highest heterozygosity in the coding segment of the gene in *B. viridis* males. A PCR-product was produced using the PPP Master Mix (Top-Bio, Czech Republic) according to manufacturer's instructions and purified[94]. The cDNA amplicons were cloned using the TOPO-TA cloning kit and DH5α-T1R Competent Cells (Thermo Fisher Scientific). Plasmid DNA was isolated from bacterial colonies using the EZNA Plasmid DNA Mini Kit I (Omega Bio-Tek Inc.), sequenced, and tested to contain the correct insert by BLAT search on the green toad genome browser (see Data availability). The sequence of the *bod1l* clone was compared to the reference genome. Subsequently, the *bod1l* locus was re-amplified and purified[94]. A 20-year-old tissue suspension (bone marrow), obtained from a male European green toad from Germany, treated with colchicine and fixed in methanol-acetic acid[31] was dropped onto a microscope slide[95], followed by FISH. The probe was labelled with digoxigenin-11-dUTP (Jena Bioscience) using DecaLabel DNA Labelling Kit (Thermo Fisher Scientific). The FISH-TSA followed[96] with minor modifications[97]. Mitotic metaphase spreads were counterstained with ProLongTM Diamond Antifade Mountant with the fluorescent 4',6-diamidino-2-phenylindole, DAPI stain (ThermoFisher Scientific). Microscopy using Leica Microsystem (Wetzlar, Germany) and processing of metaphase images followed[98].

Analysis of potential alternative splicing of bod1l

For the entire *bod1l* region, including the 3'- and 5-' UTR-regions (scf1:566495375-567417162), aligned RNAseq-reads were extracted from each RNA-seq alignment file, the splice site locations were retrieved and counted. Using all the splice sites, a reference file of them was built. The splice event counts were merged for all samples using the reference and stored as csv text file (Supplementary file 7).

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

We used Alphafold2[99] via its cloud-based application and its local installation ColabFold[99] and analyzed the structure predictions by RaptorX[100] for structural and potential functional amino-acid differences between the X and the Y-copy of *B. viridis*. The molecular data viewer iCn3D[101] was used to inspect the structures and highlight specific amino acid residues (Supplementary text 1).

AmpliSeq custom DNA panel

After providing pre-publication access to the *B. viridis* genome-browser, the Illumina concierge design service team developed two primer pools covering ca. 80 kb of the *bod1l*-region (scf1:566,785,000-566,865,000) and 10 kb for a control gene (*amh*), known to play roles in sexual development and sex determination in other vertebrates and lying on the same scaffold (scf1:324181812-324192250). Overall, the theoretical coverage was 92.3% for *bod1l*, and 98.2% for *Amh*.

Wet lab approaches followed the "AmpliSeq for Illumina On-Demand, Custom, and Community Panels" (Document #1000000036408 v09; pp. 5-23[102]) for two primer pools with 193-384 primer pairs in the standard workflow. Briefly, this included ten steps, (i) quantification of DNA samples of multiple sexed green toads (QuantiFluor® dsDNA System kit, Promega), (ii) target amplification in two separate wells for the two primer pools, (iii) partial digestion of amplicons, (iv) ligation of individual indices/sample, (v) library clean up and (vi) amplification, (vii) 2nd clean up, (viii) checking libraries on the Bioanalyser for expected insert size and concentration (QuantiFluor® dsDNA System kit, Promega), (ix) dilution and mixing, and (x) sequencing, performed by Novogene on a NovaSeq PE250 for individual libraries obtained for 54 diploid green toads (ca. 100 Mio bp raw data per library).

We included Palearctic green toads by the following numbers, taxa and sexes: 8 *B. siculus* (4 ♂♂, 3 ♀♀), 9 *B. viridis* (5 ♂♂, 4 ♀♀), 8 *B. balearicus* (5 ♂♂, 3 ♀♀), 9 *B. shaartusiensis* (4 ♂♂, 5 ♀♀), 10 *B. turanensis* (4 ♂♂, 5 ♀♀). Species' nomenclature follows[32].

The sequence reads from target enrichment were mapped against the reference genome using MINIMAP2[74] (-I 100G -a -x sr -t 4). SAMTOOLS[83] was used to add read group and sample names, to fix mates and sort the bam file output. All BAM files were merged and variant calling for the target regions was performed by SAMTOOLS and BCFTOOLS[87] with parameters to cope for extreme sequencing depth (samtools view -bu input.bam scf1:324181812-324192250 scf1:566783149-

566866298 | bcftools mpileup --annotate DP4 --max-depth 1000000000 -f bufVir1d.fa - | bcftools call -mv -Ob -o calls.bcf). The output was converted to vcf format and summarized by an awk script that filtered SNP position and calculated allele frequency from DP4 annotation for each sample. The resulting table was further inspected in Excel and SNPs that were heterozygous in males while homozygous in females were filtered. To fine map the region of the sex marker male specific SNPs of the different species were intersected (resulting in male specific SNPs present in 2, 3, 4, 5 species) and plotted along the target regions in our genome browser.

Data availability

All data generated in this study have been deposited in publicly accessible databases. Genome assembly, whole genome sequencing (WGS) data and RNA-seq reads have been submitted to GenBank on 25 February 2022 (submission ID: SUB11111520, Bioproject: PRJNA292872, Biosample: SAMN26135295). Alternatively, the IGB-Green Toad genome browser http://genomes.igb-berlin.de/Greentoad/ provides easy access to the assembly and its annotation.

Code availability

Scripts and parameters are deposited (upon manuscript acceptance) at https://github.com/HMPNK/BUFVIR.

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. Beukeboom LW, Perrin N. The evolution of sex determination. (Oxford, UK: Oxford University Press, 2014).
6. 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).

7. Perrin, N. Sex reversal: a fountain of youth for sex chromosomes? *Evolution* **63**, 3043-9 (2009).
8. 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).
9. 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).
10. Kostmann, A., Kratochvíl, L. & Rovatsos, M. Poorly differentiated XX/XY sex chromosomes are widely shared across skink radiation. *Proc Biol Sci* **288**, 20202139 (2021).
11. Pan, Q. *et al.* The rise and fall of the ancient northern pike master sex-determining gene. *Elife* **10** (2021).
12. 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).
13. 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).
14. Schartl, M., Schmid, M. & Nanda, I. Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma* **125**, 553–571 (2016).
15. 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).
16. 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).
17. 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).
18. AmphibiaWeb. See www. amphibiaweb.org (accessed June 2023).
19. 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).
20. Cauret, C.M.S. *et al.* Developmental Systems Drift and the Drivers of Sex Chromosome Evolution. *Mol Biol Evol* **37**, 799-810 (2020).
21 Cauret CMS, et al. (2023) Functional dissection and assembly of a small, newly evolved, W chromosome-specific genomic region of the African clawed frog *Xenopus laevis*. *Plos Genetics* **19**(10), e1010990 (2023).
22 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).
23. Bogart, J.P. Karyotypes. in *Evolution of the genus Bufo* (ed. Blair, W.F.) 171-232 (Austin, Texas: University of Texas Press, 1972).
24. Schmid, M. Chromosome banding in Amphibia I. Constitutive heterochromatin and nucleolus organizer regions in *Bufo* and *Hyla*. *Chromosoma (Berl.)* **66**, 361-388 (1978).
25. Piprek, R.P., Kloc, M. & Kubiak, J.Z. Bidder's organ--structure, development and function. *Int J Dev Biol* **58**, 819-27 (2014).
26 Harms, W. Untersuchungen über das Biddersche Organ der männlichen und weiblichen Kröten. *Z Anat Entwickl Gesch* **69**, 598–629 (1923).
27. 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).

28. 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).
29. 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).
30. Malone, J.H. & Fontenot, B.E. Patterns of reproductive isolation in toads. *PLoS One* **3**, e3900 (2008).
31. 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).
32. Ueda, H. Offspring of sex-reversed males in *Bufo viridis*. *Sci. Rep. Lab. Amphib. Biol. Hiroshima. Univ.* **10**(1990).
33. 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).
34. Brelsford, A. *et al.* Homologous sex chromosomes in three deeply divergent anuran species. *Evolution* **67**, 2434-2440 (2013).
35. Tamschick, S. *et al.* Sex chromosome conservation, *DMRT1* phylogeny and gonad morphology in diploid Palearctic green toads (*Bufo viridis* subgroup). *Cytogenet Genome Res* **144**, 315-324 (2014).
36. 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,** 20172667 (2018).
37. Burgin, C. *et al.* Mammal Diversity Database (Version 1.11) dataset. Zenodo http://dx.doi.org/10.5281/zenodo.7830771 (2023).
38. NCBI. See www.ncbi.nlm.nih.gov (accessed June 2023).
39. 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).
40. Stöck M., et al. (2006): Evolution of mitochondrial relationships and biogeography of Palearctic green toads (*Bufo viridis* subgroup) with insights in their genomic plasticity. *Mol Phylogen Evol* **41**, 663-689 (2006).
41. Dufresnes C., *et al.* Phylogeography of Aegean green toads (*Bufo viridis* subgroup): continental hybrid swarm vs. insular diversification with discovery of a new island endemic. *BMC Evol Biol* **18**, 67. (2018).
42. Schartl, M. & Herpin, A. Sex Determination in Vertebrates. in *Encyclopedia of Reproduction (Second Edition)* (ed. Skinner, M.K.) 159-167 (Academic Press, Oxford, 2018).
43. Gosner, K.L. A simplified table for staging anuran embryos and larvae with notes on identification. *Herpetologica*, **16**, 183–190 (1960).
44. Burren, O.S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol* **18**, 165 (2017).
45. Jackson, D.A., Dickinson, P. & Cook, P.R. The size of chromatin loops in HeLa cells. *EMBO J* **9**, 567-571 (1990).
46. 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).
47. Bayley, R. *et al.* H3K4 methylation by SETD1A/BOD1L facilitates RIF1-dependent NHEJ. *Mol Cell* **82**, 1924-1939.e10 (2022).
48. Giovanni, C. *et al.* The BOD1L subunit of the mammalian SETD1A complex sustains the expression of DNA damage repair genes despite restraining H3K4 trimethylation. Preprint at https://www.biorxiv.org/content/10.1101/2023.04.06.535882v2 (2023).

49. Hayashi, S. *et al.* Neofunctionalization of a Noncoding Portion of a DNA Transposon in the Coding Region of the Chimerical Sex-Determining Gene *dm-W* in *Xenopus* Frogs. *Mol Biol Evol* **39**, msac138 (2022).
50. Burgos, M., *et al.* Non-coding RNAs: lncRNAs, miRNAs, and piRNAs in sexual development. *Sex Dev*, **15**, 335–350 (2021).
51. Joshi M, & Rajender S. Long non-coding RNA (lncRNAs) in spermatogenesis, male infertility. *Reprod Biol Endocrinol* **18**, 103 (2020).
52. Feng B., et al. lncRNA DMRT2-as acts as a transcriptional regulator of *dmrt2* involving in sex differentiation in the Chinese tongue sol (*Cynoglossus semilaevis*). *Comp Biochem Physiol B Biochem Mol Biol* **253**, 110542 (2021).
53. Bowles, J. & Koopman, P. Retinoic acid, meiosis and germ cell fate in mammals. *Development* **134**, 3401-11 (2007).
54. Higgs, M.R. *et al. BOD1L* Is Required to Suppress Deleterious Resection of Stressed Replication Forks. *Mol Cell* **59**, 462-77 (2015).
55. Higgs, M.R. *et al.* Histone Methylation by *SETD1A* Protects Nascent DNA through the Nucleosome Chaperone Activity of *FANCD2*. *Mol Cell* **71**, e6, 25-41 (2018).
56. Rodríguez-Marí, A. & Postlethwait, J.H. The role of Fanconi anemia/BRCA genes in zebrafish sex determination. *Methods Cell Biol* **105**, 461-90 (2011).
57. Smallwood, S.A. & Kelsey, G. *De novo* DNA methylation: a germ cell perspective. *Trends Genet* **28**, 33-42 (2012).
58. Youds, J.L. & Boulton, S.J. The choice in meiosis – defining the factors that influence crossover or non-crossover formation. *Journal of Cell Science* **124**, 501-513 (2011).
59. Ki, B.S. *et al.* Epigenetic regulator Cfp1 safeguards male meiotic progression by regulating meiotic gene expression. *Experimental & Molecular Medicine* **54**, 1098-1108 (2022).
60. Fan, H.-Y. & Sun, Q.-Y. Chapter 12 - Oocyte Meiotic Maturation. in *The Ovary (Third Edition)* (eds. Leung, P.C.K. & Adashi, E.Y.) 181-203 (Academic Press, 2019).
61. Dupont, S. & Capel, B. The Chromatin State during Gonadal Sex Determination. *Sex Dev* **15**, 308-316 (2021).
62. Garcia-Moreno, S.A. *et al.* CBX2 is required to stabilize the testis pathway by repressing Wnt signaling. *PLoS Genet* **15**, e1007895 (2019).
63. Nakamura, M. Sex determination in amphibians. *Semin Cell Dev Biol* **20**, 271-82 (2009).
64. Wolf, J.C. *et al.* Effects of 17β-estradiol exposure on *Xenopus laevis* gonadal histopathology. *Environ Toxicol Chem* **29**, 1091–1105 (2010).
65. Peng, Y., Leung, H.C., Yiu, S.M. & Chin, F.Y. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420-8 (2012).
66. Ruan, J. & Li, H. Fast and accurate long-read assembly with *wtdbg2*. *Nat Methods* **17**, 155-158 (2020).
67. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540-546 (2019).
68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
69. Durand, N.C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-8 (2016).
70. Dudchenko, O. *et al. De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).

71. Durand, N.C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101 (2016).
72. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
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. Li, H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**, 3094-3100 (2018).
75. Fan Nie, *et al. De novo* diploid genome assembly using long noisy reads via haplotype-aware error correction and inconsistent overlap identification. bioRxiv 2022.09.25.509436; https://doi.org/10.1101/2022.09.25.509436 (2022).
76. Zhang, H. et al. Fast alignment and preprocessing of chromatin profiles with Chromap. Nat Comm **12,** 6566 (2021).
77. Zhou C., McCarthy S.A. & R. Durbin YaHS: yet another Hi-C scaffolding tool, Bioinformatics **39**, btac808 (2023).
78. Kolmogorov M, et al. Chromosome assembly of large and complex genomes using multiple references. *Genome Research* **28**, 1720-1732 (2018).
79. Brelsford, A., Dufresnes, C., & Perrin, N. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. Heredity, **116**, 177–181 (2016).
80. Rochette N.C., Rivera-Colón A.G. & J.M. Catchen. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol Res* **28**, 4737–4754. (2019).
81. Feron, R. *et al.* RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol Ecol Resour* **21**, 1715-1731 (2021).
82. https://github.com/RomainFeron/paper-sexdetermination-greentoad-2021.
83. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
84. Rimmer, A. *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet* **46**, 912-918 (2014).
85. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
86. Sahlin, K., Lim, M.C.W. & Prost, S. NGSpeciesID: DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecol Evol* **11**, 1392-1398 (2021).
87. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10,** giab008 (2021).
88. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-52 (2000).
89. Minh, B.Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534 (2020).
90. 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).
91. Tamschick S., *et al.* Sex reversal assessments reveal different vulnerability to endocrine disruption between deeply diverged anuran lineages. *Scientific Reports* **6**, 23825 (2016).

92. 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).
93. 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).
94. 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).
95. 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).
96. 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).
97. Knytl, M., Tlapakova, T., Vankova, T. & Krylov, V. *Silurana* chromosomal evolution: a new piece to the puzzle. *Cytogenet Genome Res* **156**, 223-228 (2018).
98. 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**, 4489 (2019).
99. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).
100. Peng, J. & Xu, J. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins* **79 Suppl 10**, 161-71 (2011).
101. Wang, J. *et al.* iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* **36**, 131-135 (2020).
102. 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 and Nathalie Granath (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, developed all primers, did some marker testing by PCRs, co-worked in the wet lab for the AmpliSeq custom DNA panel, wrote the 1st manuscript draft.

All authors contributed text to the manuscript, and read and improved the final version.

Competing interests

The authors declare to have no competing interests.

Figures

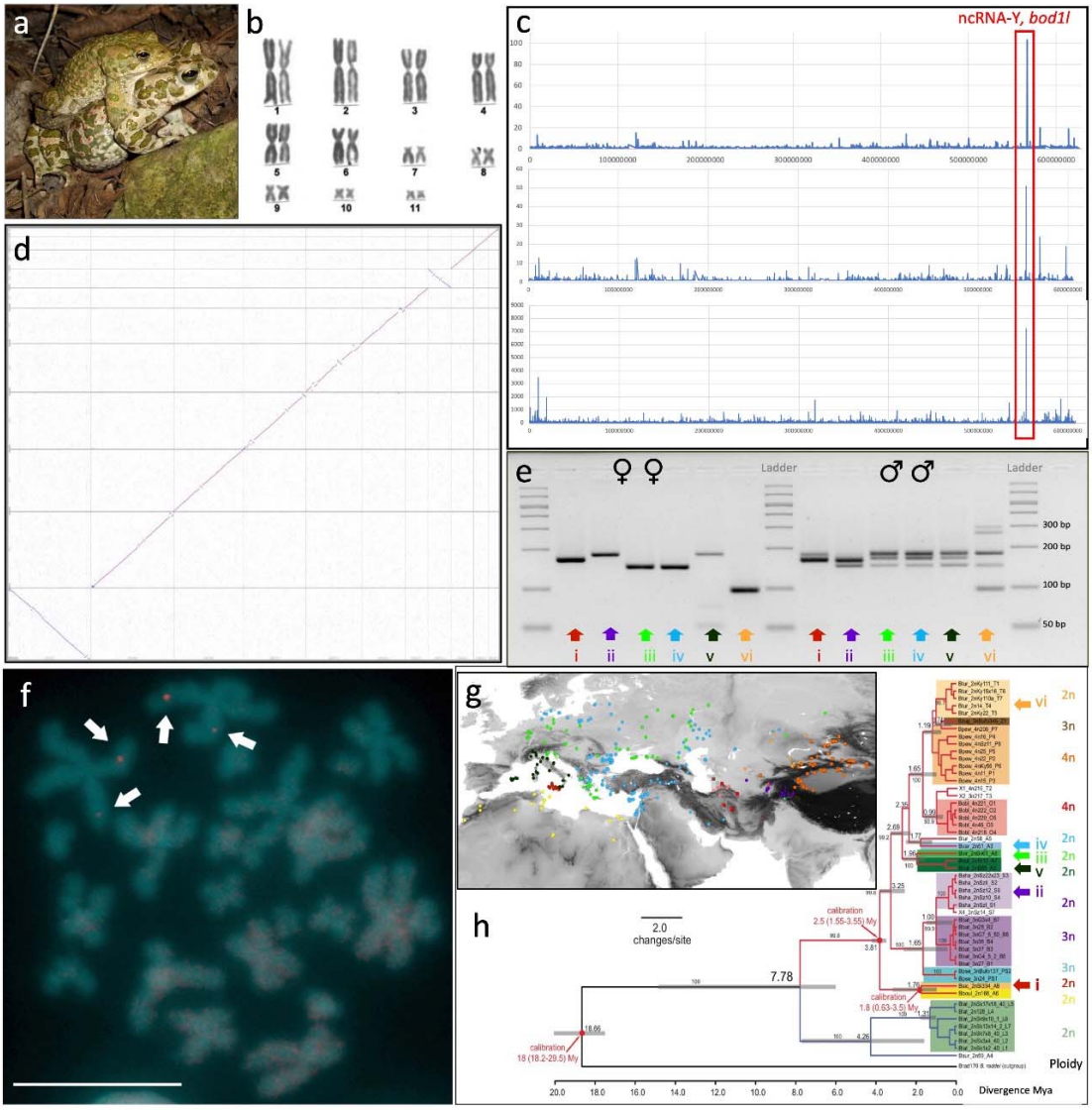


Figure 1. The chromosome-scale reference-genome of *Bufo*(*tes*) *viridis* and the sex-diagnostic marker. a, Breeding pair of *B. viridis* in amplexus (© M. Stöck). **b,** Giemsa-stained karyotype of *B. viridis*, changed from[31]. **c,** above: Mapping the pool-seq short reads to the *B. viridis* reference: genome-wide a single region on scaffold 1 (scf1:566,783,058-566,848,882), where the number of heterozygous SNPs (y-axis) on scaffold 1 (N = 103) 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 comparison with a ZZ/ZW system see Supplementary Fig. 3; **c,** middle: male heterozygous SNPs on scaffold 1 of the male genome assembly identifies the same region (note a slight shift of the peak showing the same locus due to different reference genomes); **c,** below: male-specific pool-seq coverage on the male genome (for these analyses on the entire female and male genomes: Supplementary file 2); y-axis for heterozygous SNPs are counts of SNPs (above and middle subfigure) or Y-specifically pool-sex covered base pairs (lower subfigure) in a cluster, where the maximal distance of variants is 9000 bp.

d, *B. viridis*-assembly (y-axis) compared to *B. bufo (x-axis)* assembly by the Vertebrate Genome Project (VGP, GenBank accession number GCA_905171765.1) showing high synteny of both bufonid genomes across all 11 chromosomes. **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 (note, the highest band in males is weak)*, (iii) *viridis*, (iv) *variabiilis*, (v) *balearicus*, (vi) *turanensis*. **f,** FISH-TSA on a metaphase spread from 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[36], with the phylogenetic positions of the diploid taxa, in which BvXY1 is sex-diagnostic, indicated by color-coded arrows as in (**e**); the three ploidy levels are provided to the right of the leaves, note that the sex marker is seemingly conserved only in the diploid green toads; all polyploids are of hybrid origin[36].

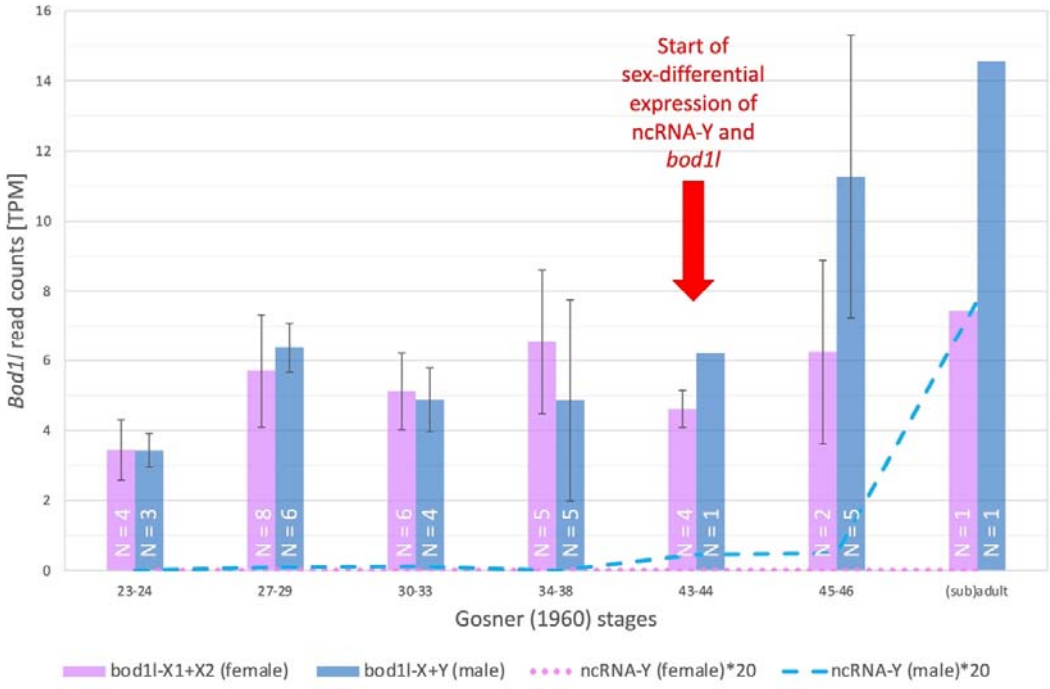


Figure 2. Additive sex-specific XX- or XY-expression of *bod1l* and expression of the ncRNA-Y in females and males in seven developmental stages, with numbers (N) of females and males (N) per stage shown on bars, genetically sexed using PCR-marker BvXY1. Gosner[43] stages 23-24, 10 days after fertilization, Gosner stages 27-29: 15 days after fertilization, Gosner stages 30-33: 18 days after fertilization, Gosner stages 34-38: 34 days after fertilization, Gosner stages 43-44: beginning of metamorphosis, Gosner stages 45-46, up to six weeks after metamorphosis, the latter

with anatomically visible differentiation of testes and Bidder's organs in males, and ovaries in females, subadult male and adult female. Numbers of available tadpole transcriptomes varied by stage; TPM: Transcript reads Per Million RNAseq reads; note that TPM for the ncRNA-Y were multiplied 20x. See Supplementary Fig. 8 for comparison of additive sex-specific expression of *bod1l* and ncRNA-Y and haplotype-specific *bod1l*-expression in males.

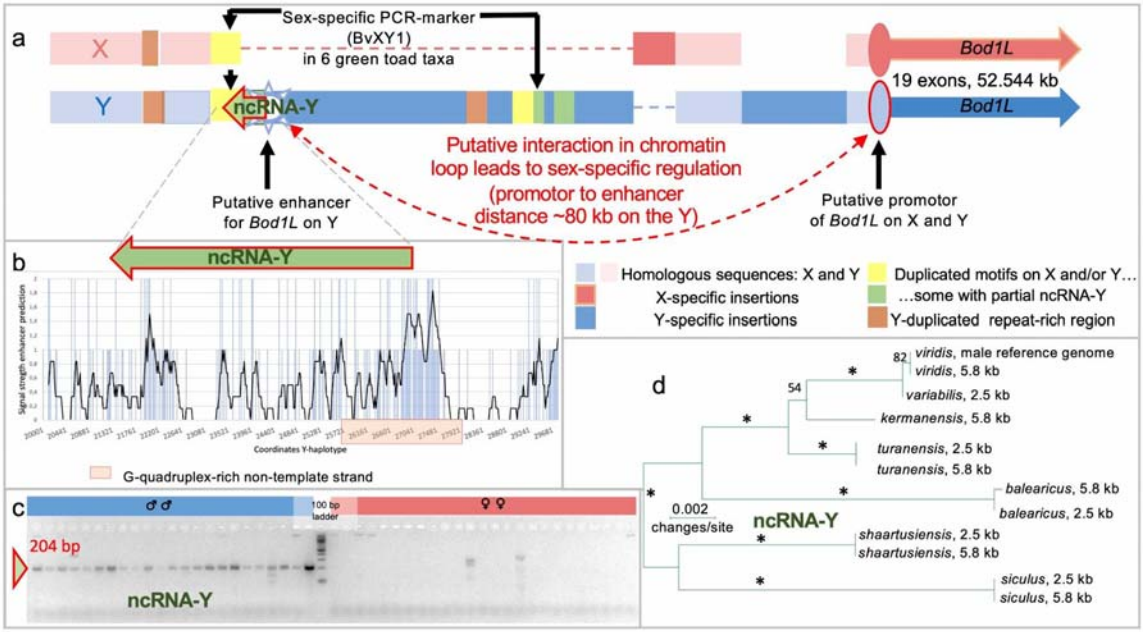
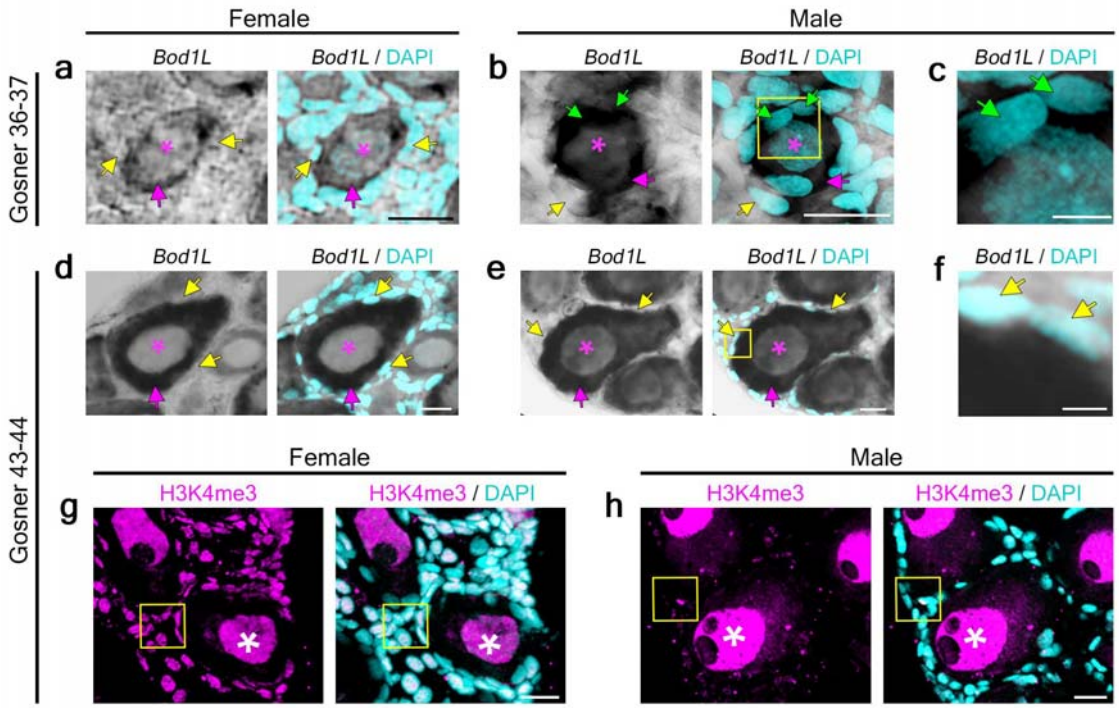


Figure 3. Structure of the X- and Y-haplotype of the candidate sex locus of *Bufo*(*tes*) *viridis* with focus on the 5'-end of *bod1l* and the Y-specific ncRNA. a, X-chromosomal (above, pink) and Y-chromosomal haplotype (below, blue) with color-coded homologous or X- and Y-specific INDELs, and the Y-specific long (ca. 5.8 kb) non-coding RNA (ncRNA-Y), flanking a putative enhancer at the 5'-end of the gene, ca. 80 kilobases before its promoter. **b,** Graph of the signal of the AI-based software *i-enhancer*, reaching from 0 to 2, predicting strong enhancer properties of the Y-specific 5'-end of *bod1l*, including a G-quadruplex-rich region on the non-template strand, overlapping with the ncRNA-Y. **c,** Y- and thus male-specific expression of the ncRNA-Y (204 bp PCR-product from cDNA) in larval green toads, from 10 days after fertilization (Gosner stages 23-24) until early metamorphosis (Gosner 46); in females, only non-specific products occurred, since X-copies of *bod1l* do not contain this ncRNA (for details and positive controls of females: Supplementary Fig. 9). **d,** Unrooted Maximum-Likelihood-based phylogenetic tree, showing that the ncRNA-Y evolved in a common ancestor of diploid Palearctic green toads sharing this XY-system, with *siculus* and *viridis* spanning the most divergent lineages; two PCR-products (2.5 and 5.8 kb), directly amplified by independent PCRs from six related diploid green toad taxa, clustered by Y-gametologue and thus species (* symbolized 100% bootstrap support; only support values >50% are shown).



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

Figure 4. *Bod1l*-expression and H3K4me3-methylation in primordial gonads of *B. viridis*, i.e. in the female early ovary, and in male Bidder's organ, contributing to the testes anlage (Gosner stages 36-37) as well as in early metamorphosis (Gosner stages 43-44). a-f, *Bod1l* expression shown by RNA *in-situ* hybridization (ISH) on 20 µm transverse cryosections of female (a, d) and male (b, c, e, f, h) gonads at developmental stages Gosner 36-37 (a-c) and 43-44 (d-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 *bod1l* RNA ISH; green arrows: nuclei of somatic cells stained by *bod1l* RNA ISH. Yellow boxes in b and e indicate zoomed images shown in c and f, respectively. **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; yellow boxes indicate somatic cells. Scale bars (a, b, d, e, g, h): 20 µm; (c, f) 5 µm. **i,** Summary of *bod1l*-expression and H3K4me3-methylation in somatic cells and germ cells in Gosner 36-37 and Gosner 43-44 male and female gonads.

Additional information

Supplementary File 1: Description of Supplementary Data.

Supplementary File 2: Comparison of female and male *B. viridis* genome assemblies.

Supplementary File 3: Analyses of SNPs causing 6 non-synonymous changes in the coding region of *bod1l* in 49 male and female transcriptomes as compared to the female reference genome. None of these changes was 100% associated with genetic sex as in the 5'-region of *bod1l*.

Supplementary File 4: Percentages of covered parts of ca. 80 kb of the *bod1l*-region by targeted sequencing in six diploid green toad taxa, using a panel based on the female *B. viridis* reference genome, compared to ca. 20 kb of the gene for *anti-Muellerian hormone* (*amh*). Taxa abbreviations: Bbal - *balearicus*, Bsha - *shaartusiensis*, Bsic - *siculus*; Btur - *turanensis*, Bvir - *viridis* and *variabilis*.

Supplementary File 5: SNP-comparisons in ca. 80 kb of the *bod1l*- (scf1: 566,783,513 - 566,865,063) and ca. 10 kb of the *amh*-regions (scf1: 324,181,734 – 324,192,327) by targeted sequencing in six diploid green toad taxa in a panel based on the female *B. viridis* reference genome. The coordinates of the cluster of male specific SNPs present in all five species are colored in red (scf1: 566,790,085 - 566,790,204). Values for color-coded SNP-comparisons with the reference genome: >=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 6: 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 7: 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.

Supplementary Figure 6: Female/male pool-seq coverage over haplotypes in *B. viridis*. a female pool-sequencing, **b** male pool-sequencing.

Supplementary Figure 7: X- and Y-haplotype comparisons by dot plots in the *bod1l*-region. a X-*versus*-Y dot plot, **b** Y-*versus*-Y, **c** X-*versus*-X.

Supplementary Figure 8: Comparison of additive sex-specific expression of bod1l and ncRNA-Y and haplotype-specific expression in males.

Supplementary Figure 9: a, Y- and thus male-specific expression of the ncRNA-Y (204 bp PCR-product from cDNA) in larval green toads. PCR-products from cDNA (204 bp), obtained with primers NcRNAY_Ex1to2_F/_R; to the left males,

to the right females; each sex is shown from Gosner stage 23-34 (10 days after fertilization) until early metamorphosis (Gosner 43-44) and one adult gonad for each sex; in females, only non-specific products occurred, since X-copies of *bod1l* do not contain and thus do not express this ncRNA. **b,** Gel image of a control-PCR using the females from (a).

Supplementary Figure 10: Graph of the signal of the AI-based software i-enhancer. Signal bars reaching from 0 to 2, predicting strong enhancer properties of the Y-specific 5'-end of *bod1l*, including a G-quadruplex-rich region on the non-template strand, overlapping with the ncRNA-Y.

Supplementary Figure 11: *Vasa*, *bod1l* expression and H3K4me3-methylation in *B. viridis* gonads, overview images and additional representative examples.

a-m, v*asa* and *bod1L* expression shown by RNA *in-situ* hybridization (ISH) on 20 µm transverse cryosections of female (**a, b, g, h**) and male (**c, d, f, i-m**) gonads, Gosner 36-37. DAPI was used to stain nuclei. Sense controls for RNA ISH are shown in **a, c, e**, **g**, right panels. Red boxes in **a, c, g, i** (left panel) indicate zoomed images shown in **b, d, h, k,** respectively. Yellow boxes in **b, d, h, k** (right panel) indicate zoomed images shown in **Fig. 4a-b,** respectively. Red boxes in l**, m** (left panel) indicate zoomed images shown in right panels of l**, m**. *Bod1L* staining on three different cryosections of male gonads are shown in k**, l, m**. Magenta asterisks indicate nuclei of germ cells; magenta arrows: cytoplasm of germ cells; yellow arrows: somatic cells not stained by RNA ISH; green arrows: somatic cells stained by RNA ISH. **n, o,** H3K4me3 and DAPI staining on 20 µm transverse cryosections of female (**l**) and male (**m**) *B. viridis* Bidder's organ and gonads, Gosner stage 43-44. Yellow boxes in **l, m** indicate zoomed images shown in **Fig. 4g**, **h,** respectively. White asterisks indicate H3K4me3-positive germ cell nuclei; white arrowheads, H3K4me3-positive somatic cell nuclei; yellow arrowheads, H3K4me3-negative somatic cell nuclei. Scale bars: (a, c, g, i) 50 µm, (b, d-f, h, k-o) 20 µm.

Supplementary Tables

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

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

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

Supplementary Texts

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