Comparative genomics of the balanced lethal system in Triturus newts

France, J.M.

Linkage Mapping vs Association

A Comparison of Two RADseq-based Approaches to Identify Markers for Homomorphic Sex Chromosomes in Large Genomes

JAMES FRANCE[1,2], WIESŁAW BABIK[3], KATARZYNA DUDEK[3], MARZENA MARSZAŁEK[3], BEN WIELSTRA[1,2]

[1] Institute of Biology Leiden, Leiden University, Leiden, The Netherlands

[2] Naturalis Biodiversity Center, Leiden, The Netherlands

[3] Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland

(in revision)

This chapter is available as a preprint at ***Authorea:*** *(10.22541/au.172137196.69932431)*

Abstract

Reliable tools for the identification of genetic sex are invaluable in many fields of biology, but their design requires knowledge of sex-linked sequences, which is lacking in many taxa. Restriction-site associated DNA sequencing (RADseq) is widely used to identify sex-linked markers, but multiple distinct strategies are employed, and it is often not obvious which is most suitable. In this study we compare two approaches for using RADseq to identify sex-linked markers. We use the common newt, *Lissotriton vulgaris*, as our study system, providing a challenging combination of homomorphic sex chromosomes and an exceptionally large genome. We attempt an associative approach, sequencing 60 adult newts of known-sex individuals, and compare this to a linkage mapping approach utilizing a family of 146 offspring with unknown sex. After optimization for a highly paralogous genome, the associative approach identifies five Y-chromosome linked markers in *L. vulgaris* and we design a robust PCR protocol for molecular sexing of four more related species. Via the linkage approach we construct a high-density map featuring 10,763 markers, matching the observed karyotype of *L. vulgaris* and showing broad synteny with the Iberian ribbed newt (*Pleurodeles waltl*). However, without incorporating the markers identified via the association-based approach, we cannot confidently distinguish a sex-determining region in the linkage map, either by analysing marker density or by identifying clusters of paternal markers. We conclude that linkage mapping alone is unlikely to yield sex-linked markers in organisms with very small sex-determining regions, however association-based RADseq can still be effective under these conditions.

Introduction

Sex-linked genetic markers are vital to both applied and fundamental biology. In agriculture and aquiculture molecular markers boost productivity by, for instance, enabling the early selection of fruit bearing female date palms (Intha & Chaiprasart 2020) or aiding the maintenance of all-male stocks of tilapia (Curzon et al. 2021). Molecular sex identification is viable even on small and degraded samples, making it invaluable for ecology, conservation and forensic biology. Examples include determining the sex of tiger prey from hairs recovered from scat (De et al. 2019), monitoring elephant sex ratios by genotyping dung (Vidya et al. 2003) and identifying the illegal poaching of female pheasants (An et al. 2007). As sex determining regions of the genome have been identified as drivers of speciation (Dufresnes & Crochet 2022; Johnson & Lachance 2012; Payseur et al. 2018), their identification and study is of particular importance to evolutionary biology.

Many taxa have highly conserved sex determination systems (Cortez et al. 2014; Ellegren 2010) enabling a single method of molecular sex identification to be used across an enormous range of species with little modification. For example, all birds possess a ZW chromosome system, and a single primer pair based on the CHD1 gene allows for sex identification across the neognathae (which includes over 99% of bird species) (Fridolfsson & Ellegren 1999). Similarly, amplification of the SRY gene identifies the presence of the Y-chromosome in a wide range of eutherian mammals (Hrovatin & Kunej 2018).

However, such conservation is far from universal, and other branches of the tree of life have experienced frequent turnover of sex chromosomes (Ma & Rovatsos 2022), often made obvious by transitions between male and female heterogametey (Bachtrog et al. 2014; van Doorn & Kirkpatrick 2010). Groups notable for rapid evolution in sex determination systems include fish (Kitano & Peichel 2012), squamate reptiles (Ezaz et al. 2010) and amphibians (Miura 2017). In addition, while the majority of plants are hermaphroditic (or monoecious), dioecy (two fully separate sexes) has independently evolved on numerous occasions (Renner 2014). Rapid turnover complicates molecular sex identification, as new markers may have to be identified on a lineage-by-lineage basis. Exacerbating this, evolutionally young sex chromosomes are typically not highly differentiated, resulting in homomorphic chromosomes with only a small region in which sex-linked markers may be found (Charlesworth et al. 2005; Wright et al. 2016).

Several sequencing approaches are employed for the identification of sex-linked markers. Recent studies have tended to employ either whole genome sequencing (WGS) (Darolti et al. 2019; Keinath et al. 2018; Rafati et al. 2020) or restriction-site associated DNA sequencing (RADseq) (Gamble et al. 2015; Hime et al. 2019; Hu et al. 2019). WGS is a powerful technique, but the cost of sequencing the entire genome of

multiple individuals of both sexes may be prohibitive. This is particularly the case for organisms with exceptionally large genomes, such as salamanders, lungfish and many genera of dioecious plants such as gingko, mistletoe and yew, that all have genome sizes in excess of 10 Gbp (Gregory 2024; Pellicer & Leitch 2020). For such gigantic genomes, RADseq may be a superior technique. RADseq targets restriction site-bounded sequences scattered randomly throughout the genome (Miller et al. 2007), giving genome-wide data, while requiring orders of magnitude less sequencing than WGS.

RADseq is employed to identify sex-linked makers via multiple methodologies (Gamble 2016). The most conceptually simple approach is association-based (Gamble & Zarkower 2014). A number of individuals from both sexes are sequenced, and the recovered RAD markers are screened for those present in one sex and absent in another. The RAD marker set can also be screened for SNPs present only in one sex, and markers present at twice the copy number in one sex than the other (Brelsford et al. 2017; Trenkel et al. 2020). This approach typically requires 10-30 individuals per sex, with decreasing sample sizes increasing the risk of generating false positives.

As RADseq involves many thousands of markers scattered randomly across the genome it is ideal for building high density linkage maps. Sex-linked regions can then be identified by quantitative trait locus analysis (QTL) (Peng et al. 2016), detecting clusters of SNPs unique to the heterogametic parent (Hu et al. 2021) or by locating regions of reduced recombination (Brelsford et al. 2016). The map may also be a valuable resource for questions beyond that of sex determination, for example, anchoring scaffolds of a whole genome assembly (Lee et al. 2019). A linkage map requires more investment than an association-based approach, as a linkage family (or families) must be bred and typically at least 100 individuals must be sequenced. However, as the sex of the offspring does not have to be known (Brelsford et al. 2016), a linkage mapping approach may be more feasible in cases where large numbers of adults are not available to be morphologically sexed and juveniles are readily bred but difficult to sex.

Little literature is available on the relative performance of different approaches for using RADseq to identify sex-linked markers. Most studies employ a single methodology, and while this does provide a list of strategies that have proved successful in at least one situation, the publication bias against negative results means that the limitations of these approaches remain obscure. The publications that do compare different tools for sex-linked maker discovery tend to either be reviews, which aggregate results generated in wildly different contexts (Palmer et al. 2019), or comparisons of bioinformatic approaches (Trenkel et al. 2020). This presents an issue for researchers designing such studies, as it is not clear which sequencing strategy is more likely to yield useful sex-linked markers. In this study we aim to contrast two approaches to sex-linked marker discovery using RADseq (linkage mapping, and presence/absence association)

by applying both to a single, challenging species, the common newt (*Lissotriton vulgaris*).

Lissotriton vulgaris one of the most widely distributed amphibian species in Europe, ranging from Ireland to Siberia (Sparreboom 2014). It is part of the smooth newt species complex, which includes six closely related newt species found in Europe and western Asia (Pabijan et al. 2017; Wielstra et al. 2018). The wider genus *Lissotriton* includes four additional species more distantly related to *L. vulgaris* (Babik et al. 2005). Like all salamanders, *Lissotriton* have gigantic genomes, estimated at 27.7-32.0 Gbp (Gregory 2024; Litvinchuk et al. 2007). *Lissotriton* possess XY sex-determination systems with little to no heteromorphism (Schmid et al. 1979; Zbożeń & Rafiński 1993). No Y-linked marker has previously been reported in any Salamander, however RADseq studies have identified the ancestral amphibian ZW system in the family Cryptobranchidae (Hime et al. 2019; Hu et al. 2019, 2021), and the first salamander whole genome assembly revealed a tiny 300 Kbp W-linked region in the axolotl (Keinath et al. 2018).

We first attempt to identify Y-linked markers via the associate approach by performing RADseq on a group of known sex *L. vulgaris*. We then test the linkage mapping approach to identify a Y-linked region, gathering RADseq data from a full-sibling *L. vulgaris* family with offspring of unknown sex. Finally, we validate candidate markers by PCR amplification in multiple taxa within the genus *Lissotriton*.

Methods & Materials

Sample acquisition

For identification of Y-linked markers by association, samples from 60 sexed adult *L. vulgaris* (30 male, 30 female) were collected from the Kraków metropolitan area in Poland, these samples are also reported in Babik et al. (in press). For construction of the linkage map an *L. vulgaris* family was bred consisting of 2 parents (1 adult male and 1 adult female, collected in Kraków, Poland) and 146 offspring of unknown sex. For validation of candidate markers via PCR, 12 samples (6 male, 6 female) of both *L. vulgaris* and *L. montandoni* and an additional 2 samples (1 male, 1 female) of multiple taxa belonging to the smooth newt complex (*L. v. ampelensis*, *L. v. meridionalis*, *L. graecus, L. kosswigi* and *L. schmidtleri*) as well as more distantly related *Lissotriton* species (*L. boscai*, *L. helveticus* and *L. italicus*,) were obtained from localities across Europe and Anatolia (see online supporting information for a full details of samples). Samples from adults consisted of tail tips and for offspring the freshly hatched larvae were collected whole. Samples were stored in 96% ethanol.

DNA extraction, library preparation and RAD-sequencing

Whole genomic DNA was extracted from the selected tissue samples with the Promega Wizard™ Genomic DNA Purification Kit (Promega, Madison, WI, USA), according to the salt-based extraction protocol of Sambrook and Russel (2001). Double digest RADseq libraries were prepared according to the Adapterama III High-Throughput 3RAD protocol (Bayona-Vásquez et al. 2019) from 100 ng of genomic DNA, using restriction enzymes *Eco*RI, *Xba*I, and *Nhe*I. Fragments in the range of 490-600 bp were excised using Pippin Prep, the libraries were pooled equimolarly and 150 bp paired-end sequencing was performed by Novogene (Cambridge, UK) on the Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) platform, targeting a yield of 1 Gbp per sample for the linkage map family and 2 Gbp for the known-sex adults.

RADseq data processing

The Stacks package v2.54 (Catchen et al. 2013; Rochette et al. 2019) was employed to process raw reads from all RADseq samples. Reads were demultiplexed and trimmed via the process_radtags program. The denovo_map.pl pipeline was then used to group reads into putative loci. Default settings were used except for the parameter M, which controls how many mismatched bases two read-pairs may have and still be assigned to the same locus. If M is too low, Stacks may not correctly group reads from the same locus together (especially if samples are genetically divergent), however if M is too high it may result in reads from paralogous loci being inappropriately aggregated together (Paris et al. 2017). As our RADseq analysis was based on two sample-sets which would not be expected to exhibit great genetic diversity (a captive bred linkage map family and a group of wild-caught newts from a relatively small area) a low value of M would seem appropriate. Accordingly, we selected the default value of M=2 for the linkage family, as this low value will minimise the distortions caused by mapping paralogs together as falsely heterozygous loci.

However, we hypothesised that an overly high value of M may actually be helpful when selecting markers for molecular sex identification. This increases the chance sex-linked loci with autosomal paralogs will be assigned reads even in the opposite sex, and so filtered out in subsequent analysis. This is desirable as these loci would likely give false positives in PCR based genotyping (due to amplification of the paralog). Consequently, we ran our analysis of the known-sex RADseq data three times, with values of M=2, M=6 and M=10.

Sex-associated presence/absence marker discovery

The bam files produced from each of the three runs of denovo_map.pl were processed with the depth function of SAMtools (Danecek et al. 2021) to produce a table of the number of reads of each marker in each of the sexed-adult samples. A custom R script was then used to identify candidate Y-linked markers which had reads in at least 90% of male samples, and less than 10% of female samples.

We aimed to minimise the likelihood of candidate Y-linked markers failing in PCR validation by avoiding candidates with a large number of paralogous sequences present in the genome. Primers designed for such markers would have a high chance of amplifying products from autosomal paralogs, resulting in false positives in female samples. Therefore, a BLAST (Camacho et al. 2009) search was then conducted for each candidate against the catalogue of all RAD markers found in the run of the same M value, and the number of potential paralogous hits (> 80% sequence similarity with a query coverage of > 25%) recorded. Candidate markers were ranked based on absence of residual reads in females, number of potential paralogs and average read depth in males. The ten highest ranked candidate markers from each run were selected for PCR screening.

After removing any duplicate markers (where the same sequence was selected from multiple runs of the pipeline), primers were designed with Primer 3 (Untergasser et al. 2012) targeting an optimal primer length of 20 bp and melting temperature of 60°C. To facilitate the design of a multiplex PCR, for each marker we attempted to design two primer pairs, amplifying both a shorter (ca. 100 bp) and a longer (ca. 200 bp) product. For candidate markers that did not consist of a continuous sequence, as the RAD fragments were longer than 2 x 150 read, the primer pairs amplifying the shorter products were derived entirely from the forward read, whereas the longer product would incorporate an additional sequence of unknown length between the forward and reverse reads. Sequences of all primers are found in Supplementary Table S2.

Sex associated marker validation

The primers designed for candidate sex-associated markers were tested via PCR amplification with 2x QIAGEN multiplex master mix (QIAGEN B.V, Venlo, Netherlands). After optimisation a final PCR protocol was designed consisting of a 95°C hot start for 10 minutes, followed by 35 cycles of denaturation for 30 seconds at 95°C, 60 seconds annealing at 61°C and 45 seconds extension at 72°C, with a final extension of 10 minutes at 72°C. All primers were used at a final concentration of 0.1 µM. Initial screening was against a single male/female pair of *L. vulgaris*. Markers showing male specificity were validated against a panel of six male and six female *L. vulgaris*. Validated makers were then tested against a male/female pairs of both *L. montandoni* and *L. helveticus* (as representatives of the *L. vulgaris* complex, and the wider *Lissotriton*

genus, respectively). Any markers with multispecies sex specificity were tested in male/female pairs of all available species of *Lissotriton*. Finally, a multiplex PCR was designed, combining the most broadly male-specific markers with an autosomal control marker, CDK-17, which amplifies a product of 537 bp.

Linkage map construction

The joint VCF file produced by Stacks was filtered with VCFtools (Danecek et al. 2011) to exclude indels and SNPs with greater than 5% missing data, a mean depth of less than 10, or a minor allele frequency of less than 0.2. The thin function of VCFtools was used to select a single SNP per marker. Lep-MAP 3 (Rastas 2017) was then used to construct paternal, maternal and sex averaged linkage maps. To incorporate the candidate Y-linked markers identified in the known-sex adults into the linkage map, a custom R script was used to translate the presence/absence of reads for these markers into pseudo-SNP genotype calls for all samples (markers with no reads were assigned an artificial AA genotype whereas markers with reads were assigned as AT). These calls were appended to the call file produced by the first stage of the Lep-MAP 3 pipeline (the ParentCall2 module) and incorporated into all subsequent steps. Initial linkage groups were created with the SeparateChromosomes2 module, with a LOD limit of 20 and distortion LOD set to 1. Unplaced markers were then added with the JoinSingles2All module with a LOD limit of 15. The markers were then ordered with the OrderMarkers2 module, using 20 merge iterations, 8 polish iterations, a minError value of 0.02 and the scale setting M/N 2. The informative mask options 23 and 13 were used for the paternal and maternal maps respectively.

Linkage map comparison with *Pleurodeles waltl* genome

Initial validation of the linkage map was performed by BLASTing the sequences of the mapped makers against the genome assembly of the Iberian ribbed newt (*Pleurodeles waltl*) (Brown et al. 2025), using a word size of 11 and requiring a minimum E value of 1e-20. To account for the high degree of paralogy typical of newt genomes, the blast results were filtered to include only hits that exceeded the significance of the next best hit by five orders of magnitude, following the methodology of Purcell et al. (2014). The filtered blast hits were then used to create an Oxford plot via a custom R script.

Marker density analysis

In an XY system, the sexed-linked region is not expected to undergo recombination in males, and so markers within this region should exhibit extreme genetic linkage when transmitted from father to offspring. When a paternal linkage map is constructed, these markers will form a region of high marker density. However, as the sex-linked region does undergo recombination in females, these markers should be spread over a wider region when the corresponding maternal map is constructed. We attempted to locate the sex-linked region by plotting marker density in both paternal and maternal linkage maps via a custom R script, and identifying a peak which is exclusive to the paternal map.

Analysis of paternal specific markers and SNPs

To be placed on the linkage map a marker must have a SNP that is heterozygous in at least one the parents. SNPs which are heterozygous in the father but homozygous in the mother are termed paternal specific SNPs. RAD loci which include only paternal specific SNPs are termed paternal specific markers. While a large number of paternal (and maternal) specific SNPs and markers will be randomly scattered across the genome, the Y-linked region is expected to be particularly enriched in paternal specific SNPs and markers, as in an XY system this region is, by definition, heterozygous in males. To identify this enriched region, we first plotted the number of paternal SNPs and markers against the total number of markers per linkage group. We then used a custom R script to divide each linkage group into bins of 2 cM and plot the probability (assuming the markers were randomly and independently distributed) of obtaining the measured number of paternal specific markers and SNPs, using a binomial distribution for the markers and a Poisson distribution for the SNPs.

Results

Sex association in known-sex adults

After demultiplexing and filtering, the 60 known-sex individuals yielded a total of 901 million reads (per sample median: 13.7 M, interquartile range: 11.5-16.5 M). The three runs of the denovo_map.pl pipeline yielded differing results as parameter M was varied. Increasing M decreased the total number of loci identified in the dataset (M=2: 1,541,940, M=6: 1,026,619, M=10: 911,354) and increased the mean adjusted sequencing depth per sample (M=2: 28.2, M=6: 30.6, M=10: 31.7) and the proportion of loci found in at least 50% of samples (M=2: 8.61%, M=6: 12.56%, M=10: 13.31%).

As expected, in the initial sets of candidate markers (M=2: 32, M=6: 26, M=10: 19, after duplicates removed, 35 unique sequences in total) selected by screening for presence in males and absence in females a high degree of paralogy was observed, with the number of BLAST hits per marker varying from 1 to 500. The proportion of candidate markers without paralogous hits increased with parameter M (M=2: 19%, M=6: 38%, M=10: 53%). After the 10 highest ranked candidate markers from each run were selected, the majority of candidates appeared in multiple runs. Removing duplicates left 14 unique candidate markers for PCR validation. In total 25 primers pairs were designed (Supplementary Table S2), as for three markers Primer3 failed to generate a valid primer pair for the 'short' product under the given conditions.

Sex-associated marker validation

Of the 14 candidate markers, six show validated male-specific amplification in *L. vulgaris*, with an additional marker showing initial male-specificity screening but failing in the 12 individual validation panel. Three markers retain sex-specificity in *L. montandoni* (Table 1). Two markers, LvY-79267 and LvY-51393 show broad male-specificity within the smooth newt species complex. However, both markers fail to amplify in males in at least one taxon within this group, with LvY-79267 failing in *L. v. meridionalis* and LvY-51393 failing in *L. v. ampelensis*. In the more distantly related species (*L. helveticus*, *L. italicus* and *L. boscai)* no sex-specific amplification is observed with any primer pair, with markers failing to amplify in either sex or amplifying in both sexes (all gels resulting from screening and validation found in Supplementary Figures 1-4). If the candidate markers from each run are considered separately, higher values of parameter M give more useful results. Five out of ten candidates selected from M=10 are sex-specific in *L. vulgaris* compared to four out of ten from M=6 and two out of ten selected from M=2. A final multiplex mix (Table 2), amplifying fragments from both LvY-79267 and LvY-51393 as well as the autosomal control marker CDK-17, shows robust sex-identification across all taxa within the *L. vulgaris* species complex included in this study (Fig. 1).

Marker	Ranked in Stacks run M=			Primer pair	Male specific amplification in *Lissotriton* taxa									
					L. vulgaris species complex							Other		
	2	6	10		*L. vulgaris*	*L. montandoni*	*L. v. ampelensis*	*L. v. meridionalis*	*L. schmidtleri*	*L. graecus*	*L. kosswigi*	*L. helveticus*	*L. italicus*	*L. boscai*
lvY-36220	yes	yes	yes	short	○									
				long	×									
lvY-65590	yes	yes	yes	short	○									
				long	○									
lvY-81842	yes	yes	yes	short	○†	○						×		
				long	○†	×						×		
lvY-123701	yes	yes	yes	short	+	×						×		
				long	+	×						×		
lvY-138925	yes	yes	yes	short	○									
				long	○									
lvY-143365	yes	yes	yes	short	○									
				long	○									
lvY-79267		yes	yes	short	+	+	+	×	+	+	+	×	×	×‡
				long	+	+	+	×	+	+	+	×	×	×‡
lvY-115632		yes	yes	short	+	+	×	×	○	+	+	×	○	×
				long	+	+	○	○	○	○	○	×	×	×‡
lvY-128014		yes	yes	short	+	×‡						×		
				long	○									
lvY-51393	yes		yes	short	+	+	×	+	+	+	+	○	○	×‡
				long	+	×	×	+	×	+	+	○	○	×‡
lvY-99941		yes		long	○									
lvY-11521	yes			long	○									
lvY-28978	yes			long	○									
lvY-102891	yes			short	○									
				long	○									

Table 1: Summary of results of PCR screening of primer pairs designed for candidate Y-linked markers in *Lissotriton* newts. Results are indicated as: + amplification only in male samples, ○ amplification in both male and female samples, × no amplification in either sex. Twenty-five primer pairs were tested in a male-female pair of *L. vulgaris*, followed by validation of successful markers in a 12 individual panel. Nine primer pairs, covering five candidate markers, show confirmed male-specific amplification in *L. vulgaris*, with a further two primer pairs, marked with †, showing initial male-specificity but failing in the wider panel. The successful primer sets were then tested in male-female pairs of *L. montandoni* and *L. helveticus*, and the three markers which show sex-specificity in *L. montandoni* were then tested in male-female pairs of all available *Lissotriton* taxa. While markers lvY-79267 and lvY-51393 demonstrate broad male-specificity within the *L. vulgaris* species complex no marker shows any sex-specificity in more distantly related *Lissotriton* taxa. In several cases, marked with ‡, the PCR results are difficult to interpret due to faint amplification of multiple off-target bands.

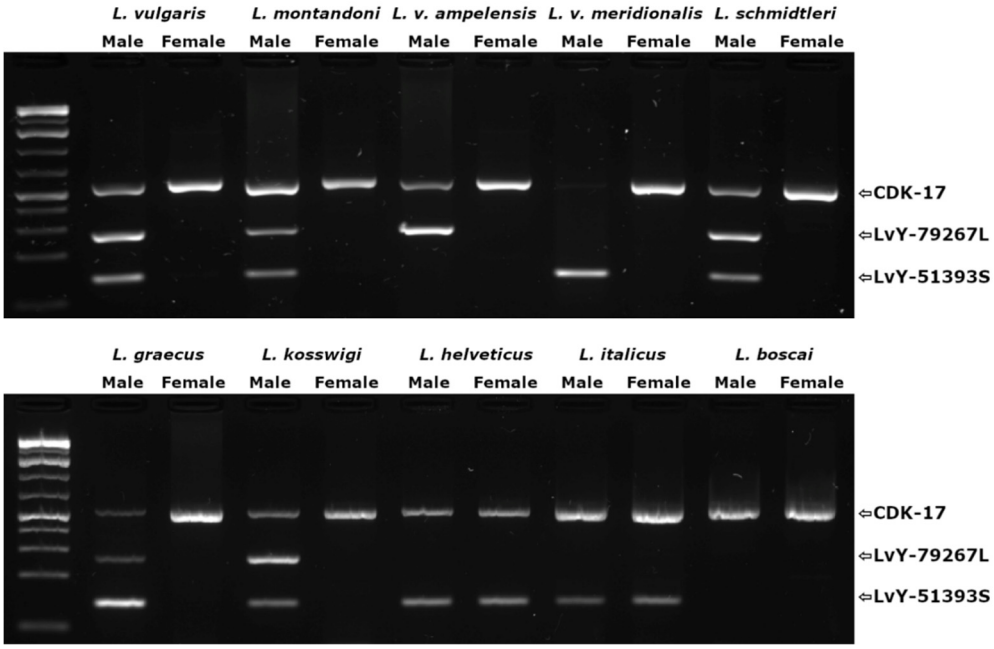

Figure 1: The results of the multiplex PCR designed for molecular sex identification in the *L. vulgaris* species complex. The three primer sets amplify a control marker CDK-17 (537 bp) and two Y-linked markers LvY-79267-Long (ca. 240 bp) and LvY-51393-Short (124 bp). Within the *L. vulgaris* complex (*L. vulgaris, L. montandoni, L. schmidtleri, L. graecus* and *L. kosswigi*) amplification of the Y-linked markers is observed only in male samples. In more distantly related *Lissotriton* species (*L. helveticus, L. italicus* and *L. boscai)* sex-specific amplification is not observed. As male amplification of one of the Y-linked markers is not observed in each of the non-nominate subspecies of *L. vulgaris* (LvY-51393-Short *in L. v. ampelensis* and LvY-79267-Long in *L. v. meridionalis*) we recommend using both diagnostic primer pairs for reliable sex-identification.

Primer Pair	Forward Sequence	Reverse Sequence	Product (bp)
CDK-17	GGCATGGGAAGAACAGAAGA	CCATCTGCTTGGACTGTTGA	537
lvY-51393-short	GACCACTGTAGAGGAGGTTGG	GCTGCCTGTTTCTGGATGTC	124
lvY-79267-long	CAAGGCCAAAATGATCCCGC	TGTGCATTGACCATAAAGCCC	ca. 240

Table 2: Primer sequences used for the sex diagnostic multiplex PCR for use within the *L. vulgaris* species complex, as demonstrated in Fig. 1. CDK-17 is a control marker which amplifies in all species, lvY-51393-short and lvY-79267-long amplify only in males, however inclusion of both is recommended as some taxa may fail to amplify one the markers.

Linkage map construction

The 148 individuals of the linkage map family yielded a total of 1.34 billion reads (per sample median: 7.91 M, interquartile range: 7.04-8.92 M). The denovo_map.pl pipeline produced a total of 414,146 RAD loci of which 137,538 (33.2%) were present in at least 50% of individuals. After filtering with VCFtools a total of 16,738 markers (each with 1 representative SNP) were available for linkage map construction. The final linkage maps consist of 12 linkage groups. The sex-averaged map contains 10,763 markers and has total length of 1,366 cM (Fig. 2). Respectively, the paternal and maternal maps) contain 7,484 and 7,452 markers and have total lengths of 1,300 and 1,688 cM (Sup. Figs. 5-6, Sup. Table S1). The sex-averaged and paternal maps include 32 Y-linked presence/absence markers. In both maps these form a tight cluster, spanning less than 2 cM, located at one end of linkage group 5.

Figure 2: Sex-averaged linkage map for *L. vulgaris* based on a full-sib family of 146 offspring. The linkage map is composed of 10,763 RAD markers in 12 linkage groups, ordered by length in centimorgans. Thirty-two Y-linked presence/absence markers (highlighted in red), first identified in known-sex adult *L. vulgaris* are located within a 2 cM region of Group 5, identifying this as the Y-chromosome.

Comparison with *Pleurodeles waltl* genome

Five hundred and seventy-nine (5.4%) of the markers placed on the linkage map, including two Y-linked markers, can be aligned with sequences within the *P. waltl* genome assembly (Fig. 3). Synteny between the taxa appears strongly conserved, with each linkage group reciprocally matching a single *P. waltl* chromosome, 472 (82%) *L. vulgaris* markers mapping to their orthologous chromosome, and large blocks of conserved synteny are observed within each chromosome/linkage group pair. Evidence of a large inversion is also seen on linkage group 10. We observe no clear pattern in the 18% of markers mapping to non-orthologous chromosomes, indicating that these are a likely a result of misalignment of paralogous sequences, rather than evidence of any large-scale genomic rearrangements. Linkage group 5 is clearly orthologous to *P. waltl* chromosome 5. The two Y-linked presence/absence markers that have identifiable orthologs both align with sequences close to one end of chromosome 5, with start co-ordinates of 35.2 and 62.5 Mbp (the overall length of chromosome 5 is 1.91 Gbp).

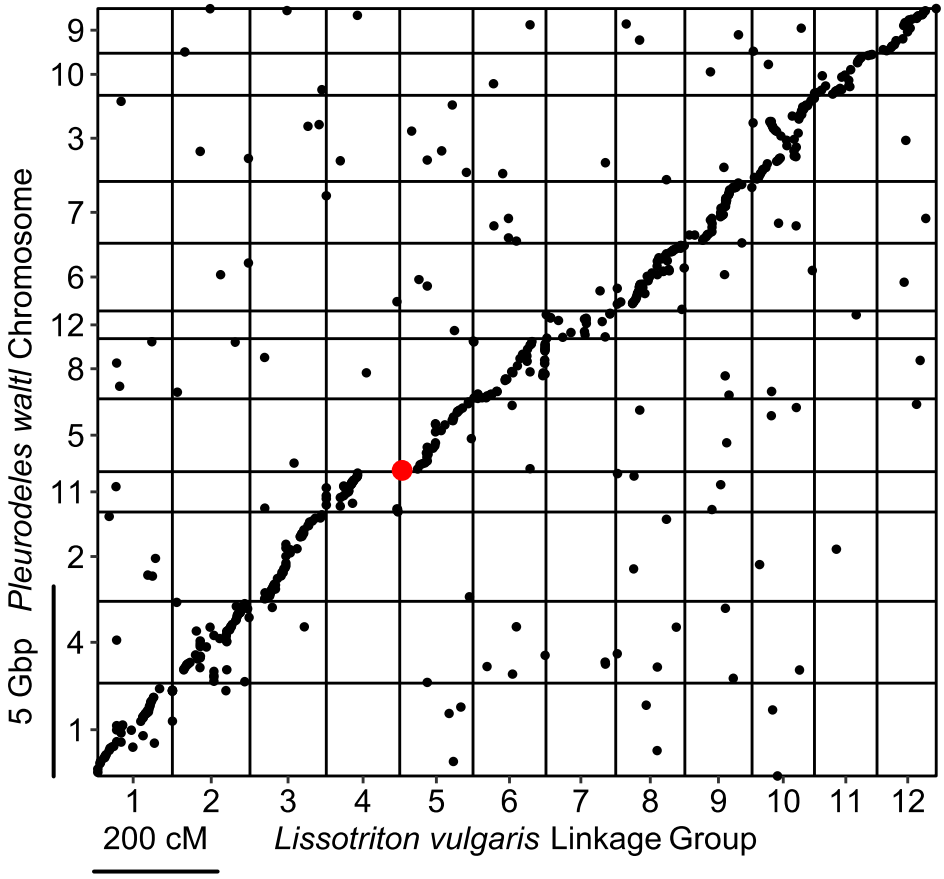

Figure 3: Oxford plot comparing the locations of 579 RAD markers in the *L. vulgaris* linkage map with their orthologs within the *P. waltl* genome, as assembled by (Brown et al. 2025). Two Y-linked presence/absence markers are highlighted in red. Four hundred and seventy-two markers (82%) map to orthologous chromosomes, demonstrating broad synteny between the two newt genera.

Marker density analysis

Marker density differs significantly between paternal and maternal linkage maps (Fig. 4). In the male map each linkage group is dominated by a single, tight cluster of markers, indicating large regions of reduced recombination. The clusters are usually in the centre of the group, suggesting that most recombination events occur near the ends of the chromosomes. In the maternal map marker density is more uniform, although areas of increased marker density are observed towards the end of some linkage groups. As expected, the Y-linked presence/absence markers are found in a region that shows high marker density in the paternal map but not in the maternal map. However such regions can be observed across the linkage map, which is an expected consequence of the differing rates of recombination in male and female meiosis. While

linkage group 5 does show the greatest difference in length and average marker density between the paternal and maternal map of any of the groups, the peak of marker density six other paternal linkage groups exceeds that of the Y-linked region. Peak marker density in the paternal map is found in linkage group 10, where 337 markers map to a single point.

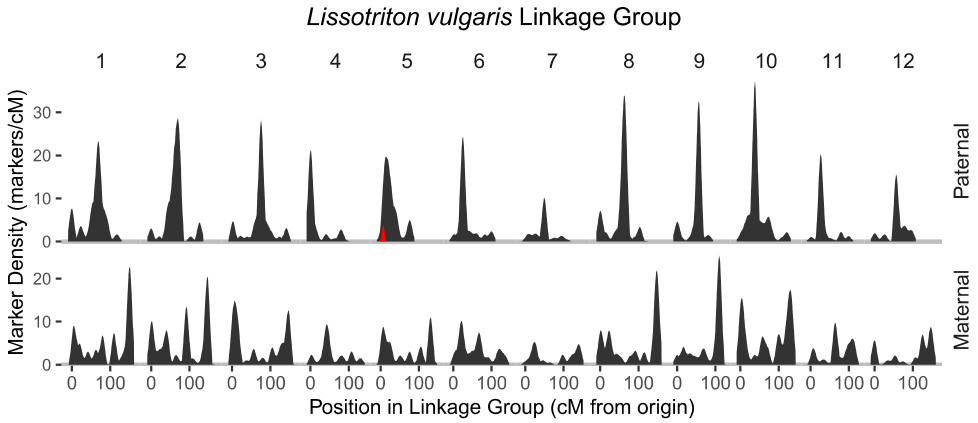

Figure 4: Marker density in the paternal and maternal *L. vulgaris* linkage maps. Total maker density is shown in black, and density of Y-linked presence/absence markers is shown in red, highlighting the Y-linked region on linkage group 5. Markers were aggregated into 2 cM bins and a Gaussian smoothing function with a 10 cM range was applied.

Paternal specific markers and SNPs

3,389 paternal specific markers and 9,414 paternal specific SNPs were located within the sex-averaged linkage map. Distribution across linkage groups was extremely uniform, with a linear trend observed between the number of total markers and paternal specific markers/SNPs within each linkage group (Fig. 5). Neither linkage group 5 nor any other group deviated significantly from this trend. The region containing the majority of the Y-linked presence absence markers has a significantly elevated concentration of both paternal specific markers ($P = 1.38 \times 10^{-4}$) and SNPs ($P = 7.77 \times 10^{-10}$), however more significant concentrations are present at multiple locations throughout the linkage map (Fig. 6). Eighteen 2 cM bins have a more significant concentration of paternal specific markers and eight have a more significant concentration of paternal SNPs.

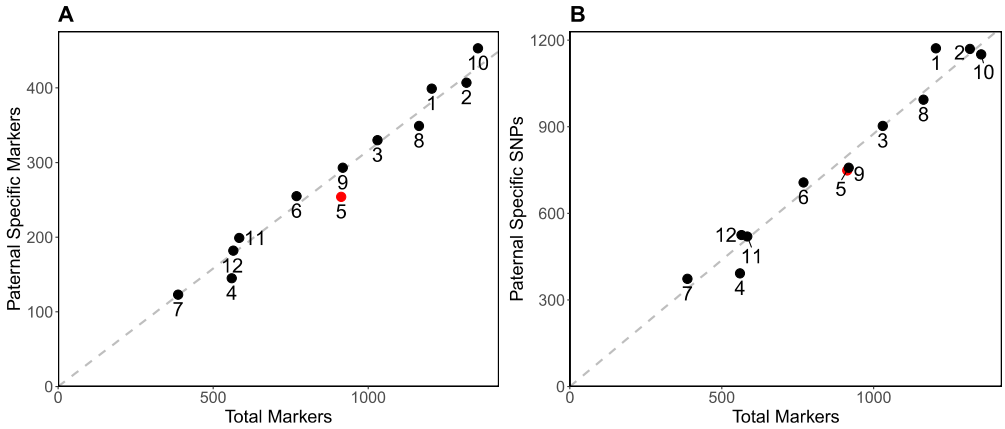

Figure 5: Plots showing the number of the paternal specific markers (A) and SNPs **(B)** against the total number of markers in each group of the sex-averaged *L. vulgaris* linkage map. Linkage group 5, baring the Y-linked presence/absence markers is highlighted in red.

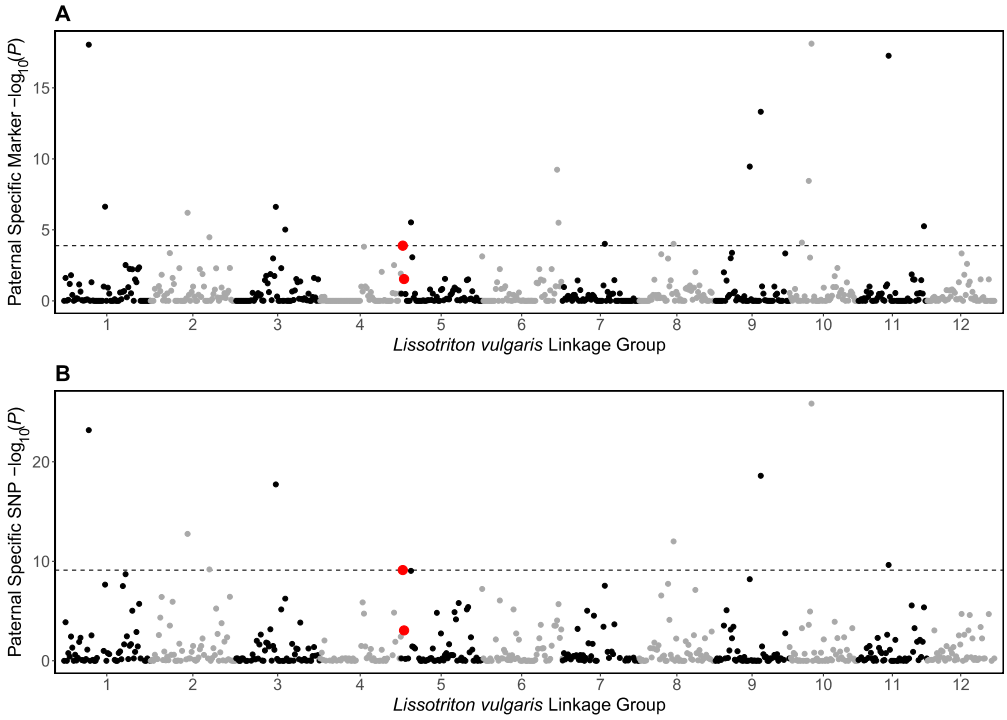

Figure 6: Manhattan style plot showing enrichment in paternal specific markers (A) and SNPs **(B)** in the sex-averaged *L. vulgaris* linkage map, divided into 689 bins of length 2 cM. Enrichment is shown as $-\log_{10}$ of the probability of a bin containing the observed number of paternal specific markers and SNPs. The 32 Y-linked presence absence markers are located in two adjacent bins highlighted in red. As expected, these bins show significant enrichment in paternal-specific markers ($p = 1.38 \times 10^{-4}$) and SNPs ($p = 7.77 \times 10^{-10}$). However other regions of the genome show even greater enrichment, with 18 and 8 bins exceeding the significance of the Y-linked region (shown as the dashed horizontal line) in paternal specific markers and SNPs respectively.

Discussion

We successfully identify sex-linked presence/absence markers in the smooth newt, *L. vulgaris* via an associative RADseq approach, confirming that this method remains effective even in exceptionally large genomes, such as those of Salamanders. As such genomic gigantism is often the result of accumulation of repetitive elements (Lee & Kim 2014; Sun et al. 2012) there is high chance that any given sequence will have multiple paralogs throughout the genome. Our results indicate that the efficiency of the discovery process can be significantly enhanced by aggressively filtering out candidate markers showing such paralogy. Out of 14 markers we test via PCR, five show validated sex association in *L. vulgaris*. Compared to the two previous studies using a similar methodology in salamanders, this is a notably high success rate. Hime et al. (2019) screened 43 loci for sex association in the hellbender, *Cryptobranchus alleganiensis*, with four successful in PCR. Hu et al. (2019) designed 100 candidate primer pairs to yield four reliable W-linked markers in the Chinese giant salamander, *Andrias davidianus*. In addition, our experience indicates that thoughtful optimisation of the upstream bioinformatics increases the chance of identifying a useful marker. Increasing the value parameter M in the denovo_map.pl/ustacks programs of the Stacks package from 2 to 10 vastly reduced the chance of a marker with paralogs being selected as a sex-linked candidate and doubled the number of candidates that gave sex-specific PCR amplification.

To our knowledge these presence/absence markers are the first tool for genetic sex identification described in any species of newt (subfamily Pleurodelinae) or salamander within the family Salamandridae. Two markers, LvY-79267 and LvY-51393, are particularly notable, as we show that in combination they allow for molecular sex identification across the smooth newt species complex with a simple a multiplex PCR protocol. As *Lissotriton* take 2-3 years to reach sexual maturity, and are difficult to morphologically sex as juveniles (Sparreboom 2014), a robust genetic assay of sex will be of significant benefit for researchers interested in the conservation, ecology and behaviour of these species.

The sex-specificity of the markers decreases with phylogenetic distance. Five markers are male-specific in *L. vulgaris*. Three of these retain specificity in *L. montandoni* – the most basal species within the smooth newt complex, which is notable as previous cytological studies were unable to identify any sex-chromosome in this species (Zbożeń et al. 1993). No marker is found to amplify sex-specifically in more distantly related *Lissotriton* species. Variation in the Y-chromosome may contribute to the differing degrees of reproductive isolation within the genus (Johnson et al. 2012; Yoshida et al. 2014) – while species within the smooth newt complex hybridise readily,

L. vulgaris and *L. helveticus* (the palmate newt) show almost complete reproductive isolation, despite co-occurring over a large area of western Europe (Miralles et al. 2024).

We were unable to obtain known-sex samples of two *Lissotriton* taxa. The Caucasian smooth newt *L. lantzi* is a member of the smooth newt complex (Wielstra et al. 2018) and so we predict that the identified markers will also be sex-specific in this species. *L. maltzani* has recently been recognised as a separate species from its close relative the Iberian newt, *L. boscai* (Sequeira et al. 2020; Speybroeck et al. 2020), but can be expected to show a similar, non-specific result with the markers described above.

We construct a high-density linkage map for *L. vulgaris* and identify a Y-determining region at the end of linkage group 5 via the incorporation of the markers identified above. The resulting map matches the observed karyotype of *L. vulgaris* (Wickbom 1945), and the number and density of markers is significantly increased compared to *L. vulgaris* x *L. montandoni* linkage maps published by (Niedzicka et al. 2017). A disadvantage of RADseq-based linkage maps is that the information they provide typically lacks context; it is difficult to directly relate a given linkage group to a particular chromosome, and the mapped sequences are unlikely to have any known function. To provide useful context, we attempt to align the *L. vulgaris* linkage map with the *P. waltl* genome assembly (Brown et al. 2025). Surprisingly, given the challenges of aligning short, non-coding sequences with a highly paralogous genome that diverged over 60 million years ago (Marjanović & Laurin 2014; Zhang & Wake 2009) a useful comparison is possible and shows that genome-level synteny between the two genera is highly conserved.

However, we are unable to confidently identify a sex-determining region of this linkage map without using additional information. We do observe considerable variation in intra-group marker density between the paternal and maternal maps, but this is not restricted to any one linkage group. In general, we show that *L. vulgaris* conforms to an extremely widespread pattern where male and female meiosis differ significantly, with males experiencing relatively more recombination near the telomeres and less recombination closer to the centromere of chromosomes (Sardell & Kirkpatrick 2020). Linkage group 5 varies slightly, as the telomeric region in which the Y-linked markers cluster shows reduced paternal recombination. This could be interpreted as restricted recombination between the X and Y chromosomes, however a similar phenomenon can also be observed in the autosomal linkage group 4.

Brelsford et al. (2016) reported a significant excess of paternal specific SNPs on the linkage group corresponding to the Y-chromosome of the European tree frog, *Hyla arborea*. In our study however, we do not observe RAD markers baring paternal specific SNPs to be more prevalent on a particular linkage group or form an obvious cluster within any linkage group. The likely explanation is that the sex-determining region of *L. vulgaris* is simply too small to be observable with this methodology - especially given

the enormous size of the overall genome and the consequently low per base pair recombination rate.

As recombination frequency, and the density of RAD markers will vary over the length of each chromosome, it is not possible to estimate the size of the Y-linked region in *L. vulgaris* with any accuracy. However, we can identify orthologous loci for two Y-linked presence/absence sequences within *P. waltl* chromosome 5, which corresponds to our linkage group 5. The orthologs are separated by just 27.3 Mbp, approximately 1.4% of the overall chromosome length. If this is reflective of the situation in the *L. vulgaris* Y-linked region, it would explain the absence of an obvious cluster of paternal SNPs. A region this small would contain too few RAD markers to stand out against the genetic background, which will contain many other clusters of paternal (and maternal) specific SNPs.

It is somewhat unexpected that the non-recombing region is not more prominent. A cytological study by Schmid et al. (1979) reported that, in male *L. vulgaris*, no chiasmata were observed in any region of the long arm of chromosome 5, which was hence identified as the Y-chromosome, despite being largely homomorphic. While we do observe reduced paternal recombination at the end of linkage group 5 where the Y-linked markers are placed, there is still a significant degree of recombination occurring within this region, the markers have not been collapsed to a single point on the paternal map. In addition, if the non-recombing Y-linked region covered the entirety of the long arm, we would expect a far great greater density of paternal specific SNPs and markers than we observe. It is possible that *L. vulgaris* shows regional diversity in the structure and recombination frequency of the Y-chromosome, as is noted in other amphibians, including *Hyla arborea* (Dufresnes et al. 2014) and the alpine newt *Ichthyosaura alpestris* (Herrero & López-Fernández 1986). Such diversity may explain the discrepancy between the karyology, performed on a population gathered near Ulm, Germany, and our linkage map, generated from a Polish population collected in the vicinity of Kraków.

We conclude that, while linkage maps are of great benefit for locating previously discovered sex-linked markers within a genome, their utility for identifying sex-linked regions without *a priori* knowledge is strongly dependent on the size of the region of supressed recombination. In species with very large genomes and small sex-linked regions, the technique is unlikely to be successful. However, we show that an associative RADseq approach can still be highly effective even in these situations, especially when measures are taken to suppress the selection of markers with autosomal paralogs.

Data Availability

All raw reads can be found as a part of the NCBI accession associated with Bioproject: PRJNA1118769. (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1118769) (France et al. 2024a). Information on samples and the positions and sequences of all markers in the *Lissotriton* RADseq linkage map can be found in a .xlsx file hosted together with scripts and bioinformatic pipelines used for analysis in the associated Zenodo repository (https://doi.org/10.5281/zenodo.13870462) (France et al. 2024b). Scripts are also available at an associated GitHub repository (https://github.com/Wielstra-Lab/Lissotriton_RADseq_Y).

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 802759). Samples were collected in accordance with the Polish General and Regional Inspectorates of Environmental Protection permits OP-I.6401 .32.2020.GZ, GDOŚ DZP-WG.6401.24.2021.TŁ and all experiments were accepted by II Local Ethical Committee for Animal Experiments in Kraków, permit 64/2020.

References

An J, Lee M, Min M-S, Lee M-H, Lee H (**2007**) A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea. *Forensic Science International*. 167: 59–61. DOI: 10.1016/j.forsciint.2005.12.031.

Babik W, Branicki W, Crnobrnja-Isailović J, Cogălniceanu D, Sas I, Olgun K, Poyarkov NA, Garcia-París M, Arntzen JW (**2005**) Phylogeography of two European newt species — discordance between mtDNA and morphology. *Molecular Ecology*. 14: 2475–2491. DOI: 10.1111/j.1365-294X.2005.02605.x.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, Perrin N, Ross L, Valenzuela N, Vamosi JC, Consortium TT of S (**2014**) Sex determination: why so many ways of doing it? *PLOS Biology*. 12: e1001899. DOI: 10.1371/journal.pbio.1001899.

Bayona-Vásquez NJ, Glenn TC, Kieran TJ, Pierson TW, Hoffberg SL, Scott PA, Bentley KE, Finger JW, Louha S, Troendle N, Diaz-Jaimes P, Mauricio R, Faircloth BC (**2019**) Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*. 7: e7724. DOI: 10.7717/peerj.7724.

Brelsford A, Dufresnes C, Perrin N (**2016**) 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. DOI: 10.1038/hdy.2015.83.

Brelsford A, Lavanchy G, Sermier R, Rausch A, Perrin N (2017) Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. *Molecular Ecology Resources*. 17: 752–759. DOI: 10.1111/1755-0998.12624.

Brown T, Mishra K, Elewa A, Iarovenko S, Subramanian E, Araus AJ, Petzold A, ... Simon A (2025) Chromosome-scale genome assembly reveals how repeat elements shape non-coding RNA landscapes active during newt limb regeneration. *Cell Genomics*. 100761. DOI: 10.1016/j.xgen.2025.100761.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics*. 10: 421. DOI: 10.1186/1471-2105-10-421.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*. 22: 3124–3140. DOI: 10.1111/mec.12354.

Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. *Heredity*. 95: 118–128. DOI: 10.1038/sj.hdy.6800697.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H (2014) Origins and functional evolution of Y chromosomes across mammals. *Nature*. 508: 488–493. DOI: 10.1038/nature13151.

Curzon AY, Shirak A, Zak T, Dor L, Benet-Perlberg A, Naor A, Low-Tanne SI, Sharkawi H, Ron M, Seroussi E (2021) All-male production by marker-assisted selection for sex determining loci of admixed *Oreochromis niloticus* and *Oreochromis aureus* stocks. *Animal Genetics*. 52: 361–364. DOI: 10.1111/age.13057.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics*. 27: 2156–2158. DOI: 10.1093/bioinformatics/btr330.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaScience*. 10: giab008. DOI: 10.1093/gigascience/giab008.

Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, Fuller RC, Bourne GR, Larkin DM, Breden F, Mank JE (2019) Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proceedings of the National Academy of Sciences*. 116: 19031–19036. DOI: 10.1073/pnas.1905298116.

De R, Joshi BD, Shukla M, Pandey P, Singh R, Goyal SP (2019) Understanding predation behaviour of the tiger (*Panthera tigris tigris*) in Ranthambore tiger Reserve, Rajasthan, India: use of low-cost gel based molecular sexing of prey hairs from scats. *Conservation Genetics Resources*. 11: 97–104. DOI: 10.1007/s12686-017-0963-2.

van Doorn GS, Kirkpatrick M (2010) Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics*. 186: 629–645. DOI: 10.1534/genetics.110.118596.

Dufresnes C, Brelsford A, Perrin N (2014) First-generation linkage map for the European tree frog (*Hyla arborea*) with utility in congeneric species. *BMC Research Notes*. 7: 850. DOI: 10.1186/1756-0500-7-850.

Dufresnes C, Crochet P-A (2022) Sex chromosomes as supergenes of speciation: why amphibians defy the rules? *Philosophical Transactions of the Royal Society B: Biological Sciences*. 377: 20210202. DOI: 10.1098/rstb.2021.0202.

Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution*. 25: 283–291. DOI: 10.1016/j.tree.2009.12.004.

Ezaz T, Sarre SD, O'Meally D, Marshall Graves JA, Georges A (2010) Sex chromosome evolution in lizards: independent origins and rapid transitions. *Cytogenetic and Genome Research*. 127: 249–260. DOI: 10.1159/000300507.

France J, Babik W, Dudek K, Marszałek M, Wielstra B (2024a) [dataset] *Lissotriton vulgaris* linkage mapping and Y-chromosome identification. *NCBI*. http://www.ncbi.nlm.nih.gov/bioproject/1118769.

France J, Babik W, Dudek K, Marszałek M, Wielstra B (2024b) [dataset] *Lissotriton* RADseq Samples and Linkage Map Details and Sequences. *Zenodo*. DOI: 10.5281/zenodo.13870462.

Fridolfsson A-K, Ellegren H (1999) A simple and universal method for molecular sexing of non-ratite birds. *Journal of Avian Biology*. 30: 116–121. DOI: 10.2307/3677252.

Gamble T (2016) Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Molecular Ecology*. 25: 2114–2116. DOI: 10.1111/mec.13648.

Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D (2015) Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Molecular Biology and Evolution*. 32: 1296–1309. DOI: 10.1093/molbev/msv023.

Gamble T, Zarkower D (2014) Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Molecular Ecology Resources*. 14: 902–913. DOI: 10.1111/1755-0998.12237.

Gregory RT (2024) Animal genome size database. https://www.genomesize.com/.

Herrero P, López-Fernández C (1986) The meiotic system of Iberian species of the genus *Triturus* (Amphibia: Caudata). *Caryologia*. 39: 385–395. DOI: 10.1080/00087114.1986.10797801.

Hime PM, Briggler JT, Reece JS, Weisrock DW (2019) Genomic data reveal conserved female heterogamety in giant salamanders with gigantic nuclear genomes. *G3: Genes|Genomes|Genetics*. 9: 3467–3476. DOI: 10.1534/g3.119.400556.

Hrovatin K, Kunej T (2018) Genetic sex determination assays in 53 mammalian species: Literature analysis and guidelines for reporting standardization. *Ecology and Evolution*. 8: 1009–1018. DOI: 10.1002/ece3.3707.

Hu Q, Chang C, Wang Q, Tian H, Qiao Z, Wang L, Meng Y, Xu C, Xiao H (2019) Genome-wide RAD sequencing to identify a sex-specific marker in Chinese giant salamander *Andrias davidianus*. *BMC Genomics*. 20: 415. DOI: 10.1186/s12864-019-5771-5.

Hu Q, Liu Y, Liao X, Tian H, Ji X, Zhu J, Xiao H (2021) A high-density genetic map construction and sex-related loci identification in Chinese Giant salamander. *BMC Genomics*. 22: 230. DOI: 10.1186/s12864-021-07550-0.

Intha N, Chaiprasart P (2020) Development of molecular markers for sex determination in *Phoenix dactylifera L. Acta Horticulturae*. 425–432. DOI: 10.17660/ActaHortic.2020.1299.64.

Johnson NA, Lachance J (2012) The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences*. 1256: E1–E22. DOI: 10.1111/j.1749-6632.2012.06748.x.

Keinath MC, Timoshevskaya N, Timoshevskiy VA, Voss SR, Smith JJ (2018) Miniscule differences between sex chromosomes in the giant genome of a salamander. *Scientific Reports*. 8: 17882. DOI: 10.1038/s41598-018-36209-2.

Kitano J, Peichel CL (2012) Turnover of sex chromosomes and speciation in fishes. *Environmental Biology of Fishes*. 94: 549–558. DOI: 10.1007/s10641-011-9853-8.

Lee B-Y, Kim M-S, Choi B-S, Nagano AJ, Au DWT, Wu RSS, Takehana Y, Lee J-S (2019) Construction of high-resolution RAD-seq based linkage map, anchoring reference genome, and QTL mapping of the sex chromosome in the marine medaka *Oryzias melastigma*. *G3 Genes|Genomes|Genetics*. 9: 3537–3545. DOI: 10.1534/g3.119.400708.

Lee S-I, Kim N-S (2014) Transposable elements and genome size variations in plants. *Genomics & Informatics*. 12: 87–97. DOI: 10.5808/GI.2014.12.3.87.

Litvinchuk SN, Rosanov JM, Borkin LJ (2007) Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the Salamandridae (Urodela, Amphibia). *Genome*. 50: 333–342. DOI: 10.1139/G07-010.

Ma W-J, Rovatsos M (2022) Sex chromosome evolution: The remarkable diversity in the evolutionary rates and mechanisms. *Journal of Evolutionary Biology*. 35: 1581–1588. DOI: 10.1111/jeb.14119.

Marjanović D, Laurin M (2014) An updated paleontological timetree of lissamphibians, with comments on the anatomy of Jurassic crown-group salamanders (Urodela). *Historical Biology*. 26: 535–550. DOI: 10.1080/08912963.2013.797972.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*. 17: 240–248. DOI: 10.1101/gr.5681207.

Miralles A, Secondi J, Pabijan M, Babik W, Lemaire C, Crochet P-A (2024) Inconsistent estimates of hybridization frequency in newts revealed by SNPs and microsatellites. *Conservation Genetics*. 25: 215–225. DOI: 10.1007/s10592-023-01556-9.

Miura I (2017) Sex determination and sex chromosomes in Amphibia. *Sexual Development*. 11: 298–306. DOI: 10.1159/000485270.

Niedzicka M, Dudek K, Fijarczyk A, Zieliński P, Babik W (2017) Linkage map of *Lissotriton* newts provides insight into the genetic basis of reproductive isolation. *G3 Genes|Genomes|Genetics*. 7: 2115–2124. DOI: 10.1534/g3.117.041178.

Pabijan M, Zieliński P, Dudek K, Stuglik M, Babik W (2017) Isolation and gene flow in a speciation continuum in newts. *Molecular Phylogenetics and Evolution*. 116: 1–12. DOI: 10.1016/j.ympev.2017.08.003.

Palmer DH, Rogers TF, Dean R, Wright AE (2019) How to identify sex chromosomes and their turnover. *Molecular Ecology*. 28: 4709–4724. DOI: 10.1111/mec.15245.

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. 8: 1360–1373. DOI: 10.1111/2041-210X.12775.

Payseur BA, Presgraves DC, Filatov DA (2018) Introduction: Sex chromosomes and speciation. *Molecular Ecology*. 27: 3745–3748. DOI: 10.1111/mec.14828.

Pellicer J, Leitch IJ (2020) The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist*. 226: 301–305. DOI: 10.1111/nph.16261.

Peng W, Xu J, Zhang Y, Feng Jianxin, Dong C, Jiang L, Feng Jingyan, Chen B, Gong Y, Chen L, Xu P (2016) An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*). *Scientific Reports*. 6: 26693. DOI: 10.1038/srep26693.

Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M (2014) Convergent genetic architecture underlies social organization in ants. *Current Biology*. 24: 2728–2732. DOI: 10.1016/j.cub.2014.09.071.

Rafati N, Chen J, Herpin A, Pettersson ME, Han F, Feng C, Wallerman O, ... Andersson L (2020) Reconstruction of the birth of a male sex chromosome present in Atlantic herring. *Proceedings of the National Academy of Sciences*. 117: 24359–24368. DOI: 10.1073/pnas.2009925117.

Rastas P (2017) Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*. 33: 3726–3732. DOI: 10.1093/bioinformatics/btx494.

Renner SS (2014) The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*. 101: 1588–1596. DOI: 10.3732/ajb.1400196.

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*. 28: 4737–4754. DOI: 10.1111/mec.15253.

Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*. CSHL Press ISBN: 978-0-87969-576-7.

Sardell JM, Kirkpatrick M (2020) Sex differences in the recombination landscape. *The American Naturalist*. 195: 361–379. DOI: 10.1086/704943.

Schmid M, Olert J, Klett C (1979) Chromosome banding in amphibia. *Chromosoma*. 71: 29–55. DOI: 10.1007/BF00426365.

Sequeira F, Bessa-Silva A, Tarroso P, Sousa-Neves T, Vallinoto M, Gonçalves H, Martínez-Solano I (2020) Discordant patterns of introgression across a narrow hybrid zone between two cryptic lineages of an Iberian endemic newt. *Journal of Evolutionary Biology*. 33: 202–216. DOI: 10.1111/jeb.13562.

Sparreboom M (2014) *Salamanders of the old world: The salamanders of Europe, Asia and northern Africa*. KNNV Publishing DOI: 10.1163/9789004285620.

Speybroeck J, Beukema W, Dufresnes C, Fritz U, Jablonski D, Lymberakis P, Martínez-Solano I, Razzetti E, Vamberger M, Vences M, Vörös J, Crochet P-A (2020) Species list of the European herpetofauna – 2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica. *Amphibia-Reptilia*. 41: 139–189. DOI: 10.1163/15685381-bja10010.

Sun C, Shepard DB, Chong RA, López Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL (2012) LTR retrotransposons contribute to genomic gigantism in Plethodontid salamanders. *Genome Biology and Evolution*. 4: 168–183. DOI: 10.1093/gbe/evr139.

Trenkel VM, Boudry P, Verrez-Bagnis V, Lorance P (2020) Methods for identifying and interpreting sex-linked SNP markers and carrying out sex assignment: application to thornback ray (*Raja clavata*). *Molecular Ecology Resources*. 20: 1610–1619. DOI: 10.1111/1755-0998.13225.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research*. 40: e115. DOI: 10.1093/nar/gks596.

Vidya TNC, Kumar VR, Arivazhagan C, Sukumar R (2003) Application of molecular sexing to free-ranging Asian elephant (*Elephas maximus*) populations in southern India. *Current Science*. 85: 1074–1077.

Wickbom T (1945) Cytological studies on Dipnoi, Urodela, Anura, and Emys. *Hereditas*. 31: 241–346. DOI: 10.1111/j.1601-5223.1945.tb02756.x.

Wielstra B, Canestrelli D, Cvijanović M, Denoël M, Fijarczyk A, Jablonski D, Liana M, Naumov B, Olgun K, Pabijan M, Pezzarossa A, Popgeorgiev G, Salvi D, Si Y, Sillero N, Sotiropoulos K, Zieliński P, Babik W (2018) The distributions of the six species constituting the smooth newt species complex (*Lissotriton vulgaris sensu lato* and *L. montandoni*) – an addition to the New Atlas of Amphibians and Reptiles of Europe. *Amphibia-Reptilia*. 39: 252–259. DOI: 10.1163/15685381-17000128.

Wright AE, Dean R, Zimmer F, Mank JE (2016) How to make a sex chromosome. *Nature Communications*. 7: 12087. DOI: 10.1038/ncomms12087.

Yoshida K, Makino T, Yamaguchi K, Shigenobu S, Hasebe M, Kawata M, Kume M, Mori S, Peichel CL, Toyoda A, Fujiyama A, Kitano J (2014) Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLOS Genetics*. 10: e1004223. DOI: 10.1371/journal.pgen.1004223.

Zbożeń J, Rafiński J (1993) A comparative study of mitotic and meiotic chromosomes of the Montandon's newt, *Triturus montandoni* (Urodela: Salamandridae) from Poland and Rumania. *Genetica*. 88: 69. DOI: 10.1007/BF02424453.

Zhang P, Wake DB (2009) Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*. 53: 492–508. DOI: 10.1016/j.ympev.2009.07.010.

Supporting Information

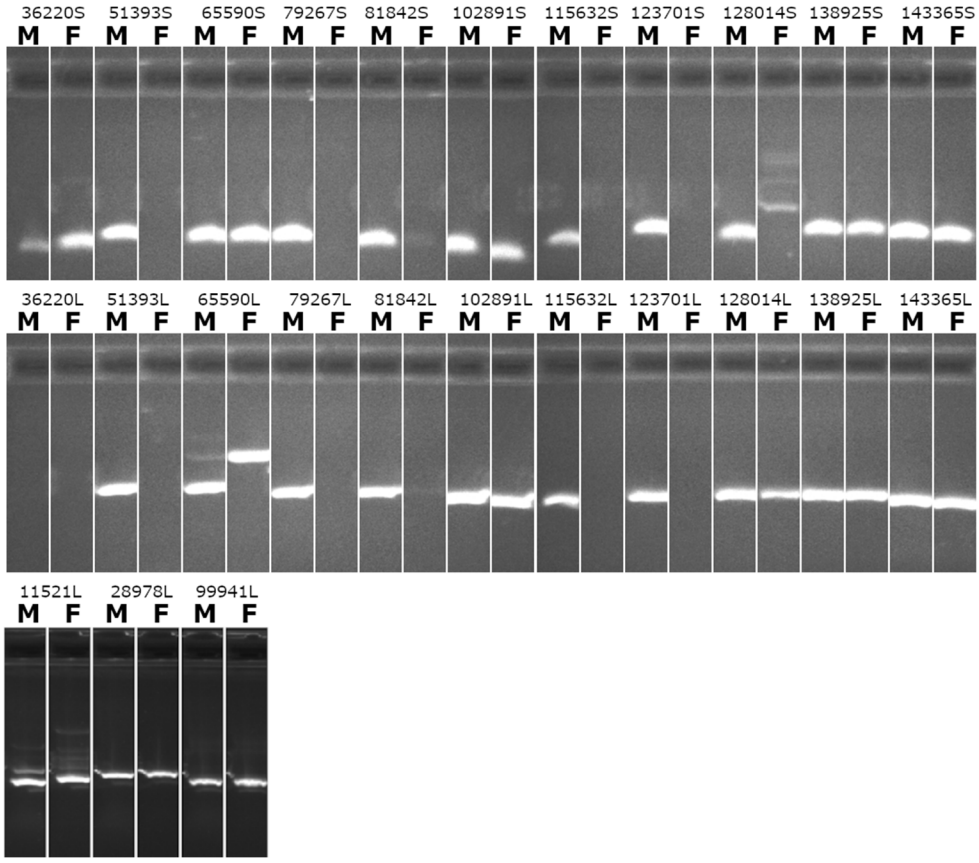

Figure S1: PCR screening of 25 primer pairs designed for candidate Y-linked markers for male specific amplification in *L. vulgaris*. Label M indicates the male sample and label F indicates female. Markers are indicated by number followed by either S (for primer pairs designed for the short product – c.a. 100 bp) or L (for primer pairs designed for the long product – c.a. 200 bp). 10 primer pairs, and 5 markers showed amplification only in the male sample. One further pair, LvY-128014-short showed strong amplification in the male and only weak amplification in the female. A single primer pair, LvY-36220-long, failed to amplify in either sample. The other 25 primer pairs amplified in both the male and female samples.

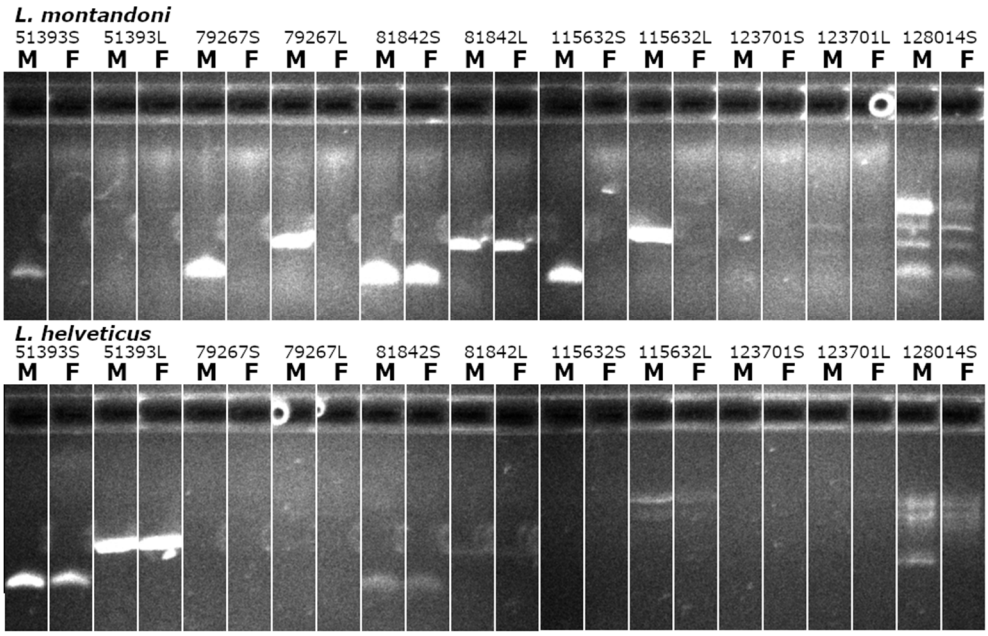

Figure S2: PCR screening of the 11 primer pairs showing successful male-specific amplification in *L. vulgaris*, in *L. montandoni* and *L. helveticus*. Label M indicates the male sample and label F indicates female. Markers are indicated by number followed by either S (for primer pairs designed for the short product – c.a. 100 bp) or L (for primer pairs designed for the long product – c.a. 200 bp). Five primer pairs, designed for three marker sequences, show male specific amplification in *L. montandoni*. Only two primer pairs show strong amplification in L. helveticus, and no male specificity is observed.

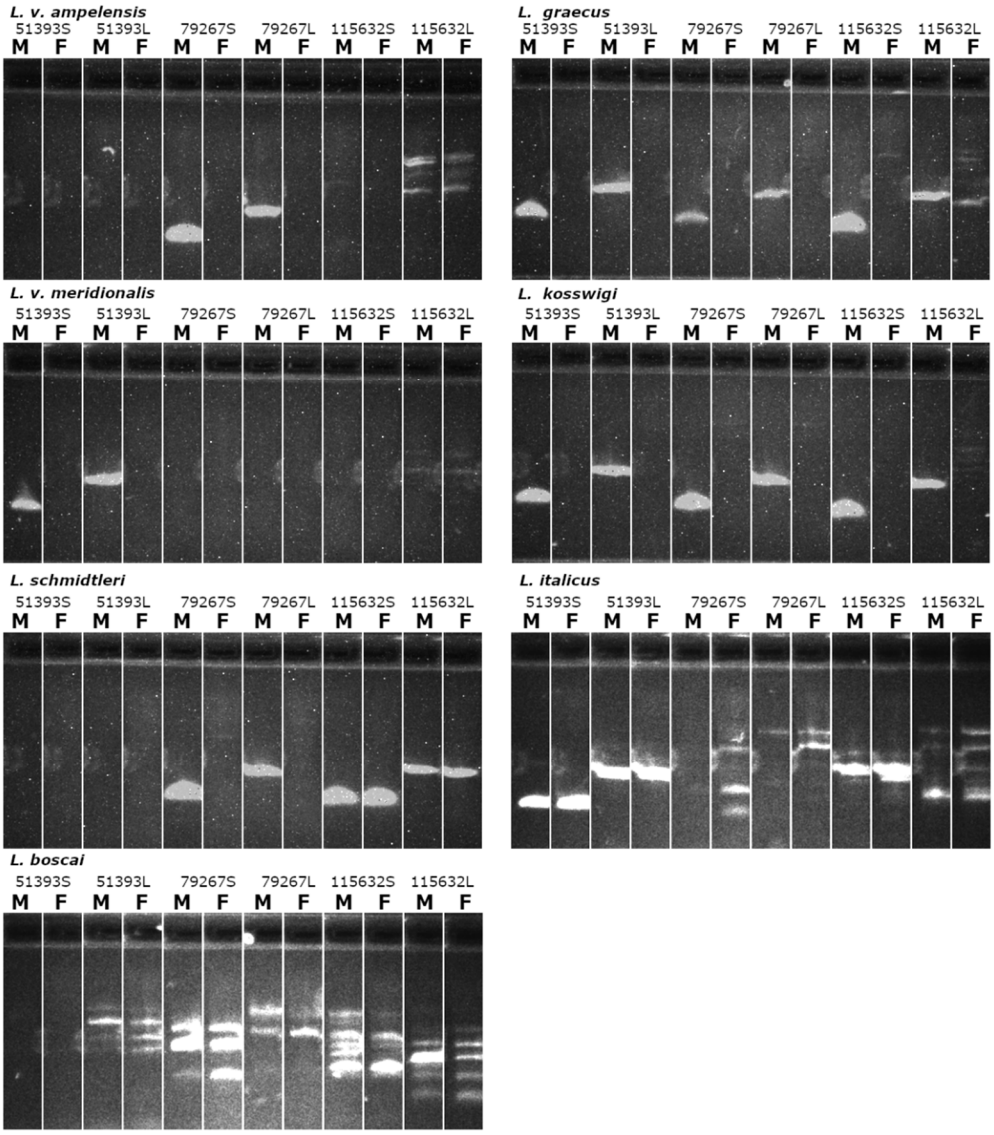

Figure S3: PCR screening in seven further *Lissotriton* taxa, using the primer pairs for the three markers that show male specificity in *L. montandoni*. For the five taxa in the *L. vulgaris* species complex, at least one marker shows male-specific amplification. In the two more distantly related species (*L. italicus* and *L. boscai*) no male-specific amplification is observed, and multiple non-target bands appear.

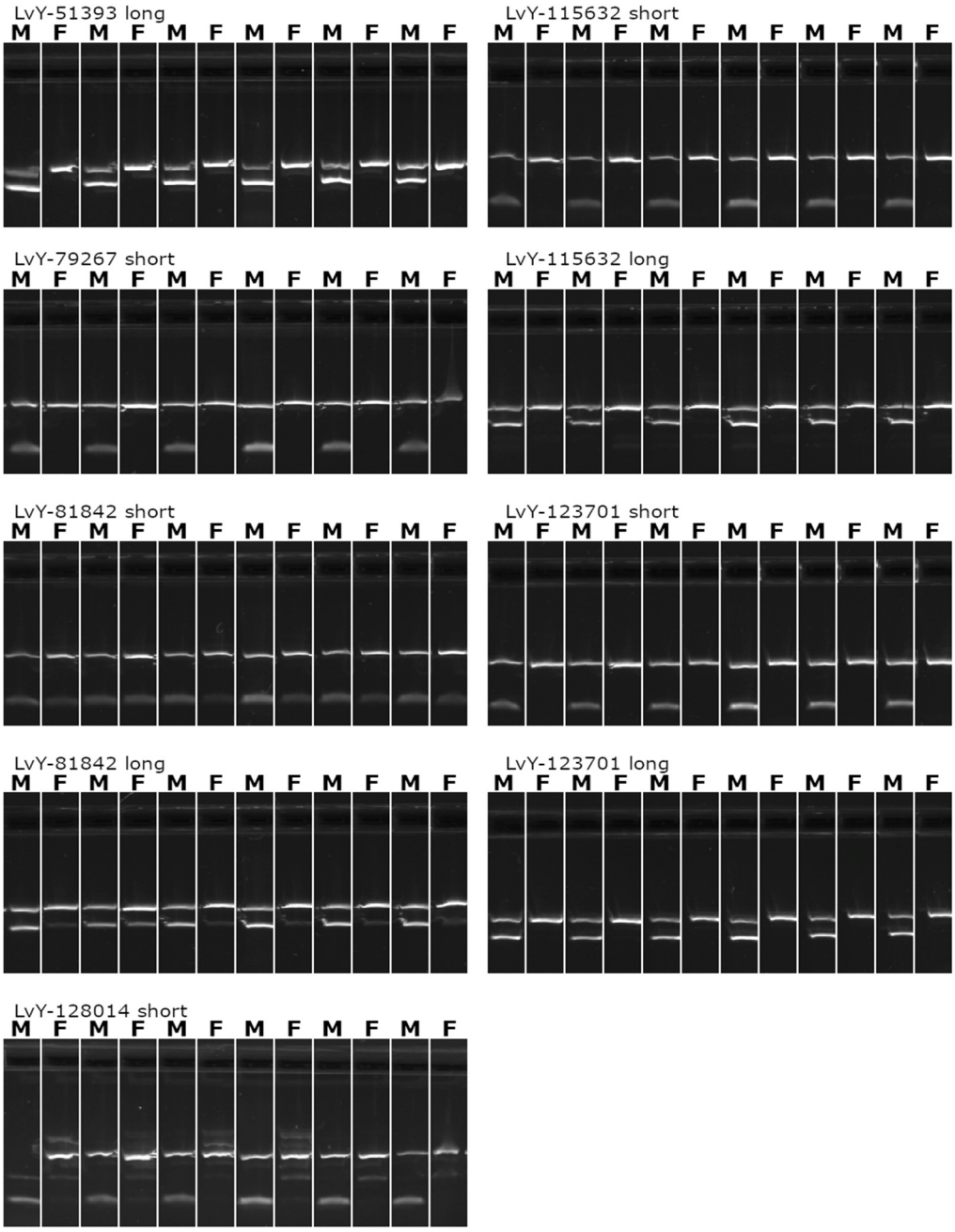

Figure S4: **PCR validation of primer pairs showing male specificity in *L. vulgaris*,** in a 12 individual panel (six male, six female, not including the two individuals previously used for screening). Primers amplifying CDK-17 were included as a control, and the resultant product appears above the test bands in all cases due to greater length (517 bp compared to 100-250 bp). Both primer pairs designed for LvY-81842 appear to amplify in their product in females and thus fail validation. All other primer pairs are validated as male specific.

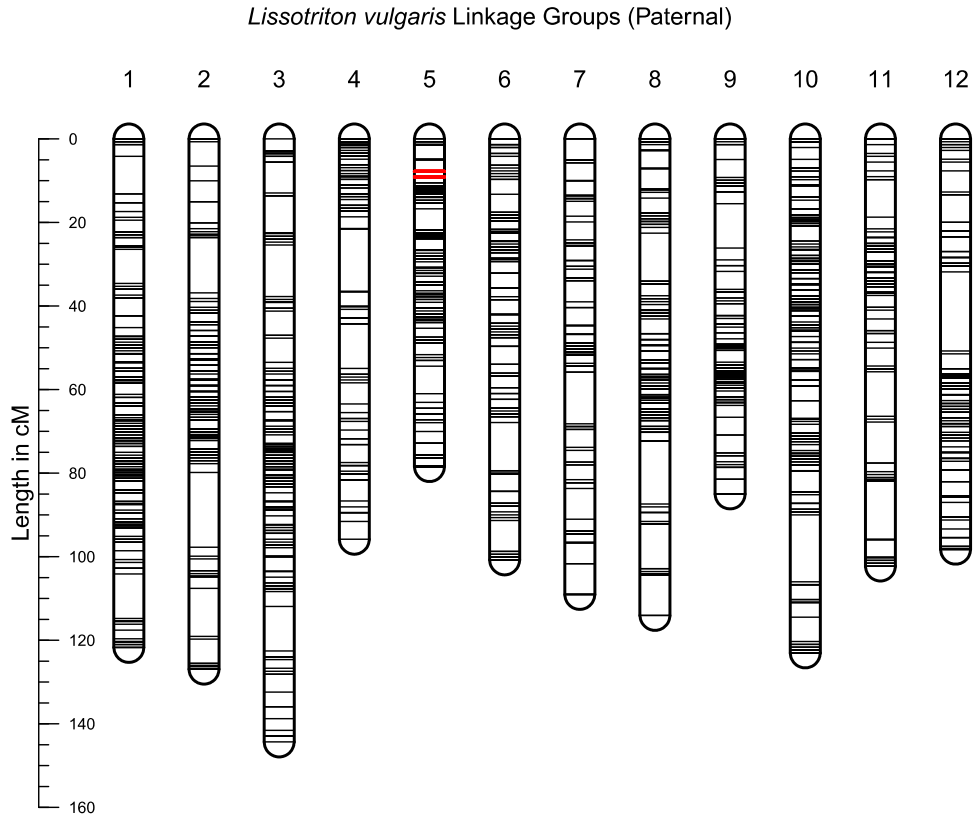

Figure S5: The paternal *L. vulgaris* linkage map, displaying 7,484 RAD markers across 12 linkage groups, including 32 Y-linked markers highlighted in red on linkage group 5. Groups are ordered according to the length of the corresponding group in the sex averaged map.

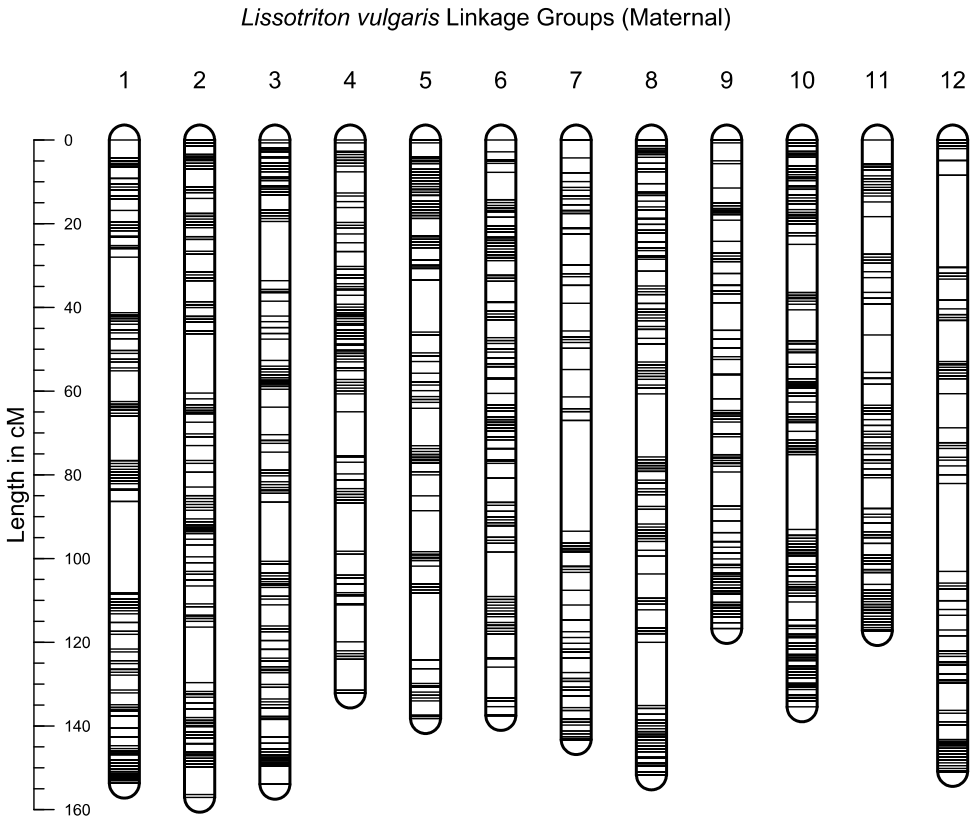

Figure S6: The maternal *L. vulgaris* linkage map, displaying 7,452 RAD markers across 12 linkage groups. Groups are ordered according to the length of the corresponding group in the sex averaged map.

Group	Sex-averaged		Paternal		Maternal	
	N markers	Length (cM)	N markers	Length (cM)	N markers	Length (cM)
1	1205	127.0	820	121.7	863	153.7
2	1317	125.4	924	126.9	916	157.0
3	1030	123.4	718	144.4	749	153.9
4	560	118.9	418	95.8	343	132.1
5	877	118.4	666	78.5	531	138.2
6	769	118.0	522	100.8	535	137.5
7	387	112.7	250	109.1	287	143.3
8	1164	111.9	831	114.1	789	151.7
9	918	108.5	626	85.0	648	116.7
10	1354	101.1	897	123.1	972	135.4
11	584	101.1	392	102.3	411	117.2
12	565	99.6	387	98.2	408	150.9

Table S1: Characteristics of the sex-averaged, paternal and maternal *Lissotriton vulgaris* linkage maps, including the number of markers and the length (in centimorgans) of each group.

Table S2: Sequences of all primers used in this study, CDK-17 is an autosomal marker used as a control, all others are candidate Y-linked markers developed for *Lissotriton vulgaris*.

Primer Pair	Forward Primer Sequence	Reverse Primer Sequence	Product (bp)
CDK-17	GGCATGGGAAGAACAGAAGA	CCATCTGCTTGGACTGTTGA	537
lvY-11521-long	GCATTTGGGCAGCTTCATTC	CAATTCAGGCACACACCAGC	>200
lvY-28978-long	TCATGCATAGCCAAAGAGTTTGTC	CCCTGATGACACTTGATCGC	>200
lvY-102891-short	CTAGATGCGCATCCACTGGG	CTGACATTAAGCAAGCCGCC	87
lvY-102891-long	GCGGCTTGCTTAATGTCAGG	CCCATAGTCTCCATGCCCTC	>200
lvY-99941-long	TTGCTGTGTGTACGTGCCAG	CGTTTGGATGGGATACAAGCAG	>200
lvY-138925-short	TGCCAATGACCAGCTCCTAC	TGGTAGCTACTCCTGGTGAAG	115
lvY-138925-long	TGCCAATGACCAGCTCCTAC	TCCACGAAGAACTGATAGAACTC	>200
lvY-81842-short	CTAGAATCTGCGGCGTCATG	TGAAGGTCACACTTTCCGCG	92
lvY-81842-long	TCAGTATGCCGTCTAGCTGC	ACCAGAGCCCCCGTTTATTG	>200
lvY-143365-short	TAGGGATCAGTTGGGGGAAC	CCGCAAAGCAAAAGAGACCC	106
lvY-143365-long	CCAGCATAAGGTGAGGAGGG	TACTGAAAAACCTGGCCCCC	>200
lvY-51393-short	GACCACTGTAGAGGAGGTTGG	GCTGCCTGTTTCTGGATGTC	124
lvY-51393-long	GACCACTGTAGAGGAGGTTGG	GATCCGTGGAGGTCGGTAAC	>200
lvY-128014-short	TTTTTGGGGGCTCTGCAGG	TGCTCAGTGTCTGTATCCTCTC	91
lvY-128014-long	GCGAGTAGATGGAAGGGTGG	TTGTTTGTCTTGCCCTTTGG	>200
lvY-65590-short	GCAGTGCAGTTCAGAGCATG	AGCCAGCACAAACAGATAGAG	104
lvY-65590-long	GCAGTGCAGTTCAGAGCATG	CAAAGCCTGTGTGCCAACTC	>200
lvY-36220-short	CTAGACTCACGCACACACCC	CCTCCTCCTCTCTCCCTAGC	97
lvY-36220-long	ACTGGTGCTAGGGAGAGAGG	GGCTTTCTTTCTCAGCACAGC	>200
lvY-123701-short	AGGCCTCAGTTCTTCTTGGG	GGTCCACTGTCCACATTGTG	126
lvY-123701-long	TGTTGCATTAGTCCTCTCCCC	GCAATTACGGACTCAGCGTTC	>200
lvY-115632-short	ACTCTACTGATACTTGCCATGCC	TGTCATCGAGCTTAGGCCAC	95
lvY-115632-long	TGTGGCCTAAGCTCGATGAC	ATTCCTCAGGGCTGTTGCAG	>200
lvY-79267-short	CAAGGCCAAAATGATCCCGC	ACTCTGGGAGCAGTAGTCAC	107
lvY-79267-long	CAAGGCCAAAATGATCCCGC	TGTGCATTGACCATAAAGCCC	>200