

# RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data

Romain Feron<sup>1,2,3</sup>  | Qiaowei Pan<sup>1,2</sup>  | Ming Wen<sup>1,4</sup> | Boudjema Imarazene<sup>1</sup> | Elodie Jouanno<sup>1</sup> | Jennifer Anderson<sup>1,5</sup>  | Amaury Herpin<sup>1</sup>  | Laurent Journot<sup>6</sup> | Hugues Parrinello<sup>6</sup> | Christophe Klopp<sup>7</sup>  | Verena A. Kottler<sup>8</sup> | Alvaro S. Roco<sup>8</sup> | Kang Du<sup>9,10</sup>  | Susanne Kneitz<sup>8</sup> | Mateus Adolphi<sup>10</sup> | Catherine A. Wilson<sup>11</sup> | Braedan McCluskey<sup>11</sup> | Angel Amores<sup>11</sup> | Thomas Desvignes<sup>11</sup>  | Frederick W. Goetz<sup>12</sup>  | Ato Takanashi<sup>13</sup> | Mari Kawaguchi<sup>13</sup> | Harry William Detrich III<sup>14</sup>  | Marcos A. Oliveira<sup>15</sup>  | Rafael H. Nóbrega<sup>15</sup>  | Takashi Sakamoto<sup>16</sup> | Masatoshi Nakamoto<sup>16</sup> | Anna Wargelius<sup>17</sup> | Ørjan Karlsen<sup>17</sup> | Zhongwei Wang<sup>8,18</sup> | Matthias Stöck<sup>19</sup>  | Robert M. Waterhouse<sup>2,3</sup>  | Ingo Braasch<sup>20</sup> | John H. Postlethwait<sup>11</sup>  | Manfred Schartl<sup>9,10</sup>  | Yann Guiguen<sup>1</sup> 

<sup>1</sup>INRAE, LPGP, Rennes, France

<sup>2</sup>Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

<sup>3</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>4</sup>State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan Normal University, Changsha, China

<sup>5</sup>Department of Organismal Biology, Systematic Biology, Uppsala University, Uppsala, Sweden

<sup>6</sup>Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Univ. Montpellier, Montpellier, France

<sup>7</sup>SIGENAE, Mathématiques et Informatique Appliquées de Toulouse, INRAE, Castanet Tolosan, France

<sup>8</sup>Physiological Chemistry, Biocenter, University of Wuerzburg, Wuerzburg, Germany

<sup>9</sup>Department of Chemistry and Biochemistry, The Xiphophorus Genetic Stock Center, Texas State University, San Marcos, TX, USA

<sup>10</sup>Developmental Biochemistry, Biocenter, University of Wuerzburg, Wuerzburg, Germany

<sup>11</sup>Institute of Neuroscience, University of Oregon, Eugene, OR, USA

<sup>12</sup>Environmental and Fisheries Sciences Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

<sup>13</sup>Department of Materials and Life Sciences, Faculty of Science and Technology, Sophia University, Tokyo, Japan

<sup>14</sup>Department of Marine and Environmental Sciences, Marine Science Center, Northeastern University, Nahant, MA, USA

<sup>15</sup>Reproductive and Molecular Biology Group, Department of Structural and Functional Biology, Institute of Biosciences, São Paulo State University, Botucatu, Brazil

<sup>16</sup>Department of Aquatic Marine Biosciences, Tokyo University of Marine Science and Technology, Tokyo, Japan

<sup>17</sup>Institute of Marine Research, Bergen, Norway

<sup>18</sup>Institute of Hydrobiology, Chinese Academy of Sciences, Beijing, China

<sup>19</sup>Leibniz-Institute of Freshwater Ecology and Inland Fisheries, IGB, Berlin, Germany

<sup>20</sup>Department of Integrative Biology, Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA

**Correspondence**

Yann Guiguen and Romain Feron, INRAE,  
LPGP, Rennes, France.  
Email: romainferon91@gmail.com

**Funding information**

National Institute of Health (USA),  
Grant/Award Number: R01GM085318  
and R35GM139635; Deutsche  
Forschungsgemeinschaft, Grant/Award  
Number: SCHA 408/10-1; National  
Science Foundation (USA) Office of Polar  
Programs, Grant/Award Number: OPP-  
1543383, PLR-1247510 and PLR-1444167;  
Agence Nationale de la Recherche,  
Grant/Award Number: ANR-10-INBS-09  
and SCHA 408/10-1; Schweizerischer  
Nationalfonds zur Förderung der  
Wissenschaftlichen Forschung, Grant/  
Award Number: PP00P3\_170664

**Abstract**

The study of sex determination and sex chromosome organization in nonmodel species has long been technically challenging, but new sequencing methodologies now enable precise and high-throughput identification of sex-specific genomic sequences. In particular, restriction site-associated DNA sequencing (RAD-Seq) is being extensively applied to explore sex determination systems in many plant and animal species. However, software specifically designed to search for and visualize sex-biased markers using RAD-Seq data is lacking. Here, we present RADSex, a computational analysis workflow designed to study the genetic basis of sex determination using RAD-Seq data. RADSex is simple to use, requires few computational resources, makes no prior assumptions about the type of sex-determination system or structure of the sex locus, and offers convenient visualization through a dedicated R package. To demonstrate the functionality of RADSex, we re-analysed a published data set of Japanese medaka, *Oryzias latipes*, where we uncovered a previously unknown Y chromosome polymorphism. We then used RADSex to analyse new RAD-Seq data sets from 15 fish species spanning multiple taxonomic orders. We identified the sex determination system and sex-specific markers in six of these species, five of which had no known sex-markers prior to this study. We show that RADSex greatly facilitates the study of sex determination systems in nonmodel species thanks to its speed of analyses, low resource usage, ease of application and visualization options. Furthermore, our analysis of new data sets from 15 species provides new insights on sex determination in fish.

**KEYWORDS**

computational workflow, fish, RAD-Sequencing, sex determination, visualization

**1 | INTRODUCTION**

Sexual reproduction is widespread in animals (Goodenough & Heitman, 2014) and, in gonochoristic species, involves segregation of male and female gonadal functions into separate individuals for their entire lives (Bachtrog et al., 2014). Males produce small, mobile gametes while females produce large, immobile gametes, and this difference has fuelled divergent evolution of morphological, physiological and behavioural traits between the two sexes. Individuals acquire sex-specific traits during sexual development, which starts with sex determination, the process that controls whether an individual develops male or female reproductive organs. Sex determination can be triggered by genetic factors (genetic sex determination, GSD), as in virtually all mammals and birds (Bachtrog et al., 2014); by environmental factors (environmental sex determination, ESD), such as temperature in many nonavian reptiles (Pezaro et al., 2017); or a combination of the two, as in several fish species (Piferrer et al., 2005). When genetic factors are involved, sex determination is entirely or partially controlled by one or multiple master sex-determining (MSD) genes located on sex chromosomes. Genetic sex determination systems can involve male heterogamy (XX/XY and XX/X0), female heterogamy (ZZ/ZW and ZZ/Z0), or multiple loci on different chromosomes (polygenic systems; Bachtrog et al., 2014).

In male- and female-heterogametic systems, sex chromosomes can be morphologically different (i.e., heteromorphic) or similar (i.e., homomorphic; Bachtrog et al., 2014). Initially, the overwhelming majority of knowledge on the structure and evolution of sex chromosomes came from studies in mammals (Wallis et al., 2008) and in *Drosophila* (Salz & Erickson, 2010), where it has been observed that suppression of recombination between the X and Y chromosomes led to degeneration of the Y. These findings have spawned both theoretical and empirical interest in sex chromosomes as models to study the consequences of recombination suppression and associated processes (Bachtrog, 2008; Charlesworth & Charlesworth, 2000; Corcoran et al., 2016; Doorn & Kirkpatrick, 2007; Gu et al., 2017; Huylmans et al., 2017; Muyle et al., 2012; Peichel et al., 2004). Recent studies investigating sex chromosomes in insects (Blackmon et al., 2017), nonavian reptiles (Modi & Crews, 2005), amphibians (Miura, 2017) and fishes (Kikuchi & Hamaguchi, 2013) have found many homomorphic sex chromosomes displaying varying levels of differentiation, with the extreme case of a sex locus restricted to allelic variation of a single nucleotide as reported in the Japanese pufferfish (Kamiya et al., 2012). These results called into question the single unified concept of sex chromosome evolution and highlighted the importance of obtaining a broader understanding of sex determination and sex chromosomes in many species across the tree of

life. The first step in this process is identifying sex-specific genomic sequences, that is sequences from the sex locus that are found in only one of the two sexes because of allelic divergence between the sex chromosomes or because of large insertions in the hemizygous chromosome. These sex-specific sequences can then be aligned to a reference genome to locate the sex locus, identify candidate MSD genes and other genes involved in sex determination, and characterize patterns of differentiation between the sex chromosomes. In addition, facilitating the identification of such sequences has practical applications: sex is an important factor in ecological (Benestan et al., 2017) and conservation studies (Ancona et al., 2017), as well as agriculture (Al-Ameri et al., 2016; Liao et al., 2017; Spigler et al., 2008) and animal production (Dan et al., 2013; Yano et al., 2013), yet the sex of an individual cannot always be easily determined by noninvasive methods.

Until recently, discovery of sex loci was not readily feasible mainly due to technical barriers preventing the precise and high-throughput identification of sex-specific genomic sequences, especially in non-model species. Genetic mapping has for a long time been the traditional approach for studying genetic sex determination (Palmer et al., 2019), but recent advances in sequencing technologies have enabled the exploration of a much broader spectrum of taxa by directly comparing genomes from phenotypically distinct males and females. In essence, this process is akin to comparing genomic differences between two populations or between a mutant and a wild type genotype, and therefore methods from molecular genetics, population genetics and ecology can be applied. A popular representational approach to comparing the genetics of populations is restriction site-associated DNA sequencing, or RAD-Seq (Andrews et al., 2016). RAD-Seq generates short sequences for a small but consistent fraction of the genome and thus allows the sequencing and comparison of multiple individuals from several populations at relatively low cost (Davey et al., 2011) and without requiring additional genomic resources. RAD-Seq has been successfully used to identify sex-specific sequences in non-model species from diverse taxa, including fish (Drinan et al., 2018), amphibians (Bewick et al., 2013), nonavian reptiles (Gamble, 2016; Gamble et al., 2015, 2017, 2018; Gamble & Zarkower, 2014; Nielsen et al., 2018, 2019), invertebrates (Carmichael et al., 2013; Mathers et al., 2015; Pratlong et al., 2017) and plants (Kafkas et al., 2015), and to identify the sex chromosomes and the sex locus in some of these species (Wilson et al., 2014). Although dedicated pipelines have been developed to analyse RAD-Seq data specifically for sex determination (Gamble & Zarkower, 2014), most of the above-mentioned studies have used *STACKS* (Catchen et al., 2011, 2013; Hohenlohe et al., 2012; Rochette & Catchen, 2017) to cluster RAD-Seq reads into polymorphic markers subsequently filtered with custom in-house scripts. However, *STACKS* requires substantial computational resources to run on large data sets and depends on multiple parameters that can greatly influence the outcome of the analysis (Paris et al., 2017; Rodríguez-Ezpeleta et al., 2016; Shafer et al., 2018), for instance in the context of sex determination, the classification of a marker as a sex-specific sequence (e.g., from a large sex-specific insertion) or as a sex-specific allele (e.g., an allele specific to one sex in a polymorphic

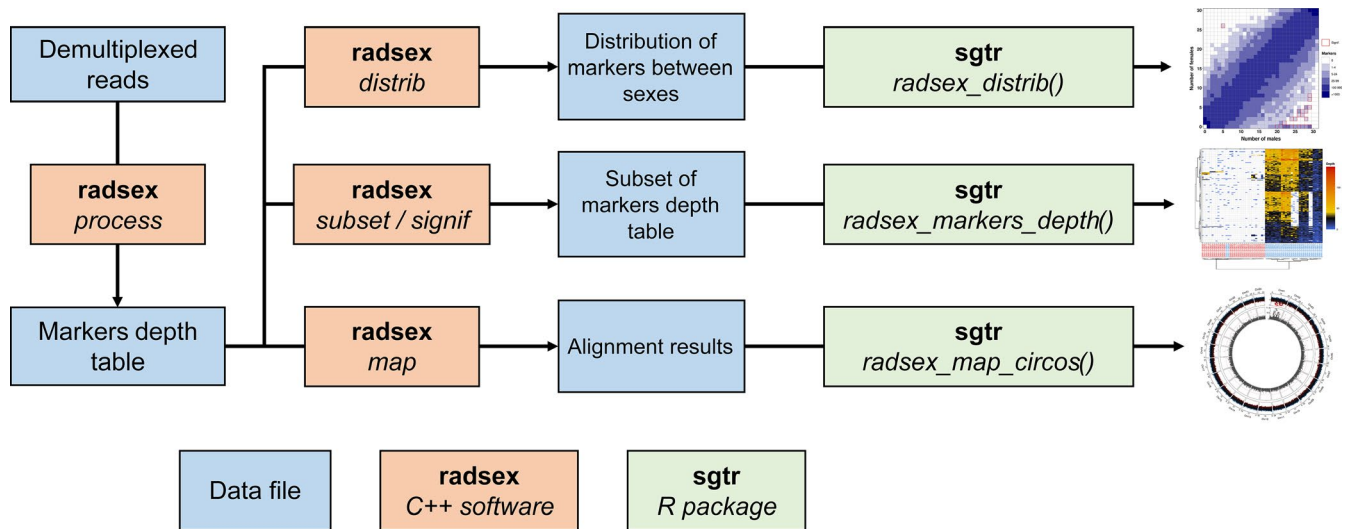
locus; Utsunomia et al., 2017). In addition, the current visualization tools for RAD-Seq data make the interpretation of results from some data sets difficult, for instance data sets containing strong population structure, individuals with misassigned phenotypic sex or sex-bias in sequencing depth.

To overcome these limitations, we developed RADSEX, a RAD-Seq data analysis computational workflow specifically designed to study sex determination. RADSEX is simple to use, makes no assumptions about whether sex-biased markers are sex-biased sequences or sex-biased alleles in a polymorphic locus, requires few resources, and offers helpful visualization tools in the form of an R package. To demonstrate the relevance of RADSEX, we analysed a previously published data set from the Japanese medaka, *Oryzias latipes*, for which we replicated previous findings, including the identification of two sex-reversed individuals, and we uncovered a previously overlooked Y-specific polymorphism in the sequenced population. We then used RADSEX to analyse new data sets that we generated from 15 fish species spanning multiple systematic orders. We identified the sex-determination system in six of these species as well as multiple sex markers, and we identified the sex chromosome in one species for which a reference genome was available. Our results show that RADSEX is well suited to study genetic sex determination in nonmodel species thanks to its speed, resource usage and ease of use. Furthermore, our multispecies analysis provides insights into the genetic mechanisms of sex determination in fishes and highlights some of the limitations of RAD-Seq in detecting small sex loci.

## 2 | MATERIAL AND METHODS

### 2.1 | Overview of the RADSEX analysis workflow

The underlying principle of RADSEX is to group identical RAD-Seq reads from all individuals in a data set into nonpolymorphic markers, and then consider the presence or absence of each marker in each individual. Markers created by RADSEX thus differ from that of other software such as *STACKS* (Catchen et al., 2011) which typically attempt to reconstruct genotypes from polymorphic markers. The RADSEX workflow (Figure 1) includes the command-line software RADSEX implemented in C++ (<https://github.com/SexGenomicsToolkit/radsex>) and the R package SGTR (<https://github.com/SexGenomicsToolkit/sgtr>) to visualize results from RADSEX. The RADSEX software includes several commands, starting with *process*, a data processing command that takes as input a set of demultiplexed reads (i.e., one fasta or fastq file containing short reads of consistent size from either single or double digest RAD-Seq protocols for each individual). The other commands perform analyses using the output of *process*: *distrib* computes the distribution of marker presence between males and females; *signif* extracts all markers significantly associated with sex; *map* aligns markers to a reference genome sequence; *depth* computes the distribution of marker depths in each individual; *freq* computes the distribution of marker presence in all individuals; and *subset* filters markers based on presence in males and females.



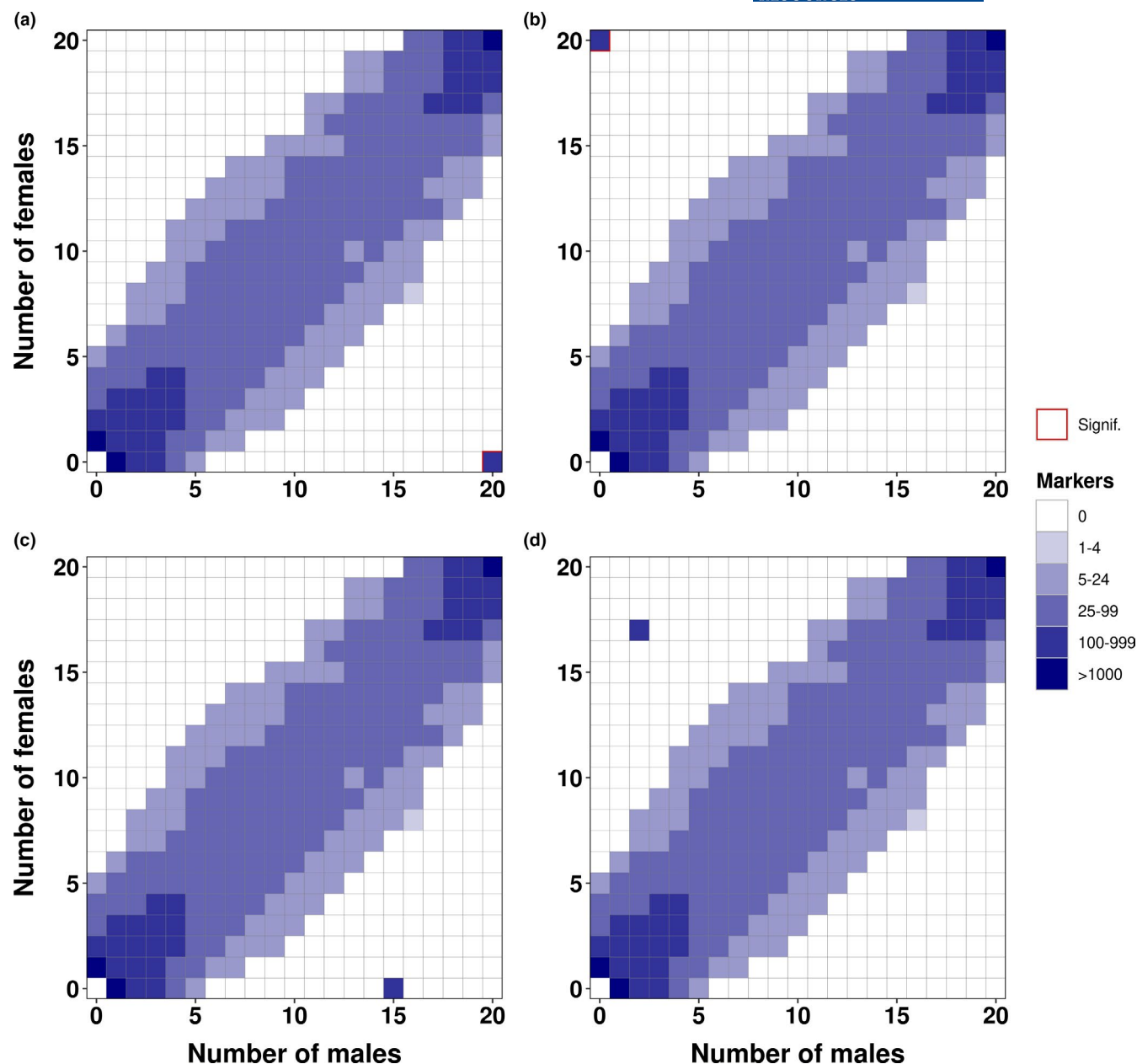
**FIGURE 1** Typical RADSex workflow for sex determination system identification. The *process* command from RADSEX computes a table containing the sequencing depth of each marker in each individual. This table is then used as input for the *distrib* command to compute the distribution of markers between males and females, and this distribution is visualized with the *radsex\_distrib()* function of SGTR. Markers significantly associated with sex are extracted with the *signif* command and their depth in each individual is visualized with the *radsex\_markers\_depth()* function of SGTR. When a genome assembly is available, markers can be aligned with the *map* command and results are visualized with the *radsex\_map\_circos()* function of SGTR.

The first step of the workflow is to compute a table of marker depths. The *process* command creates a table with identifier, sequence and depth in each individual for all unique sequences in the entire data set. This table summarizes all the sequence information present in the data set, and each row constitutes a marker for RADSEX (Figure S1). Sequences are read from each individual read file and are stored in a hash table with the sequence as key and an array of individual sequencing depths as values. Each time a sequence is read from an individual read file, its depth is incremented in the corresponding value column of the hash table. By design, and in contrast to markers typically generated by other tools, markers from RADSEX are not polymorphic. Instead, each allele at a polymorphic locus is treated as a separate marker. The table created by *process* is stored in a plain text tabulated file (Table S1) and is the main input data for all subsequent analyses performed by RADSEX. The *process* command can be parallelized, each thread processing one individual input file at a time.

The next step is to identify markers significantly associated with sex. The *distrib* command uses the table of marker depths generated with *process* to compute the distribution of markers between males and females. Phenotypic sex for each individual is given by a user-supplied tabulated file with the identifier of each individual in the first column and its phenotypic sex in the second column. In all RADSEX commands, a marker is defined as present in a given individual if the depth of this marker in this individual is higher than a user-specified threshold given by the parameter *--min-depth* (or *-d*) (Figure S1). Following this definition, the *distrib* command enumerates marker presence in every possible combination of number of males and number of females (e.g., for a population of one male and two females: 0♂1♀; 0♂2♀; 1♂0♀; 1♂1♀; 1♂2♀), and calculates

the probability of association with sex for each combination using Pearson's chi-squared test of independence. Yates' correction for continuity is applied to account for low individual numbers in some of the tests. The associated *p*-value is obtained from the Cumulative Density Function of the chi-squared distribution with one degree of freedom, as implemented in SAMTOOLS (Li et al., 2009). Bonferroni correction is applied by multiplying the *p*-value by the total number of markers present in at least one individual. Markers for which association with sex is significant (i.e.,  $p < .05$ , or a user-specified threshold) after Bonferroni correction can be obtained with the *signif* command and exported either in the same format as the marker depths table or in fasta format. The results of *distrib* (Table S2) can be visualized using the *radsex\_distrib()* function from SGTR, which generates a tile plot with the number of males on the horizontal axis, the number of females on the vertical axis, and the number of markers represented by a tile's colour for each number of males and females. Tiles for which association with sex is significant are highlighted with a red border. Simplified examples of tile plots for several sex-determining systems are presented in Figure 2.

Next, RADSEX can filter and cluster markers based on depth across individuals to identify consistent absence or presence of markers in some individuals. Markers can be filtered based on presence in males and females using the *radsex subset* command and exported in the same format as the table of marker depths (Figure S1). After extracting a subset of markers, a heatmap showing the depth of each marker in each individual can be generated using the *radsex\_markers\_depth()* function from SGTR. This function optionally allows the user to cluster both individuals and markers based on depth values and display the resulting cladograms along the heatmap. Distances are computed using the base R *dist* function and



**FIGURE 2** Tile plots showing the theoretical result of *distrib* in four scenarios. Each plot shows the distribution of RADSex markers between phenotypic males (horizontal axis) and phenotypic females (vertical axis), with color intensity for a tile indicating the number of markers present in the corresponding number of males and females. Tiles for which association with phenotypic sex is significant (Chi-squared test,  $p < .05$  after Bonferroni correction) are highlighted with a red border. (a) Tile plot for an XX/XY sex determination (SD) system showing markers present in all males and absent from all females in the bottom right corner. (b) Tile plot for a ZZ/ZW SD system showing markers present in all females and absent from all males in the top left corner. (c) Tile plot for an XX/XY SD system with five male outliers showing markers present in all but five males and absent from all females at position (15, 0). (d) Tile plot for a ZZ/ZW SD system with two male and three female outliers showing markers present in all but three females and absent from all but two males at position (2, 17).

clustering is performed with the base R *hclust* function; both distance calculation (euclidean, maximum, binary, etc.) and clustering methods (complete, average, centroid, etc.) can be specified by the user. An application of the *subset* command is the detection of outliers in the sequenced population (e.g., sex-reversed individuals or individuals with misassigned sex). In such cases, some markers may have biased presence in individuals from one sex without being

statistically associated with phenotypic sex. When clustering individuals based on depth for these sex-biased markers, real outliers will be grouped with individuals from the other sex (see Figure 4b for an example from a real data set). Markers obtained with the *signif* command can also be used as input for the *radsex\_markers\_depth()* function from *SGTR* if they were exported as a table of marker depths.



When a genome assembly is available, markers can be aligned to it using the *radsex map* command to locate genomic regions differentiated between sexes. Markers obtained from either the *process*, *signif* or *subset* commands are aligned to the genome using the *BWA mem* library (Li, 2013), and only markers uniquely aligned with a mapping quality higher than a user-specified threshold are retained. Two metrics are computed for each retained marker: (i) the probability of association with sex  $p_{\text{association with sex}}$ , estimated from a chi-squared test with Bonferroni correction as described for the *distrib* command, and (ii) the sex-bias ( $S$ ), defined as  $S = \sigma/\sigma_{\text{total}} - \varphi/\varphi_{\text{total}}$ , where  $\sigma$  and  $\varphi$  are the number of males and females in which the marker is present, and  $\sigma_{\text{total}}$  and  $\varphi_{\text{total}}$  are the total number of males and females in the population, inferred from the user-supplied sex information file. Sex-bias thus ranges from  $-1$  for a marker present in all females and absent from all males to  $+1$  for a marker present in all males and absent from all females, and is zero for a marker present in the same proportion of males and females. Using the output of *radsex map* (Table S3), the *radsex\_map\_circos()* function of *SGTR* generates a circular plot in which sex-bias, on the top track, and  $-\log_{10}(-p_{\text{association with sex}})$ , on the bottom track, are plotted against genomic position using R's *circlize* package (Gu et al., 2014). Each sector of the plot represents a linkage group, with the last sector containing unplaced scaffolds, which are important to include due to frequent assembly problems with sex chromosomes. Each metric can also be displayed in a Manhattan plot using the *radsex\_map\_manhattan()* function; in addition, a linear plot showing both metrics for a given genomic region can be generated using the *radsex\_map\_region()* function from *SGTR*.

Finally, *RADSEX* provides two commands to assess the distribution of marker depths in the entire data set. First, the *freq* command computes the number of individuals in which each marker is present for a given minimum depth value and generates a count table of markers present in each number of individuals. Second, the *depth* command computes the minimum, maximum, median and average marker depth for each individual using markers present in more than 75% of individuals (or a user-specified threshold). Results from *radsex freq* and *radsex depth* can be visualized with the *radsex\_freq()* and *radsex\_depth()* functions from *SGTR*.

## 2.2 | Sample collection

General information on the different species, species collectors and samples is given in Table S5. Detailed information on sample collection for each data set is provided in Methods S1.

## 2.3 | Genomic DNA extraction

Genomic DNA (gDNA) of *Lepisosteus oculatus*, *Plecoglossus altivelis*, *Gadus morhua*, *Notothenia rossii*, *Sander vitreus*, *Tinca tinca*, *Carassius auratus* and all *Danio* species was extracted from fin clips stored in ethanol using NucleoSpin Kits for Tissue (Macherey-Nagel) following

the producer's protocol. gDNA of *Gymnotus carapo*, *Gymnocorymbus ternetzi*, *Hippocampus abdominalis* and *Poecilia sphenops* was extracted from individual fin clips stored in ethanol or from pooled organs of individual fish for *Poecilia sphenops* with a phenol/chloroform protocol. Fin clips or tissues were lysed in 1 ml extraction buffer (0.1 M EDTA pH 8, 0.2% SDS, 0.2 M NaCl) containing 200 µg/ml proteinase K for 3 hr at 80°C, and 500 µl phenol was added to each sample. After mixing and incubating at room temperature for 10 min, 500 µl chloroform/isoamyl alcohol (24:1) was added. Samples were incubated with occasional inversion at room temperature for 10 min and then centrifuged for 10 min at 5,000 g (5°C). The upper layer was transferred to a new tube and one volume of chloroform/isoamyl alcohol (24:1) was added. The samples were kept at room temperature for 10 min with occasional inverting, and subsequently centrifuged for 10 min at 5,000 g (5°C). The supernatant was transferred to a glass vial on ice and 2.5 times the sample volume of cold 100% ethanol was carefully added. The DNA was spooled with a glass rod and dissolved in TE buffer pH 8.

All gDNA concentrations were quantified with a NanoDrop ND2000 spectrophotometer (Thermo Scientific) and/or Qubit (ThermoFisher) before processing gDNA for construction of RAD libraries.

## 2.4 | RAD-Seq and demultiplexing of RAD-Seq reads

RAD libraries were constructed for each species from individual fish gDNA using a single restriction enzyme, *SbfI*, following the standard protocol as previously described (Amores et al., 2011). Libraries were sequenced as single end 100-bp reads on one lane of an Illumina HiSeq 2500. Quality control and demultiplexing of the reads was performed with the *process\_radtags.pl* wrapper script from *STACKS* version 1.44 (Catchen et al., 2011) using the following options: discard reads with low-quality scores (-q), remove any reads with an uncalled base (-c), and rescue barcodes and RAD-seq reads (-r). Demultiplexing results for each species are summarized in Table S7.

## 2.5 | Performance measurements

Runtime and peak memory usage of *RADSEX* (version 1.1.2) was measured with the "benchmark" directive of *SNAKEMAKE* (Mölder et al., 2021) on the Genotoul computational platform using four threads. Performance was measured for the *process*, *distrib* and *signif* commands run sequentially with a minimum depth of 1, because this value would require the longest runtime and memory.

## 2.6 | Software used in the analyses

All analyses were performed using version 1.1.2 of *RADSEX* and figures were generated using version 1.1.2 of the *SGTR* R package in

R version 3.5.2. Both RADSEX and SGTR are released under a GPLv3 licence; releases and source code are available at <https://github.com/SexGenomicsToolkit/radsex> for RADSEX and <https://github.com/SexGenomicsToolkit/sgtr> for SGTR. Both RADSEX and SGTR are also available on the conda channel Bioconda. A complete documentation for RADSEX is available at <https://sexgenomicstoolkit.github.io/html/radsex/introduction.html>.

### 3 | RESULTS

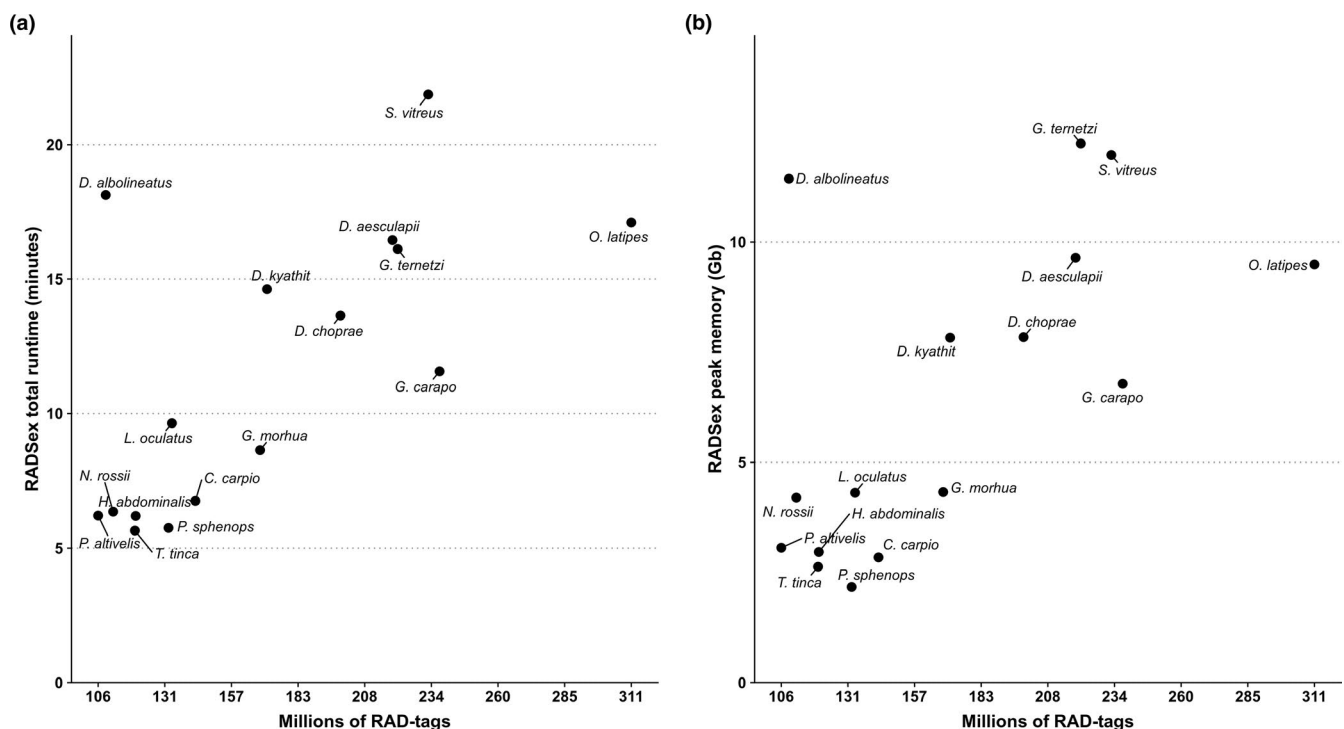
#### 3.1 | Performance of RADSEX

The runtime and peak memory usage of *radsex process*, *distrib* and *signif* were measured on data sets from 15 fish species that we generated for this study. Apart from one outlier data set, *Danio albolineatus*, RADSEX's runtime was linearly and positively correlated to the number of RAD-Seq reads, ranging from 276 s for *Tinca tinca*, whose data set contained ~120 million RAD-Seq reads, to 1,240 s (20 min and 40 s) for *Sander vitreus*, whose data set contained ~232 million RAD-Seq reads (Figure 3a). Runtime and peak memory usage for all data sets are provided in Table S4. RADSEX's peak memory usage was also linearly and positively correlated to the number of RAD-Seq reads, ranging from 2.2 Gb for *Poecilia sphenops* to 12.24 Gb for *Gymnocorymbus ternetzi* (Figure 3b). The high runtime and memory usage of RADSEX for the *Danio albolineatus* data set is related to an extremely high number of markers found in only one individual, which

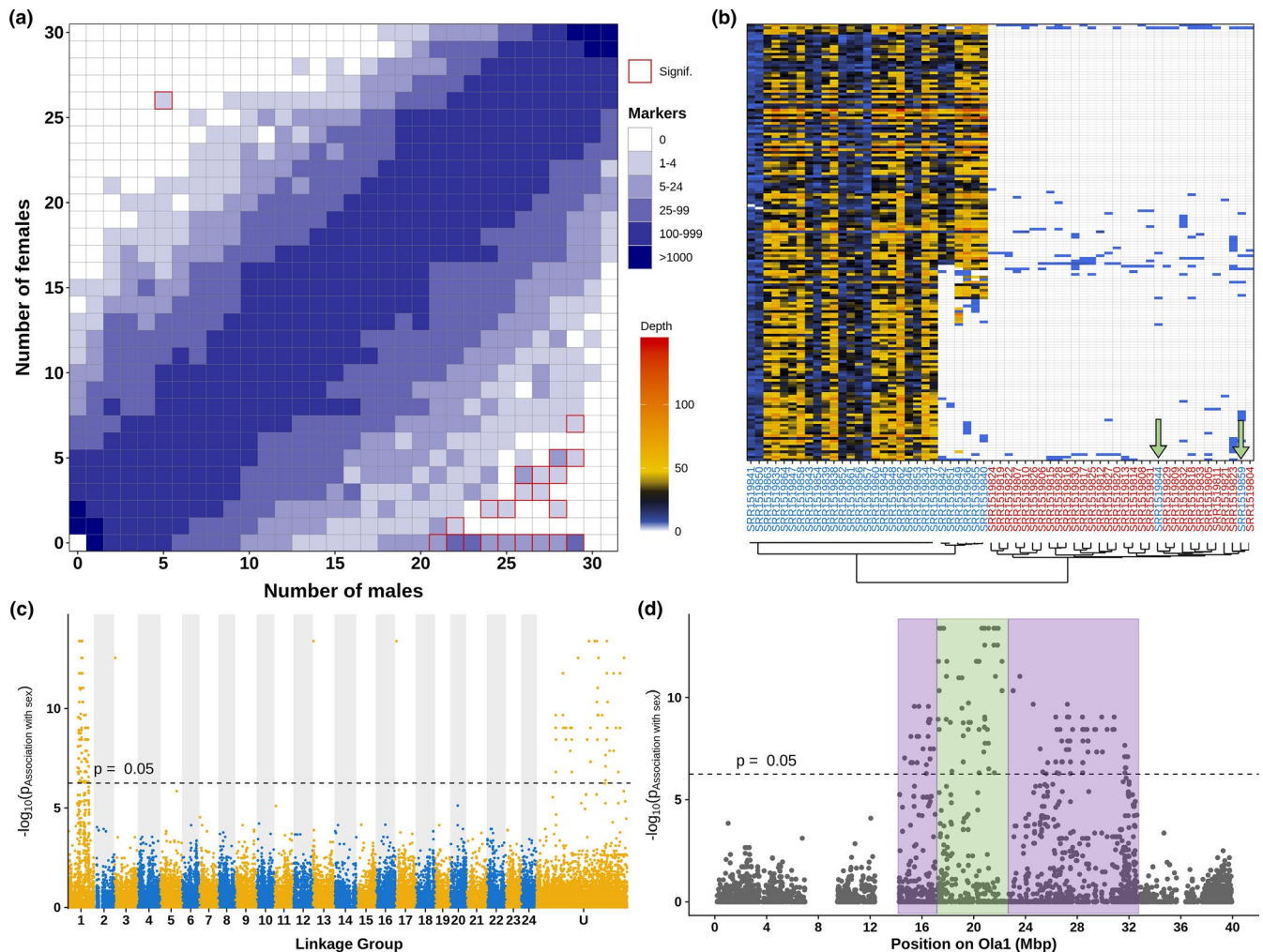
could indicate low sequencing quality of this fish. This hypothesis is supported by the fact that two thirds of the reads from this individual were discarded during multiplexing for this data set. It is worth noting that the peak memory usage and most of the runtime were in the *process* command, which needs to be run only once for each data set because filtering parameters (i.e., minimum depth for marker presence) can be applied in the later steps of the RADSEX pipeline.

#### 3.2 | RADSEX validation on a published data set

To assess whether RADSEX can recover previous results on sex determination, we analysed a data set that has been used to characterize a major sex-determination locus (Wilson et al., 2014). This data set consists of RAD-Seq reads from 31 males and 30 females from the *carpio* strain of *Oryzias latipes*, the Japanese medaka, which has a known XX/XY sex determination system with chromosome 1 as the sex chromosome (Matsuda et al., 2002; Nanda et al., 2002). We analysed this *O. latipes* data set with RADSEX using a minimum depth of 10 (-d 10) to define the presence of a marker in an individual. This value of minimum depth was chosen based on the median sequencing depth in each individual computed with *radsex depth*: the average median sequencing depth was 34, and thus a stringent minimum depth of 10 can be used to discard markers from potential sequencing errors but still retain real markers. In total, we found 121,492 markers present in at least one individual with a minimum depth of 10, among which 194 markers were significantly associated with male



**FIGURE 3** Performance of the RADSEX workflow (*process*, *signif*, and *distrib*). (a) Total runtime of RADSEX (in minutes) plotted against the number of RAD-seq reads for each dataset. (b) Peak memory of RADSEX in gigabytes plotted against the number of RAD-seq reads for each dataset.



**FIGURE 4** Validation of the RADSex approach on the *Oryzias latipes* dataset. (a) Tile plot showing the distribution of RADSex markers between males (horizontal axis) and females (vertical axis) using the original sex assignment and a minimum depth of 10 ( $d = 10$ ). Color intensity indicates the number of markers present in the corresponding number of males and females, and a red border indicates significant association with sex (Chi-squared test,  $p < .05$  after Bonferroni correction). (b) Heatmap showing individual sequencing depth for all markers present in at least 20 males and absent ( $d = 10$ ) from all females. Each row corresponds to a marker, and each column corresponds to an individual. Males (blue labels) and females (red labels) clustered in two separate groups, except for two male outliers, identified by the green arrows, which clustered with the females. In addition, six males forming a sub-cluster within the male cluster had low or null depth for half of the markers. (c) Manhattan plot showing  $-\log_{10}(P)$  of a chi-squared test on number of males and number of females, after Bonferroni correction, for all RADSex markers aligned to the genome of *O. latipes*, with odd linkage groups (LGs) in yellow with light background and even LGs in blue with shaded background. Unplaced scaffolds were concatenated in a super scaffold (U) by order of decreasing size. The dashed line indicates a  $p$ -value of .05 after Bonferroni correction. Most markers significantly associated with sex were aligned to LG01 (Ola1) or to unplaced scaffolds. (d) The  $-\log_{10}(P)$  of a chi-squared test on the number of males and number of females is plotted against genomic position on LG01 of *O. latipes* (Ola1) for all RADSex markers aligned to this LG. The dashed line indicates a  $p$ -value of .05 after Bonferroni correction. Markers significantly associated with sex were aligned to a continuous region between 14.7 and 31.9 Mb. Markers present in all males apart from the two outliers and absent in all females were aligned in a continuous region spanning from 17 to 23 Mb (green box), while male-biased markers absent from the sub-cluster of six males were aligned in two regions, from 14 to 17 Mb and from 24 to 31.9 Mb (purple boxes).

phenotype (*radsex signif*,  $p < .05$ , chi-squared test with Bonferroni correction, highlighted tiles in Figure 4a). Among these 194 markers, 165 were present in at least 20 of the 31 males. In addition, a single marker present in 26 of the 30 females and five males was significantly associated with female phenotype (Figure 4a). However, we did not find any marker simultaneously present in more than 29 of the 31 phenotypic males and absent from all phenotypic females. To

understand why no marker was found exclusively in all males given that medaka has an XY/XX sex determination system, we extracted individual depth for all 165 markers present in at least 20 males and none of the females using *radsex subset*. We clustered both markers and individuals separately based on these depth values and generated a heatmap of the results using the *radsex\_markers\_depth()* function of *SGTR* (Figure 4b). As expected, phenotypic males and females



clustered in two separate groups, except for two phenotypic males (green arrows in Figure 4b) that clustered with the phenotypic females, showing zero or very low depth for the extracted markers. These two outlier phenotypic males had already been found to be sex-reversed genetic XX females by retrospective genotyping in Wilson et al. (2014). After correcting the designated phenotypic sex of these two individuals, *radsex signif* yielded 232 markers significantly associated with male phenotype and no marker significantly associated with female phenotype ( $p < .05$ , chi-squared test with Bonferroni correction), as expected for an XY/XX system. In addition, six males forming a subcluster within the male cluster from *sgtr radsex\_markers\_depth()* had low or null depth for almost half of the extracted markers, suggesting that several Y chromosome haplotypes could be present in the sequenced population.

To identify the sex chromosome and to delimit the sex-differentiated region, we aligned the RADSex markers to the assembly used by Wilson et al. (MEDAKA1, available at [http://77.235.253.122:8012/Oryzias\\_latipes/Info/Index](http://77.235.253.122:8012/Oryzias_latipes/Info/Index)) using *radsex map*. Among the 232 markers significantly associated with male phenotype after reassignment of the sex-reversed fish, 131 (56%) aligned to LG01 of *O. latipes* (Ola1), three (1%) aligned to other chromosomes, 54 (23%) aligned to unplaced scaffolds, and 44 (19%) were not uniquely aligned with mapping quality higher than 20 (Figure 4c). On Ola1, markers significantly associated with male phenotype aligned to a continuous region spanning from 14.7 to 31.9 Mb (Figure 4d). Within this region, markers found in all males and no females aligned between 17 and 23 Mb (green box in Figure 4d) and markers present in all but the six males identified in the clustering step and absent from all females aligned between 14.7 and 31.9 Mb (purple boxes in Figure 4d). This result supports the hypothesis that multiple Y chromosome haplotypes are present in this medaka population, with one region located between 17 and 23 Mb on Ola1 showing strong X/Y differentiation in all Y haplotypes, and a wider region located between 14.7 and 31.9 Mb on Ola1 showing strong X/Y differentiation in only one Y haplotype.

To evaluate whether markers created by RADSex in the medaka data set represent biological alleles, we computed the number of markers at each alignment starting position on the genome sequence from the results of *radsex map*. The vast majority (94%) of positions had one (65%) or two (29%) markers aligned, and less than 1% of positions had more than three markers aligned, as expected for alleles in biological loci (Figure S2A). The original study by Wilson et al. (2014) identified 248 single nucleotide polymorphisms (SNPs) associated with male phenotype aligned to a region ranging from 14.3 to 32.5 Mb on Ola1. To compare sex-biased RADSex markers with the sex-biased SNPs from Wilson et al. (2014), the genomic position and  $p$ -value of association with sex were retrieved for all 248 sex-biased SNPs from table S1 from Wilson et al. (2014). Among these SNPs, 153 (62%) were located within a RADSex marker significantly associated with sex, 89 (36%) were located within a RADSex marker not significantly associated with sex, and six (2%) were located at a position where no RADSex marker was aligned (Figure S2B). The discrepancy for these 89 SNPs can be explained by two

factors. First, although  $p$ -values of association with sex for RADSex markers were correlated with  $p$ -values for SNPs from Wilson et al. (2014) aligned within a marker,  $p$ -values were generally slightly lower for Wilson et al. (2014; Figure S2C). Second, the number of markers used for Bonferroni correction was higher for RADSex (121,492 markers, threshold indicated by the blue dashed line in Figure S2C) than for Wilson et al. (2014; ~700 SNPs after filtering out SNPs present in <75% males or females, threshold indicated by the red horizontal dashed line in Figure S2C). Using the same threshold for Bonferroni correction (red vertical dashed line in Figure S2C), the number of SNPs located within RADSex markers not significantly associated with sex dropped to 22 (9%). In addition, RADSex markers significantly associated with sex were aligned at 67 new positions (38%) that did not contain any sex-biased SNPs from Wilson et al. (Figure S2D), 40 of which (60%) were aligned to Ola1 between 14.6 and 32.5 Mb.

Overall, these results indicate that markers identified as significantly associated by RADSex are consistent with sex-biased SNPs from Wilson et al. (2014), but the threshold for association with sex is more stringent in RADSex. Despite this increased stringency, RADSex identified markers significantly associated with sex for 67 genomic locations that were not identified in the original study. Furthermore, our analysis revealed a previously unidentified Y-specific polymorphism in the sequenced population. This example highlights the effectiveness and versatility of RADSex and its visualization tools to identify, explore and explain patterns in data sets that were previously overlooked.

### 3.3 | Analysis of 15 new ray-finned fish RAD-Seq data sets

RAD-Seq data sets were generated for 15 ray-finned species in which no markers for genetic sex determination were available when we initiated this work. For each species, we used RADSex to search for sex-biased markers with a minimum marker depth ( $-d$ ) of 1, 2, 5 and 10. We report results for a minimum marker depth of 10 (Table 1) because it is the most stringent, except for two species that had the lowest sequencing depths, for which we report results for a minimum marker depth of 2 (Figure S3). We identified an XX/XY sex-determining system in six species (*Cyprinus carpio*, *Gymnotus carapo*, *Plecoglossus altivelis*, *Tinca tinca*, *Gadus morhua* and *Poecilia sphenops*) with a few male and/or female outliers in four of these species (Figure 5). Markers significantly associated with phenotypic sex for each species are provided in Table S6. An XX/XY sex-determining system is in agreement with previous reports for *Cyprinus carpio* (Gomelsky, 2003), *Gymnotus carapo* (da Silva et al., 2014), *Plecoglossus altivelis* (Watanabe et al., 2004), and *Gadus morhua* (Haugen et al., 2012; Whitehead et al., 2012). The XX/XY sex-determining system inferred from RADSex results for *Poecilia sphenops* conflicts with published findings on an ornamental population for that species that showed a female heterogametic system (Nanda et al., 1993). However, alternative population-specific sex

TABLE 1 Summary of RADSex results for the 15 data sets analysed

Species	SD system identified with <i>radsex</i>	Number of sex markers (minimum depth)	SD system previously described
<i>Cyprinus carpio</i>	XX/XY with male outliers	7 ( <i>d</i> = 10)	XX/XY (Gomelsky, 2003)
<i>Danio aesculapii</i>	No SD identified	0 ( <i>d</i> = 10)	Unknown
<i>Danio albolineatus</i>	No SD identified	0 ( <i>d</i> = 10)	Unknown
<i>Danio choprae</i>	No SD identified	0 ( <i>d</i> = 10)	Unknown
<i>Danio kyathit</i>	No SD identified	0 ( <i>d</i> = 10)	Unknown
<i>Gadus morhua</i>	XX/XY with outliers	3 ( <i>d</i> = 10)	XX/XY (Haugen et al., 2012)
<i>Gymnocorymbus ternetzi</i>	No SD identified	0 ( <i>d</i> = 2)	Potentially XX/XY (David & Pandian, 2016)
<i>Gymnotus carapo</i>	XX/XY with female outliers	8 ( <i>d</i> = 10)	XX/XY (da Silva et al., 2014)
<i>Hippocampus abdominalis</i>	No SD identified	0 ( <i>d</i> = 10)	Unknown
<i>Lepisosteus oculatus</i>	No SD identified	0 ( <i>d</i> = 2)	Unknown
<i>Notothenia rossii</i>	No SD identified	0 ( <i>d</i> = 10)	Unknown
<i>Plecoglossus altivelis</i>	XX/XY	47 ( <i>d</i> = 10)	XX/XY (Watanabe et al., 2004)
<i>Poecilia sphe-nops</i>	XX/XY with female outliers	7 ( <i>d</i> = 10)	ZZ/ZW (Nanda et al., 1993) XX/XY (Volf & Scharl, 2001)
<i>Sander vitreus</i>	No SD identified	0 ( <i>d</i> = 2)	Unknown
<i>Tinca tinca</i>	XX/XY	6 ( <i>d</i> = 10)	Unknown

Note: Species where a sex-determination (SD) system was identified with *Radsex* are highlighted with a grey background. The number of markers significantly associated with phenotypic sex is also given for each species, with the value of *d* indicating the minimal number of reads for a marker to be considered present in an individual in the analysis.

determination systems have also been described for *Poecilia sphe-nops* (Volf & Scharl, 2001). Because these studies were performed on ornamental fish and laboratory strains, previously derived conclusions may not apply to wild populations such as the one we analysed. Lastly, to the best of our knowledge, sex determination was not previously characterized in the tench, *Tinca tinca*.

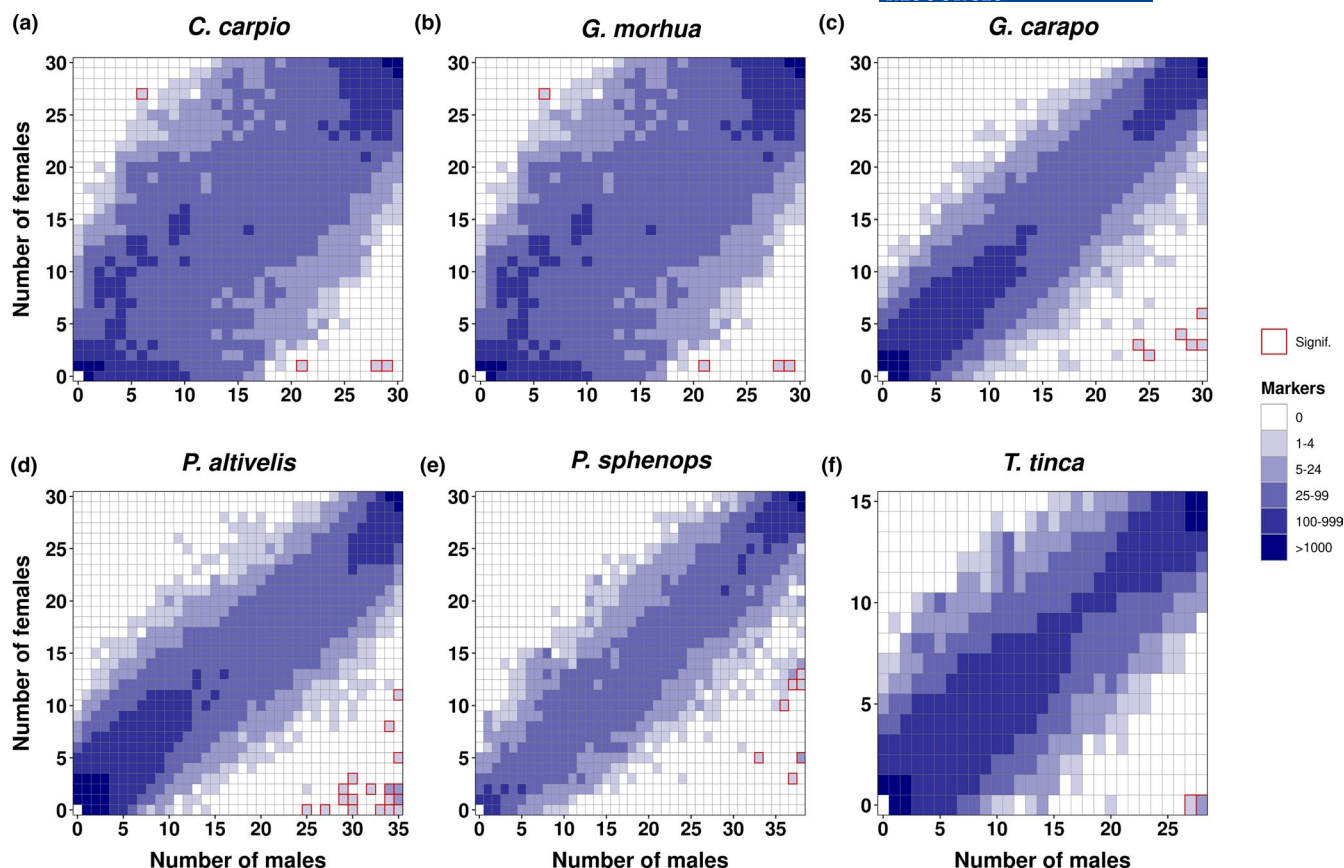
In *Gadus morhua*, three markers were significantly associated with male phenotype, but they were also found in two females and none of the three markers was present in more than 27 of the 34 males (Figure 5b). Furthermore, multiple markers were found pre-dominantly in males but not significantly associated with sex, with eight markers consistently present with depth >5 in 20 males and absent from all females (these markers are highlighted with arrows in Figure 6a). The three markers significantly associated with sex (red-outlined boxes) and most other markers strongly but not signifi-cantly associated with sex aligned to cod linkage group 11 (Gmo11; Figure 6b,c). Markers strongly associated with sex aligned to a re-gion spanning from 0 to 20 Mb on Gmo11, and the three markers significantly associated with sex aligned between 10 and 15 Mb on this chromosome in a region that was previously characterized as containing the sex locus in this species (Kirubakaran et al., 2019; Star et al., 2016). These results confirm the previously identified XX/XY sex determining system and Gmo11 as the sex chromosome in Atlantic cod (Haugen et al., 2012; Whitehead et al., 2012), but our findings suggest an additional sex determination complexity in the studied Atlantic cod aquaculture population that had not been ob-served before in wild populations of Atlantic cod (Kirubakaran et al., 2019; Star et al., 2016). The incomplete sex-linkage of sex-biased

markers may be explained by a Y-chromosome polymorphism and/or the existence of sex-reversed genetic XX females. Although Y-chromosome population differences have rarely been explored, at least in fish species, the results from our re-analysis of a medaka population suggest that such polymorphisms may be more frequent than expected. In addition, female-to-male sex reversal could be a consequence of intensive aquaculture rearing conditions leading to stress-induced sex reversal (Geffroy & Douhard, 2019), which has been observed in other fish species reared in laboratory or aquacul-ture facilities (Pan et al., 2019).

In the nine other species (Table 1; Figure S4), we did not find any marker significantly associated with phenotypic sex, which indicates that these species either (i) have a small undifferentiated sex locus, (ii) lack a genetic sex determination mechanism, or (iii) rely on a com-plex combination of environmental and genetic sex determination factors that prevented the detection of a clear sex-biased signal in the data set.

## 4 | DISCUSSION

We demonstrate that RADSex is an efficient and—thanks to its simple usage and visualization options—user-friendly tool to iden-tify sex-determination systems and sex-biased markers from RAD-Seq data. Unlike other popular RAD-Seq analysis software, such as STACKS (Catchen et al., 2011) and PYRAD (Eaton, 2014), which cata-logue polymorphic loci, RADSex creates monomorphic markers. RADSex therefore counts as a separate marker each allele at a locus

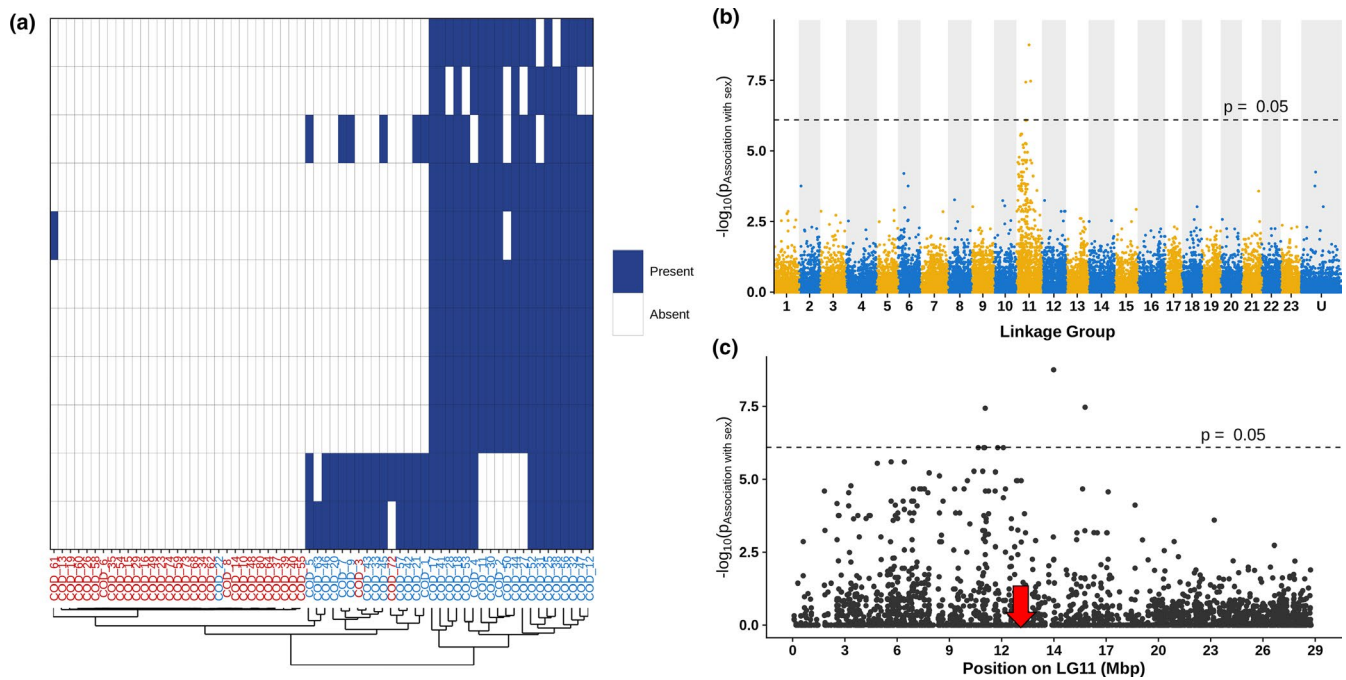


**FIGURE 5** Tile plots showing the distribution of markers between males and females for six species (a–f) with markers significantly associated with sex. The number of markers, represented by tile color intensity, is plotted for every combination of number of males (horizontal axis) and number of females (vertical axis) for each species. Combinations of number of males and number of females for which association with sex is significant ( $p < .05$ , Chi-squared test with Bonferroni correction) are highlighted in red. Distributions were computed with a minimum depth of 10 ( $-d\ 10$ ).

defined as polymorphic by other software. Working with monomorphic markers allows statistical comparison of populations using straightforward presence/absence tests, yielding results that are comparable between data sets. In contrast, the outcome of grouping sequences into polymorphic markers can be sensitive to parameter values (Paris et al., 2017; Rodríguez-Ezpeleta et al., 2016; Shafer et al., 2018). Leveraging this property of monomorphic markers, RADSex offers a simple and reproducible way to identify *de novo* genetic differences between females and males and to locate these differences on a genome when a reference assembly is available. In addition to this simplicity, RADSex requires few resources and provides comprehensive visualization tools to assist the interpretation of its output.

In particular, the tile plot visualization provides a comprehensive overview of the RAD-Seq results across all individuals and allows the user to quickly identify straightforward cases of genetic sex determination systems. For instance, XX/XY or ZZ/ZW monofactorial systems can be easily detected in the absence of any major environmental effects on sex, as demonstrated by our analyses on the ayu (*Plecoglossus altivelis*) and the tench (*T. tinca*) data sets: in both species, the tile plot revealed complete sex-linkage of several markers significantly associated with male phenotype, indicating a male heterogametic sex

determination system. Besides these straightforward cases, the tile plot visualization is useful to identify the sex-determination system in more complex data sets, for instance in the presence of outliers and sequencing biases. These complex data sets can then be further explored with the marker depths heatmap visualization, for instance to confirm whether sex-biased markers are absent or present in specific individuals. Using this heatmap, we were able to quickly identify sex-reversed individuals responsible for the noncomplete sex-linkage of sex-biased markers in a publicly available medaka data set (Wilson et al., 2014), as well as outliers in our own common carp (*Cyprinus carpio*), Atlantic cod (*Gadus morhua*), banded knifefish (*Gymnotus carapo*) and common molly (*Poecilia sphenops*) data sets. These outliers could arise from human errors in assessing phenotypic sex; in particular, in the common molly, males are externally indistinguishable from females until puberty, which occurs late in some males, and immature testes resemble nonreproductively active ovaries. Because in this species sex was determined from secondary sex characters and macroscopy of the gonads, but not histology, it is possible that the relatively high number of females in which we found male-biased markers were immature males that were mis-sexed as females. The presence of outliers, however, could also hint at a more complex sex determination system, for instance polygenic sex determination (Moore & Roberts,



**FIGURE 6** Results of RADSex analysis on *Gadus morhua*. (a) Tile plot showing the presence of markers in each individual for all markers present with depth higher than five in at least 20 males and in fewer than two females. Each row corresponds to a marker and each column corresponds to an individual, with females labeled in red and males labeled in blue. (b) Manhattan plot showing  $-\log_{10}(P)$  of a chi-squared test on the number of males and number of females for all RADSex markers aligned to the genome of *G. morhua*, with odd linkage groups (LGs) in yellow with light background and even LGs in blue with shaded background. Unplaced scaffolds were concatenated in a super scaffold (U) by decreasing order of size. The dashed line indicates a  $p$ -value of .05 after Bonferroni correction. (c)  $-\log_{10}(P)$  of a chi-squared test on the number of males and the number of females for all RADSex markers aligned to LG11 of *G. morhua*. The dashed line indicates a  $p$ -value of .05 after Bonferroni correction. Markers significantly associated with sex ( $p < 0.05$ ) and other markers strongly associated with sex were aligned to the first 20 Mb of the chromosome. The position of the candidate master sex determining gene is indicating by a red arrow.

2013), autosomal modifiers that are polymorphic in the population (Wu, 1983), or simple monofactorial genetic sex determination systems with some sex-reversed individuals triggered by environmental factors (Baroiller & D'Cotta, 2016; Dupoué et al., 2019; Wessels et al., 2017). With a high frequency of outliers, sex-biased markers may no longer be significantly associated with phenotypic sex, yet these markers can still be biologically relevant.

Because RADSex creates nonpolymorphic markers, the number and distribution of markers between individuals is only affected by a single parameter controlling the minimum depth to consider a marker present or absent in an individual. An overly high minimum depth value can lead to false negatives (i.e., markers considered absent from an individual because of insufficient sequencing depth) and a very low minimum depth value can create false positives (i.e., markers considered present in an individual because of sequencing errors). Moreover, because the probability of association with phenotypic sex for a marker is adjusted using Bonferroni correction, the stringency of the test for significance increases with the total number of markers identified in the data set, which is affected by the minimum depth parameter. In the 15 data sets that we generated, median sequencing coverage in an individual was between 30× and 100×, except for three species for which it was lower than 30× (Figure S3). For all species except the three species with low sequencing coverage, we found that

minimum depth values between 5 and 10 provided a good balance between stringency and minimizing false negatives. Some data sets, however, may require using low minimum depth values, for instance the three data sets for which median individual sequencing depth was low, or species with large genome sizes such as some sharks (Hara et al., 2018) or some amphibians (Sclavi & Herrick, 2019) for which sequencing all individuals of a population with sufficient depth can be costly. In such cases, the memory efficiency of RADSex enables the analysis of large data sets with minimum depth as low as 1, which can require a lot of memory with other software. In addition to sequencing errors, common sources of biases in RAD-Seq analyses include highly repeated sequences (Catchen et al., 2011), PCR duplicates and allele drop-out (Andrews et al., 2016). RADSex considers the presence and absence of markers and highly repeated sequences are treated like any other, meaning that RADSex is able to identify sex-biased repeated sequences. PCR duplicates can produce false negatives if one allele at a heterozygous locus is under-amplified in one individual, resulting in a sequencing depth lower than the minimum depth threshold for this allele in this individual; this effect can be mitigated by using a low minimum depth threshold. Allele drop-out would also generate false positives and cannot be avoided with RADSex; however, the visualization tools included in RADSex help identify scenarios where the sex signal is tempered by false negatives.



The numbers of markers significantly associated with phenotypic sex can provide an estimate of the size of the sex-determining region, with a low or high number of sex-specific markers reflecting a small or a large nonrecombining region, respectively. It is worth noting, however, that other factors can affect this number, for instance the level of differentiation between males and females within the nonrecombining region, or bias in GC content within this region (Sigeman et al., 2018; Smeds et al., 2015) potentially leading to an over- or under-representation of RAD-Seq markers depending on the restriction enzyme used for library construction. An extreme case is complete absence of markers associated with phenotypic sex for any value of minimum depth, which we observed in nine of our data sets. These null results may be linked to two key features of sex determination in teleosts. First, early cytological studies have shown that the majority of teleost fish carry homomorphic sex chromosome pairs (Devlin & Nagahama, 2002), and more recent genomic analyses have revealed that small sex loci are common in teleosts, with the extreme case of a single SNP being the sole difference between the sexes in Fugu (Kamiya et al., 2012). In such cases with very little differentiation between the sex chromosomes, the fragmented resolution of RAD-Seq means that the likelihood of obtaining RAD sequences from the sex locus can be very low. Although this problem might be overcome by increasing the number of RAD-Seq reads, for example by using a restriction enzyme that cuts more frequently, genome-wide approaches such as pooled sequencing of males and females (Feron et al., 2020; Gammerdinger et al., 2016; Wen et al., 2019) or individual whole genome sequencing (Star et al., 2016) may be necessary to identify the sex-determining region. Second, a complete absence of sex-biased markers can reflect a complex sex determination system, for instance (i) a polygenic system, (ii) strong polymorphism on the sex chromosomes leading to multiple W or Y haplotypes each shared by only a fraction of individuals from the heterogametic sex, or (iii) a system involving a strong environmental effect that would weaken the association between genetic markers and phenotypic sex. In fish, many environmental components can interfere with genetic sex determination mechanisms (Heule et al., 2014; Kikuchi and Hamaguchi, 2013). The visualization tools included in RADSex facilitate the identification of sex-reversed individuals and sex-chromosome polymorphism as demonstrated in our re-analysis of the Medaka data set. The association between phenotypic sex and genomic regions, however, may still be difficult to detect when sex reversal is frequent, especially when associated with phenotyping uncertainty. For such systems, pooled sequencing may not be well suited because male and female pools will each contain heterogeneous genotypes, and individual sequencing remains expensive despite decreasing sequencing costs. A hybrid strategy using both RAD-Seq and pooled sequencing may prove to be the most efficient for these complex sex-determination systems, as demonstrated in the goldfish (Wen et al., 2019), a species with a strong thermal effect on sex determination (Goto-Kazeto et al., 2006). Finally, in some species, sex is determined entirely by environmental factors and therefore does not involve any genetic mechanism (Martínez-Juárez & Moreno-Mendoza, 2019).

One of the strengths of RAD-Seq is the ability to compare populations without a reference genome, which may not always be available for nonmodel species. When a genome is available, however, locating sex-biased markers provides information on the size and level of differentiation of the sex-determining region and can help to identify candidate genes involved in sex determination. The Manhattan plot, circos plot and contig plot visualizations included in RADSex quickly display the location of the sex-biased markers over the whole genome or over selected regions to identify the sex-determining region. Furthermore, the sex-bias metric and probability of association with sex computed with RADSex can be more effective than  $F_{ST}$  computed on polymorphic loci to detect variability in the sex-determining region. This was illustrated by our re-analysis of a publicly available Japanese medaka data set (Wilson et al., 2014), in which we detected multiple Y chromosome haplotypes that were not found in the original study. Intraspecific sex chromosome polymorphisms have been reported in different taxa, including mammals in which Y chromosome polymorphisms are widely used for human evolution studies (Jobling & Tyler-Smith, 2003) and some fruit flies in which the Y chromosome polymorphism is thought to control male fitness (Chippindale & Rice, 2001). In fish, sex chromosome polymorphisms have been reported in different populations of the guppy, *Poecilia reticulata*, and in other closely related species of guppies (Nanda et al., 2014), and different sex chromosomes have even been found within the same species, for instance in tilapias (Cnaani et al., 2008). To the best of our knowledge, however, there are no reports of intraspecific population differences of a single sex chromosome in fish from genomic data, and our results provide the first example of such a Y chromosome polymorphism in this clade.

As part of this study, we used RADSex to investigate the sex determination systems of 15 species sampled broadly across the ray-finned fish phylogeny. We identified XX/XY sex-determining systems in six of these 15 species but did not find evidence of a ZZ/ZW sex-determining system in any of the other nine species. Although our sampling is too limited to make general inferences about teleost sex determination, the predominance of XX/XY over ZZ/ZW sex-determining systems in our data sets is in agreement with previous findings that, in teleosts, transitions from ZZ/ZW to XX/XY sex-determining systems are more frequent than the reverse (Pennell et al., 2018), which would result in XY sex-determining systems being prevalent in this clade.

All of our data sets were generated using single-digest RAD-Seq, but the workflow accepts double-digest RAD-Seq data as well. Furthermore, although all species included in this study are fish, we did not use any fish-specific assumptions in RADSex's implementation, and therefore the computational workflow can be applied to RAD-Seq data from any species. Finally, we specifically developed RADSex to study sex determination, and this design decision is reflected in the wording of this paper. However, the computational workflow was designed to be generic, and therefore both the RADSEX software and the *sgTR* R package could be used for other major quantitative trait loci with contrasting binary phenotypes, whether naturally occurring or induced by mutagenesis.



## ACKNOWLEDGEMENTS

We thank all colleagues who helped us improve the beta version of RADSex by testing it and providing feedback. We thank Daniel Jeffries for his feedback on the software and enriching discussions on our results. This project was supported by funds from the "Agence Nationale de la Recherche," the "Deutsche Forschungsgemeinschaft" (ANR/DFG, PhyloSex project, 2014-2016, SCHA 408/10-1, M.S.), the National Institutes of Health (USA) (grants R01GM085318 and R35GM139635, J.H.P.), and the National Science Foundation (USA) Office of Polar Programs (grants PLR-1247510 and PLR-1444167 to H.W.D. and grant OPP-1543383 to J.H.P., T.D. and H.W.D.). The MGX core sequencing facility was supported by France Genomique National infrastructure, funded as part of the "Investissement d'avenir" programme managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). R.F. was partially supported by Swiss National Science Foundation grant PP00P3\_170664 to R.M.W. Common carp were provided by the PEARL INRA 1036 U3E experimental facilities that are supported by the ANAEE-France National Infrastructure. We thank Allyse Ferrara, Quenton Fontenot, and the Bayosphere Laboratory at Nicholls State University (Thibodeaux, LA) for support with spotted gar sample collection and preparation, and Dan Rosauer and Jonathan Meerbeek from the Iowa Department of Natural Resources for walleye sampling. The authors thank the captain and crew of the ARSV *Laurence M. Gould* and the personnel of the US Antarctic Support Contractor for assistance in Chile, at sea, at Palmer Station, Antarctica, and logistically in Denver, CO. This is publication # 414 from the Northeastern University Marine Science Center. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing computing and storage resources.

## AUTHOR CONTRIBUTIONS

R.F. designed and implemented RADSex and sgtr, with feedback from Q.P., M.W. and Y.G., J.H.P., and M.Sch. designed the PhyloSex project and advised on results interpretation. Q.P., M.W., B.I., J.A., A.H., K.K., A.S.R., K.D., S.K., C.K., J.H.P., M.Sch. and Y.G. participated in the analysis of the results. H.P. and L.J. prepared libraries and performed the sequencing. R.F., Q.P., J.H.P., M.Sch. and Y.G. drafted the manuscript. R.F., Q.P., R.M.W., J.H.P., M.Sch. and Y.G. revised the manuscript. Y.G., M.Sch., J.H.P., E.J., S.K., M.W., M.A., C.W., B.M., A.A., T.D., F.W.G., M.K., H.W.D., M.O., R.N., T.S., M.N., M.St., A.W., Ø.K. and I.B. collected, sexed and/or extracted and prepared gDNA samples. All authors approved the final manuscript.

## DATA AVAILABILITY STATEMENT

All RAD-Seq experiments have been submitted to GenBank under the Bioproject PRJNA548074. A computational workflow implementing all the analyses performed in this study, including generating figures, is available at <https://github.com/RomainFeron/paper-sexdetermination-radsex>. RADSex is released under GPLv3 licence;

source code, installation instructions and documentation are available at <https://github.com/SexGenomicsToolkit/radsex>. sgtr is released under GPLv3 licence and is available at <https://github.com/SexGenomicsToolkit/sgtr>.

## ORCID

Romain Feron  <https://orcid.org/0000-0001-5893-6184>  
 Qiaowei Pan  <https://orcid.org/0000-0002-8501-0405>  
 Jennifer Anderson  <https://orcid.org/0000-0002-0713-6897>  
 Amaury Herpin  <https://orcid.org/0000-0002-0630-4027>  
 Christophe Klopp  <https://orcid.org/0000-0001-7126-5477>  
 Kang Du  <https://orcid.org/0000-0002-1497-8945>  
 Thomas Desvignes  <https://orcid.org/0000-0001-5126-8785>  
 Frederick W. Goetz  <https://orcid.org/0000-0002-6300-1655>  
 Harry William Detrich  <https://orcid.org/0000-0002-0783-4505>  
 Marcos A. Oliveira  <https://orcid.org/0000-0002-7088-3211>  
 Rafael H. Nóbrega  <https://orcid.org/0000-0001-9796-5076>  
 Matthias Stöck  <https://orcid.org/0000-0003-4888-8371>  
 Robert M. Waterhouse  <https://orcid.org/0000-0003-4199-9052>  
 John H. Postlethwait  <https://orcid.org/0000-0002-5476-2137>  
 Manfred Scharl  <https://orcid.org/0000-0001-9882-5948>  
 Yann Guiguen  <https://orcid.org/0000-0001-5464-6219>

## REFERENCES

- Al-Ameri, A. A., Al-Qurainy, F., Gaafar, A.-R.-Z., Khan, S., & Nadeem, M. (2016). Molecular identification of sex in phoenix dactylifera using inter simple sequence repeat markers. *BioMed Research International*. <https://doi.org/10.1155/2016/4530846>
- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., & Postlethwait, J. H. (2011). Genome evolution and meiotic maps by massively parallel DNA sequencing: Spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 188(4), 799–808. <https://doi.org/10.1534/genetics.111.127324>
- Ancona, S., Dénes, F. V., Krüger, O., Székely, T., & Beissinger, S. R. (2017). Estimating adult sex ratios in nature. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372(1729). <https://doi.org/10.1098/rstb.2016.0313>
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Bachtrog, D. (2008). The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics*, 179(3), 1513–1525. <https://doi.org/10.1534/genetics.107.084012>
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., & Tree of Sex Consortium. (2014). Sex determination: Why so many ways of doing it? *PLOS Biology*, 12(7), e1001899. <https://doi.org/10.1371/journal.pbio.1001899>
- Baroiller, J.-F., & D'Cotta, H. (2016). The reversible sex of gonochoristic fish: Insights and consequences. *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, 10(5–6), 242–266. <https://doi.org/10.1159/000452362>
- Benestan, L., Moore, J.-S., Sutherland, B. J. G., Le Luyer, J., Maaroufi, H., Rougeux, C., Normandeau, E., Rycroft, N., Atema, J., Harris, L. N., Tallman, R. F., Greenwood, S. J., Clark, F. K., & Bernatchez, L. (2017). Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex

- determination systems. *Molecular Ecology*, 26(24), 6767–6783. <https://doi.org/10.1111/mec.14217>
- Bewick, A. J., Chain, F. J. J., Zimmerman, L. B., Sesay, A., Gilchrist, M. J., Owens, N. D. L., Seifertova, E., Krylov, V., Macha, J., Tlapakova, T., Kubickova, S., Cernohorska, H., Zarsky, V., & Evans, B. J. (2013). A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*. *Genome Biology and Evolution*, 5(6), 1087–1098. <https://doi.org/10.1093/gbe/evt073>
- Blackmon, H., Ross, L., & Bachtrog, D. (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. *The Journal of Heredity*, 108(1), 78–93. <https://doi.org/10.1093/jhered/esw047>
- Carmichael, S. N., Bekaert, M., Taggart, J. B., Christie, H. R. L., Bassett, D. I., Bron, J. E., Skuce, P. J., Gharbi, K., Skern-Mauritzen, R., & Sturm, A. (2013). Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*) using RAD sequencing. *PLoS One*, 8(10), e77832. <https://doi.org/10.1371/journal.pone.0077832>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping Loci de novo from short-read sequences. *G3*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1403), 1563–1572. <https://doi.org/10.1098/rstb.2000.0717>
- Chippindale, A. K., & Rice, W. R. (2001). Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), 5677–5682. <https://doi.org/10.1073/pnas.101456898>
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D. J., Tomasino, E., Coutanceau, J.-P., Pepey, E., Shirak, A., & Kocher, T. D. (2008). Genetics of sex determination in tilapiine species. *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, 2(1), 43–54. <https://doi.org/10.1159/000117718>
- Corcoran, P., Anderson, J. L., Jacobson, D. J., Sun, Y., Ni, P., Lascoux, M., & Johannesson, H. (2016). Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Research*, 26(4), 486–498. <https://doi.org/10.1101/gr.197244.115>
- da Silva, M., Matoso, D. A., Artoni, R. F., & Feldberg, E. (2014). New approach data in electric fish (Teleostei: Gymnotus): Sex chromosome evolution and repetitive DNA. *Zebrafish*, 11(6), 528–535. <https://doi.org/10.1089/zeb.2013.0966>
- Dan, C., Mei, J., Wang, D., & Gui, J.-F. (2013). Genetic differentiation and efficient sex-specific marker development of a pair of Y- and X-linked markers in yellow catfish. *International Journal of Biological Sciences*, 9(10), 1043–1049. <https://doi.org/10.7150/ijbs.7203>
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510. <https://doi.org/10.1038/nrg3012>
- David, C. J., & Pandian, T. J. (2016). A rapid PCR assay for identification of males in widow tetra, *Gymnocorymbus ternetzi* (BOULENGER, 1895). *European Journal of Biomedical*, 3(6), 436–440.
- Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture*, 208(3), 191–364. [https://doi.org/10.1016/S0044-8486\(02\)00057-1](https://doi.org/10.1016/S0044-8486(02)00057-1)
- Drinan, D. P., Loher, T., & Hauser, L. (2018). Identification of genomic regions associated with sex in Pacific Halibut. *The Journal of Heredity*, 109(3), 326–332. <https://doi.org/10.1093/jhered/esx102>
- Dupoué, A., Lourdais, O., Meylan, S., Brischoux, F., Angelier, F., Rozen-Rechels, D., Marcangeli, Y., Decencière, B., Agostini, S., & Le Galliard, J.-F. (2019). Some like it dry: Water restriction overrides heterogametic sex determination in two reptiles. *Ecology and Evolution*, 9(11), 6524–6533. <https://doi.org/10.1002/ece3.5229>
- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Eché, C., Valière, S., Donnadiou, C., Haffray, P., Bestin, A., Morvezen, R., Acloque, H., Euclide, P. T., Wen, M., Jouano, E., Schartl, M., Postlethwait, J. H., Schraidt, C., ... Guiguen, Y. (2020). Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca Flavescens*. *Molecular Ecology Resources*, 20(2), 531–543. <https://doi.org/10.1111/1755-0998.13133>
- Gamble, T. (2016). Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Molecular Ecology*, 25(10), 2114–2116. <https://doi.org/10.1111/mec.13648>
- Gamble, T., Castoe, T. A., Nielsen, S. V., Banks, J. L., Card, D. C., Schield, D. R., Schuett, G. W., & Booth, W. (2017). The discovery of XY sex chromosomes in a Boa and Python. *Current Biology*, 27(14), 2148–2153.e4. <https://doi.org/10.1016/j.cub.2017.06.010>
- Gamble, T., Coryell, J., Ezaz, T., Lynch, J., Scantlebury, D. P., & 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(5), 1296–1309. <https://doi.org/10.1093/molbev/msv023>
- Gamble, T., McKenna, E., Meyer, W., Nielsen, S. V., Pinto, B. J., Scantlebury, D. P., & Higham, T. E. (2018). XX/XY sex chromosomes in the South American Dwarf Gecko (*Gonatodes humeralis*). *The Journal of Heredity*, 109(4), 462–468. <https://doi.org/10.1093/jhered/esx112>
- Gamble, T., & Zarkower, D. (2014). Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Molecular Ecology Resources*, 14(5), 902–913. <https://doi.org/10.1111/1755-0998.12237>
- Gammerding, W. J., Conte, M. A., Baroiller, J.-F., D'Cotta, H., & Kocher, T. D. (2016). Comparative analysis of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*. *BMC Genomics*, 17(1), 808. <https://doi.org/10.1186/s12864-016-3163-7>
- Geffroy, B., & Douhard, M. (2019). The adaptive sex in stressful environments. *Trends in Ecology & Evolution*, 34(7), 628–640. <https://doi.org/10.1016/j.tree.2019.02.012>
- Gomelsky, B. (2003). Chromosome set manipulation and sex control in common carp: A review. *Aquatic Living Resources*, 16(5), 408–415. [https://doi.org/10.1016/S0990-7440\(03\)00085-8](https://doi.org/10.1016/S0990-7440(03)00085-8)
- Goodenough, U., & Heitman, J. (2014). Origins of eukaryotic sexual reproduction. *Cold Spring Harbor Perspectives in Biology*, 6(3). <https://doi.org/10.1101/cshperspect.a016154>
- Goto-Kazeto, R., Abe, Y., Masai, K., Yamaha, E., Adachi, S., & Yamauchi, K. (2006). Temperature-dependent sex differentiation in goldfish: Establishing the temperature-sensitive period and effect of constant and fluctuating water temperatures. *Aquaculture*, 254(1), 617–624. <https://doi.org/10.1016/j.aquaculture.2005.10.009>
- Gu, L., Walters, J. R., & Knipple, D. C. (2017). Conserved patterns of sex chromosome dosage compensation in the lepidoptera (WZ/ZZ): Insights from a Moth Neo-Z chromosome. *Genome Biology and Evolution*, 9(3), 802–816. <https://doi.org/10.1093/gbe/evx039>
- Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, 30(19), 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>
- Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S. D., Tatsumi, K., Tanaka, K., Motone, F., Kageyama, Y., Nozu, R.,

- Adachi, N., Nishimura, O., Nakagawa, R., Tanegashima, C., Kiyatake, I., Matsumoto, R., Murakumo, K., Nishida, K., ... Kuraku, S. (2018). Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nature Ecology & Evolution*, 2(11), 1761–1771. <https://doi.org/10.1038/s41559-018-0673-5>
- Haugen, T., Almeida, F. F. L., Andersson, E., Bogerd, J., Male, R., Skaar, K. S., Schulz, R. W., Sørhus, E., Wijgerde, T., & Taranger, G. L. (2012). Sex differentiation in Atlantic cod (*Gadus morhua* L.): Morphological and gene expression studies. *Reproductive Biology and Endocrinology*, 10, 47. <https://doi.org/10.1186/1477-7827-10-47>
- Heule, C., Salzburger, W., & Böhne, A. (2014). Genetics of Sexual Development: An Evolutionary Playground for Fish. *Genetics*, 196(3), 579–591. <https://doi.org/10.1534/genetics.114.161158>
- Hohenlohe, P. A., Catchen, J., & Cresko, W. A. (2012). Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. *Methods in Molecular Biology*, 888, 235–260. [https://doi.org/10.1007/978-1-61779-870-2\\_14](https://doi.org/10.1007/978-1-61779-870-2_14)
- Huylmans, A. K., Macon, A., & Vicoso, B. (2017). Global dosage compensation is ubiquitous in lepidoptera, but counteracted by the masculinization of the Z chromosome. *Molecular Biology and Evolution*, 34(10), 2637–2649. <https://doi.org/10.1093/molbev/msx190>
- Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics*, 4(8), 598–612. <https://doi.org/10.1038/nrg1124>
- Kafkas, S., Khodaeiaminjan, M., Güney, M., & Kafkas, E. (2015). Identification of sex-linked SNP markers using RAD sequencing suggests ZW/ZZ sex determination in *Pistacia vera* L. *BMC Genomics*, 16, 98. <https://doi.org/10.1186/s12864-015-1326-6>
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., & Kikuchi, K. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genetics*, 8(7), e1002798. <https://doi.org/10.1371/journal.pgen.1002798>
- Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution. *Developmental Dynamics: an Official Publication of the American Association of Anatomists*, 242(4), 339–353. <https://doi.org/10.1002/dvdy.23927>
- Kirubakaran, T. G., Andersen, Ø., De Rosa, M. C., Andersenstuen, T., Hallan, K., Kent, M. P., & Lien, S. (2019). Characterization of a male specific region containing a candidate sex determining gene in Atlantic cod. *Scientific Reports*, 9(1), 116. <https://doi.org/10.1038/s41598-018-36748-8>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-Bio]. Retrieved from <http://arxiv.org/abs/1303.3997>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liao, Z., Yu, Q., & Ming, R. (2017). Development of male-specific markers and identification of sex reversal mutants in papaya. *Euphytica*, 213(2), 53. <https://doi.org/10.1007/s10681-016-1806-z>
- Martínez-Juárez, A., & Moreno-Mendoza, N. (2019). Mechanisms related to sexual determination by temperature in reptiles. *Journal of Thermal Biology*, 85, 102400. <https://doi.org/10.1016/j.jtherbio.2019.102400>
- Mathers, T. C., Hammond, R. L., Jenner, R. A., Hänfling, B., Atkins, J., & Gómez, A. (2015). Transition in sexual system and sex chromosome evolution in the tadpole shrimp *Triops cancriformis*. *Heredity*, 115(1), 37–46. <https://doi.org/10.1038/hdy.2015.10>
- Matsuda, M., Nagahama, Y., Shinomiya, A. I., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., & Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417(6888), 559–563. <https://doi.org/10.1038/nature751>
- Miura, I. (2017). Sex determination and sex chromosomes in amphibia. *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, 11(5–6), 298–306. <https://doi.org/10.1159/000485270>
- Modi, W. S., & Crews, D. (2005). Sex chromosomes and sex determination in reptiles. *Current Opinion in Genetics & Development*, 15(6), 660–665. <https://doi.org/10.1016/j.gde.2005.09.009>
- Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, 10, 33. <https://doi.org/10.12688/f1000research.29032.1>
- Moore, E. C., & Roberts, R. B. (2013). Polygenic sex determination. *Current Biology*, 23(12), R510–512. <https://doi.org/10.1016/j.cub.2013.04.004>
- Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., & Marais, G. A. B. (2012). Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLOS Biology*, 10(4), e1001308. <https://doi.org/10.1371/journal.pbio.1001308>
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., Schmid, M., & Scharl, M. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), 11778–11783. <https://doi.org/10.1073/pnas.182314699>
- Nanda, I., Scharl, M., Epplen, J. T., Feichtinger, W., & Schmid, M. (1993). Primitive sex chromosomes in poeciliid fishes harbor simple repetitive DNA sequences. *The Journal of Experimental Zoology*, 265(3), 301–308. <https://doi.org/10.1002/jez.1402650311>
- Nanda, I., Schories, S., Tripathi, N., Dreyer, C., Haaf, T., Schmid, M., & Scharl, M. (2014). Sex chromosome polymorphism in guppies. *Chromosoma*, 123(4), 373–383. <https://doi.org/10.1007/s00412-014-0455-z>
- Nielsen, S. V., Banks, J. L., Diaz, R. E., Trainor, P. A., & Gamble, T. (2018). Dynamic sex chromosomes in Old World chameleons (Squamata: Chamaeleonidae). *Journal of Evolutionary Biology*, 31(4), 484–490. <https://doi.org/10.1111/jeb.13242>
- Nielsen, S. V., Daza, J. D., Pinto, B. J., & Gamble, T. (2019). ZZ/ZW sex chromosomes in the endemic Puerto Rican Leaf-Toed Gecko (*Phyllodactylus wirshingi*). *Cytogenetic and Genome Research*, 157(1–2), 89–97. <https://doi.org/10.1159/000496379>
- Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology*, 28(21), 4709–4724. <https://doi.org/10.1111/mec.15245>
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., Parrinello, H., Journot, L., Klopp, C., Lluch, J., Roques, C., Postlethwait, J., Scharl, M., Herpin, A., & Guiguen, Y. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics*, 15(8), e1008013. <https://doi.org/10.1371/journal.pgen.1008013>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, 8(10), 1360–1373. <https://doi.org/10.1111/2041-210X.12775>
- Peichel, C. L., Ross, J. A., Matson, C. K., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Mori, S., Schluter, D., & Kingsley, D. M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology*, 14(16), 1416–1424. <https://doi.org/10.1016/j.cub.2004.08.030>
- Pennell, M. W., Mank, J. E., & Peichel, C. L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species.



- Molecular Ecology*, 27(19), 3950–3963. <https://doi.org/10.1111/mec.14540>
- Pezaro, N., Doody, J. S., & Thompson, M. B. (2017). The ecology and evolution of temperature-dependent reaction norms for sex determination in reptiles: A mechanistic conceptual model. *Biological Reviews of the Cambridge Philosophical Society*, 92(3), 1348–1364. <https://doi.org/10.1111/brv.12285>
- Piferrer, F., Blázquez, M., Navarro, L., & González, A. (2005). Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *General and Comparative Endocrinology*, 142(1–2), 102–110. <https://doi.org/10.1016/j.ygcen.2005.02.011>
- Pratlong, M., Haguénauer, A., Chéneseau, S., Brener, K., Mitta, G., Toulza, E., & Pontarotti, P. (2017). Evidence for a genetic sex determination in Cnidaria, the Mediterranean red coral (*Corallium rubrum*). *Royal Society Open Science*, 4(3), 160880. <https://doi.org/10.1098/rsos.160880>
- Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12), 2640–2659. <https://doi.org/10.1038/nprot.2017.123>
- Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: Effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, 16(4), 991–1001. <https://doi.org/10.1111/1755-0998.12518>
- Salz, H. K., & Erickson, J. W. (2010). Sex determination in *Drosophila*: The view from the top. *Fly*, 4(1), 60–70. <https://doi.org/10.4161/fly.4.1.11277>
- Scavi, B., & Herrick, J. (2019). Genome size variation and species diversity in salamanders. *Journal of Evolutionary Biology*, 32(3), 278–286. <https://doi.org/10.1111/jeb.13412>
- Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2018). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907–917.
- Sigeman, H., Ponnikas, S., Videvall, E., Zhang, H., Chauhan, P., Naurin, S., & Hansson, B. (2018). Insights into avian incomplete dosage compensation: Sex-biased gene expression coevolves with sex chromosome degeneration in the common whitethroat. *Genes*, 9(8), 373–373. <https://doi.org/10.3390/genes9080373>
- Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., Nater, A., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., & Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*, 6, 7330. <https://doi.org/10.1038/ncomms8330>
- Spigler, R. B., Lewers, K. S., Main, D. S., & Ashman, T.-L. (2008). Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity*, 101(6), 507–517. <https://doi.org/10.1038/hdy.2008.100>
- Star, B., Tørresen, O. K., Nederbragt, A. J., Jakobsen, K. S., Pampoulie, C., & Jentoft, S. (2016). Genomic characterization of the Atlantic cod sex-locus. *Scientific Reports*, 6, 31235. <https://doi.org/10.1038/srep31235>
- Utsunomia, R., Scacchetti, P. C., Hermida, M., Fernández-Cebrián, R., Taboada, X., Fernández, C., Bekaert, M., Mendes, N. J., Robledo, D., Mank, J. E., Taggart, J. B., Oliveira, C., Foresti, F., & Martínez, P. (2017). Evolution and conservation of Characidium sex chromosomes. *Heredity*, 119(4), 237–244. <https://doi.org/10.1038/hdy.2017.43>
- van Doorn, G. S., & Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature*, 449(7164), 909–912. <https://doi.org/10.1038/nature06178>
- Volff, J. N., & Schartl, M. (2001). Variability of genetic sex determination in poeciliid fishes. *Genetica*, 111(1–3), 101–110.
- Wallis, M. C., Waters, P. D., & Graves, J. (2008). Sex determination in mammals—Before and after the evolution of SRY. *Cellular and Molecular Life Sciences*, 65(20), 3182–3195. <https://doi.org/10.1007/s00018-008-8109-z>
- Watanabe, T., Yamasaki, K., Seki, S., & Taniguchi, N. (2004). Detection of ayu sex-linked DNA markers using homologous clones. *Fisheries Science*, 70(1), 47–52. <https://doi.org/10.1111/j.1444-2906.2003.00769.x>
- Wen, M., Feron, R., Pan, Q., Guguin, J., Jouanno, E., Herpin, A., Klopp, C., Cabau, C., Zahm, M., Parrinello, H., Journot, L., Burgess, S. M., Omori, Y., Postlethwait, J. H., Schartl, M., & Guiguen, Y. (2019). Sex chromosome and sex locus characterization in the goldfish, *Carassius auratus*. *BMC Genomics*, 21, 552.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., & Knorr, C. (2017). DdRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics*, 18(1), 531. <https://doi.org/10.1186/s12864-017-3930-0>
- Whitehead, J. A., Benfey, T. J., & Martin-Robichaud, D. J. (2012). Ovarian development and sex ratio of gynogenetic Atlantic cod (*Gadus morhua*). *Aquaculture*, 324–325, 174–181. <https://doi.org/10.1016/j.aquaculture.2011.10.039>
- Wilson, C. A., High, S. K., McCluskey, B. M., Amores, A., Yan, Y.-L., Titus, T. A., Anderson, J. L., Batzel, P., Carvan, M. J., Schartl, M., & Postlethwait, J. H. (2014). Wild sex in zebrafish: Loss of the natural sex determinant in domesticated strains. *Genetics*, 198(3), 1291–1308. <https://doi.org/10.1534/genetics.114.169284>
- Wu, C. I. (1983). The fate of autosomal modifiers of the sex-ratio trait in *Drosophila* and other sex-linked meiotic drive systems. *Theoretical Population Biology*, 24(2), 107–120. [https://doi.org/10.1016/0040-5809\(83\)90035-7](https://doi.org/10.1016/0040-5809(83)90035-7)
- Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications*, 6(3), 486–496. <https://doi.org/10.1111/eva.12032>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Feron R, Pan Q, Wen M, et al.

RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol Ecol Resour*. 2021;00:1–17. <https://doi.org/10.1111/1755-0998.13360>