MR. ROMAIN FERON (Orcid ID : 0000-0001-5893-6184)
DR. CHRISTOPHE KLOPP (Orcid ID : 0000-0001-7126-5477)
DR. MATTHIAS STÖCK (Orcid ID : 0000-0003-4888-8371)
DR. ROBERT M. WATERHOUSE (Orcid ID : 0000-0003-4199-9052)

Article type : Resource Article

# RADSex: a computational workflow to study sex determination using Restriction Site-Associated DNA Sequencing data

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

**Affiliations**

1. INRAE, LPGP, 35000, Rennes, France
2. Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
3. Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
4. State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan Normal University, Changsha, China
5. Department of Systematic Biology, EBC, Uppsala University, Norbyvägen 18D 752 36 Uppsala, Sweden

6. Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Univ. Montpellier, F-34094 Montpellier, France
7. SIGENAE, Mathématiques et Informatique Appliquées de Toulouse, INRAE, Castanet Tolosan, France
8. Physiological Chemistry, Biocenter, University of Wuerzburg, 97074 Wuerzburg, Germany
9. The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas, USA
10. Developmental Biochemistry, Biocenter, University of Wuerzburg, 97074 Wuerzburg, Germany
11. Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA
12. Environmental and Fisheries Sciences Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd East, Seattle, WA 98112, USA
13. Department of Materials and Life Sciences, Faculty of Science and Technology, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
14. Department of Marine and Environmental Sciences, Marine Science Center, Northeastern University, Nahant, Massachusetts 01908
15. Reproductive and Molecular Biology Group, Department of Structural and Functional Biology, Institute of Biosciences, São Paulo State University, Botucatu, São Paulo, Brazil
16. Department of Aquatic Marine Biosciences, Tokyo University of Marine Science and Technology, Tokyo 108-8477, Japan
17. Institute of Marine Research, P.O. Box 1870, Nordnes, NO-5817, Bergen, Norway
18. Institute of Hydrobiology, Chinese Academy of Sciences, Beijing, China.
19. Leibniz-Institute of Freshwater Ecology and Inland Fisheries, IGB, Berlin, Germany
20. Department of Integrative Biology, Michigan State University, East Lansing, MI, USA

*Corresponding authors: yann.guiguen@inrae.fr; + 33 2 48 50 09; romainferon91@gmail.com;

# ABSTRACT

The study of sex determination and sex chromosome organisation in non-model 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-analyzed a published dataset of Japanese medaka, *Oryzias latipes*, where we uncovered a previously unknown Y chromosome polymorphism. We then used RADSex to analyze new RAD-Seq datasets 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 non-model species thanks to its speed of analyses, low resource usage, ease of application, and visualization options. Furthermore, our analysis of new datasets from 15 species provides new insights on sex determination in fish.

**Keywords:** sex determination, RAD-Sequencing, fish, computational workflow, visualization

# 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 fueled divergent evolution of morphological, physiological, and behavioral 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), like in virtually all mammals and birds (Bachtrog *et al.*, 2014); by environmental factors (Environmental Sex Determination, ESD), like temperature in many non-avian reptiles (Pezaro, Doody, & Thompson, 2017); or a combination of the both, as in several fish species (Piferrer, Blázquez, Navarro, & González, 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, Waters, & Graves, 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; L. Gu, Walters, & Knipple, 2017; Huylmans, Macon, & Vicoso, 2017; Muyle *et al.*, 2012; Peichel *et al.*, 2004). More recent studies investigating sex chromosomes in insects (Blackmon, Ross, & Bachtrog, 2017), non-avian 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, *i.e.,* 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, Dénes, Krüger, Székely, & Beissinger, 2017), as well as agriculture (Al-Ameri, Al-Qurainy, Gaafar, Khan, & Nadeem, 2016; Liao, Yu, & Ming, 2017; Spigler, Lewers, Main, & Ashman, 2008) and animal production (Dan, Mei, Wang, & Gui, 2013; Yano *et al.*, 2013), yet the sex of an individual cannot always be easily determined by non-invasive 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, Rogers, Dean, & Wright, 2019), but recent advances in sequencing technologies enable 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, Good, Miller, Luikart, & Hohenlohe, 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, Loher, & Hauser, 2018), amphibians (Bewick *et al.*, 2013), non-avian reptiles (Gamble, 2016; Gamble *et al.*, 2017, 2015, 2018; Gamble & Zarkower, 2014; S. V. Nielsen, Banks, Diaz, Trainor, & Gamble, 2018; Nielsen, Daza, Pinto, & Gamble, 2019), invertebrates (Carmichael *et al.*, 2013; Mathers *et al.*, 2015; Pratlong *et al.*, 2017), and plants (Kafkas, Khodaeiaminjan, Güney, & Kafkas, 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 analyze RAD-Seq data specifically for sex determination (Gamble & Zarkower, 2014), most of the above mentioned studies have used Stacks (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Hohenlohe, Catchen, & Cresko, 2012; Rochette & Catchen, 2017) to cluster RAD-Seq reads into polymorphic markers subsequently filtered with custom in-house scripts. Stacks, however, requires substantial computational resources to run on large datasets and depends on multiple parameters that can greatly influence the outcome of the analysis (Paris, Stevens, & Catchen, 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 datasets difficult, for instance, datasets containing strong population structure, individuals with mis-assigned 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 analyzed a previously published dataset 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 analyze new datasets 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 non-model species thanks to its speed, resource usage, and ease of use. Furthermore, our multi-species 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.

# MATERIAL AND METHODS

## Overview of the RADSex analysis workflow

The underlying principle of RADSex is to group identical RAD-Seq reads from all individuals in a dataset into non-polymorphic 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 (**Fig. 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.

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 dataset. This table summarizes all the sequence information present in the dataset, and each row constitutes a marker for *radsex* (**Supp. Fig. 1**). Sequences are read from each individual reads 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 (**Supp. Table 1**) 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) (**Supp. Fig. 1**). 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.,* $p < 0.05$ (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* (**Supp. Table 2**) 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 color 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 **Fig. 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 (**Supp. Fig. 1**). 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 clustering is performed with the base R *hclust* function; both distance calculation (*e.g.,* euclidean, maximum, binary ...) and clustering methods (*e.g.*, complete, average, centroid ...)

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 **Fig. 4.B** for an example from a real dataset). 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: 1) 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 2) the sex-bias (*S*), defined as $S = ♂ / ♂_{total} - ♀ / ♀_{total}$, where ♂ and ♀ are the number of males and females in which the marker is present, and $♂_{total}$ and $♀_{total}$ 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* (**Supp. Table 3**), 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, Gu, Eils, Schlesner, & Brors, 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 dataset. 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*.

## Sample collection

General information on the different species, species collectors, and samples is given in **Supp. Table 5**. Detailed information on sample collection for each dataset is provided in the supplementary material (**Supplementary Methods 1**).

## Genomic DNA (gDNA) extraction

Genomic DNA (gDNA) of *Lepisosteus oculatus*, *Plecoglossus altivelis*, *Gadus morhua*, *Notothenia rossii*, *Sander vitreus*, *Tinca tinca*, *Carassius auratus*, and all *Danios* were extracted from fin clips stored in ethanol using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) following the producer's protocol. Genomic DNA of *Gymnotus carapo*, *Gymnocorymbus ternetzi*, *Hippocampus abdominalis*, and *Poecilia sphenops* were extracted from individual fin clips stored in ethanol or from pooled organs of individual fish for *P. 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 three hours 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 5000 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 5000 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, Wilmington, Delaware) and/or Qubit (ThermoFisher, France) before processing gDNA for RAD libraries construction.

## RAD-Sequencing 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, Catchen, Ferrara, Fontenot, & Postlethwait, 2011). Libraries were sequenced as single end 100 bp reads on one lane of 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 **Supp. Table 7**.

## Performance measurements

Runtime and peak memory usage of *radsex* (version 1.0.0) was measured with the "benchmark" directive of Snakemake (Mölder *et al.*, 2021) on the Genotoul computational platform (Toulouse, France) using 4 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.

## 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 license; 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.

# RESULTS

## Performance of RADSex

The runtime and peak memory usage of *radsex process, distrib,* and *signif* were measured on datasets from 15 fish species that we generated for this study. Apart from one outlier dataset, *Danio albolineatus*, RADSex's runtime was linearly and positively correlated to the number of RAD-seq reads, ranging from 276 seconds for *Tinca tinca*, whose dataset contained ~ 120 M. RAD-seq reads, to 1240 seconds (20 min and 40 s) for *Sander vitreus*, whose dataset contained ~ 232 M. RAD-seq reads (**Fig. 3.A**). Runtime and peak memory usage for all datasets are provided in **Supp. Table 4**. 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 *G. ternetzi* (**Fig. 3.B**). The high runtime and memory usage of *radsex*

for the *Danio albolineatus* dataset 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 dataset. 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 dataset because filtering parameters (*i.e.,* minimum depth for marker presence) can be applied in the later steps of the RADSex pipeline.

## RADSex validation on a published dataset

To assess whether RADSex can recover previous results on sex determination, we analyzed a dataset that has been used to characterize a major sex-determination locus (Wilson *et al.*, 2014). This dataset consists of RAD-Seq reads from 31 males and 30 females from the *carbio* 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 analyzed this *O. latipes* dataset 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 phenotype (*radsex signif*, $p < 0.05$, chi-squared test with Bonferroni correction, highlighted tiles in **Fig. 4.A**). 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 5 males was significantly associated with female phenotype (**Fig. 4.A**). 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* (**Fig. 4.B**). As expected, phenotypic males and females clustered in two separate groups, except for two phenotypic males (green arrows in **Fig. 4.B**) 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.* (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 ($p < 0.05$, chi-squared test with Bonferroni correction), as expected for an XY/XX system. In addition, six males forming a sub-cluster 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) 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 (**Fig. 4.C**). On Ola1, markers significantly associated with male phenotype aligned to a continuous region spanning from 14.7 Mb to 31.9 Mb (**Fig. 4.D**). Within this region, markers found in all males and no females aligned between 17 and 23 Mb (green box in **Fig. 4.D**) and markers present in all but the six males identified in the clustering step and absent from all females aligned between 14.7 Mb and 31.9 Mb (purple boxes in **Fig. 4.D**). 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 dataset 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 (**Supp. Fig. 2.A**). The original study by Wilson *et al.* (2014) identified 248 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 Supplementary 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 (**Supp. Fig. 2.B**). 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) (**Supp. Fig. 2.C**). Second, the number of markers used for Bonferroni correction was higher for RADSex (121,492 markers, threshold indicated by the blue dashed line in **Supp. Fig. 2.C**) 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 **Supp. Fig. 2C**). Using the same threshold for Bonferroni correction (red vertical dashed line in **Supp. Fig. 2.C**), the number of SNPs located within RADSex markers not significantly associated with sex dropped to 22 SNPs (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.* (**Supp. Fig. 2.D**), 40 of which (60%) were aligned to Ola1 between 14.6 Mb 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.* (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 datasets that were previously overlooked.

## Analysis of 15 new ray-finned fish RAD-sequencing datasets

RAD-Seq datasets 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 (**Supp. Fig. 3**). We identified an XX/XY sex determining system in six species (*Cyprinus carpio*, *Gymnotus carapo*, *Plecoglossus atlivelis*, *Tinca tinca*, *Gadus morhua*, and *Poecilia sphenops*) with a few male and/or female outliers in four of these species (**Fig. 5**). Markers significantly associated with phenotypic sex for each species are provided in **Supp. Table 6**. An XX/XY SD system is in agreement with previous reports for *C. carpio* (Gomelsky, 2003), *G. carapo* (da Silva, Matoso, Artoni, & Feldberg, 2014), *P. altivelis* (Watanabe, Yamasaki, Seki, & Taniguchi, 2004), and *G. morhua* (Haugen *et al.*, 2012; Whitehead, Benfey, & Martin-Robichaud, 2012). The XX/XY SD system inferred from RADSex results for *P. sphenops* conflicts with published findings on an ornamental population for that species that showed a female heterogametic system (Nanda, Schartl, Epplen, Feichtinger, & Schmid, 1993). However, alternative population-specific sex determination

systems have also been described for *P. sphenops* (Volff & Schartl, 2001). Because these studies were performed on ornamental fish and laboratory strains, previously derived conclusions may not apply to wild populations like the one we analyzed. Lastly, to the best of our knowledge, sex determination was not previously characterized in the tench, *Tinca tinca*.

In *G. 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 **(Fig. 5.B)**. Furthermore, multiple markers were found predominantly in males but not significantly associated with sex, with eight markers consistently present with depth higher than five in 20 males and absent from all females (these markers are highlighted with arrows in **Fig. 6.A**). The three markers significantly associated with sex (red-outlined boxes) and most other markers strongly but not significantly associated with sex aligned to cod linkage group 11 (Gmo11) (**Fig. 6.B and 6.C**). Markers strongly associated with sex aligned to a region spanning from 0 to 20 Mb on Gmo11, and the three markers significantly associated with sex aligned between 10 to 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 observed 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 aquaculture facilities (Pan *et al.*, 2019).

In the nine other species (**Table 1** and **Supp. Fig. 4**), we did not find any marker significantly associated with phenotypic sex, which indicates that these species either 1) have a small undifferentiated sex locus, 2) lack a genetic sex determination mechanism, or 3) rely on a complex combination of environmental and genetic sex determination factors that prevented the detection of a clear sex-biased signal in the dataset.

# DISCUSSION

We demonstrate that RADSex is an efficient and -- thanks to its simple usage and visualization options -- user-friendly tool to identify sex-determination systems and sex-biased markers from RAD-Seq data. Unlike other popular RAD-Seq analysis software, *e.g.,* Stacks (Catchen *et al.*, 2011) and PyRad (Eaton, 2014), which catalog polymorphic loci, RADSex creates monomorphic markers. RADSex therefore counts as a separate marker each allele at a locus 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 datasets. 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 absence of any major environmental effects on sex, as demonstrated by our analyses on the ayu (*P. altivelis*) and the tench (*T. tinca*) datasets: 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 datasets, for instance in the presence of outliers and sequencing biases. These complex datasets 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 non-complete sex-linkage of sex-biased markers in a publicly available medaka dataset (Wilson *et al.*, 2014), as well as outliers in our own common carp (*C. carpio*), Atlantic cod (*G. morhua*), banded knifefish (*G. carapo*), and common molly (*P. sphenops*) datasets. 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 non-reproductively 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, 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 non-polymorphic 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 dataset, which is affected by the minimum depth parameter. In the 15 datasets that we generated, median sequencing coverage in an individual was between 30x and 100x, except for three species for which it was lower than 30x (**Supp. Fig. 3**). 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 datasets, however, may require using low minimum depth values, for instance the three datasets for which median individual sequencing depth was low, or species with large genome sizes like 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 datasets with minimum depth as low as one, which can require a lot of memory with other software. In addition to sequencing errors, common sources of biases in RAD-Sequencing analyses include highly repeated sequences (Catchen *et al.*, 2011), PCR duplicates, and allele drop-out (Andrews *et al.*, 2016). RADSex considers 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 identifying 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 non-recombining 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 non-recombining 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 a complete absence of markers associated with phenotypic sex for any value of minimum depth, which we observed in nine of our datasets. 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 like pooled sequencing of males and females (Feron *et al.*, 2020; Gammerdinger, Conte, Baroiller, D'Cotta, & Kocher, 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 1) a polygenic system, 2) 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 3) 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 (Kikuchi and Hamakuchi 2013, Heule *et al.*, 2014). 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 dataset. 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 case of 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 non-model 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 reanalysis of a publicly available Japanese medaka dataset (Wilson *et al.*, 2014), in which we detected multiple Y chromosome haplotypes that were not found in the original study. Intra-specific 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 intra-specific 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 SD systems in six of these 15 species but did not find evidence of a ZZ/ZW SD 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 SD systems in our datasets is in agreement with previous findings that in teleosts, transitions from ZZ/ZW to XX/XY SD systems are more frequent than the reverse (Pennell, Mank, & Peichel, 2018), which would result in XY SD systems being prevalent in this clade.

All of our datasets 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 manuscript. 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, MS), the National Institutes of Health (USA) grants R01GM085318 and R35GM139635, JHP), and the National Science Foundation (USA) Office of Polar Programs (grants PLR-1247510 and PLR-1444167 to HWD and grant OPP-1543383 to JHP, TD, and HWD). The MGX core sequencing facility was supported by France Genomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). RF was partially supported by Swiss National Science Foundation grant PP00P3_170664 to RMW. 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 Bayousphere Lab at Nicholls State University (Thibodeaux, LA) for support with spotted gar sample collection and preparation, 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. Part of this work was supported by a contribution from the Marine Science Center at Northeastern University (contribution #TBD). We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi: 10.15454/1.5572369328961167E12) for providing computing and storage resources.

# 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*, *2016*. doi: 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. doi: 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). doi: 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. doi: 10.1038/nrg.2015.28

Bachtrog, D. (2008). The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics*, *179*(3), 1513–1525. doi: 10.1534/genetics.107.084012

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., … Consortium, T. T. of S. (2014). Sex Determination: Why So Many Ways of Doing It? *PLOS Biology*, *12*(7), e1001899. doi: 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. doi: 10.1159/000452362

Benestan, L., Moore, J.-S., Sutherland, B. J. G., Le Luyer, J., Maaroufi, H., Rougeux, C., … 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. doi: 10.1111/mec.14217

Bewick, A. J., Chain, F. J. J., Zimmerman, L. B., Sesay, A., Gilchrist, M. J., Owens, N. D. L., … 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. doi: 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. doi: 10.1093/jhered/esw047

Carmichael, S. N., Bekaert, M., Taggart, J. B., Christie, H. R. L., Bassett, D. I., Bron, J. E., …

Sturm, A. (2013). Identification of a sex-linked SNP marker in the salmon louse (Lepeophtheirus salmonis) using RAD sequencing. *PloS One*, *8*(10), e77832. doi: 10.1371/journal.pone.0077832

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. doi: 10.1111/mec.12354

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 (Bethesda, Md.)*, *1*(3), 171–182. doi: 10.1534/g3.111.000240

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. doi: 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. doi: 10.1073/pnas.101456898

Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., … 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. doi: 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*. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1016/S0044-8486(02)00057-1

Doorn, G. S. van, & Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature*, *449*(7164), 909–912. doi: 10.1038/nature06178

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. doi: 10.1093/jhered/esx102

Dupoué, A., Lourdais, O., Meylan, S., Brischoux, F., Angelier, F., Rozen-Rechels, D., … 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. doi: 10.1002/ece3.5229

Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics (Oxford, England)*, *30*(13), 1844–1849. doi: 10.1093/bioinformatics/btu121

Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., … 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. doi: 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. doi: 10.1111/mec.13648

Gamble, T., Castoe, T. A., Nielsen, S. V., Banks, J. L., Card, D. C., Schield, D. R., … Booth, W. (2017). The Discovery of XY Sex Chromosomes in a Boa and Python. *Current Biology: CB*, *27*(14), 2148-2153.e4. doi: 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. doi: 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. doi: 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. doi: 10.1111/1755-0998.12237

Gammerdinger, 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. doi: 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. doi: 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. doi: 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). doi: 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. doi: 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. doi: 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. doi: 10.1093/bioinformatics/btu393

Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S. D., … Kuraku, S. (2018). Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nature Ecology & Evolution*, *2*(11), 1761–1771. doi: 10.1038/s41559-018-0673-5

Haugen, T., Almeida, F. F., Andersson, E., Bogerd, J., Male, R., Skaar, K. S., … Taranger, G. L. (2012). Sex differentiation in Atlantic cod (Gadus morhua L.): Morphological and gene expression studies. *Reproductive Biology and Endocrinology: RB&E*, *10*, 47. doi: 10.1186/1477-7827-10-47

781 Hohenlohe, P. A., Catchen, J., & Cresko, W. A. (2012). Population genomic analysis of model
782 and nonmodel organisms using sequenced RAD tags. *Methods in Molecular Biology*
783 *(Clifton, N.J.)*, *888*, 235–260. doi: 10.1007/978-1-61779-870-2_14
784 Huylmans, A. K., Macon, A., & Vicoso, B. (2017). Global Dosage Compensation Is Ubiquitous
785 in Lepidoptera, but Counteracted by the Masculinization of the Z Chromosome.
786 *Molecular Biology and Evolution*, *34*(10), 2637–2649. doi: 10.1093/molbev/msx190
787 Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker
788 comes of age. *Nature Reviews Genetics*, *4*(8), 598–612. doi: 10.1038/nrg1124
789 Kafkas, S., Khodaeiaminjan, M., Güney, M., & Kafkas, E. (2015). Identification of sex-linked
790 SNP markers using RAD sequencing suggests ZW/ZZ sex determination in Pistacia vera
791 L. *BMC Genomics*, *16*, 98. doi: 10.1186/s12864-015-1326-6
792 Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., … Kikuchi, K. (2012). A
793 trans-species missense SNP in Amhr2 is associated with sex determination in the tiger
794 pufferfish, Takifugu rubripes (fugu). *PLoS Genetics*, *8*(7), e1002798. doi:
795 10.1371/journal.pgen.1002798
796 Karlsen, Ø., Meeren, T. van der, Rønnestad, I., Mangor-Jensen, A., Galloway, T. F., Kjørsvik,
797 E., & Hamre, K. (2015). Copepods enhance nutritional status, growth and development in
798 Atlantic cod (Gadus morhua L.) larvae—Can we identify the underlying factors? *PeerJ*,
799 *3*, e902. doi: 10.7717/peerj.902
800 Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome
801 evolution. *Developmental Dynamics: An Official Publication of the American*
802 *Association of Anatomists*, *242*(4), 339–353. doi: 10.1002/dvdy.23927
803 Kirubakaran, T. G., Andersen, Ø., De Rosa, M. C., Andersstuen, T., Hallan, K., Kent, M. P., &
804 Lien, S. (2019). Characterization of a male specific region containing a candidate sex
805 determining gene in Atlantic cod. *Scientific Reports*, *9*(1), 116. doi: 10.1038/s41598-018-
806 36748-8
807 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
808 *ArXiv:1303.3997 [q-Bio]*. Retrieved from http://arxiv.org/abs/1303.3997
809 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project
810 Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools.
811 *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. doi:
812 10.1093/bioinformatics/btp352
813 Liao, Z., Yu, Q., & Ming, R. (2017). Development of male-specific markers and identification of

814 sex reversal mutants in papaya. *Euphytica*, *213*(2), 53. doi: 10.1007/s10681-016-1806-z
815 Martínez-Juárez, A., & Moreno-Mendoza, N. (2019). Mechanisms related to sexual
816 determination by temperature in reptiles. *Journal of Thermal Biology*, *85*, 102400. doi:
817 10.1016/j.jtherbio.2019.102400
818 Mathers, T. C., Hammond, R. L., Jenner, R. A., Hänfling, B., Atkins, J., & Gómez, A. (2015).
819 Transition in sexual system and sex chromosome evolution in the tadpole shrimp Triops
820 cancriformis. *Heredity*, *115*(1), 37–46. doi: 10.1038/hdy.2015.10
821 Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., … Sakaizumi,
822 M. (2002). DMY is a Y-specific DM-domain gene required for male development in the
823 medaka fish. *Nature*, *417*(6888), 559–563. doi: 10.1038/nature751
824 Miura, I. (2017). Sex Determination and Sex Chromosomes in Amphibia. *Sexual Development:*
825 *Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of*
826 *Sex Determination and Differentiation*, *11*(5–6), 298–306. doi: 10.1159/000485270
827 Modi, W. S., & Crews, D. (2005). Sex chromosomes and sex determination in reptiles. *Current*
828 *Opinion in Genetics & Development*, *15*(6), 660–665. doi: 10.1016/j.gde.2005.09.009
829 Mölder F, Jablonski KP, Letcher B, … Köster, J. (2021) Sustainable data analysis with
830 Snakemake [version 1; peer review: awaiting peer review]. *F1000Research*. 10:33.
831 doi:10.12688/f1000research.29032.1
832 Moore, E. C., & Roberts, R. B. (2013). Polygenic sex determination. *Current Biology: CB*,
833 *23*(12), R510-512. doi: 10.1016/j.cub.2013.04.004
834 Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., & Marais, G. A. B. (2012).
835 Rapid De Novo Evolution of X Chromosome Dosage Compensation in Silene latifolia, a
836 Plant with Young Sex Chromosomes. *PLOS Biology*, *10*(4), e1001308. doi:
837 10.1371/journal.pbio.1001308
838 Nanda, I., Schartl, M., Epplen, J. T., Feichtinger, W., & Schmid, M. (1993). Primitive sex
839 chromosomes in poeciliid fishes harbor simple repetitive DNA sequences. *The Journal of*
840 *Experimental Zoology*, *265*(3), 301–308. doi: 10.1002/jez.1402650311
841 Nanda, Indrajit, Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., … Schartl, M.
842 (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y
843 chromosome of the medaka, Oryzias latipes. *Proceedings of the National Academy of*
844 *Sciences of the United States of America*, *99*(18), 11778–11783. doi:
845 10.1073/pnas.182314699
846 Nanda, Indrajit, Schories, S., Tripathi, N., Dreyer, C., Haaf, T., Schmid, M., & Schartl, M.

(2014). Sex chromosome polymorphism in guppies. *Chromosoma*, *123*(4), 373–383. doi: 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. doi: 10.1111/jeb.13242

Nielsen, Stuart 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. doi: 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. doi: https://doi.org/10.1111/mec.15245

Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., … 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. doi: 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. doi: 10.1111/2041-210X.12775

Peichel, C. L., Ross, J. A., Matson, C. K., Dickson, M., Grimwood, J., Schmutz, J., … Kingsley, D. M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology: CB*, *14*(16), 1416–1424. doi: 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. doi: 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. doi: 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. doi: 10.1016/j.ygcen.2005.02.011

Pratlong, M., Haguenauer, A., Chenesseau, 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. doi: 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. doi: 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. doi: 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. doi: 10.4161/fly.4.1.11277

Sclavi, B., & Herrick, J. (2019). Genome size variation and species diversity in salamanders. *Journal of Evolutionary Biology*, *32*(3), 278–286. doi: 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*, 907–917. doi: 10.1111/2041-210X.12700@10.1111/(ISSN)2041-210x.PracticalToolsFieldMethodsMEE32018

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). doi: 10.3390/genes9080373

Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., … Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*, *6*, 7330. doi: 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. doi: 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. doi: 10.1038/srep31235

Utsunomia, R., Scacchetti, P. C., Hermida, M., Fernández-Cebrián, R., Taboada, X., Fernández, C., … Martínez, P. (2017). Evolution and conservation of Characidium sex

chromosomes. *Heredity*, *119*(4), 237–244. doi: 10.1038/hdy.2017.43

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. a. M. (2008). Sex determination in mammals—Before and after the evolution of SRY. *Cellular and Molecular Life Sciences: CMLS*, *65*(20), 3182–3195. doi: 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. doi: 10.1111/j.1444-2906.2003.00769.x

Wen, M., Feron, R., Pan, Q., Guguin, J., Jouanno, E., Herpin, A., … Guiguen, Y. (2019). Sex chromosome and sex locus characterization in the goldfish, Carassius auratus. *BioRxiv*, 2019.12.20.875377. doi: 10.1101/2019.12.20.875377

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. doi: 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. doi: 10.1016/j.aquaculture.2011.10.039

Wilson, C. A., High, S. K., McCluskey, B. M., Amores, A., Yan, Y., Titus, T. A., … Postlethwait, J. H. (2014). Wild sex in zebrafish: Loss of the natural sex determinant in domesticated strains. *Genetics*, *198*(3), 1291–1308. doi: 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. doi: 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. doi: 10.1111/eva.12032

# DATA ACCESSIBILITY

All RAD-Sequencing 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-

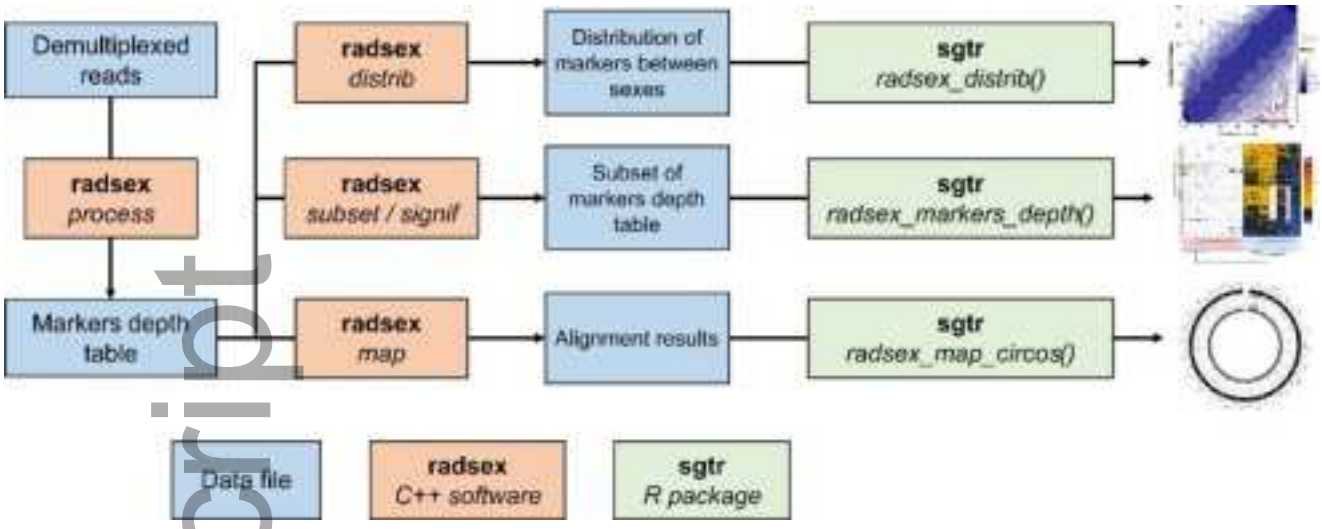
sexdetermination-radsex. RADSex is released under GPLv3 license; source code, installation instructions, and documentation are available at https://github.com/SexGenomicsToolkit/radsex. *sgtr* is released under GPLv3 license and available at https://github.com/SexGenomicsToolkit/sgtr.

# AUTHOR CONTRIBUTIONS

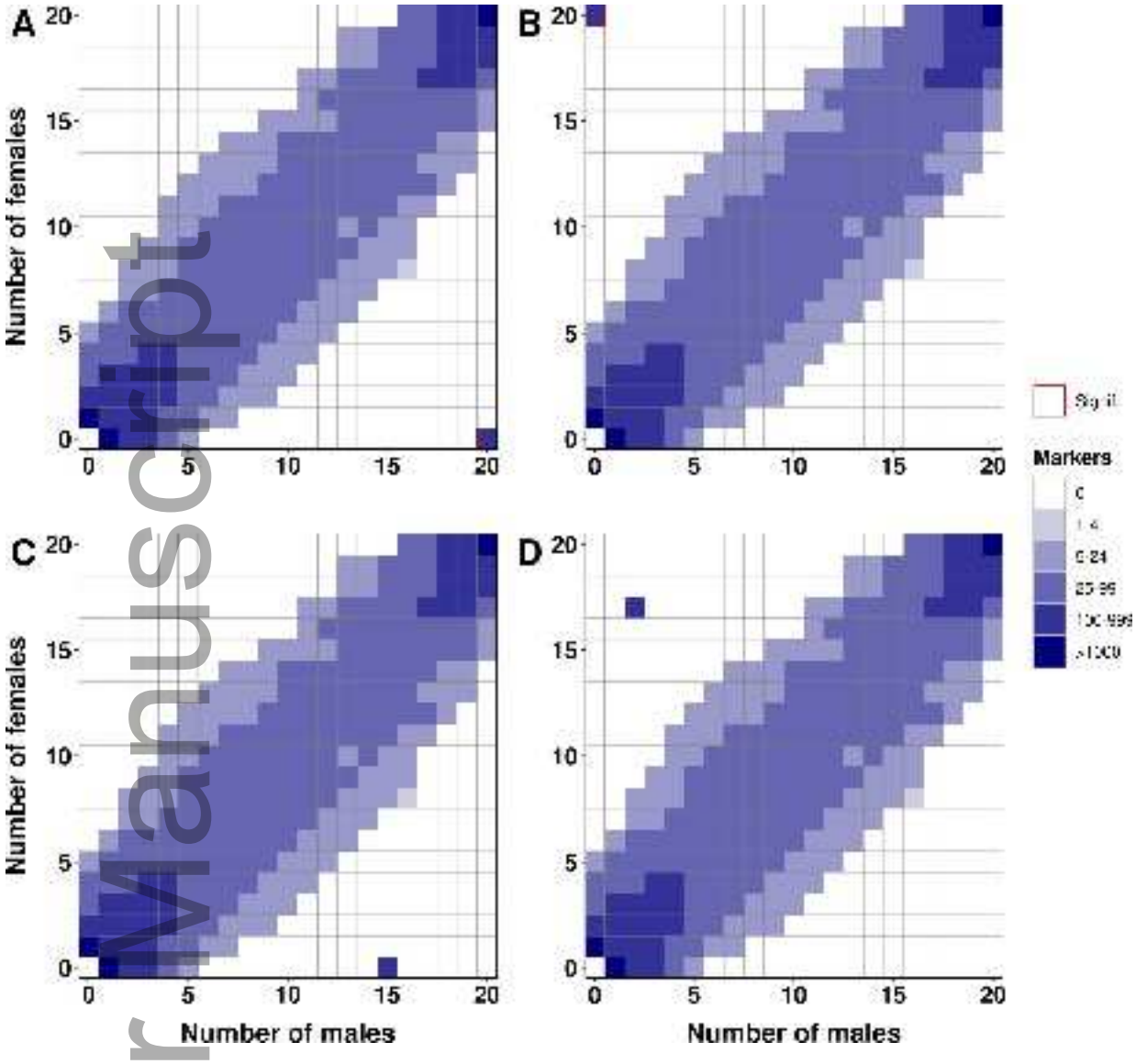
RF designed and implemented RADSex and sgtr, with feedback from QP, MW, and YG. YG, JHP, and MSch designed the PhyloSex project and advised on results interpretation. QP, MW, BI, JA, AH, KK, ASR, KD, SK, CK, JHP, MSch and YG participated in the analysis of the results. HP and LJ prepared libraries and performed the sequencing. RF, QP, JHP, MSch and YG drafted the manuscript. RF, QP, RMW, JHP, MSch, and YG revised the manuscript. YG, MSch, JHP, EJ, SK, MW, MA, CW, BM, AA, TD, FWG, MK, HWD, MO, RN, TS, MN, MSt, AW, ØK and IB collected, sexed and/or extracted and prepared gDNA samples. All authors approved the final manuscript.

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

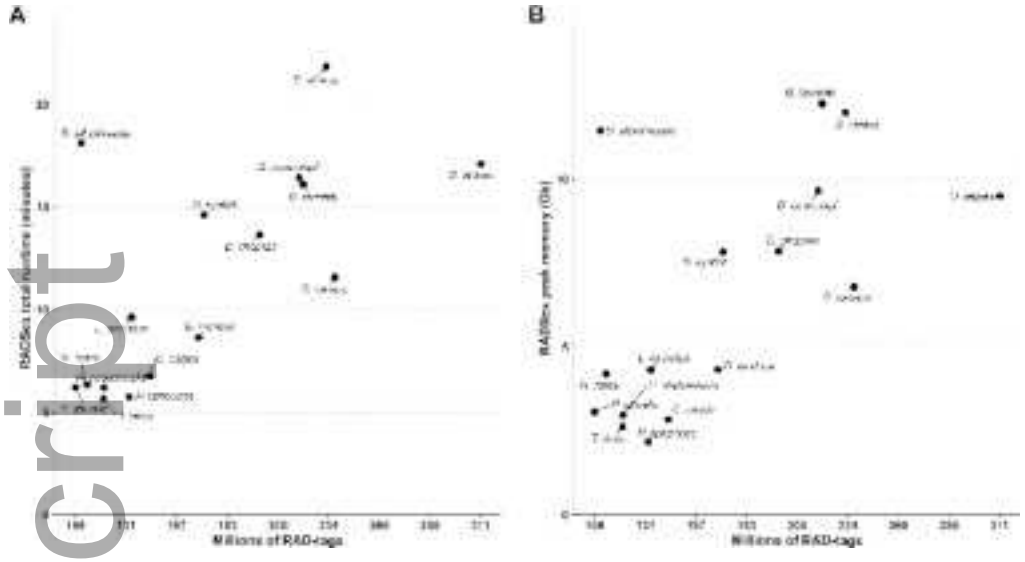
**Table 1: Summary of RADSex results for the 15 datasets analyzed.** Species where a sex-determination system was identified with *radsex* are highlighted with a grey background. The number of markers significantly associated with phenotypic sex is also given for each species, with the value of '*d*' indicating the minimal number of reads for a marker to be considered present in an individual in the analysis.



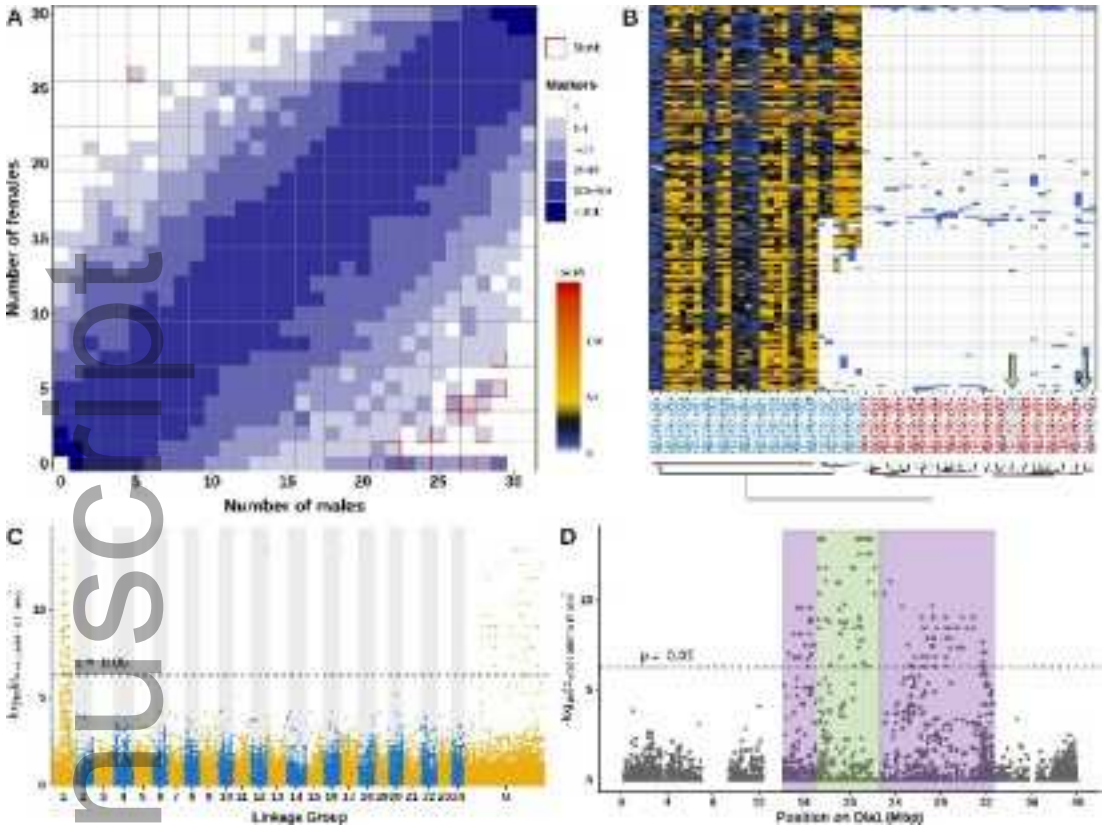

men_13360_f1.tif



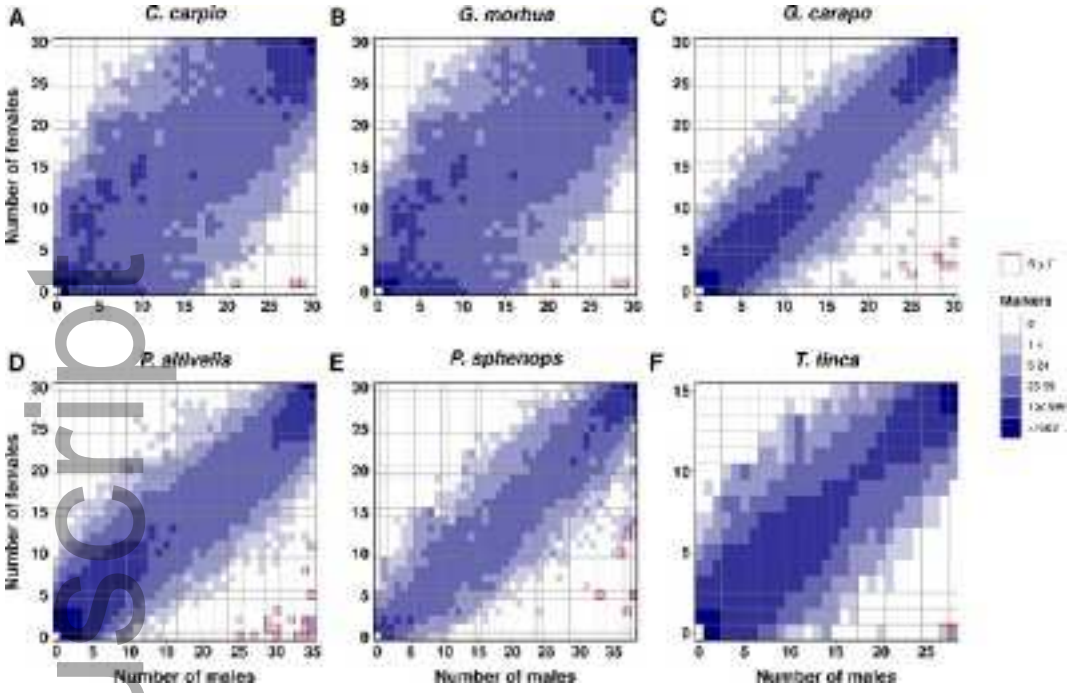

men_13360_f2.tif



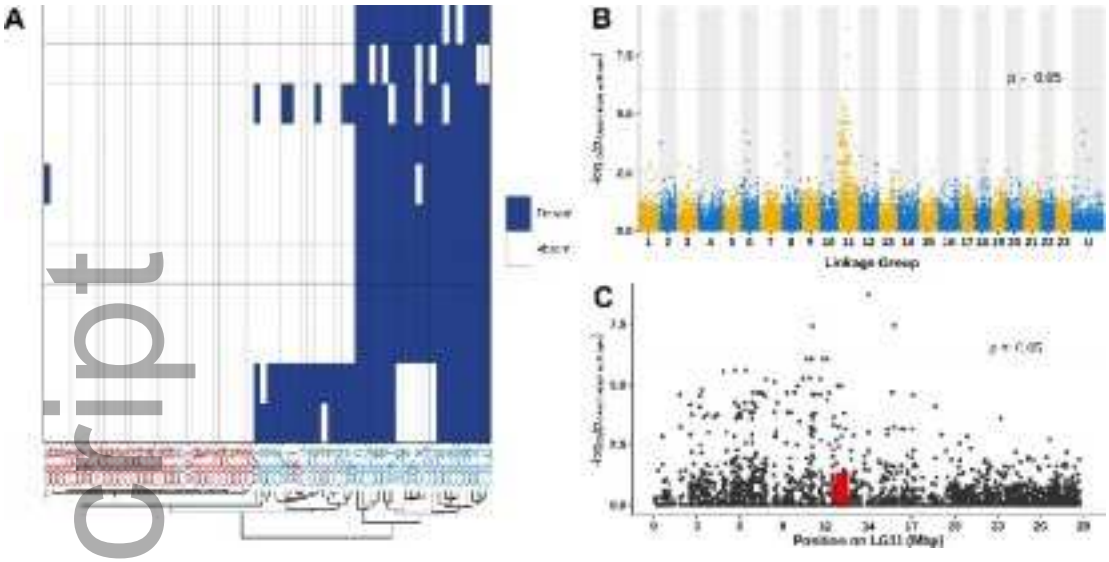

men_13360_f3.tif




men_13360_f4.tif




men_13360_f5.tif




men_13360_f6.tif