Conservation Genetics Resources
https://doi.org/10.1007/s12686-023-01331-5

METHODS AND RESOURCES

An R-based tool for identifying sex-linked markers from restriction site-associated DNA sequencing with applications to elasmobranch conservation

Floriaan Devloo-Delva[1,2,3] · Thierry Gosselin[4] · Paul A. Butcher[5] · Peter M. Grewe[1] · Charlie Huveneers[6] · Robin B. Thomson[1] · Jonathan M. Werry[7] · Pierre Feutry[1]

Received: 27 June 2022 / Accepted: 27 October 2023

Abstract
Identifying sex-linked markers from genomic data has both theoretical and applied importance, especially in conservation. Yet, few methods and tools exist to detect such markers from Restriction-site-Associated DNA sequencing reads and even fewer tools can identify sex-linked markers from existing genotyped data. Here, we describe a new R function that can identify sex-linked markers in species with partially non-recombining sex chromosomes. We test the accuracy and speed of our function with an example dataset from a species of conservation concern, the White Shark, *Carcharodon carcharias*. We further compare our method against other approaches and find that our method detects more sex-linked markers that can be reliably mapped to reference genomes. Overall, we provide a conservation and fisheries-relevant tool that can reliably and efficiently assign sex from genetic data in species with a heterogametic sex and we demonstrate its utility by developing a sex-identification PCR test for White Sharks.

Keywords RADseq · ddRAD · DArTseq · SNPs

Main text

Sex-linked markers (SLMs) are important in both theoretical and applied biological sciences, especially in conservation. For example, such markers can provide valuable insight into sex chromosome turnover rates (Charlesworth and Mank 2010; Kitano and Peichel 2012). Further, in the field of population genetics they allow the inference of sex-specific demographic events due to their sex-specific inheritance, for instance the comparison of mitochondrial and Y chromosome markers can reveal sex-specific migration rates (Petit et al. 2002; Wilson Sayres 2018). Lastly, sex ratio is a key component ecological and demographic studies, particularly when females and males differ in life-history traits (e.g. Tsai et al. 2014; Pillans et al. 2021). In species with a heterogametic sex, markers on sex chromosomes provide a way to identify sex from DNA samples when sexual dimorphism is absent and destructive sampling is not an option, e.g. for Threatened species (Stovall et al. 2018; Suda et al. 2019) or when sampling in no-take zones.

Despite the importance of SLMs, only a few methods and tools exist to identify them in non-model species. Most approaches are focussed on presence-absence and heterozygosity patterns of Restriction-site-Associated DNA sequencing (RADseq) data to differentiate the heterogametic (XY or ZW) from the homogametic (XX or ZZ) sex (Gamble and Zarkower 2014; Fowler and Buonaccorsi 2016; Gamble

✉ Floriaan Devloo-Delva
floriaan.devloo-delva@csiro.au

1 CSIRO Environment, Hobart 7000 TAS, Australia

2 CSIRO Australian National Fish Collection, Hobart 7000 TAS, Australia

3 School of Natural Sciences – Quantitative Marine Science, University of Tasmania, Hobart 7001 TAS, Australia

4 Independent Researcher, Montréal, Québec, Canada

5 NSW Department of Primary Industries, National Marine Science Centre, Coffs Harbour, NSW 2450, Australia

6 Southern Shark Ecology Group, College of Science and Engineering, Flinders University, Adelaide, SA 5042, Australia

7 Griffith Centre for Coastal Management, Griffith University, Gold Coast, QLD 4222, Australia

Published online: 16 November 2023

2016; Hill et al. 2018). Most of the workflows following this approach use demultiplexed FASTQ files and are based in *Stacks* or *RADtools* (see Gamble and Zarkower 2014; Fowler and Buonaccorsi 2016), which can be computationally intensive, or the faster *RADSex* software (based in C++; Feron et al. 2021). Currently, the only methods that can identify SLMs from genotyped single nucleotide polymorphism (SNP) data are outliers detection methods (e.g. *BayeScan*; Foll and Gaggiotti 2008) where the data is partitioned by sex (e.g. Benestan et al. 2017; Trenkel et al. 2020).

Many elasmobranchs (sharks and rays) are threatened with extinction (Dulvy et al. 2014, 2021). Their slow life history, low fitness and low connectivity between populations, which is often male biased (see Phillips et al. 2021), has instigated many conservation genomic studies in elasmobranchs (see Ovenden et al. 2018; Green et al. 2022; Devloo-Delva et al. 2023a). However, to date, no studies have used these available genomic resources to investigate the value of SLMs for sex identification and population genetics. In this study, we introduce a tool that can analyse the existing genomic data, such as RADseq, Diversity Arrays Technologies (DArTseq), and genotyping by sequencing (GBS) to look for differential signals between heterogametic and homogametic individuals and identify SLMs on the sex chromosomes.

Specifically, we have designed a function, 'sexy_markers', as part of the *radiator* package (Gosselin et al. 2020) in the R environment (R Core Team 2020) which tests three different scenarios to find markers on the sex chromosomes under the assumption that one sex is heterogametic (see Supplementary Material Sect. 1): (i) markers are only present in females or males, (ii) markers are homozygous in one sex while exhibiting an intermediate range of heterozygosity in the other sex, (iii) markers have double the read depth in females or males. Here, the first scenario identifies markers on the sex chromosome unique to the heterogametic sex and the latter two detect markers on the sex chromosomes shared by the heterogametic and homogametic sexes. This function also allows the re-assignment of genetic sex when markers on the heterogametic sex chromosome are identified. The function is based on a visual identification of SLMs from genotyped RAD-type data after minimal data-quality filtering. *Radiator* is developed to import bi-allelic SNP data from vcf and csv files from several genotype callers: *DArTsoft14*, *Stacks*, *GATK*, *platypus*, *samtools* and *ipyrad*.

We demonstrated the workflow and accuracy of this function using a White Shark example (*Carcharodon carcharias*, listed as Vulnerable to extinction; Rigby et al. 2019), with a DArTseq dataset of 558 individuals and 23,393 SNPs (Bruce et al. 2018; Hillary et al. 2018). We further compare our function to alternative approaches that use genotyped SNP data as input (i.e. fixed allele differences): 'gl.report.sexlinked' from *dartR* package (Gruber et al. 2018; Mijangos et al. 2022), *OutFLANK* (Whitlock and Lotterhos 2015), and *PCadapt* (Luu et al. 2017). These outlier methods identify markers with differences in allele frequencies between the sexes, i.e. at homologous regions between the sex chromosomes (Robledo-Ruiz et al. 2023). The SLMs were validated using polymerase chain reactions (PCR) with primers designed from SLMs that were mapped to the reference genomes from Marra et al. (2019) and the Vertebrate Genome Project (VGP; https://vgp.github.io/genomeark/; NCBI RefSeq accession GCF_017639515.1). Autosomal primers in the beta-actin gene were also designed from the reference genome to act as a positive control between sexes. Primer sequences and PCR conditions are described in the Supplementary Material (Sect. 4).

Overall, we found nine Y-linked and 406 X-linked markers using the 'sexy_markers' function in less than 5 min computation time. The nine heterogametic SLMs allowed us to assign sex to 43 individuals with unknown visual sex and showed a 6.7% phenotypic – genotypic sex discrepancy across the 402 sexed sharks that passed quality filtering. The latter is most likely explained by human error, although hermaphroditic elasmobranchs have been described sporadically (reviewed in Adolfi et al. 2019). Further, the outlier methods identified 131 and 2720 SLMs for *OutFLANK* and *PCadapt* respectively (10 markers in common), but only *PCadapt* had 16 markers in common with the 'sexy_markers' approach (see Supplementary Material Sect. 2). We were able to confidently blast 179 SLMs (seven Y-linked and 172 X-linked markers) to 49 scaffolds from the Marra et al. (2019) genome, of which 47 SLMs mapped to five scaffolds (i.e. putative sex scaffolds). Eight Y-linked and 215 X-linked markers had confident BLAST hits (see Supplementary Material Sect. 3) to eight scaffolds from the VGP genome, with the majority (199 SLMs) mapping to three scaffolds. Overall, we conclude that 48% of the 415 identified SLMs were located in close proximity on putative sex chromosome scaffolds. These markers were considered as a reference to test the accuracy of the 'sexy_markers' function with suboptimal data (Fig. 1). By randomly sampling 6, 12, 24, 48, 72, 96, 120, 144, 252 and 348 individuals for 200, 1000, 2000, 5000, 7500, 10000, 12500, 15000, 17500 and 20000 markers, we showed that too few individuals (< 100) and too few markers (< 10,000 or 50% of the total data) will identify false positive SLMs (i.e. not in common with the SLMs from the full data; Fig. 1A-B). This result was more pronounced when the female:male sex ratio was skewed (2:1 or 1:2; Fig. 1C-F). The Y- and X-linked markers were validated through multiplex PCR (Fig. 2; Supplementary Material Sect. 4). PCR results showed that males amplified for the Y-chromosome fragments, while X and beta actin fragments were present in both sexes. The same individuals

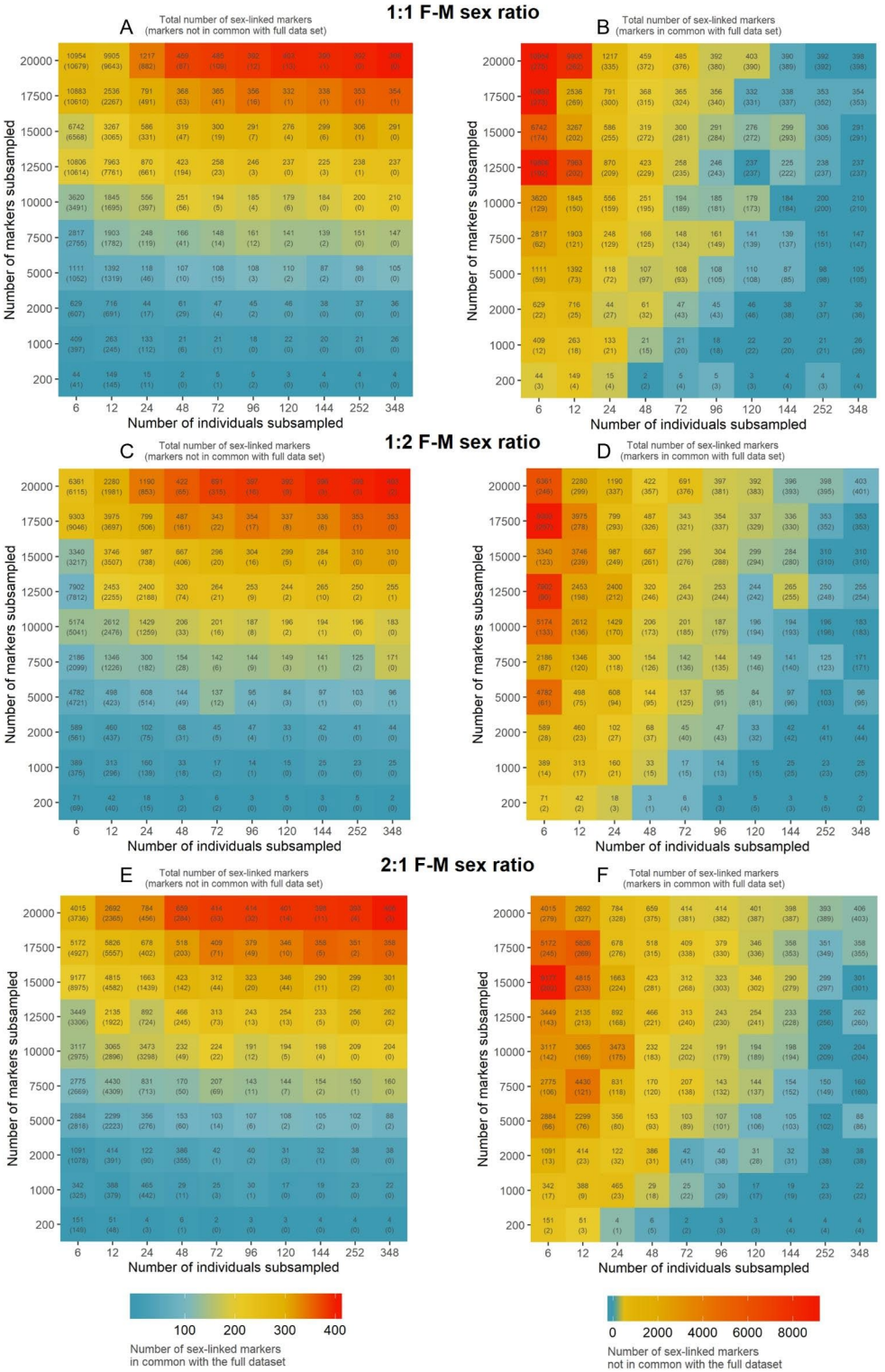

Fig. 1 Robustness test for the 'sexy_markers' R function on White Shark, *Carcharodon carcharias*, data. The heatmaps show data with varying numbers of markers and individuals, and a female:male (F-M) sex ratio of 1:1 **(A,B)**, 1:2 **(C,D)** or 2:1 **(E,F)**. Each square is coloured according to correctly identified sex-linked markers **(A,C,E)** or false positive detections **(B,D,F)**.

that had a phenotypic – genotypic sex mismatch based on the heterogametic SLMs (70 base pairs) also showed this discrepancy for the Y-chromosome fragment (655 base pairs; Fig. 2).

In general, these results confirm that the White Shark has partially non-recombining sex chromosomes (X and Y), males being the heterogametic sex. This is the first study to validate male heterogamety in the White Shark with sufficiently high sample size (n = 558); an observation also obtained using karyotyping (Maddock and Schwartz 1996), where the authors suggested the White Shark and several other elasmobranchs possess X and Y sex chromosomes,

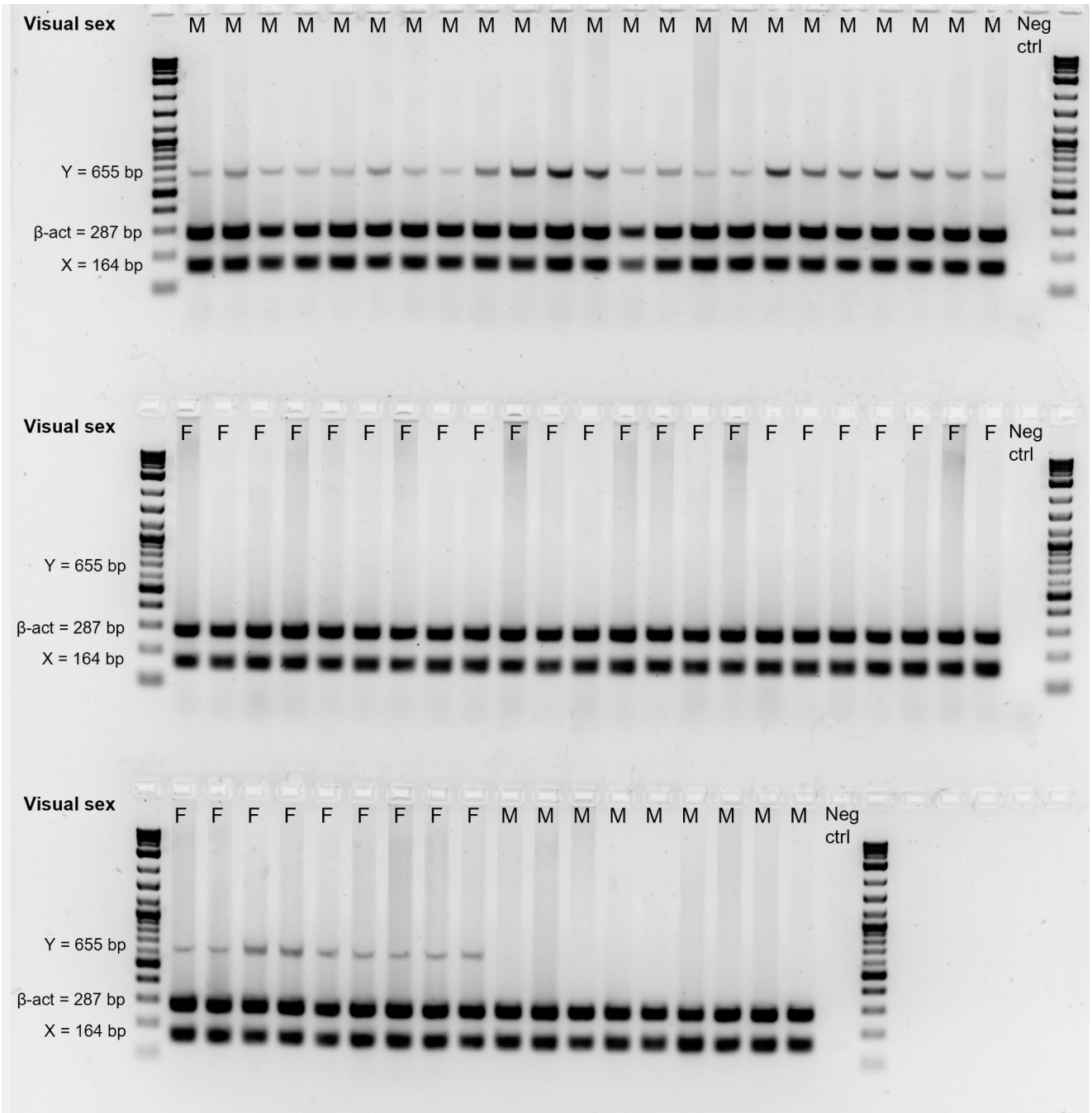

Fig. 2 PCR results of Y- and X-chromosome amplicons between female (F) and male (M) White Sharks, *Carcharodon carcharias*, showing 655 and 164 base pairs (bp) fragments, respectively. Conserved beta-actin amplicons (287 bp) serve as a positive control. PCR products were visualised on a 2% agarose gel, stained with SYBR safe (Invitrogen, USA) and a TriDye 1 kb DNA ladder (New England BioLabs, UK) for size reference. The first row shows males that were identified as males by both visual assignment and PCR assay. The second row shows females that were identified as females by both visual assignment and PCR assay. The last row shows the individuals that had a discrepancy between phenotypic and genotypic sex

albeit with low samples sizes (n = 1). Further, we showed the utility and robustness of the 'sexy_markers' function for species with distinct sex chromosomes, where species with larger non-recombining regions have a higher chance of finding Y/W-linked markers. Importantly, the function takes genotyped SNP data as input (whereas other software require demultiplexed FASTQ files), which allows a more versatile use of previously published datasets for comparative studies. Finally, we developed a quick (~ 2-hour) PCR assay to identify the sex of sampled White Sharks. This

tool will prove useful for sex identification in species that do not display obvious morphological differences between sexes. For instance, most juvenile sharks without developed external sex organs or samples obtained from processed carcasses (e.g. fisheries or fin trade) could be sexed using our method. Future studies include applying the R function on other species, as well as utilising the sex-linked markers for population genetic studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-023-01331-5.

Acknowledgements This work was supported by the Marine Biodiversity Hub a collaborative partnership supported through funding from the Australian Government's National Environmental Science Program. This research was funded in part through the Ord River Research Offset grant from CSIRO, secured by Richard Pillans. Floriaan Devloo-Delva was supported by a joint UTAS/CSIRO scholarship and the Quantitative Marine Science program. Thierry Gosselin was supported by the CSIRO Frohlich fellowship. We acknowledge the following people and agencies for their invaluable help obtaining a large portion of the samples: Barry Bruce and Russell Bradford from CSIRO, David Harasti and Christopher Gallen from the New South Wales Department of Primary Industry, Mike Travers from the Department of Primary Industries and Regional Development (Western Australia), Paul Rogers and Crystal Beckmann from the South Australian Research and Development Institute, Michael Drew from Flinders University, numerous fisheries compliance officers from the Department of Primary Industries and Regions (South Australia), Malcolm Francis from the National Institute of Water and Atmospheric Research (New Zealand), and Clinton Duffy from the Department of Conservation (New Zealand). We thank Barry Bruce and Russell Bradford for securing funding and organising all sample contributions. We are grateful to Rasanthi Gunasekera for the laboratory processing of all tissue samples. We also thank Mark Bravington for the fruitful discussions regarding the sex marker algorithm and Toby Patterson, Clinton Duffy, Russell Bradford, and anonymous reviewer for insightful comments on the manuscript.

Author contributions FD, TG, PMG, RBT, and PF designed the study. Samples were provided by PAB, CH, and JMW. PMG and PF processed the samples and created the SNP data. FD, TG and RBT developed the R function. FD analysed the data and tested the robustness of the function. FD and PMG developed the PCR assay. FD prepared the manuscript with feedback from all co-authors. All authors agree to be accountable for all aspects of the work.

Funding Open access funding provided by CSIRO Library Services.

References

Adolfi MC, Nakajima RT, Nóbrega RH, Schartl M (2019) Intersex, hermaphroditism, and gonadal plasticity in vertebrates: evolution of the Müllerian duct and Amh/Amhr2 signaling. Annu Rev Anim Biosci 7:149–172

Benestan L, Moore JS, Sutherland BJG, Le Luyer J, Maaroufi H, Rougeux C, Normandeau E, Rycroft N, Atema J, Harris LN, Tallman RF, Greenwood SJ, Clark FK, Bernatchez L (2017) Sex matters in massive parallel sequencing: evidence for biases in genetic parameter estimation and investigation of sex determination systems. Mol Ecol 26:6767–6783

Bruce B, Bradford R, Bravington MW, Feutry P, Grewe PM, Gunasekera RM, Harasti D, Hillary R, Patterson T (2018) A national assessment of the status of white sharks. ed. M.B.H. National Environmental Science Programme. 64. CSIRO, Hobart, Australia

Charlesworth D, Mank JE (2010) The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics 186:9–31

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available online at https://www.r-project.org/. *Vienna, Austria*. Vienna, Austria

Devloo-Delva F, Burridge CP, Kyne PM, Brunnschweiler JM, Chapman DD, Charvet P, Chen X, Cliff G, Daly R, Drymon JM, Espinoza M, Fernando D, Garcia Barcia L, Glaus K, González Garza BI, Grant MI, Gunasekera RM, Hernandez SI, Hyodo S, Jabado RW, Jaquemet S, Johnson G, Ketchum JT, Magalon H, Marthick JR, Mollen FH, Mona S, Naylor GJP, Nevill JEG, Phillips NM, Pillans RD, Postaire BD, Smoothey AF, Tachihara K, Tillet BJ, Valerio-Vargas JA, Feutry P (2023a) From rivers to ocean basins: the role of ocean barriers and philopatry in the genetic structuring of a cosmopolitan coastal predator. Ecol Evol 13:e9837

Devloo-Delva F, Gosselin T, Butcher PA, Grewe PM, Huverneers C, Thomson RB, Werry JM, Feutry P (2023b) An R-based tool for identifying sex-linked markers from restriction site-Associated DNA sequencing with applications to elasmobranch conservation. v1. CSIRO. https://doi.org/10.25919/c9mm-8960. Data Collection

Dulvy NK, Fowler SL, Musick JA, Cavanagh RD, Kyne PM, Harrison LR, Carlson JK, Davidson LNK, Sonja V (2014) Extinction risk and conservation of the world's sharks and rays. eLife 3:e00590

Dulvy NK, Pacoureau N, Rigby CL, Pollom RA, Jabado RW, Ebert DA, Finucci B, Pollock CM, Cheok J, Derrick DH, Herman KB, Sherman CS, VanderWright WJ, Lawson JM, Walls RHL, Carlson JK, Charvet P, Bineesh KK, Fernando D, Ralph GM, Matsushiba JH, Hilton-Taylor C, Fordham SV, Simpfendorfer CA (2021) Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. Curr Biol, 1–24

Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, Herpin A, Journot L, Parrinello H, Klopp C, Kottler VA, Roco AS, Du K, Kneitz S, Adolfi M, Wilson CA, McCluskey B, Amores A, Desvignes T, Goetz FW, Takanashi A, Kawaguchi M, Detrich HW, Oliveira M, Nobrega R, Sakamoto T, Nakamoto M, Wargelius A, Karlsen Ø, Wang Z, Stöck M, Waterhouse RM, Braasch I, Postlethwait JH, Schartl M, Guiguen Y (2021) RADSex: a computational workflow to study sex determination using restriction site-associated DNA sequencing data. Mol Ecol Resour 21:1715–1731

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics 180:977–993

Fowler BLS, Buonaccorsi VP (2016) Genomic characterization of sex-identification markers in Sebastes carnatus and Sebastes chrysomelas rockfishes. Mol Ecol 25:2165–2175

Gamble T (2016) Using RAD-seq to recognize sex-specific markers and sex chromosome systems. Mol Ecol 25:2114–2116

Gamble T, Zarkower D (2014) Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Mol Ecol Resour 14:902–913

Gosselin T, Lamothe M, Devloo-Delva F, Grewe P (2020) RADseq data exploration, manipulation and visualization using R

Green ME, Simpfendorfer CA, Devloo-Delva F (2022) Population structure and connectivity of chondrichthyans. In: Carrier JC, Simpfendorfer CA, Heithaus MR, Yopak KE (eds) Biology of sharks and their relatives. CRC Press, Boca Raton, FL, USA, pp 523–544

Gruber B, Unmack PJ, Berry OF, Georges A (2018) Dartr: an r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 18:691–699

Hill PL, Burridge CP, Ezaz T, Wapstra E (2018) Conservation of sex-linked markers among conspecific populations of a viviparous skink, Niveoscincus Ocellatus, exhibiting genetic and temperature-dependent sex determination. Genome Biol Evol 10:1079–1087

Hillary RM, Bravington MV, Patterson TA, Grewe P, Bradford R, Feutry P, Gunasekera R, Peddemors V, Werry J, Francis MP, Duffy CAJ, Bruce BD (2018) Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. Sci Rep 8:2261

Kitano J, Peichel CL (2012) Turnover of sex chromosomes and speciation in fishes. Environ Biol Fish 94:549–558

Luu K, Bazin E, Blum MG (2017) Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour 17:67–77

Maddock MB, Schwartz FJ (1996) Elasmobranch cytogenetics: methods and sex chromosomes. Bull Mar Sci 58:147–155

Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Pavinski Bitar P, Richards VP, Komissarov A, Rayko M, Kliver S, Stanhope BJ, Winkler C, O'Brien SJ, Antunes A, Jorgensen S, Shivji MS (2019) White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proceedings of the National Academy of Sciences*, 116, 4446–4455

Mijangos JL, Gruber B, Berry O, Pacioni C, Georges A (2022) dartR v2: an accessible genetic analysis platform for conservation, ecology and agriculture. Methods Ecol Evol 13:2150–2158

Ovenden JR, Dudgeon C, Feutry P, Feldheim K, Maes GE (2018) Genetics and genomics for fundamental and applied research on elasmobranchs. In: Carrier JC, Heithaus MR, Simpfendorfer CA (eds) Shark Research: Emerging Technologies and Applications for the Field and Laboratory. CRC Press, Boca Raton, USA, pp 235–254

Petit E, Balloux F, Excoffier L (2002) Mammalian population genetics: why not Y? Trends Ecol Evol 17:28–33

Phillips NM, Devloo-Delva F, McCall C, Daly-Engel T (2021) Reviewing the genetic evidence for sex-biased dispersal in elasmobranchs. Rev Fish Biol Fish 31:821–841

Pillans R, Rochester WA, Babcock RC, Thomson DP, Haywood MD, Vanderklift M (2021) Long-term acoustic monitoring reveals site fidelity, reproductive migrations and sex specific differences in habitat use and migratory timing in a large coastal shark (*Negaprion acutidens*). Front Mar Sci 8:616633

Rigby CL, Barreto R, Carlson J, Fernando D, Fordham S, Francis MP, Herman K, Jabado RW, Liu KM, Lowe CG, Marshall A, Pacoureau N, Romanov E, Sherley RB, Winker H (2019) *Carcharodon carcharias*. https://doi.org/10.2305/IUCN.UK.2019-3.RLTS.T3855A2878674.en. Downloaded on 29 May 2020. The IUCN Red List of Threatened Species 2019. e.T3855A2878674

Robledo-Ruiz DA, Austin L, Amos JN, Castrejón-Figueroa J, Harley DKP, Magrath MJL, Sunnucks P, Pavlova A (2023) Easy-to-use R functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex assignment. Mol Ecol Resour 00:1–21

Stovall WR, Taylor HR, Black M, Grosser S, Rutherford K, Gemmell NJ (2018) Genetic sex assignment in wild populations using genotyping-by-sequencing data: a statistical threshold approach. Mol Ecol Resour 18:179–190

Suda A, Nishiki I, Iwasaki Y, Matsuura A, Akita T, Suzuki N, Fujiwara A (2019) Improvement of the Pacific bluefin tuna (*Thunnus orientalis*) reference genome and development of male-specific DNA markers. Sci Rep 9:1–12

Trenkel VM, Boudry P, Verrez-Bagnis V, Lorance P (2020) Methods for identifying and interpreting sex-linked SNP markers and carrying out sex assignment: application to thornback ray (Raja clavata). Mol Ecol Resour 20:1610–1619

Tsai W-P, Sun C-L, Punt AE, Liu K-M (2014) Demographic analysis of the shortfin mako shark, *Isurus oxyrinchus*, in the Northwest Pacific using a two-sex stage-based matrix model. ICES J Mar Sci 71:1604–1618

Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of Fst. Am Nat 186:S24–S36

Wilson Sayres MA (2018) Genetic diversity on the sex chromosomes. Genome Biol Evol 10:1064–1078