A rapid approach for sex assignment by RAD-seq using a reference genome

GBS approach for sexing

Diego M. Peralta[1,2*], Juan I. Túnez[1,3], Ulises E. Rodríguez Cruz[4], Santiago G. Ceballos[5,6]

[1]Grupo de Investigación en Ecología Molecular, Instituto de Ecología y Desarrollo Sustentable (INEDES-CONICET-UNLu-CIC), Luján, Argentina

[2]Departamento de Ecología de la Diversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México

[3]Departamento de Ciencias Básicas, Universidad Nacional de Luján, Luján, Argentina

[4]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México

[5]Instituto de Ciencias Polares, Ambiente y Recursos Naturales, Universidad Nacional de Tierra del Fuego, Ushuaia, Argentina

[6]Centro Austral de Investigaciones Científicas (CADIC-CONICET), Ushuaia, Argentina

*Corresponding author

E-mail: peraltadm@ecologia.unam.mx

Abstract

Sex identification is a common objective in molecular ecology. While many vertebrates display sexual dimorphism, determining the sex can be challenging in certain situations, such as species lacking clear sex-related phenotypic characteristics or in studies using non-invasive methods. In these cases, DNA analyses serve as valuable tools not only for sex determination but also for validating sex assignment based on phenotypic traits. In this study, we developed a bioinformatic framework for sex assignment using genomic data obtained through GBS, and having an available closely related genome assembled at the chromosome level. Our method consists of two *ad hoc* indexes that rely on the different properties of the mammalian heteromorphic sex chromosomes. For this purpose, we mapped RAD-seq loci to a reference genome and then obtained missingness and coverage depth values for the autosomes and X and Y chromosomes of each individual. Our methodology successfully determined the sex of 165 fur seals that had been phenotypically sexed in a previous study and 40 sea lions sampled in a non-invasive way. Additionally, we evaluated the accuracy of each index in sequences with varying average coverage depths, with Index Y proving greater reliability and robustness in assigning sex to individuals with low-depth coverage. We believe that the approach presented here can be extended to any animal taxa with known heteromorphic XY/ZW sex chromosome systems and that it can tolerate various qualities of GBS sequencing data.

Keywords

Population genomics, Conservation genetics, Sex identification, Genomic resource, Genotyping-by-sequencing

Introduction

Data on the genetic diversity, population size, sex ratio, kinship and distribution of free-ranging wildlife are essential for establishing effective conservation strategies [1]. In particular, information on the sex of individuals is of great importance in population ecology and conservation [2,3]. For example, it contributes to the understanding of population structure and dynamics in monitoring programs [4] or allows to identify and infer socio-ecological patterns in behavioral studies of living and extinct species [5,6], respectively. Knowledge of the sex ratio of effective individuals is particularly relevant for species or populations of conservation concern, as unbalanced sex ratios may have negative impacts on population growth and resilience [7]. However, sex identification is hampered in non-sexually dimorphic species (especially during the juvenile period) and in elusive ones due to the difficulty of direct observation [3,8]. As a result, the determination of sex through DNA analysis applied to biological samples (mainly those collected in a non-invasive way) has become a useful tool in molecular ecology and conservation genetics [9].

In mammals, molecular sexing has traditionally relied on the use of the polymerase chain reaction (PCR) for amplification of specific fragments on the Y chromosome or amplification of homologous fragments from both X and Y chromosomes [9–13]. Nevertheless, these methods depend on amplification reliability and external controls [14,15]. Other commonly used approaches are based on differences in size polymorphisms among X and Y homologous fragments such as Amelogenin gene size [16], or on the existence of unique restriction sites in a Y fragment such as zinc finger genes [17–19]. However, these methods do not apply to all taxa [15].

Restriction Associated DNA sequencing (RAD-seq) has emerged as a popular method in the field of population genetics because it overcomes the limitations of traditional techniques [20] and allows

73 working with wild populations and non-model species, among other advantages. Since the advent
74 of RAD-seq, several studies on sex assignment and sex chromosome systems, among others, have
75 benefited from this technique. It improved the understanding of sex chromosomes systems,
76 revealed the influence of sex-linked markers in population genetic studies, or gave tools to assign
77 sex to samples from different sources with successful results [21,22]. For example, Carmichael *et al*.
78 [23] identified a sex-linked SNP marker in a salmon's parasite thus confirming the presence of a
79 genetic mechanism for sex determination and pointed out that this information could be useful in
80 the development of control strategies. Benestan *et al*. [24] stated that an unbalanced sex ratio in
81 the samples and the presence of sex-linked markers (not removed from the dataset) may lead to a
82 potential bias in the correct interpretation of population structure. In the context of individual sex
83 identification, Stovall *et al*. [25] successfully assigned the sex to 86 New Zealand fur seals
84 (*Arctocephalus forsteri*) through the discovery and statistical validation of sex-specific loci using a
85 de novo RAD-seq approach. Their findings demonstrated that monomorphic loci (frequently ignored
86 in the SNP datasets) are relevant when identifying sex-linked markers. Finally, Trenkel *et al*. [26]
87 applied three sex-determining methods on 1680 rays of unknown sex-determination system and
88 offered recommendations based on the error of each method. The main difference between the
89 former two works and the third one is that it uses a reference genome.
90
91 In this context, the growing popularity of methods like RAD-seq [27] takes place simultaneously
92 with the positive trend in generation of new reference genomes. In fact, the availability of genomes
93 assembled at chromosome-level has shown a remarkable ten-fold increase from 2015 to the present
94 day, mainly due to the coordinated and standardized efforts of international genome initiatives [28].
95 Taking mammals and birds as an example, there are 655 chromosome-level genome assemblies (of
96 which 32% contain sex chromosomes) from 379 species that are currently available at NCBI

97 (https://www.ncbi.nlm.nih.gov/datasets/genome/. Last accession 08-24-2023). Although there is
98 overrepresentation within the clades of these groups (i.e., primates, artiodactyls, rodents and
99 carnivores among mammals; and passeriformes, gallinaceous poultry and waterfowl among birds),
100 this drawback is likely to be overcome by the continuous addition of new taxa.

101 In addition to the above, the number of genome assemblies containing sex chromosomes is
102 continuously increasing [29], allowing for more comprehensive analyses of population and
103 functional genomics. In particular, this contributed to an enhanced understanding of sex
104 determination systems and their significant variability. In the animal kingdom, these are typically
105 exemplified by the X/Y and Z/W configurations, in which one of the chromosomal pairs may
106 potentially vary in size and carries sex-determining factors. The fundamental distinction between
107 XY/ZW systems lies in the identity of the heterogametic sex (i.e., X/Z for males and Z/W for females)
108 [29], even though various deviations have been reported. For instance, monomorphic forms of
109 XY/ZW with chromosomal pairs being similar in shape and genetic content, as observed in some
110 lizards [30]; a reduced number of sex chromosomes as observed in two subspecies of the Japanese
111 rodent *Tokudaia osimensis* due to the lack of the Y chromosome [31]; or the presence of an
112 additional X chromosome, as reported for the African pygmy mouse *Mus minutoides* [32].
113 Additionally, the spectrum encompasses a broad range of aberrations such as aneuploidies and neo-
114 sex chromosomes, among others [29,33].

115

116 Considering the accuracy of RAD-seq and the increasing availability of reference genome assemblies
117 including sex chromosomes in many species, here we propose a simple methodology for sexing
118 individuals based on the different properties of the mammalian sex chromosomes X and Y and the
119 possibility of mapping RAD-seq loci to a reference genome. Thus, we developed two indexes aiming
120 at identifying the sex of animals with heteromorphic XY/ZW sex chromosomes, based on a closely

121 related genome assembled at the chromosome level. The combination of both indexes allowed us
122 to accurately assign sex to the fur seal dataset from Stovall *et al*. [25] and to the sea lion dataset
123 from Peralta *et al*. [34].
124

125 Methods

126 Samples

127 In this study we used GBS (Genotyping-by-sequencing) data from 165 live fur seals
128 (*Arctocephalus forsteri*) whose sex had been determined phenotypically (93 females and 72 males).
129 These were obtained from a GBS dataset of 255 dead and live individuals [25] available at NCBI
130 Sequence Read Archive (BioProject Accession no.: PRJNA419445). Then, we applied the method to
131 the RAD-seq dataset of 40 sea lions (*Otaria flavescens*) of unknown sex from Peralta *et al*. [34]
132 available at NCBI Sequence Read Archive (BioProject Accession no.: PRJNA1022835). Samples were
133 collected by either invasive or non-invasive methods from individuals along the coast of Argentina
134 [35].
135

136 RAD sequencing and filtering

137 The RAD-seq library from Peralta *et al*. [34] was constructed according to the protocol described in
138 Roesti *et al*. [36,37], adopted from Hohenlohe *et al*. [38]. The restriction digestion was made using
139 only the Sbf1 enzyme, followed by the addition of a specific barcode to the restricted DNA. The
140 library was single-end sequenced to 150 bp in a single lane on an Illumina HiSeq 4000 at the
141 Genomics and Cell Characterization Core Facility, University of Oregon, USA.
142 Individual reads of both data sets were mapped to the California sea lion (*Zalophus californianus*)
143 representative genome mZalCal1.pri.v2 (GCA_009762305.2) using BWA MEM v0.7.17 [39] with

144 default parameters. Unmapped reads were removed from the SAM alignment files using SAMtools
145 v1.7 [40]. For detailed methodology see Peralta *et al*. [34]. Regarding the selection of the
146 representative genome, it is worth mentioning that the Pinnipedia superfamily is characterized by
147 a pronounced karyotypic uniformity, particularly within Otariidae, having 2n = 36 and closely
148 agreeing karyotypes [41]. This allowed us to use the *Z. californianus* genome as representative of
149 both species.

150 The filtering steps (gstacks, STACKS v2.64) [42,43] were aimed to retain as much sex-linked
151 markers as possible. Therefore, considering that Y-linked loci are present only in males, we set the
152 minimum proportion of individuals across populations required to process a locus (-R) to 0.3, which
153 assumes a minimum of 30% for males in our dataset. We also filtered loci with a maximum observed
154 heterozygosity greater than 0.7 (--max-obs-het), and used the flag to create a vcf file containing all
155 sites (variable and fixed) within the RAD loci (--vcf-all).

156

157 Sex identification

158 Sex identification of individuals was carried out using the chromosome-level assembly of the
159 California sea lion genome as reference. Since male mammals carry a single copy of the X
160 chromosome, they are expected to have about 50% of coverage depth for loci mapping to the X
161 chromosomes compared to an autosomal chromosome, while in females this value should be similar
162 when compare the X and any autosome. In regard to the Y chromosome, females are expected to
163 have a high value of missingness (defined as the percentage of total number of loci found in the
164 population that are missing in an individual), while in males it should be similar to the missingness
165 of any other chromosome. Theoretically, however, it should be more similar to that of the X
166 chromosome because the heterogametic sex has only a single copy of each sex chromosome. Taking
167 all this into account we constructed two different indexes for sex identification. We defined Index X

as the average coverage depth of the X chromosome divided by the average coverage depth of the autosomes. On the other hand, to create Index Y, we first defined the individual completeness as one minus missingness and used the completeness values to get an index with expected values ranging between 0 and 1 (See Table 1). Index Y could be interpreted as the expected missingness for the Y chromosome in relation to the overall missingness in that individual.

To calculate missingness and coverage depth we used different options of VCFtools [44] over the output files of STACKS “populations.all.vcf“ [42,43]. We filtered RAD-loci in the Y (NC_045613.1) and X (NC_045612.1) chromosomes, and all the autosomes to obtain the missingness per individual with the option “--missing-indv” and the average coverage depth with the option “--depth” in VCFtools. We used all the positions in the RAD loci because the indexes do not depend on the presence or absence of SNPs in the loci but rather on the presence or absence of the loci themselves. A summary of the step-by-step process is depicted in the workflow in Fig 1, while the detailed procedure is available in the GitHub repository at https://github.com/Dieggarp/sexing (S1 File). Next, we calculated both indexes with the equations given in Table 1 using R [45]. This last step was executed using an automated R script, “sexing.R” (S1 File), also included in the GitHub repository. This script creates two files, first "final_sexing.csv", which contains the assignment of sex for each individual and second "sexing_plots.pdf", which contains three dispersion plots (Index X vs Y, Index X vs coverage depth, Index Y vs coverage depth) to visualize the results. We only used values of Index Y for assigning sex to each individual because Index X may lead to sex misidentification (see Discussion).

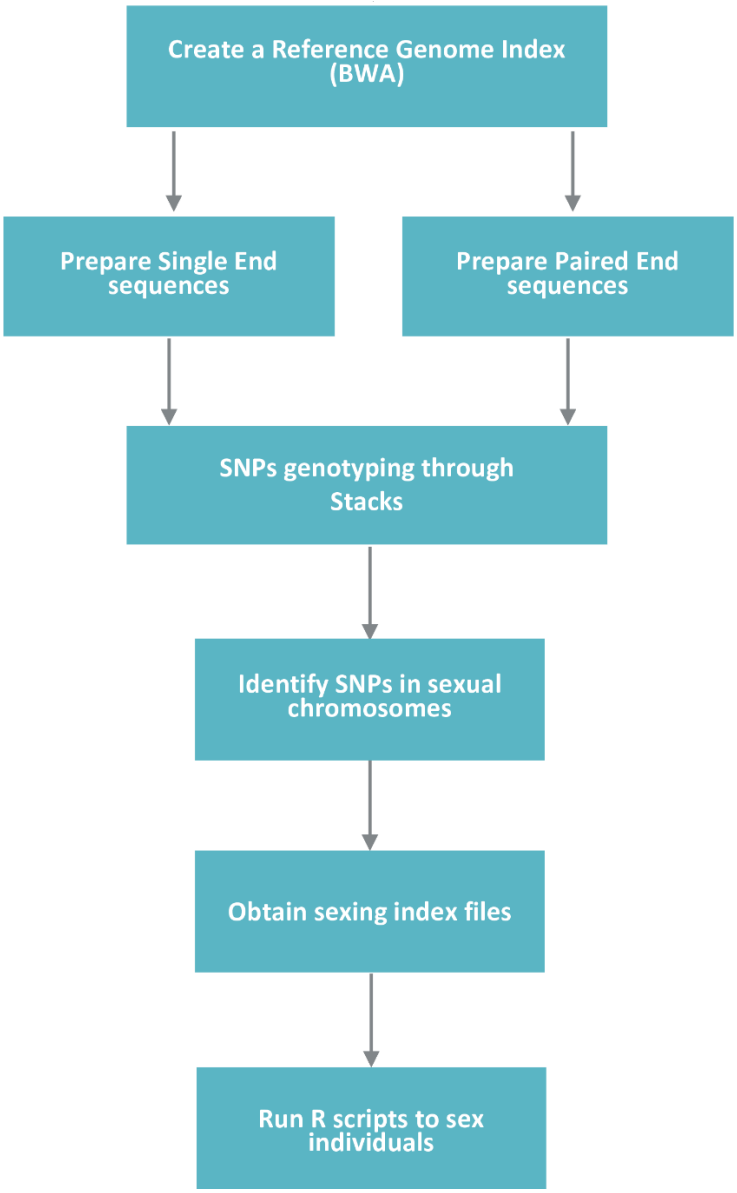

Fig 1. Sexing workflow. Illustration of a step-by-step summary of the entire sexing procedure, starting with the alignment to a reference genome and ending with the creation of two files that indicate the sex assigned to each individual.

Table 1. Index X and Y calculation.

		Expected Value	
	Equation	Female	Male

Index X	$= \frac{Depth_X}{Depth_A}$	1	0.5
Index Y	$= \frac{Com_X - Com_y}{Com_X}$	1	0

Indexes X and Y for sex identification. Index X is computed by dividing the average coverage depth of the X chromosome by the average coverage depth of the autosomes, expecting to find equal values of average coverage depth for the X chromosome in females, obtaining an Index X = 1, and half in males, Index X = 0.5. Index Y is computed by taking the difference in completeness between the X and Y chromosomes and dividing it by the completeness of the X chromosome. As a result, females would exhibit a complete difference in completeness between the sex chromosomes, obtaining an Index Y = 1, given the absence of the Y chromosome (Y missingness = 1). In contrast, males would exhibit no difference in completeness between these chromosomes, resulting in Index Y = 0 due to the presence of both sex chromosomes.

DepthX and DepthA: average coverage depth across loci for the X chromosome and autosomes, respectively. ComX and ComY: completeness for the X and Y chromosomes, respectively. Com = one minus missingness, where missingness is the proportion of total number of loci in the population that are absent in that individual.

Results

RAD loci filtering

After quality and coverage filtering the *A. forsteri* dataset, we obtained an average of 2.9 million reads per sample. A total of 560,778 widely shared loci were recovered with an average coverage depth of 11.5 X ± 7.1. On average, 31,775.1±8,096.47 loci were found on autosomal chromosomes, 19,574 on the X chromosome and 551 on the Y chromosome. The former filtering with population parameters using "-r" of 0.3 resulted in 180,343 shared loci (present in at least 30% of samples), containing 16,690,597 sites (fixed and variable) used for subsequent analysis.

220 For the *O. flavescens* dataset we obtained an average of 3.2 million reads per sample. A total of
221 818,244 widely shared loci were recovered with an average coverage depth of 13.8 X ± 13.1. On
222 average, 46,414.8±12,839.35 loci were found on autosomal chromosomes, 26,457 on the X
223 chromosome and 1,169 on the Y chromosome. The former filtering with population parameters
224 applying "-r" of 0.3 resulted in 127,983 shared loci (present in at least 30% of samples) containing
225 18,438,603 sites (fixed and variable), which were used for subsequent analysis.
226

227 ## Genomic sexing

228 Of the 165 fur seals from Stovall *et al*. [25], we successfully sexed 164 (99.4%) individuals with
229 both indexes and one with Index Y only, resulting in 95 females and 70 males (vs. 93 females and 72
230 males that had been phenotypically identified) (S1 Table). Interestingly, our two indexes sexed as
231 females two individuals phenotypically classified as males [25]. Thus, there was 98.8% of
232 coincidence between our method and sex assignment by phenotypic traits (but see discussion),
233 100% for females and 97.2% for males.
234 The dispersion plot for both indexes (Fig 2) showed two tight clouds of points in accordance with
235 what was expected for males and females. On the one hand, tightly clustered points at the
236 intersection of value 1 for both indexes were inferred to belong to females. On the other hand,
237 points located almost entirely near the intersection of value 0.65 for Index X and around 0 for Index
238 Y were inferred to belong to males. The value of the outlier point lying within the lower cloud of
239 points was lower than the others (-1.6) for Index Y and close to 1 for Index X. It is worth noting that
240 this individual also showed the lowest average coverage depth (1.4X), allowing us to test how this
241 value affected both indexes. We found that low values of average coverage depth led to a
242 remarkable decline in the performance of Index X (i.e., an individual of low average coverage depth
243 was misidentified) but not of Index Y (S1 Fig). It is also interesting to remark that in male individuals,

samples with very low average coverage depth can even give negative values for Index Y, thus improving sex identification.

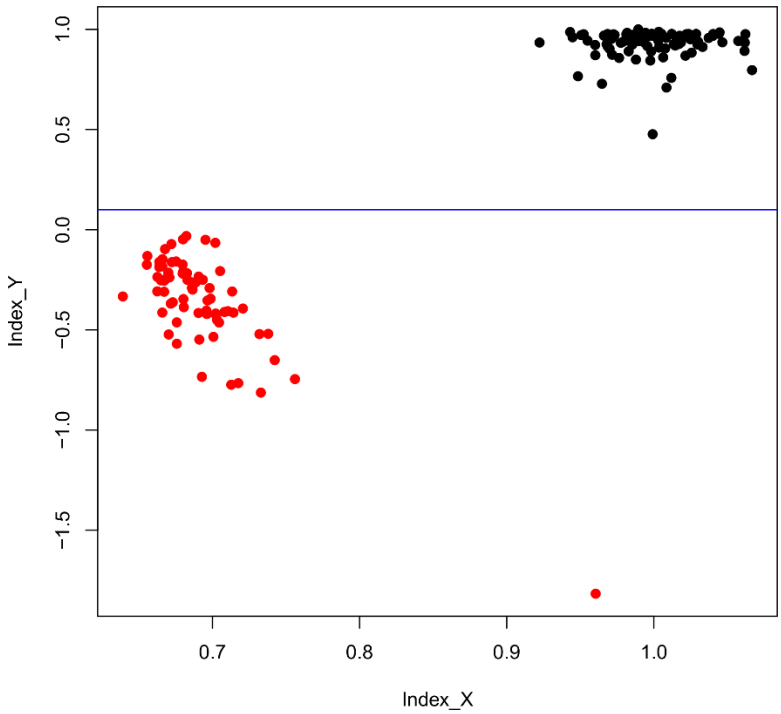

Figure 2. Dispersion plot of Index Y and X for fur seals. Dispersion plot of both indexes based on the dataset from Stovall *et al.* [25]. Red points: males; black points: females. The blue line indicates that Index Y achieves a complete separation between sexes.

Of the 40 individuals analyzed from the South American sea lions, 32 were successfully sexed using both indexes, resulting in 19 males and 13 females. The remaining individuals could only be sexed by Index Y, resulting in two females and six males. (S2 Table).

The dispersion plot for the X and Y indexes (Fig 3) showed two clouds of points representing males and females. The cloud of points corresponding to females was clustered at the intersection of value 1 for both X and Y, while that of males was near the intersection of values 0.5 for Index X and 0 for Index Y. However, some values were above 0.5 for Index X and below 0 for Index Y. Once more we

evaluated the effect of coverage depth on our indexes, finding that Index X was the only one affected (S2 Fig) and that again male samples with very low average coverage depth (<10X) yielded negative values for Index Y, providing a better separation of sexes.

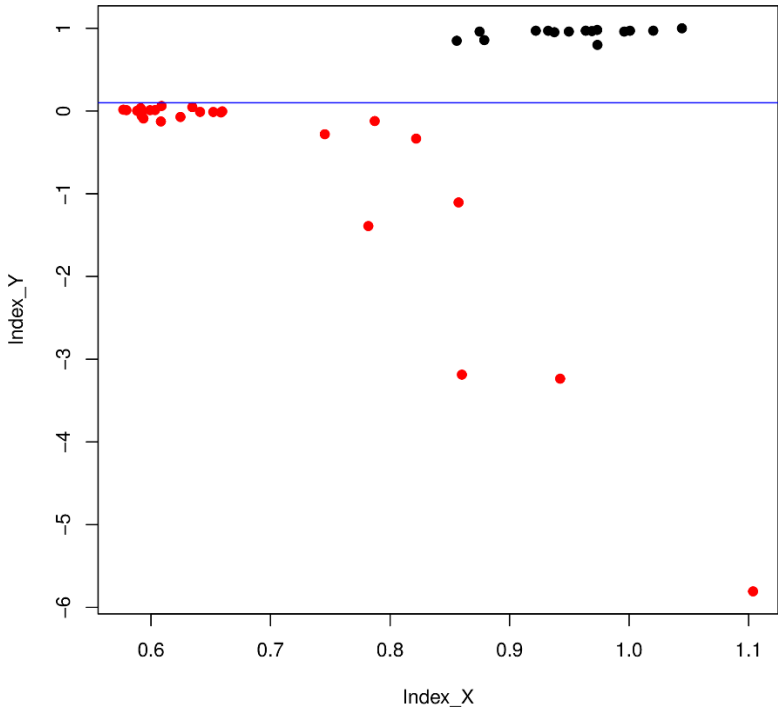

Figure 3. Dispersion plot of Index Y and X for sea lions. Dispersion plot of both indexes based on the dataset from Peralta *et al*. [34]. Red dots: males; black points: females. The blue line indicates that Index Y achieves a complete separation between sexes.

Discussion

Sexing animals from a RADseq approach

Here we provide a simple framework for sexing different types of samples using two *ad hoc* indexes. Our method will allow researchers working with RAD-seq or other similar GBS methodology and having an available reference genome, to sex individuals in an easy and reliable way.

273 After applying our indexes to the individuals in the Stovall *et al*. [25] dataset, we obtained 95
274 females and 70 males. The dispersion plot constructed with the indexes (Fig 2) showed that Index Y
275 located females near the value 1 and males below 0. Index X grouped almost all individuals into two
276 different clouds of points along the abscissa, except for one individual that could be differentiated
277 only by the Index Y. It is important to highlight that our indexes identified as females two individuals
278 previously labeled as males. In fact, Stovall *et al*. [25] reported that three individuals phenotypically
279 labeled as males lacked male-specific genetic markers at the time of genotypic identification,
280 thereby assuming that their sex had been misidentified. These cases of misidentifications or
281 mislabeled data were further confirmed by one of the authors (Neil J. Gemmell, pers. comm.).
282
283 Our sea lion dataset yielded 25 males and 15 females, but eight of them could only be identified
284 with Index Y because Index X could not discriminate between sexes. Certainly, the dispersion plot
285 (Fig 3) showed that Index Y proved to be successful for sex determination while Index X failed in
286 several cases. These results support the notion that Index Y is more reliable and robust than Index
287 X.
288 The samples that could only be sexed by Index Y shared low levels of average coverage depth,
289 with values less than 5X in some cases. This may be linked to the origin of the samples and the extent
290 of DNA degradation, as indicated in Supplementary Table 2. Specifically, only one of the samples
291 from the dataset of Peralta *et al*. [34] was obtained from a live animal while the remaining ones
292 corresponded to deceased animals at different stages of decomposition. In this regard, we found
293 that Index X requires an average coverage depth above 10X to correctly assign the sex of samples;
294 below this value it tended to fail, with a bias toward males. A possible explanation may be that the
295 difference in coverage depth between the X chromosome and the autosomes that characterizes
296 males is reduced when average coverage decreases. On the contrary, the Y index seems to

297 overcome this problem, as low coverage depth would not interfere with correct sex assignment.
298 Index Y even seems to benefit from low coverage depth because males tend to show more negative
299 values at lower coverage, resulting in a more pronounced distinction between sexes. In this sense,
300 the influence of the different coverage depths on our indexes was evident in both the fur seal and
301 our sea lion datasets. In turn, these inferences become more robust when contrasting the
302 phenotypic sex of the seals with the efficacy of each of our indexes, confirming our suspicions about
303 their different performance.

304 Overall, we suggest that Index Y is more reliable in assigning sex to individuals with low depth
305 coverage, which is of particular relevance for studies involving non-invasive samples where DNA is
306 usually degraded and on which RAD-seq has a weaker performance [46]. However, its joint use with
307 Index X may provide a more accurate sexing result or can serve as a double check.

308

309 Our Index X is similar to the approach developed by Pečnerová *et al*. [5], who determined the
310 sex of 98 woolly mammoths from frozen bone samples. They mapped the entire genome of
311 mammoths against the chromosome-level LoxAfr4 assembly of a female African savannah elephant
312 to compare the number of reads mapping to an autosome and to chromosome X. Like in our
313 method, males were expected to have about 50% of reads mapping to chromosome X compared to
314 the autosome, while females were expected to have a comparable number of reads mapping to
315 both chromosomes. The main difference with the present study is that we developed Index X using
316 a RAD-seq approach and compared values of coverage depth for all the sites (fixed and variable) in
317 the RAD-loci identified, while Pečnerová *et al*. [5] used the number of total reads mapping to
318 chromosome 8 and X, respectively, for each individual. Moreover, the method described by these
319 authors requires the chromosomes under comparison to be of the similar length because it uses the
320 number of total reads, while ours does not because it measures average coverage depth per loci.

321 Another method that deserves mention was used by Stovall *et al*. [25]. They employed *a de novo*
322 approach to identify SNPs representing a useful tool if no chromosome-level reference genome is
323 available for the species under study. However, multiple statistical validations must be performed
324 to identify sex-related SNPs and assign them to individuals; this procedure needs to be repeated for
325 each new study, thereby hampering its implementation. Moreover, it may be influenced by
326 sequencing standards such as coverage depth and sample size. In comparison, the method proposed
327 here is robust, accurate and easy to use even with samples having low average coverage depth and
328 can be applied to new species with minimal modifications. Lastly, the accuracy of their sexing
329 method compared to phenotypic sexing was 98.9% and 95.8% for females and males, respectively,
330 while in the present study it was 100% and 97.2% for females and males, respectively.
331
332 Sex-associated markers have been used in studies covering a wide spectrum of topics such as
333 sex-associated SNPs and sex-determining systems in different taxa including mammals, fishes,
334 cnidarians, frogs, lobsters, magnolias and cannabis, among others [21,23–26,47–51]. In this context,
335 it is worth mentioning the bioinformatic tool "RADsex", which is used to identify sex-associated
336 markers through statistical analysis and prediction models using RAD-seq data [21]. Among other
337 benefits, this method also provides the possibility to assign sex to individuals based on genomic
338 information.
339
340 Finally, the method described here is simple and useful for determining the sex of samples from
341 different sources, but it can only be applied in species with heteromorphic sex chromosomes (e.g.,
342 XY/ZW). In these cases, one sex is heterogametic and sex chromosomes differ from each other in
343 their features, such as size, gene content, repeat content, or structural differences. These

344 characteristics give meaning to our indices and enable the detection of differences between males
345 and females, achieving effective sex assignment.

346 The availability of a chromosome-level reference genome from a phylogenetically closely-related
347 species may be regarded as another limitation of our method. However, reference genomes with
348 these characteristics are progressively becoming more prevalent for non-model species, due to the
349 progressive decrease in costs and DNA input requirements for genome assemblies, advancements
350 in bioinformatics tools and improvements in annotation pipelines [52–55]. Additionally, ongoing
351 advances in assembly technologies are addressing the challenges associated with assembling sex
352 chromosomes [29,55].

353 All the above mentioned, along with the existence of numerous international consortia and
354 genome projects focused on the creation of high-quality reference genomes for species across the
355 phylogenetic tree of life (e.g., the 10,000 Plants Genome Sequencing Project, Earth BioGenome
356 Project, Global Invertebrates Genomics Alliance, Vertebrate Genome Project), promise the
357 publication of thousands of genome assemblies containing sex chromosomes within the next
358 decade [29].

359

360

361 # Conclusion

362 We developed a simple methodology for determining the sex of different individuals using
363 genomic RAD-seq data. Our approach comprises two *ad hoc* indices that successfully assigned the
364 sex of sea lions and fur seals. Genomic libraries were created using DNA obtained from various
365 sample types, including non-invasive ones, which sometimes hamper sex identification. We believe
366 that this method could be extrapolated to any animal species with known heteromorphic XY/ZW

sex chromosome systems, for which a chromosome-level reference genome of a closely-related species is available.

Acknowledgments

We are grateful to the following people and organizations for their assistance during sampling: the park rangers of the National Parks and Protected Areas; the government agencies and their staff (Administración de Parques Nacionales, Secretaría de Medio Ambiente y Desarrollo Sustentable de Río Negro, Secretaría de Ambiente de Santa Cruz, Secretaría de Ambiente, Desarrollo Sostenible y Cambio Climático de Tierra del Fuego); the staff of the NGO Fundación Cadace; Adolfo Imbert, Dario Urruti and Agustín "Chepan" Szczepañsk from Centro Hípico del Fin del Mundo, Ushuaia; Lida Pimer; and Proyecto IMMA. We wish to thank Prof. Neil Gemmell for their time and willingness to answer our questions and the anonymous reviewer for manuscript feedback.

Data availability

All relevant data are within the manuscript and its Supporting Information files. Moreover a detailed sequential overview of the methodological procedure is available at: https://github.com/Dieggarp/sexing. RAD-seq sequence reads belonging to Peralta et al. [34] are deposited in GenBank under the BioProject Accession no.: PRJNA1022835.

Supporting information

Supplementary File S1: Step-by-step protocol. It will be also available on protocols.io.

Supplementary figures: Contains S1 Fig. and S2 Fig. with the dispersion plots of Indexes Y and X vs average coverage depth of both datasets.

S1 Table: Table of the parameters and index calculation for the fur seal (*Arctocephalus forsteri*).

S2 Table: Table of the parameters and index calculation for the sea lion (*Otaria flavescens*).

References

1. Sastre N, Francino O, Lampreave G, Bologov V V, López-Martín JM, Sánchez A, et al. Sex identification of wolf (Canis lupus) using non-invasive samples. Conserv Genet. 2009;10: 555–558.
2. Piggott MP, Taylor AC. Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. Wildl Res. 2003;30: 1–13.
3. Peppin L, McEwing R, Ogden R, Hermes R, Harper C, Guthrie A, et al. Molecular sexing of African rhinoceros. Conserv Genet. 2010;11: 1181–1184.
4. Hill D. Handbook of biodiversity methods: survey, evaluation and monitoring. Cambridge University Press; 2005.
5. Pečnerová P, Díez-del-Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, et al. Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. Curr Biol. 2017;27: 3505–3510.
6. Nardelli M, Ibañez E, Dobler D, Illia G, Abba AM, Túnez JI. Genetic approach reveals a polygynous-polyandrous mating system and no social organization in a small and isolated population of the screaming hairy armadillo, Chaetophractus vellerosus. Genetica. 2020;148. doi:10.1007/s10709-020-00092-5
7. Engen S, Lande R, SÆther B-E. Demographic stochasticity and Allee effects in populations with two sexes. Ecology. 2003;84: 2378–2386.
8. Ahlering MA, Hailer F, Roberts MT, Foley C. A simple and accurate method to sex savannah, forest and Asian elephants using noninvasive sampling techniques. Mol Ecol Resour. 2011;11: 831–834.
9. Villesen P, Fredsted T. Fast and non-invasive PCR sexing of primates: apes, old world monkeys, new world monkeys and Strepsirrhines. BMC Ecol. 2006;6: 1–6.
10. Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of XY homologous gene amelogenin. Biotechniques. 1993;15: 636–638.
11. Richard KR, McCarrey SW, Wright JM. DNA sequence from the SRY gene of the sperm whale

423 (Physeter macrocephalus) for use in molecular sexing. Can J Zool. 1994;72: 873–877.
424 12. Durnin ME, Palsbøll PJ, Ryder OA, McCullough DR. A reliable genetic technique for sex
425 determination of giant panda (Ailuropoda melanoleuca) from non-invasively collected hair
426 samples. Conserv Genet. 2007;8: 715–720.
427 13. Cotts L, Moratelli R, Silva DA, Carvalho EF, Loiola S, Amaral CRL. Integrating molecular and
428 morphological data in the secondary sexual identification of museum specimens of
429 Tamandua tetradactyla (Xenarthra, Pilosa). Forensic Sci Int Genet Suppl Ser. 2019;7: 296–
430 298.
431 14. Kamimura S, Nishiyama N, Ookutsu S, Goto K, Hamana K. Determination of bovine fetal sex
432 by PCR using fetal fluid aspirated by transvaginal ultrasound-guided amniocentesis.
433 Theriogenology. 1997;47: 1563–1569.
434 15. Fernando P, Melnick DJ. Molecular sexing eutherian mammals. Mol Ecol Notes. 2001;1: 350–
435 353.
436 16. Bailey DMD, Affara NA, Ferguson-Smith MA. The XY homologous gene amelogenin maps to
437 the short arms of both the X and Y chromosomes and is highly conserved in primates.
438 Genomics. 1992;14: 203–205.
439 17. Aasen E, Medrano JF. Amplification of the ZFY and ZFX genes for sex identification in humans,
440 cattle, sheep and goats. Bio/technology. 1990;8: 1279–1281.
441 18. Munshi-South J, Tchignoumba L, Brown J, Abbondanza N, Maldonado JE, Henderson A, et al.
442 Physiological indicators of stress in African forest elephants (Loxodonta africana cyclotis) in
443 relation to petroleum operations in Gabon, Central Africa. Divers Distrib. 2008;14: 995–1003.
444 19. Martinelli AB, DE MORAES-BARROS N, Alvarenga CS, Chaves PB, Santos LAD, Fagundes V. A
445 PCR-RFLP assay for gender assignment in the three-toed sloths (Bradypus, Pilosa,
446 Bradypodidae). Mol Ecol Resour. 2010;10: 732–734.
447 20. Davey JL, Blaxter MW. RADseq: Next-generation population genetics. Brief Funct Genomics.
448 2010;9: 416–423. doi:10.1093/bfgp/elq031
449 21. Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, et al. RADSex: A computational
450 workflow to study sex determination using restriction site-associated DNA sequencing data.
451 Mol Ecol Resour. 2021;21: 1715–1731.
452 22. Gamble T, Zarkower D. Identification of sex-specific molecular markers using restriction site-
453 associated DNA sequencing. Mol Ecol Resour. 2014;14: 902–913.
454 23. Carmichael SN, Bekaert M, Taggart JB, Christie HRL, Bassett DI, Bron JE, et al. Identification

455 of a sex-linked SNP marker in the salmon louse (Lepeophtheirus salmonis) using RAD
456 sequencing. PLoS One. 2013;8: e77832.
457 24. Benestan L, Moore J, Sutherland BJG, Le Luyer J, Maaroufi H, Rougeux C, et al. Sex matters
458 in massive parallel sequencing: Evidence for biases in genetic parameter estimation and
459 investigation of sex determination systems. Mol Ecol. 2017;26: 6767–6783.
460 25. Stovall WR, Taylor HR, Black M, Grosser S, Rutherford K, Gemmell NJ. Genetic sex assignment
461 in wild populations using genotyping-by-sequencing data: A statistical threshold approach.
462 Mol Ecol Resour. 2018;18: 179–190. doi:10.1111/1755-0998.12767
463 26. Trenkel VM, Boudry P, Verrez-Bagnis V, Lorance P. Methods for identifying and interpreting
464 sex-linked SNP markers and carrying out sex assignment: application to thornback ray (Raja
465 clavata). Mol Ecol Resour. 2020;20: 1610–1619.
466 27. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq
467 for ecological and evolutionary genomics. Nat Rev Genet. 2016;17: 81–92.
468 doi:10.1038/nrg.2015.28
469 28. Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, et al. The era of
470 reference genomes in conservation genomics. Trends Ecol Evol. 2022;37: 197–202.
471 29. Carey SB, Lovell JT, Jenkins J, Leebens-Mack J, Schmutz J, Wilson MA, et al. Representing sex
472 chromosomes in genome assemblies. Cell genomics. 2022;2.
473 30. Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, et al. The genome of the
474 green anole lizard and a comparative analysis with birds and mammals. Nature. 2011;477:
475 587–591.
476 31. Sutou S, Mitsui Y, Tsuchiya K. Sex determination without the Y chromosome in two Japanese
477 rodents Tokudaia osimensis osimensis and Tokudaia osimensis spp. Mamm genome.
478 2001;12: 17–21.
479 32. Saunders PA, Perez J, Ronce O, Veyrunes F. Multiple sex chromosome drivers in a mammal
480 with three sex chromosomes. Curr Biol. 2022;32: 2001–2010.
481 33. Graves JAM. Evolution of vertebrate sex chromosomes and dosage compensation. Nat Rev
482 Genet. 2016;17: 33–46.
483 34. Peralta DM, Ibañez EA, Lucero S, Cappozzo HL, Ceballos SG, Túnez JI. (Under review) Low
484 genetic variability and high isolation of a post-harvest South American pinniped population
485 as revealed by genome-wide SNPs. bioRxiv. 2023. doi:10.1101/2023.12.15.571663
486 35. Peralta DM, Cappozzo HL, Ibañez EA, Lucero S, Failla M, Túnez JI. Phylogeography of Otaria

487 flavescens (Carnivora: Pinnipedia): unravelling genetic connectivity at the southernmost limit
488 of its distribution. Biol J Linn Soc. 2021;134: 11–27. doi:10.1093/biolinnean/blab053
489 36. Roesti M, Hendry AP, Salzburger W, Berner D. Genome divergence during evolutionary
490 diversification as revealed in replicate lake–stream stickleback population pairs. Mol Ecol.
491 2012;21: 2852–2862.
492 37. Roesti M, Moser D, Berner D. Recombination in the threespine stickleback genome—
493 patterns and consequences. Mol Ecol. 2013;22: 3014–3027.
494 38. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics
495 of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet.
496 2010;6. doi:10.1371/journal.pgen.1000862
497 39. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform.
498 Bioinformatics. 2010;26: 589–595.
499 40. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
500 population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:
501 2987–2993.
502 41. Árnason Ú. Comparative chromosome studies in Pinnipedia. Hereditas. 1974;76: 179–225.
503 42. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: Analytical methods for paired-end
504 sequencing improve RADseq-based population genomics. Mol Ecol. 2019;28: 4737–4754.
505 43. Rochette NC, Catchen JM. Deriving genotypes from RAD-seq short-read data using Stacks.
506 Nat Protoc. 2017;12: 2640–2659. doi:10.1038/nprot.2017.123
507 44. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call
508 format and VCFtools. Bioinformatics. 2011. doi:10.1093/bioinformatics/btr330
509 45. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria.
510 2019;0. Available: https://www.r-project.org/
511 46. Andrews KR, De Barba M, Russello MA, Waits LP. Advances in Using Non-invasive, Archival,
512 and Environmental Samples for Population Genomic Studies. In: Hohenlohe PA, Rajora OP,
513 editors. Population Genomics: Wildlife. Cham: Springer International Publishing; 2018. pp.
514 63–99. doi:10.1007/13836_2018_45
515 47. Wessels S, Krause I, Floren C, Schütz E, Beck J, Knorr C. ddRADseq reveals determinants for
516 temperature-dependent sex reversal in Nile tilapia on LG23. BMC Genomics. 2017;18: 1–13.
517 48. Kafkas S, Khodaeiaminjan M, Güney M, Kafkas E. Identification of sex-linked SNP markers
518 using RAD sequencing suggests ZW/ZZ sex determination in Pistacia vera L. BMC Genomics.

519 2015;16: 1–11.
520 49. Pratlong M, Haguenauer A, Chenesseau S, Brener K, Mitta G, Toulza E, et al. Evidence for a
521 genetic sex determination in Cnidaria, the Mediterranean red coral (Corallium rubrum). R
522 Soc open Sci. 2017;4: 160880.
523 50. Yang X, Luo W, Xia Y, Zeng X. Using sex-linked markers via Genotyping-by-Sequencing to
524 identify XX/XY sex chromosomes in the Spiny frog (Quasipaa boulengeri). Genes (Basel).
525 2022;13: 575.
526 51. Gilchrist EJ, Hegebarth D, Wang S, Quilichini TD, Sawler J, Toh SY, et al. A rapid method for
527 sex identification in Cannabis sativa using high resolution melt analysis. Botany. 2021;101:
528 284–290.
529 52. Taylor RS, Manseau M, Redquest B, Keobouasone S, Gagné P, Martineau C, et al. Whole
530 genome sequences from non-invasively collected caribou faecal samples. Conserv Genet
531 Resour. 2021; 1–16.
532 53. Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of
533 threatened species. Genes (Basel). 2019;10: 846.
534 54. Wright B, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. From reference
535 genomes to population genomics: Comparing three reference-aligned reduced-
536 representation sequencing pipelines in two wildlife species. BMC Genomics. 2019;20: 1–10.
537 doi:10.1186/s12864-019-5806-y
538 55. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and
539 error-free genome assemblies of all vertebrate species. Nature. 2021;592: 737–746.
540