When sex scarce: A rapid approach to sexing individuals by RAD-seq using a reference genome

GBS approach for sexing

Diego M. Peralta[12*], Juan I. Túnez[13], Ulises E. Rodríguez Cruz[4], Santiago G. Ceballos[56]

[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: peraltadd@gmail.com

Abstract

Assigning sex to individuals without previous information is a common objective of molecular ecology. Here, we developed a framework for sexing animals by using two indexes based on the different properties of the mammalian sexual chromosomes. We mapped RAD-seq loci to a reference genome to obtain missingness and coverage depth from chromosomes Y, X and autosomal, which allowed identifying the sex of fur seals from a previous study with previous sex information. Moreover, we sexed 38 sea lions sampled non-invasively, allowing us to discuss our indexes' reliability at different coverage depths. We believe this approach could extrapolate to any mammal species or taxa with known XY sex chromosome systems and different qualities of the GBS sequencing.

Author summary

When we start a new behavioral study, one of the arising questions is whether males and females have the same behavior facing different situations. However, it is not always easy to identify the sex of an animal. Some species do not present differences at a glance between sexes, and at other times, differences appear at completing sex maturity, which complicates knowing their sex. But, a more challenging scenario appears, for example, at working with complex animals to catch or even see, where we have to settle with their feathers, hairs or excrement. For that reason, the possibility of identifying the sex of several individuals using genetics has become of great interest. In this work, we developed a computational methodology for sexing animals genetically using the biological differences of sex chromosomes.

Keywords

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

52 **Introduction**

53

54 Population size, sex ratio, kinship and distribution of free-ranging wildlife are among the most
55 relevant data to achieve effective conservation strategies (Sastre et al. 2009). However, when
56 the subject under study is an elusive or endangered species and using non-invasive
57 methodologies is necessary (Piggott and Taylor 2003), information about sex is not always
58 present. In this regard, identifying the sex of individuals is an important component in
59 population ecology and conservation (Peppin et al. 2010), for example, improving the
60 understanding of population structure and dynamics in monitoring programs (Hill 2005) or
61 infering socio-ecological patterns in behavioral studies of living and extinct species (Pečnerová
62 et al. 2017; Nardelli et al. 2020). Indeed, having knowledge of the sex ratio within a population
63 is significant for species or populations of conservation concern, as sub-optimal sex ratios can
64 have a negative impact on population growth and resilience (Engen et al. 2003). However, the
65 lack of sexual dimorphism of some species, particularly in juveniles, and the few opportunities
66 for direct observation of elusive species, usually hamper sex identification (Peppin et al. 2010;
67 Ahlering et al. 2011). This circumstance has turned to identify the sex of a DNA sample into an
68 important tool in molecular ecology and conservation genetics (Villesen and Fredsted 2006).

69

70 Different traditional approaches have achieved molecular sexing in mammals. The most
71 popular methods use the polymerase chain reaction (PCR) to amplify specific fragments in the Y
72 chromosome, or homologous fragments from both X and Y chromosomes (Sullivan et al. 1993;
73 Richard et al. 1994; Villesen and Fredsted 2006; Durnin et al. 2007; Cotts et al. 2019).
74 Nevertheless, these methods depend on amplification reliability and external controls.
75 (Kamimura et al. 1997; Fernando and Melnick 2001). Other commonly approaches are the
76 differences in size polymorphisms among X and Y homologous fragments, e.g. Amelogenin gene
77 size (Bailey et al. 1992), or the existence of unique restriction sites in a Y fragment, like zinc finger

genes (Aasen and Medrano 1990; Munshi-South et al. 2008; Martinelli et al. 2010), although, they do not apply to every taxa (Fernando and Melnick 2001).

The advantages of the Restriction site associated DNA (RAD-seq) markers to overcome the limitations of traditional techniques (Davey and Blaxter 2010) and the possibility of working with wild populations and non-model species became this method popular in the field of population genetics. In this regard, since the appearance of RAD-seq several works have attempted to use its potential to understand sex chromosomes systems better or to sexing samples of different sources with successful results (Carmichael et al. 2013; Gamble and Zarkower 2014; Benestan et al. 2017; Feron et al. 2021). For example, Stovall et al. (2018) assigned sex to 86 individuals of New Zealand fur seals (*Arctocephalus forsteri*) by the discovery and statistical validation of sex-specific loci in a *de novo* RAD-seq assay.

Considering the confidence of RAD-seq and the increase of reference genomes assembly for multiple species, our objective was: to develop a simple methodology for sexing individuals based on the different properties of the mammalian sexual chromosomes X and Y and the possibility of mapping RAD-seq loci to a reference genome. Thus, we create two indexes that can potentially identify the sex of any animal with XY sex chromosomes, given the availability of a closely related genome assembled at the chromosome level. Combining these indexes allowed us to accurately assign sex to the fur seal's dataset from Stovall et al. (2018) and the sea lion's from Peralta (2020).

Methods

Samples

We used the GBS (Genotyping-by-sequencing) data analyzed in Stovall et al. (2018), consisting of 165 fur seals (*Arctocephalus forsteri*) of known sex. This dataset was downloaded from the NCBI Sequence Read Archive (BioProject Accession no.: PRJNA419445).

After that, we applied the method to 38 individuals from Peralta (2020) Rad-seq library of *O. flavescens*. These samples were obtained by invasive and non-invasive methods along the coast of Argentina (Peralta et al. 2021).

SNP calling and filtering

Individual reads of both data sets were mapped to the California sea lion representative genome mZalCal1.pri.v2 (GCA_009762295.2). Detailed methodology in Peralta (2020). The SNPs calling and filtering steps (*gstacks*, STACKS v2.53) were aimed to retain as many markers associated with sex chromosomes as possible. Thus, considering that Y linked loci are present only in males, we set the minimum proportion of individuals across populations to process a locus (-R) to 0.3, which assumes a minimum of 30 % of males in our data set.

Sex identification

Sex identification of individuals was carried out using the chromosome-level assembled genome of the California sea lion as reference. Since males carry a single copy of the X chromosome in mammals, specimens belonging to males are expected to have about 50% of coverage depth for loci mapping to the X chromosomes compared to an autosomal chromosome. On the other hand, females are expected to have a similar coverage depth for the X and any autosomal chromosome. In the case of the Y chromosome, female specimens are expected to have a high value of missingness (measured as the percentage of the total number of loci found in the population that are absent in the individual). Whereas the missingness level in male specimens for the Y chromosome should be similar to the overall missingness found at any other chromosome. However, it should be theoretically more similar to the missingness

level found at the X chromosome since males have a single copy for each sex chromosome. Considering this, we used two different indexes for sex identification based on the different properties of the two sex chromosomes. We defined the Index X as the average coverage depth of the X chromosome divided by the average coverage depth of the autosomal chromosome 10. For the Index_Y (based on the Y chromosome), we first defined the individual coverage as one minus the missingness, and used the coverage 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 regarding the overall missingness for that individual.

To obtain values of missingness and coverage depth we used different options of VCFtools (Danecek et al. 2011) over the output file of STACKS “populations.snps.vcf“ (Catchen et al. 2013). We filtered SNP loci in the sexual Y (NC_045613.1) and X (NC_045612.1) chromosomes, and the autosomal chromosome 10 (NC_045604.1). We then obtained the missingness per individual for the X and Y chromosomes with the option “--missing-indv” of VCFtools and the average coverage depth with the option “--depth”. The step-by-step procedure is described in the supplementary file “Sexing.sh” (SI_sexing.sh). We then calculate both indexes using R (R Core Team 2019) and following the equations described in Table 1. We attach an automatized R script, “Sexing.R” (SI_sexing.R), to perform this last step. 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. For assigning sex to each individual, we only use values of Index Y because of the misidentification of using Index X.

Results

Genomic sexing

From the 165 fur seals from Stovall et al. (2018), we successfully sexed 164 individuals with both indexes and one with Index Y only, resulting in 95 females and 70 males (Table S1). Interesting, two individuals assigned as males in Stovall et al. (2018) were sexed as females by our two indexes.

The dispersion plot of both Index values (Figure 1) show two tight clouds of dots accordingly with our expectation for males and females. On the one hand, the dots that formed a tight cloud at the intersection of value 1 of both indexes are inferred to belong to females. On the other hand, the dots located almost in their totality near the intersection of values 0.65 of Index X and with values close to zero or negatives of Index Y are inferred to belong to males. The dot that appears as an outlier within the lower cloud presents lower values for Index Y than the others (-1.6) and a value close to 1 for the Index X. It is worth noting this individual also showed the lowest value of coverage depth, which led us to test how coverage depth affects both indexes. We found that low values of coverage depth decline the performance of Index X badly (an individual of low coverage depth was even misidentified) but not of Index Y (Figure S1). Moreover, something interesting is that at low values of coverage depth, the Index Y shows more negative values in male individuals, which even may improve its performance to identify sex.

From the 38 individuals analyzed of the Sea Lion population, 30 were successfully sexed using both indexes, resulting in 19 males and 11 females. The remaining individuals could only be sexed by the Index Y, resulting in two females and six males. (Table S2).

The dispersion plot of both Index values (Figure 2) showed two clouds of dots representing males and females. The female cloud was located at the intersection of value 1 for Index X and Y, while the male cloud was near the intersection of values 0.5 of Index X and 0 of Index Y. However, some dots present values above 0.5, for Index X, and lower than 0 for Index Y. Last, we again studied the effect of low coverage depth in our indexes to prove if they were also

181 affected in this second case. Accordingly, we found that Index X was affected, whereas Index Y
182 continued to perform without any problem (Figure S2). Once more, we found that at low
183 coverage depth, Index Y values for males turn more negatives, improving its correct
184 identification.
185
186 **Discussion**
187
188 **Sexing animals from a RADseq approach**
189
190 This work provide a simple framework for sexing different types of samples using two ad hoc
191 indexes. Our method will allow researchers working with RAD-seq or other similar GBS
192 methodology, counting with a related reference genome, to assign sex to their individuals in an
193 easy way.
194 After applying our indexes, we assigned the sex to all the individuals in the Stovall et al. (2018)
195 dataset, finding 95 females and 70 males. The dispersion plot created with the indexes (Figure
196 1) allows us to observe how Index Y locates individuals assigned as females near the value 1 in
197 the plot, while individuals assigned as males were located below 0. In the case of Index X, it
198 allows to split almost all individuals into two different clouds along the abscissa, except for one
199 individual that was only differentiated by the Y-index. A point to notice is that we found two
200 individuals labelled as males that were identified as females accordingly to our indexes. Still, one
201 of the authors confirmed these cases could correspond to misidentifications or mislabeled data
202 in personal communication.
203 Regarding our dataset of sea lions, we found 25 males and 13 females. However, eight of our
204 individuals could only be sexed through the Index Y because their values of Index X were unable
205 to discriminate between males and females. In this case, the dispersion plot (Figure 2) confirms
206 the reliability of Index Y, which differentiate all the individuals, while the Index X assign several

207 cases wrongly. The impossibility of assigning the whole individuals through the two indexes led
208 us to think that Index X is not as robust as Index Y.

209 The main similitude among the samples that could only be sexed by Index Y was the low
210 values of coverage depth, in some cases lower than 5X. Indeed, when we tested the
211 performance of Index X at different coverage depths, we found that it needs coverage over 10X
212 to assign the sex of a sample correctly. When this value is below 10X, this index tends to fail,
213 mainly for the assignment of males. A possible explanation may be the disappearance of the
214 difference in coverage depth between the X chromosome and the autosomal that characterize
215 males when the mean coverage decreases. On the contrary, the Y index seems to overcome this
216 problem since low coverage depth would not interfere with the correct assignment of females
217 and males. Indeed, the Index Y seems to benefit from low coverage depth because, at lower
218 coverage, males tend to show more negative values, which separate them more from females.
219 In this sense, we observed the influence of the different coverage depths over our indexes in
220 both the fur seals dataset and our sea lion's one. These results are very important because the
221 fur seals dataset already counted with information about sex, confirming our suspicions of the
222 performances of Index X and Y.

223 Overall, we suggest that Index Y is more reliable in assigning sex to individuals with low depth
224 coverage, which may take actual relevance in studies working with non-invasive samples where
225 DNA is usually degraded, and RAD-seq performed less effectively (Andrews et al. 2018).
226 However, its combination with Index X may allow a better sexing or act as a double check in the
227 process.

228

229 It is worth noting that our Index X is similar to the approach achieved by Pečnerová et al.
230 (2017), who identified the sex of 98 woolly mammoths from frozen bone samples. Their
231 methodology consisted of mapping the entire genome data of mammoth with the chromosome-
232 level LoxAfr4 assembly of a female African savannah elephant to compare the number of reads

233 mapping between an autosome chromosome and the X chromosome. As in our method,
234 specimens belonging to males were expected to have about 50% of reads mapping to
235 chromosome X compared to the autosomal, while females were expected to have a comparable
236 number of reads mapping to both chromosomes. In this regard, the main difference in our
237 methodologies lies in the fact that we developed our index from a RAD-seq approach and
238 compared values of coverage depth for those loci where SNPs were previously found. In
239 comparison, Pečnerová et al. (2017) used the number of total reads mapping the chromosome
240 8 and chromosome X for each individual. Indeed, their methodology requires that chromosomes
241 under comparison have similar lengths because they observe total reads, whereas our method
242 is independent of that because it measures average coverage depth per loci.
243
244 Finally, although our method offers a simple alternative to identify the sex of different types
245 of samples, its main limitation is the necessity of a chromosome-level reference genome from a
246 near phylogenetic species. This premise could result in a problem, but lastly, the assembling of
247 genomes is more and more common in non-model species, and even it could be achieved from
248 non-invasive samples (Taylor et al. 2021). Indeed, the reduction in costs and the input DNA
249 requisites, and the improvements in the bioinformatic assembly and annotation pipelines based
250 on non-model species become new reference genomes more attainable (Brandies et al. 2019;
251 Wright et al. 2019). In turn, as Brandies et al. (2019) remarks, in recent years, international
252 consortia and genome projects have been formed to create high-quality reference genomes for
253 species spanning the phylogenetic tree of life.
254
255 **Conclusion**
256
257 Overall, in this work, we developed a straightforward methodology to identify the sex of an
258 individual using two indexes harnessing genomic data from RAD-seq. This method allowed us to

correctly assign the sex to the individuals of two datasets of fur seals and sea lions with different kinds of samples, including non-invasive ones, which hamper knowing the sex of the sampled individual. We think this method could extrapolate to any animal species or other taxa with known XY sex chromosome systems giving the availability of a closely related genome reference assembled at the chromosome level.

Acknowledgments and funding information

We are grateful to the different persons and organizations who help us during the sampling: to the park rangers of the National Parks and Protected Areas; to 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 and Secretaría de Ambiente, Desarrollo Sostenible y Cambio Climático de Tierra del Fuego); to the staff of the NGO Fundación Cadace; to Adolfo Imbert, Dario Urruti and Agustín "Chepan" Szczepañsk from Centro Hípico del Fin del Mundo, Ushuaia; to Lida Pimer; and to Proyecto IMMA. We also want to thank Prof. Neil Gemmell for their predisposition and time in answering our doubts. This work was supported by fellowships from the Consejo Nacional de Investigaciones Científicas y Técnicas and Universidad Nacional Autónoma de México, and grants from the National Geographic Society (10000-16 and EC-52880R-18), the Universidad Nacional de Luján (Finalidad 3.5 año 2017) and Sociedad Argentina para el Estudio de los Mamíferos (Premio Kravetz 2019). The authors declare that they have no conflict of interest.

References

Aasen E, Medrano JF. 1990. Amplification of the ZFY and ZFX genes for sex identification in humans, cattle, sheep and goats. Bio/technology. 8(12):1279–1281.

Ahlering MA, Hailer F, Roberts MT, Foley C. 2011. A simple and accurate method to sex savannah, forest and Asian elephants using noninvasive sampling techniques. Mol Ecol Resour. 11(5):831–834.

Andrews KR, De Barba M, Russello MA, Waits LP. 2018. Advances in Using Non-invasive,

Archival, and Environmental Samples for Population Genomic Studies. In: Hohenlohe PA, Rajora OP, editors. Population Genomics: Wildlife. Cham: Springer International Publishing. p. 63–99. https://doi.org/10.1007/13836_2018_45.

Bailey DMD, Affara NA, Ferguson-Smith MA. 1992. The XY homologous gene amelogenin maps to the short arms of both the X and Y chromosomes and is highly conserved in primates. Genomics. 14(1):203–205.

Benestan L, Moore J, Sutherland BJG, Le Luyer J, Maaroufi H, Rougeux C, Normandeau E, Rycroft N, Atema J, Harris LN. 2017. Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. Mol Ecol. 26(24):6767–6783.

Brandies P, Peel E, Hogg CJ, Belov K. 2019. The value of reference genomes in the conservation of threatened species. Genes (Basel). 10(11):846.

Carmichael SN, Bekaert M, Taggart JB, Christie HRL, Bassett DI, Bron JE, Skuce PJ, Gharbi K, Skern-Mauritzen R, Sturm A. 2013. Identification of a sex-linked SNP marker in the salmon louse (Lepeophtheirus salmonis) using RAD sequencing. PLoS One. 8(10):e77832.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. Mol Ecol. 22(11). doi:10.1111/mec.12354.

Cotts L, Moratelli R, Silva DA, Carvalho EF, Loiola S, Amaral CRL. 2019. Integrating molecular and morphological data in the secondary sexual identification of museum specimens of Tamandua tetradactyla (Xenarthra, Pilosa). Forensic Sci Int Genet Suppl Ser. 7(1):296–298.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. doi:10.1093/bioinformatics/btr330.

Davey JL, Blaxter MW. 2010. RADseq: Next-generation population genetics. Brief Funct Genomics. 9(5–6):416–423. doi:10.1093/bfgp/elq031.

Durnin ME, Palsbøll PJ, Ryder OA, McCullough DR. 2007. A reliable genetic technique for sex determination of giant panda (Ailuropoda melanoleuca) from non-invasively collected hair samples. Conserv Genet. 8(3):715–720.

Engen S, Lande R, SÆther B-E. 2003. Demographic stochasticity and Allee effects in populations with two sexes. Ecology. 84(9):2378–2386.

Fernando P, Melnick DJ. 2001. Molecular sexing eutherian mammals. Mol Ecol Notes. 1(4):350–353.

Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, Herpin A, Journot L, Parrinello H, Klopp C. 2021. RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. Mol Ecol Resour. 21(5):1715–1731.

Gamble T, Zarkower D. 2014. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Mol Ecol Resour. 14(5):902–913.

Hill D. 2005. Handbook of biodiversity methods: survey, evaluation and monitoring. Cambridge University Press.

Kamimura S, Nishiyama N, Ookutsu S, Goto K, Hamana K. 1997. Determination of bovine fetal sex by PCR using fetal fluid aspirated by transvaginal ultrasound-guided amniocentesis. Theriogenology. 47(8):1563–1569.

Martinelli AB, DE MORAES-BARROS N, Alvarenga CS, Chaves PB, Santos LAD, Fagundes V. 2010. A PCR-RFLP assay for gender assignment in the three-toed sloths (Bradypus, Pilosa, Bradypodidae). Mol Ecol Resour. 10(4):732–734.

Munshi-South J, Tchignoumba L, Brown J, Abbondanza N, Maldonado JE, Henderson A, Alonso A. 2008. Physiological indicators of stress in African forest elephants (Loxodonta africana cyclotis) in relation to petroleum operations in Gabon, Central Africa. Divers Distrib. 14(6):995–1003.

Nardelli M, Ibañez E, Dobler D, Illia G, Abba AM, Túnez JI. 2020. 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. 148(2). doi:10.1007/s10709-020-00092-5.

Pečnerová P, Díez-del-Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, Nikolskiy P, Tikhonov A, Vartanyan S, Dalén L. 2017. Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. Curr Biol. 27(22):3505–3510.

Peppin L, McEwing R, Ogden R, Hermes R, Harper C, Guthrie A, Carvalho GR. 2010. Molecular sexing of African rhinoceros. Conserv Genet. 11(3):1181–1184.

Peralta DM. 2020. Ecología molecular y comportamiento social del León Marino de América del Sur, Otaria flavescens. Universidad de Buenos Aires.

Peralta DM, Cappozzo HL, Ibañez EA, Lucero S, Failla M, Túnez JI. 2021. Phylogeography of Otaria flavescens (Carnivora: Pinnipedia): unravelling genetic connectivity at the southernmost limit of its distribution. Biol J Linn Soc. 134(1):11–27. doi:10.1093/biolinnean/blab053.

Piggott MP, Taylor AC. 2003. Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. Wildl Res. 30(1):1–13.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria. 0. https://www.r-project.org/.

Richard KR, McCarrey SW, Wright JM. 1994. DNA sequence from the SRY gene of the sperm whale (Physeter macrocephalus) for use in molecular sexing. Can J Zool. 72(5):873–877.

Sastre N, Francino O, Lampreave G, Bologov V V, López-Martín JM, Sánchez A, Ramírez O. 2009. Sex identification of wolf (Canis lupus) using non-invasive samples. Conserv Genet. 10(3):555–558.

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(2):179–190. doi:10.1111/1755-0998.12767.

Sullivan KM, Mannucci A, Kimpton CP, Gill P. 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of XY homologous gene amelogenin. Biotechniques. 15(4):636–638.

Taylor RS, Manseau M, Redquest B, Keobouasone S, Gagné P, Martineau C, Wilson PJ. 2021. Whole genome sequences from non-invasively collected caribou faecal samples. Conserv Genet Resour.:1–16.

Villesen P, Fredsted T. 2006. Fast and non-invasive PCR sexing of primates: apes, old world monkeys, new world monkeys and Strepsirrhines. BMC Ecol. 6(1):1–6.

Wright B, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. 2019. From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. BMC Genomics. 20(1):1–10. doi:10.1186/s12864-019-5806-y.

Data Accessibility and Benefit-Sharing section

RAD-seq sequence reads will be deposited in the GenBank after the publication of the manuscript.

Table

		Expected Values	
	Calculation	Females	Males
Index_X	$= \frac{Depth_X}{Depth_A}$	1	0.5
Index_Y	$= \frac{Cov_X - Cov_Y}{Cov_X}$	1	0

Table 1. Indexes X and Y for sex identification. Depth refers to the average coverage depth across loci for the X chromosome or an autosome (Chromosome 10 in our case). *Cov* refers to the coverage for X and Y chromosomes calculated as one minus the missingness, where the

missingness value is the proportion of the total number of loci found in the population that are absent in that individual.

Figures

Figure 1. Dispersion plot of Index Y and X for the Stovall et al. (2018) data set. The red dots include male individuals, while the black ones include females. The blue line shows how females and males are complete differentiated by Index Y.

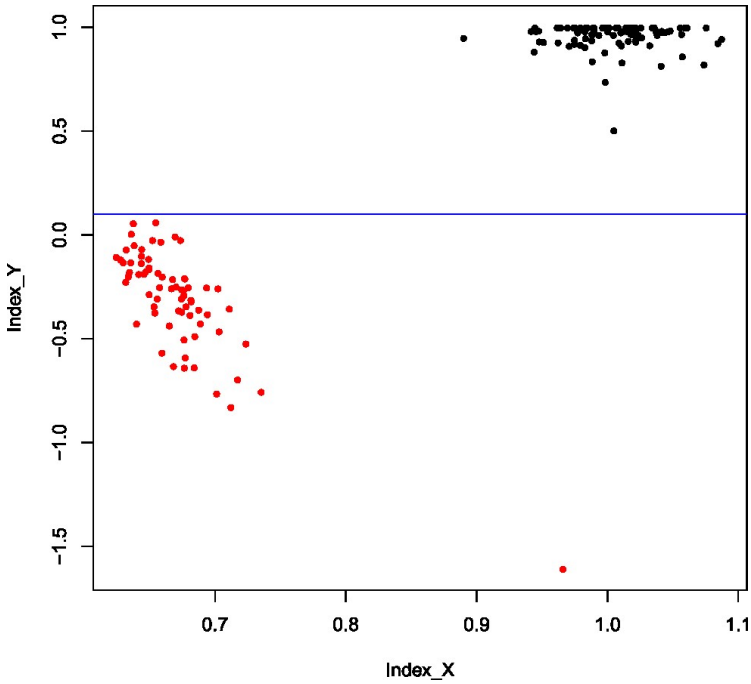
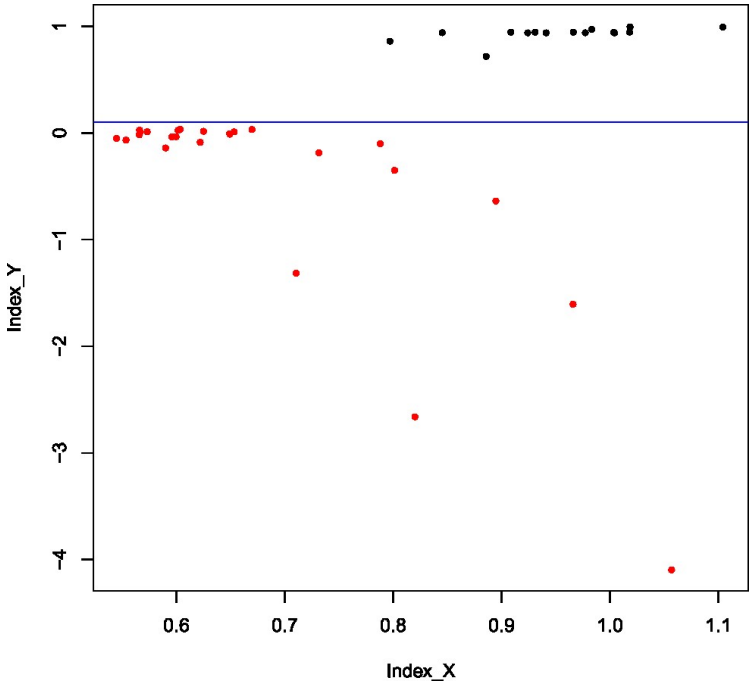


Figure 2. Dispersion plot of Index Y and X for Peralta (2020) data set. The red dots include male individuals, while the black ones include females. The blue line shows how females and males are complete differentiated by Index Y.



395