Received: 23 September 2021 | Accepted: 11 October 2021

DOI: 10.1111/1755-0998.13530

MOLECULAR ECOLOGY RESOURCES WILEY

NEWS AND VIEWS

Perspective

# A bioinformatic toolkit to simultaneously identify sex and sex-linked regions

Iulia Darolti[1] | Judith E. Mank[1,2]

[1]Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

[2]Centre for Ecology and Conservation, University of Exeter, Penryn, UK

**Correspondence**
Judith E. Mank, Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada.
Email: mank@zoology.ubc.ca

**Funding information**
ERC, Grant/Award Number: 680951

Sex chromosomes are strange things, and often exhibit unusual patterns of diversity, rates of evolution, and gene regulation (Bachtrog et al., 2011, Mank, 2013). These unique features mean that although sex chromosomes are often a relatively small proportion of the genome, they are best identified and assessed separately from the autosomal majority when carrying out genomic analyses. However, identifying and partitioning genomic regions into sex-linked and autosomal in non-model species can often be quite difficult. In this issue of *Molecular Ecology Resources*, Nursyifa et al. (2021) provide a useful method that combines sequencing depth information with clustering models to assign sex to samples at the same time as identifying sex-linked scaffolds. This method gives robust results even with more challenging or low-quality data, and thus is particularly promising in studies of non-model organisms.

KEYWORDS
genomics, non-model species, sex chromosomes, short read sequencing

There are two primary approaches to identify sex-linked regions using genomic DNA data in non-model species. It is possible to sequence pools of male and female DNA to identify regions with allele frequency differences between the sexes. This is based on the assumption that as the X and Y diverge, male-specific single nucleotide polymorphisms (SNPs) will accumulate on the Y chromosome (Palmer et al., 2019). This approach can either use a reference genome or, using reduced representation, be reference genome-free. Other methods require mapping whole-genome resequencing data from males and females to compare read depth differences between the sexes. When mapping to a female genome, the single remaining copy of the X in males leads to reduced read depth compared to diploid females (Figure 1, Darolti et al., 2019). Both these methods are quite noisy, and although they can identify the type of heterogamety and the relative proportion of the genome that is sex-linked, they are poor at identifying specific regions, with substantial type I and type II error. Other methods based on SNP segregation require laboratory-reared families, and may not be possible for many organisms.

Both of these methods also rely on comparing genomes of known male and female samples. For many organisms, however, the sex of an individual may not be obvious based on external characteristics. Without sacrificing an individual to determine sex organs, or observing sex-specific mating behaviour, it can be impossible to ascertain sex. This is particularly problematic for non-model species of conservation concern, for which invasive sampling is ethically fraught and limited genomic resources exist.

Nursyifa et al. (2021) offer a new alternative, also based on read coverage differences between the sexes, but which does not require knowing the sex of the samples a priori. The method, sex assignment through coverage (SATC) requires relatively low read depth information of at least two samples of each sex and an assembly of a low quality reference genome from them. The clever advance here is to incorporate a principal component and Gaussian mixture clustering step based on contig read depth across the samples (Figure 1). This clusters sex-linked genomic regions from autosomal ones very effectively, and also has the added benefit of identifying the sex of the samples a posteriori.

The authors tested SATC on several species of birds and therian mammals, and it performed very well, effectively identifying the sex of the samples and sex-linked contigs. However, both birds and

 © 2021 John Wiley & Sons Ltd 

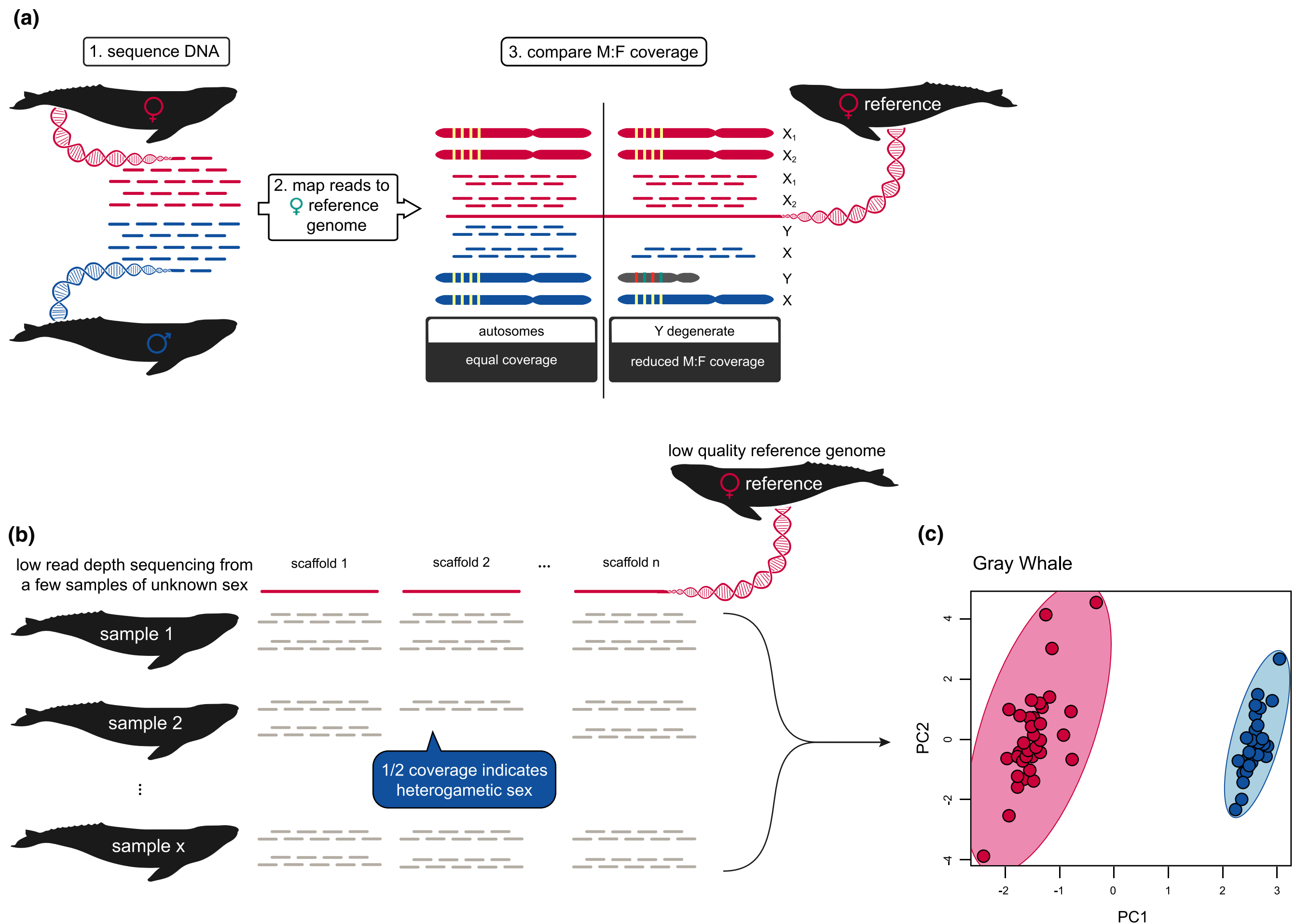


FIGURE 1 Overview of SATC, which jointly assigns sex to samples and identifies sex-linked scaffolds. (a) For XY sex chromosomes, mapping male (blue) and female (red) DNA-seq reads to a female reference genome can distinguish between autosomal and sex-linked regions. While the two sexes should have an equal read depth for autosomal regions, males are expected to show a reduced coverage compared to females in regions of Y degeneration as Y reads no longer map to the X assembly. (b) SATC uses as input a scaffold-level genome assembly and read mapping information from low read depth sequencing of a few samples of unknown sex. (c) The coverage data of each scaffold for each sample is used in a principal component analysis (PCA) and Gaussian mixture clustering to partition samples into male (blue) and female (red), and to distinguish sex-linked from autosomal scaffolds. PCA plot modified from Nursyifa et al. (2021). Figure drawn by Jacelyn Shu (jacelyndesigns.com)

mammals have very heteromorphic sex chromosomes, which are actually quite unusual among vertebrates (Bachtrog et al., 2014). Most vertebrates that have sex chromosomes show limited divergence between the X and Y or Z and W, and it will be interesting to see over time how the method performs on these types of homomorphic sex chromosomes.

## ACKNOWLEDGEMENTS

We thank Jacelyn Shu (jacelyndesigns.com) for composing Figure 1. We gratefully acknowledge funding from the ERC (grant agreement 680951), NSERC and a Canada 150 Research Chair to JEM.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

No data were used in this manuscript.

## ORCID

*Judith E. Mank* https://orcid.org/0000-0002-2450-513X

## REFERENCES

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W., & Valenzuela, N. (2011). Are all sex chromosomes created equal? *Trends in Genetics*, *27*, 350–357. https://doi.org/10.1016/j.tig.2011.05.005

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, *12*, e1001899. https://doi.org/10.1371/journal.pbio.1001899

Darolti, I., Wright, A. E., Sandkam, B. A., Morris, J., Bloch, N. I., Farre, M., Fuller, R. C., Bourne, G. R., Larkin, D. M., Breden, F., & Mank, J. E. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 19031–19036. https://doi.org/10.1073/pnas.1905298116

Mank, J. E. (2013). Sex chromosome dosage compensation: Definitely not for everyone. *Trends in Genetics*, *12*, 677–683. https://doi.org/10.1016/j.tig.2013.07.005

Nursyifa, C., Brüniche-Olsen, A., Garcia-Erill, G., Heller, R., & Albrechtsen, A. (2021). Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13491

Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology*, *28*, 4709–4724. https://doi.org/10.1111/mec.15245

**How to cite this article:** Darolti, I., & Mank, J. E. (2021). A bioinformatic toolkit to simultaneously identify sex and sex-linked regions. *Molecular Ecology Resources*, 00, 1–3. https://doi.org/10.1111/1755-0998.13530