# MOLECULAR ECOLOGY

**Supplemental Information for:**

# Easy-to-use *R* functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex-assignment

Diana A. Robledo-Ruiz, Lana Austin, J. Nevil Amos, Jesús Castrejón-Figueroa, Michael J. L. Magrath, Paul Sunnucks, Alexandra Pavlova

## Table of Contents

**Methods S1.** Identifying the X chromosome in LBP genome assembly

We searched for genes *GLA*, *HPRT1, PGK1* and *SLC16A2* found in the X chromosome of marsupials in order to identify the X chromosome in the Leadbeater's possum assembly (LBP; Deakin et al. 2009). We extracted the gene sequences from the common brushtail possum's chromosome-length assembly (*Trichosurus vulpecula*; mTriVul1.pri in NCBI) and aligned them to the LBP assembly with BLASTn v2.9.0. We set BLASTn to find a maximum of 500 alignments per query sequence (max_target_seqs) with minimum e-value ≥ 10 (Altschul et al. 1990). For each loci sequence, we kept only alignments with the smallest expected value, allowing for ties. All four genes aligned to "HiC_scaffold_3" (length 82.6 Mb) and therefore, we considered it the X chromosome.
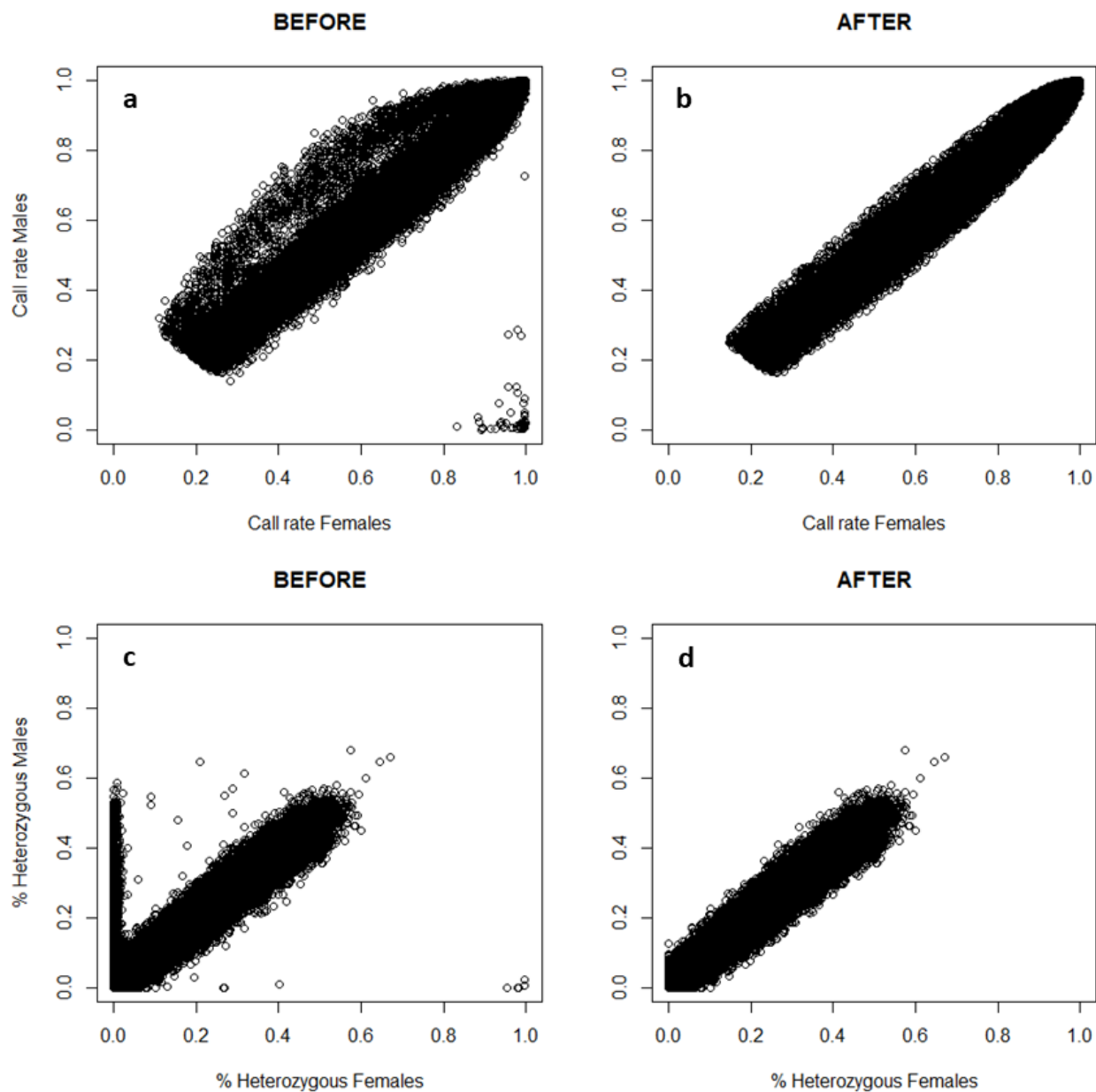
**BEFORE**

**AFTER**

**BEFORE**

**AFTER**



**Figure S1.** Plots produced by function *filter.sex.linked* on YTH data (yellow-tufted honeyeater). Top panels: plots of female call rate against male call rate in which each point represents a locus, before **(a)** and after **(b)** removing 2,220 sex-linked loci with differential call rate between the sexes. Bottom panels: plots of the proportion of heterozygous females against the proportion of heterozygous males with each point representing a locus, before **(c)** and after **(d)** removing 1,135 sex-linked loci with differential heterozygosity between the sexes.
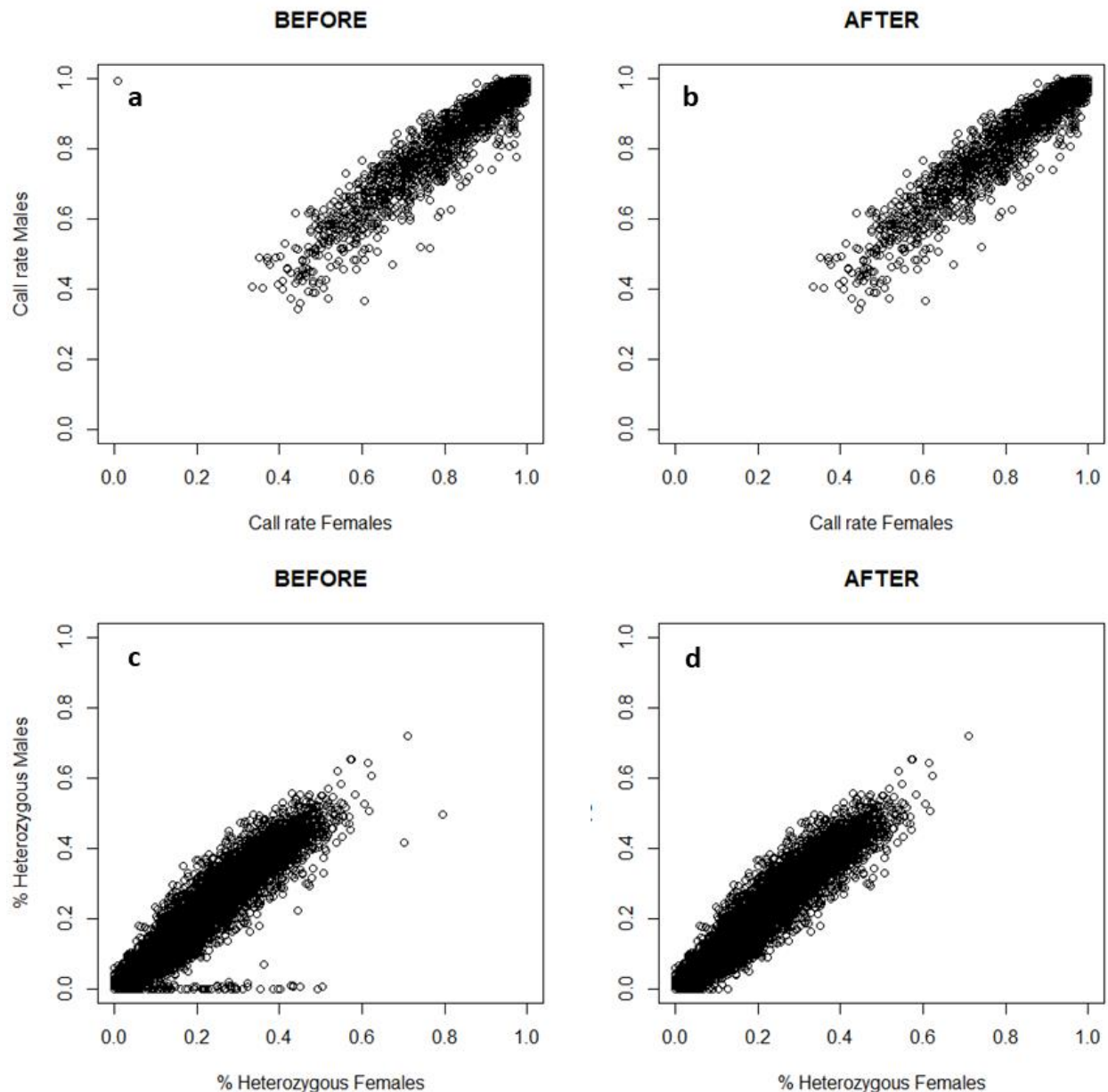
**Figure S2.** Plots produced by function *filter.sex.linked* on LBP data (Leadbeater's possum). Top panels: plots of female call rate against male call rate in which each point represents a locus, before **(a)** and after **(b)** removing 4 sex-linked loci with differential call rate between the sexes. Bottom panels: plots of the proportion of heterozygous females against the proportion of heterozygous males with each point representing a locus, before **(c)** and after **(d)** removing 67 sex-linked loci with differential heterozygosity between the sexes.

**Figure S3.** Six measures of population genetic diversity before and after removing sex-linked loci (Ho: observed heterozygosity, He: expected heterozygosity, FIS: Wright's $F_{IS}$, Polymorphism, number of private alleles, and mean AR: mean allelic richness) per population of eastern yellow robin (EYR).
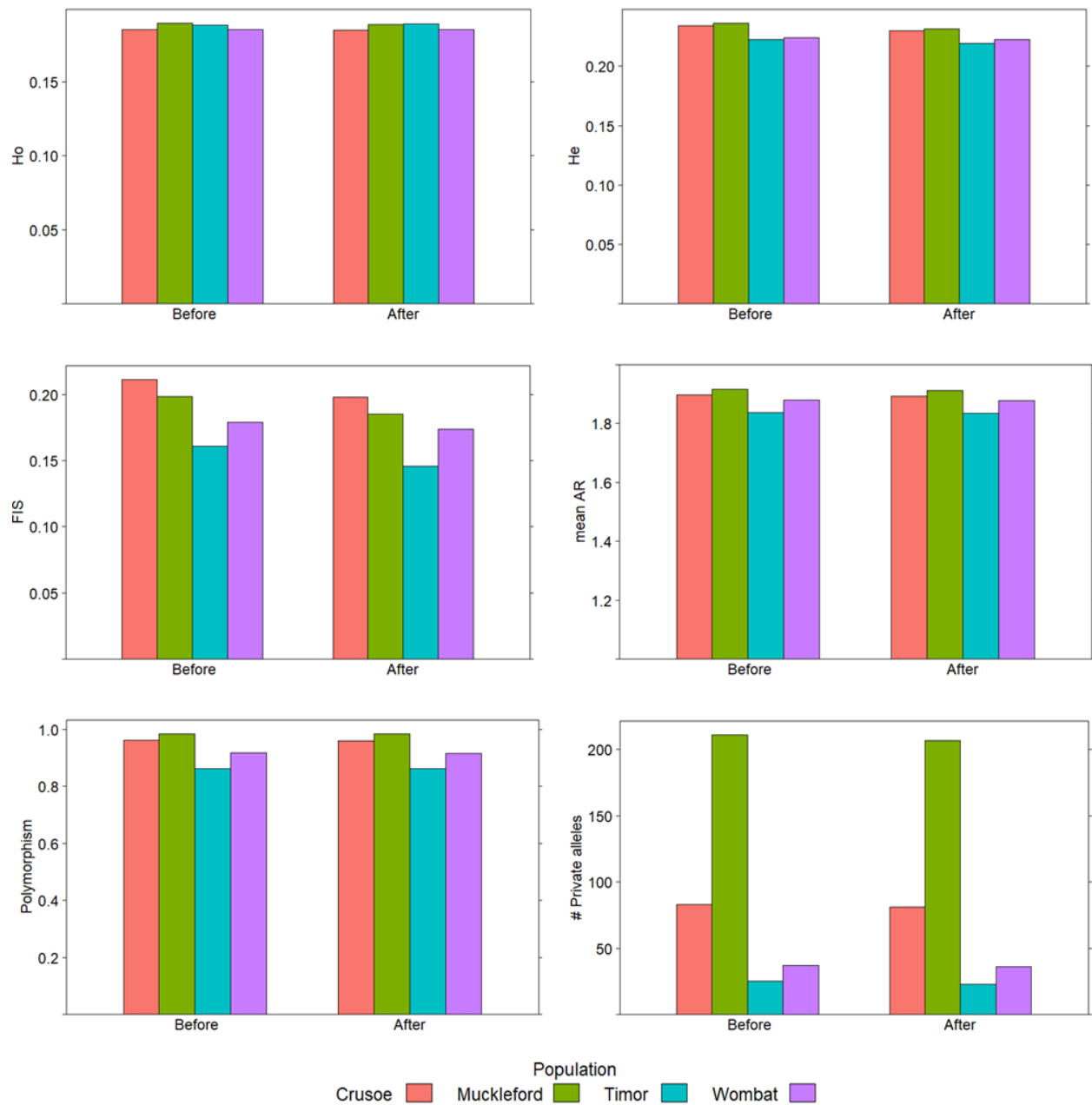
**Figure S4.** Six measures of population genetic diversity before and after removing sex-linked loci (Ho: observed heterozygosity, He: expected heterozygosity, FIS: Wright's $F_{IS}$, Polymorphism, number of private alleles, and mean AR: mean allelic richness) per population of yellow-tufted honeyeater (YTH). Note that the strong bias in the number of samples for *cassidix* resulted in an upward bias in genetic diversity estimates for this taxon due to ascertainment bias of this genetic dataset; *cassidix* was previously shown to have lowest genetic diversity of all subspecies (Pavlova et al. 2014).
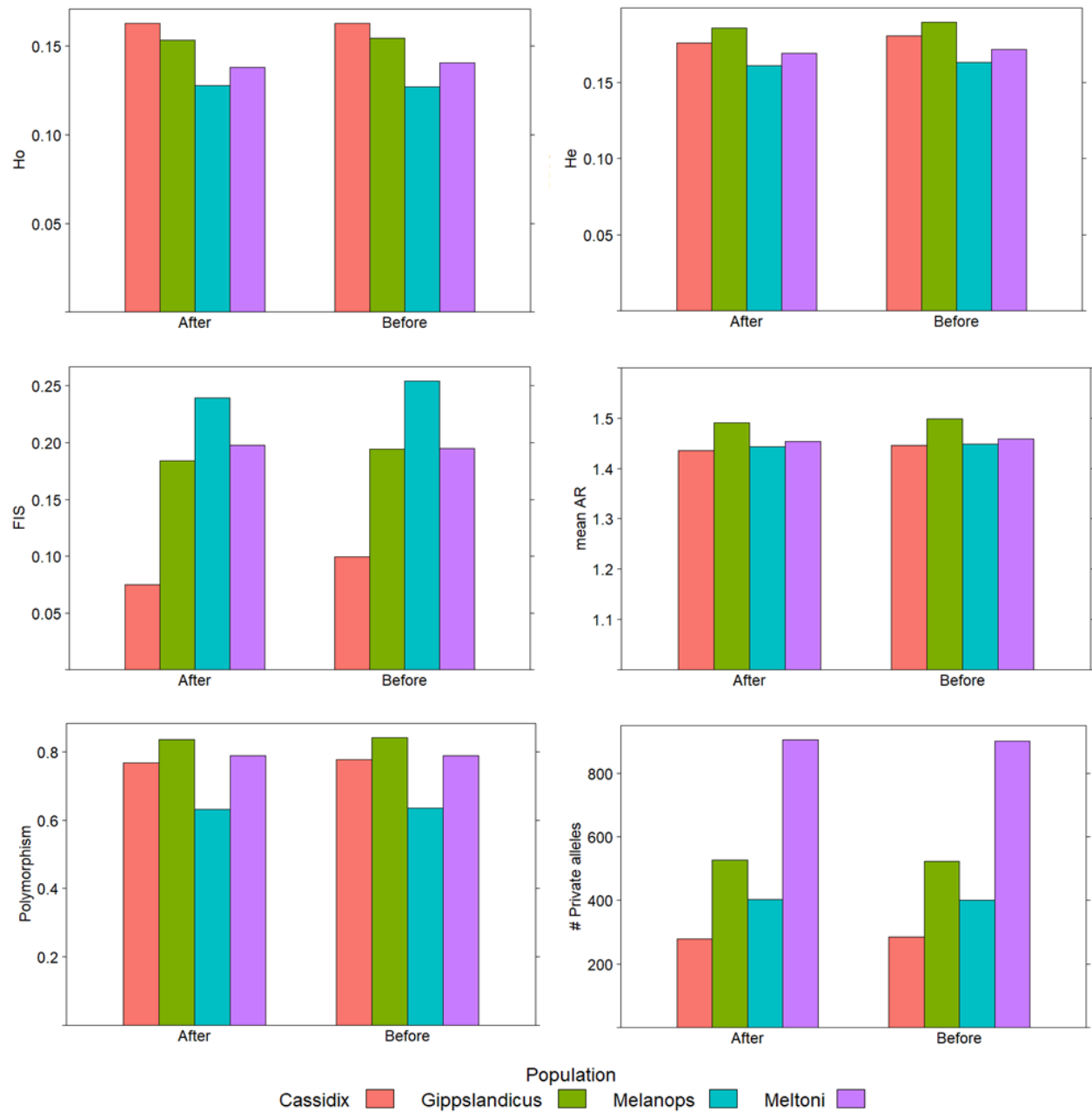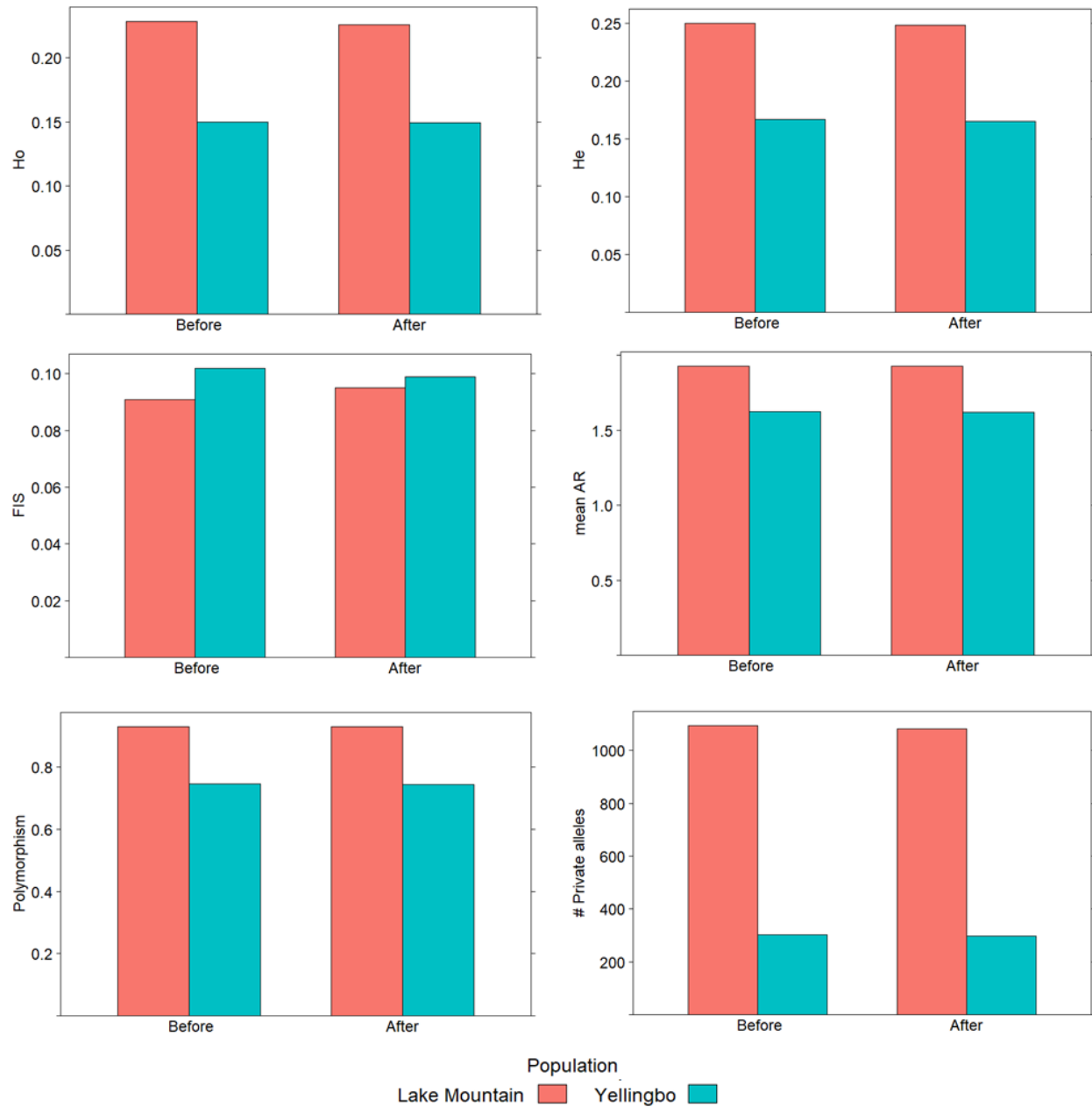
**Figure S5.** Six measures of population genetic diversity before and after removing sex-linked loci (Ho: observed heterozygosity, He: expected heterozygosity, FIS: Wright's $F_{IS}$, Polymorphism, number of private alleles, and mean AR: mean allelic richness) per population of Leadbeater's possum (LBP).

7

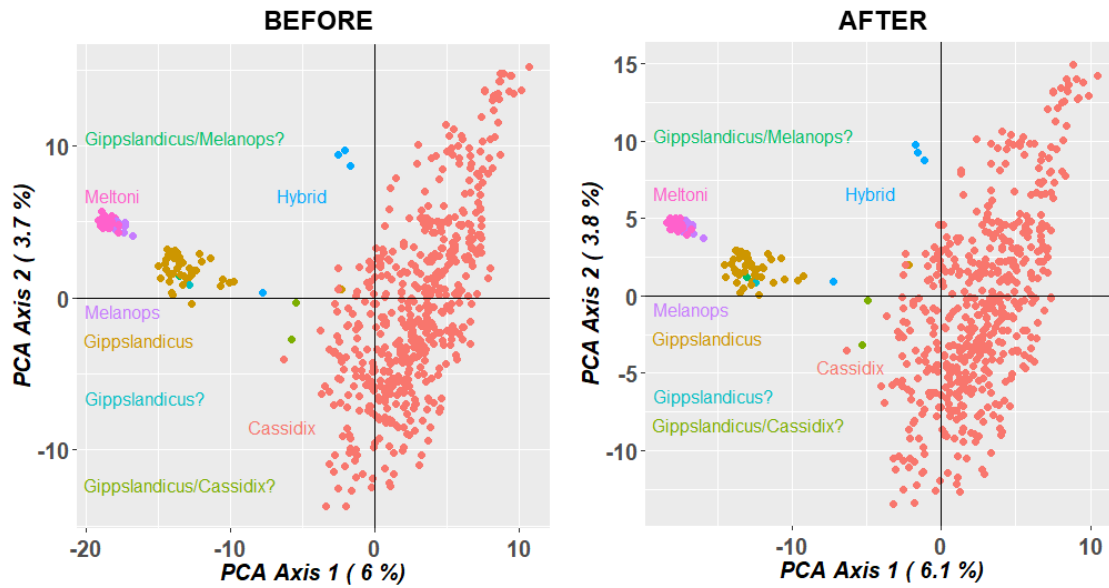**Figure S6.** Principal Component Analyses (PCA) of the genomic dataset of yellow-tufted honeyeater (YTH) before and after removing sex-linked loci.
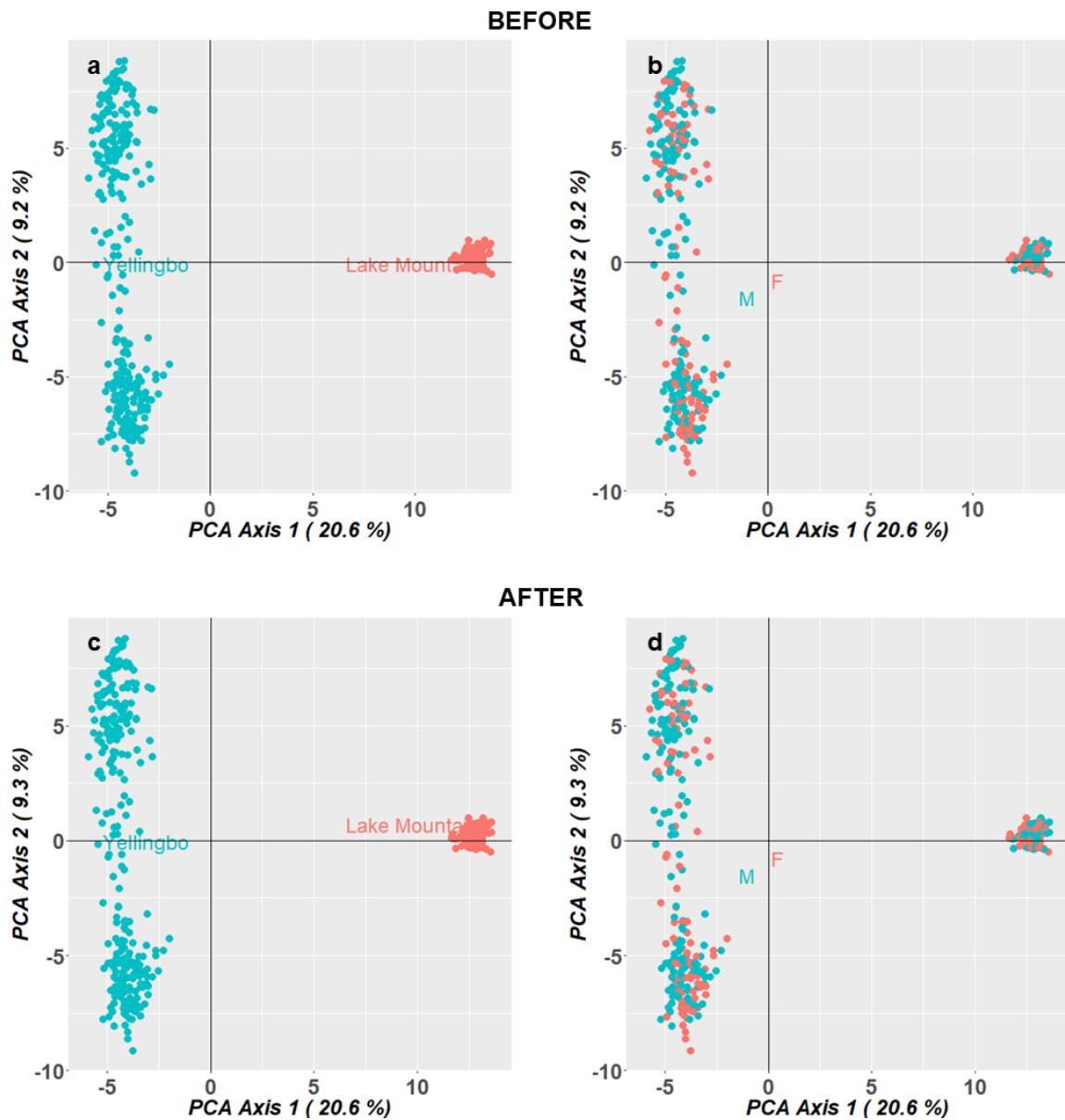
**Figure S7.** Principal Component Analyses (PCA) of the genomic dataset of Leadbeater's possum (LBP) before and after removing sex-linked loci. On (**a**) and (**c**), individuals are coloured according to their population. On (**b**) and (**d**), individuals are coloured by sex.

**Table S1.** Number of true sex-linked loci that function *filter.sex.linked* was able to identify with variable number of known-sex individuals for EYR, YTH and LBP datasets. The sex ratio of known sex-individuals was 1:1, except for 'All' which included the whole set of known-sex individuals (EYR: 352 females and 429 males, YTH: 289 females and 347 males, LBP: 164 females and 212 males).

| n | W-linked/Y-linked | Sex-biased | Z-linked/X-linked | Gametologs | Total |
|---|---|---|---|---|---|
| **EYR** | | | | | |
| **20** | 0 | 0 | 0 | 0 | 0 |
| **24** | 0 | 0 | 0 | 0 | 0 |
| **30** | 105 | 0 | 0 | 101 | 206 (6.5%) |
| **40** | 108 | 7 | 0 | 156 | 271 (8.5%) |
| **50** | 110 | 19 | 4 | 162 | 295 (7.9%) |
| **100** | 110 | 80 | 235 | 224 | 649 (20.5%) |
| **200** | 110 | 364 | 458 | 242 | 1174 (37%) |
| **400** | 110 | 1728 | 540 | 245 | 2623 (82.8%) |
| **All** | 110 | 2142 | 668 | 248 | 3168 (100%) |
| **YTH** | | | | | |
| **20** | 0 | 0 | 0 | 0 | |
| **24** | 0 | 0 | 0 | 0 | |
| **30** | 46 | 0 | 0 | 0 | 46 (1.6%) |
| **40** | 48 | 2 | 0 | 3 | 53 (1.9%) |
| **50** | 48 | 5 | 0 | 3 | 56 (2%) |
| **100** | 48 | 9 | 328 | 3 | 388 (13.6) |
| **200** | 47 | 197 | 702 | 3 | 949 (33.2%) |
| **400** | 48 | 1080 | 853 | 4 | 1985 (69.5%) |
| **All** | 48 | 1899 | 905 | 4 | 2856 (100%) |
| **LBP** | | | | | |
| **20** | 0 | 0 | 0 | 0 | 0 |
| **24** | 0 | 0 | 0 | 0 | 0 |
| **30** | 1 | 0 | 0 | 0 | 1 (1.6%) |
| **40** | 1 | 0 | 0 | 0 | 1 (1.6%) |
| **50** | 1 | 0 | 0 | 0 | 1 (1.6%) |
| **100** | 1 | 0 | 10 | 0 | 11 (17.2%) |
| **200** | 1 | 1 | 39 | 0 | 41 (64.1%) |

| All | 1 | 2 | 60 | 1 | 64 (100%) |
|-----|---|---|----|----|-----------|

**REFERENCES**

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Deakin, J. E., Chaumeil, J., Hore, T. A., & Marshall Graves, J. A. (2009). Unravelling the evolutionary origins of X chromosome inactivation in mammals: insights from marsupials and monotremes. *Chromosome Research*, *17*, 671-685.

Pavlova, A, Selwood, P, Harrisson, KA, Murray, N, Quin, B, Menkhorst, P, Smales, I, Sunnucks, P (2014) Integrating phylogeography and morphometrics to assess conservation merits and inform conservation strategies for an endangered subspecies of a common bird species. *Biological Conservation* 174, 136–146.