

Analysing Genomic Data with **dartRverse**: Accessible Tools for Conservation





MONASH University

Diana A Robledo-Ruiz (she/her)



MONASH University

Hypothetical scenario

Diana A Robledo-Ruiz (she/her)



MONASH University

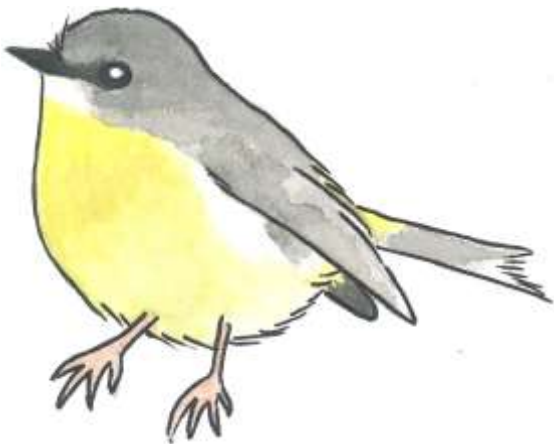
Hypothetical scenario





MONASH University

Hypothetical scenario



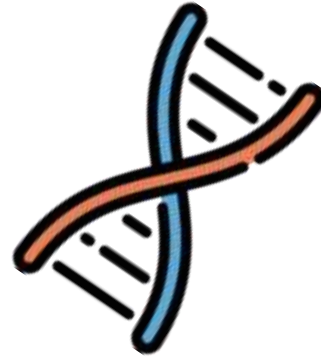
- Genetic diversity
- Genetic structure
- Parentage analyses



- SNPs from DArT, RAD

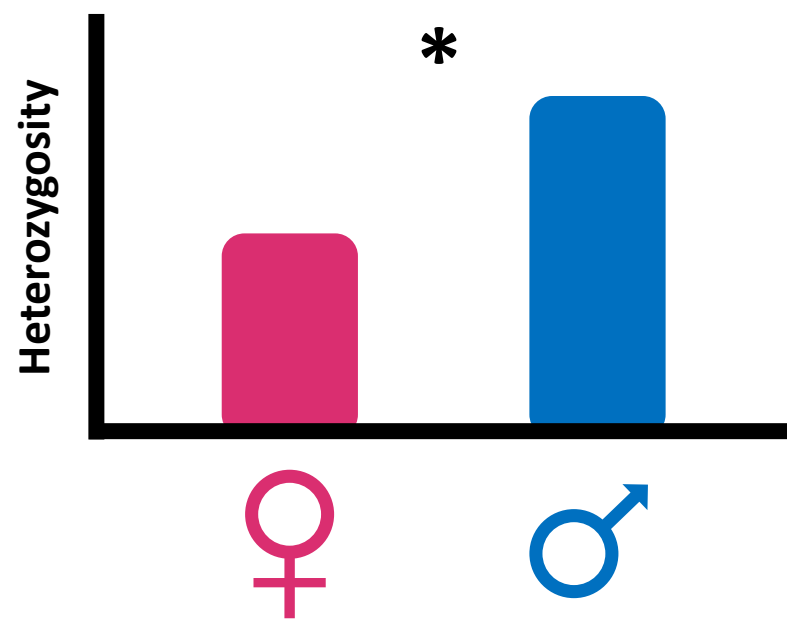


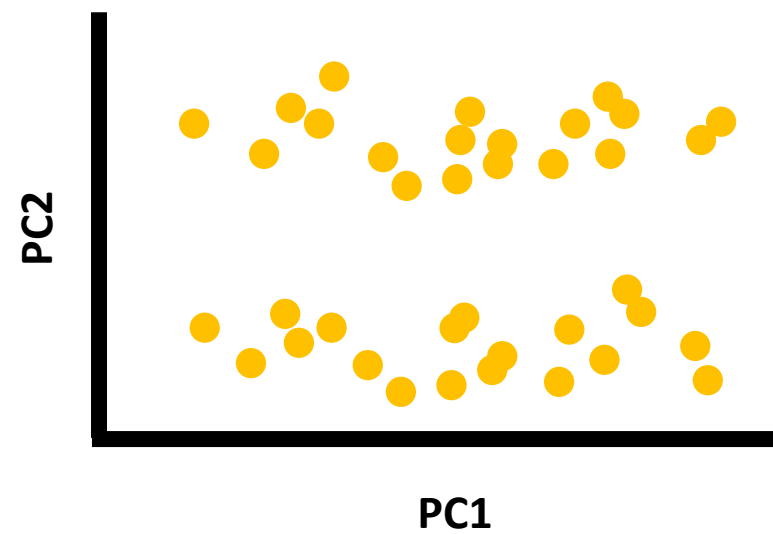
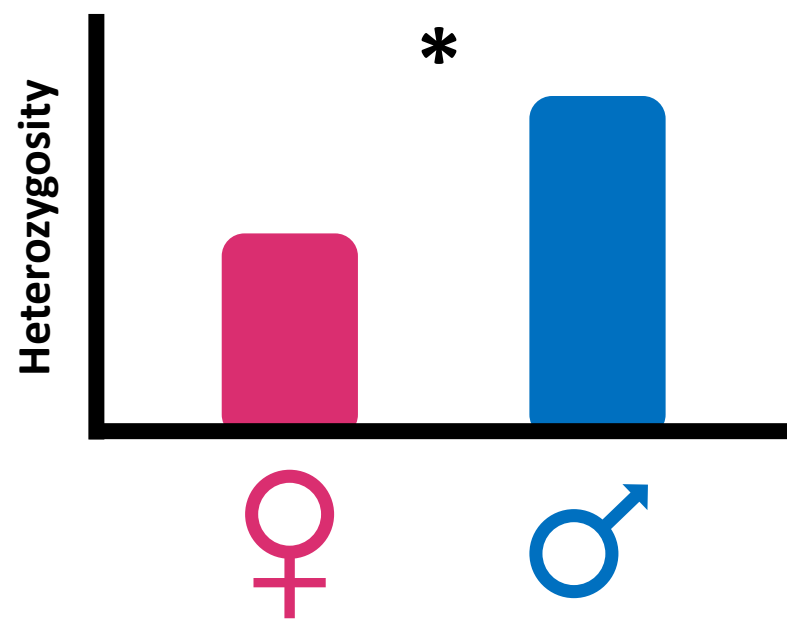
- SNPs from DArT, RAD
- Filtering

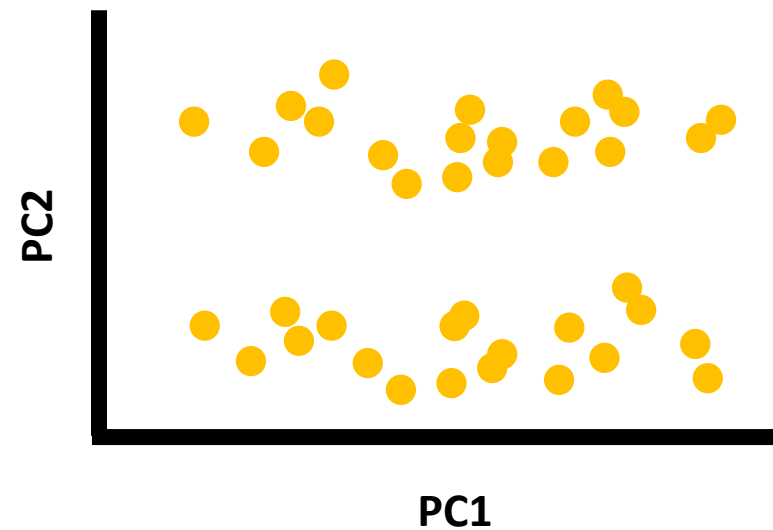
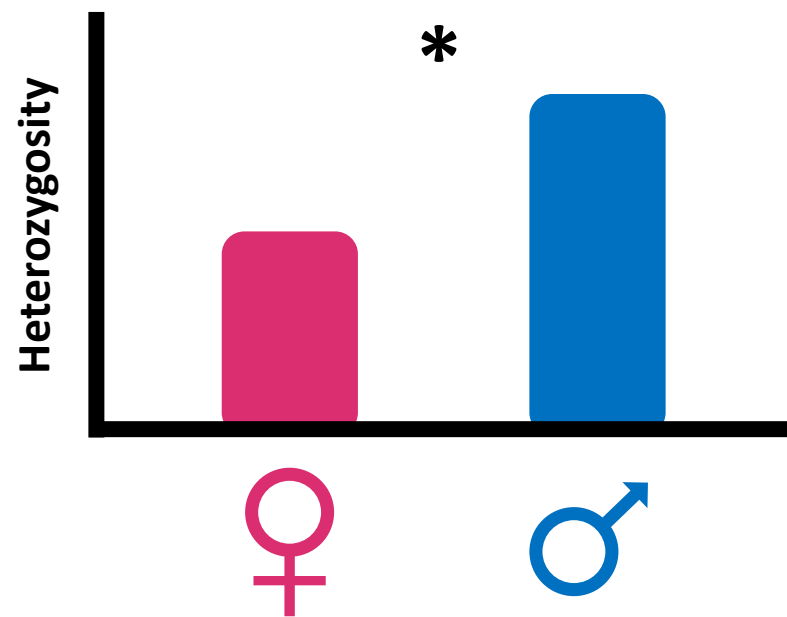


→ SNP quality: read depth
missing data
MAC

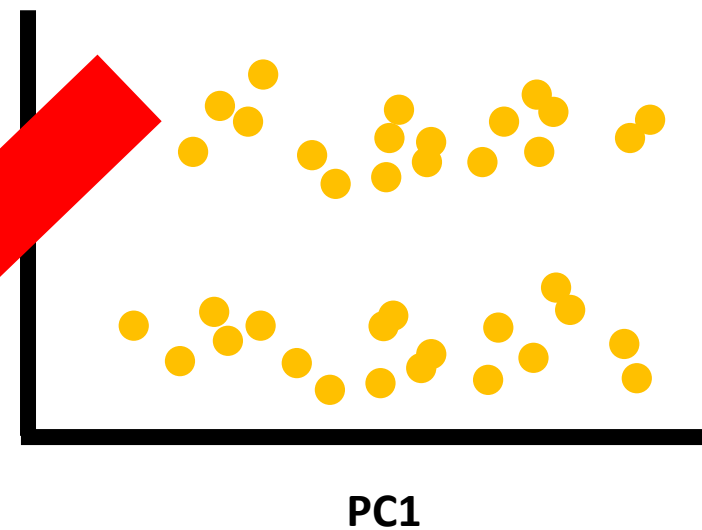
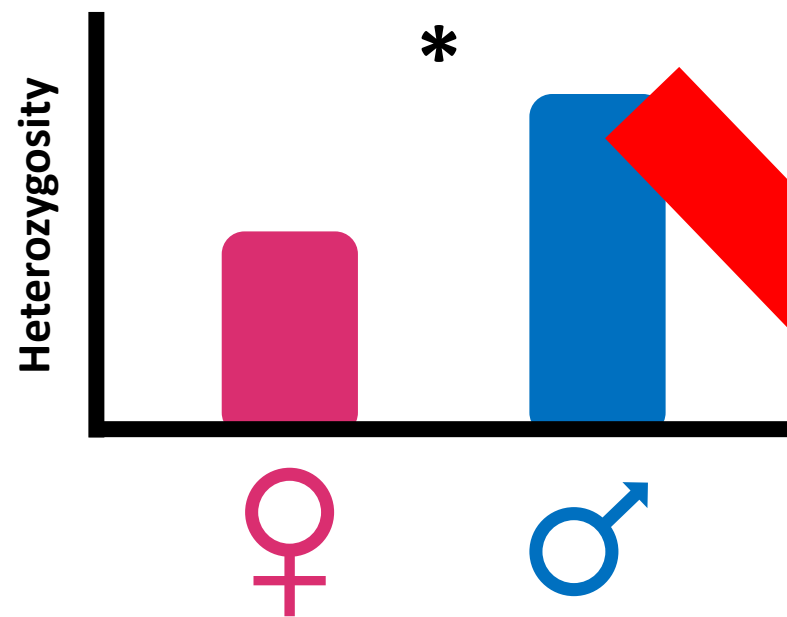
→ Individual quality: missing data



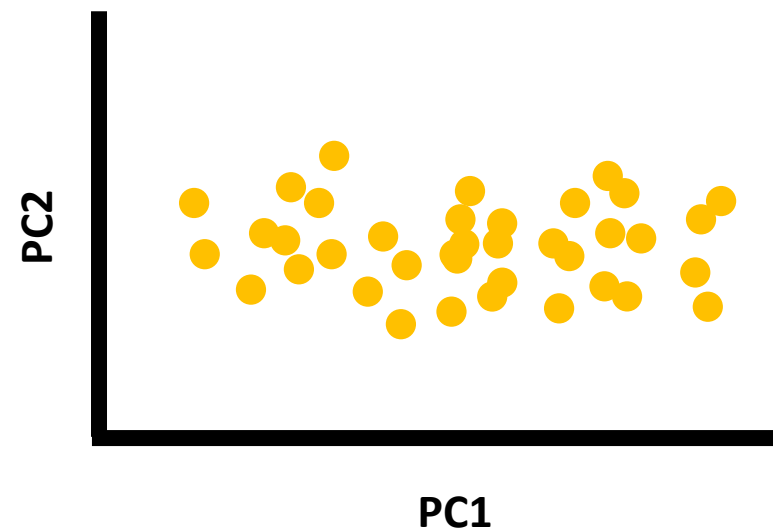
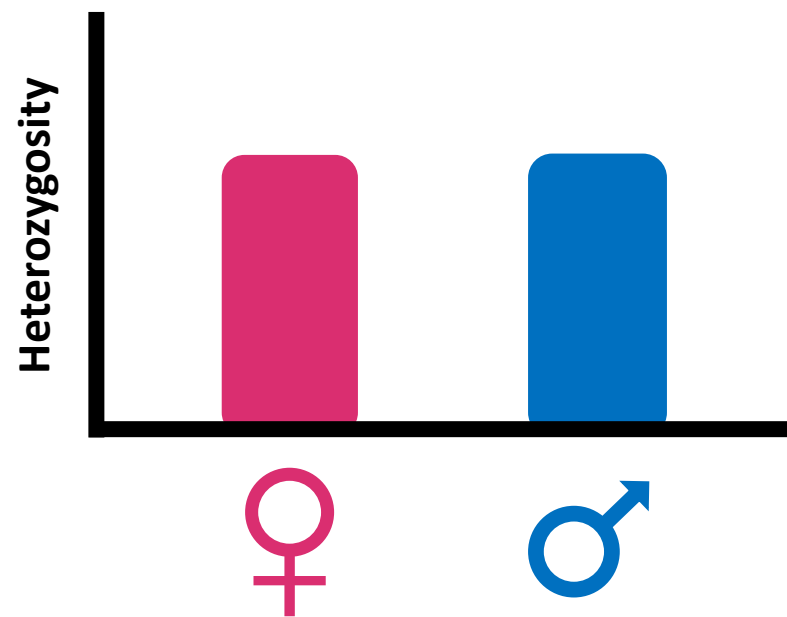




91 parent-offspring



91 parent-offspring



104 parent-offspring

What did you miss?

Filtering

- SNP quality: read depth
missing data
MAC
- Individual quality: missing data

Filtering

- Sex-linked loci
- SNP quality: read depth
missing data
MAC
- Individual quality: missing data

Filtering

- Sex-linked loci
- SNP quality: read depth
missing data
MAC
- Individual quality: missing data

Popgen analyses assume autosomal loci!

Sex-linked loci and sex chromosomes



X



Y



X



X





X

Y



X

X





X

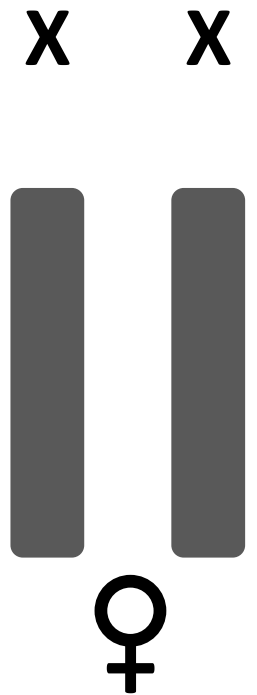
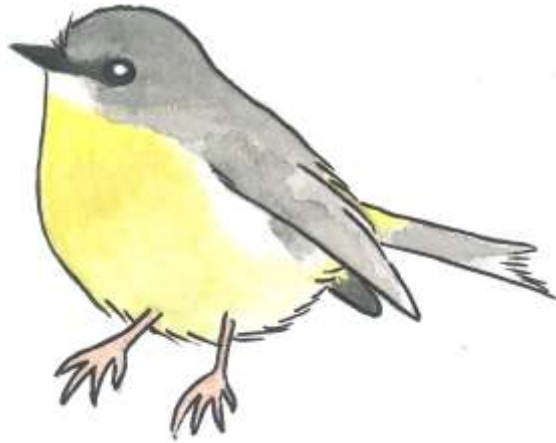
Y

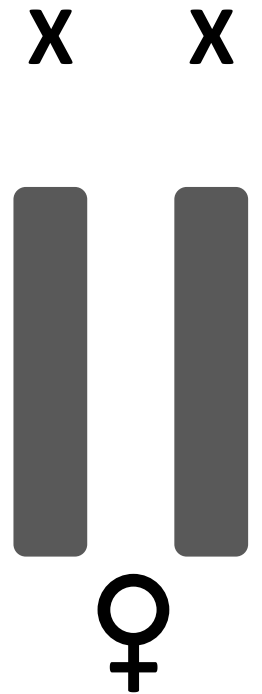
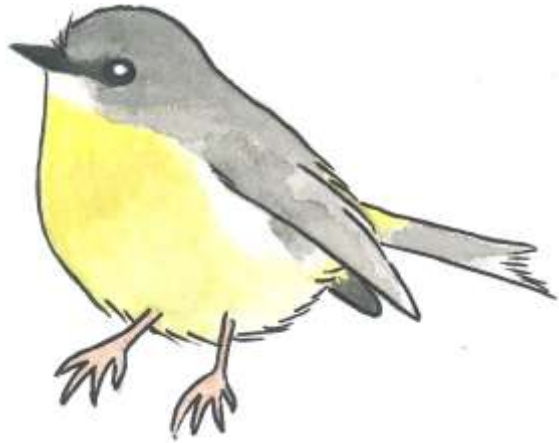


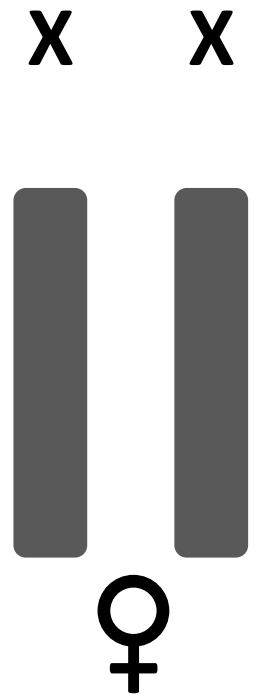
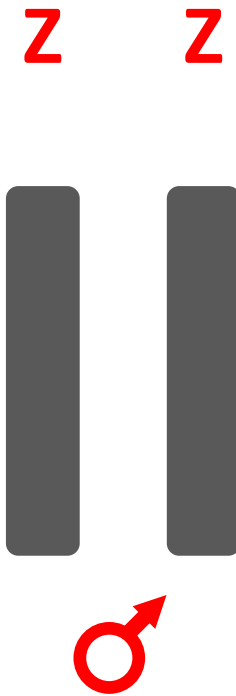
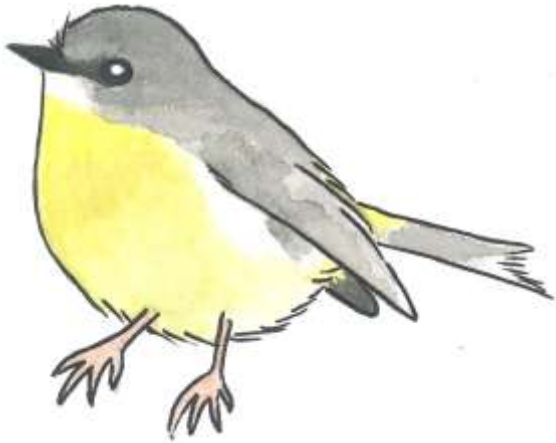
X

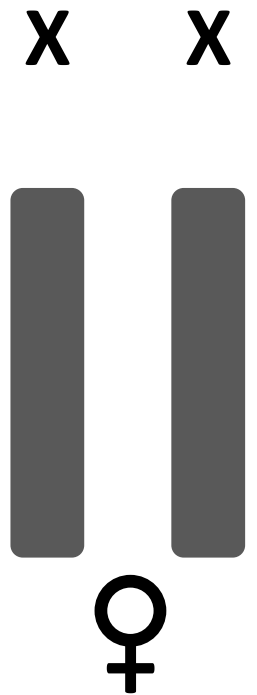
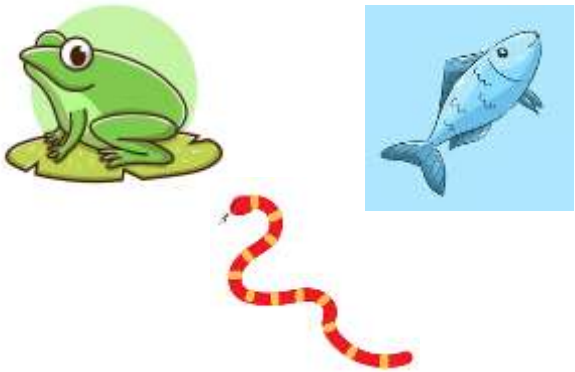
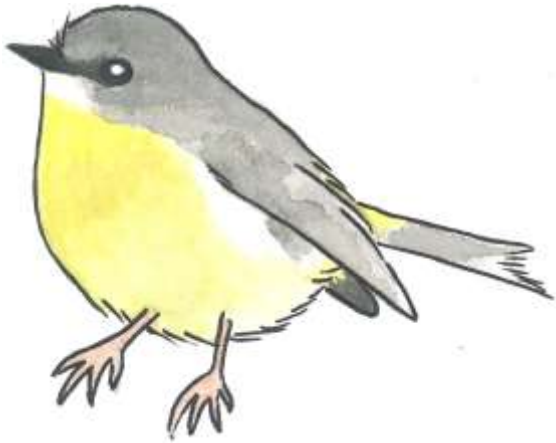
X

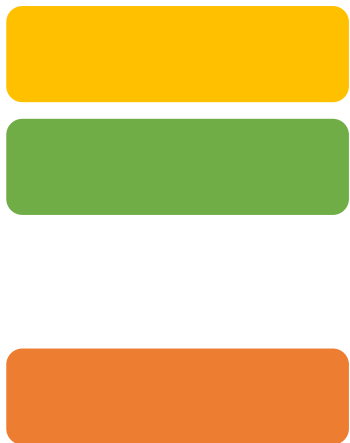


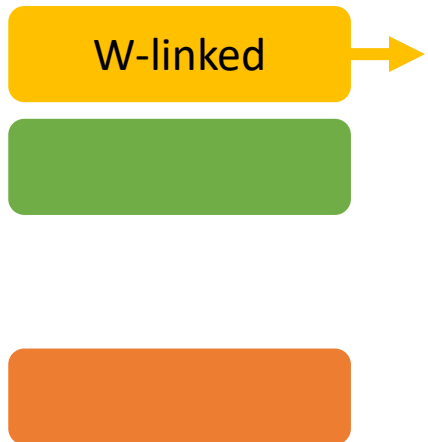


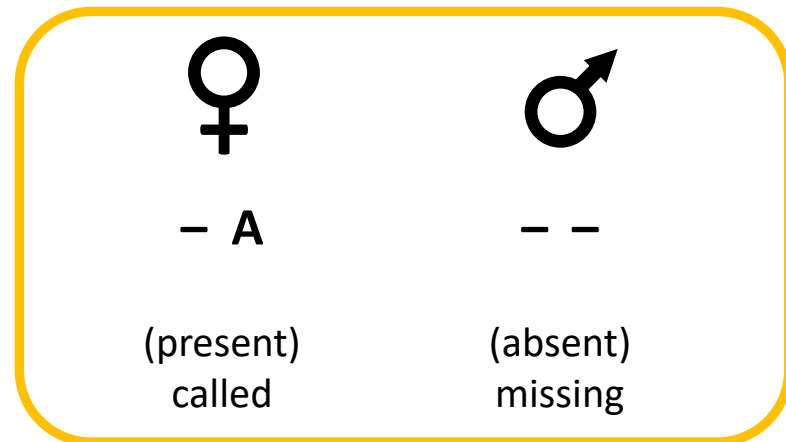
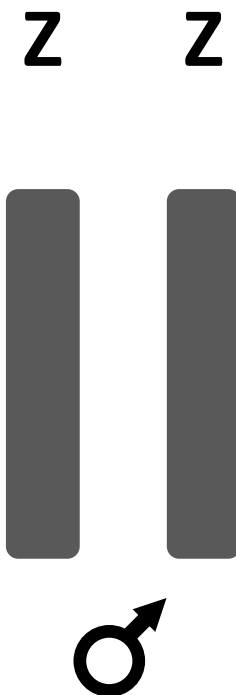
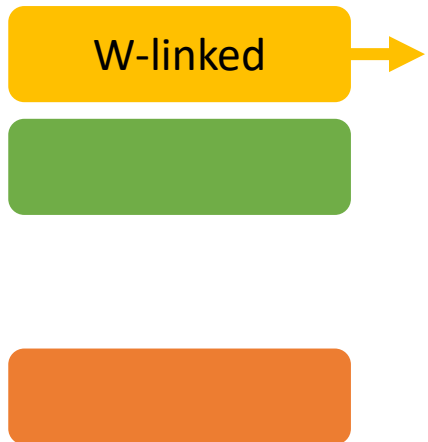












W-linked



Z

W



Z

Z



- A

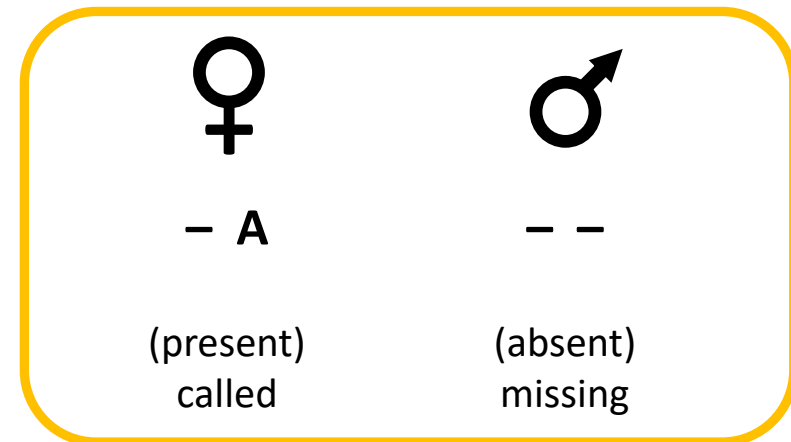
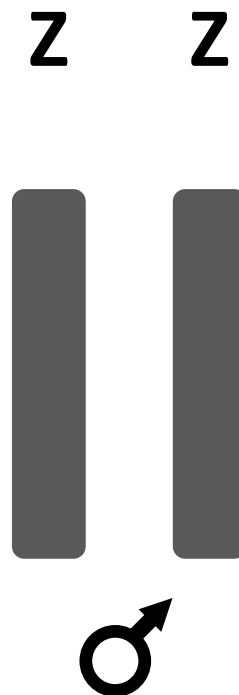
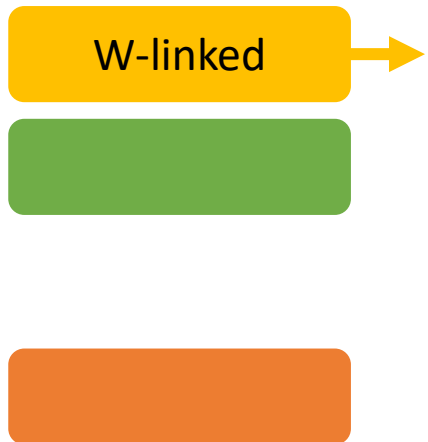
(present)
called

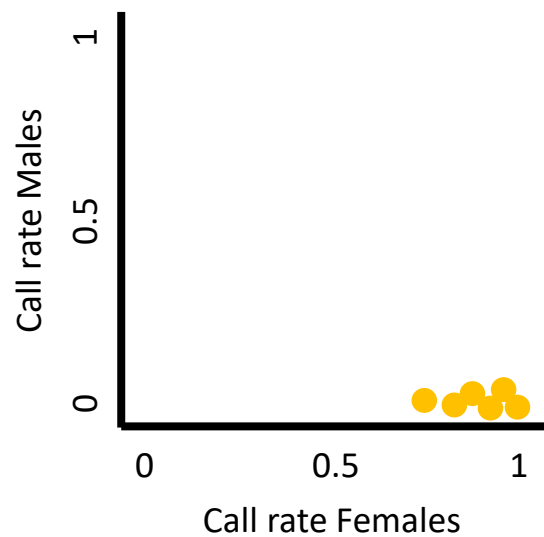
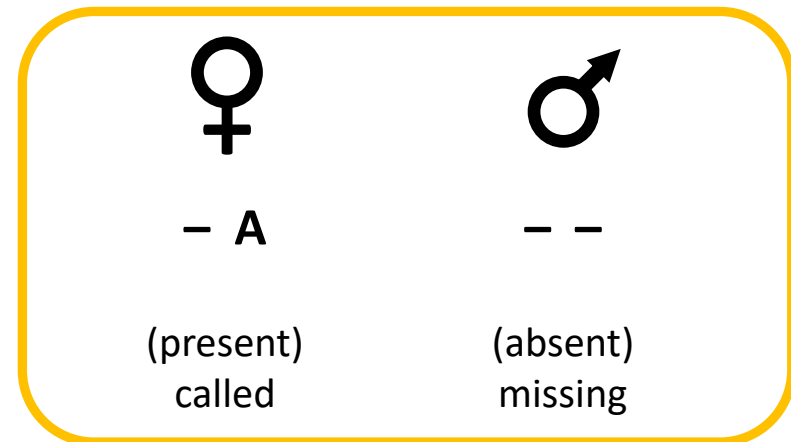
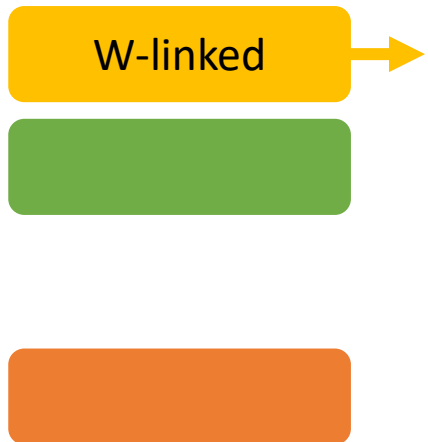


- -

(absent)
missing

0 0.5 1
Call rate Females





W-linked



Z

W



Z

Z



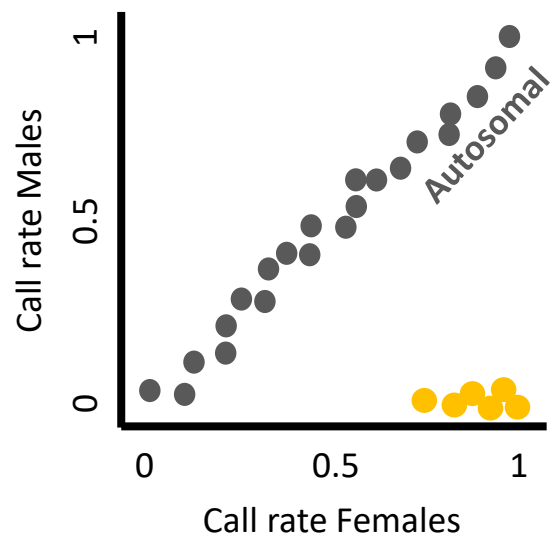
- A

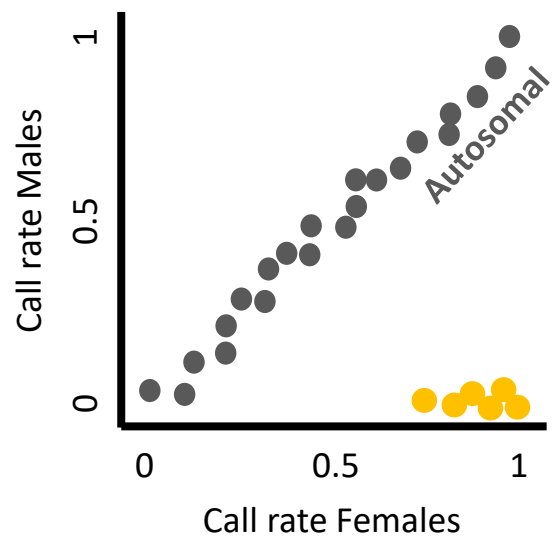
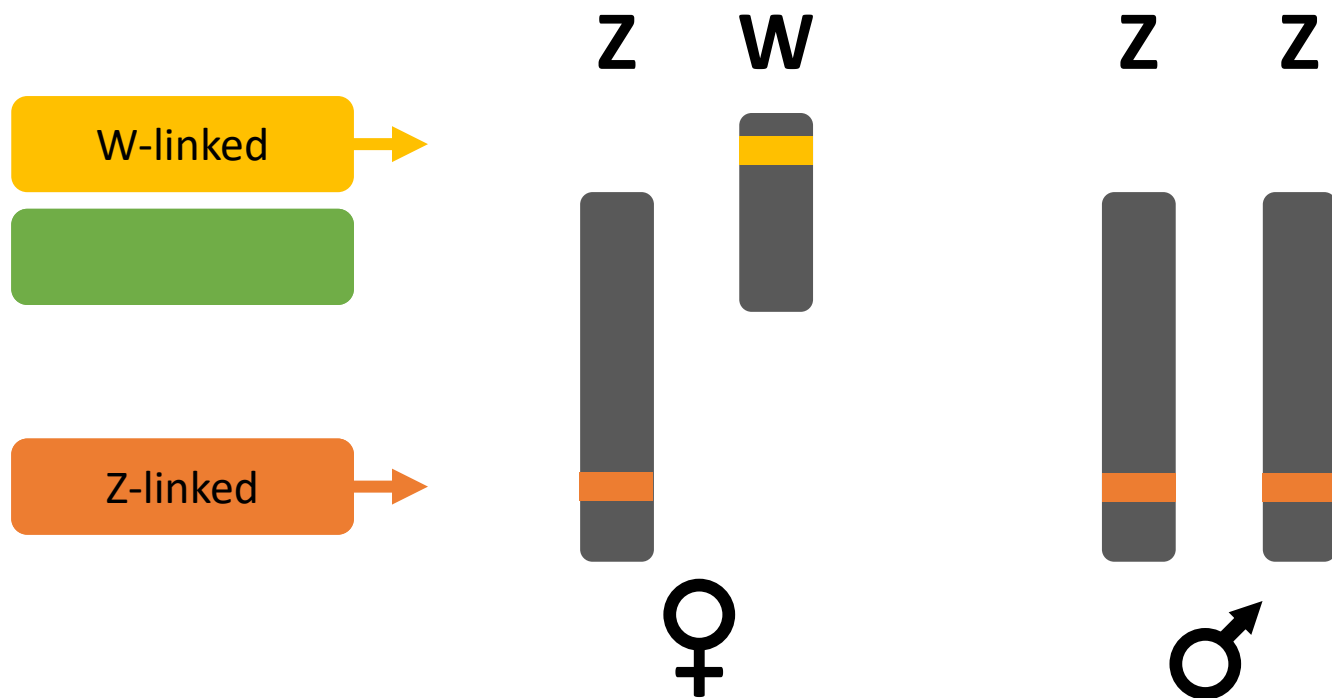
(present)
called

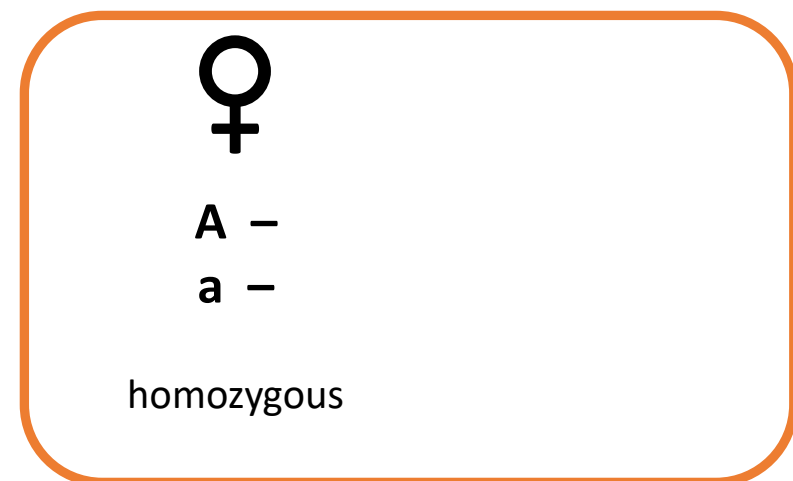
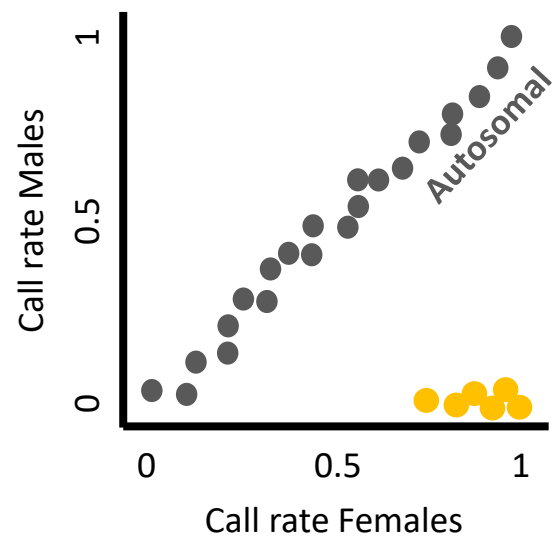
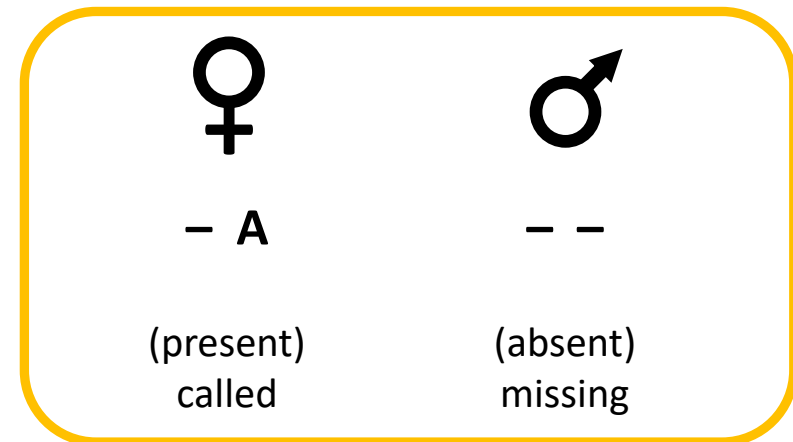
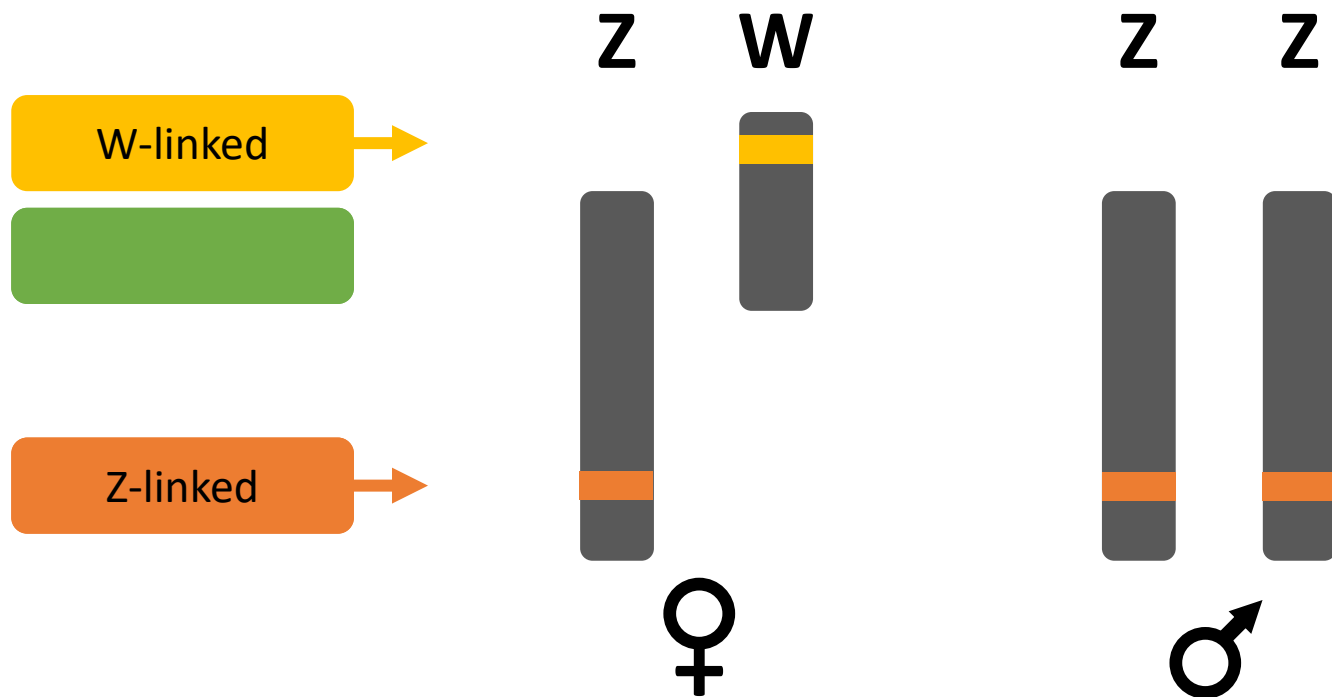


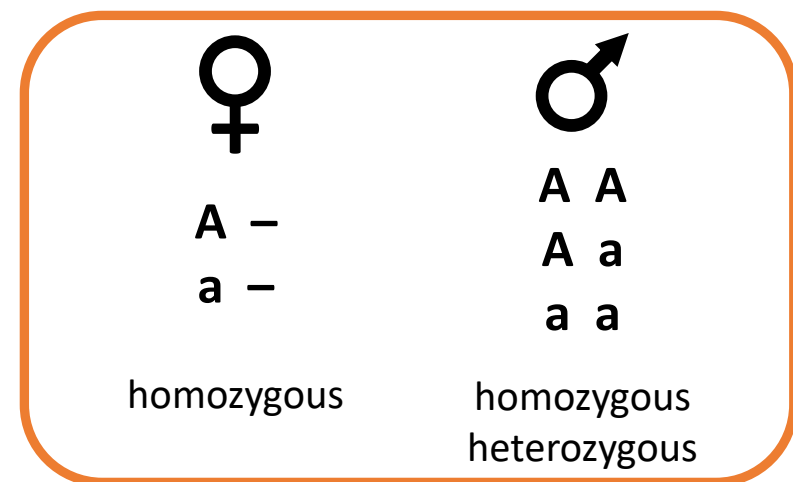
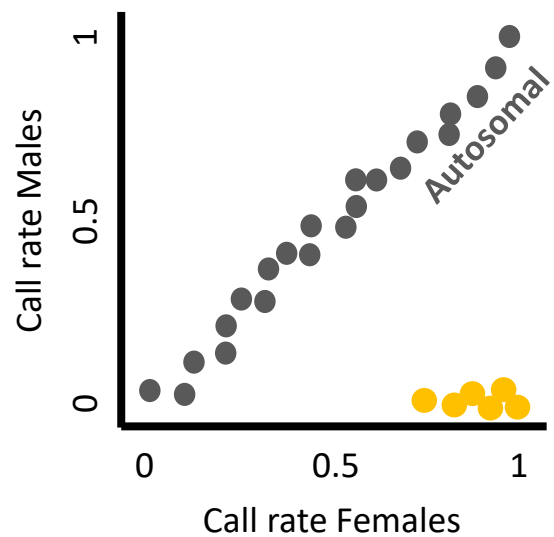
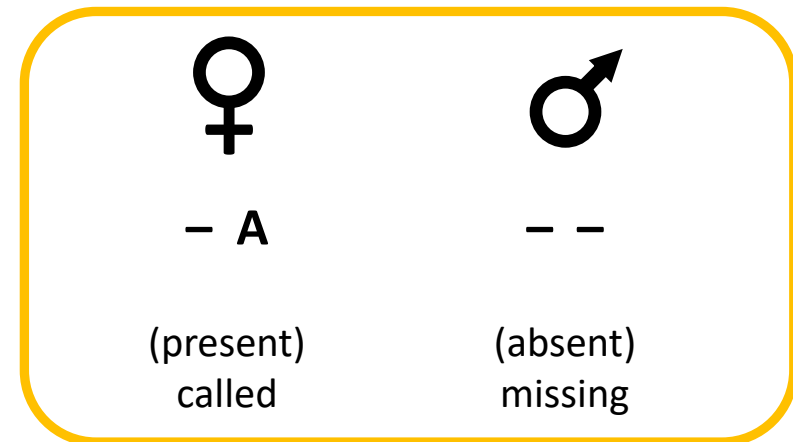
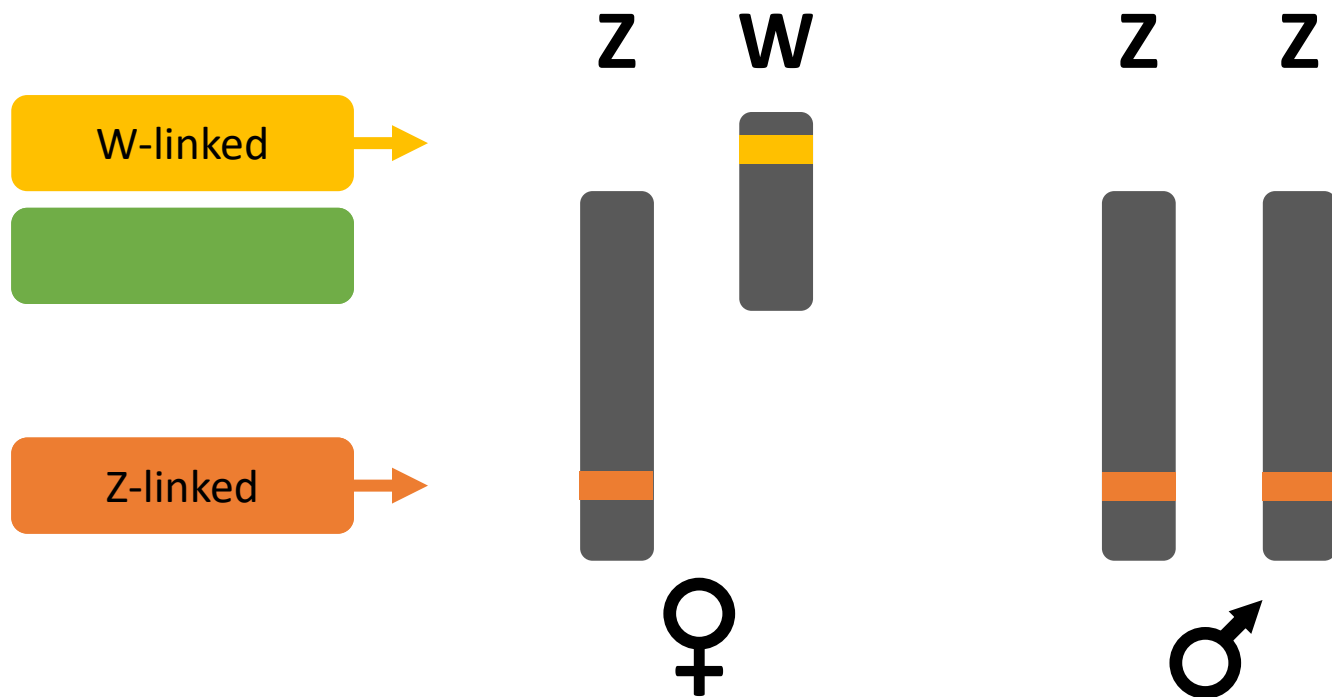
- -

(absent)
missing





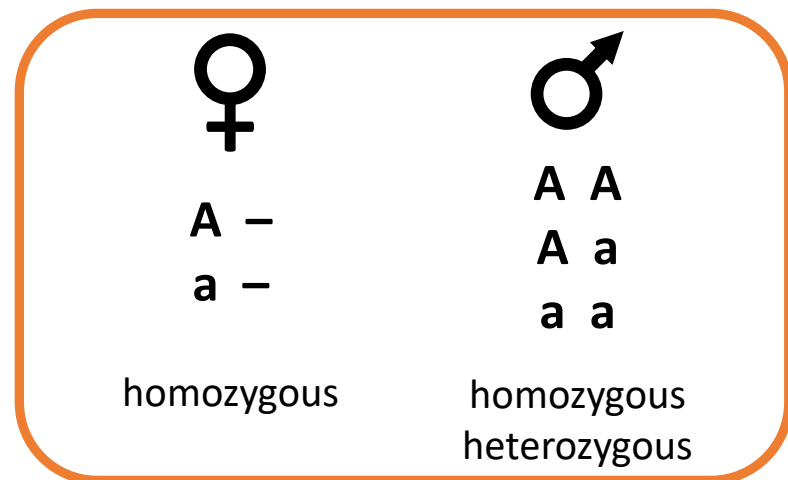
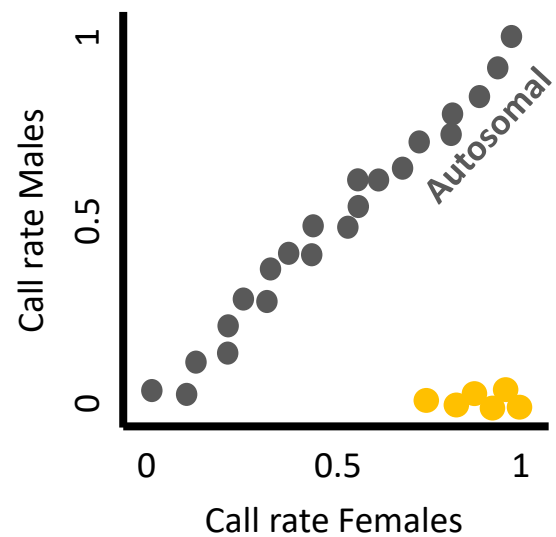
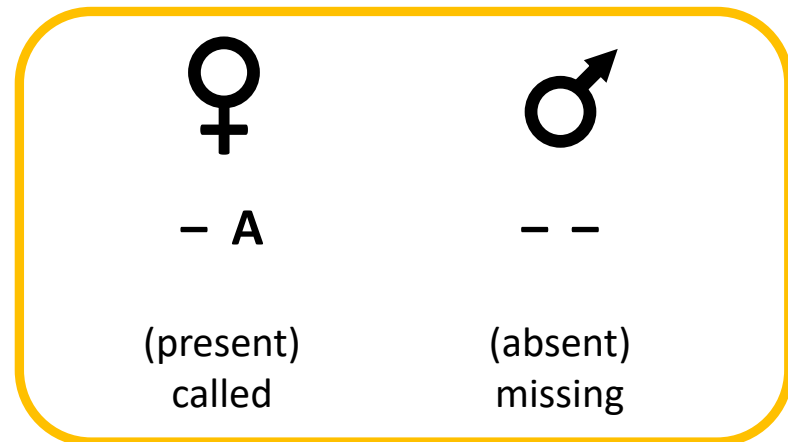
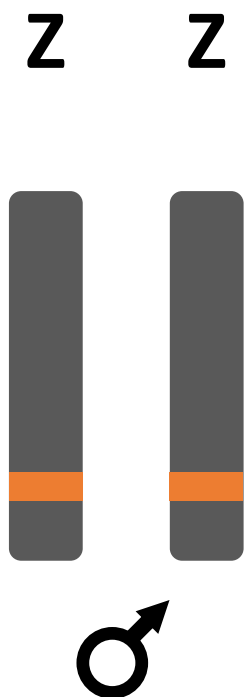
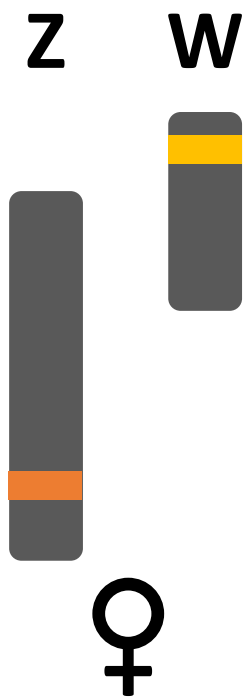


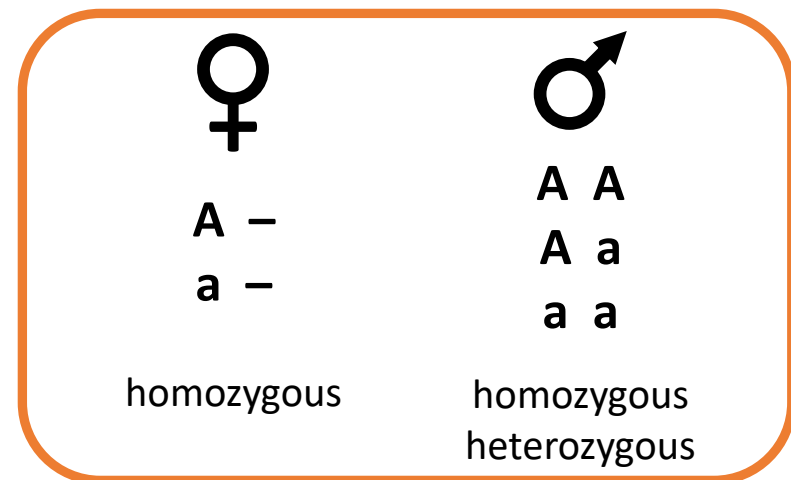
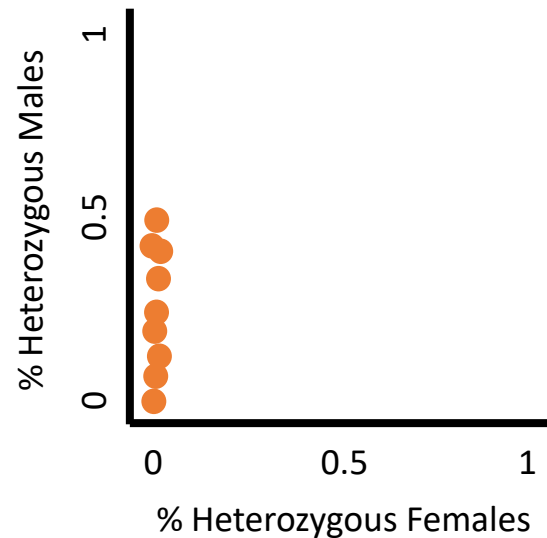
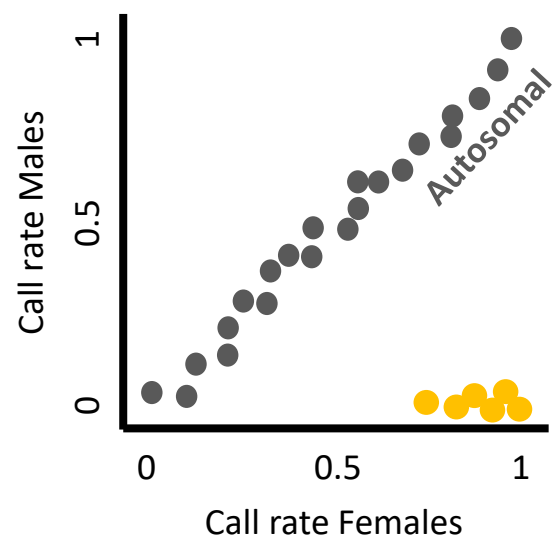
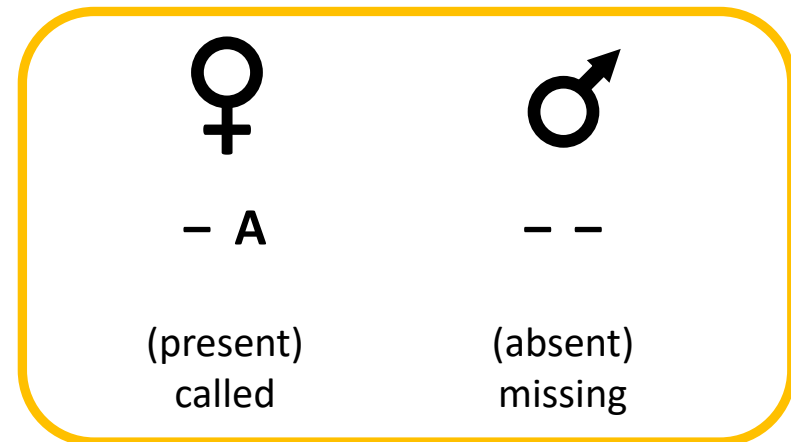
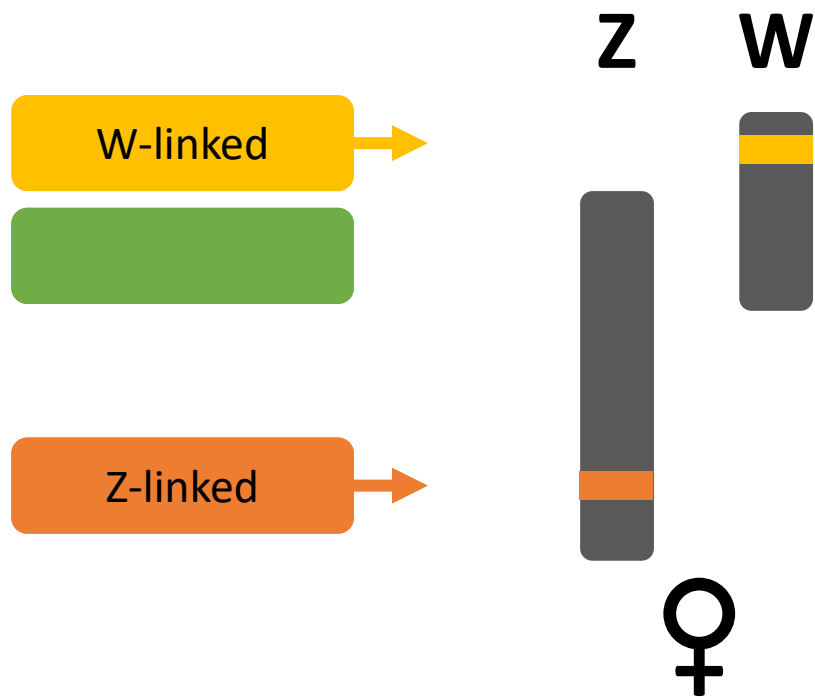


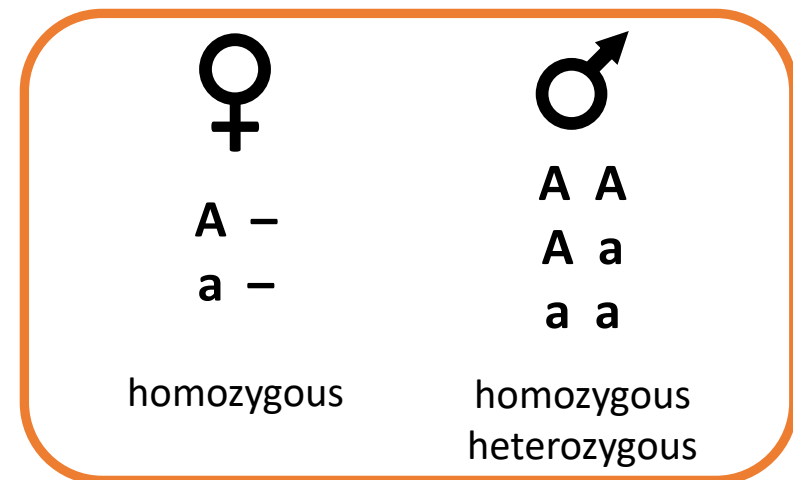
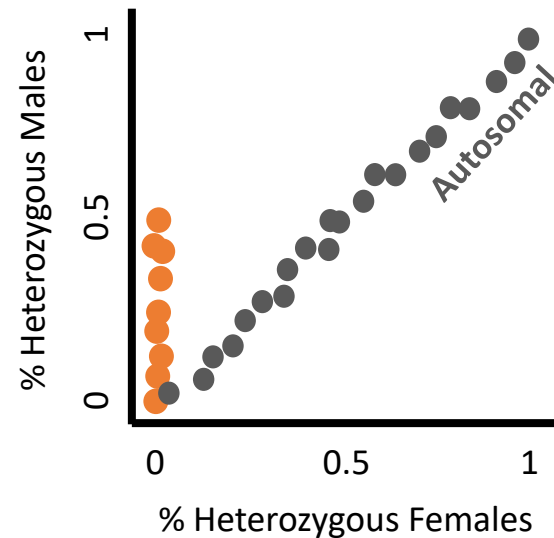
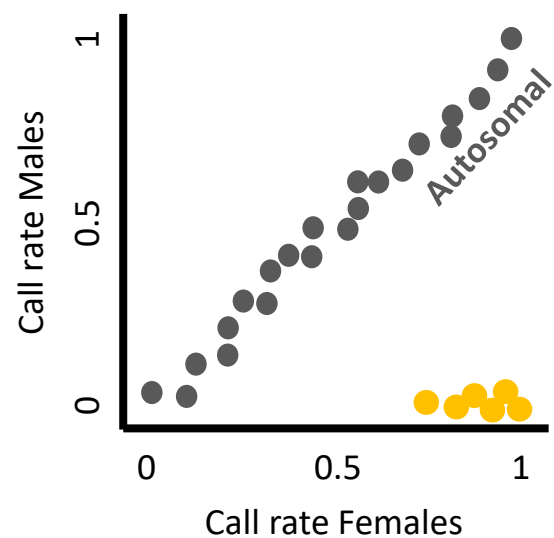
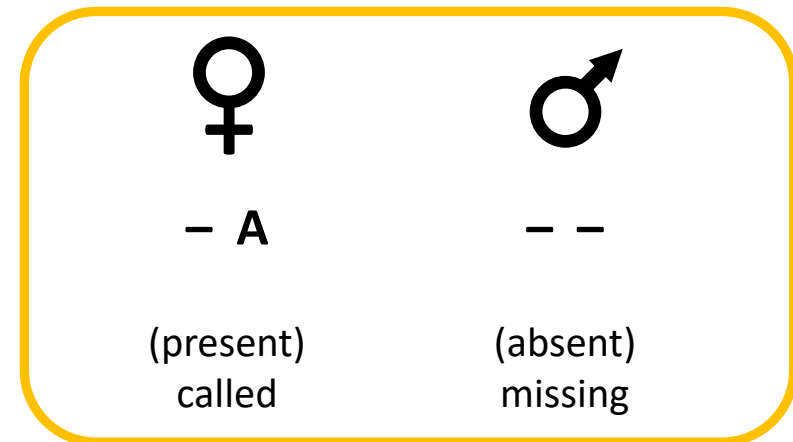
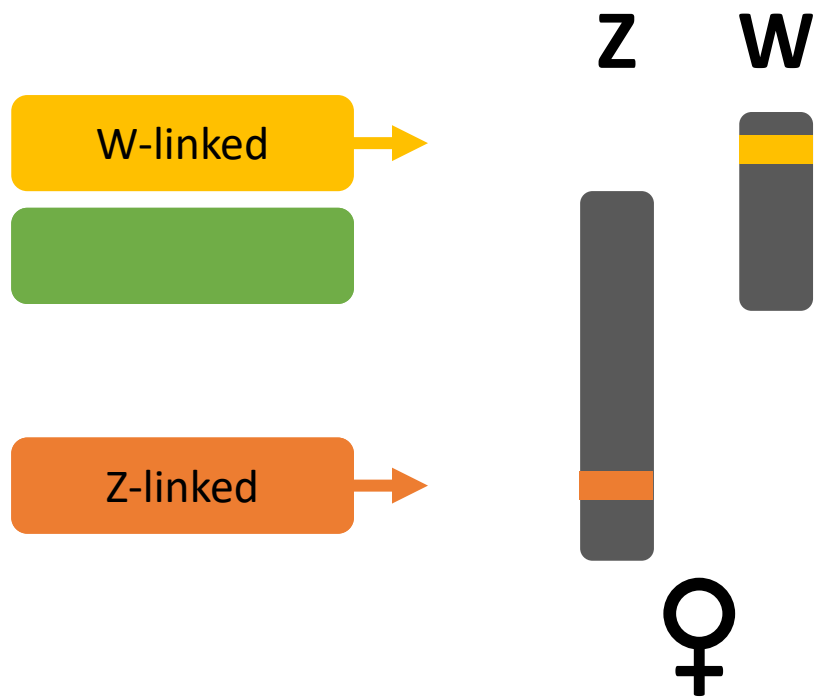
W-linked →

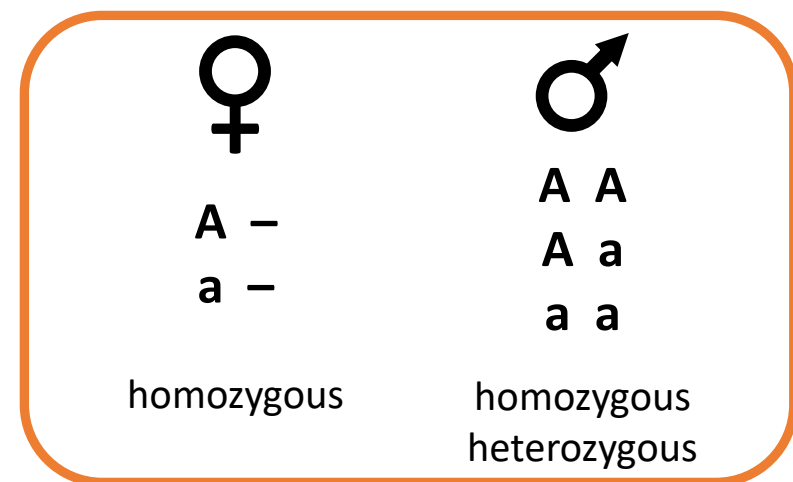
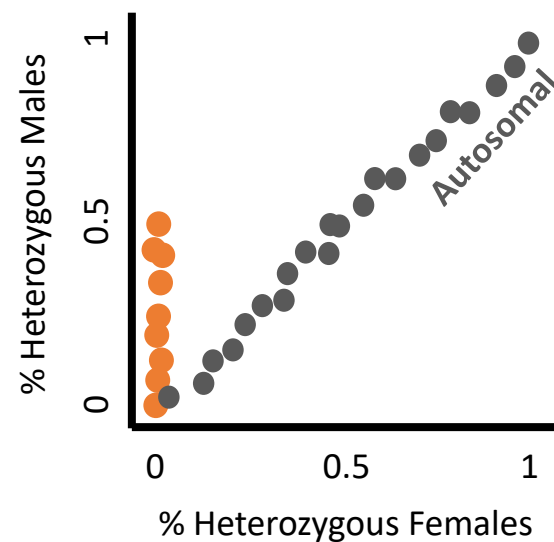
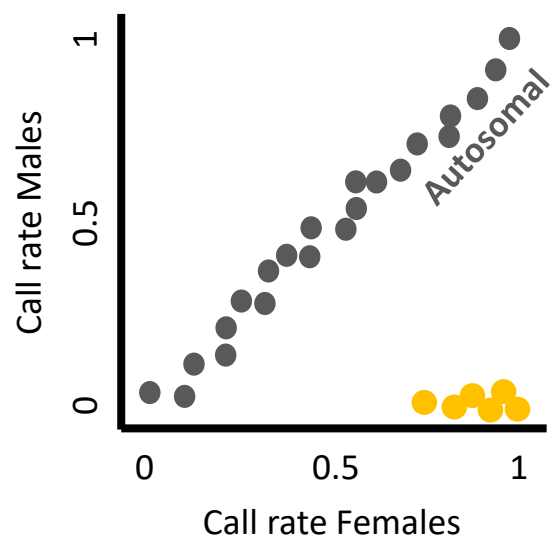
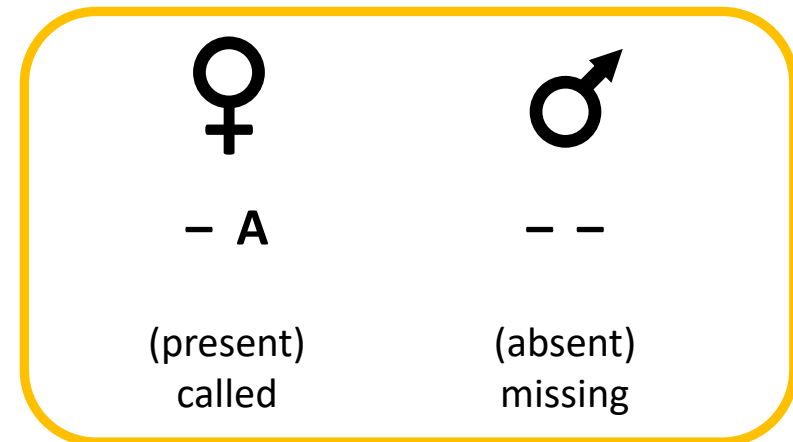
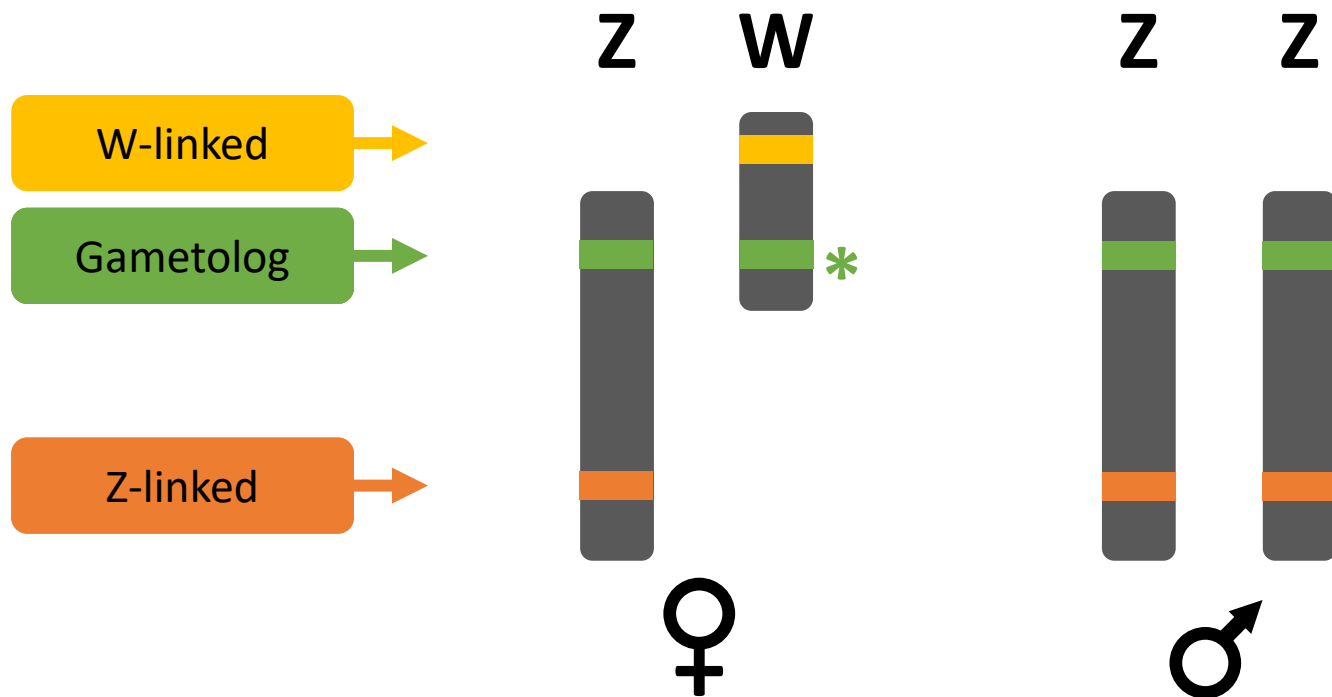


Z-linked →





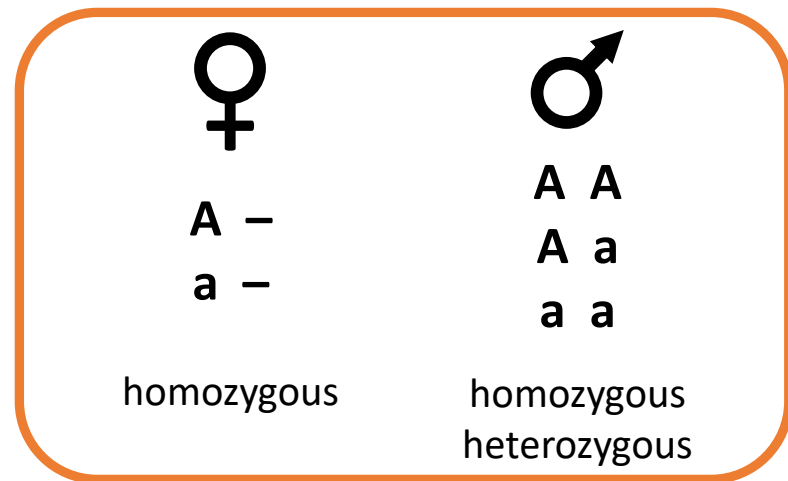
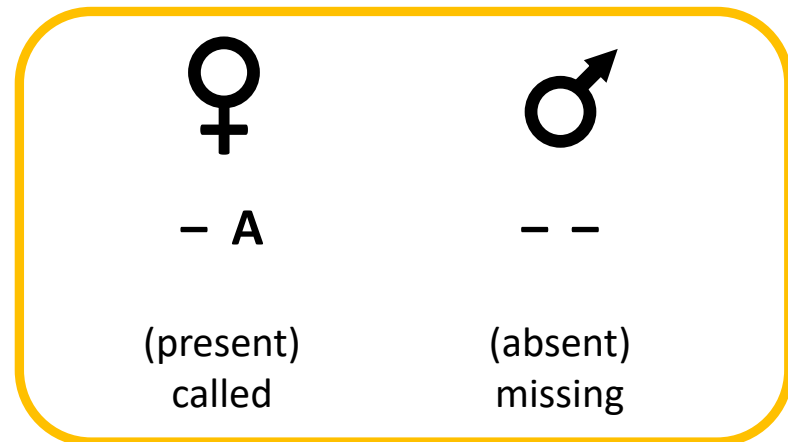
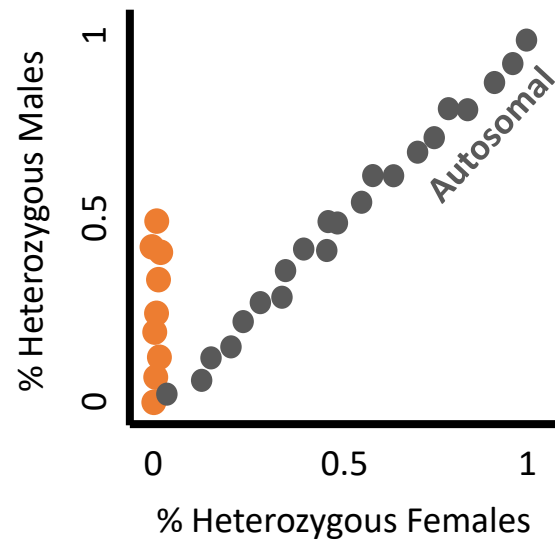
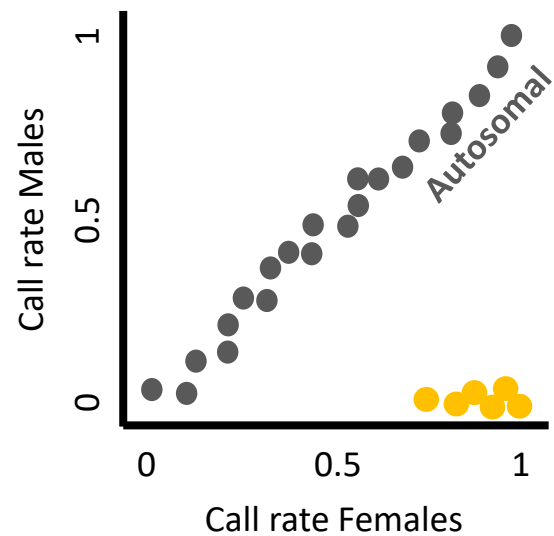
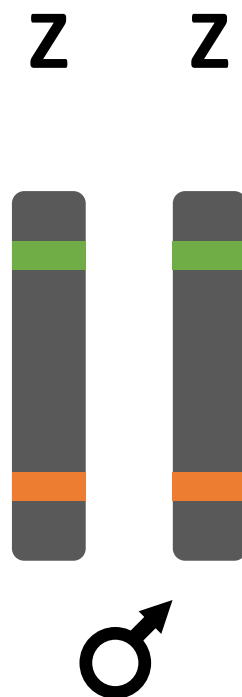




W-linked →

Gametolog →

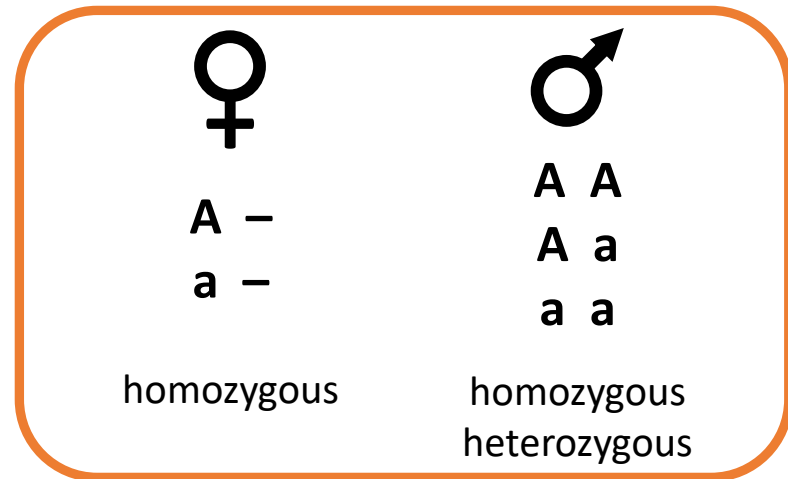
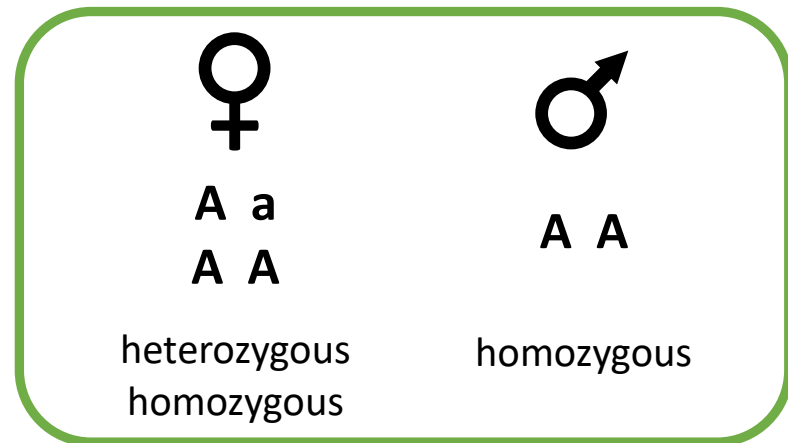
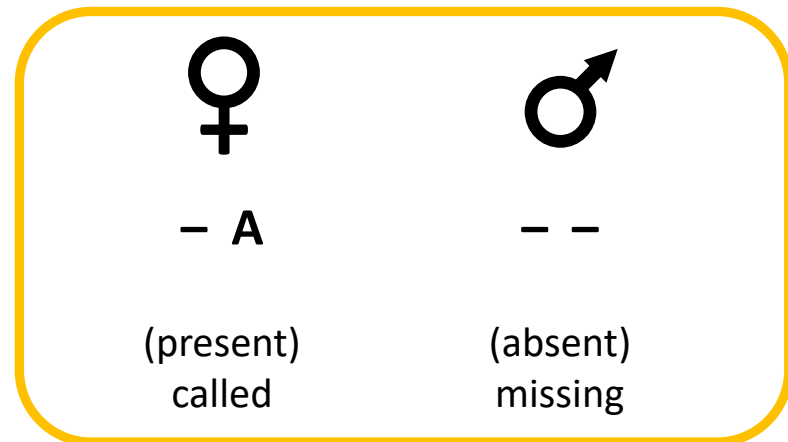
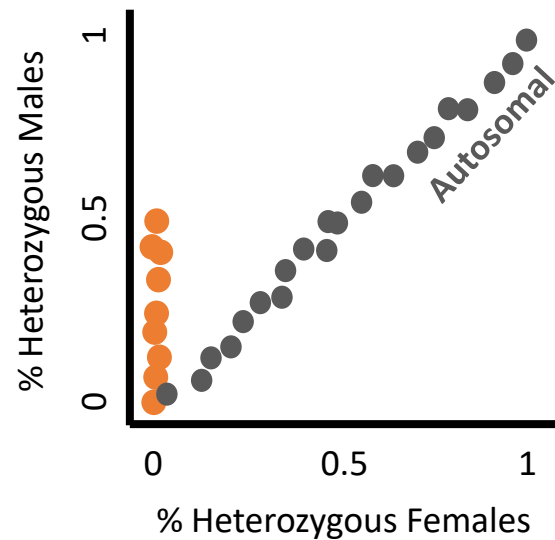
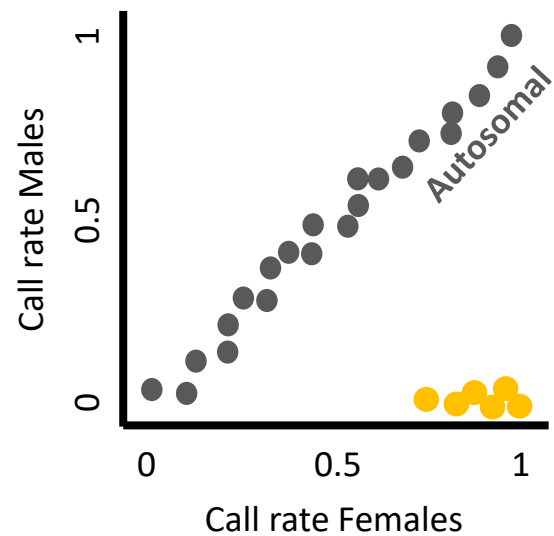
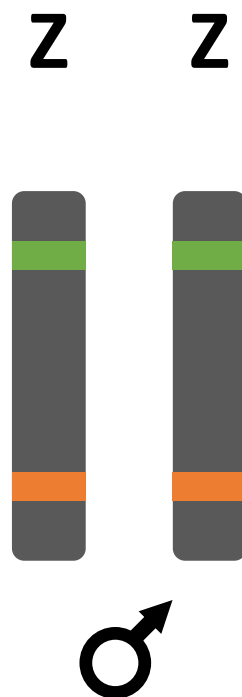
Z-linked →



W-linked →

Gametolog →

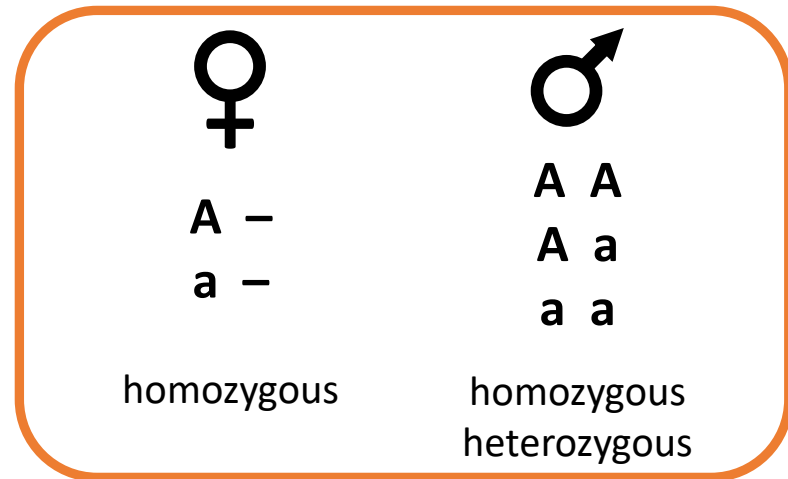
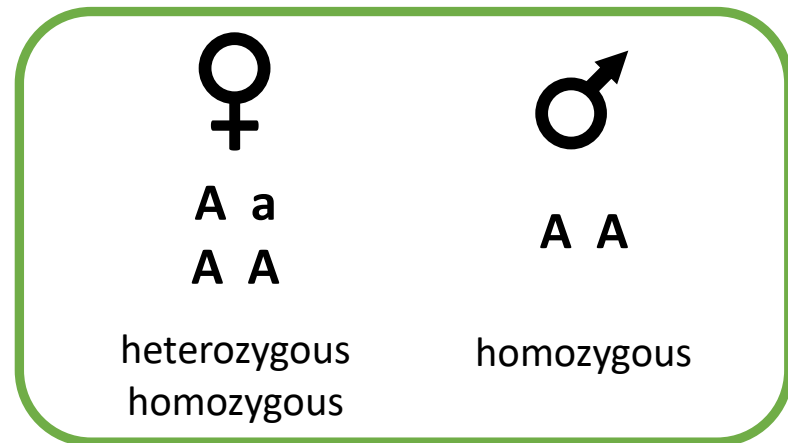
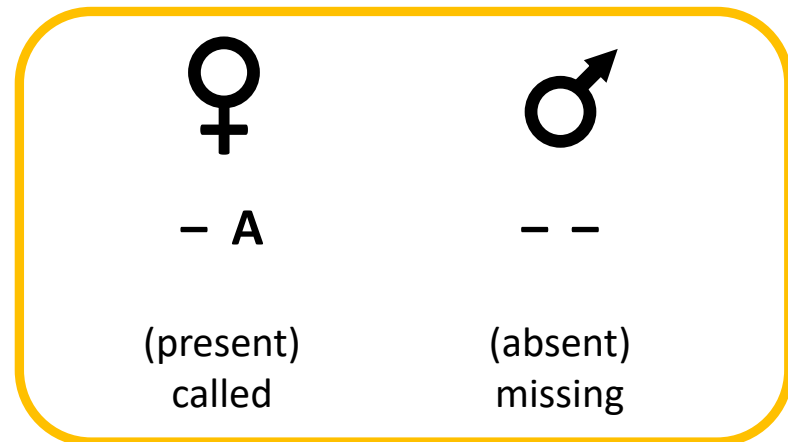
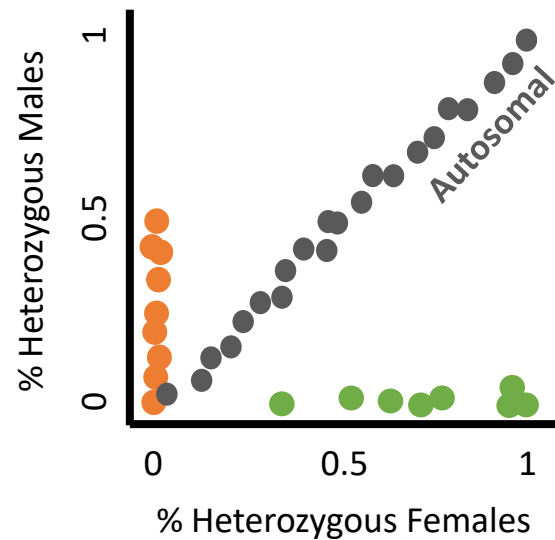
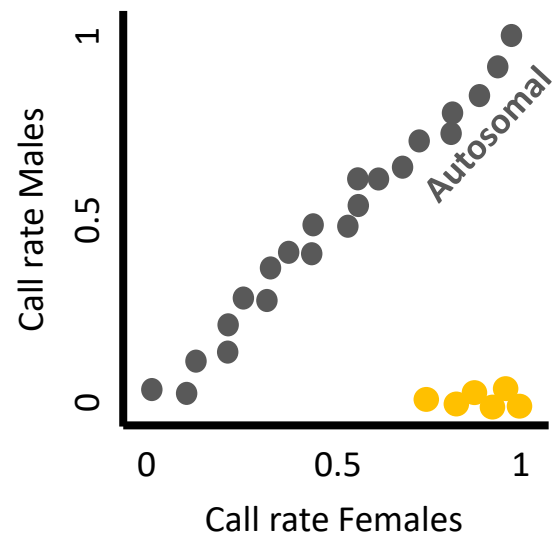
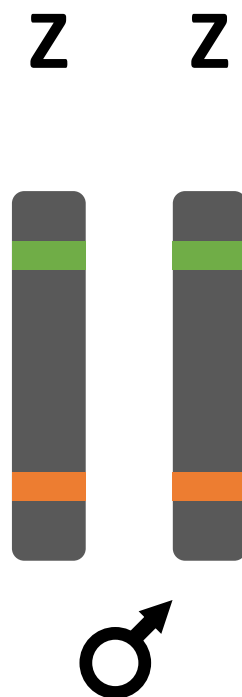
Z-linked →

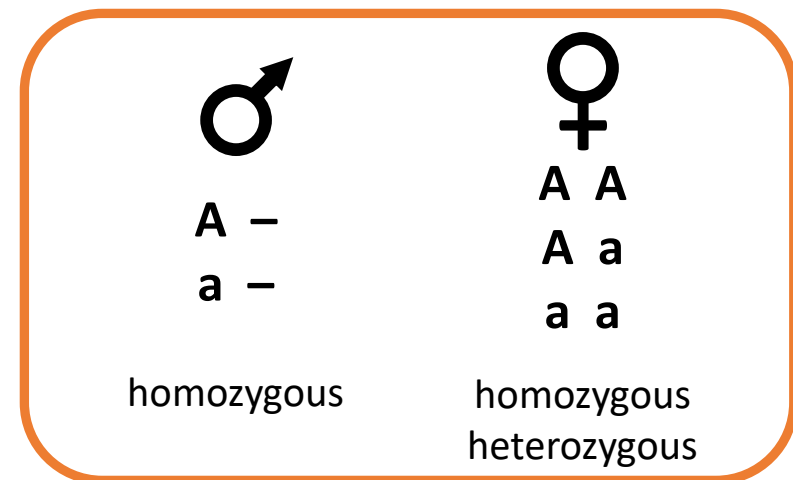
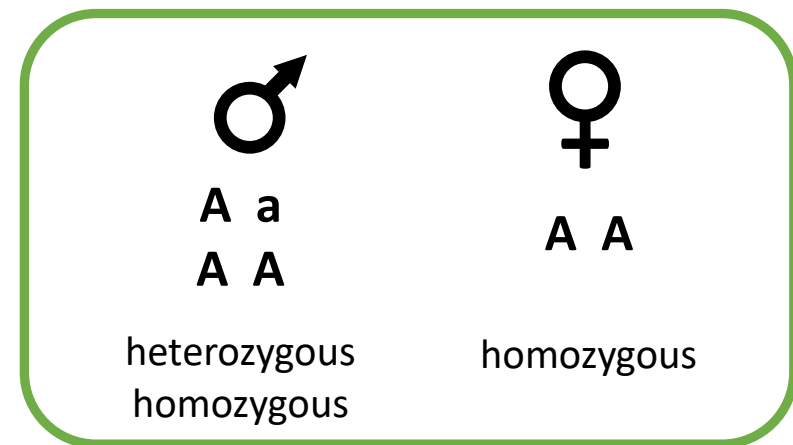
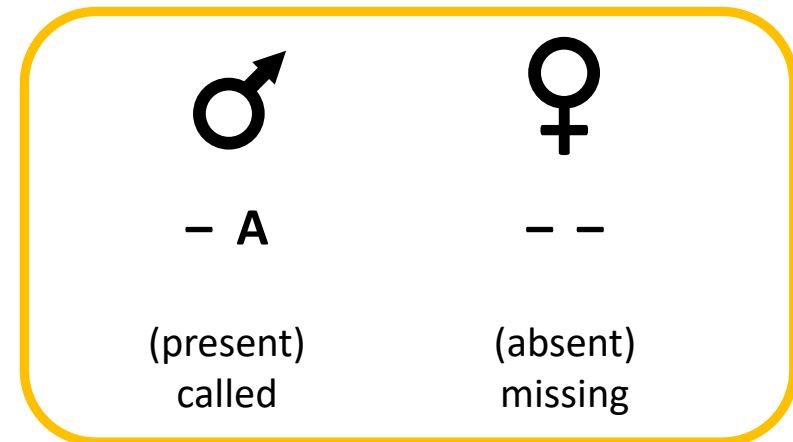
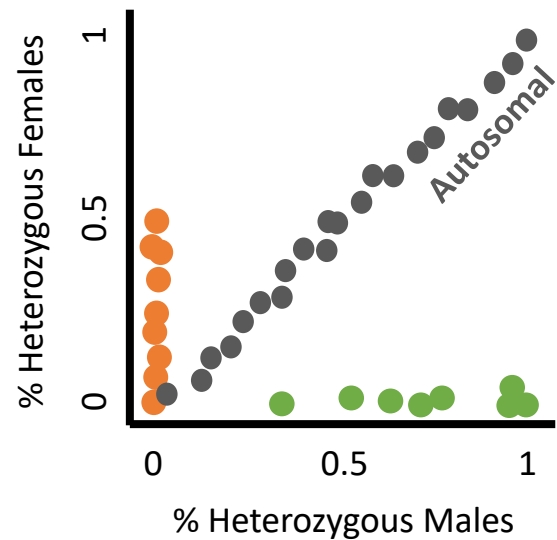
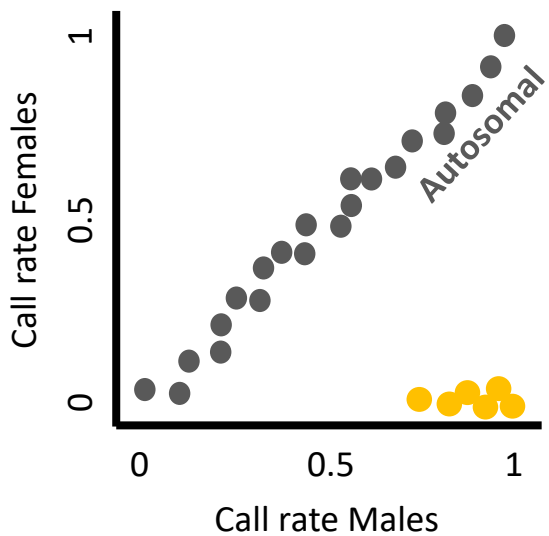
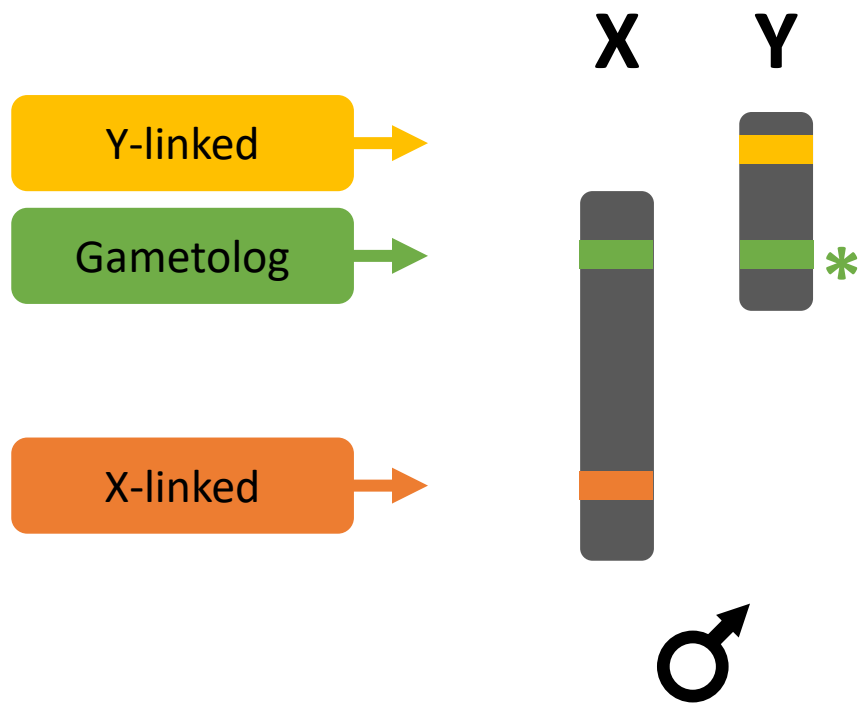


W-linked →

Gametolog →

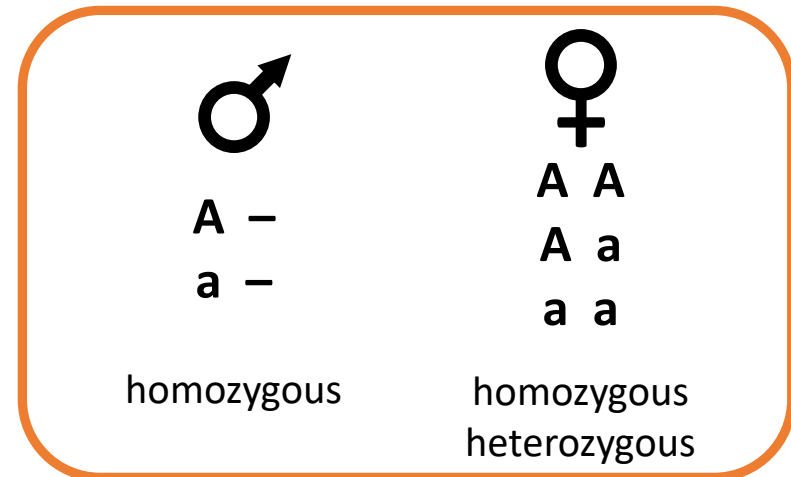
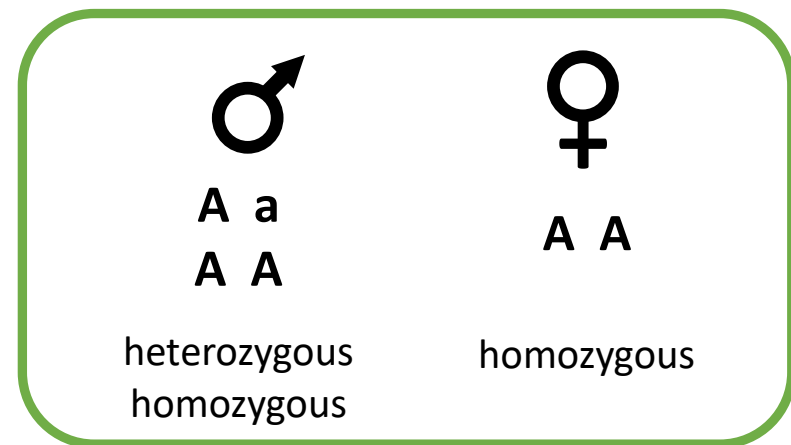
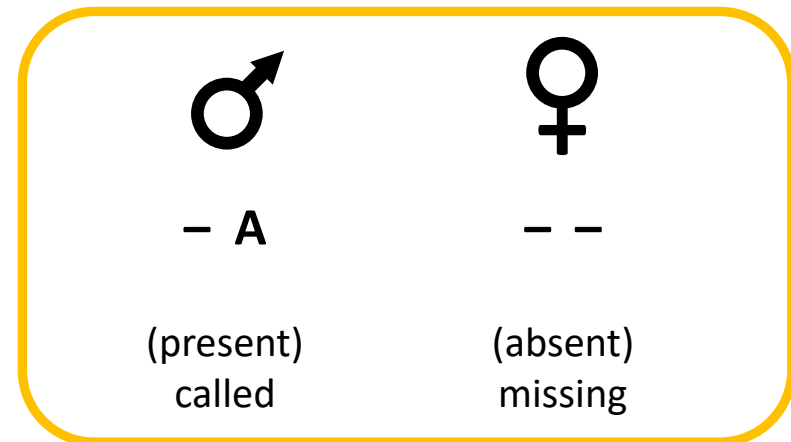
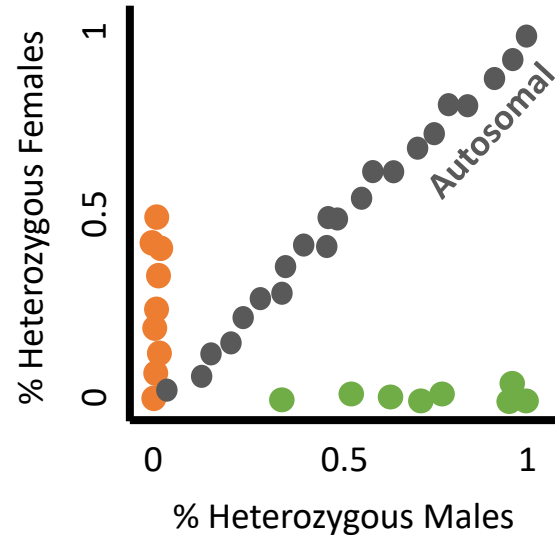
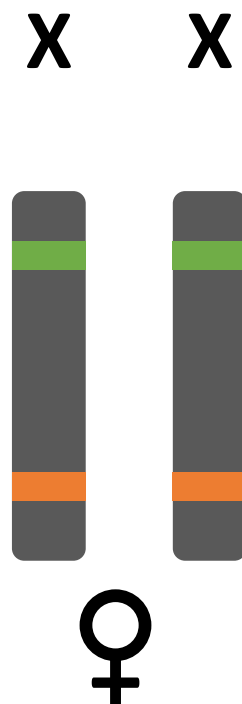
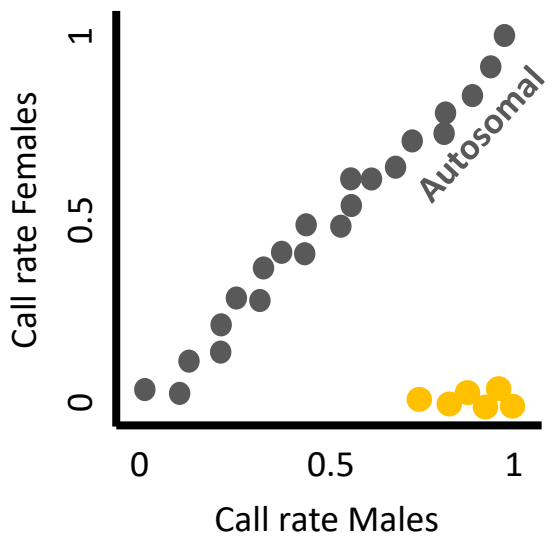
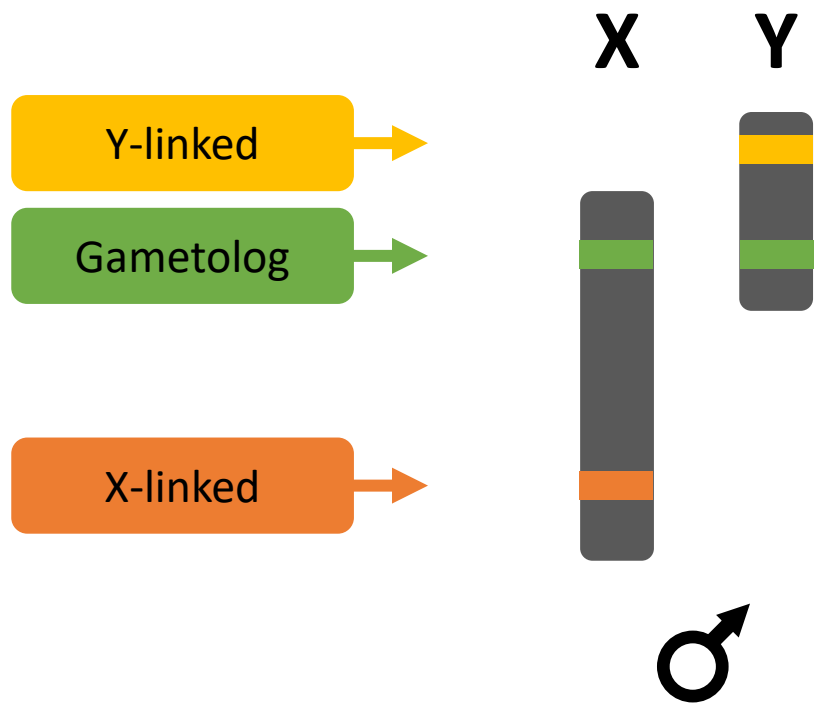
Z-linked →

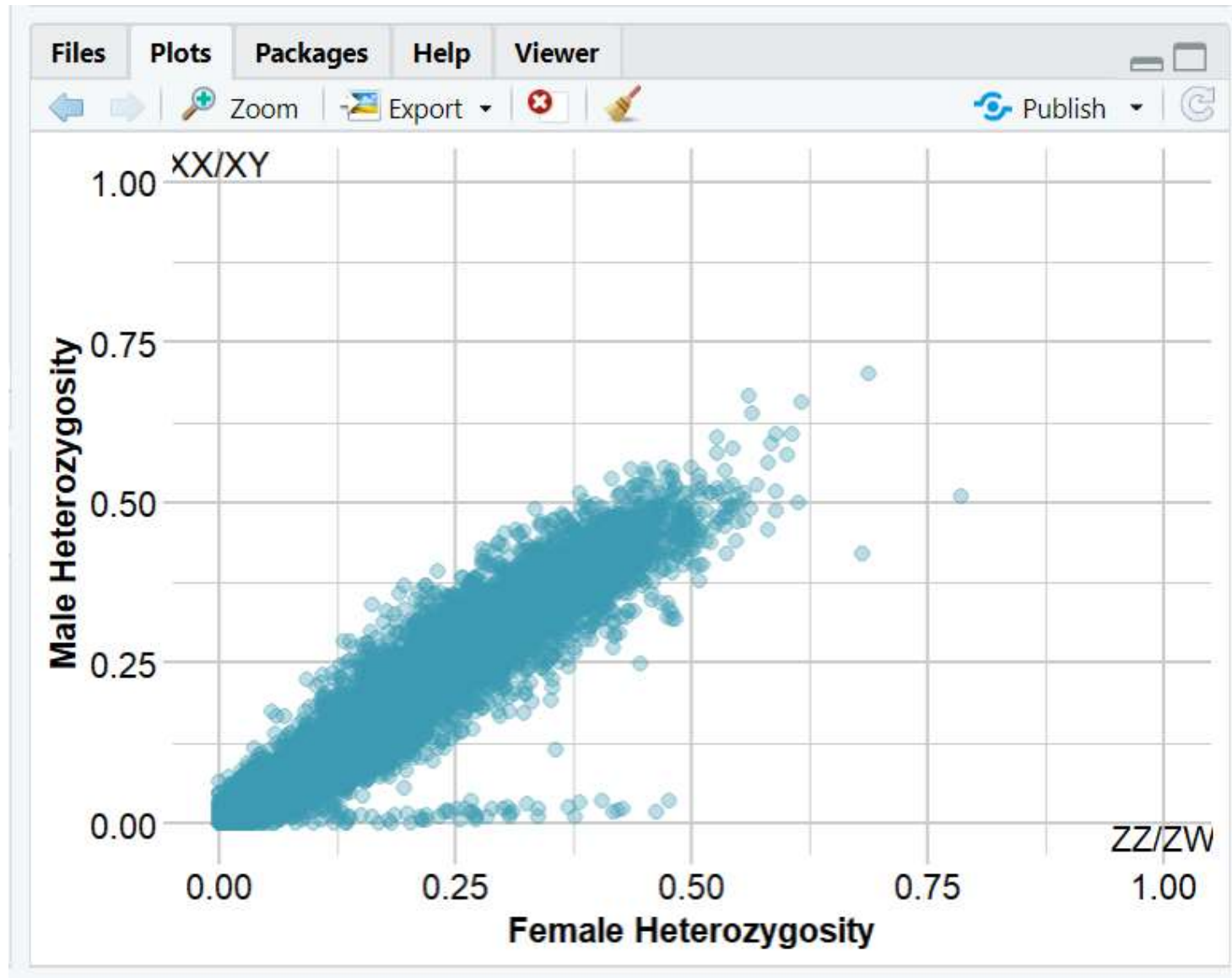


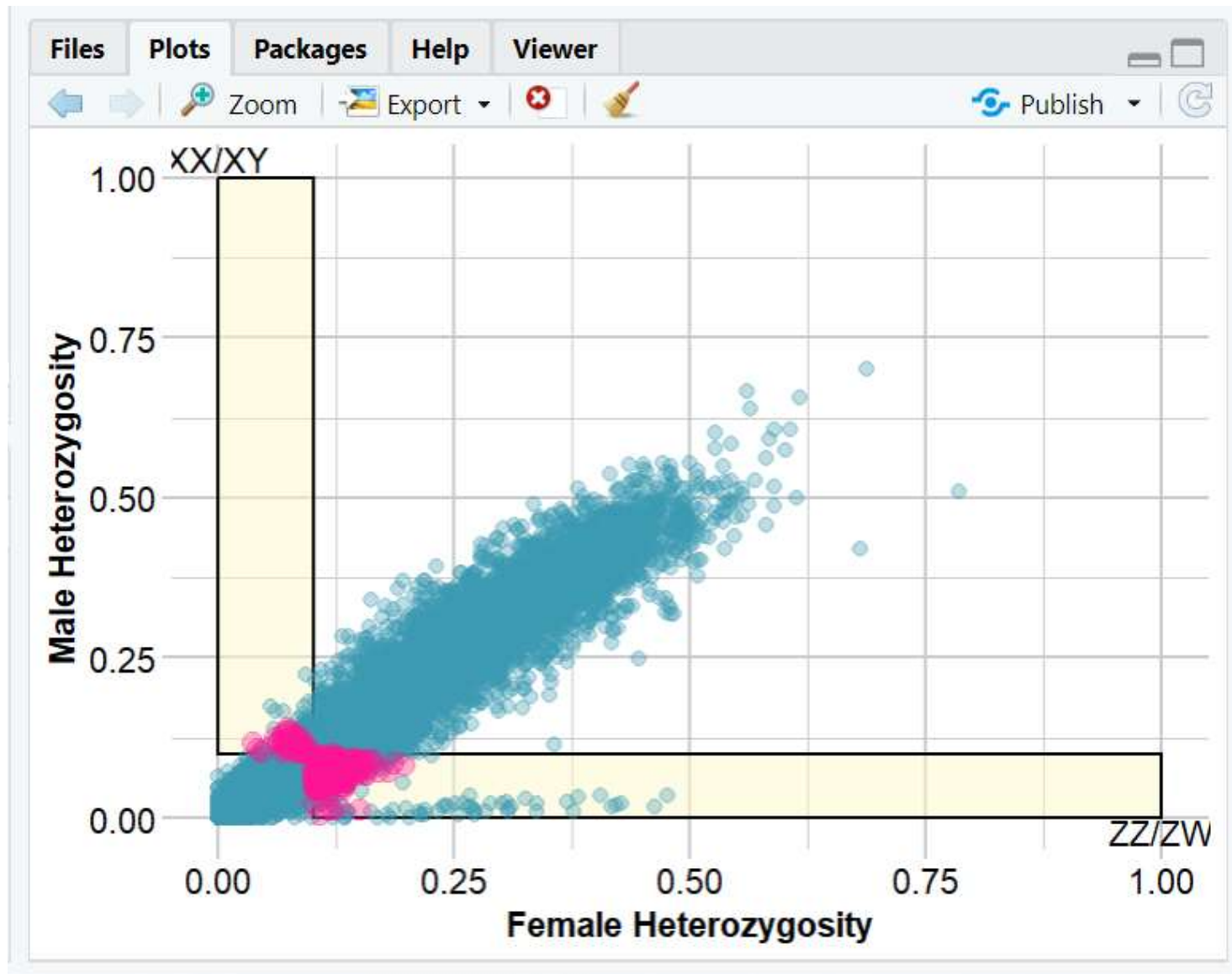


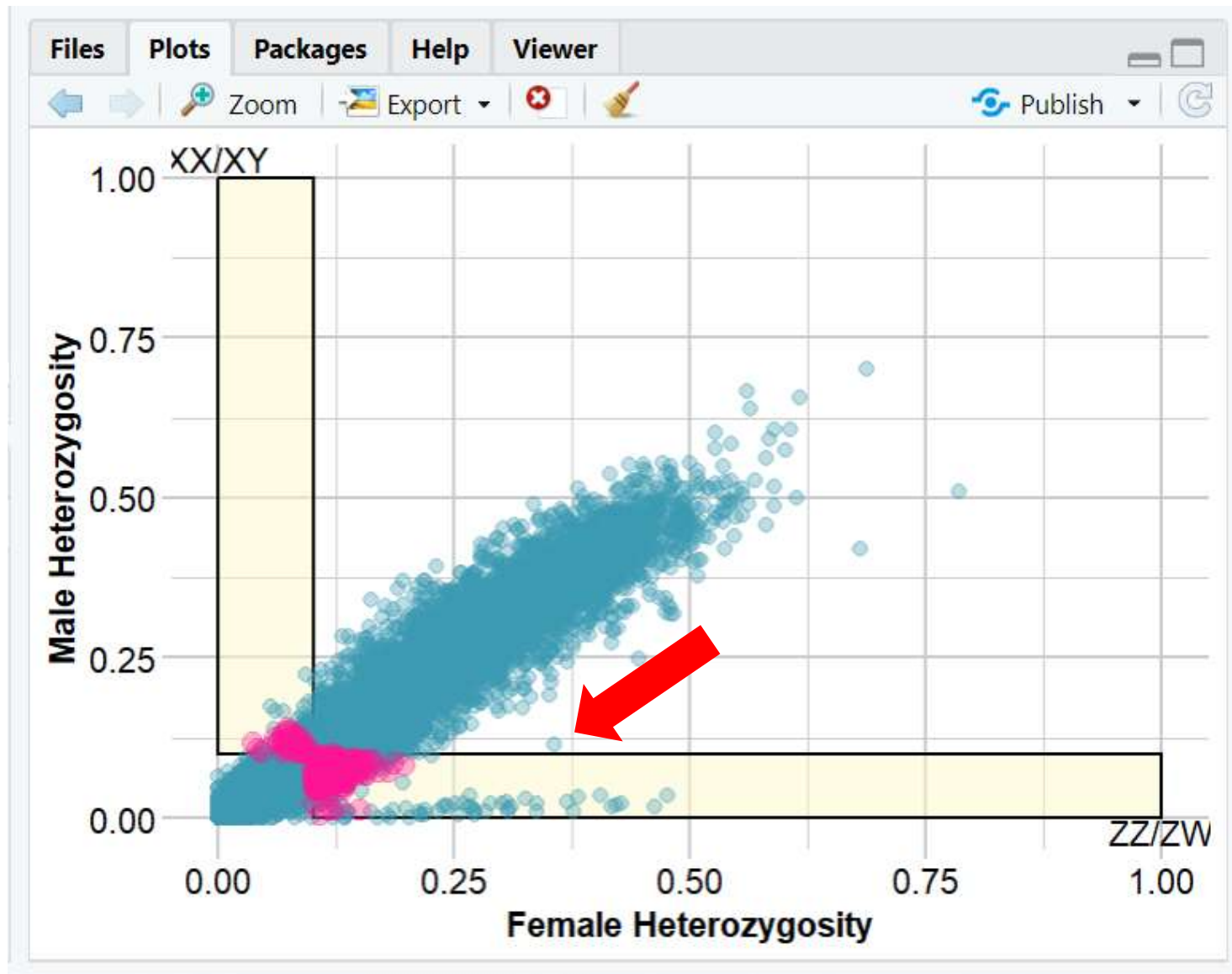
Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses







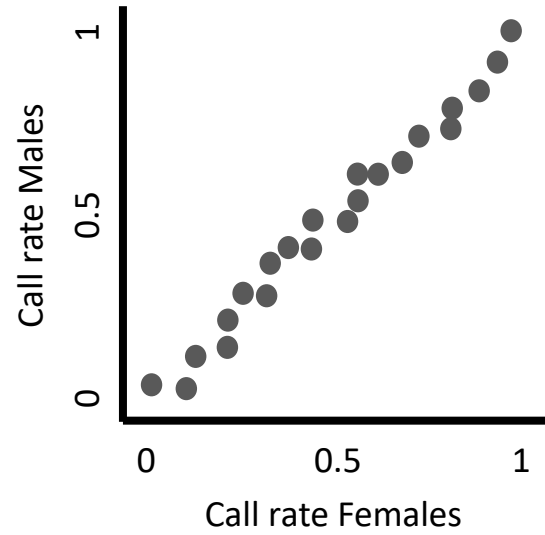


*New package with function that
does it automatically!*

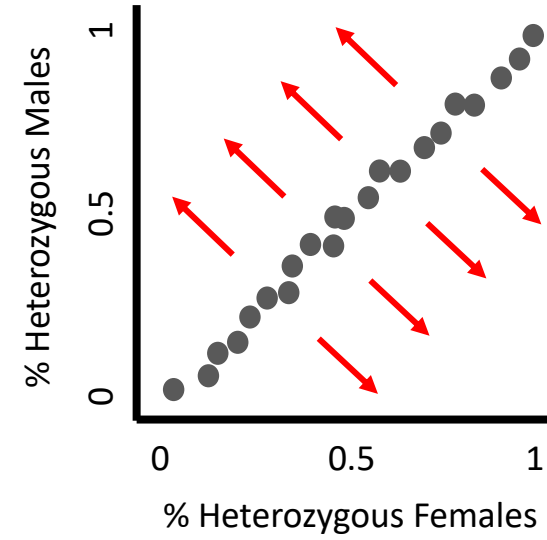
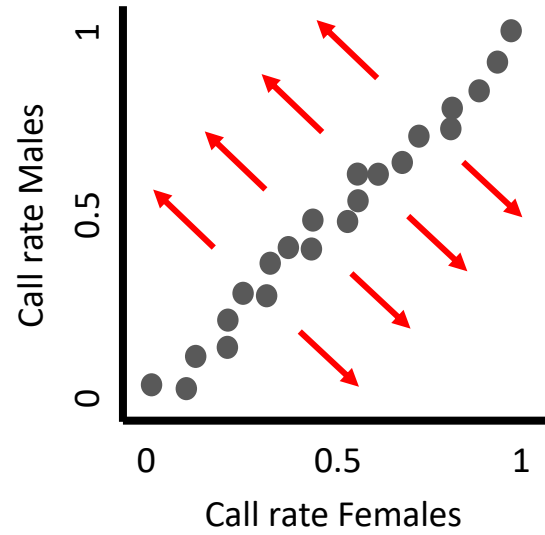


```
gl.filter.sexlinked
```

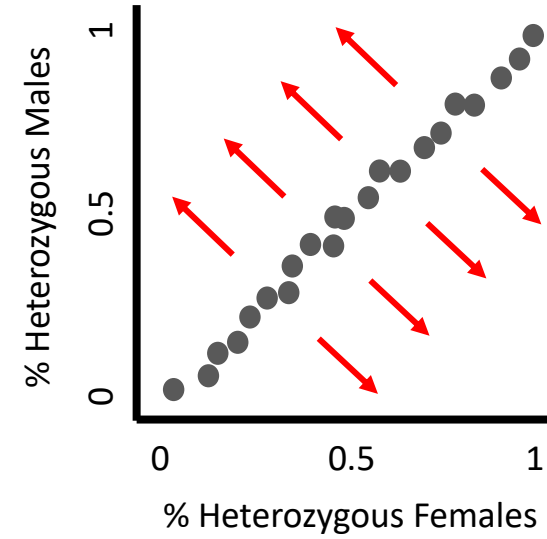
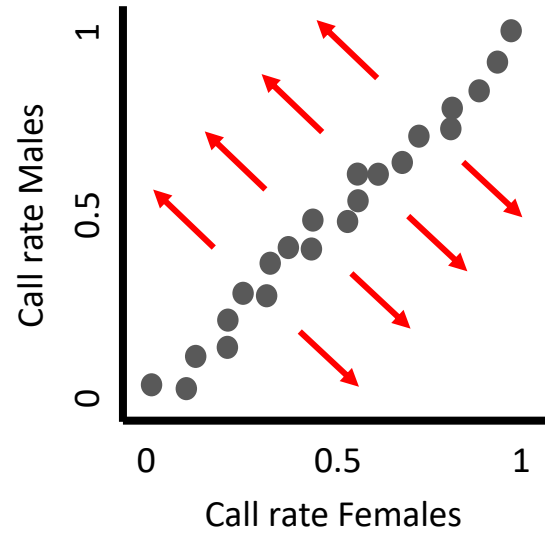
gl.filter.sexlinked



gl.filter.sexlinked

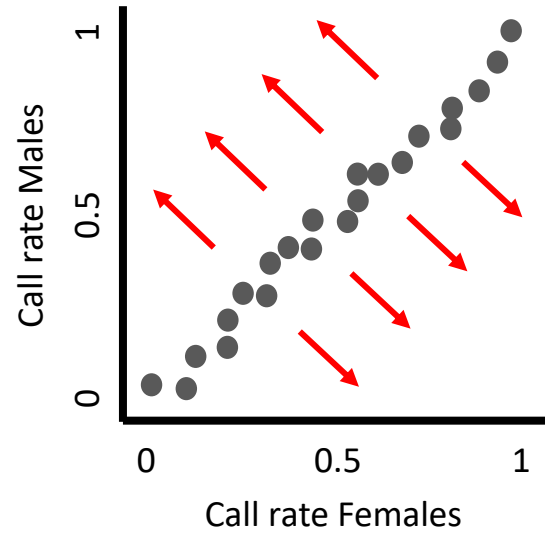


gl.filter.sexlinked

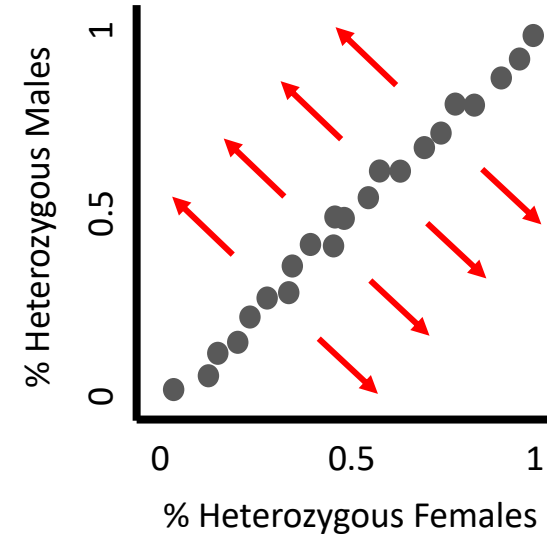


	M	F
Called	230	0
Missing	1	170

gl.filter.sexlinked

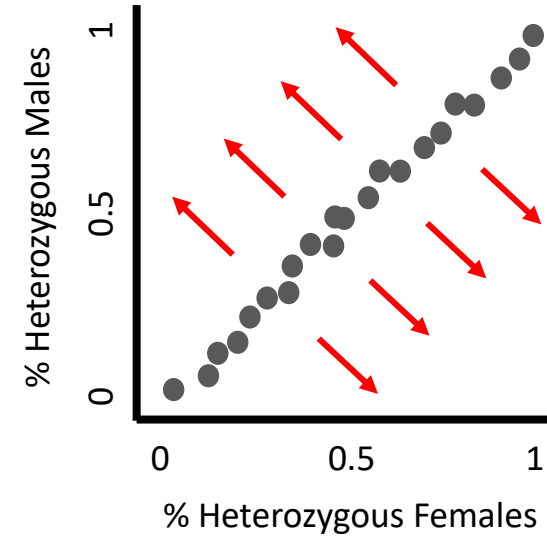
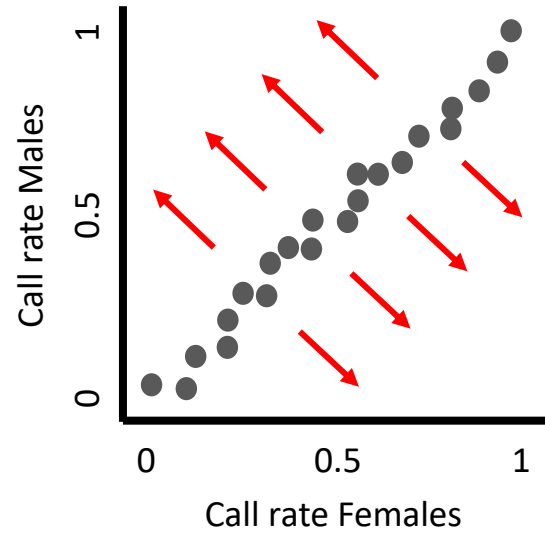


	M	F
Called	230	0
Missing	1	170



	M	F
Homo	2	228
Het	171	0

gl.filter.sexlinked



```
gl.filter.sexlinked(gl.object,
```


sex



```
gl.filter.sexlinked(gl.object,
```

`gl.filter.sexlinked(gl.object, zwxy)`



A diagram consisting of a black arrow pointing from the word 'sex' to the 'gl.object' argument in the function call 'gl.filter.sexlinked(gl.object, ^{zw}_{xy})'.

`gl.filter.sexlinked(gl.object, $\begin{matrix} \text{ZW} \\ \text{xy} \end{matrix}$)`



W-linked

Gametolog

Z-linked

Autosomal

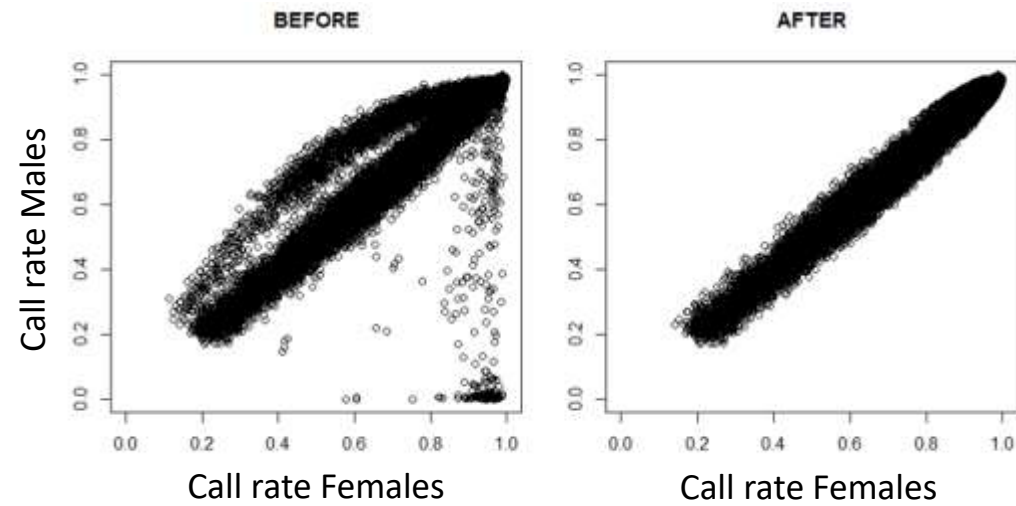
```
gl.filter.sexlinked(gl.object,sex  
                    zw  
                    xy)
```

W-linked

Gametolog

Z-linked

Autosomal



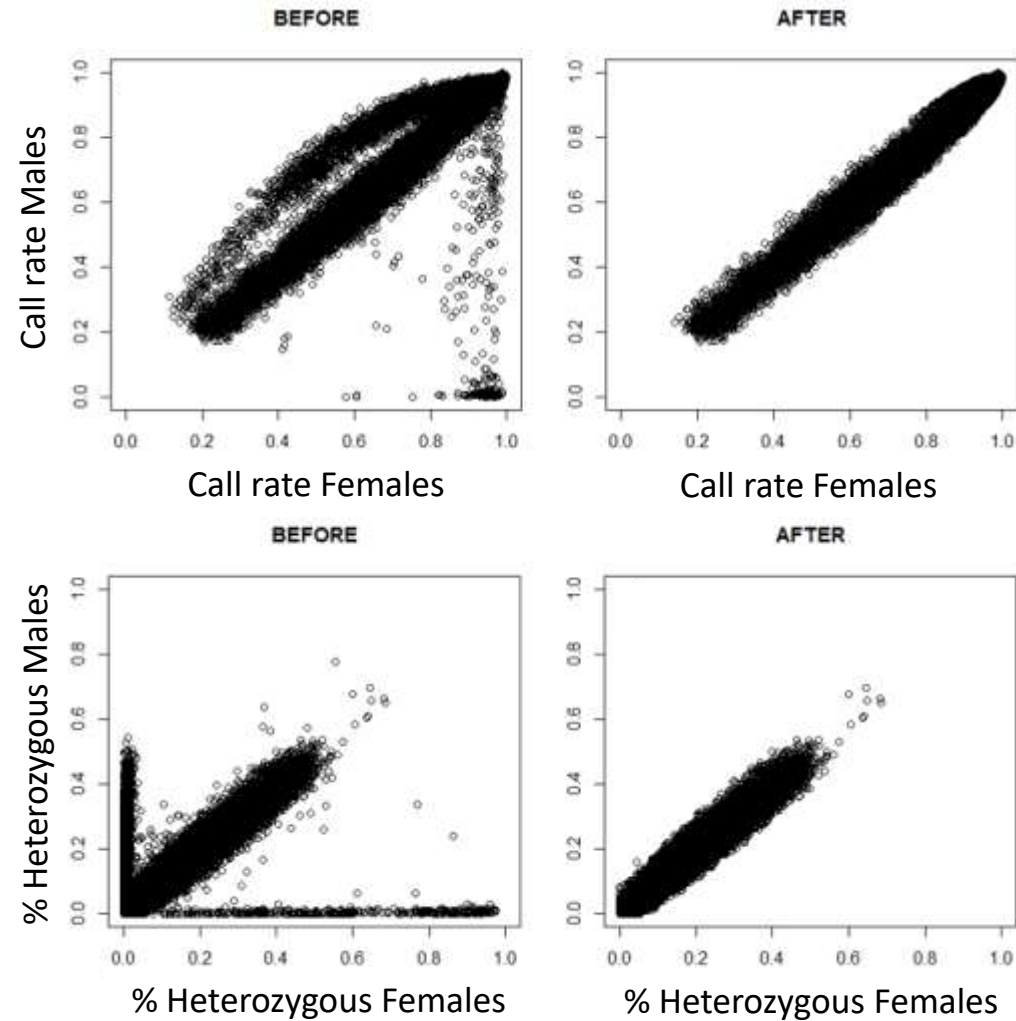
`gl.filter.sexlinked(gl.object, sexzw
xy)`

W-linked

Gametolog

Z-linked

Autosomal



`gl.filter.sexlinked(gl.object, sexzw
xy)`

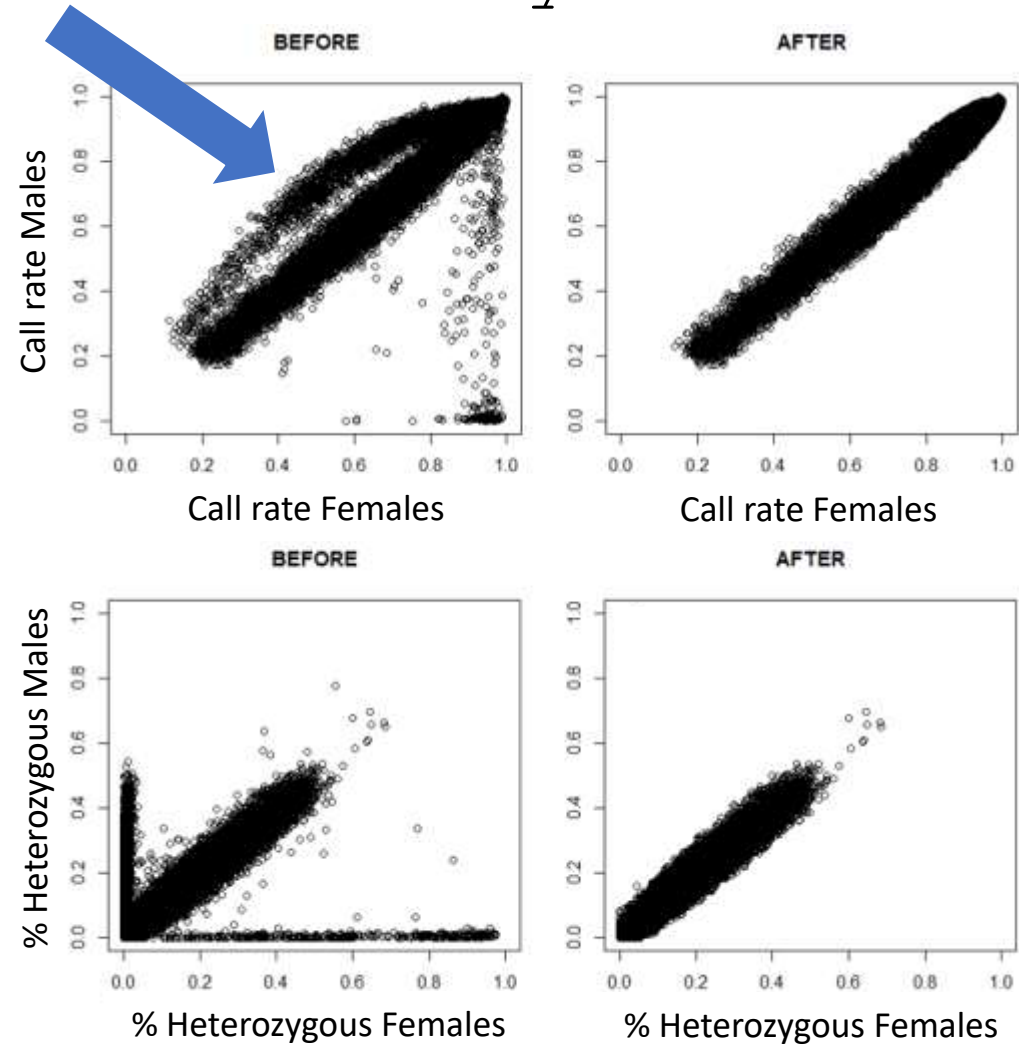
Sex-biased

W-linked

Gametolog

Z-linked

Autosomal



`gl.filter.sexlinked(gl.object, sexzw
xy)`

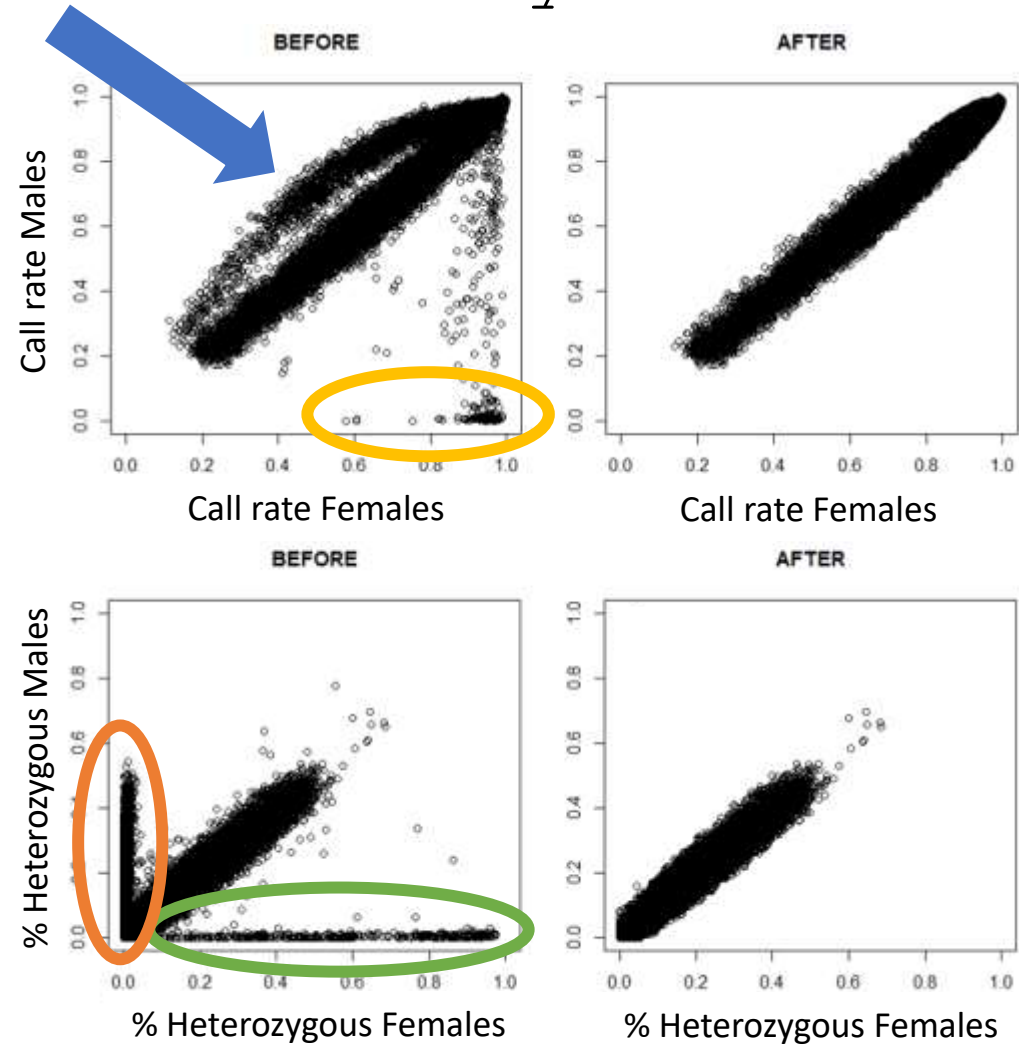
Sex-biased

W-linked

Gametolog

Z-linked

Autosomal



`gl.filter.sexlinked(gl.object, sexzw
xy)`

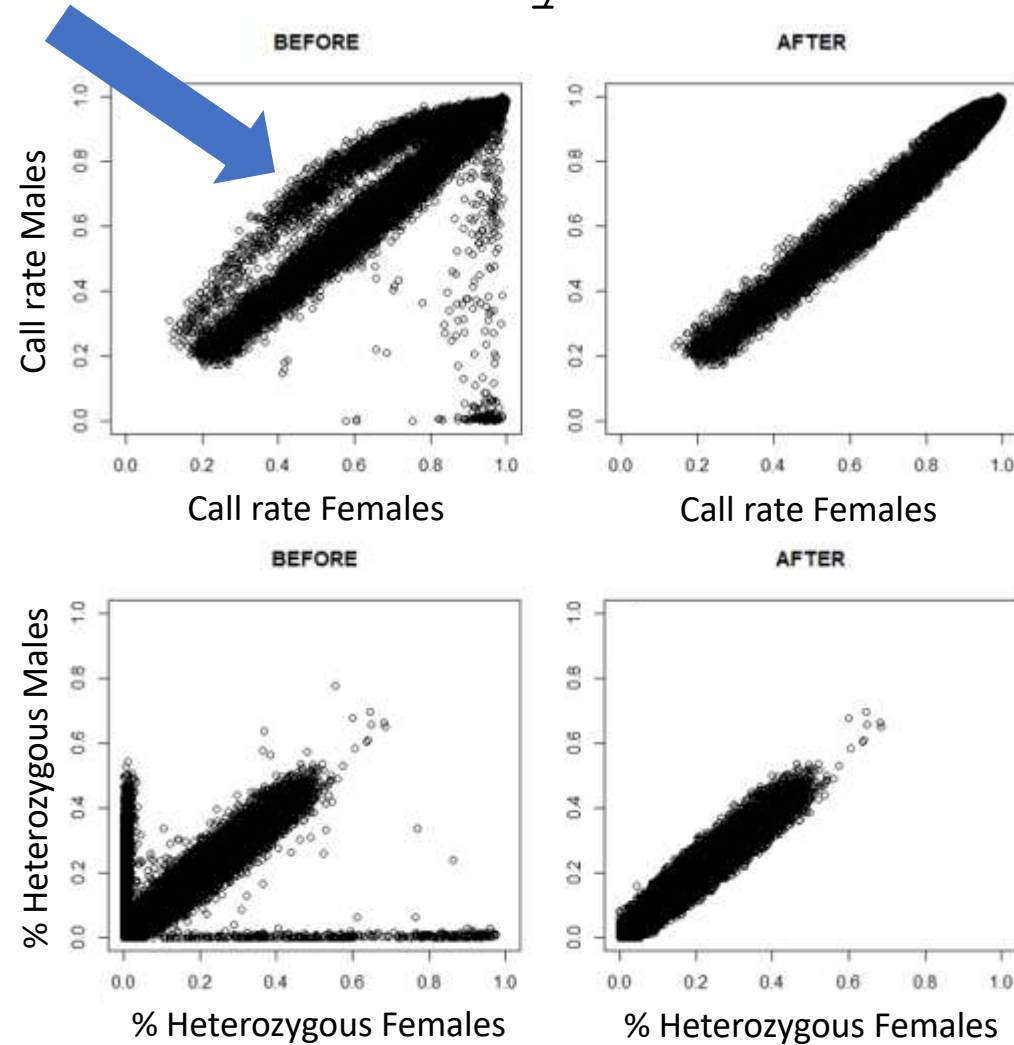
Sex-biased

W-linked

Gametolog

Z-linked

Autosomal



98.8%

Is removing sex-linked loci
necessary?



Yellow-tufted Honeyeater



Eastern Yellow Robin

Sloppy SNP filtering

- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

Sloppy SNP filtering

- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

Correct SNP filtering

- Sex-linked loci
- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

Sloppy SNP filtering

- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

Correct SNP filtering

- Sex-linked loci
- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

① *How good are standard filters at removing sex-linked loci?*

Sloppy SNP filtering

- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

Correct SNP filtering

- Sex-linked loci
- Loci read depth
- Loci missing data
- Ind. missing data
- Loci MAC

- ① *How good are standard filters at removing sex-linked loci?*
- ② *How do sex-linked loci affect popgen results?*



YTH

INITIAL DATASET

W-linked

Z-linked

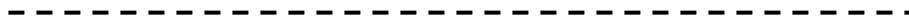
Gametolog

100%

100%

100%

Read depth



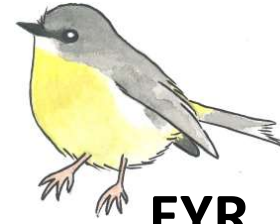
Loci
missing data



Individual
missing data
+
MAC



FINAL DATASET



EYR

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%



FINAL DATASET



YTH

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%

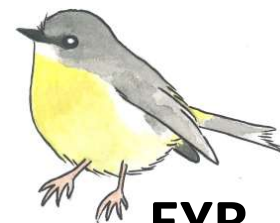
Read depth

75%

Loci
missing data

Individual
missing data
+
MAC

FINAL DATASET



EYR

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%

49%

FINAL DATASET



YTH

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%

Read depth

75%

75%

62%

Loci
missing data

60%

37%

Individual
missing data
+
MAC

57%

37%

FINAL DATASET



EYR

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%

Read depth

49%

81%

80%

Loci
missing data

75%

71%

Individual
missing data
+
MAC

75%

71%

FINAL DATASET



YTH

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%

Read depth

75%

75%

62%

Loci
missing data

60%

37%

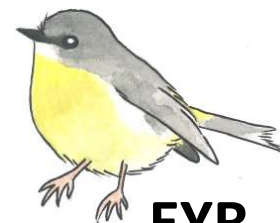
Individual
missing data
+
MAC

57%

37%

FINAL DATASET

5.7%



EYR

INITIAL DATASET

W-linked

Z-linked

Gametolog

100%

100%

100%

Read depth

49%

81%

80%

Loci
missing data

75%

71%

Individual
missing data
+
MAC

75%

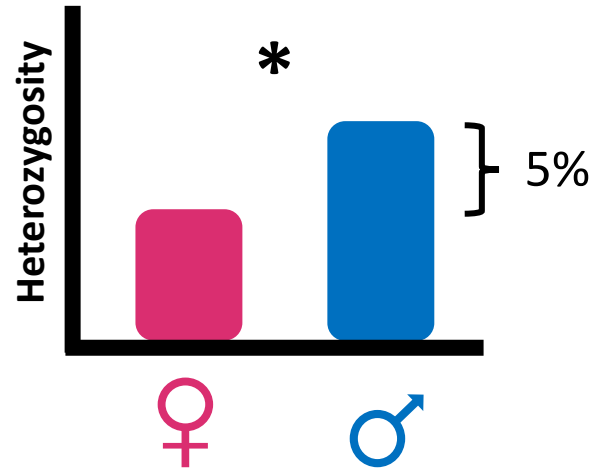
71%

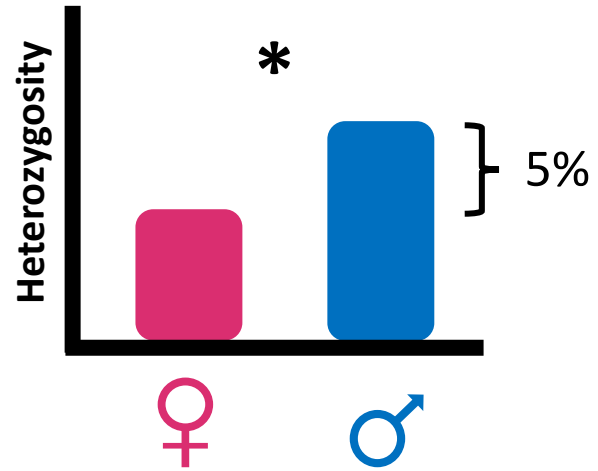
FINAL DATASET

7.8%

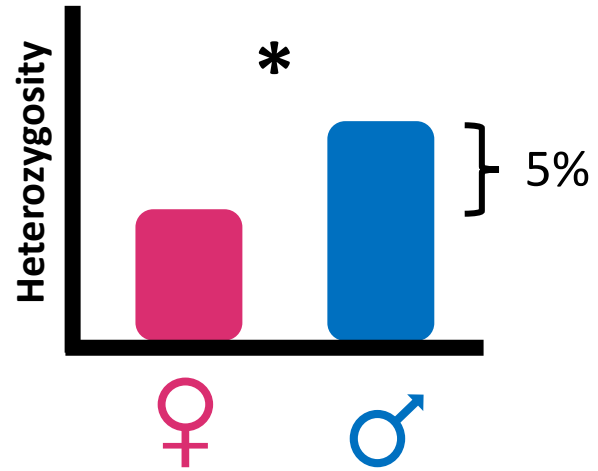
Take home message no. 1

Standard filters suck at removing all sex-linked loci

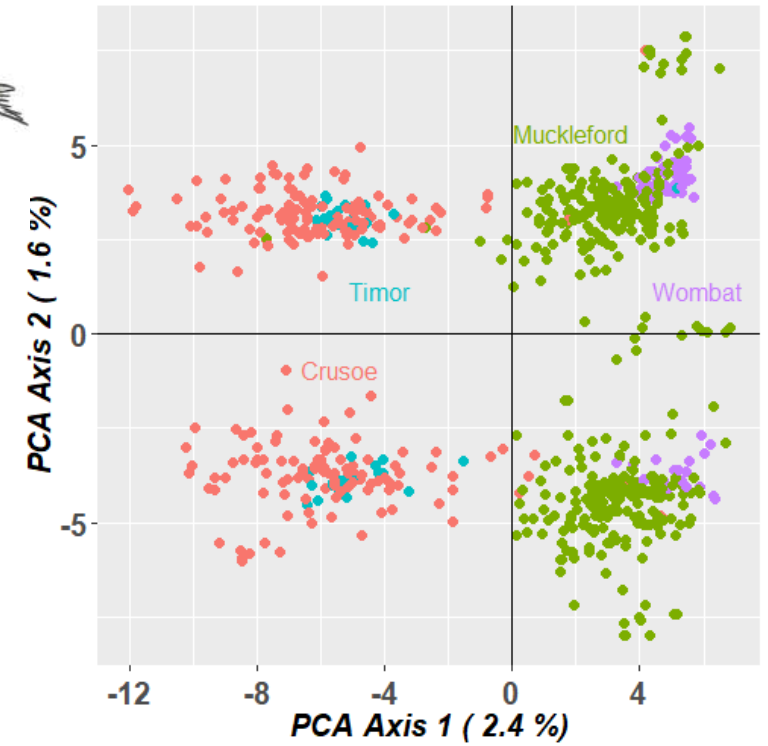
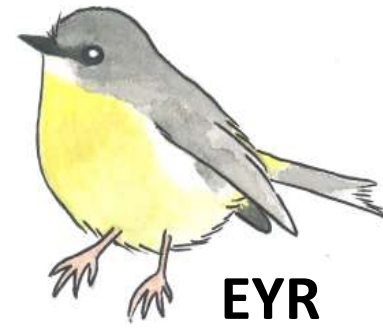


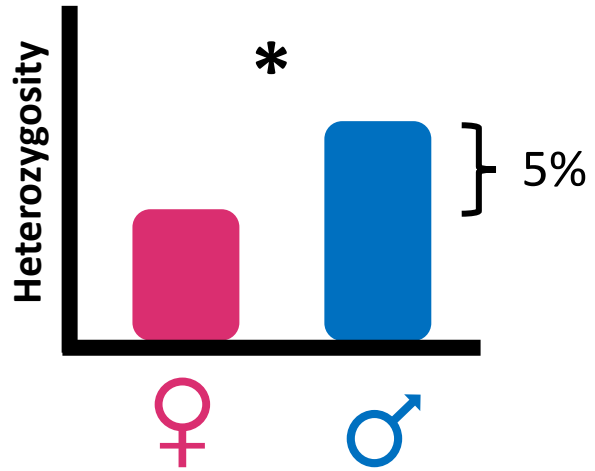


- Female philopatry
- Inbred males are dying (inbreeding depression)

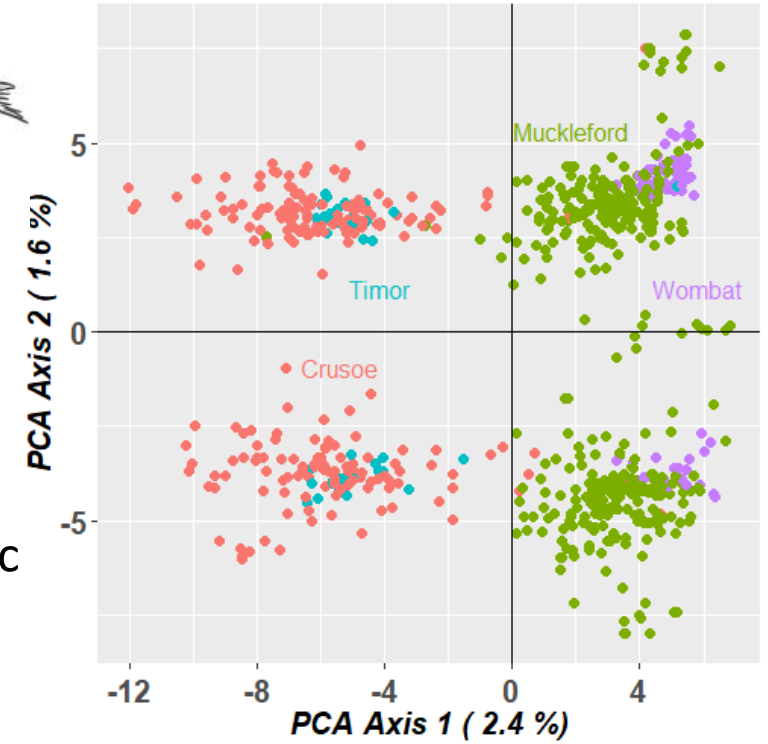
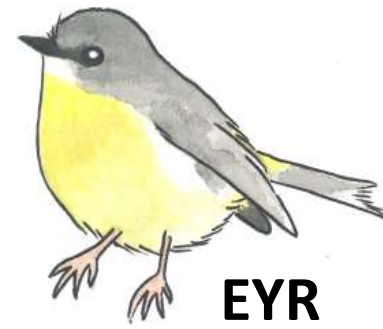


- Female philopatry
- Inbred males are dying (inbreeding depression)

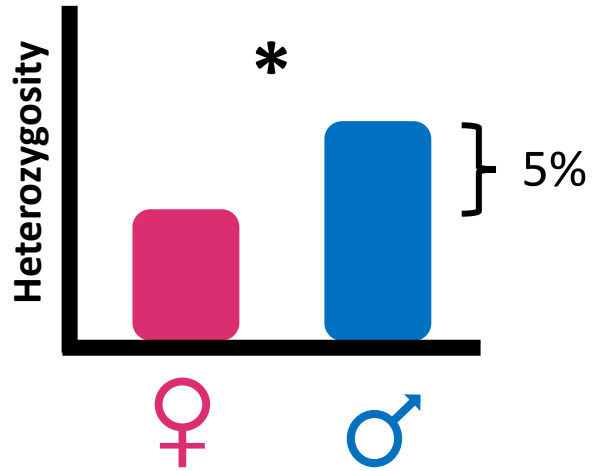




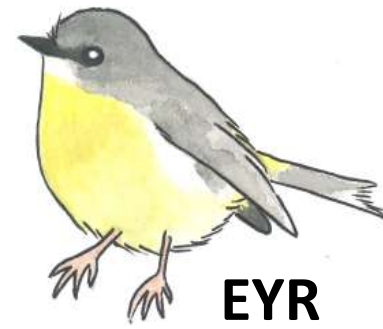
- Female philopatry
- Inbred males are dying (inbreeding depression)



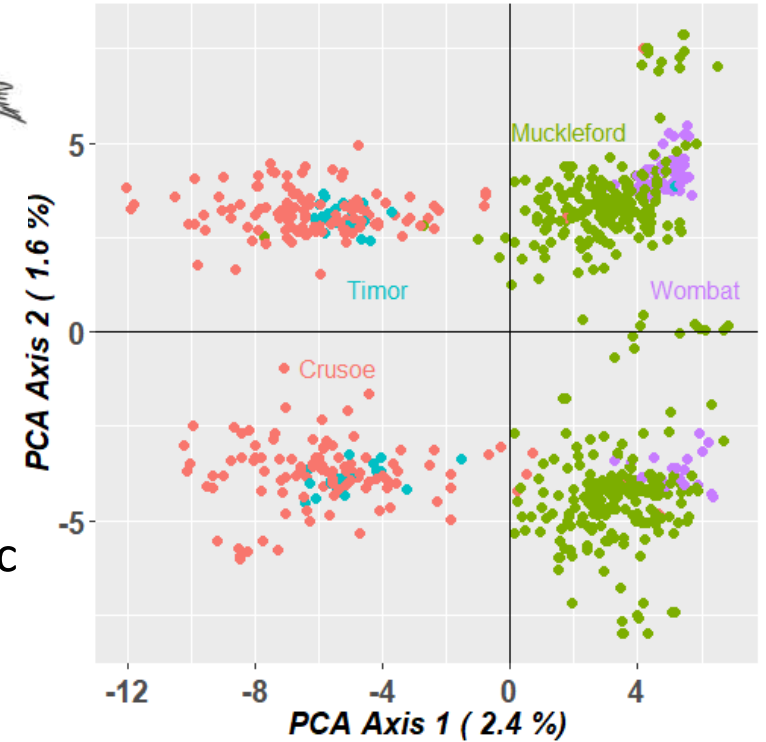
- Two cryptic sympatric species



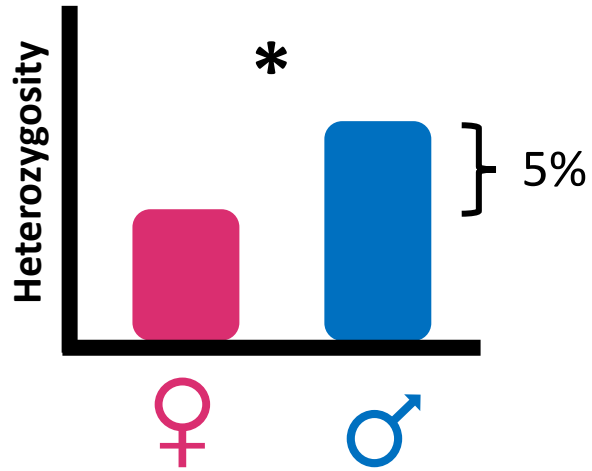
- Female philopatry
- Inbred males are dying (inbreeding depression)



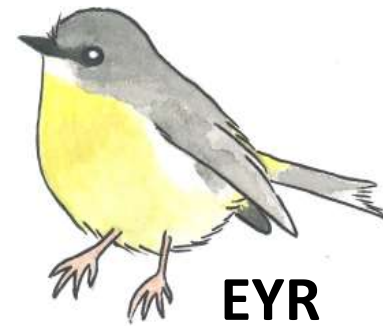
- Two cryptic sympatric species



91 parent-offspring

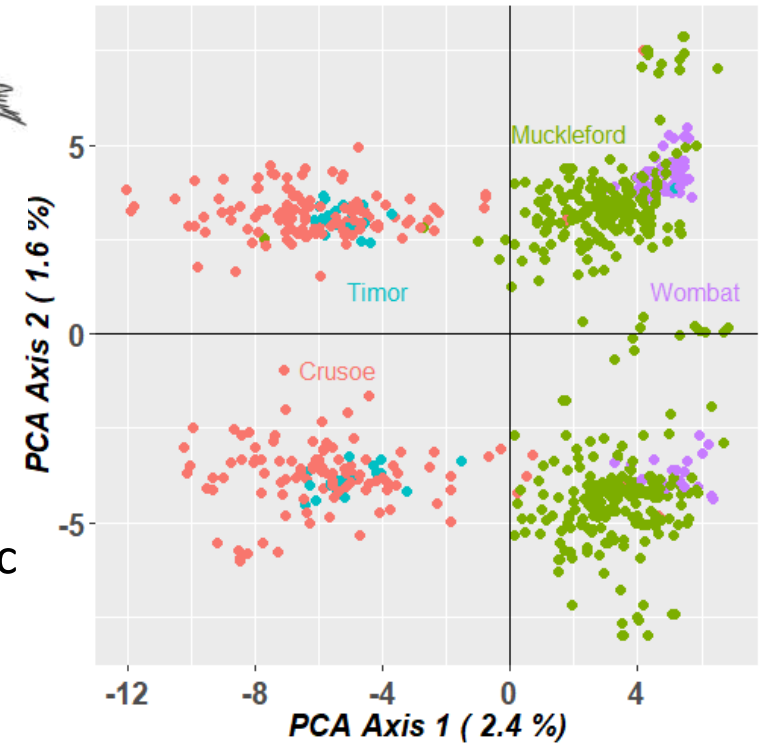


- Female philopatry
- Inbred males are dying (inbreeding depression)



EYR

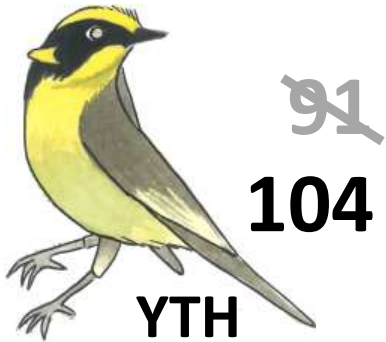
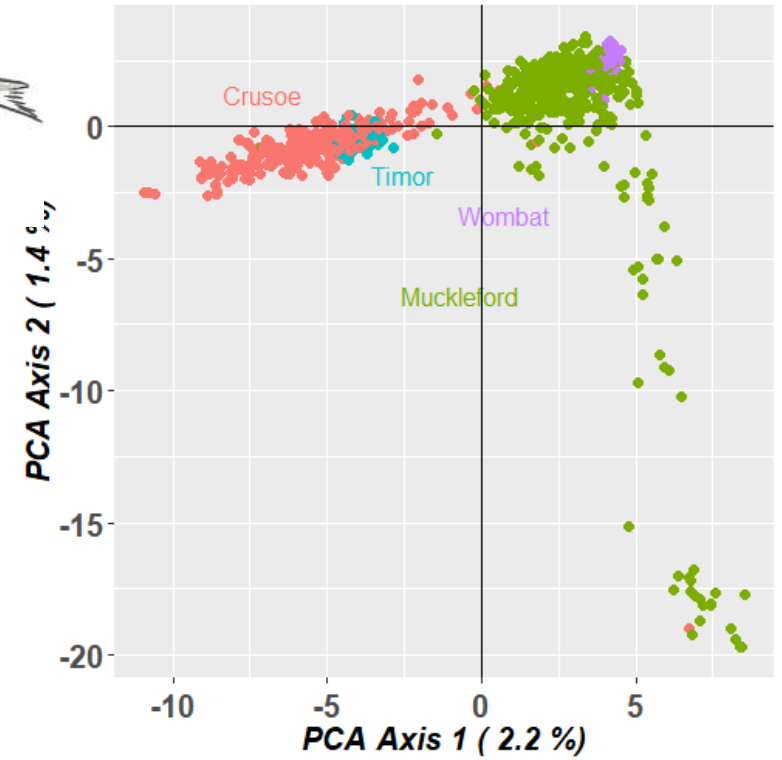
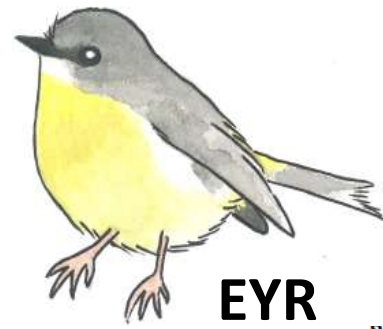
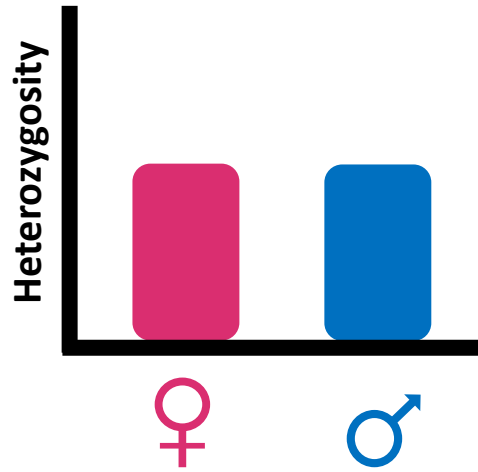
- Two cryptic sympatric species



YTH

91 parent-offspring

- Extra-pair paternity
- Captive breeding
- Translocations



~~91~~
104 parent-offspring

Take home message no. 2

Sex-linked loci can f*-up popgen analyses:
REMOVE!!!**

Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses

Applications of sex-linked loci

- 1. REMOVE THEM: Facilitate unbiased popgen analyses**
- 2. USE THEM:**
 - Sex identification of samples**

Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses

2. USE THEM:

sex -> sex-linked loci

- Sex identification of samples

Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses

2. USE THEM: sex → sex-linked loci

- **Sex identification of samples** sex-linked loci → sex

Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses

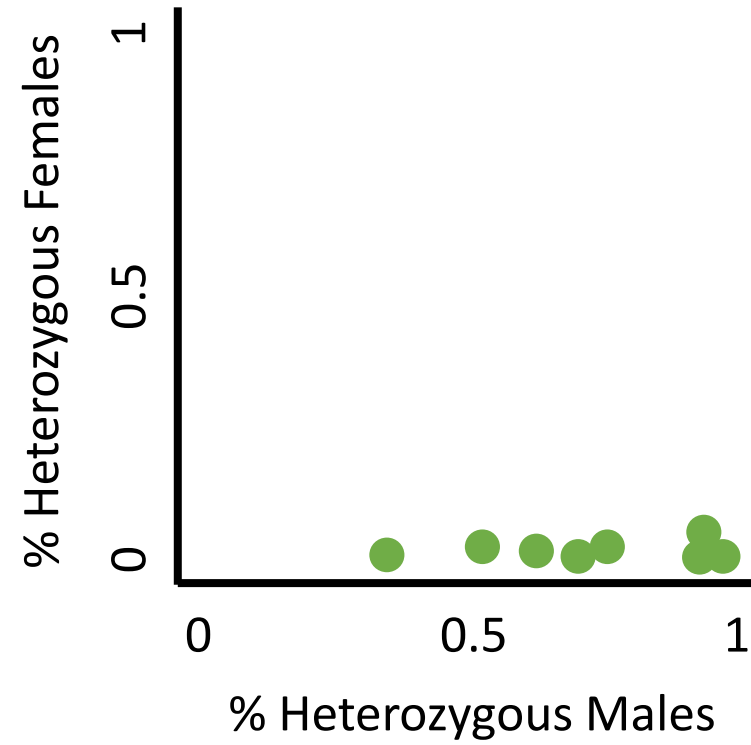
2. USE THEM:

- Sex identification of samples**

```
gl.infer.sex
```


Gametolog

`gl.infer.sex`



W-linked

Gametolog

Z-linked



`gl.infer.sex`

W-linked

Gametolog

Z-linked



`gl.infer.sex` *ML algorithm!



	W-linked sex	gametolog sex	Z-linked sex	agreed sex
IndA	F	F	F	F
IndB	M	M	M	M
IndC	F	F	M	F*

W-linked

Gametolog

Z-linked



`gl.infer.sex`

More robust!

	W-linked sex	gametolog sex	Z-linked sex	agreed sex
IndA	F	F	F	F
IndB	M	M	M	M
IndC	F	F	M	F*

Let's talk numbers

sex (15 males + 15 females)

sex (15 males + 15 females)



gl.filter.sexlinked

sex (15 males + 15 females)



`gl.filter.sexlinked`



sex (15 males + 15 females)



`gl.filter.sexlinked`



`gl.infer.sex`



sex (all)

sex (15 males + 15 females)



`gl.filter.sexlinked`



`gl.infer.sex`



sex (all)



`gl.filter.sexlinked`



sex (15 males + 15 females)



`gl.filter.sexlinked`



`gl.infer.sex`



sex (all)



`gl.filter.sexlinked`



Quality filters...

sex (15 males + 15 females)



`gl.filter.sexlinked`



`gl.infer.sex`



sex (all)



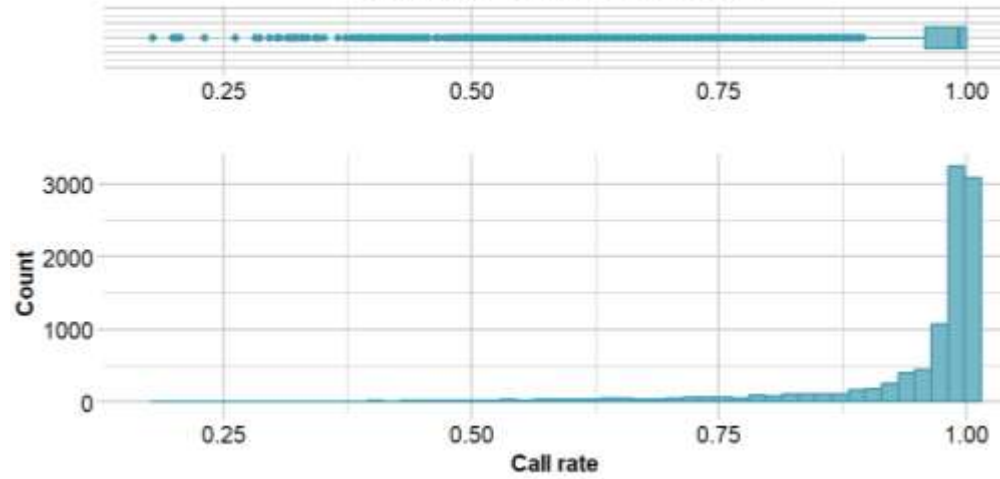
`gl.filter.sexlinked`



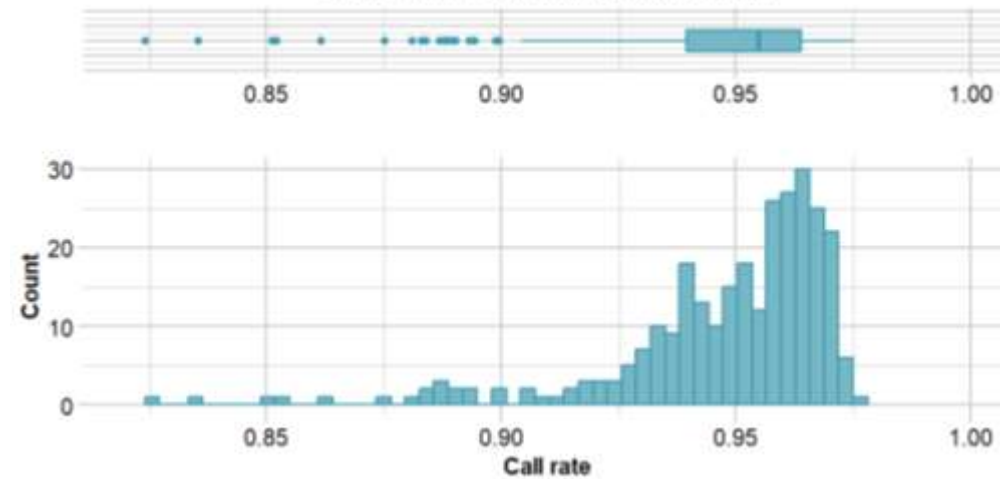
Quality filters... ?

Mammal (XY)

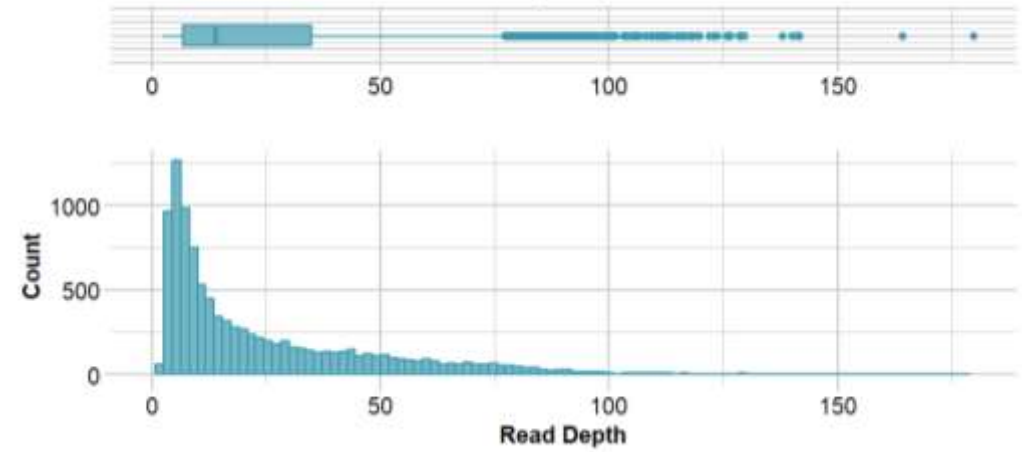
SNP data - Call Rate by Locus



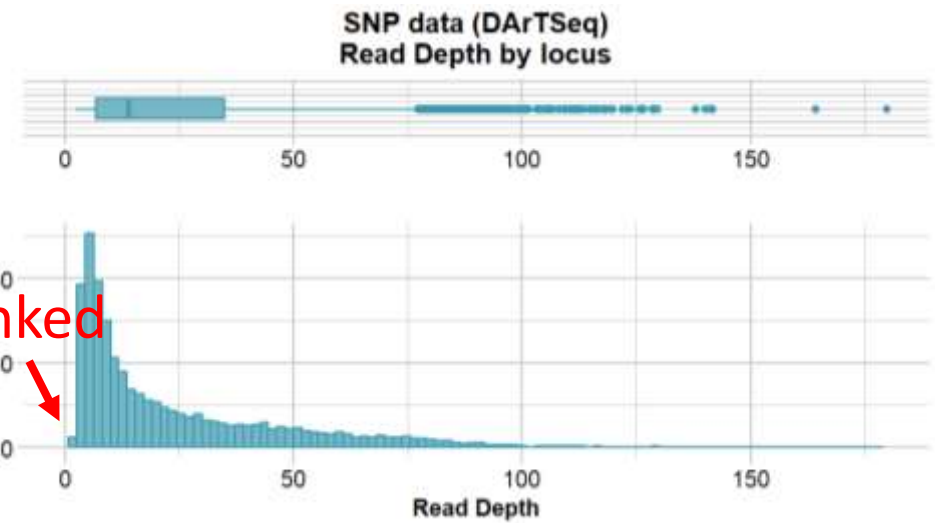
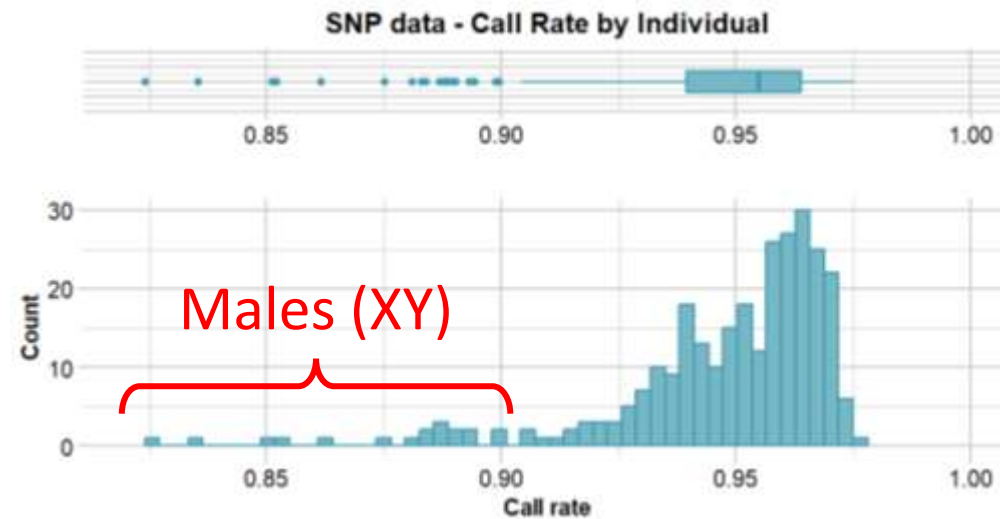
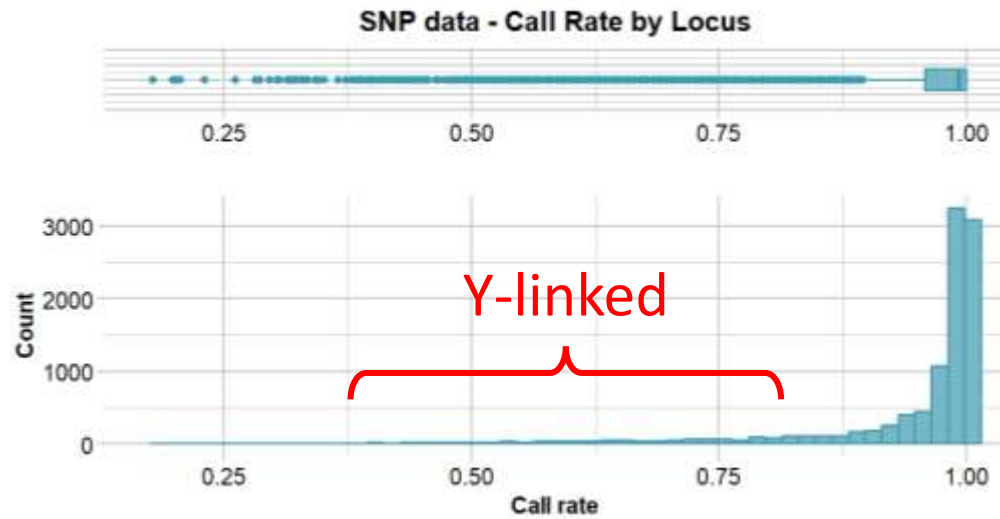
SNP data - Call Rate by Individual

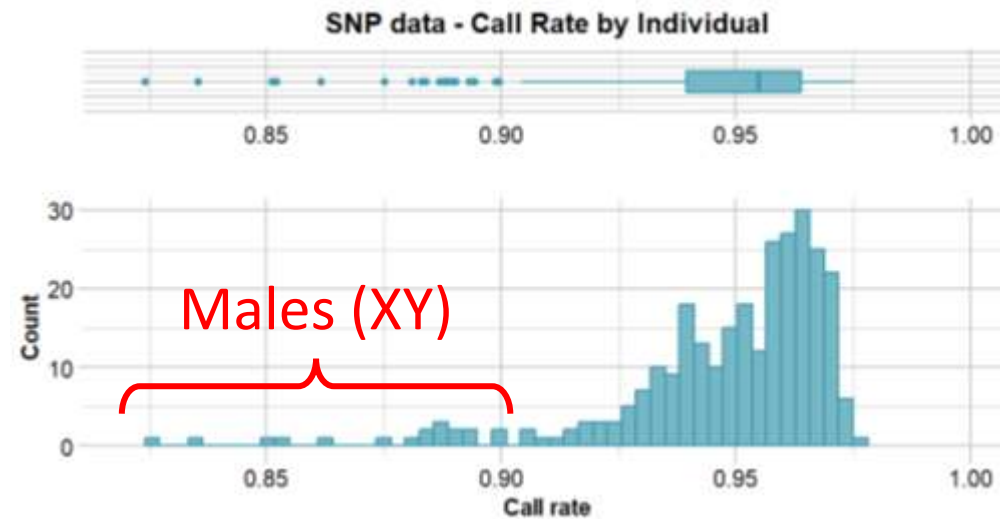
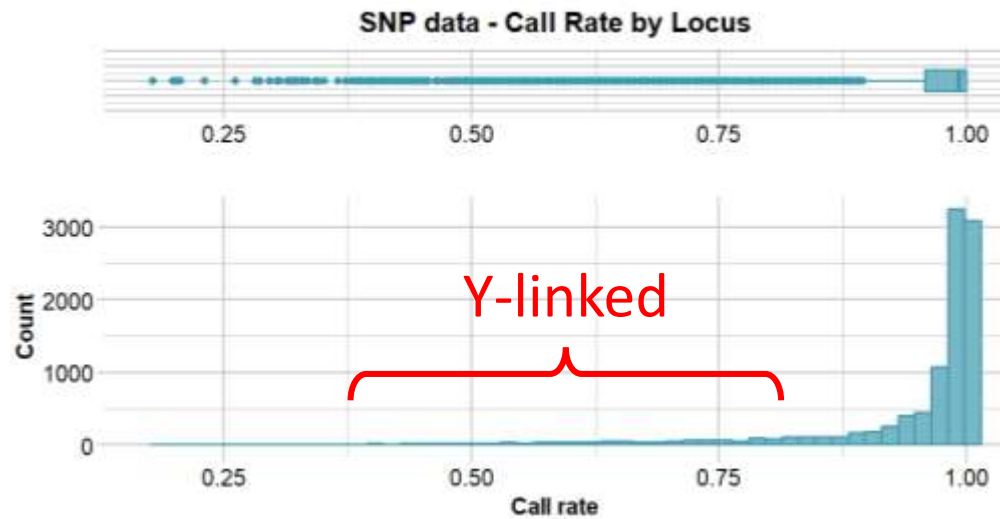


SNP data (DARTSeq)
Read Depth by locus

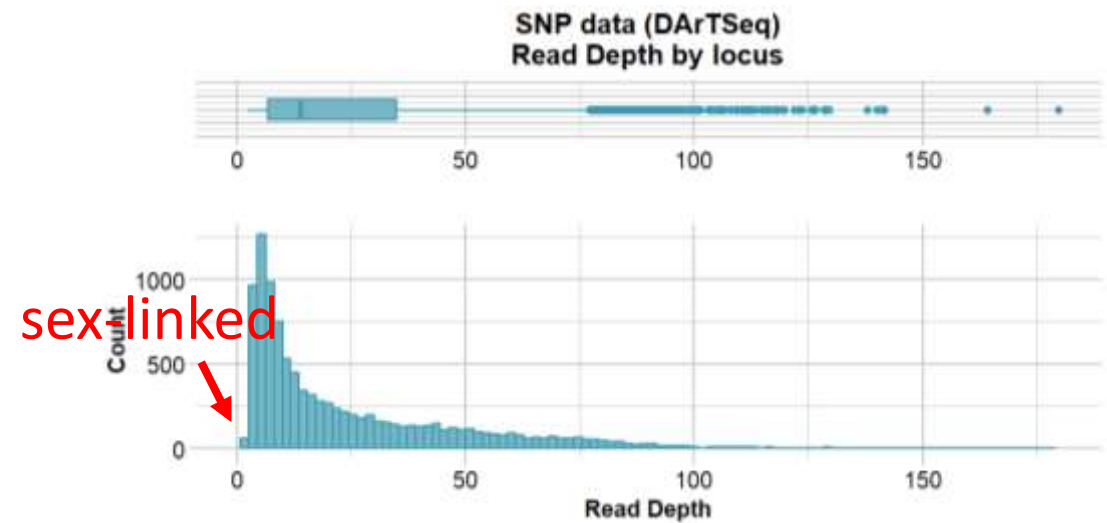


Mammal (XY)





Mammal (XY)



Remove sex-linked loci first to guarantee that:

- You won't unfairly remove inds from one sex
- Downstream filters remove bad-quality loci

Take home message no. 3

Remove sex-linked loci FIRST

Applications of sex-linked loci

- 1. REMOVE THEM: Facilitate unbiased popgen analyses**
- 2. USE THEM:**
 - Sex identification of samples**

Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses

2. USE THEM:

- Sex identification of samples**
- Study sex-biased dispersal**

Applications of sex-linked loci

1. REMOVE THEM: Facilitate unbiased popgen analyses

2. USE THEM:

- Sex identification of samples
- Study sex-biased dispersal
- Study the evolution of sex chromosomes

Chapter B3: Molecular sex identification for applications in conservation, industry, and veterinary medicine

Authors and Affiliations

Clare E. Holleley¹, Sarah L. Whiteley^{1,2}, Floriaan Devloo-Delva¹, Andreas Bachler^{1,3}, Joshua Linas⁴, Arthur Georges²

¹ National Research Collections Australia, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia.

² Institute for Applied Ecology, University of Canberra, Canberra, Australia.

³ Australian National University, Research School of Biology, Canberra, Australia.

⁴ The Unusual Pet Vets Jindalee, Veterinary Services for Reptiles, Birds and Small Mammals, Queensland, Australia.

Abstract

The ability to identify whether an organism is female or male is essential data feeding into a range of applications in biological research, animal-based industries and veterinary medicine. Many organisms show striking sexual dimorphism in secondary sexual characters and/or visible genitalia that make assigning a sex to an individual unambiguous. However, there are also many species where it is impossible to distinguish males from females on the basis of their outward appearance, and surprising complexity and lability exists in this seemingly binary trait (e.g., environmental sex reversal, natural sex change). Even in sexually dimorphic species, there are often life-stages where sex cannot be confidently assigned (e.g., embryonic

Received: 25 November 2022 | Revised: 28 March 2023 | Accepted: 4 July 2023

DOI: 10.1111/1755-0998.13844

RESOURCE ARTICLE

MOLECULAR ECOLOGY
RESOURCES

WILEY

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

Diana A. Robledo-Ruiz¹ | Lana Austin¹ | J. Nevil Amos^{1,2} |
Jesús Castrejón-Figueroa¹ | Daniel K. P. Harley³ | Michael J. L. Magrath^{3,4} |
Paul Sunnucks¹ | Alexandra Pavlova¹

¹School of Biological Sciences, Monash University, Clayton, Victoria, Australia

²Department of Energy, Environment and Climate Action, Arthur Rylah Institute for Environmental Research, Heidelberg, Victoria, Australia

³Department of Wildlife Conservation and Science, Zoos Victoria, Parkville, Victoria, Australia

⁴School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

Abstract

Identifying sex-linked markers in genomic datasets is important because their presence in supposedly neutral autosomal datasets can result in incorrect estimates of genetic diversity, population structure and parentage. However, detecting sex-linked loci can be challenging, and available scripts neglect some categories of sex-linked variation. Here, we present new *R* functions to (1) identify and separate sex-linked loci in ZW and XY sex determination systems and (2) infer the genetic sex of individu-

Check for updates

Chapter B3: Molecular sex identification for applications in conservation, industry, and veterinary medicine

Authors and Affiliations

Clare E. Holleley¹, Sarah L. Whiteley^{1,2}, Fioriaan Devloo-Delva¹, Andreas Bachler^{1,3}, Joshua Linas⁴, Arthur Georges²

¹ National Research Collections Australia, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia.

² Institute for Applied Ecology, University of Canberra, Canberra, Australia.

³ Australian National University, Research School of Biology, Canberra, Australia.

⁴ The Unusual Pet Vets Jindalee, Veterinary Services for Reptiles, Birds and Small Mammals, Queensland, Australia.

Abstract

The ability to identify whether an organism is female or male is essential data feeding into a range of applications in biological research, animal-based industries and veterinary medicine. Many organisms show striking sexual dimorphism in secondary sexual characters and/or visible genitalia that make assigning a sex to an individual unambiguous. However, there are also many species where it is impossible to distinguish males from females on the basis of their outward appearance, and surprising complexity and lability exists in this seemingly binary trait (e.g., environmental sex reversal, natural sex change). Even in sexually dimorphic species, there are often life-stages where sex cannot be confidently assigned (e.g., embryonic

Questions?

Check for updates

Received: 25 November 2022 | Revised: 28 March 2023 | Accepted: 4 July 2023

DOI: 10.1111/1755-0998.13844

RESOURCE ARTICLE

**MOLECULAR ECOLOGY
RESOURCES** WILEY

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

Diana A. Robledo-Ruiz¹ | Lana Austin¹ | J. Nevil Amos^{1,2} |
Jesús Castrejón-Figueroa¹ | Daniel K. P. Harley³ | Michael J. L. Magrath^{3,4} |
Paul Sunnucks¹ | Alexandra Pavlova¹

¹School of Biological Sciences, Monash University, Clayton, Victoria, Australia

²Department of Energy, Environment and Climate Action, Arthur Rylah Institute for Environmental Research, Heidelberg, Victoria, Australia

³Department of Wildlife Conservation and Science, Zoos Victoria, Parkville, Victoria, Australia

⁴School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

Abstract

Identifying sex-linked markers in genomic datasets is important because their presence in supposedly neutral autosomal datasets can result in incorrect estimates of genetic diversity, population structure and parentage. However, detecting sex-linked loci can be challenging, and available scripts neglect some categories of sex-linked variation. Here, we present new *R* functions to (1) identify and separate sex-linked loci in ZW and XY sex determination systems and (2) infer the genetic sex of individu-