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

ICCB
JUNE 15-20
2025

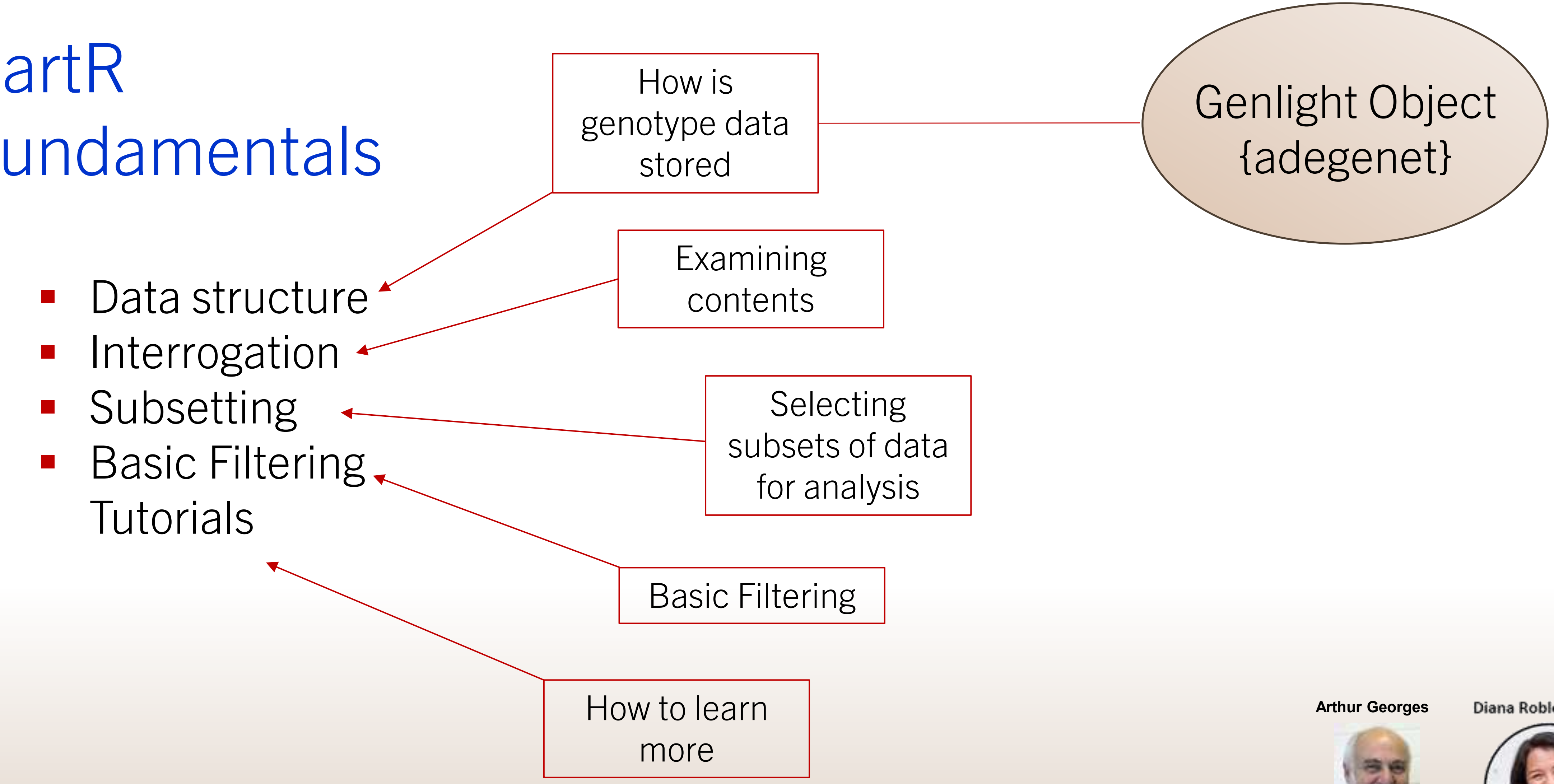
UNIVERSITY OF CANBERRA

DArT diversity arrays technology

MONASH University

Centre for Biodiversity Analysis

dartR Fundamentals



- Data structure
- Interrogation
- Subsetting
- Basic Filtering

Tutorials

How is genotype data stored

Genlight Object {adegenet}

Examining contents

Selecting subsets of data for analysis

Basic Filtering

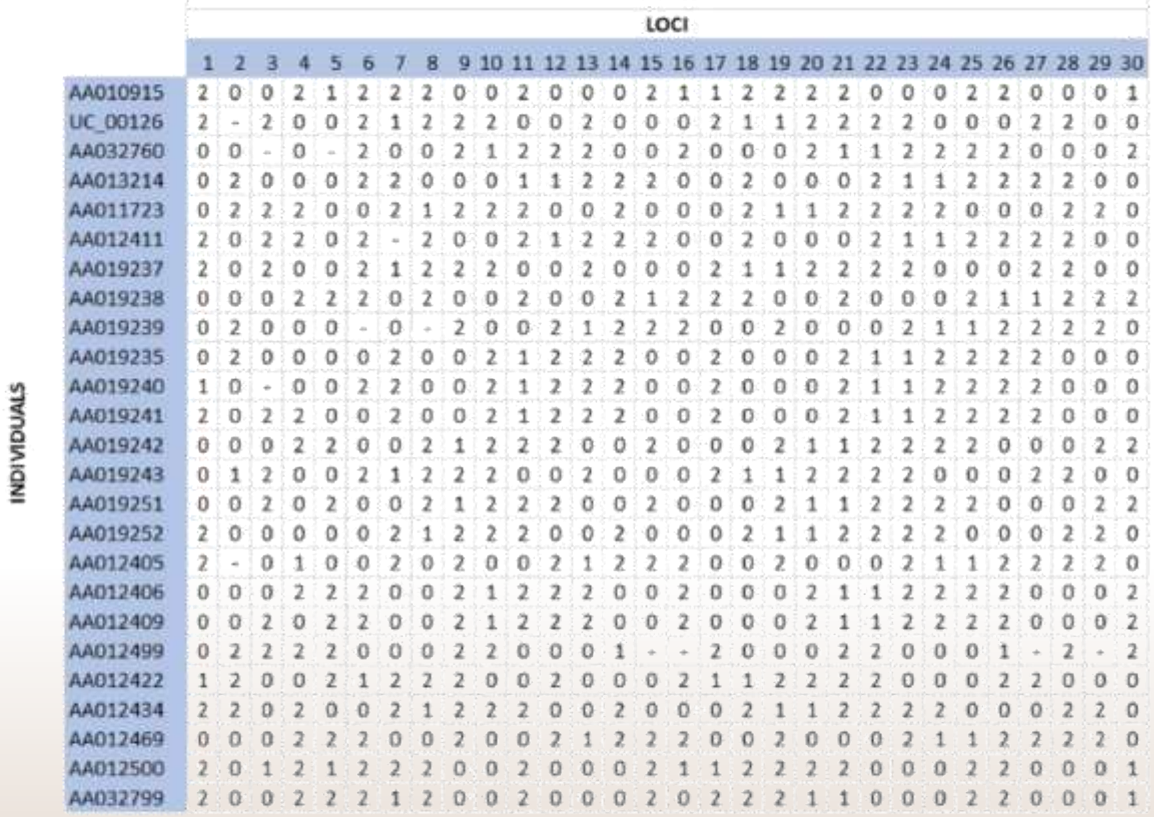
How to learn more

Arthur Georges
UNIVERSITY OF CANBERRA

Diana Robledo-Ruiz
MONASH University

dartR
Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials



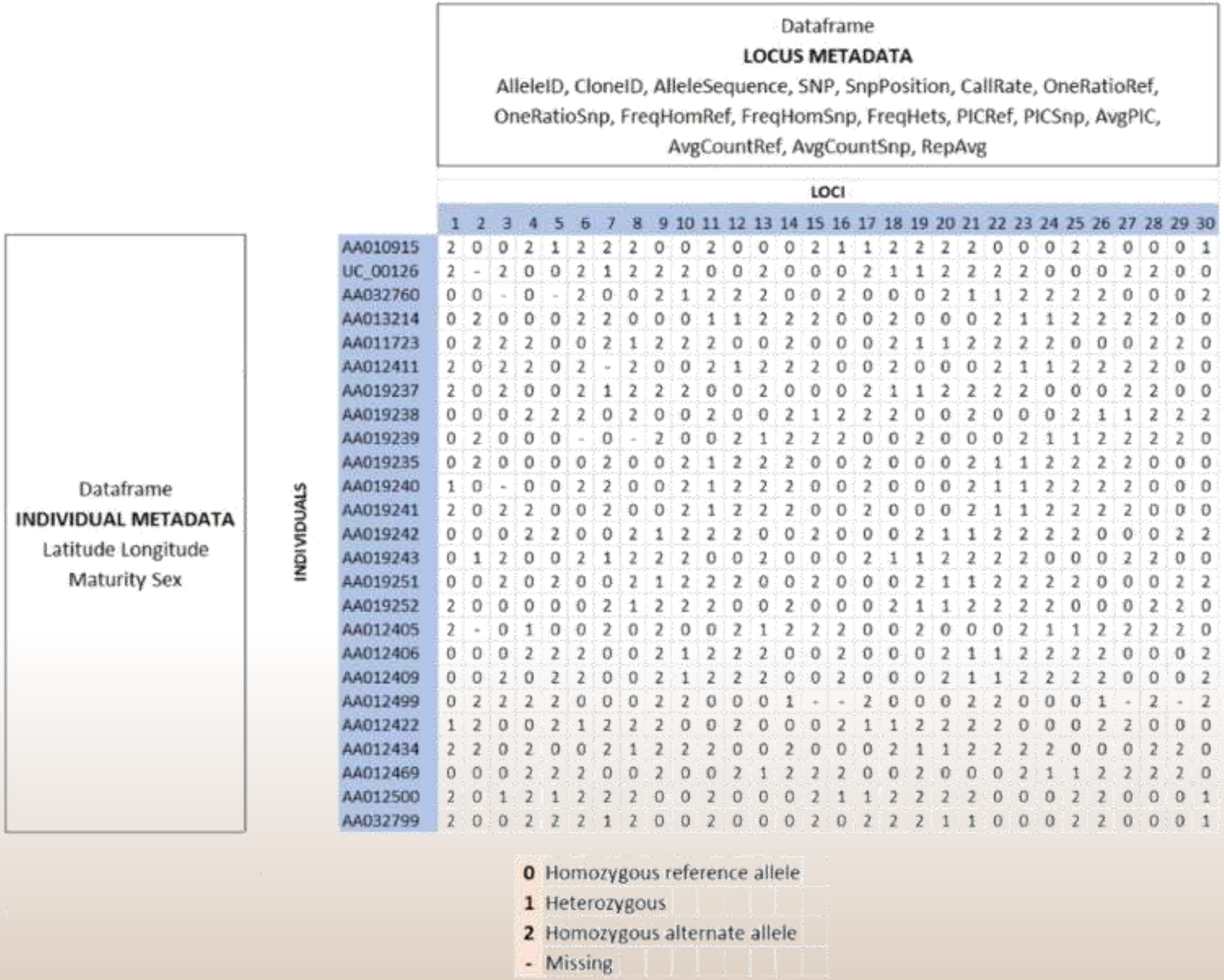
INDIVIDUALS

	LOCI																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
AA010915	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0	1
UC_00126	2	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA032760	0	0	-	0	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA013214	0	2	0	0	0	2	2	0	0	0	1	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0
AA011723	0	2	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012411	2	0	2	2	0	2	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0
AA019237	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA019238	0	0	0	2	2	2	0	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2
AA019239	0	2	0	0	0	-	0	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA019235	0	2	0	0	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019240	1	0	-	0	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019241	2	0	2	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019242	0	0	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2
AA019243	0	1	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA019251	0	0	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2
AA019252	2	0	0	0	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012405	2	-	0	1	0	0	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA012406	0	0	0	2	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA012409	0	0	2	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA012499	0	2	2	2	2	0	0	0	2	2	0	0	0	1	-	-	2	0	0	0	2	2	0	0	0	1	-	2	-	2
AA012422	1	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0
AA012434	2	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012469	0	0	0	2	2	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA012500	2	0	1	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0	1
AA032799	2	0	0	2	2	2	1	2	0	0	2	0	0	0	2	0	2	2	2	1	1	0	0	0	2	2	0	0	0	1

0	Homozygous reference allele
1	Heterozygous
2	Homozygous alternate allele
-	Missing

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials



Dataframe
LOCUS METADATA
AlleleID, CloneID, AlleleSequence, SNP, SnpPosition, CallRate, OneRatioRef, OneRatioSnp, FreqHomRef, FreqHomSnp, FreqHets, PICRef, PICSnp, AvgPIC, AvgCountRef, AvgCountSnp, RepAvg

Dataframe
INDIVIDUAL METADATA
Latitude Longitude
Maturity Sex

LOCI

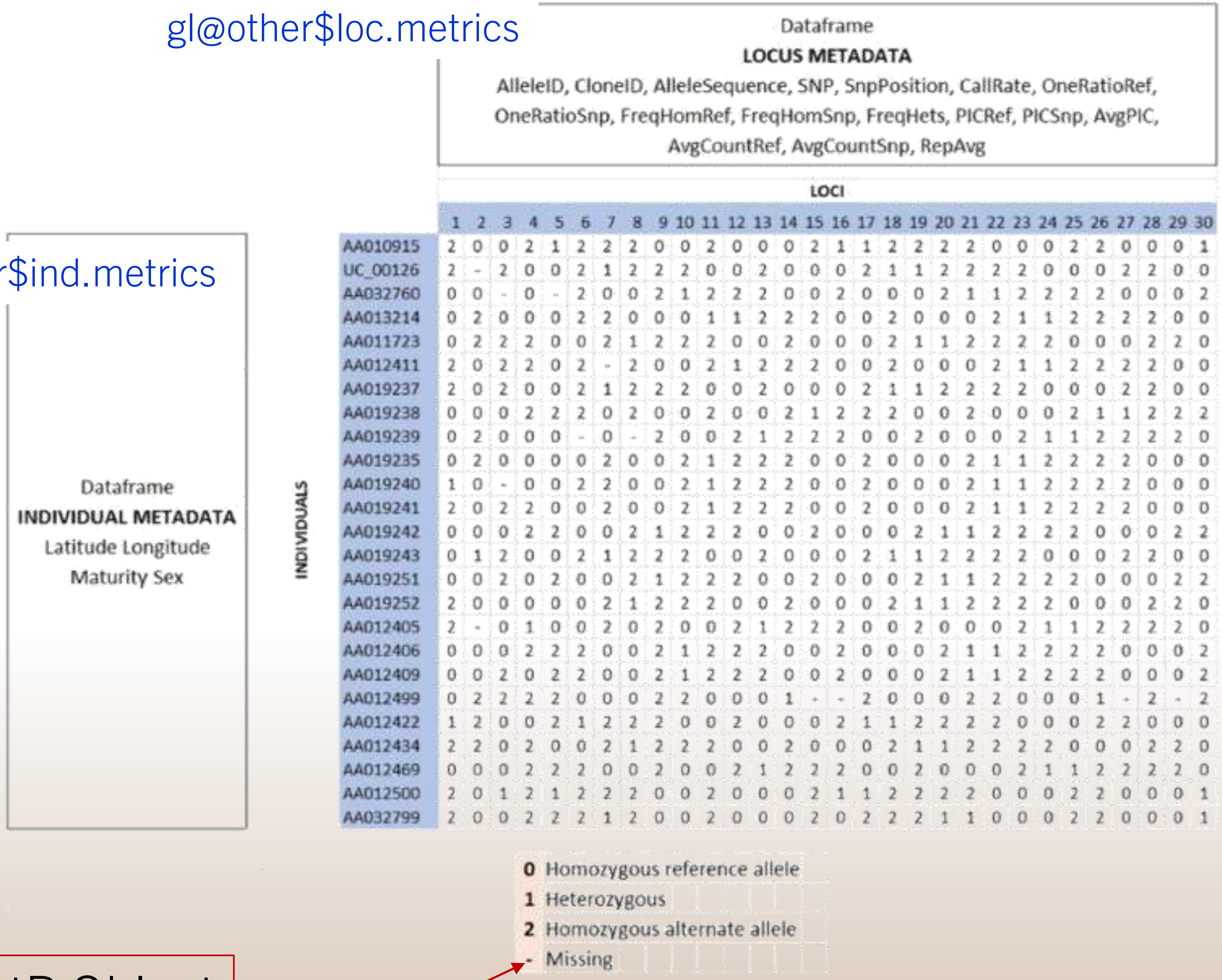
INDIVIDUALS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
AA010915	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0	1
UC_00126	2	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA032760	0	0	-	0	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA013214	0	2	0	0	0	2	2	0	0	0	1	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0
AA011723	0	2	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012411	2	0	2	2	0	2	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0
AA019237	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA019238	0	0	0	2	2	2	0	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2
AA019239	0	2	0	0	0	-	0	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA019235	0	2	0	0	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019240	1	0	-	0	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019241	2	0	2	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019242	0	0	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2
AA019243	0	1	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA019251	0	0	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2
AA019252	2	0	0	0	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012405	2	-	0	1	0	0	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA012406	0	0	0	2	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA012409	0	0	2	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA012499	0	2	2	2	2	0	0	0	2	2	0	0	0	1	-	-	2	0	0	0	2	2	0	0	0	1	-	2	-	2
AA012422	1	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0
AA012434	2	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012469	0	0	0	2	2	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA012500	2	0	1	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0	1
AA032799	2	0	0	2	2	2	1	2	0	0	2	0	0	0	2	0	2	2	2	1	1	0	0	0	2	2	0	0	0	1

0	Homozygous reference allele
1	Heterozygous
2	Homozygous alternate allele
-	Missing

dartR Fundamentals

- **Data structure**
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials



gl@other$loc.metrics

Dataframe
LOCUS METADATA
AlleleID, CloneID, AlleleSequence, SNP, SnpPosition, CallRate, OneRatioRef, OneRatioSnp, FreqHomRef, FreqHomSnp, FreqHets, PICRef, PICSnp, AvgPIC, AvgCountRef, AvgCountSnp, RepAvg

gl@other$ind.metrics

Dataframe
INDIVIDUAL METADATA
Latitude Longitude
Maturity Sex

LOCI

INDIVIDUALS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
AA010915	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0	1
UC_00126	2	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA032760	0	0	-	0	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA013214	0	2	0	0	0	2	2	0	0	0	1	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0
AA011723	0	2	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012411	2	0	2	2	0	2	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0
AA019237	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA019238	0	0	0	2	2	2	0	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2
AA019239	0	2	0	0	0	-	0	-	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA019235	0	2	0	0	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019240	1	0	-	0	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019241	2	0	2	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0
AA019242	0	0	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2
AA019243	0	1	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0
AA019251	0	0	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2
AA019252	2	0	0	0	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012405	2	-	0	1	0	0	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA012406	0	0	0	2	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA012409	0	0	2	0	2	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2
AA012499	0	2	2	2	2	0	0	0	2	2	0	0	0	1	-	-	2	0	0	0	2	2	0	0	0	1	-	2	-	2
AA012422	1	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0
AA012434	2	2	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0
AA012469	0	0	0	2	2	2	0	0	2	0	0	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0
AA012500	2	0	1	2	1	2	2	2	0	0	2	0	0	0	2	1	1	2	2	2	2	0	0	0	2	2	0	0	0	1
AA032799	2	0	0	2	2	2	1	2	0	0	2	0	0	0	2	0	2	2	2	1	1	0	0	0	2	2	0	0	0	1

0	Homozygous reference allele
1	Heterozygous
2	Homozygous alternate allele
-	Missing

NA

Flags
History
Verbosity Setting
etc

dartR Object

dartR
Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

Follow along if you like.
Maybe use your own
genlight object?

Quick and Nasty

```
testset.gl
```

```
/// GENLIGHT OBJECT /////////

// 250 genotypes,  255 binary SNPs, size: 752 Kb
7868 (12.34 %) missing data

// Basic content
 @gen: list of 250 SNPbin
 @ploidy: ploidy of each individual  (range: 2-2)

// Optional content
 @ind.names:  250 individual labels
 @loc.names:  255 locus labels
 @loc.all:  255 alleles
 @position: integer storing positions of the SNPs
 @pop: population of each individual (group size range: 1-11)
 @other: a list containing: loc.metrics  latlong  ind.metrics  history
   loc.metrics.flags
```

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee’s sessions)
- Tutorials

{adegenet} Accessors

```
gl <- testset.gl
nInd(gl)
[1] 250
nLoc(gl)
[1] 250
nPop(gl)
[1] 30
popNames(gl) or indNames(gl) or locNames(gl)
```

```
table(pop(gl))
EmmacBrisWive  EmmacBurdMist  EmmacBurnBara  EmmacClarJack  EmmacClarYate  EmmacCoopAvin  EmmacCoopCully
           10             10             11              5              5             10             10
EmmacCoopEulb  EmmacFitzAllig  EmmacJohnWari  EmmacMaclGeor  EmmacMaryBoru  EmmacMaryPetr  EmmacMDBBowm
           10              10             10             11              6              4            10
EmmacMDBCond  EmmacMDBCudg  EmmacMDBForb  EmmacMDBGwyd  EmmacMDBMaci EmmacMDBMurrMung  EmmacMDBSanf
          10            10            11             9           10               10            10
EmmacNormJack  EmmacNormLeic  EmmacNormSalt  EmmacRichCasi   EmmacRoss  EmmacRussEube  EmmacTweeUki
            6              1              1             10          10             10            10
EmsubRopeMata  EmvicVictJasp
            5              5
```

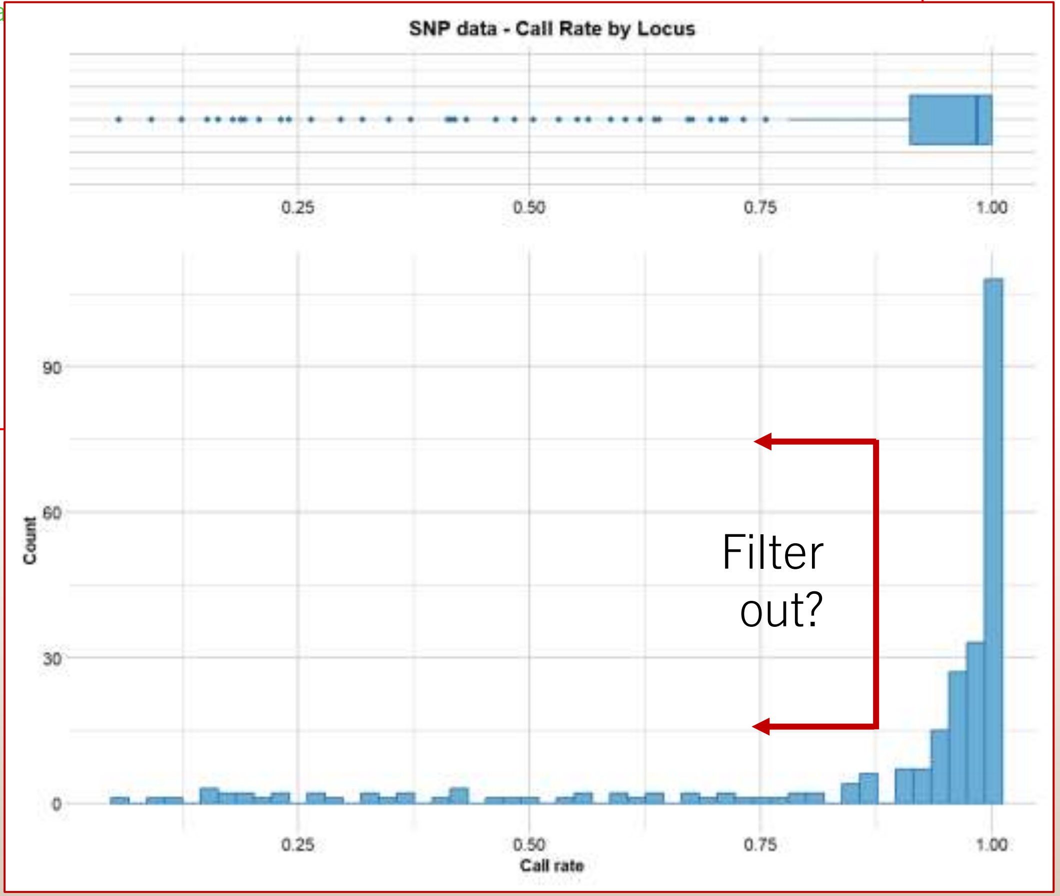
```
as.matrix(gl)[1:7,1:5]
         100049687-12-C/T 100049698-16-G/A 100049728-23-A/G 100049805-56-T/A 100049816-51-A/G
AA010915                2               NA                0                0                0
UC_00126                2               NA                0                0                0
AA032760               NA               NA                0                0                0
AA013214                2               NA                0                0                0
AA011723                2               NA                0                0                0
AA012411                2               NA                0                0                0
AA019237                2               NA                0                0                0
```

dartR Fundamentals

- Data structure
- **Interrogation**
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

gl.report functions

```
gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
[Starting gl.report.callrate
  Processing genlight object with SNP da
  Reporting Call Rate by Locus
  No. of loci = 255
  No. of individuals = 250
    Minimum      :  0.056
    1st quartile :  0.912
    Median       :  0.984
    Mean         :  0.8765804
    3r quartile  :  1
    Maximum      :  1
    Missing Rate Overall:  0.1234

Completed: gl.report.callrate
```



Filter out?

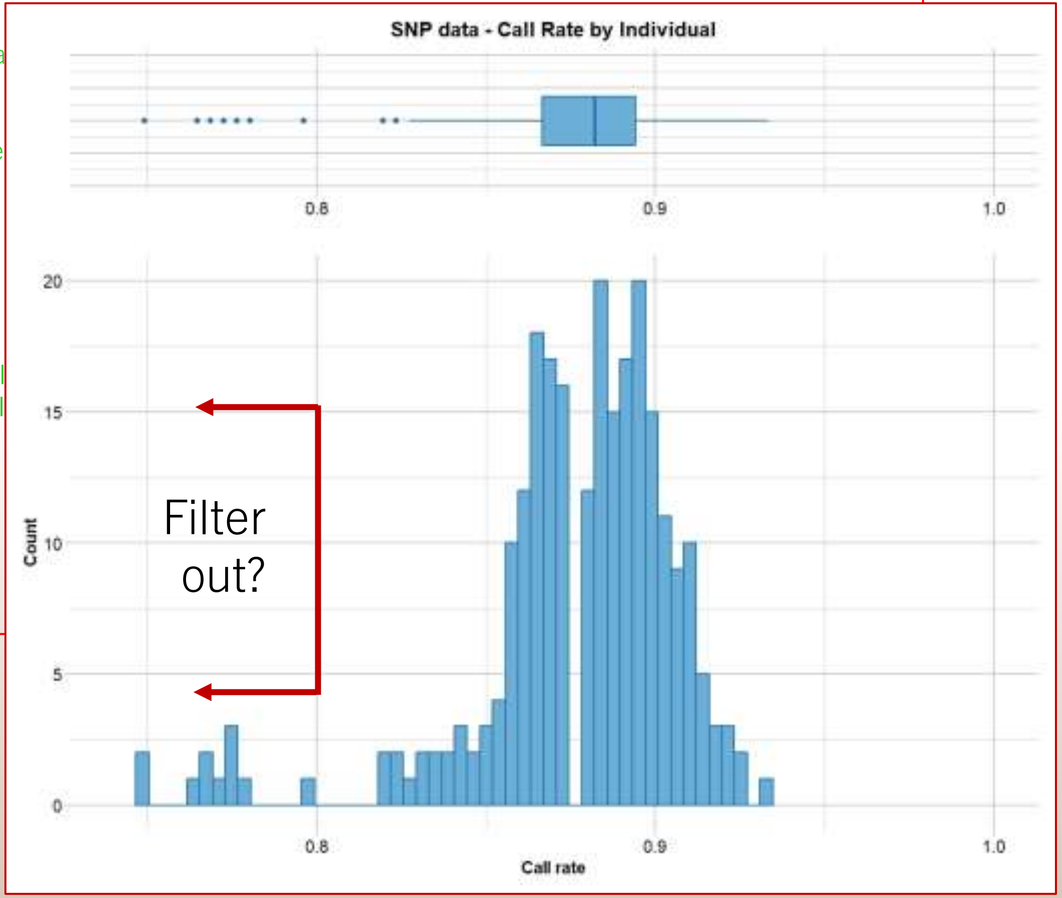
dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

gl.report functions

```
gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
```

```
Starting gl.report.callrate
  Processing genlight object with SNP da
  Reporting Call Rate by Individual

Listing 30 populations and their average
  Monitor again after filtering
      Population CallRate  N
1    EmmacBrisWive   0.8839 10
2    EmmacBurdMist   0.8808 10
3    EmmacBurnBara   0.8859 11
    ……….
Listing 20 individuals with the lowest Cal
  Use this list to see which individuals wil
Individual  CallRate
1    AA063722 0.7490196
2    AA063726 0.7490196
3    AA063732 0.7647059
 ……….
Completed: gl.report.callrate
```



Filter out?

Who will be lost?

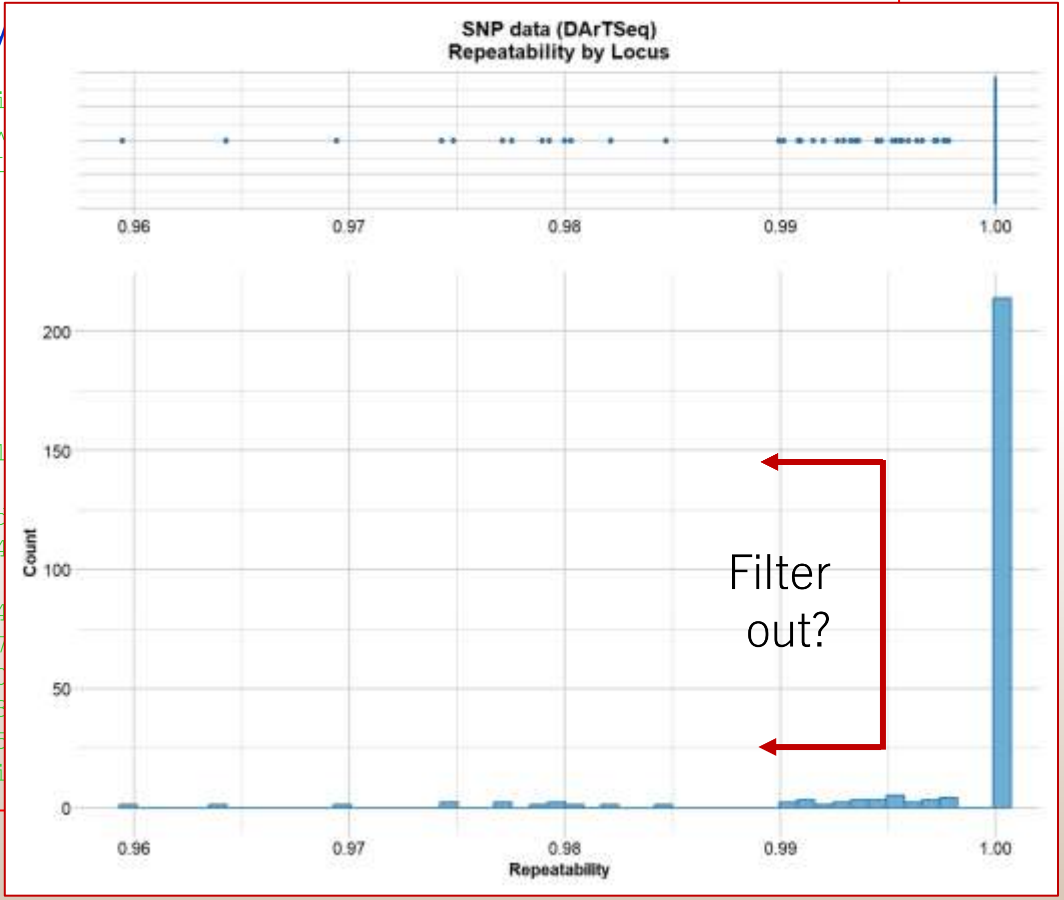
dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

gl.report functions

```
gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility
```

```
gl.report.reproducibility(gl)
Starting gl.report.reproducibi
  Processing genlight object w
  Reporting Repeatability by L
  No. of loci = 255
  No. of individuals = 250
    Minimum       :  0.959459
    1st quartile  :  1
    Median        :  1
    Mean          :  0.9981525
    3r quartile   :  1
    Maximum       :  1
    Missing Rate Overall:  0.1

   Quantile Threshold Retained
1      100%  1.000000      214
...............
17      20%  1.000000      214
18      15%  0.997674      217
19      10%  0.994536      230
20       5%  0.984694      243
21       0%  0.959459      255
Completed: gl.report.reproduci
```



Filter out?

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

You get the gist . . . . .

- gl.report.callrate()
- gl.report.reproducibility()
- gl.report.secondaries()
- gl.report.rdepth()
- gl.report.monomorphs()
- gl.report.overhang()
- gl.report.hamming()
- gl.report.overshoot()
- gl.report.locmetric()

. . . . etc

Keep up . . .

```
testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)
```

Try some of these in your own time

SNP Analysis using dartR

Interrogating and Manipulating Data

I.A.E

Institute for Applied Ecology

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

Subsetting your data

- gl.keep.ind()
- gl.drop.ind()
- gl.keep.loc()
- gl.drop.loc()
- gl.keep.pop()
- gl.drop.pop()
- gl.merge.pop()
- gl.subsample.ind()
- gl.subsample.loc()

……………………… etc

Try some of these in your own time

Keep up …

```
testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)
```

SNP Analysis using dartR

Interrogating and Manipulating Data

I A E

Institute for Applied Ecology

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee’s sessions)
- Tutorials

Recall . . . ..

- gl.**report**.callrate()
- gl.**report**.reproducibility()
- gl.**report**.secondaries()
- gl.**report**.rdepth()
- gl.**report**.monomorphs()
- gl.**report**.overhang()
- gl.**report**.hamming()
- gl.**report**.overshoot()
- gl.**report**.locmetric()

. . . . etc

Keep up . . .

```
testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)
```

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

Basic Filtering

- gl <- gl.**filter**.callrate()
- gl <- gl.**filter**.reproducibility()
- gl <- gl.**filter**.secondaries()
- gl <- gl.**filter**.rdepth()
- gl <- gl.**filter**.monomorphs()
- gl <- gl.**filter**.overhang()
- gl <- gl.**filter**.hamming()
- gl <- gl.filter.overshoot()
- gl <- gl.**report**.locmetric()

……………………… etc

Keep up …

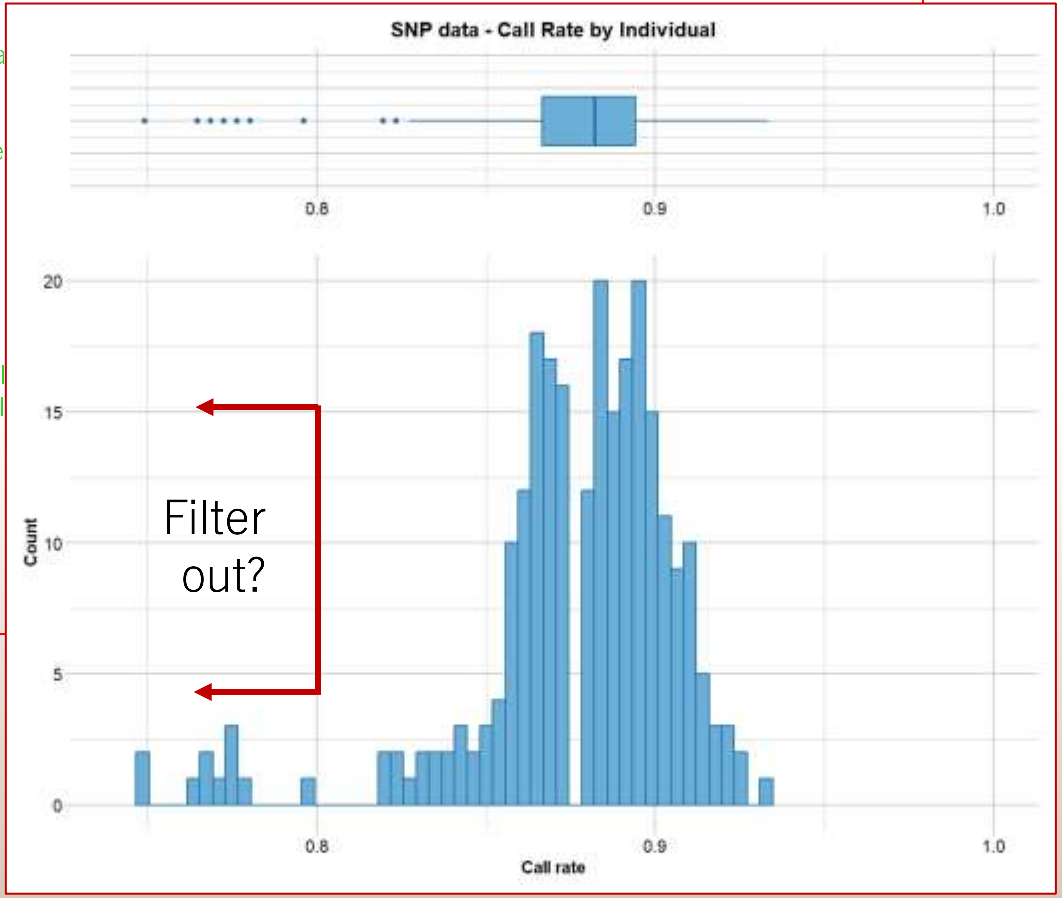
```
testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)
```

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

Recall …..

```
gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
Starting gl.report.callrate
  Processing genlight object with SNP da
  Reporting Call Rate by Individual

Listing 30 populations and their average
  Monitor again after filtering
      Population CallRate  N
1    EmmacBrisWive   0.8839 10
2    EmmacBurdMist   0.8808 10
3    EmmacBurnBara   0.8859 11
    ……….
Listing 20 individuals with the lowest Cal
  Use this list to see which individuals wil
Individual  CallRate
1    AA063722 0.7490196
2    AA063726 0.7490196
3    AA063732 0.7647059
 ……….
Completed: gl.report.callrate
```



Filter out?

Who will be lost?

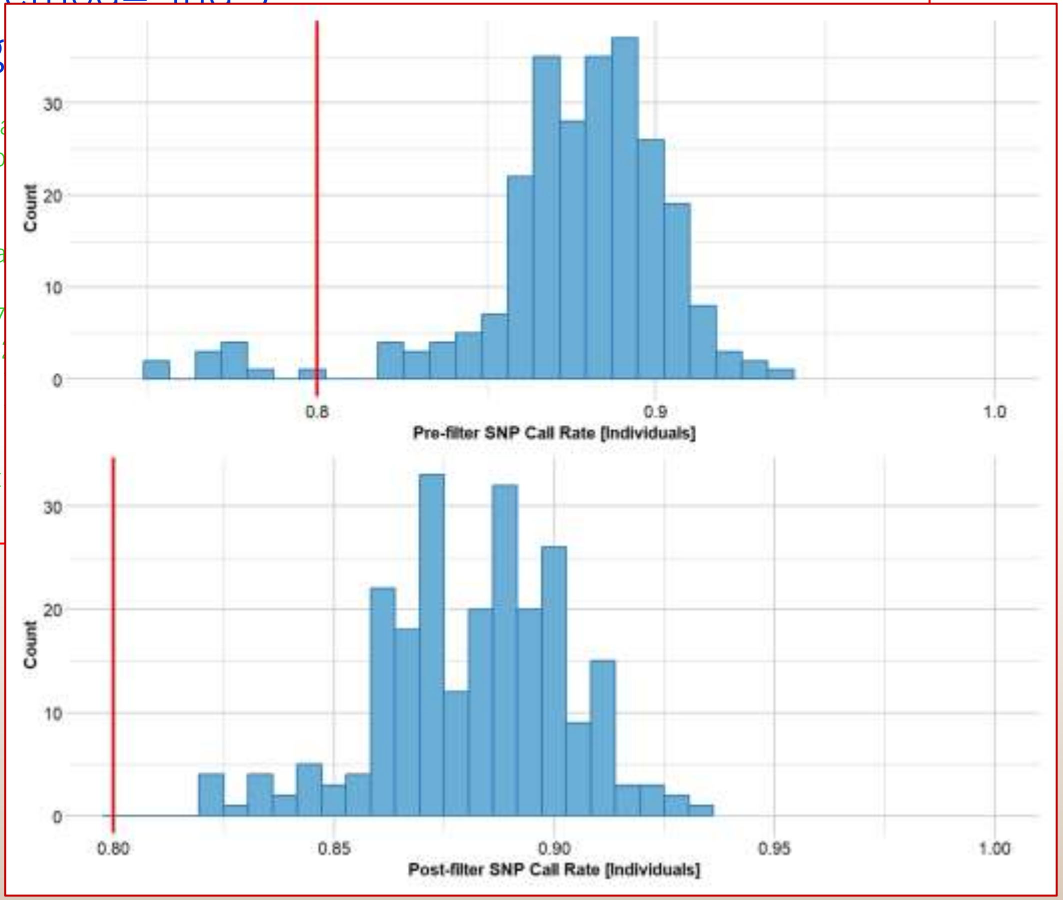
dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- **Basic Filtering**
  (refer Renee’s sessions)
- Tutorials

gl.filter functions

```
gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl <- gl.filter.callrate(g
Starting gl.filter.callrate
  Processing genlight object with SNP da
  Warning: Data may include monomorp
          calculations for filtering
  Recalculating Call Rate
  Removing individuals based on Call Ra
  Individuals deleted (CallRate <=  0.8 ):
AA032760[EmmacMDBMaci], AA0637
AA063722[EmmacCoopAvin], AA0637
AA063708[EmmacCoopAvin], AA0637
AA063714[EmmacCoopAvin], AA0637
  Note: Locus metrics not recalculated
  Note: Resultant monomorphic loci not
Completed: gl.filter.callrate
```

Who has been discarded?

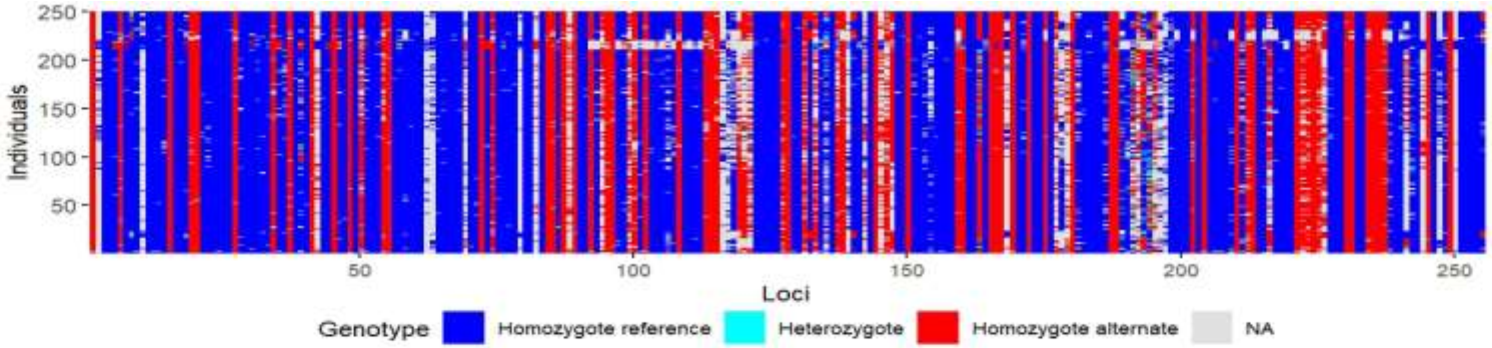


dartR Fundamentals

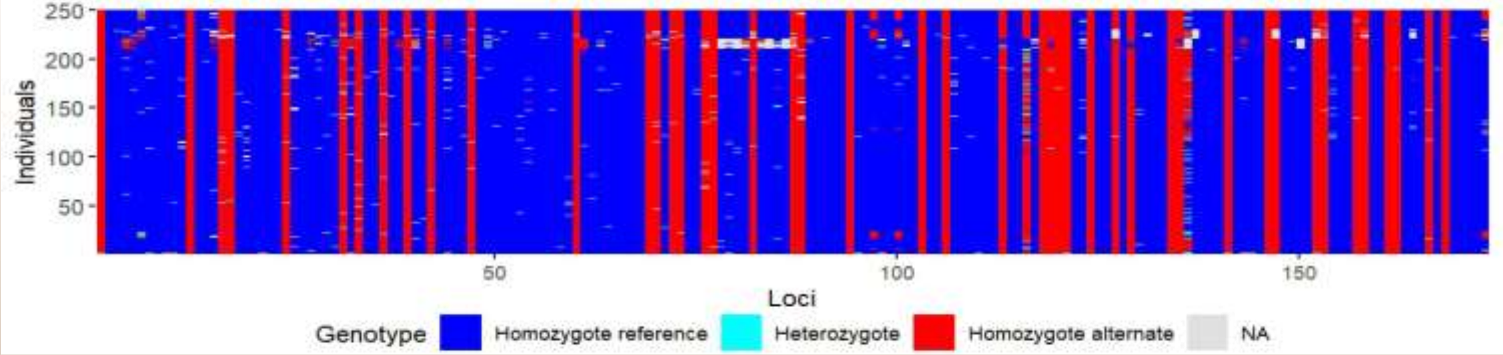
- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee’s sessions)
- Tutorials

Quick graphical evaluation

```
gl <- testset.gl
gl.smearplot(gl)
```



```
gl <- gl.filter.callrate(gl,verbose=0)
gl <- gl.filter.callrate(gl,method=“ind”,threshold=0.80,verbose=0)
gl.smearplot(gl)
```



dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- **Basic Filtering**
  (refer Renee's sessions)
- Tutorials

dartR.sexlinked

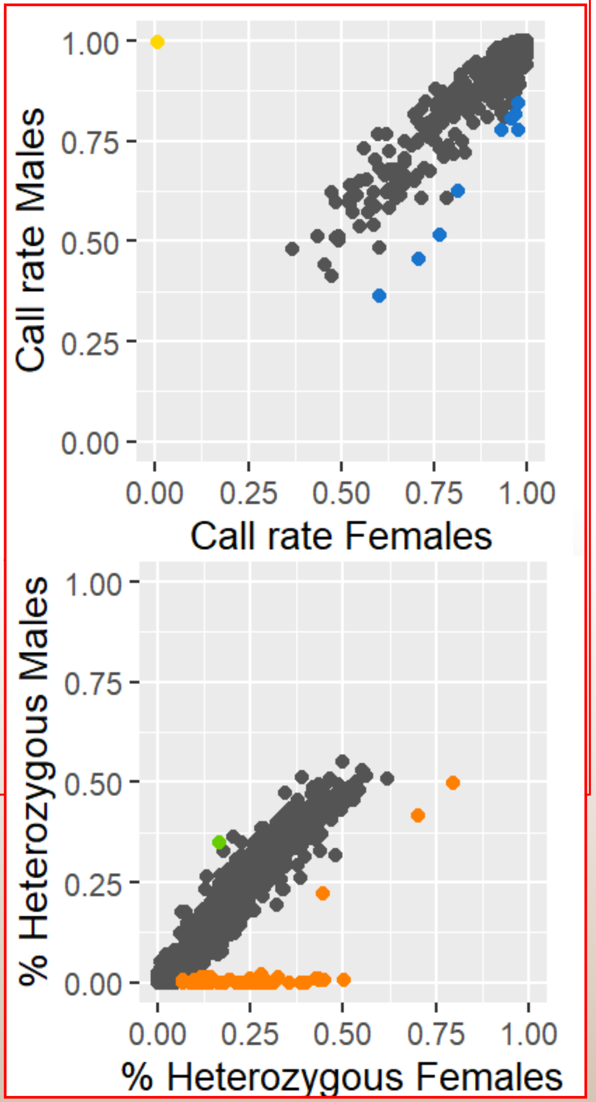
gl@other$ind.metrics

At least 15 males and 15 females

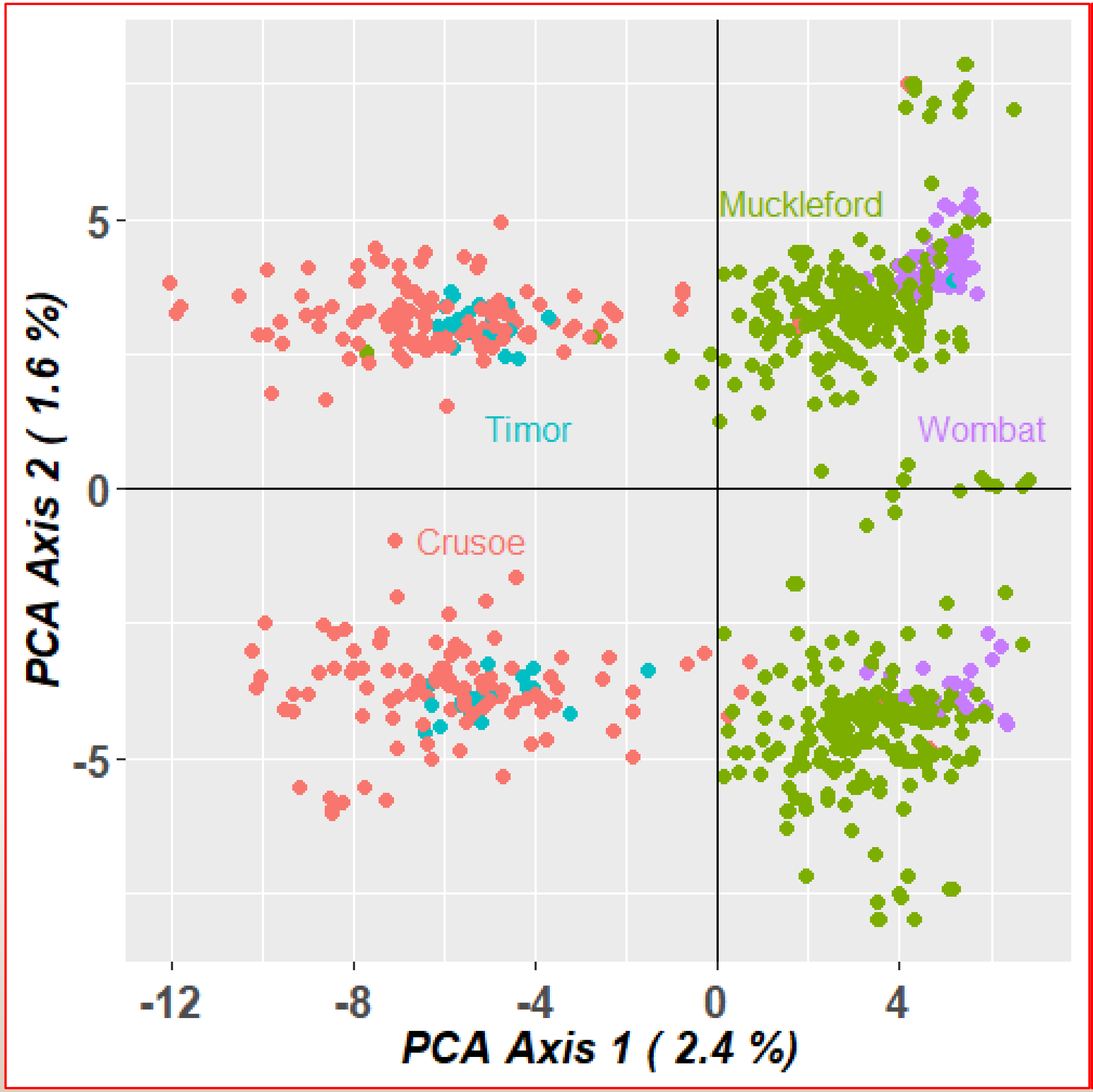
```
out <- gl.report.sexlinked(gl, system = “xy”)
```

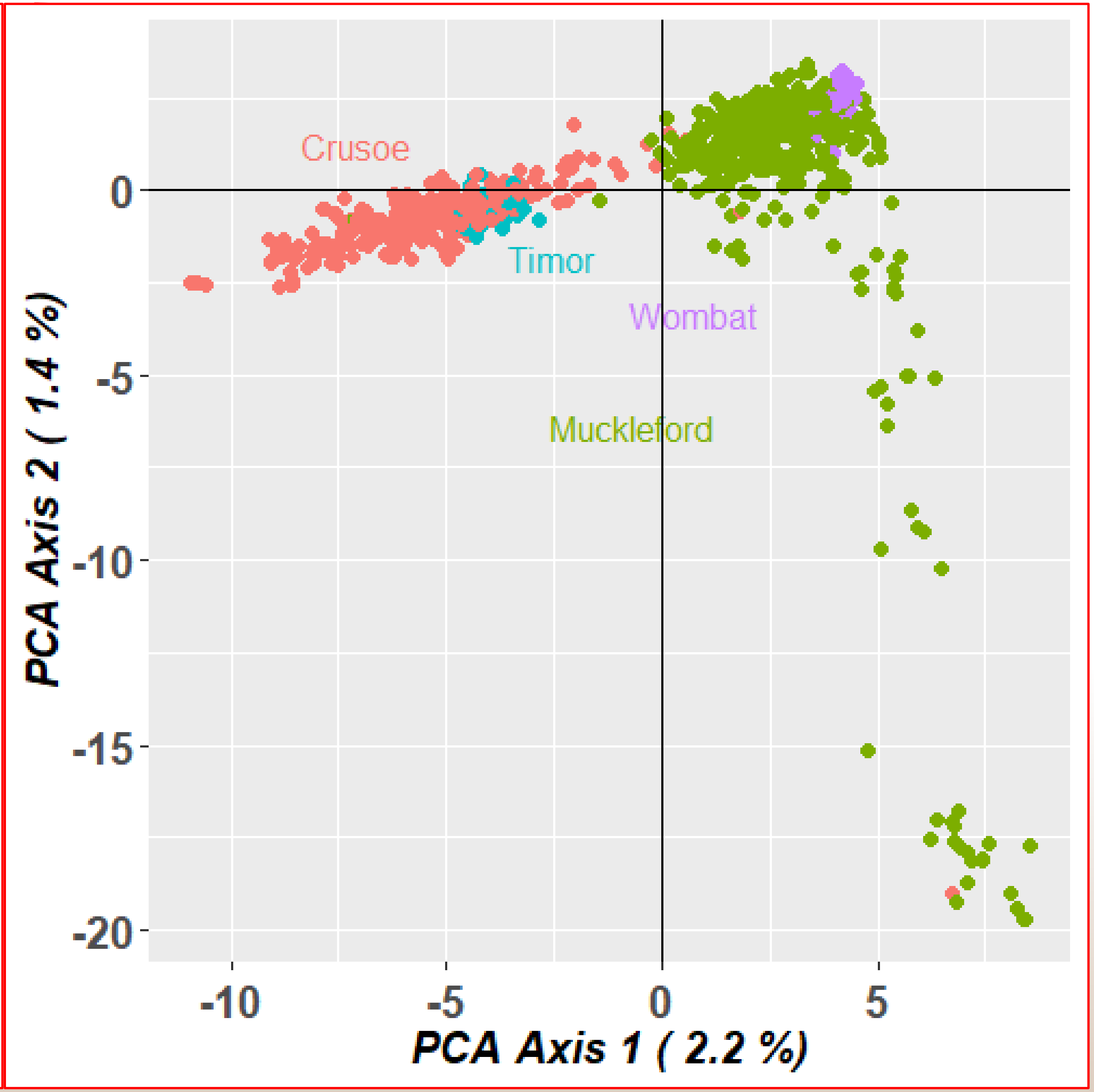
```
Starting dartR.sexlinked
Starting gl.drop.sexlinked
 Processing genlight object with SNP data
Detected 162 females and 211 males.
Starting phase 1. May take a while...
Building call rate plots.
Starting phase 2. May take a while...
Building heterozygosity plots.
Done building heterozygosity plots.
**FINISHED**
Total of analyzed loci: 1000.
Found 77 sex-linked loci:
    1 Y-linked loci (yellow)
    9 sex-biased loci (blue)
    66 X-linked loci (orange)
    1 gametologs (green)
And kept 923 autosomal loci (grey).
Completed: :: Completed: dartR.sexlinked Completed: g
```

```
View(out)
gl.drop.sexlinked(gl, system = “xy”)
gl.keep.sexlinked(gl, system = “xy”)
```



Before all other filters!


Muckleford
Timor
Wombat
Crusoe
5
0
-5
-12
-8
-4
0
4
PCA Axis 2 (1.6 %)
PCA Axis 1 (2.4 %)


Crusoe
Timor
Wombat
Muckleford
0
-5
-10
-15
-20
-10
-5
0
5
PCA Axis 2 (1.4 %)
PCA Axis 1 (2.2 %)

dartR Fundamentals

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials

Try some of these in your own time

Basic Filtering

- `gl <- gl.filter.callrate()`
- `gl <- gl.filter.reproducibility()`
- `gl <- gl.filter.secondaries()`
- `gl <- gl.filter.rdepth()`
- `gl <- gl.filter.monomorphs()`
- `gl <- gl.filter.overhang()`
- `gl <- gl.filter.hamming()`
- `gl <- gl.filter.overshoot()`
- `gl <- gl.report.locmetric()`

……………………… etc

SNP Analysis using dartR

Guide to Basic Filtering

I A E

Institute for Applied Ecology

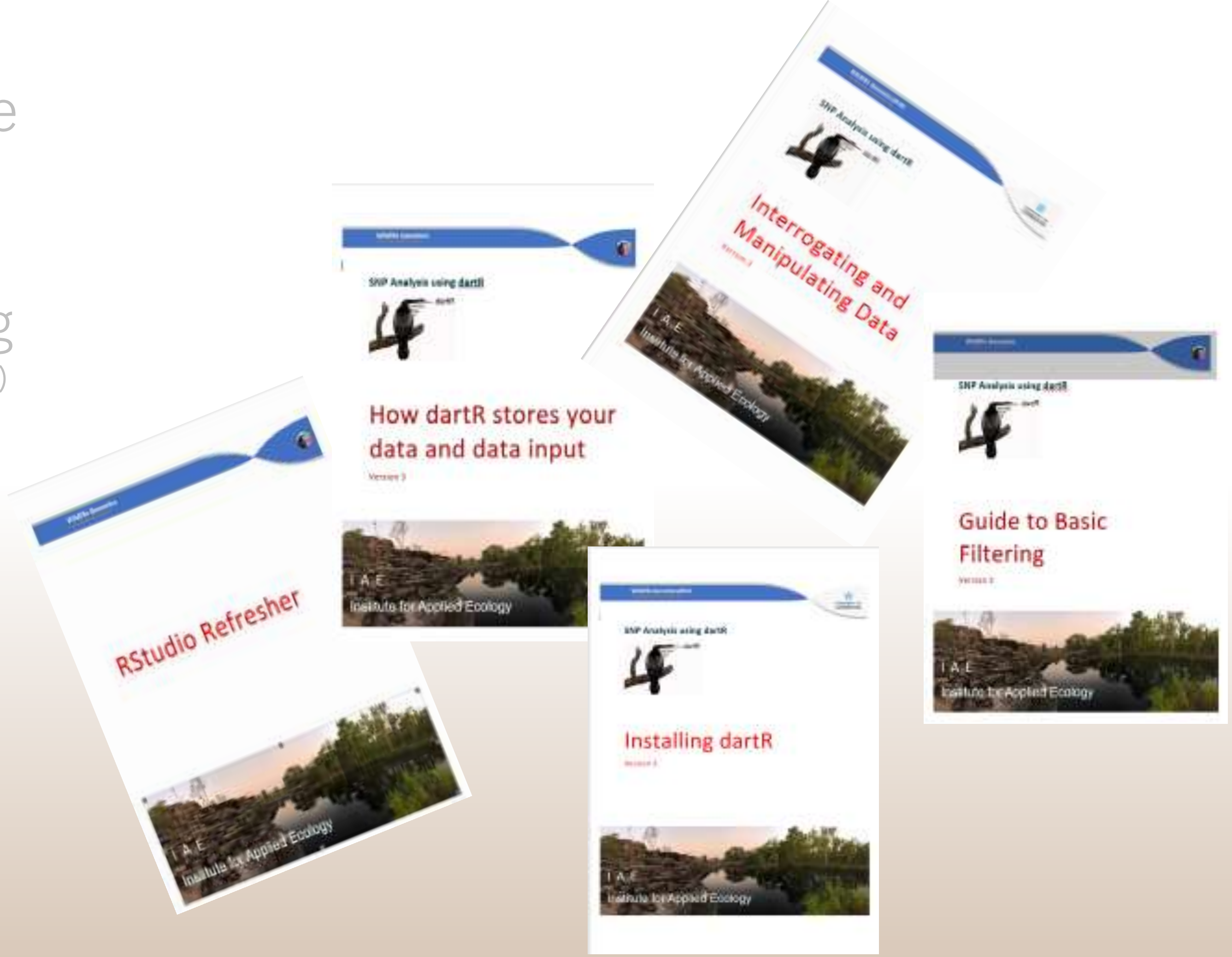
Keep up …

```
testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)

gl.report.callrate(gl,method="ind")
gl <- gl.filter.callrate(gl,
method="ind", threshold=0.80)

gl <- testset.gl
gl.smearplot(gl)
gl <- gl.filter.callrate(gl,verbose=0)
gl <- gl.filter.callrate(gl,
method="ind", threshold=0.80,
verbose=0)
gl.smearplot(gl)
```

dartR Fundamentals

Tutorials

- Data structure
- Interrogation
- Subsetting
- Basic Filtering
  (refer Renee's sessions)
- Tutorials
  - inline help
  - ?function
  - CRAN documentation



RStudio Refresher

I A E Institute for Applied Ecology

SNP Analysis using dartR

How dartR stores your data and data input

I A E Institute for Applied Ecology

SNP Analysis using dartR

Interrogating and Manipulating Data

I A E Institute for Applied Ecology

SNP Analysis using dartR

Installing dartR

I A E Institute for Applied Ecology

SNP Analysis using dartR

Guide to Basic Filtering

I A E Institute for Applied Ecology

Keep up …

testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)

gl.report.callrate(gl,method="ind")
gl <- gl.filter.callrate(gl,
method="ind", threshold=0.80)

gl <- testset.gl
gl.smearplot(gl)
gl <- gl.filter.callrate(gl,verbose=0)
gl <- gl.filter.callrate(gl,
method="ind", threshold=0.80,
verbose=0)
gl.smearplot(gl)

Conclusion
Where have we come?

Discussion

Keep up …

```
testset.gl

gl <- testset.gl
nInd(gl)
nLoc(gl)
nPop(gl)
popNames(gl)
indNames(gl)
locNames(gl)
table(pop(gl))
as.matrix(gl)[1:7,1:5]

gl <- testset.gl
gl.set.verbosity(3)
gl.report.callrate(gl)
gl.report.callrate(gl,method="ind")
gl.report.reproducibility(gl)

gl.report.callrate(gl,method="ind")
gl <- gl.filter.callrate(gl,
method=“ind”, threshold=0.80)

gl <- testset.gl
gl.smearplot(gl)
gl <- gl.filter.callrate(gl,verbose=0)
gl <- gl.filter.callrate(gl,
method=“ind”, threshold=0.80,
verbose=0)
gl.smearplot(gl)
```