Exercises in Marine Ecological Genetics

03. Genetic drift and effective population size

Photograph by Luiz Rocha

- Heterozygosity-based tests continued

- Simulate genetic drift

- Estimate N_e from genetic data

- Model the effect of a bottleneck on N_e

Martin Helmkampf

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Download course materials using git

Go to project directory and check contents

```
cd dir              # e.g. Documents/meg23_exercises
ls -l                           # -l: long format
```

📁 meg23_exercises
├── local
└── meg23_repo
    ├── code
    ├── data
    ├── other
    └── slides

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Download course materials using git

Go to project directory and check contents

```
cd dir          # e.g. Documents/meg23_exercises
ls -l                        # -l: long format
```

Update course repository

```
cd meg23_repo
git pull
```

```
meg23_exercises
├── local
└── meg23_repo
    ├── code
    ├── data
    ├── other
    └── slides
```
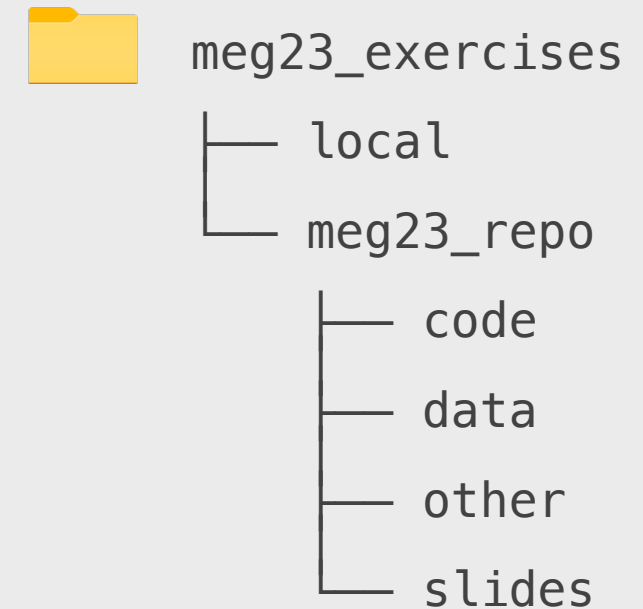
Please do not save over files in the course repository. Instead, save your own scripts to the local subdirectory (including copies of course scripts you would like to edit), e.g with

```
cp code/03_drift.R ../local/03_drift_lc.R          # cp [source] [destination]
```

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

In case of an error message …

Delete old repository and re-download from GitHub

```
rm -rf meg23_repo

git clone https://github.com/mhelmkampf/meg23_repo.git
```

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Genepop format

yellowblue.txt

```
Example data
Color
POP
ind01, yy
ind02, yy
ind03, yy
ind04, yy
ind05, yy
ind06, yy
ind07, yb
ind08, bb
ind09, bb
ind10, bb
```

puella_barbados.txt

```
Microsatellite genotypes of Hypoplectrus puella from Barbados
g2
gag010
h24
hyp001
hyp015
hyp018
e2
hyp008
hyp016
pam013
POP
barbados ind#735 , 203235 121127 210248 225231 126132 190192 157157 236244 194220 114128
barbados ind#736 , 205217 119119 204216 225229 126126 190192 157165 236236 216222 106144
barbados ind#737 , 203203 135147 224226 233235 126136 190198 157163 238250 202226 126160
barbados ind#738 , 211217 119121 228230 223231 126130 190190 157157 240244 220226 138160
barbados ind#739 , 205225 121125 208216 227231 132132 192192 158159 244256 192196 146146
barbados ind#740 , 217233 119121 216228 227233 126126 192198 157157 242256 194216 000000
barbados ind#741 , 203209 000000 216222 223229 126130 192198 157157 234236 188226 118130
barbados ind#742 , 203215 119119 222234 233233 130130 190190 157157 240246 182188 108126
barbados ind#743 , 203243 111121 216224 231231 126126 190192 157159 234249 188198 118126
barbados ind#744 , 203225 121135 206232 231231 126130 192198 157159 240249 182222 122156
barbados ind#745 , 203211 121123 224236 227231 126132 192192 157163 238238 182204 138148
barbados ind#746 , 227233 119123 216234 227233 126130 190190 157159 236245 186188 124142
barbados ind#747 , 203223 133141 236240 231231 126126 192198 157157 234236 194204 118126
barbados ind#748 , 203217 141145 222226 227235 126126 190192 145157 238240 182196 108108
barbados ind#749 , 205221 125147 210238 223231 126132 192198 157157 234234 188200 106148
barbados ind#750 , 203235 123123 230234 223231 126132 192198 157157 236256 188214 122136
barbados ind#751 , 203211 121143 212248 223229 126132 190196 157159 234242 214216 118128
barbados ind#752 , 213217 123169 204234 223231 126134 190198 157163 236248 193224 142148
```

Free text

Locus names

Sample / population

Alleles (000 = missing data)

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Genepop format

yellowblue.txt

```
Example data
Color
POP
ind01, yy
ind02, yy
ind03, yy
ind04, yy
ind05, yy
ind06, yy
ind07, yb
ind08, bb
ind09, bb
ind10, bb
```

puella_barbados.txt

```
Microsatellite genotypes of Hypoplectrus puella from Barbados
g2
gag010
h24
hyp001
hyp015
hyp018
e2
hyp008
hyp016
pam013
POP
barbados ind#735 , 203235 121127 210248 225231 126132 190192 157157 236244 194220 114128
barbados ind#736 , 205217 119119 204216 225229 126126 190192 157165 236236 216222 106144
barbados ind#737 , 203203 135147 224226 233235 126136 190198 157163 238250 202226 126160
barbados ind#738 , 211217 119121 228230 223231 126130 190190 157157 240244 220226 138160
barbados ind#739 , 205225 121125 208216 227231 132132 192192 158159 244256 192196 146146
barbados ind#740 , 217233 119121 216228 227233 126126 192198 157157 242256 194216 000000
barbados ind#741 , 203209 000000 216222 223229 126130 192198 157157 234236 188226 118130
barbados ind#742 , 203215 119119 222234 233233 130130 190190 157157 240246 182188 108126
barbados ind#743 , 203243 111121 216224 231231 126126 190192 157159 234249 188198 118126
barbados ind#744 , 203225 121135 206232 231231 126130 192198 157159 240249 182222 122156
barbados ind#745 , 203211 121123 224236 227231 126132 192192 157163 238238 182204 138148
barbados ind#746 , 227233 119123 216234 227233 126130 190190 157159 236245 186188 124142
barbados ind#747 , 203223 133141 236240 231231 126126 192198 157157 234236 194204 118126
barbados ind#748 , 203217 141145 222226 227235 126126 190192 145157 238240 182196 108108
barbados ind#749 , 205221 125147 210238 223231 126132 192198 157157 234234 188200 106148
barbados ind#750 , 203235 123123 230234 223231 126132 192198 157157 236256 188214 122136
barbados ind#751 , 203211 121143 212248 223229 126132 190196 157159 234242 214216 118128
barbados ind#752 , 213217 123169 204234 223231 126134 190198 157163 236248 193224 142148
```

Free text

Locus names

Sample / population

Alleles (000 = missing data)

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Useful R functions for population genetics

Import Genepop file as genind object

```
x <- read.genepop(file, ncode = n)          # adegenet package
```

Object name

No. of characters encoding each allele

Name of Genepop file, including relative path

(e.g. "../meg23_repo/data/msats/puella_barbados.gen")

Locus statistics (e.g. no. of alleles, heterozygosity, evenness)

```
locus_table(genind)                          # poppr package
```

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Useful R functions for population genetics

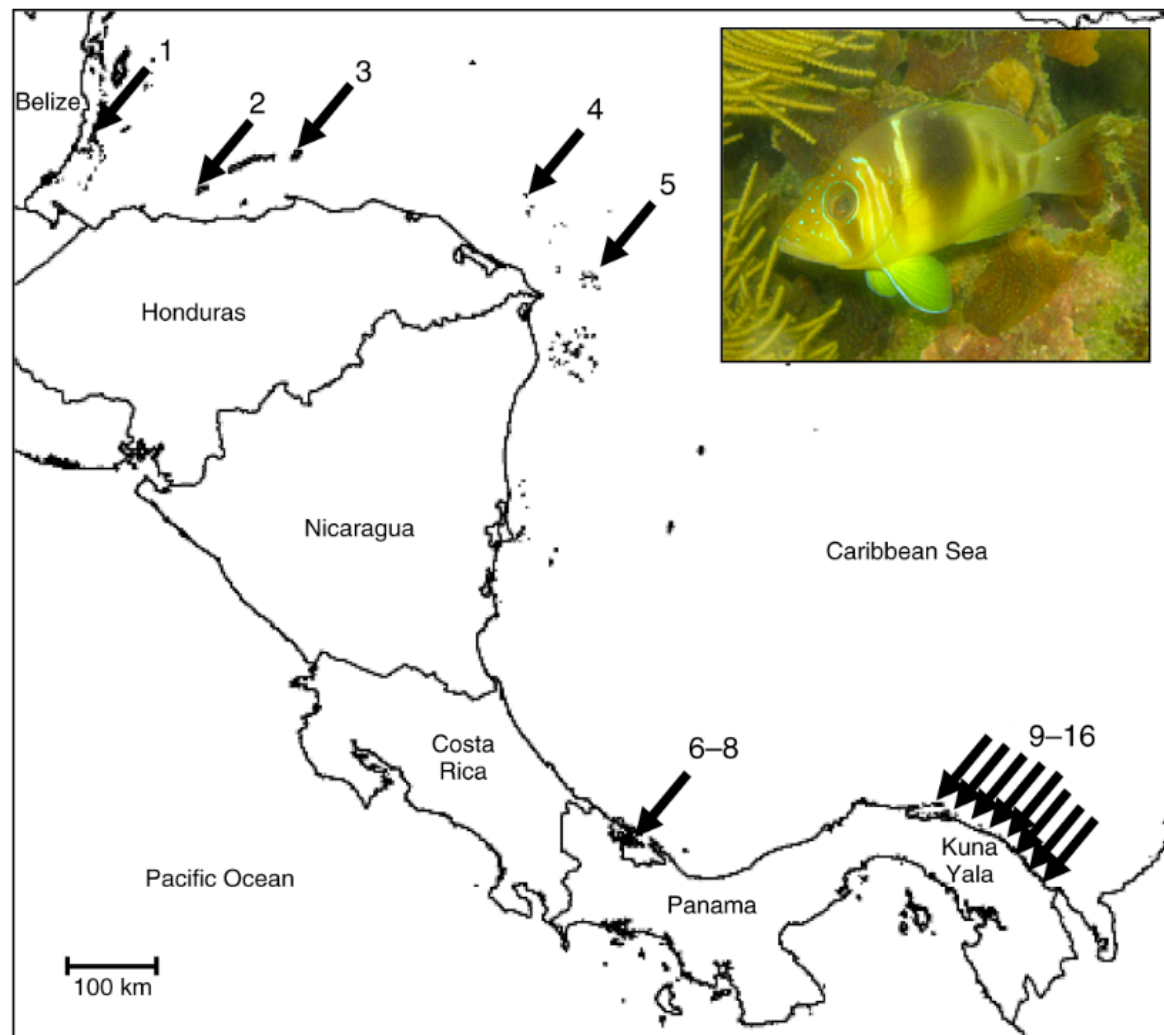Test for departure from Hardy-Weinberg equilibrium

```
hw.test(genind)                              # quick summary (pegas package)


test_HW(file, outputFile = new_file)         # detailed (genepop package)
```

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Barred hamlet (*H. puella*) microsatellite dataset

Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*)

OSCAR PUEBLA,[1,2,3] ELDREDGE BERMINGHAM,[1,2] AND FRÉDÉRIC GUICHARD[2]

- 10 highly variable microsatellite loci

- 854 individuals

- 15 Caribbean locations

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Barred hamlet (*H. puella*) microsatellite dataset

Hamlets are simultaneous hermaphrodites with external fertilization.

Photograph by Luiz Rocha

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Hamlets are simultaneous hermaphrodites with external fertilization.

Photograph by Luiz Rocha

Test whether self-fertilization occurs regularly using population genetic data

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

HETEROZYGOSITY

In one population

H_o = proportion of heterozygote individuals, observed heterozygosity

$H_e = 2pq = 1 - p^2 - q^2$, expected heterozygosity (assuming HW equilibrium)

$$F = \frac{H_e - H_o}{H_e}$$

<u>Fixation index</u>: *proportion by which heterozygosity is reduced or increased relative to the heterozygosity of a population at HW equilibrium with the same allele frequencies.*

Divided by H_e −> *proportion* (of expected heterozygosity)

Varies between -1 and 1

$F < 0$: heterozygote excess
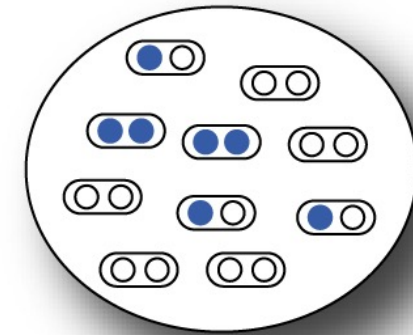
$F > 0$ heterozygote deficit (homozygote excess)

May be averaged over several loci −> reduces bias

May be extended to *k* alleles

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

HETEROZYGOSITY

In one population

H_o = proportion of heterozygote individuals, observed heterozygosity

$H_e = 2pq = 1 - p^2 - q^2$, expected heterozygosity (assuming HW equilibrium)

$$F = \frac{H_e - H_o}{H_e}$$

<u>Fixation index</u>: *proportion by which heterozygosity is reduced or increased relative to the heterozygosity of a population at HW equilibrium with the same allele frequencies.*

Divided by H_e −> *proportion* (of expected heterozygosity)

Varies between -1 and 1

$F < 0$: heterozygote excess

$F > 0$ heterozygote deficit (homozygote excess)

May be averaged over several loci −> reduces bias

May be extended to k alleles

$$\frac{0.445 - 0.1}{0.445} = 0.78$$

Simulate genetic drift

http://evolutiongenetics.georgetown.edu/simulations/

Genetic drift

Total number of replicate populations:	100
Effective population size (N_e) of each population:	100
Number of generations:	200

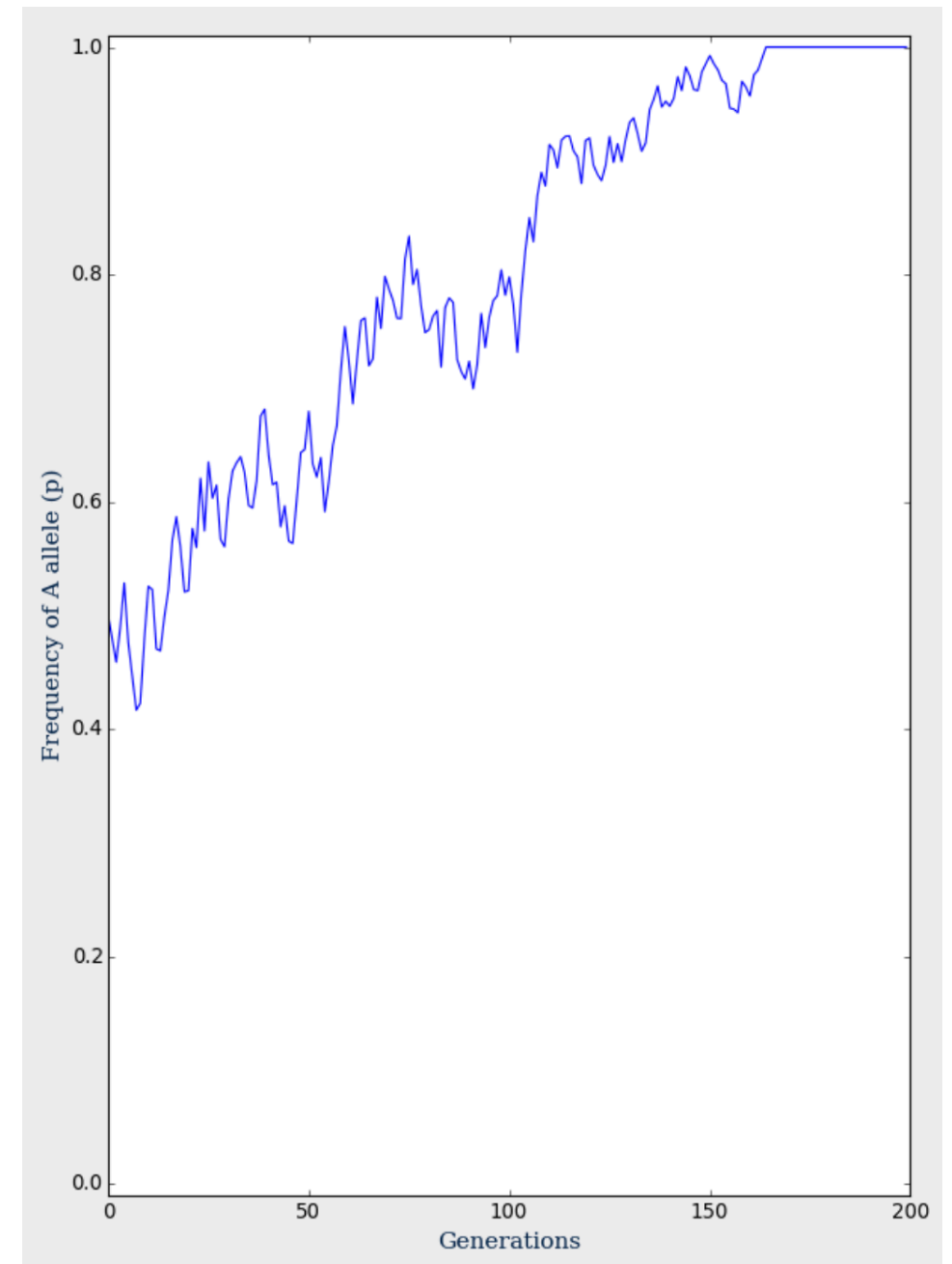**Initial allele frequency (f_A):** 0 ——●—— 1
0.5

Number of replicate populations in allele frequency plot: 1 —●———— 20
1

What do you observe when

- *N*e is high / low

- Initial allele frequency is low / medium / high
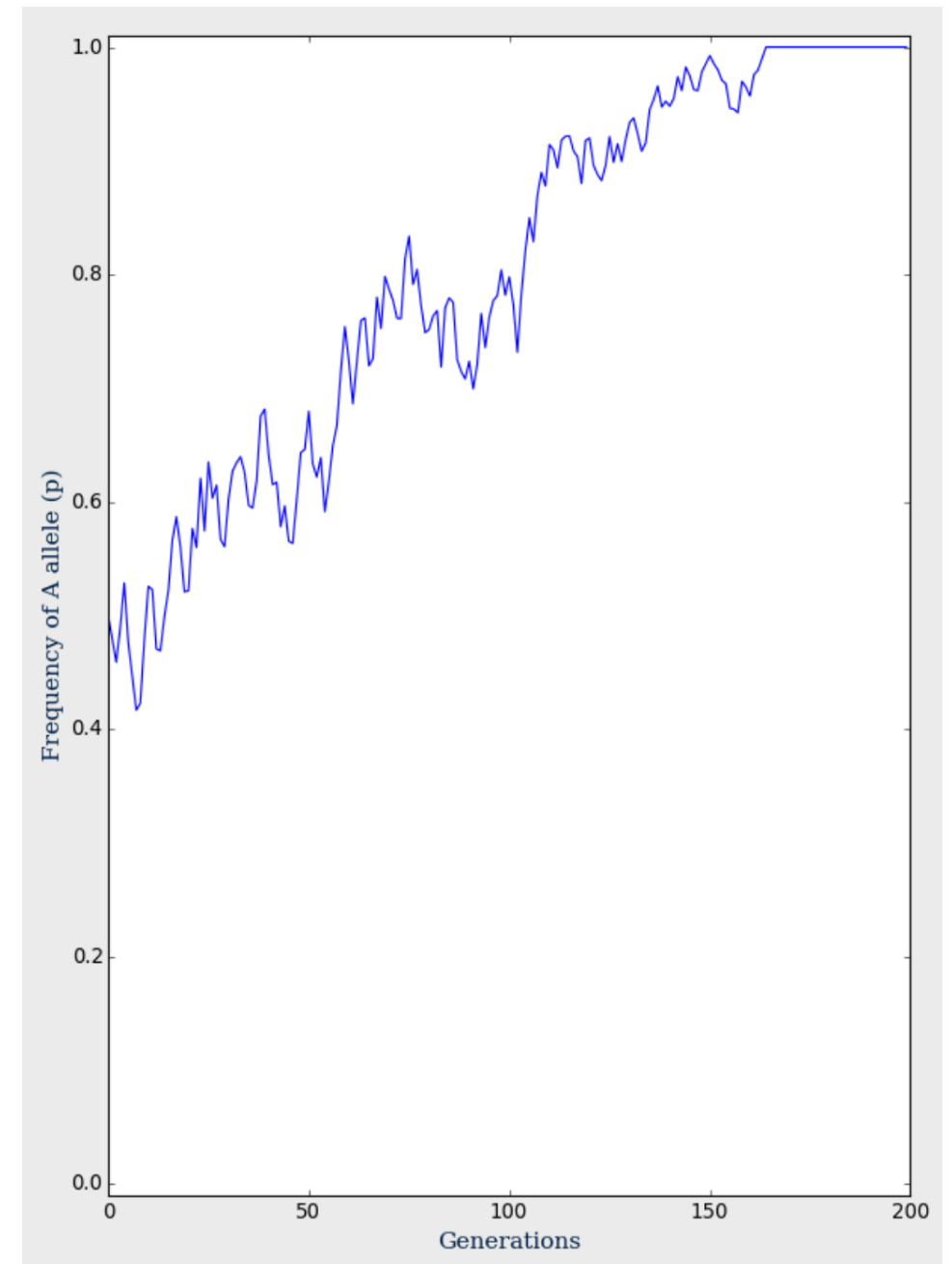
- you run one or many replicate populations?

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Simulate genetic drift

Probability of fixation = initial allele frequency *p*

(over unlimited time)

Time to fixation:

$$\bar{T}_{\text{fixed}} = \frac{-4N_e(1-p)\ln(1-p)}{p}$$

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Estimate N_e using heterozygote excess

Compared with an infinitely large population at Hardy–Weinberg equilibrium, a population generated from a number of N_m male parents and a number of N_f female parents is expected to show a deficit of homozygotes and an excess of heterozygotes at a neutral locus when N_m, N_f or both are small. This is because male and female parents are expected to have different allele frequencies owing to drift. The smaller the value of N_m or N_f, the greater the difference between paternal and maternal allele frequencies and thus the greater the excess in heterozygosity of the offspring population. There is a simple functional relationship between the N_e of the parental population and the amount of heterozygosity excess in the offspring population

— Wang et al. 2016, *Heredity*

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Estimate N_e using heterozygote excess

Exercise 3

Equation (23) suggests that measuring the heterozygosity excess, D, at a number of marker loci in a population yields an estimate of the parental population effective size. Pudovkin *et al.* (1996) proposed such a N_e estimator by accounting for the sampling effect,

$$\hat{N}_e = \frac{1}{2\hat{D}} + \frac{1}{2(\hat{D}+1)},$$

— Wang et al. 2016, *Heredity*

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Estimate N_e using heterozygote excess

Equation (23) suggests that <mark>measuring the heterozygosity excess, D, at a number of marker loci in a population yields an estimate of the parental population effective size.</mark> Pudovkin *et al.* (1996) proposed such a N_e estimator by accounting for the sampling effect,

$$\hat{N}_e = \frac{1}{2\hat{D}} + \frac{1}{2(\hat{D}+1)},$$

— Wang et al. 2016, *Heredity*

$$D_j(i) = \frac{H_j^{obs}(i) - H_j^{exp}(i)}{H_j^{exp}(i)}$$

Index of heterozygote excess

$H_j(i)$: observed / expected frequency of heterozygotes having allele *i* at locus *j*

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Estimate N_e using heterozygote excess

How reliable do you think the results are?

"This method has enough power and precision to detect a finite Nb when the number of breeders is 30 or less, the progeny sampled numbers 200 or more, and the cumulative number of independent alleles exceeds 80. The method is applicable for situations in which progeny generation is produced by only a small number of parents (as might be the case in highly fecund marine species like mussels, oysters or some fish), …"

— Zhdanova & Pudovkin 2008, *J Hered*

"All these methods have rather low precision, though they are nevertheless often the only means available for inferring Ne or Nb. As they have different advantages and disadvantages, it seems advisable to use them in conjunction and to integrate the estimates thereby obtained."

— Wang et al. 2016, *Heredity*

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Model the effect of a bottleneck on *Ne*

N_e for fluctuating population = harmonic mean of population size over time

Harmonic mean = reciprocal of the arithmetic mean of reciprocals

$$H = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}}$$

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg

Course outline

Class	Date	Topics	Script
01	Apr 14	Introduction, software installation	01_intro.R
02	Apr 21	Hardy-Weinberg equilibrium	02_hwe.R
03	Apr 28	Genetic drift and effective population size	03_drift.R
04	May 05	Population structure and gene flow	04_structure.R
05	May 12	Isolation by distance (lecture online, exercises in person)	05_ibd.R
–	May 19	Himmelfahrt break	–
06	May 26	Whole-genome sequencing and genome assembly	06_genseq.sh
07	Jun 02	Genotyping, SNPs and population genomics	07_snps.R
08	Jun 09	Recombination and linkage disequilibrium	08_linkage.R
–	Jun 16	Student presentations	–
09	Jun 23	Selection and mutation	09_selection.R
10	Jun 30	DNA barcoding	10_barcode.sh
11	Jul 07	Metabarcoding	11_meta.sh
–	Jul 14	To be determined	–

Summer 2023

Exercises in Marine Ecological Genetics
03. Genetic drift and effective population size

Carl von Ossietzky
Universität
Oldenburg