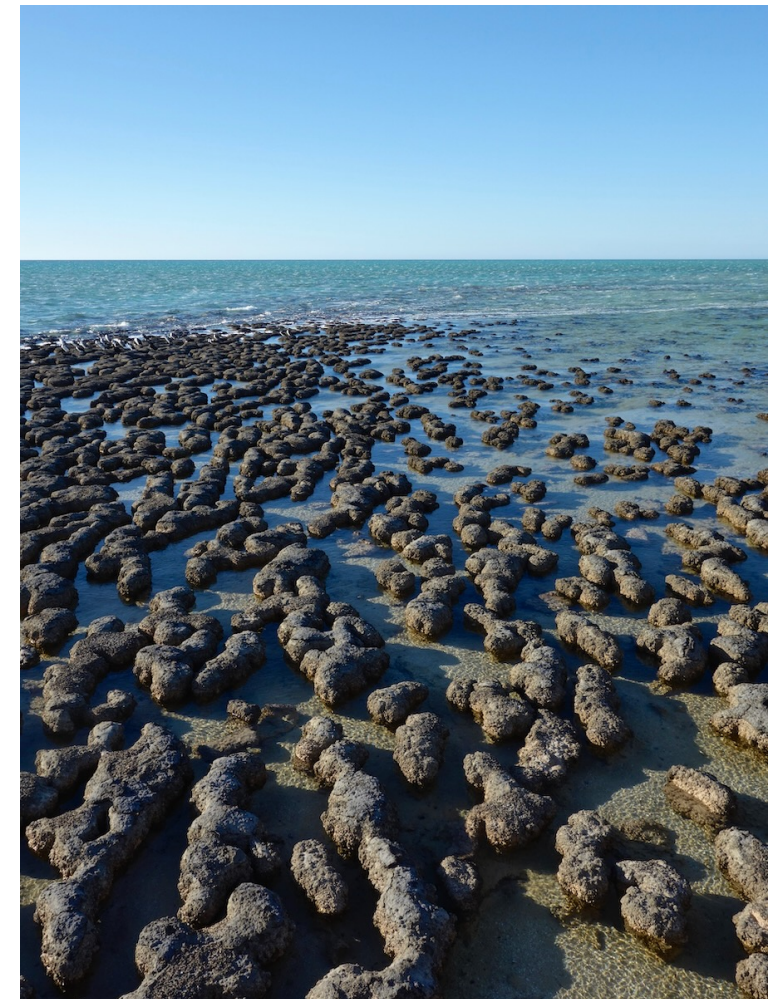# Exercises in Marine Ecological Genetics

04. Population structure and gene flow

- Model the effect of a bottleneck on N_e

- Calculate F-statistics in R

- Test for population differentiation

- Visualize population-specific F_{ST} using ggplot

Martin Helmkampf

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
04. Population structure and gene flow

Carl von Ossietzky
Universität
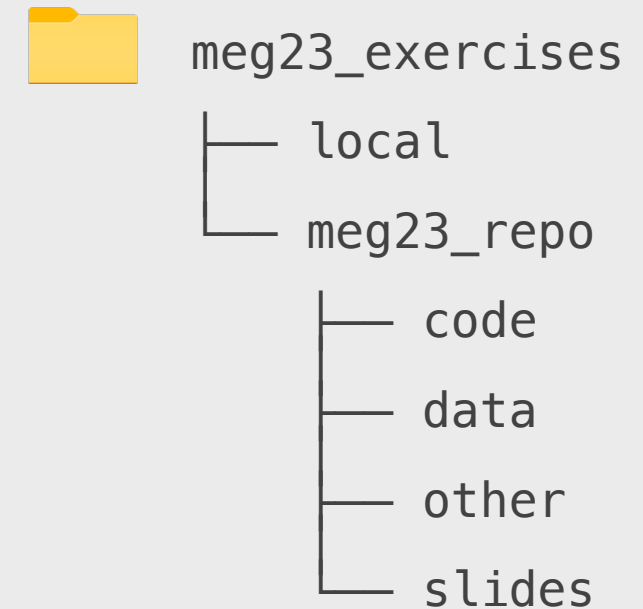Oldenburg

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/04_structure.R ../local/04_structure_lc.R    # cp [source] [destination]
```
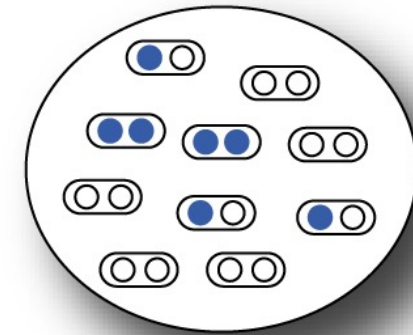
Summer 2023

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

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
```

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

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
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg
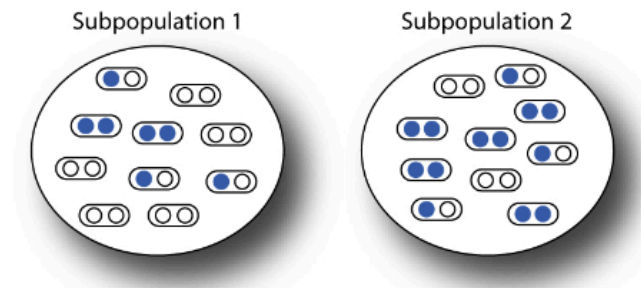
In n subpopulations

Subpopulation 1   Subpopulation 2

$$H_I = \frac{1}{n} \sum_{i=1}^{n} \hat{H}_i$$ Mean observed heterozygosity within subpopulations

$$H_S = \frac{1}{n} \sum_{i=1}^{n} 2p_i q_i$$ Mean expected heterozygosity within subpopulations (assuming HW within subpops)

$$H_T = 2\overline{p}\,\overline{q}$$ Expected heterozygosity of the total population (assuming HW within the total pop)

$\hat{H}_i$ observed heterozygosity in population i $p_i(q_i)$ frequency of allele A(a) in population i

$\overline{p}(\overline{q})$ mean frequency of allele A(a) n : number of subpopulations

$$F_{IS} = \frac{H_S - H_I}{H_S} \qquad F_{ST} = \frac{H_T - H_S}{H_T} \qquad F_{IT} = \frac{H_T - H_I}{H_T}$$

F_{IS} Average difference between observed and Hardy–Weinberg expected heterozygosity within each subpopulation (due to non-random mating)

F_{ST} Reduction in heterozygosity due to subpopulation divergence of allele frequencies
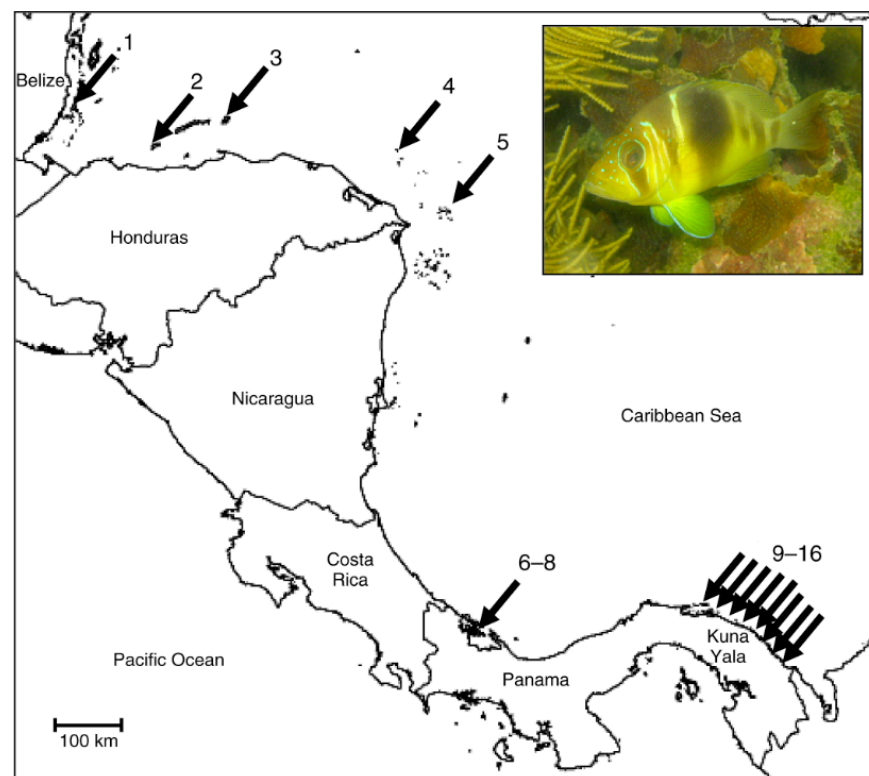
F_{IT} Combined departure from HW expected genotype frequencies due to non-random mating within subpopulations and divergence of allele frequencies among subpopulations

Summer 2023

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

Calculate global F_{ST} and test for differentiation
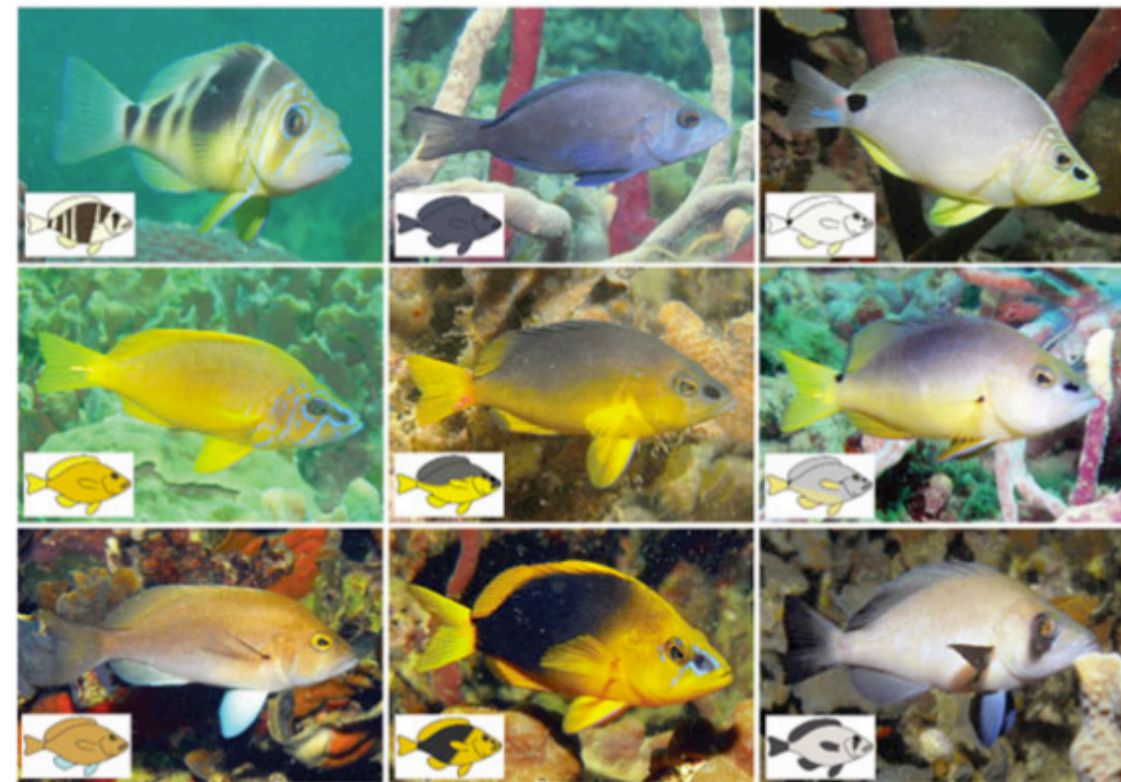
Compare the amount of genetic structure in

- Caribbean populations of *H. puella* (puella_caribbean.gen/.txt)

- a time series of *H. puella* samples from Panama (puella_temporal.gen/.txt)

- Several species of hamlets from various Caribbean locations (hamlets_caribbean.gen/.txt)

— Puebla et al. 2009, *Ecology*

— Puebla et al. 2012, *Proc R Soc B*

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

Species	$\hat{F}_{ST}$	$\widehat{N_e m}$	Reference
Amphibians			
Alytes muletansis (Mallorcan midwife toad)	0.12–0.53	1.8–0.2	Kraaijeveld-Smit et al. 2005
Birds			
Gallus gallus (broiler chicken breed)	0.19	1.0	Emara et al. 2002
Mammals			
Capreolus capreolus (roe deer)	0.097–0.146	2.2–1.4	Wang and Schreiber 2001
Homo sapiens (human)	0.03–0.05	7.8–4.6	Rosenberg et al. 2002
Microtus arvalis (common vole)	0.17	1.2	Heckel et al. 2005
Plants			
Arabidopsis thaliana (mouse-ear cress)	0.643	0.1	Bergelson et al. 1998
Oryza officinalis (wild rice)	0.44	0.3	Gao 2005
Phlox drummondii (annual phlox)	0.17	1.2	Levin 1977
Prunus armeniaca (apricot)	0.32	0.5	Romero et al. 2003
Fish			
Morone saxatilis (striped bass)	0.002	11.8	Brown et al. 2005
Sparisoma viride (stoplight parrotfish)	0.019	12.4	Geertjes et al. 2004
Insects			
Drosophila melanogaster (fruit fly)	0.112	2.0	Singh and Rhomberg 1987
Glossina pallidipes (tsetse fly)	0.18	1.1	Ouma et al. 2005
Heliconius charithonia (butterfly)	0.003	79.8	Kronforst and Flemming 2001
Corals			
Seriatopora hystrix	0.089–0.136	2.6–1.6	Maier et al. 200

Summer 2023

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

Species	$\hat{F}_{ST}$	$\widehat{N_e m}$	Reference
Amphibians			
Alytes muletansis (Mallorcan midwife toad)	0.12–0.53	1.8–0.2	Kraaijeveld-Smit et al. 2005
Birds			
Gallus gallus (broiler chicken breed)	0.19	1.0	Emara et al. 2002
Mammals			
Capreolus capreolus (roe deer)	0.097–0.146	2.2–1.4	
Homo sapiens (human)	0.03–0.05	7.8–4.6	
Microtus arvalis (common vole)	0.17	1.2	
Plants			
Arabidopsis thaliana (mouse-ear cress)	0.643	0.1	
Oryza officinalis (wild rice)	0.44	0.3	
Phlox drummondii (annual phlox)	0.17	1.2	
Prunus armeniaca (apricot)	0.32	0.5	
Fish			
Morone saxatilis (striped bass)	0.002	11.8	Brown et al. 2005
Sparisoma viride (stoplight parrotfish)	0.019	12.4	Geertjes et al. 2004
Insects			
Drosophila melanogaster (fruit fly)	0.112	2.0	Singh and Rhomberg 1987
Glossina pallidipes (tsetse fly)	0.18	1.1	Ouma et al. 2005
Heliconius charithonia (butterfly)	0.003	79.8	Kronforst and Flemming 2001
Corals			
Seriatopora hystrix	0.089–0.136	2.6–1.6	Maier et al. 200

F_{ST}	Subpopulations are …
< 0.15	very similar
0.15 – 0.25	similar
> 0.25	distinct

Summer 2023

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

Summer 2023

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

— Hench et al. 2022, *PNAS*

New R concepts: tidyverse and ggplot

Tidyverse pipes

```
x <- df %>%        # assign to new object (copy)

  mutate() %>%     # 1st operation

  arrange() %>%    # 2nd operation

  select()         # 3rd operation
```
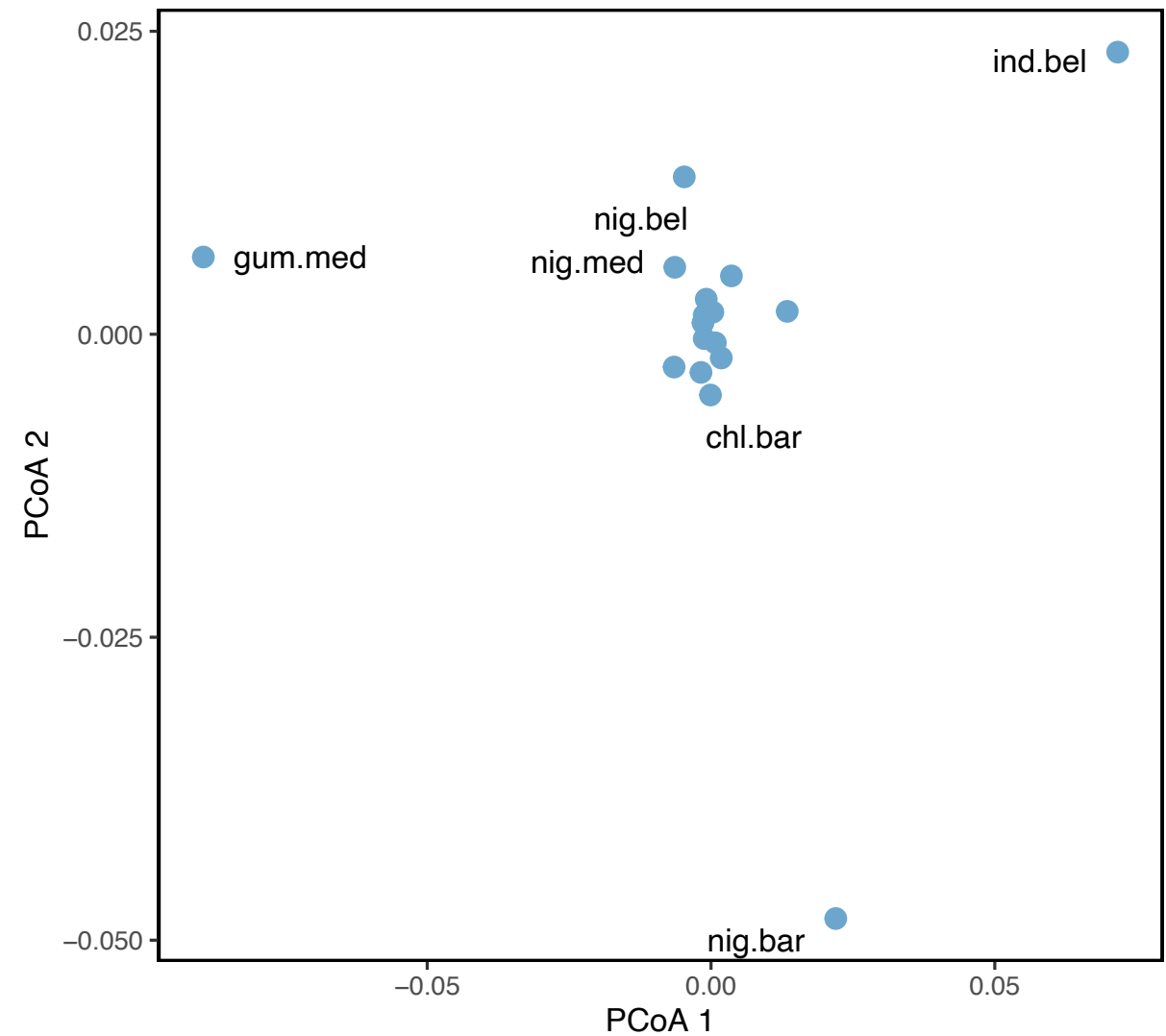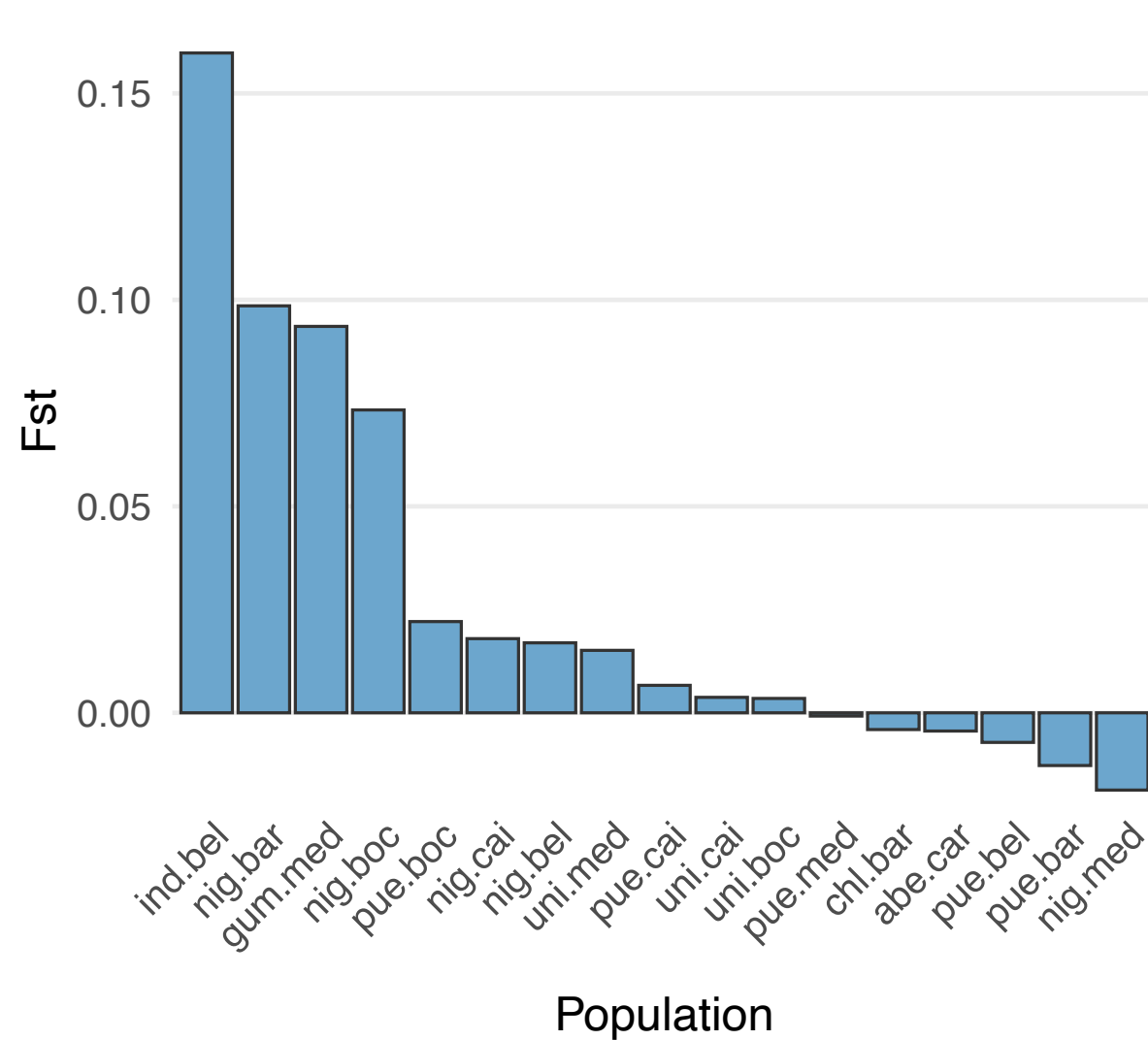
Basic ggplot syntax

```
ggplot(data = df, aes(x = var1, y = var2)) +  # mapping variables

  geom_bar() +     # geometric shapes representing the data

  labs() +         # set plot and axis labels

  guides() +       # customize plot legend

  theme()          # customize non-data, e.g. fonts, gridlines
```
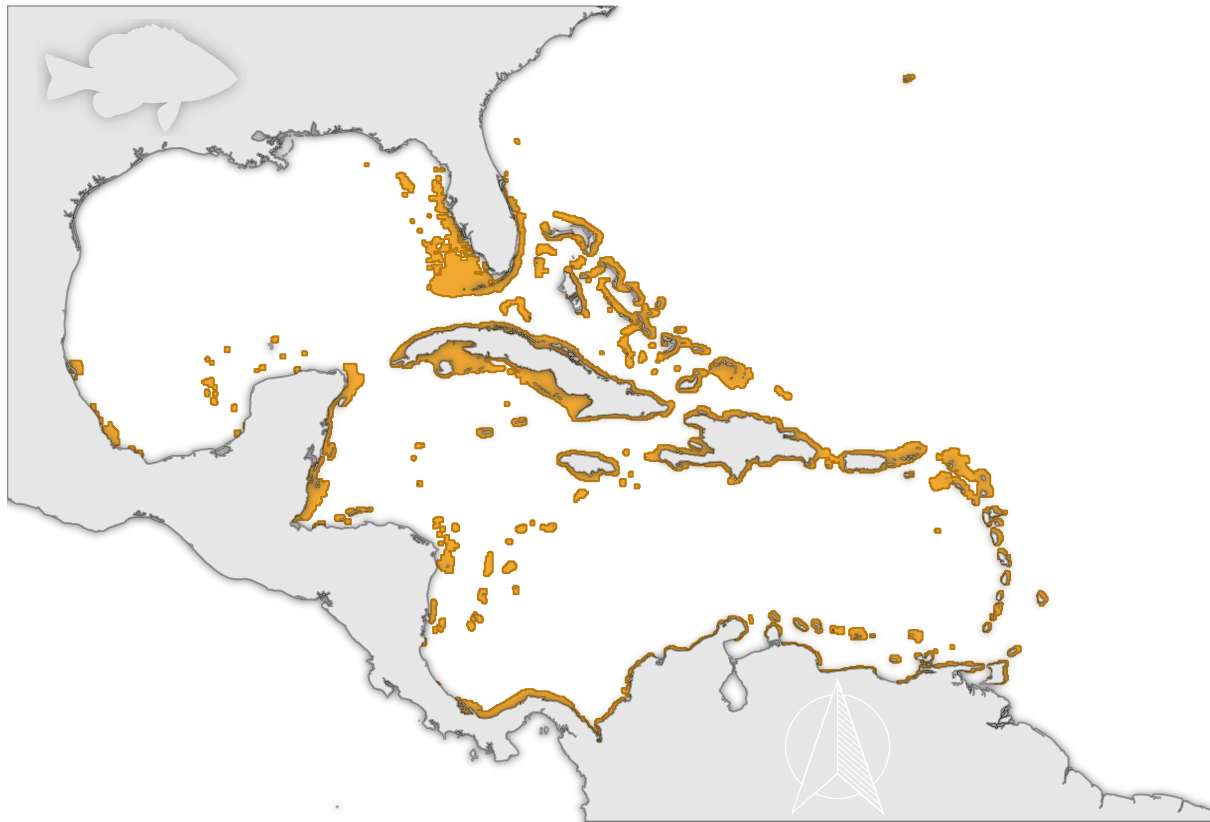
Exercises in Marine Ecological Genetics
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

Calculate population-specific *F*st

Exercises in Marine Ecological Genetics
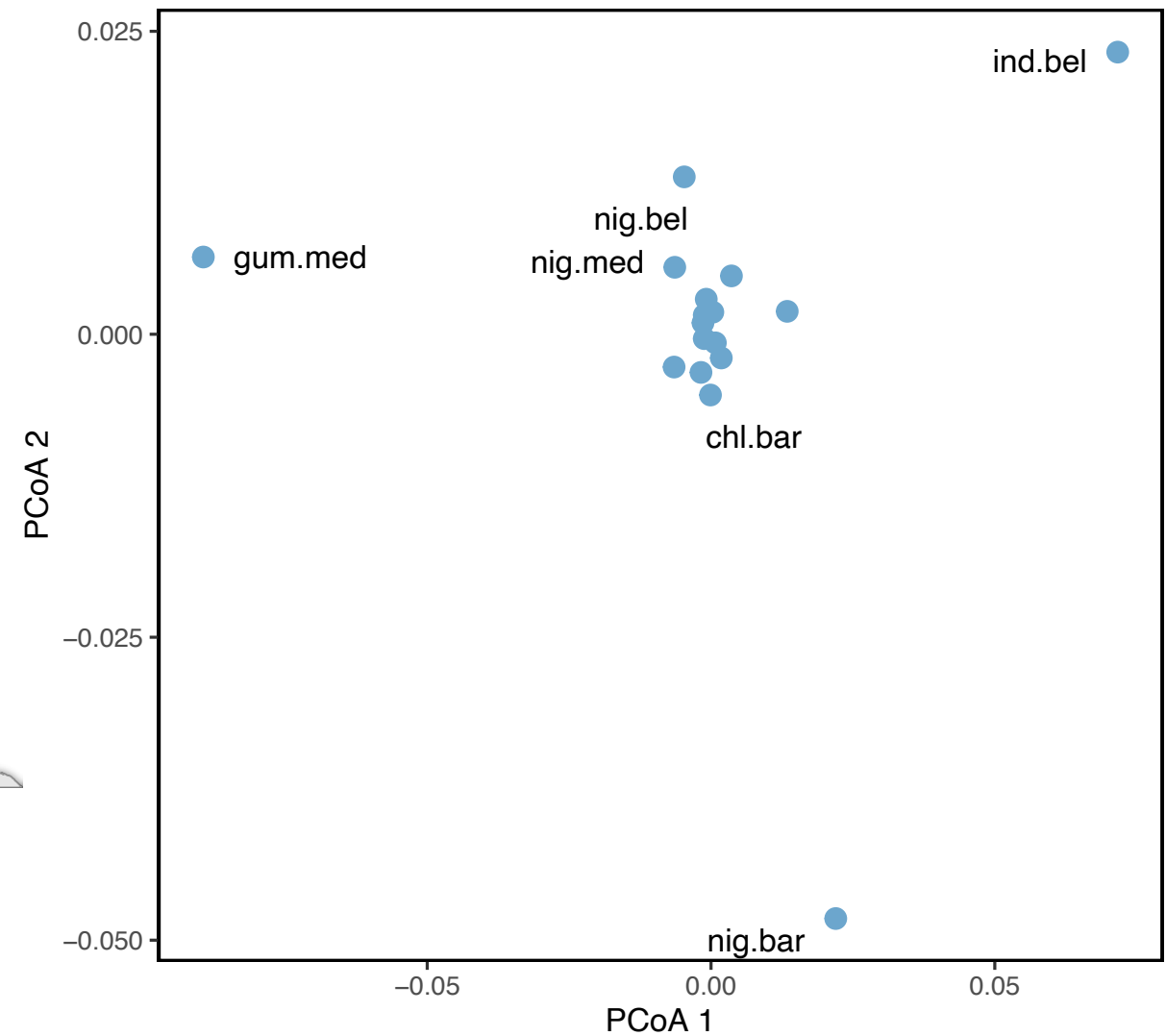04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg

Calculate population-specific *F*st

Map by Kosmas Hench

data from biogeodb.stri.si.edu (Robertson & Van Tassell 2019)

Exercises in Marine Ecological Genetics
04. Population structure and gene flow

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
04. Population structure and gene flow

Carl von Ossietzky
Universität
Oldenburg