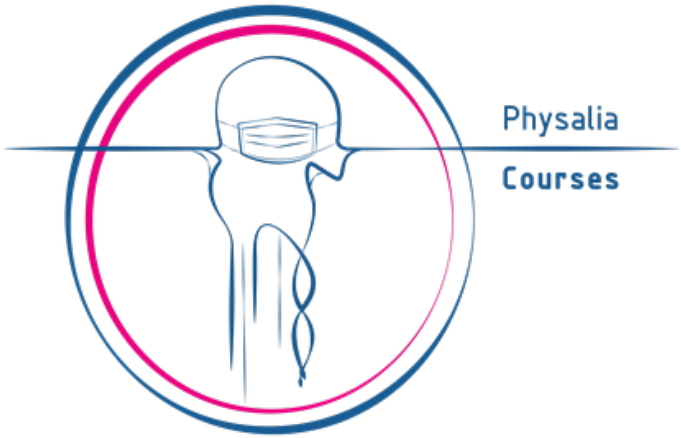


Course introduction

Population genomic inference from low-coverage whole genome sequencing data

Course goals

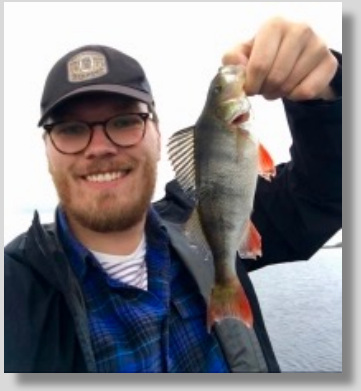
- Understand the power and challenges associated with using low-coverage whole genome sequencing data for population genomic analysis
- Become familiar with all steps involved from sample to inference
- Develop an intuition for the statistical framework implemented in ANGSD and associated programs
- Gain experience with building a bioinformatic pipeline to process low-coverage sequencing data to perform different types of population genomic analyses

Who we are

Nina Overgaard Therkildsen
Cornell University

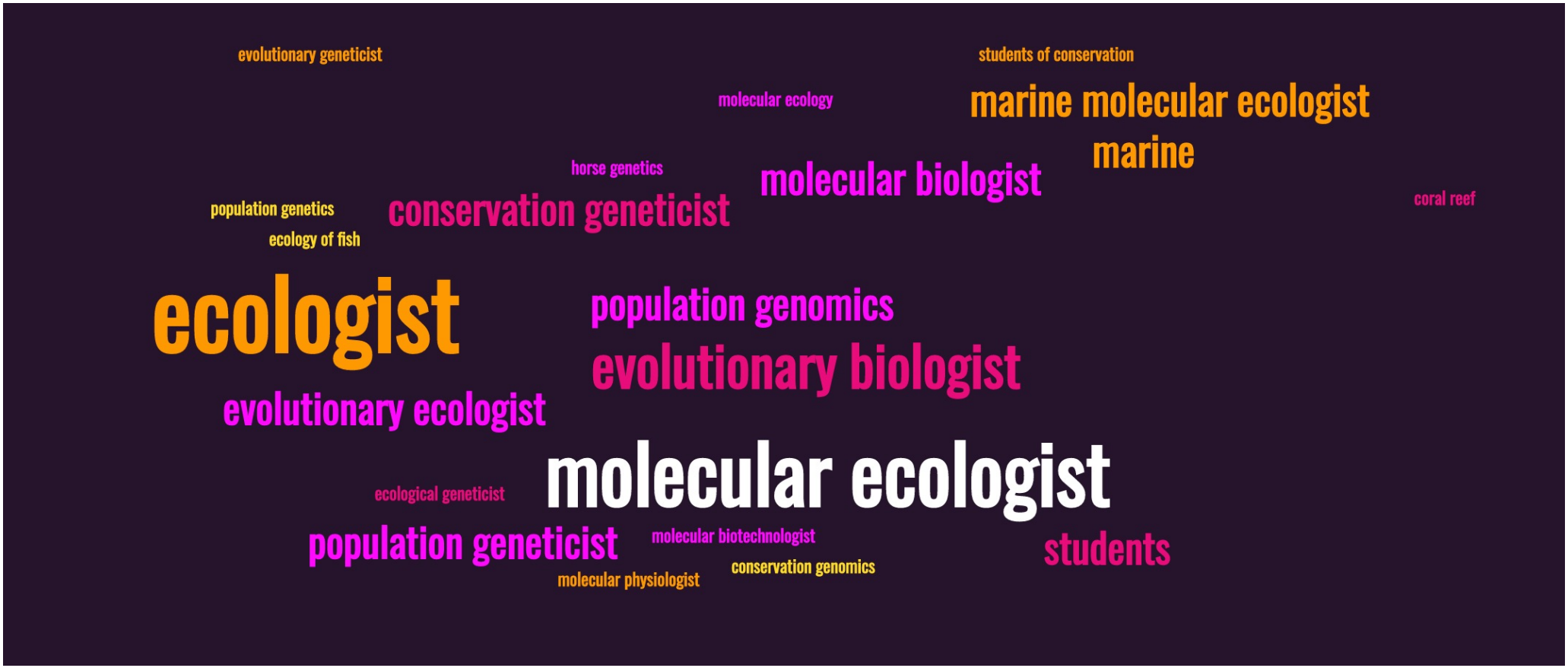
Matteo Fumagalli
Imperial College

Nicolas Lou
Cornell University

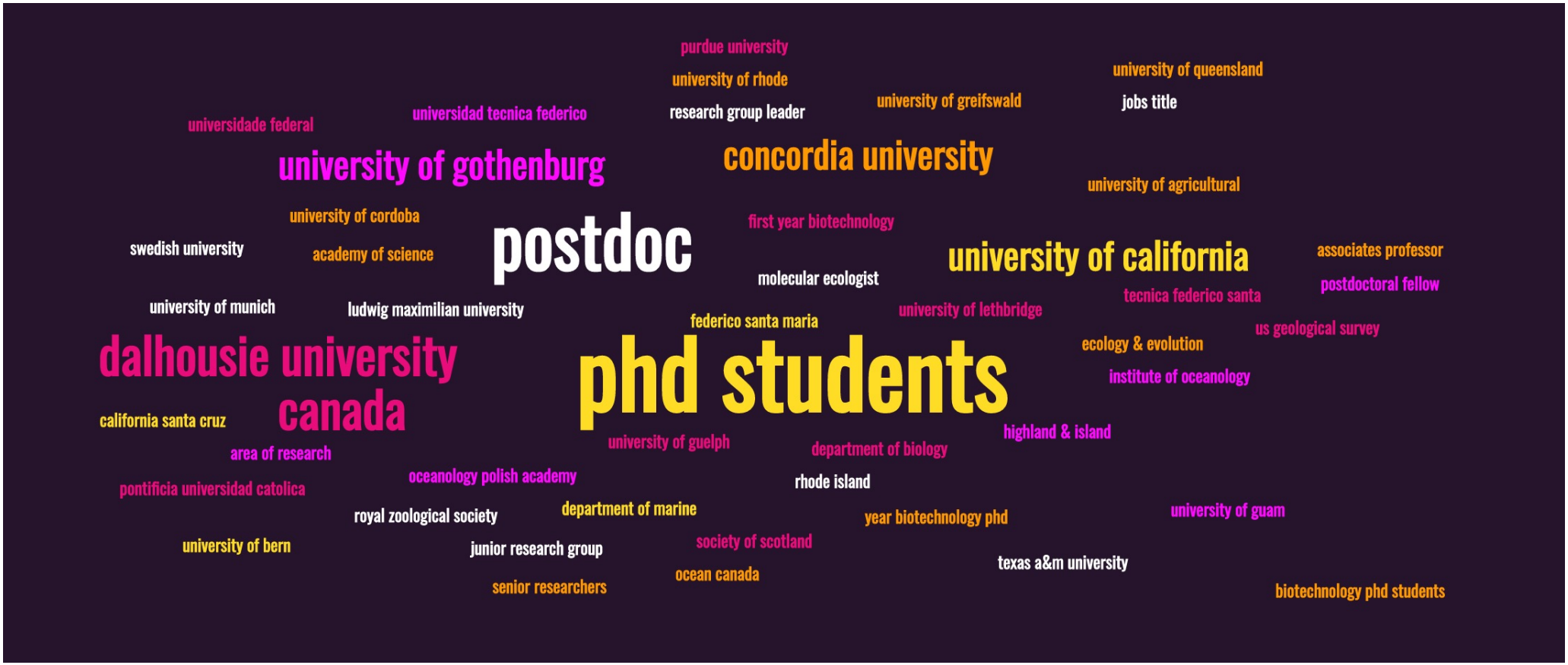


Arne Jacobs
University of Glasgow

Who you are


evolutionary geneticist
students of conservation
molecular ecology
marine molecular ecologist
marine
horse genetics
molecular biologist
coral reef
population genetics
conservation geneticist
ecology of fish
ecologist
population genomics
evolutionary biologist
evolutionary ecologist
molecular ecologist
ecological geneticist
population geneticist
molecular biotechnologist
students
conservation genomics
molecular physiologist

Who you are


purdue university
university of rhode
university of queensland
jobs title
research group leader
university of greifswald
universidad tecnica federico
universidade federal
university of gothenburg
concordia university
university of agricultural
university of cordoba
first year biotechnology
swedish university
academy of science
postdoc
university of california
associates professor
molecular ecologist
postdoctoral fellow
tecnica federico santa
university of munich
ludwig maximilian university
university of lethbridge
federico santa maria
us geological survey
ecology & evolution
dalhousie university
canada
phd students
institute of oceanology
california santa cruz
highland & island
university of guelph
department of biology
area of research
oceanology polish academy
rhode island
pontificia universidad catolica
royal zoological society
department of marine
year biotechnology phd
university of guam
university of bern
junior research group
society of scotland
texas a&m university
ocean canada
senior researchers
biotechnology phd students

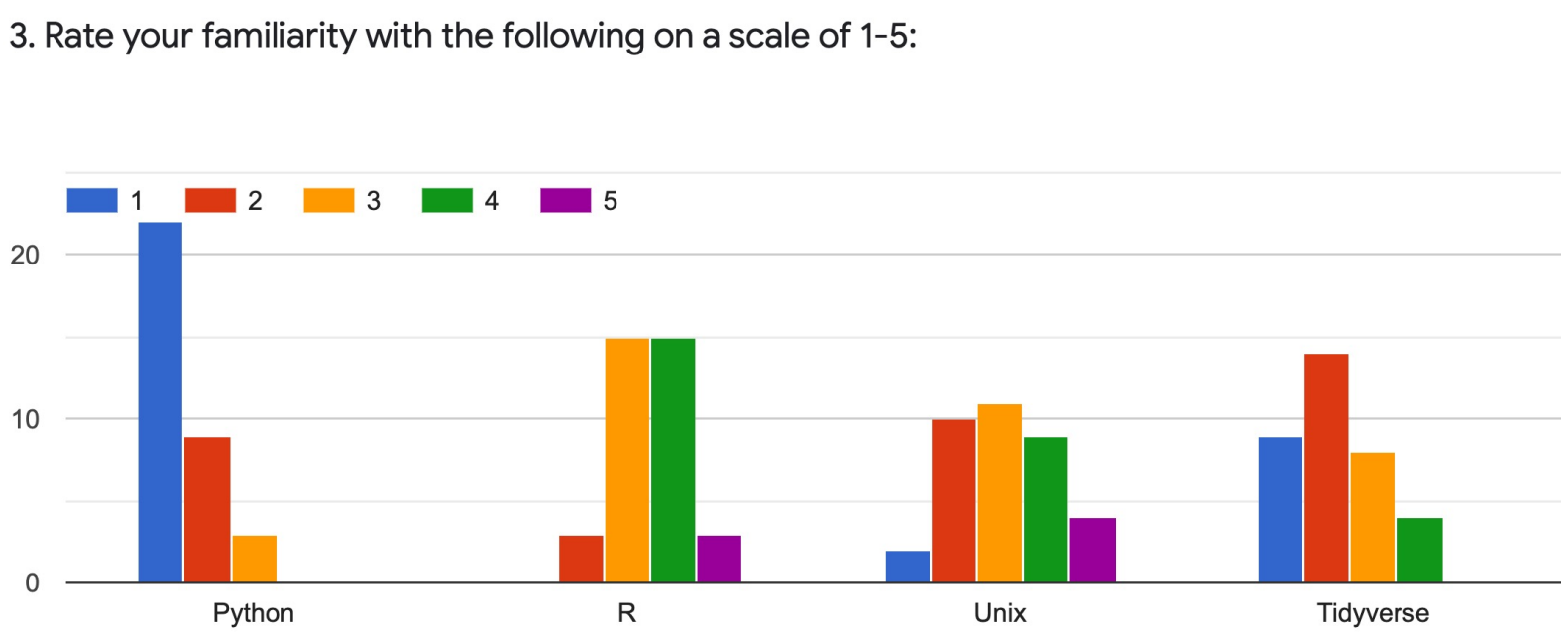
Where are you?

- Please indicate your location on this map https://www.google.com/maps/d/edit?hl=en&mid=1MKCUnLFw7h2IqcVjWS-0m02ZMbOggkP4&ll=15.382761523639108%2C0&z=3

Round of introductions

- Please share
  - Your name and current affiliations
  - What kind of organism(s) you work on
  - Why you’re interested in this course

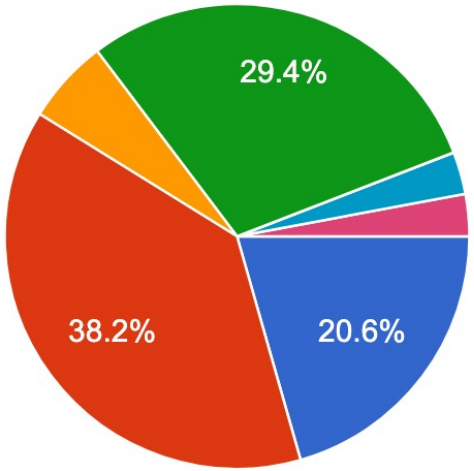
From the pre-course survey



From the pre-course survey

Which of the following data types do you work with

34 responses



- Medium-high coverage whole genome sequence data (>5x)
- Low-coverage whole genome sequence data (<5x)
- Target capture sequence data
- Reduced-representation sequence data (RAD-seq, GBS, etc)
- Microsatellite data
- RNAseq
- Other


ALL ARE
WELCOME
HERE
Graphic: https://allarewelcomehere.us/

Code of Conduct

- We are dedicated to providing a **welcoming and supportive environment for everyone**, regardless of background, identity and prior experience. Everyone in this course will be coming from a different place with different experiences and expectations. We will not tolerate any form of language or behavior used to exclude, intimidate, or cause discomfort. This applies to all course participants (instructor, students, guests). In order to foster a positive and professional learning environment, we encourage the following kinds of behaviors

Behaviors we encourage

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Show courtesy and respect towards others
- Help each other - you may well learn something or reinforce your own skills in the process

Approximate daily schedule

Berlin time	US eastern	Activity
14 – 15.15	8 - 9.15	Session 1
		BREAK
15.30 – 16.45	9.30 -10.45	Session 2
		BREAK
17.15 – 18.30	11.15 - 12.30	Session 3
		BREAK
18.45 - 20	12.45 - 2	Session 4

Course schedule

- Day 1
  - Welcome!
  - Introduction to low-coverage whole genome sequencing
  - From sample to fastq
  - From fastq to bam

- Day 2
  - Genotype likelihoods
  - SNP calling
  - Allele calling

Course schedule

- Day 3
  - Linkage disequilibrium
  - Population structure (PCA and admixture analysis)

- Day 4
  - The site frequency spectrum (1d and 2d)
  - Fst and diversity statistics
  - Overview of other applications and future perspective

Daily practicals

- Will be available in a GitHub repo that you can keep accessing after the course
- Will work in breakout rooms
- Indicate your breakout room preference each day
  - Quiet room (everyone works independently)
  - Semi-quiet room (people mostly work independently, but can ask each other questions)
  - Collaborative room (you work through the exercises together)

- Ask questions!

What we will not have time to cover

- Methylation/epigenetics (bisulphite sequencing) and gene expression
  - This course is about DNA sequence variation only

- Imputation
- Methods specific to ancient DNA
- GWAS
- Structural variant detection

Course goals

- Understand the power and challenges associated with using low-coverage whole genome sequencing data for population genomic analysis
- Become familiar with all steps involved from sample to inference
- Develop an intuition for the statistical framework implemented in ANGSD and associated programs
- Gain experience with building a bioinformatic pipeline to process low-coverage sequencing data to perform different types of population genomic analyses