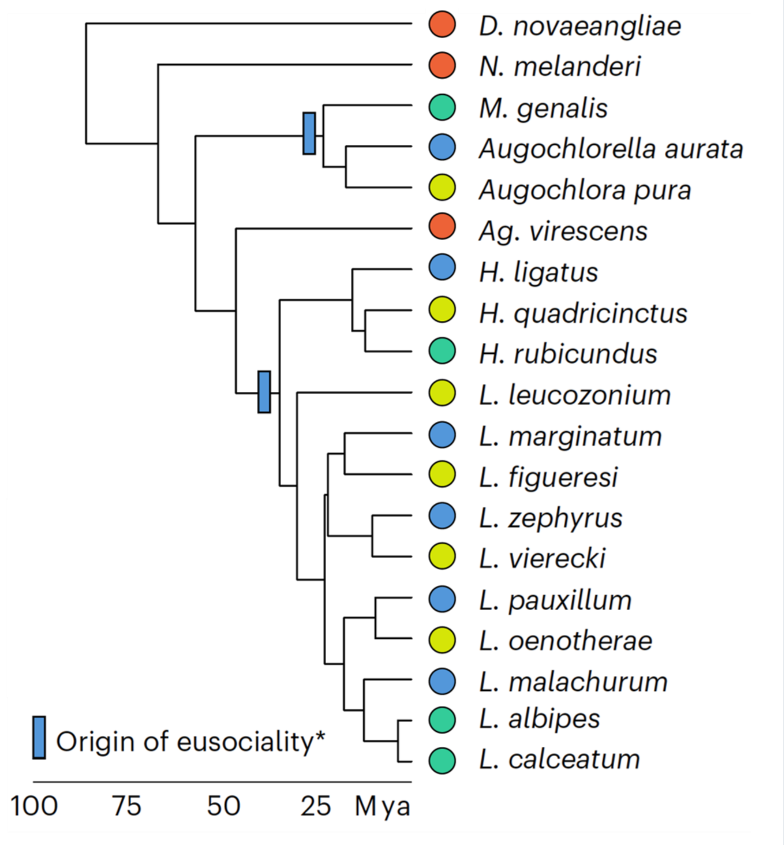
Pipeline Example

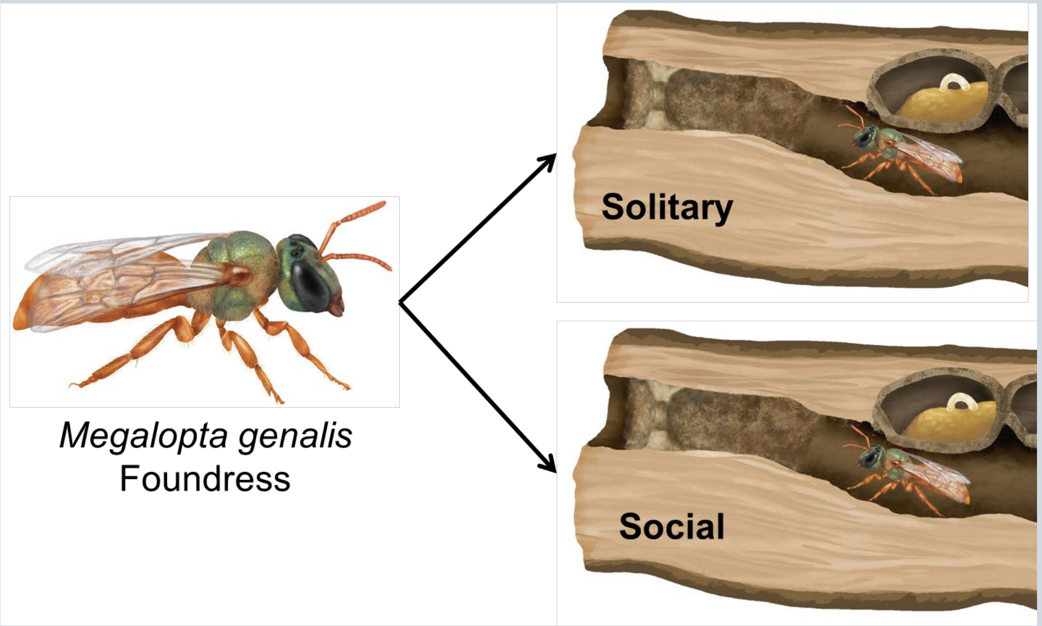
ANDREW WEBB

Sociality and Population structure of *Megalopta genalis*

Sociality of *Megalopta genalis*

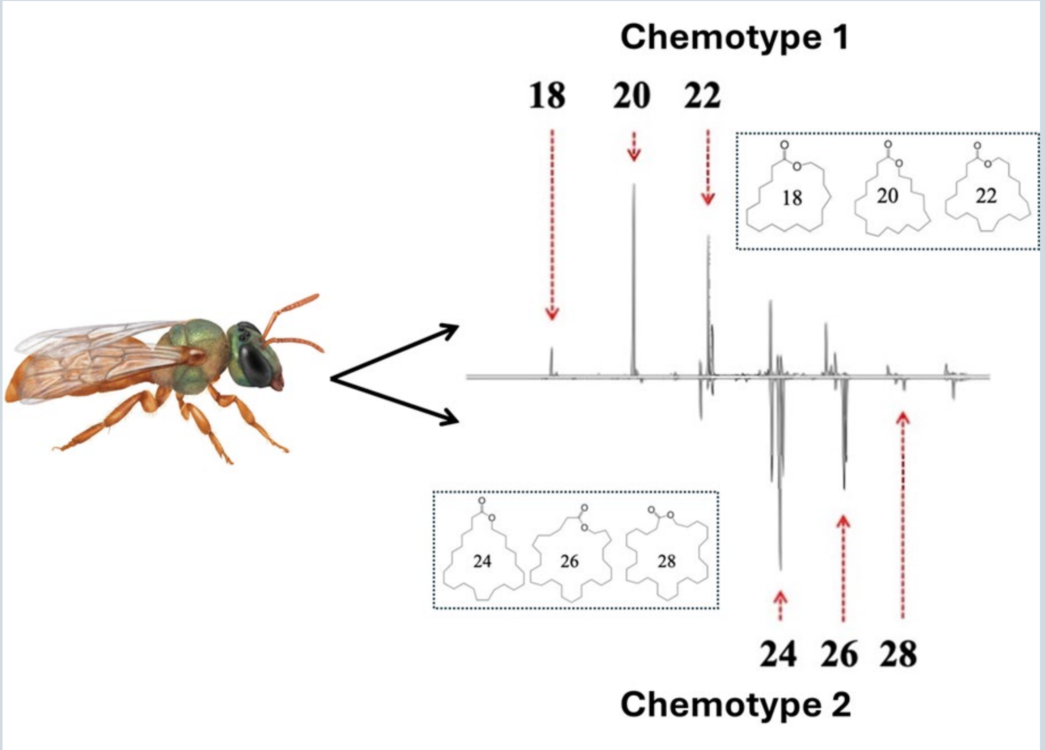


Sweat bee phylogeny

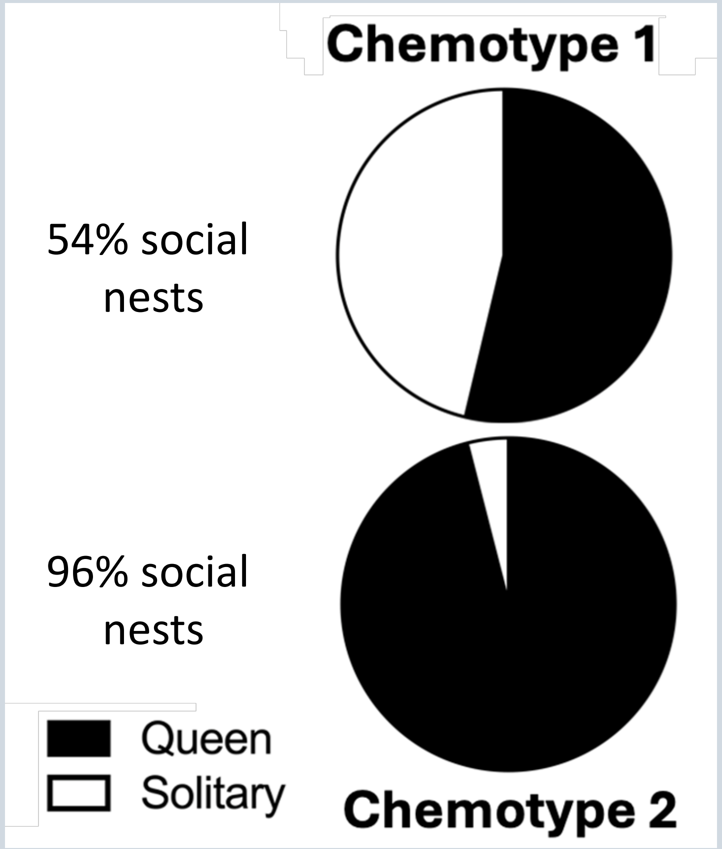


Nest Establishment

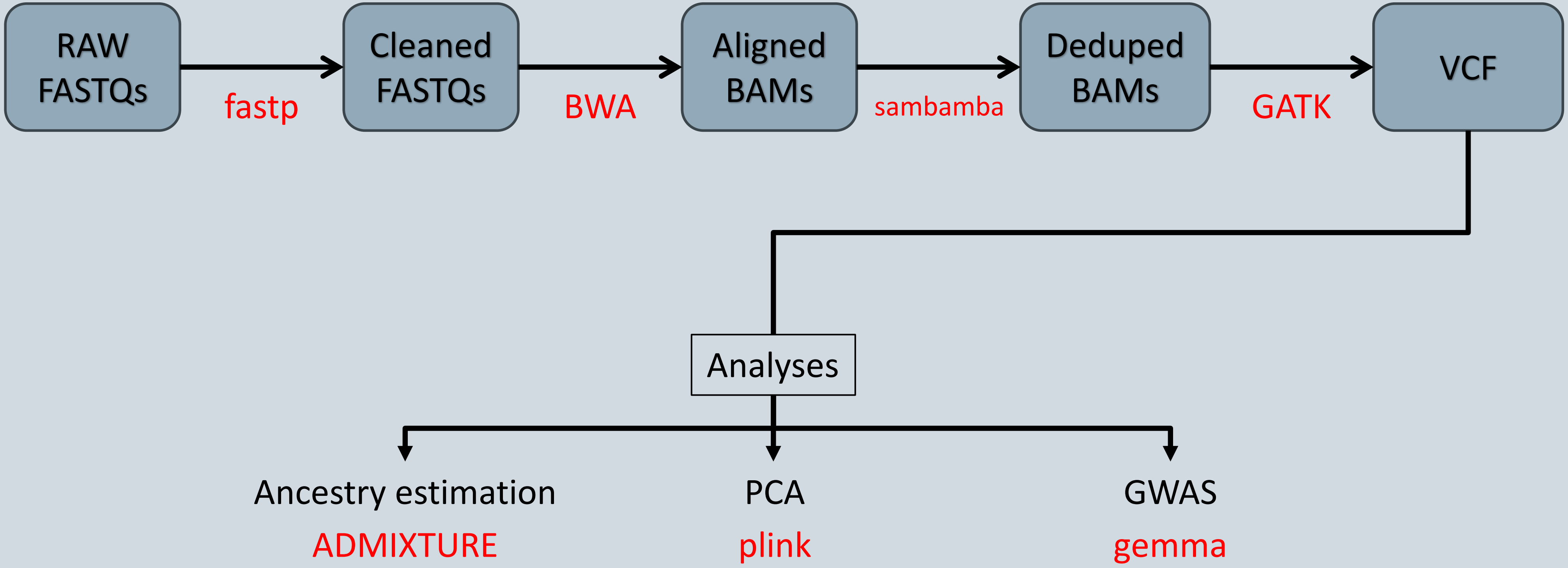
Sociality of *Megalopta genalis*



GC-MS Chromatograms of macrocyclic lactones



Our pipeline



RAW FASTQs

fastp

Cleaned FASTQs

BWA

Aligned BAMs

sambamba

Deduped BAMs

GATK

VCF

Analyses

Ancestry estimation

ADMIXTURE

PCA

plink

GWAS

gemma

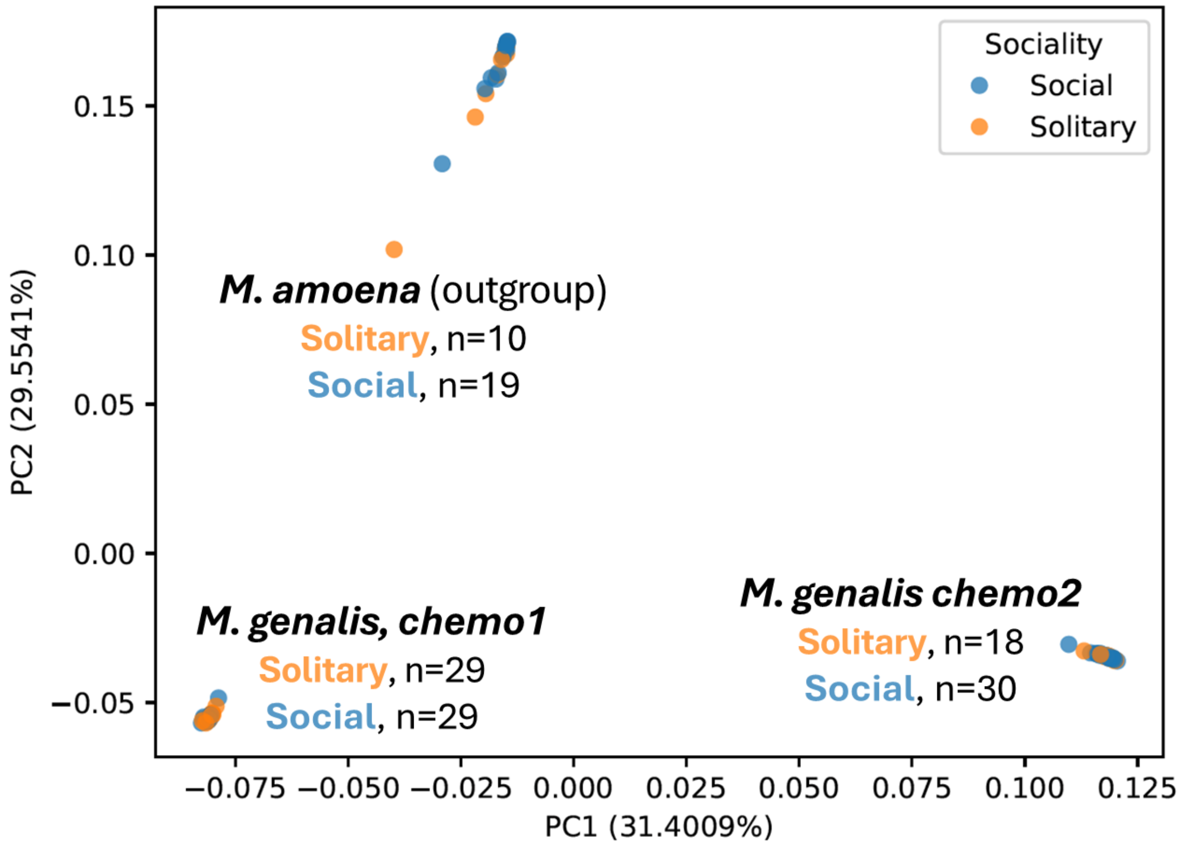
Ancestry estimation

M. genalis,
chemotype 1

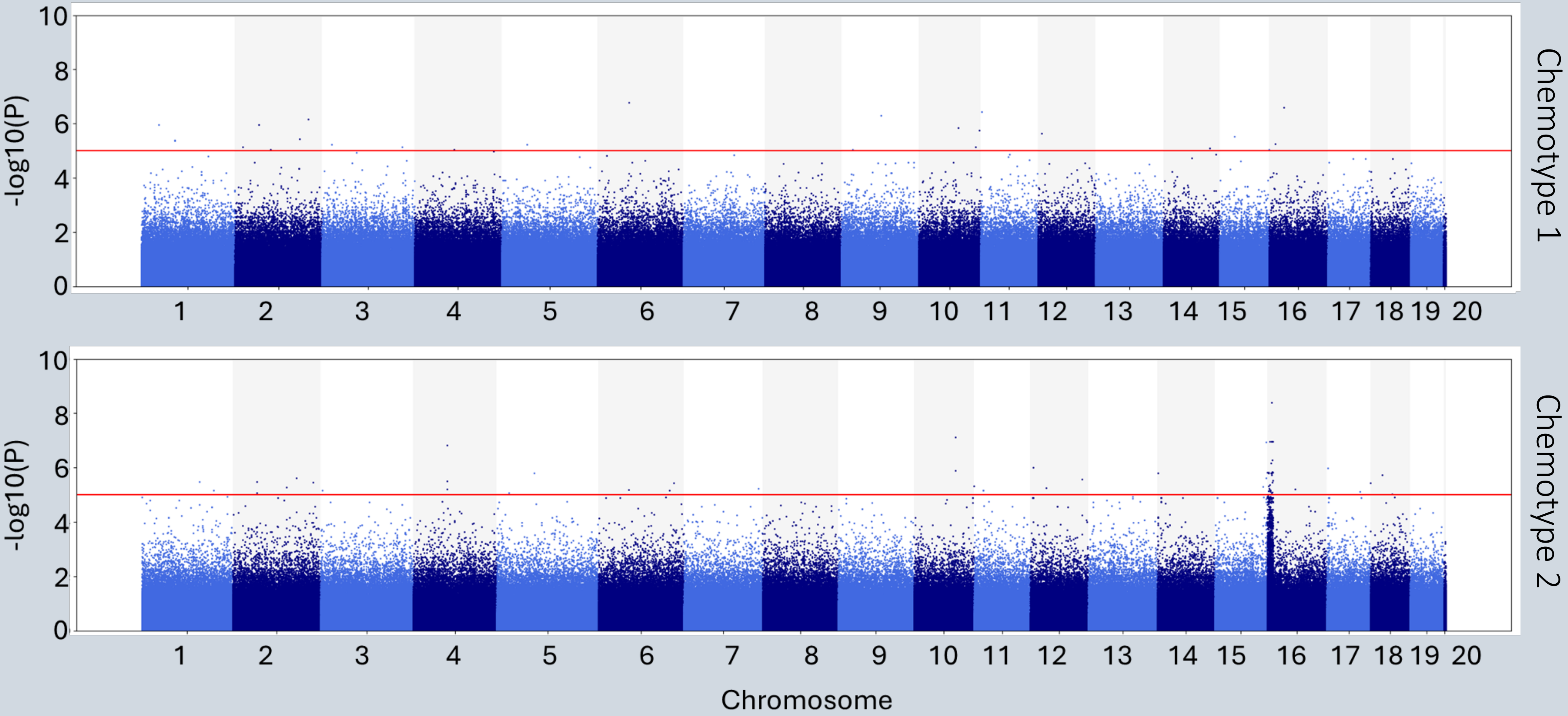
M. genalis,
chemotype 2

M. amoena
(outgroup)

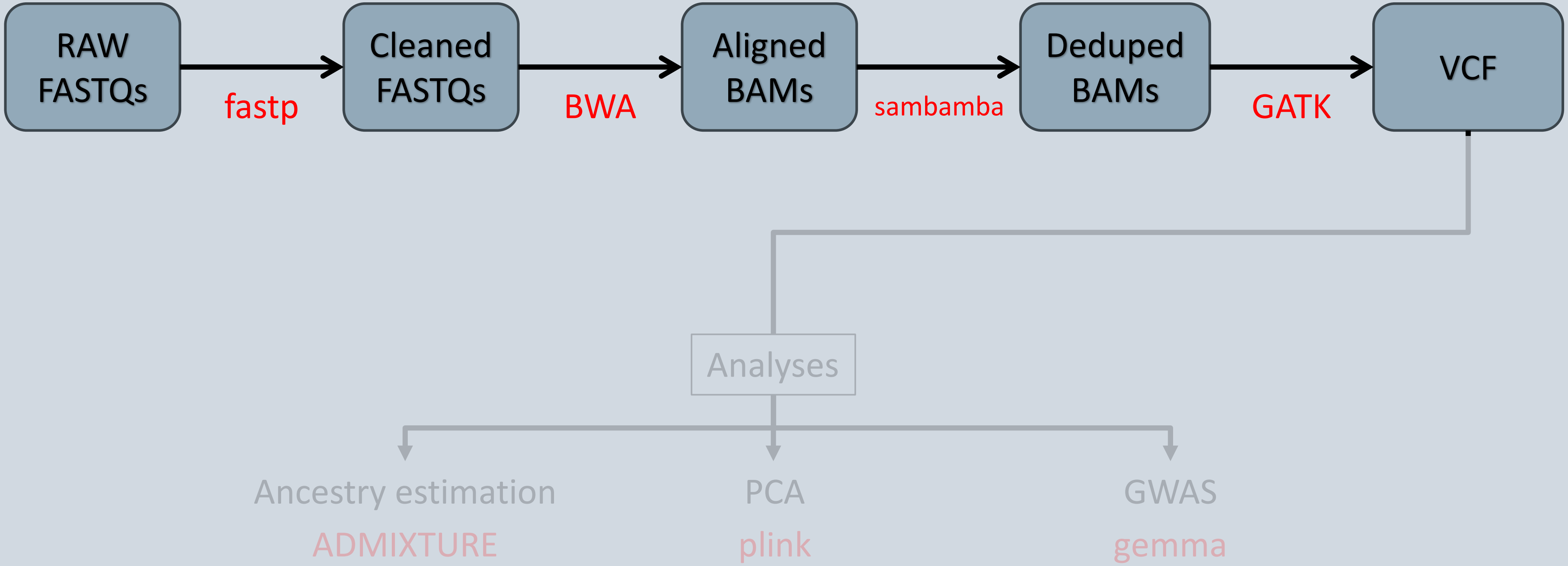
Principal Component Analysis (PCA)



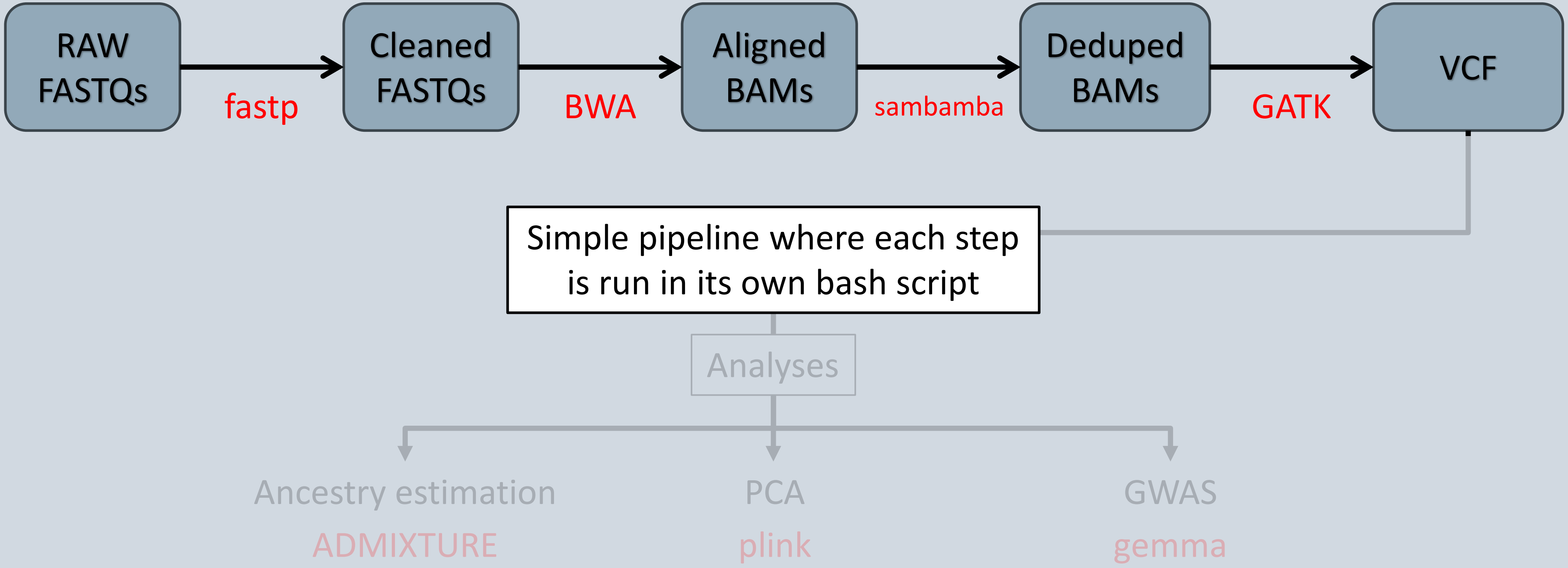
Genome-Wide Association Study (GWAS)



Cluster demo



Cluster demo



RAW FASTQs

fastp

Cleaned FASTQs

BWA

Aligned BAMs

sambamba

Deduped BAMs

GATK

VCF

Simple pipeline where each step is run in its own bash script

Analyses

Ancestry estimation

ADMIXTURE

PCA

plink

GWAS

gemma

SLURM

```
#!/bin/bash
#SBATCH --job-name=myserial-job                 # Name of the job
#SBATCH --output=myserial-job.%j.out            # Name of the output file (with jobID %j)
#SBATCH --error=myserial-job.%j.err             # Name of the error file (with jobID %j)
#SBATCH --nodes=1                               # Node count
#SBATCH --ntasks=1                              # Number of tasks across all nodes
#SBATCH --cpus-per-task=1                       # Cores per task (>1 if multi-threaded tasks)
#SBATCH --mem-per-cpu=4G                        # Memory per core (4G is default)
#SBATCH --time=00:01:00                         # Run time limit (HH:MM:SS)
#SBATCH --mail-type=all                         # Email on job start, end, and fault
#SBATCH --mail-user=<YourNetID>@princeton.edu   # Email address

echo 'Hello world!'
```