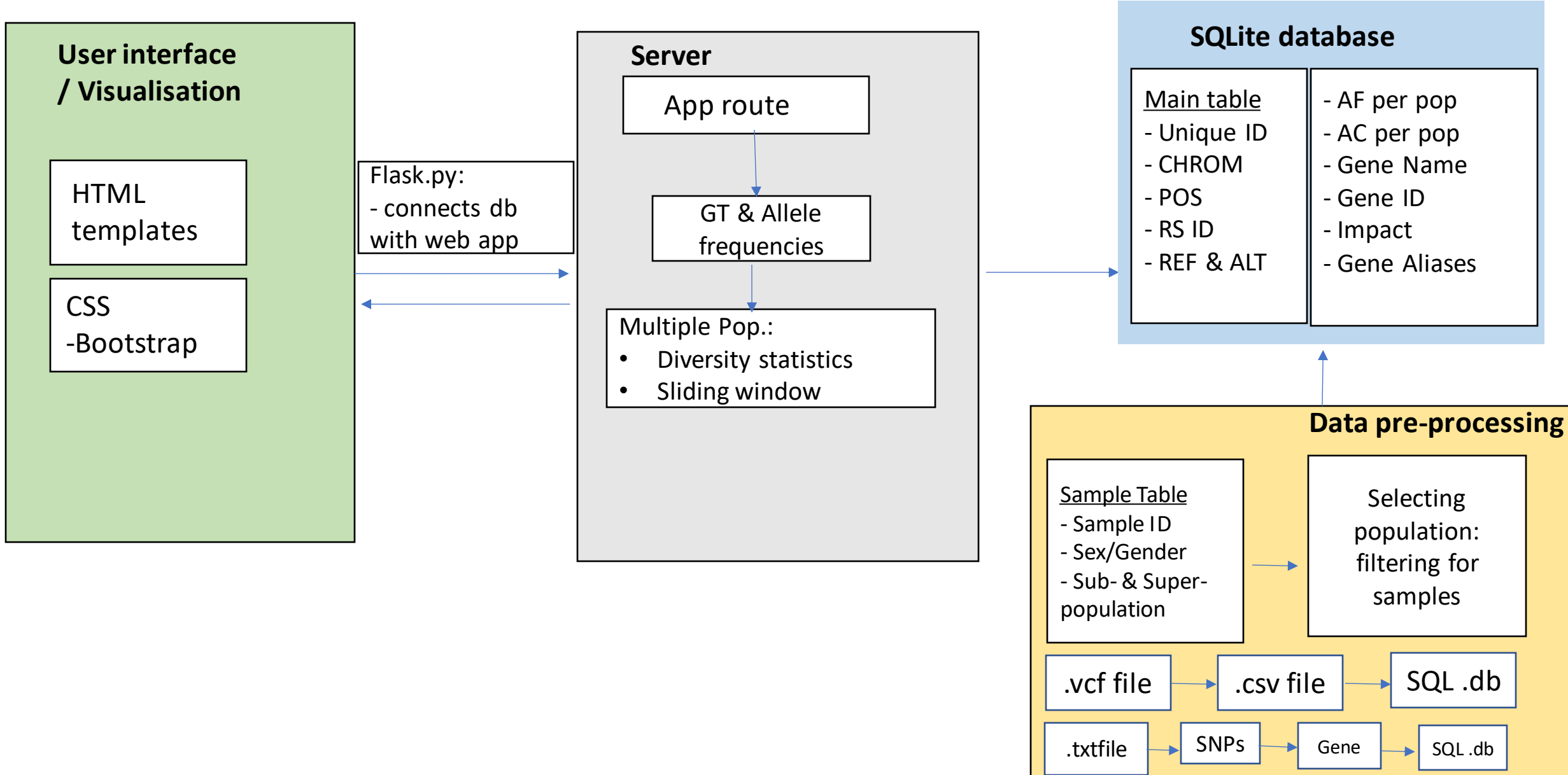# BIO727P GROUP SOFTWARE PROJECT

Team Celine
Celine, Amanah & Gracia
17.02.2021

SOFTWARE ARCHITECTURE

COMPLETED TASKS (14/02 - 18/02/2022)

- ✓ Calculate summary statistics
  - ✓ Shannon Diversity, Tajima's D – *Celine*
  - ✓ Heterozygosity Index – *Gracia*
  - ✓ FST – *Amanah*
  - ✓ Sliding window & plots – *Amanah & Gracia*

- ✓ Gene Aliases added and combine data to one table – *Gracia*

- ✓ Documentation – *Amanah*

- ✓ Flask: added routes - *Celine*
  - ✓ html pages (search results and statistics/population selection)
  - ✓ Statistics added: shannon

- ✓ Add more populations – *group decision*
  - ✓ Extract population data and calculate gt & allele freqs – *Celine & Gracia*
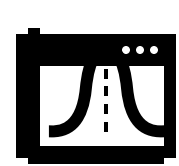
GANTT CHART UPDATE
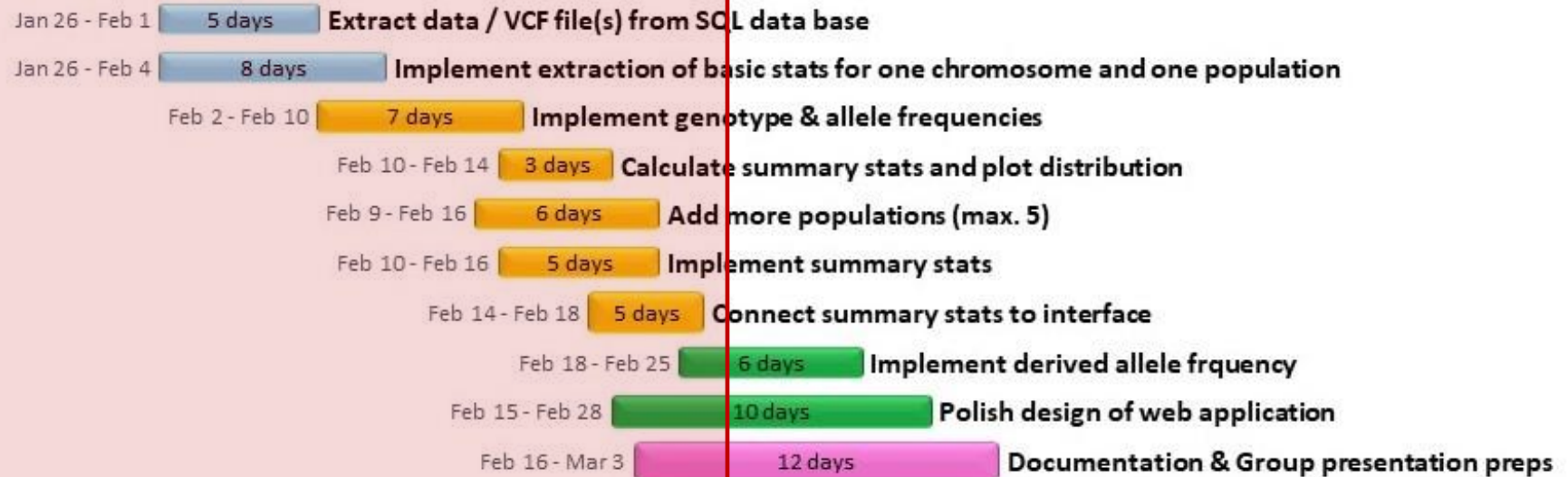
NEXT WEEK'S MILESTONES (21/02 – 25/02/2022)

➢ **Connect summary stats calculation and sliding window for a query to web app –** *Celine*

➢ **Download text file –** *Celine*

➢ **Documentation –** *Amanah & Gracia*

➢ Polishing of web app (incl. Design → HTML templates) – *Celine*

➢ Derived chimpanzee allele frequency **–** *Amanah*

==> *making sure what we are producing makes sense!*

QUESTIONS

1. Which types of genes should we keep? NcRNA (example: ENSG00000232073 ***), pseudogenes, or only DNA genes?

2. Derived allele frequency → theta threshold?

3. Is it a problem that no rsIDs for all SNPs?

4. Documentation structure?

5. Any questions for us?