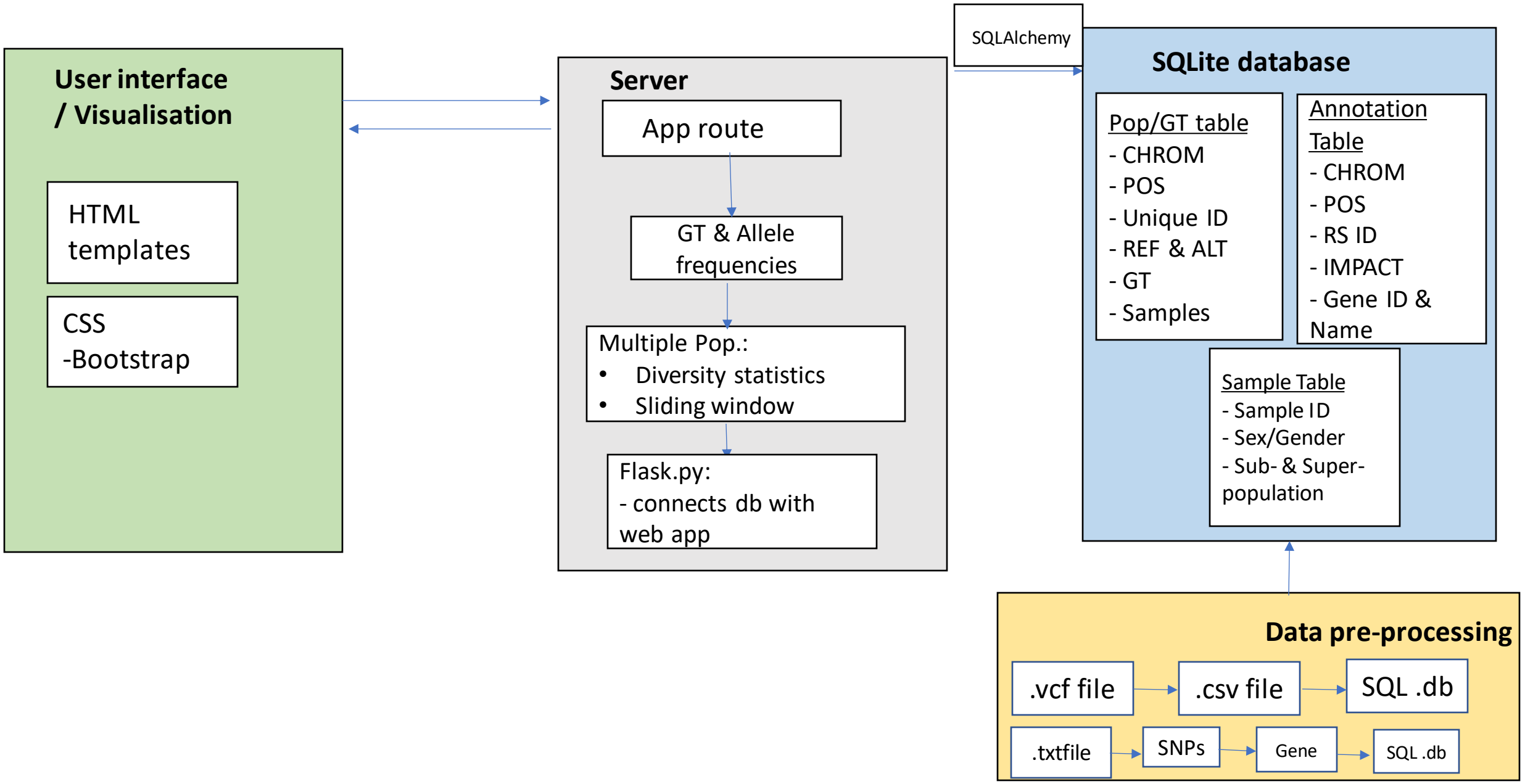
BIO727P GROUP SOFTWARE PROJECT

Team Celine
Celine, Amanah, Gracia & Pavan
03.02.2021

SOFTWARE ARCHITECTURE



Celine's Workspace – pls ignore

- need to refine indels and connect GBR table after frequencies are calculated

Genotype table – phased.vcf

- CHROM
- POS
- Unique ID {{PK}}
  22:10519265:CA:C
- REF & ALT (multiple ALTs)
- GT/sample

Gene names table

- CHROM
- POS
- RS ID
- Unique ID {{FK}} need this
- REF & ALT <delete later>
- IMPACT
- GENE name
- GENEID

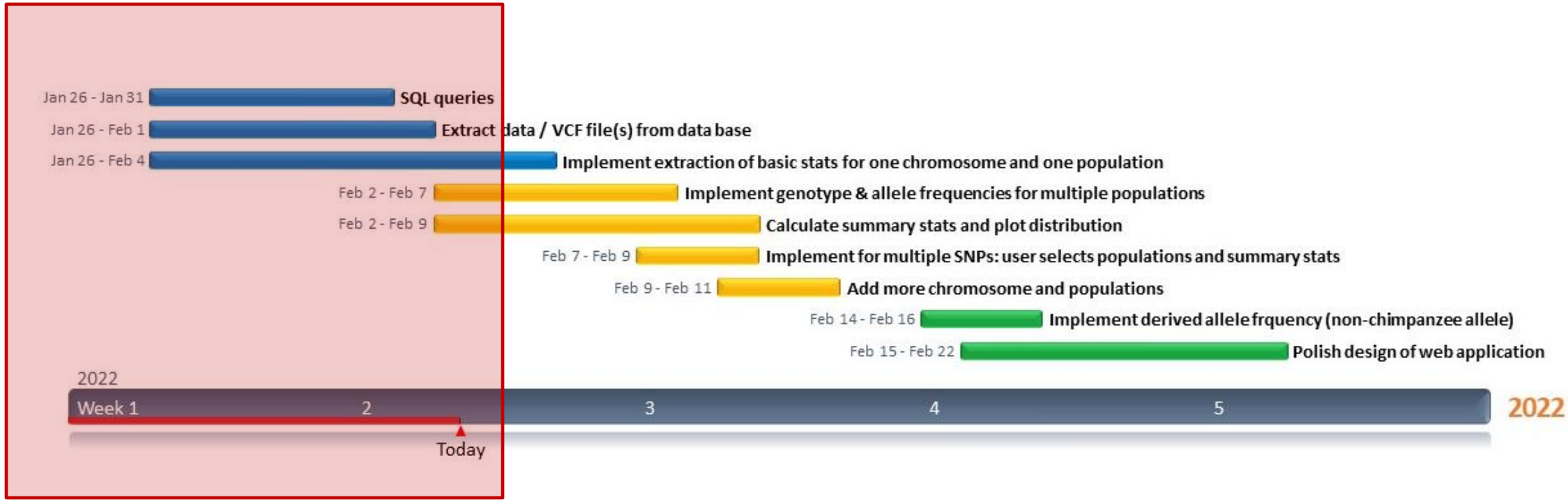
Population-sample table

- SAMPLE ID {{PK}}
- Sex
- Biosample ID
- Population code & name
- Superpopulation code % name
- DATA

COMPLETED TASKS

- ✓Extraction of VCF files
- ✓Building SQL database
- ✓Filter for one population using `bcftools` ➔ focused on GBR for now
- ✓Researched on summary statistics

GANTT CHART UPDATE



NEXT WEEK'S MILESTONES

- Properly research summary stats / inputs → Amanah & Pavan
- Implement genotype frequencies → Amanah
- Calculate allele frequencies → Pavan
- Connect flask and SQL database → Celine
- Setting up query function to search the DB → Celine
- Build very basic web app template → Celine/Gracia
- Derive summary statistics from successful GT & allele frequencies (incl. Sliding window) → Gracia

QUESTIONS

- Just SNPs? Or also indels
- Do indels count as SNPs
- Who picks the 5 populations? User or us as developers?
- Do you have any questions for us?