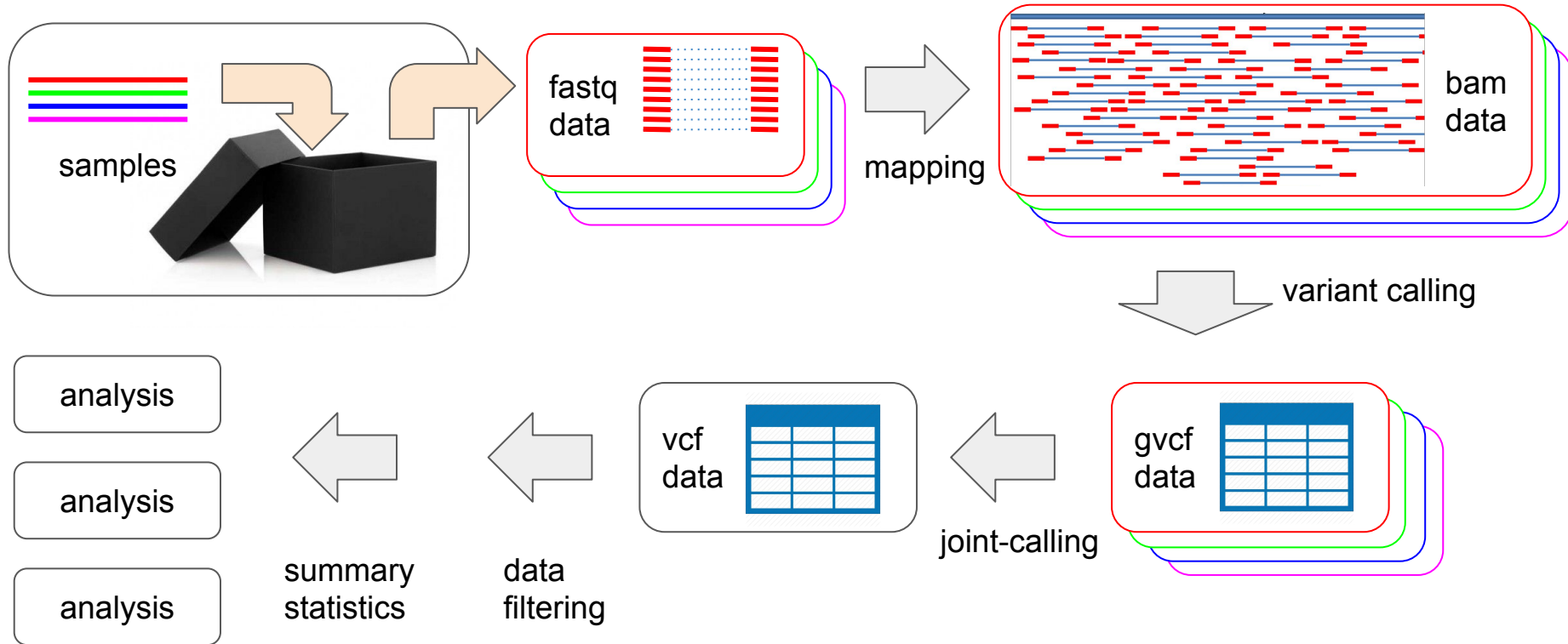


21 March, 2016:
Handling sam and vcf data, quality control

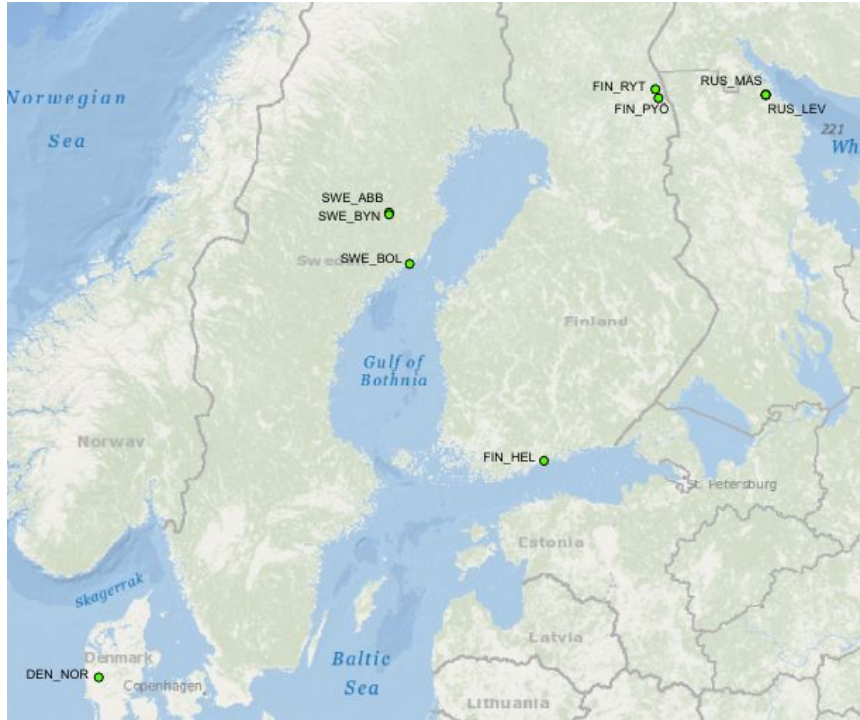
Overview of resequencing data analysis



Analysis pipeline

1. preparations: indexing, dictionary for reference
 2. alignment (or mapping)
 3. sorting and duplicate removal
 4. re-alignment around indels
 5. variant calling: genome vcf
- for each sample
6. joint-calling: vcf
 7. vcf data filtering
- combined data

Our data



- ten samples from each population
- 10X coverage

- sample names anonymized
 - popX refers to ten populations
 - sample-Y come from those

From which population is your sample?