User manual, data integration and interactions

MSc Applied Bioinformatics,

Marie Schmit

Database population

To execute the script for database population, you need to have java installed. In an IDE, open and run DataBasePopulating.java (folder Database, dataBaseParsingPopulation, src). You can also open it in a terminal, writing: javac DataBasePopulating.java.

A browser window will then appear, allowing you to choose the vcf file that you want to save in the database. Select the file, click the button "open" and wait. It can take some time, depending on the size of the vcf file.

Run REST API

Open a terminal in folder REST_API, where javascript scripts are saved. Create a new npm project: npm init.

Install the following dependencies (see also package.json):

- Express: npm install express –save
- Body parser: npm install body-parser –save
- Sqlite3: npm install sqlite3 –save
- R script: npm install r-script (or npm install r-script-with-bug-fixes)

Once the dependencies are installed, run the script, writing: node server.js. This should send the following message: Application deployed on port 3000. You can now open your browser to type urls.

Display density plot with R, plumber

To display density plot, you need to open the script variant_density.R (folder REST API) in an IDE. Make sure that you install the following packages:

- Plumber
- Ggplot2
- Jsonlite
- Httr

Deploy the R API by typing: plumber::plumb("variant_density.R")$run(port = 3001) in your console. Make sure that your server is activated (node server.js). The end-point to display density plot is usable (see section "end points of API").

End points of API

Replace every parameter :param by its value. :param? are optional.

```
http://localhost:3000/api/genomes
```

To see the list of genomes id and name contained in the database.

eg. http://localhost:3000/api/genomes

```
http://localhost:3000/api/variants/region/:genome/:chromosome/:startPosition/:endPosition/:type?/:subtype?
```

To display a list of variants located in a specific region of a specific chromosome in a specific dataset. StartPosition and endPosition gives the coordinates of this specific region. Type (SNP, InDel) and subtype (Insertion, Deletion) are optional.

Eg. http://localhost:3000/api/variants/region/9968/4/200000/420000/InDel/Insertion

```
http://localhost:3000/api/variants/zygosity/:genome/:chromosome/:zygosity/:startPosition/:endPosition/:type?/:subtype?
```

To display a list of variants having a specified zygosity, located in a specific region of a specific chromosome in a specific dataset. StartPosition and endPosition gives the coordinates of this specific region. Type (SNP, InDel) and subtype (Insertion, Deletion) are optional. Zygosity can take values: Heterozygote, Homozygote.

Eg.
http://localhost:3000/api/variants/zygosity/9968/4/Homozygote/200000/420000/InDel/Insertion

```
http://localhost:3000/api/variants/quality/:genome/:chromosome/:quality/:type?/:subtype?
```

To display the variant a having minimal quality per genome. This minimal value is set by parameter: quality.

Eg. http://localhost:3000/api/variants/quality/9968/4/40/InDel/Insertion

```
http://localhost:3000/api/variants/depth/:genome/:chromosome/:depth/:type?/:subtype?
```

To display the variant having a minimal depth per genome. This minimal value is set by parameter: depth.

Eg. http://localhost:3000/api/variants/depth/7208/1/500/SNP

```
http://localhost:3000/api/variants/meanDepthQuality/:genome/:chromosome/:type?/:subtype?
```

To calculate the mean quality and the depth per genome, per chromosome.

Eg. http://localhost:3000/api/variants/meanDepthQuality/8233/2/

```
http://localhost:3000/apiDens/density/:genome/:chromosome/:windowSize/:type?/:subtype?
```

Display array of densities for a given genome, chromosome and window. The parameter windowSize is the size of the windows that split the position data.

Eg. http://localhost:3000/apiDens/density/9968/5/100000/InDel/Insertion

```
http://127.0.0.1:3001/apiDens/density/plot/genome/chromosome/windowSize/type/subtype
```

After activation of plumber REST API in R (see section "Display density plot with R, plumber).

Where genome, chromosome, windowSize, type and subtype are replaced with the genome, chromosome of interest, the size of the windows on which density is calculated, the type (SNP, InDel) and subtype (Insertion, Deletion) of the variant (those two parameters are optional).

Eg: http://127.0.0.1:3001/apiDens/density/plot/8233/1/100000/