

# Kaka: Towards a Distributed Genotype Database

Helge Dzierzon and Samantha Baldwin

October 28 2015

# Genotype Data in PFR

- ▶ Many species: Kiwifruit, Apple, Pear, Potato, Grape, ...
- ▶ Diverse genotype data: GBS, QTL. ...
- ▶ The data is semi structured



Plant & Food  
**RESEARCH**  
RANGAHAU AHUMĀRA KAI



# PFR Wetware Distribution

- ▶ 14 PFR sites in NZ
- ▶ Data exchange across sites challenging
- ▶ Analysts distributed as well



# Current way of “Sharing”

- ▶ The K-Drive
- ▶ Excel Sheets



philipmartin.info

# Database Initiatives in PFR

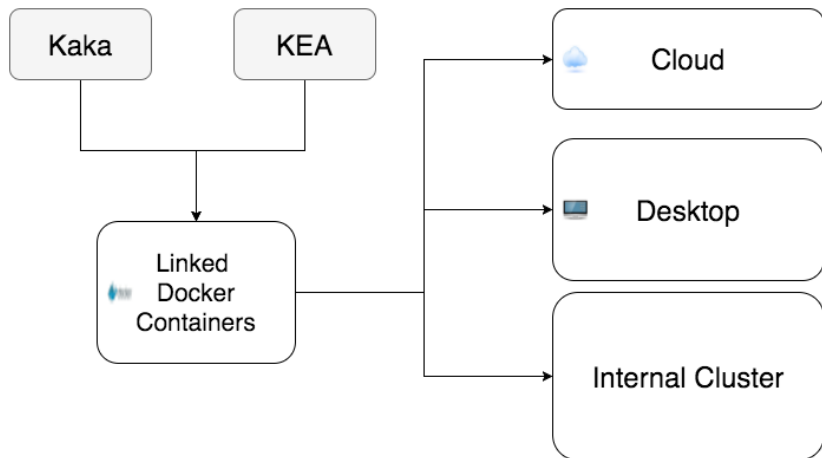
- ▶ EBrida (commercial breeder's DB)
- ▶ Kea (inhouse sample tracking)
- ▶ Ensembl (Genome Browsing)
- ▶ Kaka (Genotypes concentrates on samples)
- ▶ Kakapo (Locations)
- ▶ ...

# Kaka

- ▶ Supplement to Kea
- ▶ Python Django
- ▶ PGSQL and JSON (NoSQL)
- ▶ Continuous integration
- ▶ Containerisation
- ▶ Needs location (Chr, location on Chromosome)



# Distributed DB



## Querying via API

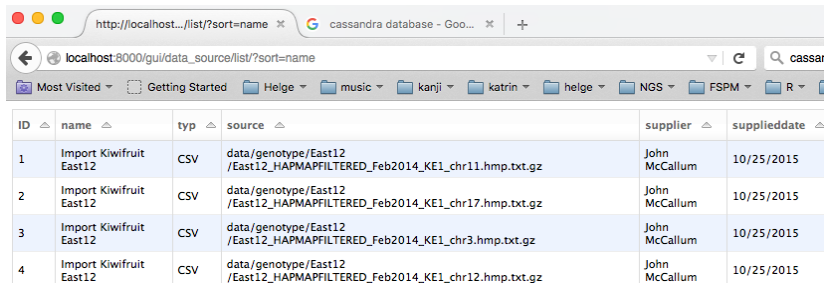
```
http://localhost/report/genotype/xlsx/?experiment=12East  
http://localhost/report/genotype/csv/?experiment=12East  
http://localhost/gui/genotype/list/
```



## Screenshots

|    | A                       | B        | C                | D     | E        | F        | G       | H         | I         | J         | K         | L         | M     |
|----|-------------------------|----------|------------------|-------|----------|----------|---------|-----------|-----------|-----------|-----------|-----------|-------|
|    | datasource_name         | ontology | study_name       | chrom | pos      | rs#      | alleles | T06.12E-0 | T06.12E-0 | T06.12E-0 | T06.12E-0 | T06.12E-0 | T06.1 |
| 1  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 30    | 30432254 | S30_3043 | A/G     | A         | A         | A         | A         | A         | A     |
| 2  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 80       | S11_80   | A/T     | A         | A         | N         | A         | A         | A     |
| 3  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 6548302  | S11_6548 | T/C     | N         | N         | T         | N         | N         | T     |
| 4  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 84       | S11_84   | T/C     | T         | T         | N         | T         | T         | T     |
| 5  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 88       | S11_88   | A/T     | A         | A         | N         | A         | A         | A     |
| 6  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 98       | S11_98   | A/G     | A         | A         | N         | A         | A         | A     |
| 7  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 112      | S11_112  | G/A     | G         | A         | N         | G         | G         | G     |
| 8  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 115      | S11_115  | C/T     | T         | T         | N         | Y         | C         | Y     |
| 9  | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 127      | S11_127  | T/C     | T         | T         | N         | T         | T         | T     |
| 10 | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 172      | S11_172  | T/G     | T         | N         | N         | T         | N         | T     |
| 11 | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 173      | S11_173  | A/G     | A         | N         | N         | A         | N         | A     |
| 12 | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 11    | 175      | S11_175  | A/C     | C         | N         | N         | C         | N         | C     |
| 13 | Import Kiwifruit East12 | Genotype | Kiwifruit 12East | 17    | 1422736  | S17_1422 | G/A     | G         | G         | G         | G         | G         | G     |

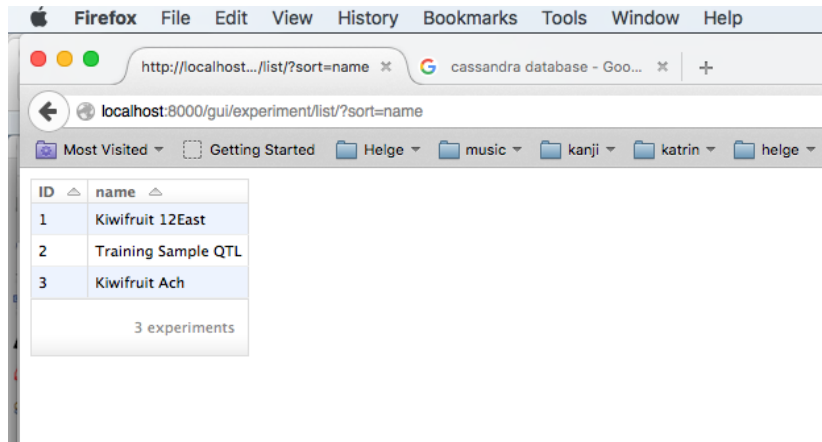
# Screenshots



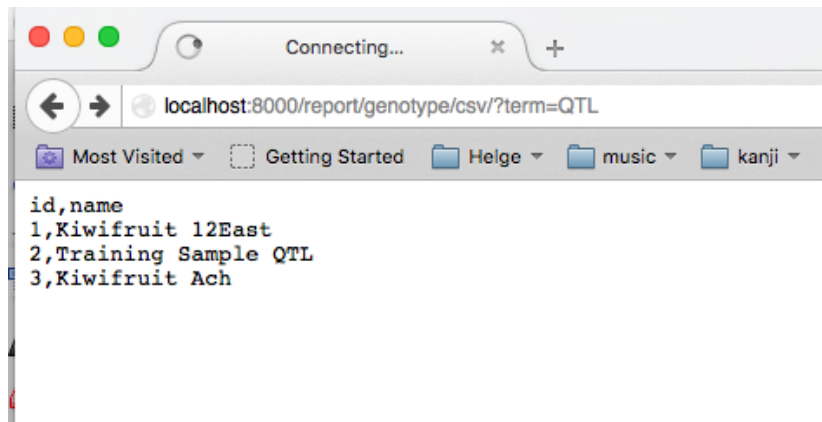
The screenshot shows a web browser window with the address bar displaying `http://localhost:8000/gui/data_source/list/?sort=name`. The browser's address bar also shows `localhost:8000/gui/data_source/list/?sort=name` and a search bar with the text `cassar`. The browser's tab bar shows several tabs, including `Most Visited`, `Getting Started`, `Helge`, `music`, `kanji`, `katrin`, `helge`, `NGS`, `FSPM`, and `R`. The main content area displays a table with the following data:

| ID | name                    | typ | source  | supplier      | supplieddate |
|----|-------------------------|-----|---|---------------|--------------|
| 1  | Import Kiwifruit East12 | CSV | data/genotype/East12/East12_HAPMAPFILTERED_Feb2014_KE1_chr11.hmp.txt.gz | John McCallum | 10/25/2015   |
| 2  | Import Kiwifruit East12 | CSV | data/genotype/East12/East12_HAPMAPFILTERED_Feb2014_KE1_chr17.hmp.txt.gz | John McCallum | 10/25/2015   |
| 3  | Import Kiwifruit East12 | CSV | data/genotype/East12/East12_HAPMAPFILTERED_Feb2014_KE1_chr3.hmp.txt.gz  | John McCallum | 10/25/2015   |
| 4  | Import Kiwifruit East12 | CSV | data/genotype/East12/East12_HAPMAPFILTERED_Feb2014_KE1_chr12.hmp.txt.gz | John McCallum | 10/25/2015   |

# Screenshots



# Screenshots



# Conclusions

- ▶ Genotype and phenotype information in one database
- ▶ It is a great tool already but needs a nerd
- ▶ It needs some input from the research community
- ▶ Kea needs to be attached as a docker container
- ▶ The data input is very manual
- ▶ It needs an interface to Ensembl

Thank you! Questions?

