Systems Genetics Software

Brian S. Yandell, UW-Madison, www.stat.wisc.edu/~yandell/PlantSysGen

January 2017

software setup

If we were going to do a live demo …

- install R and RStudio
    - see datacarpentry instructions
    - or next two slides
- install packages qtl, rrBLUP
- download or view https://github.com/byandell/thailand

install R

R: powerful programming language for data exploration, visualization, and analysis

- main R site: https://cran.r-project.org/
- Thailand mirror: http://mirrors.psu.ac.th/pub/cran/ (use this)
- Want version 3.2.2 or higher
- click on machine version to "**Download and Install R**"
- for Windows, click on "**base**"
- click on current version to install
- follow instructions

install Rstudio

RStudio: nice graphical interface for R projects

- https://www.rstudio.com/products/rstudio/download3/
- click "**Download**" under "**RStudio Desktop Personal License**"
- click on "**Installer**" for your machine
- follow instructions for easy install

install R packages

- open up Rstudio, which opens R
- On console, type

```
install.packages("qtl", "rrBLUP")
```

This will install the two main packages we will use.

qtl: R package help

- http://rqtl.org (lots of useful stuff here)
  - see R/qtl Install steps
- A brief tour of R/qtl: [pdf | code]
- A shorter tour of R/qtl: [pdf | code]
- R/qtl sample data formats
- A Guide to QTL Mapping with R/qtl: book by Karl W. Broman & Śaunak Sen

mixed model software for QTL and GWA

- EMMA (Efficient Mixed Model Association)
- EMMAX (Efficient Mixed-Model Association eXpedited)
- Convert matrix operations to vector operations
  - with one clever eigendecomposition
  - 1000x faster than TASSEL/SAS implementation
  - $O(n)$ per iteration
  - accurate and reliable
- Eleazar Eskin Lab (UCLA) standalone packages

rrBLUP: R package help

- rrBLUP
- EMMA adapted to genotypic prediction

  - `A.mat` to compute K from markers

  - `kin.blup` to predict phenotypes

- http://potatobreeding.cals.wisc.edu/software/

- Amy Jacobson slides

- Jeff Endelman Notes

R/qtl2scan

[qtl2scan](#)

- EMMA adapted to R/qtl2 package

- leave out one chromosome method

  - `calc_kinship` with `type="loco"` to compute K

  - `scan1` with `kinship` option to fit mixed model

- Karl Broman (UW-Madison) R package suite

other useful R packages

There is a very useful set of R packages that were recently developed, which are packaged together as the "Tidyverse" (http://tidyverse.org/). We will not use these directly, but they are worth knowing about. To install them type the following on the Rstudio console:

```
install.packages(tidyverse)
```

This will install a number of small packages on your system.

For the advanced R/qtl2 section, there are further packages. We will discuss that later.

organizing your work

Want to go further organizing your work?

- Karl Broman's initial steps to reproducible research
- Karl Broman's data organization tips

communication opportunities

- many online resources

  - http://www.stat.wisc.edu/~yandell/statgen/reference/: Yandell Statistical Genomics References

  - http://datacarpentry.org: spreadsheet to R in 2 days

  - https://www.lynda.com/: online skills tutorials

  - Professional Skills for Data Science

- open source tools are quite good

  - R & Rstudio for data science

  - Github for version control

- Google Drive, DropBox, Box, …

Material for this course at https://github.com/byandell/PlantSysGen