The 1,001 Genomes Project

Magnus Nordborg

University of Southern California, Los Angeles
and
Gregor Mendel Institute, Vienna
The Gregor Mendel Institute

- Devoted to basic research in plant biology
- Part of the Vienna Biocenter, with ~1500 scientists from over 40 countries
- Currently 8 research groups (~100 people), but hiring, and offers are as good as the local pastries (which are very good indeed)
The Arabidopsis Advantage: natural inbred lines
The Arabidopsis Advantage: natural inbred lines

- Collect natural genotypes
The Arabidopsis Advantage: natural inbred lines

- Collect natural genotypes
- Bulk and store (without nasty smells)
The Arabidopsis Advantage: natural inbred lines

- Collect natural genotypes
- Bulk and store (without nasty smells)
- Grow in replicate under controlled conditions
Genotype once, phenotype many times...

- Proof-of-principle paper with 107 phenotypes in 96 [sic] individuals published in 2010 (Atwell et al., Nature)

- More than 1,300 lines have been genotyped using a 250k SNP chip (Horton et al., Nature Genetics, in press)

- More than 1,000 lines are being sequenced (http://www.1001genomes.org)
1,001 Genomes Project

- 80 by Weigel (Cao et al., Nature Genet., 2011)
- 18 by Mott (Gan et al., Nature, 2011)
- 200 by Nordborg (in prep.)
- 200 by Ecker (in prep.)
- >500 by Monsanto (coordinated by Bergelson; sequenced by end of year)
Da Plan

- The reads will be hosted by Argonne and the University of Chicago.

- A unified pipeline will (somehow) be run on all the raw data to create a manageable number of unified polymorphism data sets.

- These will be published along with a database of phenotypes and information about the origin of the lines (which will of course be sent to the stock center).

- We will also produce tools for do-it-yourself GWAS and another analyses.
Integration

• We are collaborating with Paul Kersey (EBI and the “transPlant” project), Doreen Ware (Gramene) and Ed Buckler to avoid too much redundancy

• Beyond that, it is hard enough to manage the (unfunded!) project as it is...
We have barely begun...

- Over 2,000 “sequenced” genomes by summer,...
- ...hundreds of phenotypes in db already,...
- ...hundreds of thousands of in-between-ome phenotypes in process,...
- ...et cetera!
GWAS web-application

- Web-application for running GWAS on the fly and displaying results
- Multiple methods
- 250k dataset
- Interactive charts and visualizations
- HDF5 data storage
GWAS web-app
mRNA-Seq web-application

- ~25 000 phenotypes
- 2 environments (10°C & 16°C)
AtPolyDB

- Based on GDPDM (Ed Buckler)
- Database for:
  - Genotypes
  - Phenotypes & experiments
  - Germplasm & taxonomy
- GWAS results
- Gramene ontology terms are used
- Synced with internal LabCollector database
- Pleiotropy & GxE analysis