Generating Large Scale Pedigree Drawings for Genetic Studies

R. Mayani\textsuperscript{1}, G. Mehta\textsuperscript{1}, E. Deelman\textsuperscript{1}, K. Seth\textsuperscript{1}, J. Vöckler\textsuperscript{1}, F. Wang\textsuperscript{2}

\textsuperscript{1}Information Sciences Institute, University of Southern California; \textsuperscript{2}Dept. Of Psychiatry, Washington University

www.nimhgenetic.org/pedigree

**Pedigree Drawing Service**
- Current Pedigree Services provide capability to draw a pedigree interactively single individual at a time.
- Others can plot only a single pedigree from an input file.
- Large scale studies collect data for hundreds of families totaling thousands of individuals.
- Tedious repetitive work to plot pedigrees taking weeks to months.
  - Draw individual pedigree’s by hand or import each pedigree CSV and then verify results.
  - Collect all pedigrees and create single file for distribution.
- Web based bulk pedigree generation service –
  - Upload a single disease distribution with multiple pedigrees in a single file.
- Based on open source Madeline plotting software –
  - No licensing costs, can be maintained and customized as required.
- Quick generation of high quality PDF booklet with multiple pedigrees, legends and index.
  - Generate hundreds of pedigrees in minutes.
- Simple CSV format file to provide multiple pedigree.
  - Requires 5-6 set fields while rest can be used to show additional information per individual.

**Generating Pedigrees**
- Upload CSV in specified format.
- Select mode: Hatch Pattern or Color.
- Select labels to show below each individual.
  - Multiple labels allowed with custom alias.

**High Resolution Pedigree Images**

**Pedigree Generation Debugging**
- Debug windows shows plotting status for each family.
  - Number of individuals in input family and output plot at a glance.
  - Log file for each family highlighting any errors detected by the web service or Madeline.
  - Debug report provides information about missing individuals, gender validations, and other QC checks.

**Automated Quality Control**
- Analysts used to perform validation/quality control in SAS scripts manually.
  - Error prone, requires analyst training.
- Support for basic data validation and quality control.
  - Leveraging Madeline’s included QC.
- Currently enhancing service to add advanced quality control models to check or fix input pedigree data.
  - Ability to select which QC modules to run.
  - Ability to upload custom QC modules.

**Acknowledgments**
- Jay Tischfield, John Rice, Yigal Arens, Linda Brzustowicz, Scott Saccone, Bill Howells for their suggestions.
- Center for Genomic Studies and the Pedigree Service are funded by National Institute of Mental Health grant #U24MH068457.