

Generating Large Scale Pedigree Drawings for Genetic Studies

R. Mayani¹, G. Mehta¹, E. Deelman¹, K. Seth¹, J. Vöckler¹, F. Wang²

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

www.nimhgenetic.org/pedigree

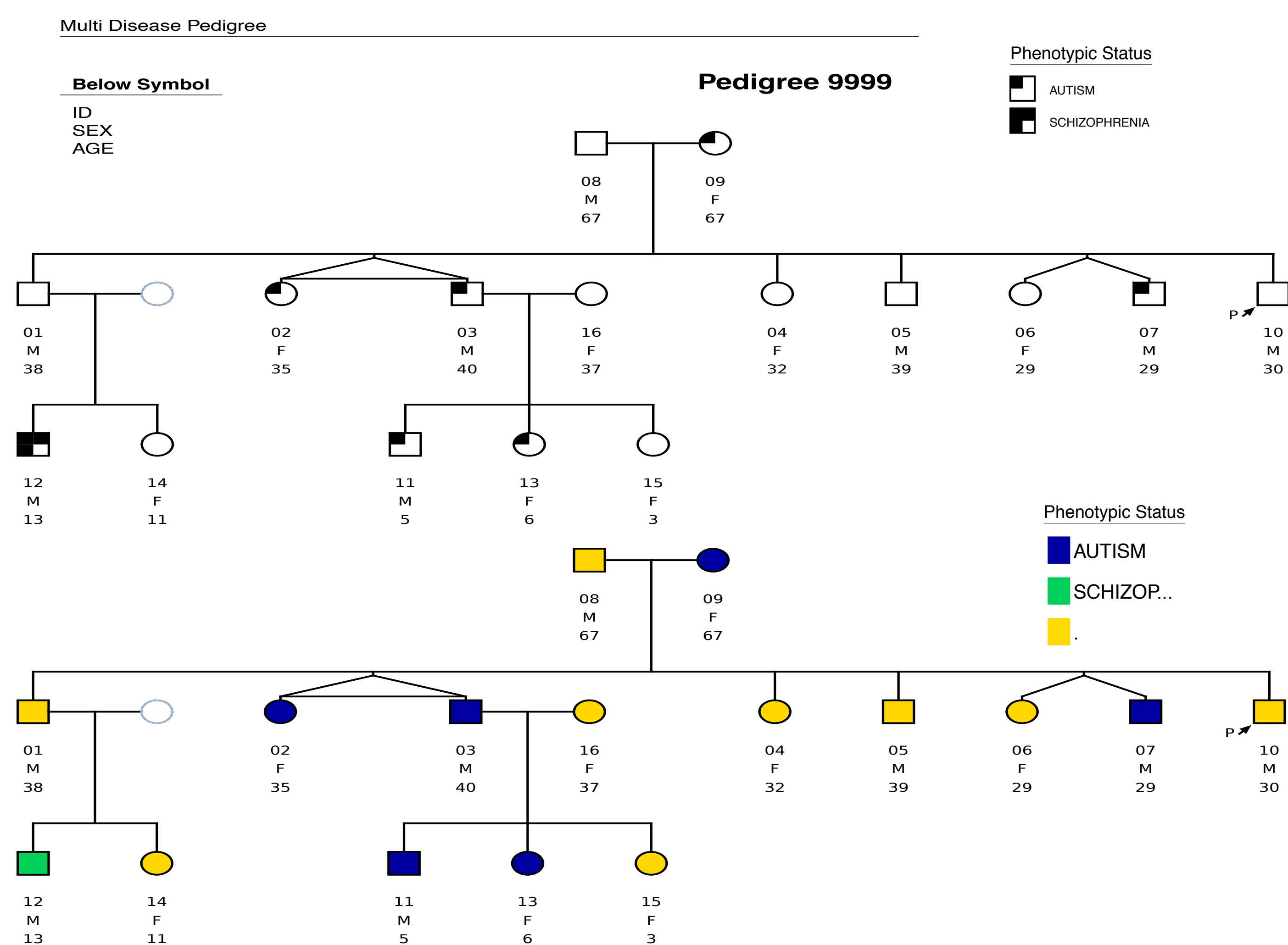
Pedigree Drawing Service

- 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
- Drop files here to upload
- Color Options: Partial Hatch, Color
- Labels: familyid, father, deceased, dztwin, dxsys, individualid, mother, affected, cell_id, gender, proband, mztwin, age
- Select Autofill option in Hatch pattern mode or assign patterns manually to affection status
 - Assign Description to each affection for printing on Legends page
 - Click Generate to produce high resolution PDF

High Resolution Pedigree Images



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

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.

Debug Console:

Submitting file nimh_study50_efam2-fake.txt
File nimh_study50_efam2-fake.txt uploaded.
Configurable parameters generated.
Sending file to Madeline for processing... this can take some time.
Pdf generated for correct families in file nimh_study50_efam2-fake.txt

Family ID	Individuals in input	Individuals plotted	Log File
9999	16	16	Madeline Log

done..

Clear Debug

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

