



Seedling Select: Cost Modeling Software for Marker-Assisted Seedling Selection (MASS) in Rosaceae Tree Fruit

Sushan Ru¹, Daniel Edge-Garza¹, Julia Harshman², Kate Evans², Sook Jung¹, Dorrie Main¹, Cameron Peace¹

¹Department of Horticulture, Washington State University, Pullman, WA, ²Department of Horticulture, Washington State University, Wenatchee, WA

INTRODUCTION

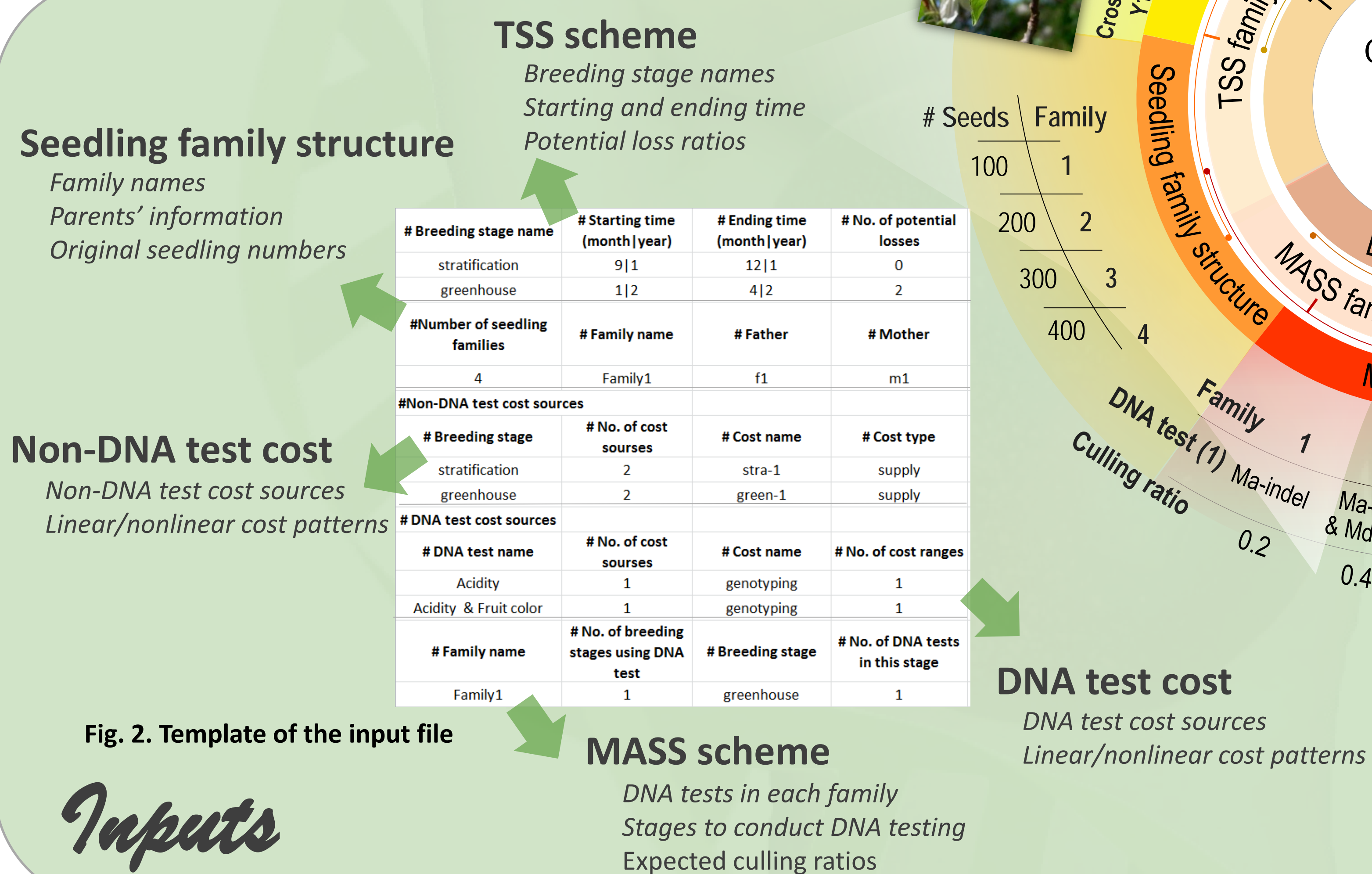
The **Rosaceae** family includes most of the temperate tree fruit crops such as apple, apricot, peach, pear, and cherry. **Seedling selection** is the process of selecting superior seedlings generated from crosses for cultivar development. **Traditional seedling selection (TSS)**, based on phenotypic information, is expensive for most rosaceous tree fruit because of their long juvenile phases and large plant sizes. **Marker-assisted seedling selection (MASS)** can reduce selection cost by selecting seedlings in early stages using trait-level predictive DNA tests. However, this cost saving varies case by case and **accurate cost estimation** before conducting MASS is very important. **Seedling Select** is a software tool to enable accurate and flexible cost estimation for rosaceous tree fruit. Currently programmed in Python™, **Seedling Select** can estimate seedling selection costs in various MASS schemes, compare the cost of MASS with TSS and recommend cost-efficient schemes. Here we report on the structure and algorithm of **Seedling Select** using a case study of the WSU apple breeding program, accesses of the software, and plans of future development.

Model

Seedling Select estimates seedling selection cost based on a MASS cost modeling pipeline proposed in this study (Fig. 1). This pipeline requires input information regarding TSS scheme, MASS scheme, seedling family structure, DNA test cost structure, and Non-DNA test structure (Fig. 2). Users fill out an input template file with above information and run **Seedling Select**.

Seedling Select will first model family size changes throughout TSS stages (TSS family structure) based on TSS scheme and seedling family structure. It will then model family size changes through MASS stages (MASS family structure) based on MASS scheme and seedling family structure. With known TSS and MASS family structure, TSS cost and the non-DNA test portion of MASS cost can be estimated according to non-DNA test cost structure, while DNA test cost in MASS can be estimated according to DNA test cost structure. Total MASS will be calculated by summing up non-DNA test cost and DNA test cost in MASS. In the end, cost comparisons will be made between TSS and MASS and cost-efficient schemes will be recommended.

Fig. 1. MASS cost modeling pipeline →

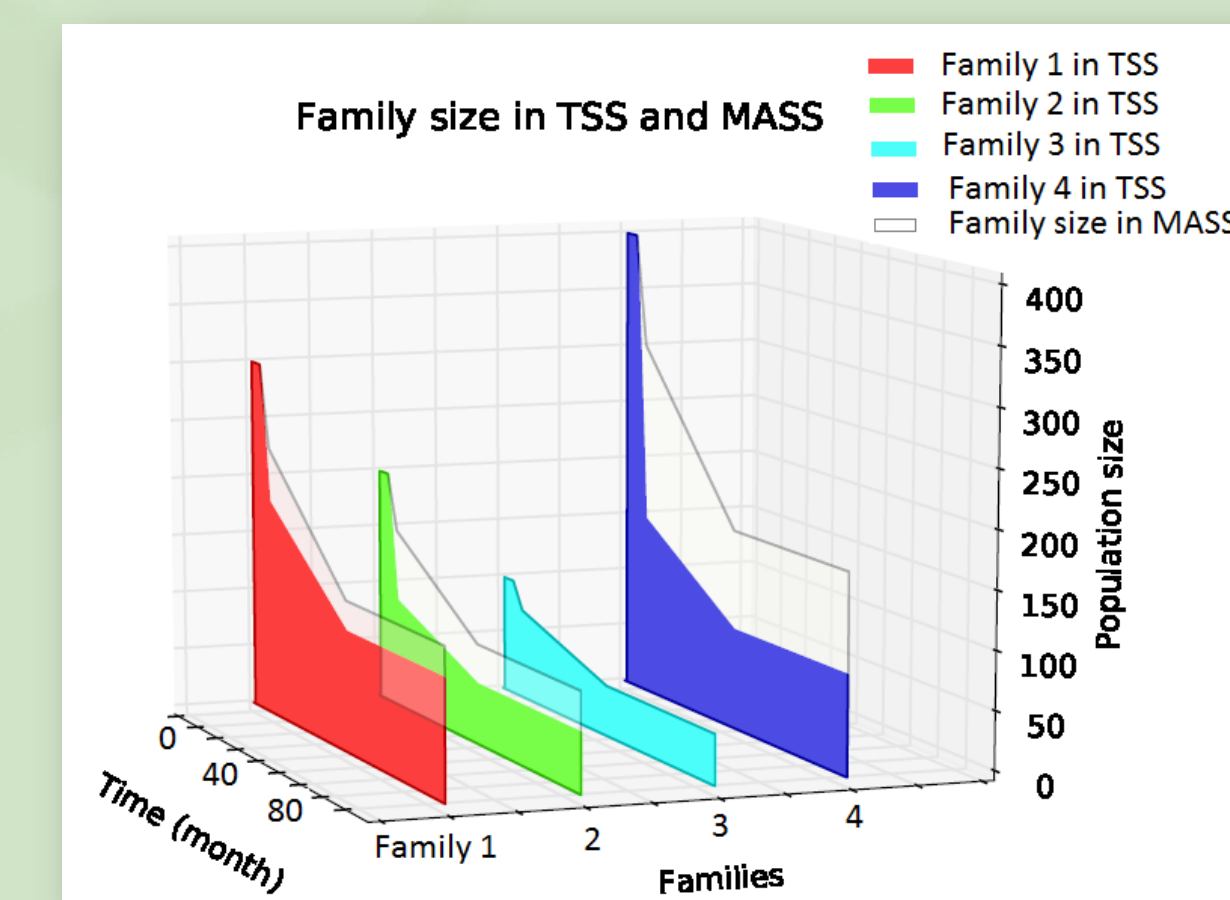
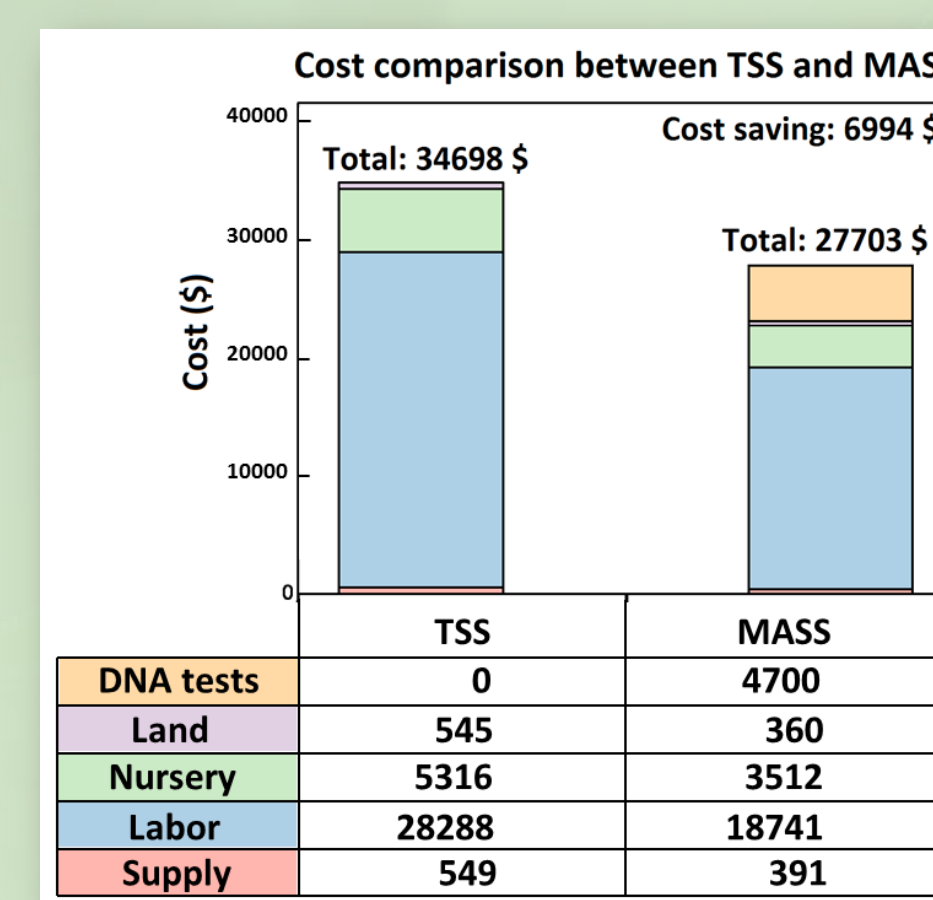
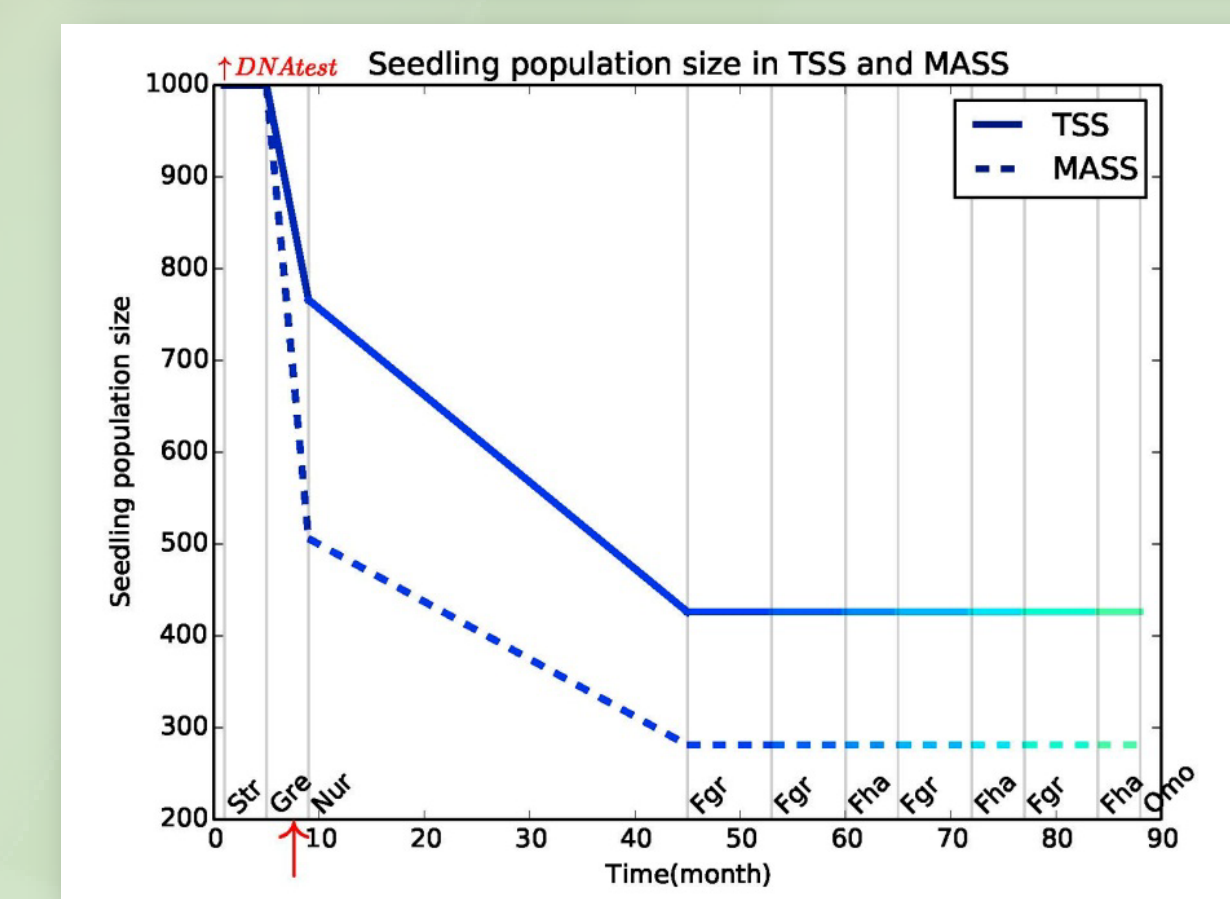


Outputs

Seedling Select generates four output files once the calculation is done. The first output file is a spreadsheet displaying both input and output data (Fig. 3), where users can find summaries of input information and cost comparisons between MASS and TSS. The other three output files are graphic files. "Population size.pdf" illustrates how population size changes throughout TSS and MASS stages. It also indicates in which stage DNA test(s) are conducted (Fig. 4). "Family structure.pdf" is a 3D plot illustrating family sizes in each stage of MASS and TSS (Fig. 5). "Cost comparison.pdf" displays cost breakdown of MASS and TSS and compares their differences (Fig. 6).

Fig. 3. Output data file: → MASS and TSS cost by stage

	TSS costs(\$)	MASS non-DNA test costs(\$)	MASS DNA test costs(\$)	MASS costs(\$)
stratification	55	55	0	55
greenhouse	270	270	4700	4970
nursery	5316	3512	0	3512
fgrow-1	588	388	0	388
fgrow-2	187	124	0	124
fharvest-2	4537	2993	0	2993
fgrow-3	187	124	0	124
fharvest-3	4537	2993	0	2993
fgrow-4	187	124	0	124
fharvest-4	4537	2993	0	2993
omoving	14297	9430	0	9430
Total	34698			23003



Software access

Seedling Select is currently available at the Genome Database for Rosaceae (www.rosaceae.org/breeders_toolbox/seedling_select_front_page). You can download the input file template from the webpage and modify it with your own data. After running **Seedling Select**, output files will be available to download.

You can also get Python scripts from the author (sushan.ru@email.wsu.edu) and run it on your own desktop with a Canopy Enthought™ environment. Canopy Enthought™ can be freely accessed at www.enthought.com/downloads/.

Future development

Future development goals of **Seedling Select** are to further improve its calculating ability and user-friendliness:

- Test current model in various rosaceous tree fruit crops as well as adjust input template and calculation models for different crops (Summer, 2014)
- Develop features for automatic modeling and comparison of all possible selection schemes (Summer, 2014)
- Build intuitive online interfaces to assist users in generating input file and operating **Seedling Select** (Fall, 2014)