The RosBREED

Pedigree-Based Breeding Information Management (BIM) System

Sook Jung[1], Taein Lee[1], Kate Evans[1], Sushan Ru[1], Amy Iezzoni[2], Gennaro Fazio[3], Cameron Peace[1*], and Dorrie Main[1*]

[1]Washington State University, [2]Michigan State University, [3] USDA- ARS. *Corresponding authors : dorrie@wsu.edu and cpeace@wsu.edu

Introduction

RosBREED is a national USDA NIFA SCRI funded project dedicated to enabling Marker-assisted Breeding (MAB) in Rosaceae. The Pedigree-Based Breeding Information Management (BIM) System is a collection of web-based breeding assistant tools facilitating the application of the RosBREED MAB Pipeline in Rosaceae. Hosted on the Genome Database for Rosaceae (GDR, www.rosaceae.org) through the Breeders Toolbox, the BIM System is beginning to enable integration of a range of information sources to provide knowledge to use in breeding decisions. Current available features of the Toolbox include:

1. Data management which includes: uploading, browsing, searching and downloading (1) genotypic and phenotypic data of user-chosen varieties (*Reference Germplasm Database*), and (2) reported marker-locus-trait associations for apple, peach, cherry and strawberry (*Trait Locus Warehouse*)
2. Data translation: generating input files to use in breeding software (e.g. generating input files for Pedimap™ - a tool for exploring and visualizing the flow of phenotypes and alleles through pedigrees)
3. Providing a decision support tool for identifying efficient crosses (*Cross Assist)*, which leverages knowledge on DNA-based breeding values

In this poster, the overall structure of BIM System along with details of current available features especially *Cross Assist* are discussed. Future functionalities are also described.

BIM System Structure

The Marker-Assisted Breeding (MAB) Pipeline is a systematic approach to translating genomics research discoveries into practical outcomes for breeders. To ease and streamline such translation, the BIM System consists of a series of software and database modules that support the MAB Pipeline.

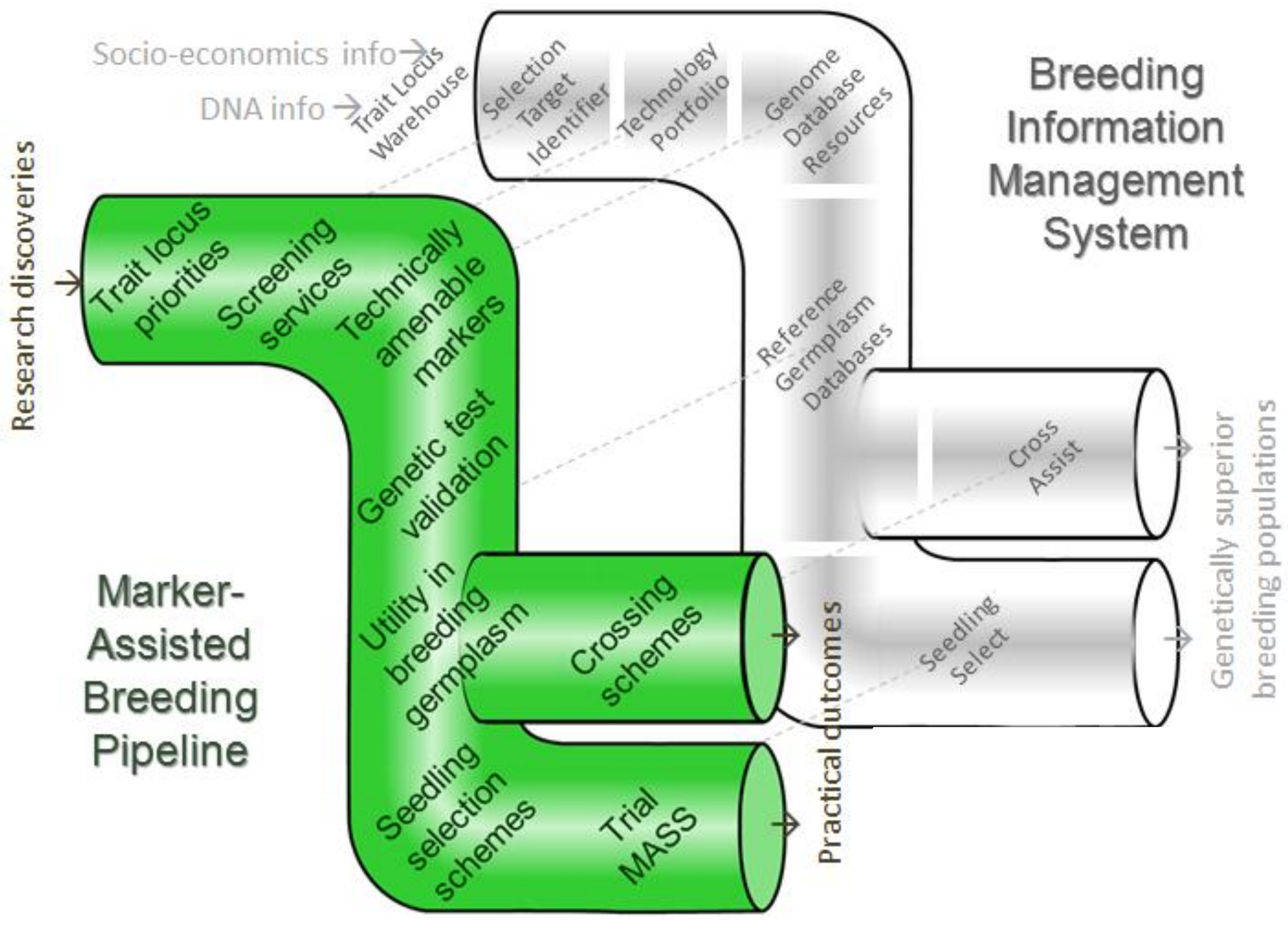


Fig. 1. The RosBREED BIM System and MAB Pipeline

BIM System Current Features

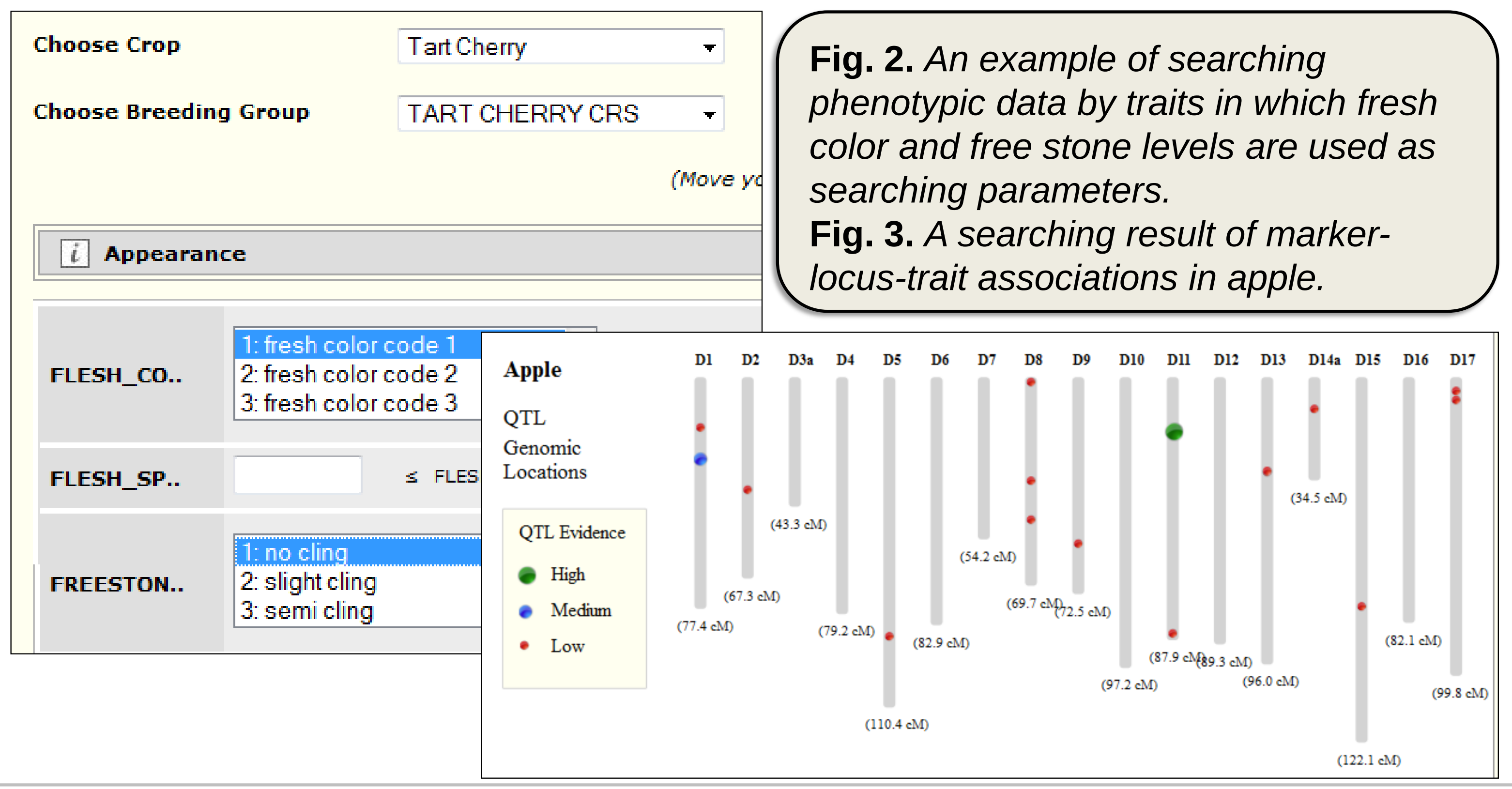
1. Data Management

- ***Reference Germplasm Database* (Fig. 2.)**

An available functionality of the *Reference Germplasm Database* in the BIM System is to serve as an information warehouse for uploading, archiving, browsing, searching and downloading phenotypic and genotypic data for RosBREED demonstration breeding programs. Data is available for over 5,344 accessions from apple, peach, cherry and strawberry.

- ***Trait Locus Warehouse* (Fig. 3.)**

This module allows access to reported marker-locus-trait associations for Rosaceae. *The database is continually updated.*


Choose Crop Tart Cherry

Choose Breeding Group TART CHERRY CRS

(Move yo

Appearance

FLESH_CO.. 1: fresh color code 1 / 2: fresh color code 2 / 3: fresh color code 3

FLESH_SP.. ≤ FLES

FREESTON.. 1: no cling / 2: slight cling / 3: semi cling

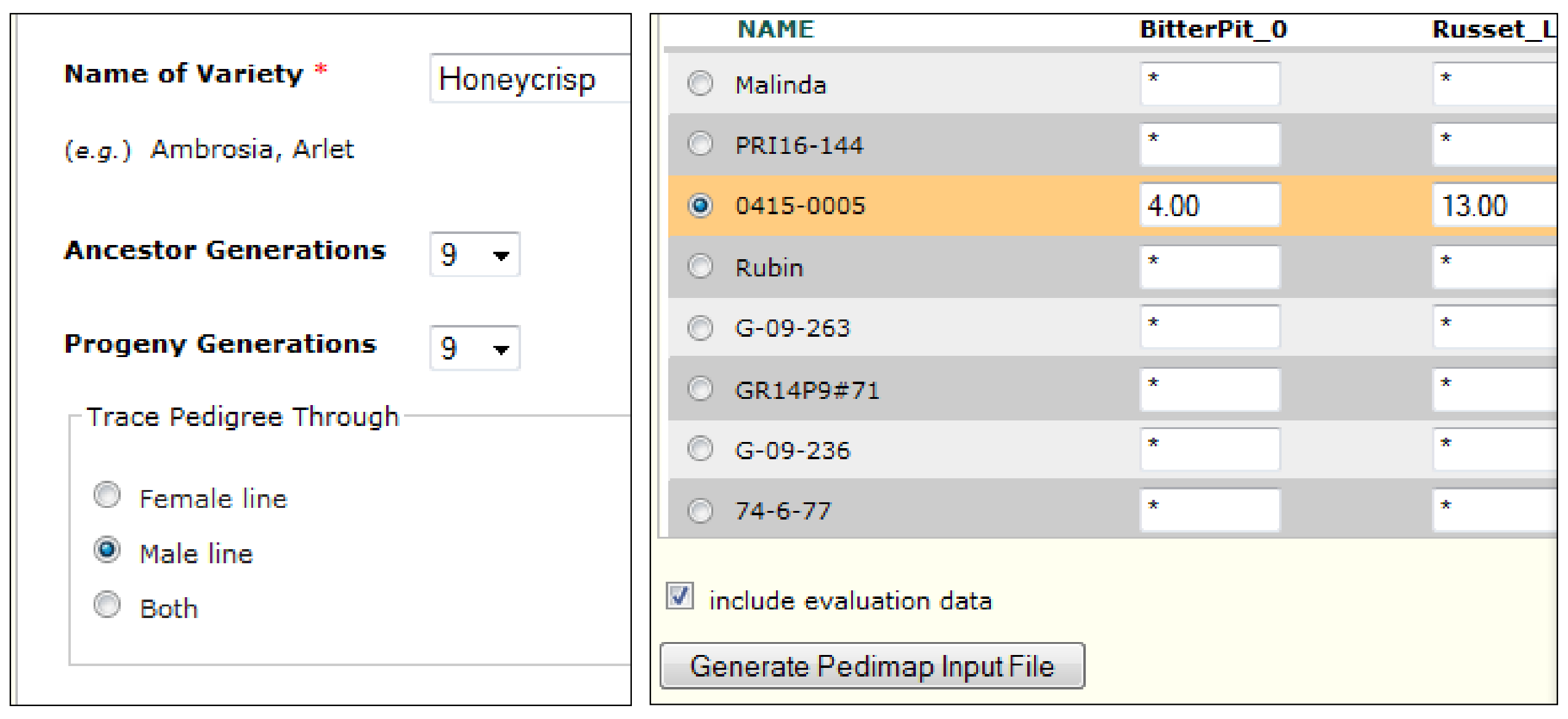
Fig. 2. *An example of searching phenotypic data by traits in which fresh color and free stone levels are used as searching parameters.*

Fig. 3. *A searching result of marker-locus-trait associations in apple.*

2. Data Translation

- ***Genome Database Resources***

An available function of the Genome Database Resources is to generate input files for Pedimap™, which is a critical tool used in RosBREED for Pedigree-Based Analysis. This functionality makes the transition from data collection to data utilization more efficient and accurate (**Fig. 4**.).


Name of Variety * Honeycrisp

(e.g.) Ambrosia, Arlet

Ancestor Generations 9

Progeny Generations 9

Trace Pedigree Through

○ Female line
● Male line
○ Both

	NAME	BitterPit_0	Russet_L
○	Malinda	*	*
○	PRI16-144	*	*
●	0415-0005	4.00	13.00
○	Rubin	*	*
○	G-09-263	*	*
○	GR14P9#71	*	*
○	G-09-236	*	*
○	74-6-77	*	*

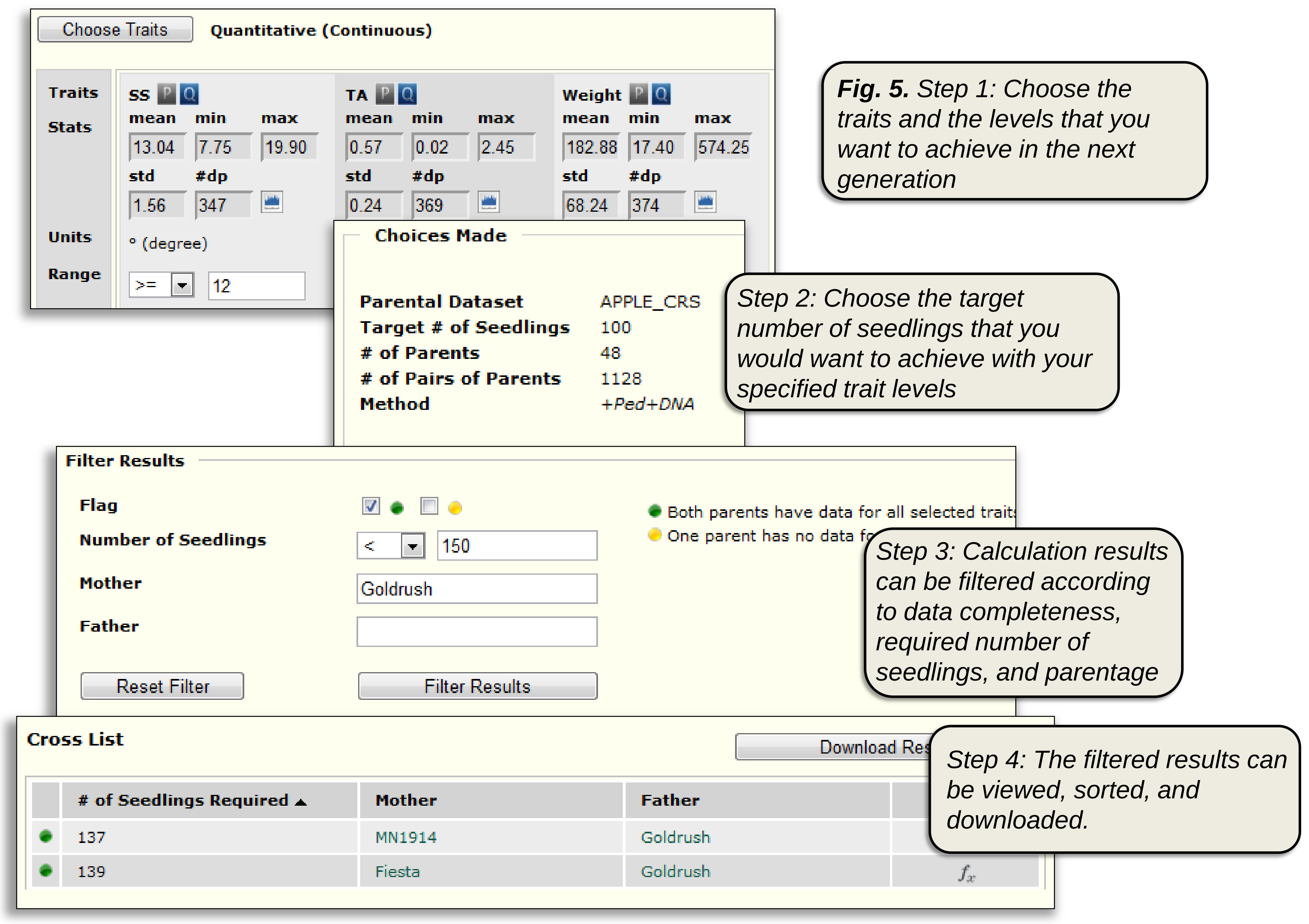
☑ include evaluation data

Fig. 4. Setting parameters for Pedimap™ input files.

3. Decision Support

- ***Cross Assist***

Cross Assist is a RosBREED decision support tool which determines efficient parent combinations ("crosses") among a breeder's available parents. This efficiency is determined by the number of seedlings that would need to be created to result in a target number of seedlings predicted to each perform within specific trait thresholds. An efficient cross would be one that achieves the trait thresholds with relatively fewer seedlings. Three methods of determination of efficient crosses are offered by *Cross Assist*, which are "*Phenotype*" (uses only phenotypic information of individuals in the dataset), "*+Pedigree*" (uses both phenotypic and pedigree information), and "*+Ped+DNA*" (uses phenotypic, pedigree information and information provided by DNA-based functional genotypes).



Quantitative (Continuous)

Traits	SS			TA			Weight		
Stats	mean	min	max	mean	min	max	mean	min	max
	13.04	7.75	19.90	0.57	0.02	2.45	182.88	17.40	574.25
	std	#dp		std	#dp		std	#dp	
	1.56	347		0.24	369		68.24	374	
Units	° (degree)								
Range	>=	12							

Choices Made

Parental Dataset	APPLE_CRS
Target # of Seedlings	100
# of Parents	48
# of Pairs of Parents	1128
Method	*+Ped+DNA*

Filter Results

Flag ☑ ● ☐ ●

Number of Seedlings < 150

Mother Goldrush

Father

● Both parents have data for all selected trait
● One parent has no data fo

Cross List

	# of Seedlings Required ▲	Mother	Father	
●	137	MN1914	Goldrush	
●	139	Fiesta	Goldrush	f_x

Fig. 5. *Step 1: Choose the traits and the levels that you want to achieve in the next generation*

Step 2: Choose the target number of seedlings that you would want to achieve with your specified trait levels

Step 3: Calculation results can be filtered according to data completeness, required number of seedlings, and parentage

Step 4: The filtered results can be viewed, sorted, and downloaded.

BIM System Future Features

Selection Target Identifier is being developed as an online-driven database of economically weighted market classes, fruit quality traits, and other high-impact traits.

Technology Portfolio is being developed as an online tool to support breeding programs in choosing appropriate technologies and service providers for their genetic screening needs.

Genome Database Resources will be improved to further facilitate the development of genetic tests with the potential to distinguish all available functional alleles for chosen QTLs.

Reference Germplasm Database is currently serving as a breeding data warehouse, however, whose functionalities will be extended to assisting breeders with validation of genetic tests, quantification of allele functionality, and determination of the distribution of the functional alleles in the user's germplasm.

Seedling Select is being developed as a decision-support tool to assist marker-assisted seedling selection in Rosaceae.

Conclusions

BIM System currently contains four modules : *Reference Germplasm Database*, *Trait Locus Warehouse*, *Genome Database Resources* and *Cross Assist* which together assist MAB in Rosaceae through data management, data translation and cross decision support. Module development and improvement in progress includes a *Selection Target Identifier*, *Technology Portfolio*, advanced *Genome Database Resources*, advanced *Reference Germplasm Database*, and *Seedling Select* which will further ease and streamline the translation from genomics research discoveries into practical outcomes for breeders.

Acknowledgements

We acknowledge with thanks funding from the USDA's National Institute of Food and Agriculture – Specialty Crop Research Initiative project, "RosBREED: Enabling marker-assisted breeding in Rosaceae" (2009-51181-05808), the Washington Tree Fruit Research Commission and the Department of Horticulture at Washington State University and RosBREED data providers.