

A Molecular Selection Tool -MOSEL

Graham McLaren

Background

Discussions with breeders at ICRISAT, IRRI and CIMMYT indicate that the lack of a reliable, easy to use application to display and integrate genotype, pedigree and phenotype information and allow selection of lines for promotion to the next cycle of breeding is a constraint to the adoption and effectiveness of molecular breeding.

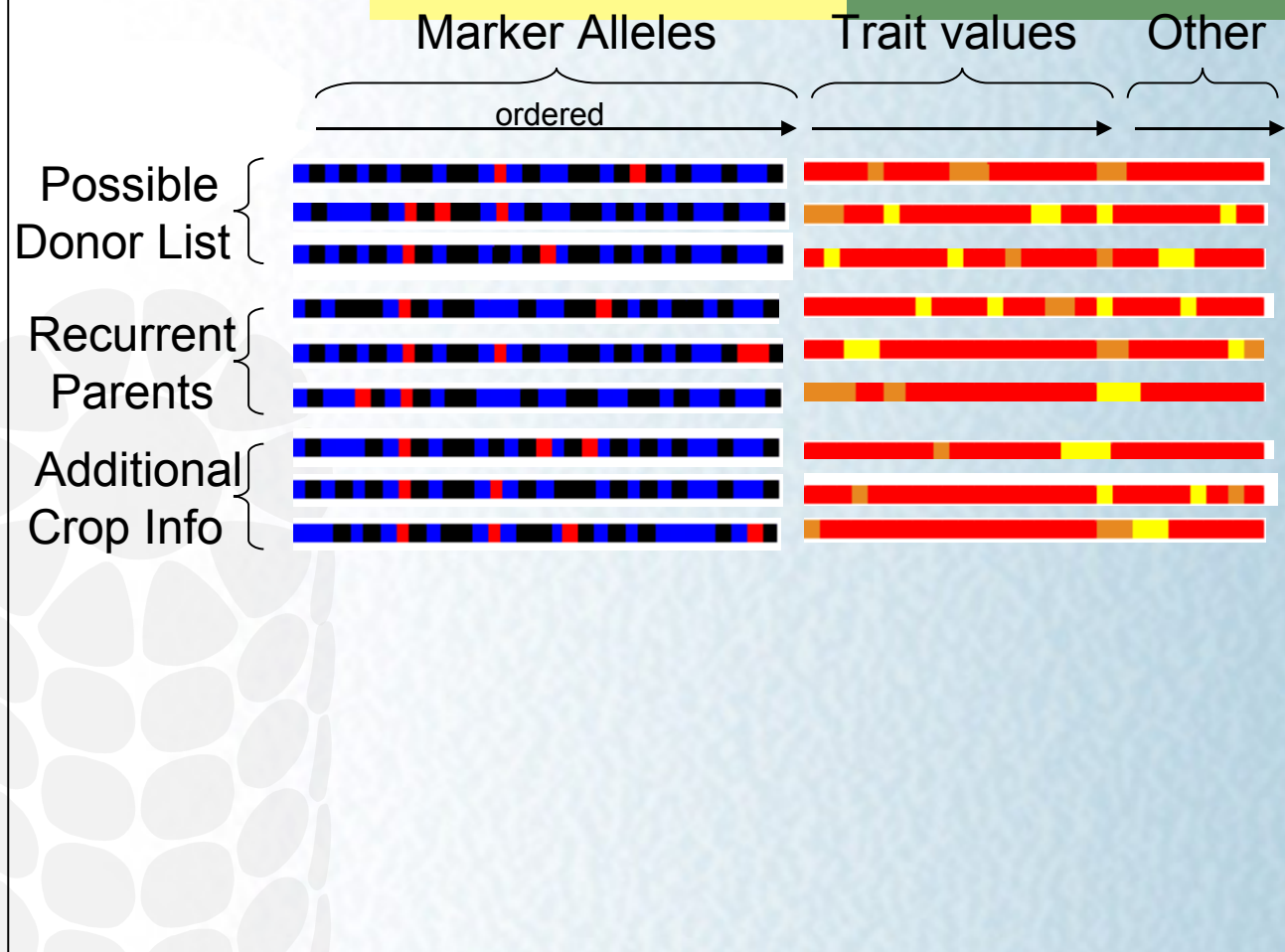
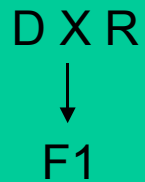
Purpose

- Access pedigree, genotype and phenotype data for test and reference lines.
- Access marker and QTL data from mapping studies
- Integrate the data sources into a single display
- Show the genetic proximity of test lines to foreground and background target genotypes
- Allow selection of test lines based on marker and/or phenotypic traits

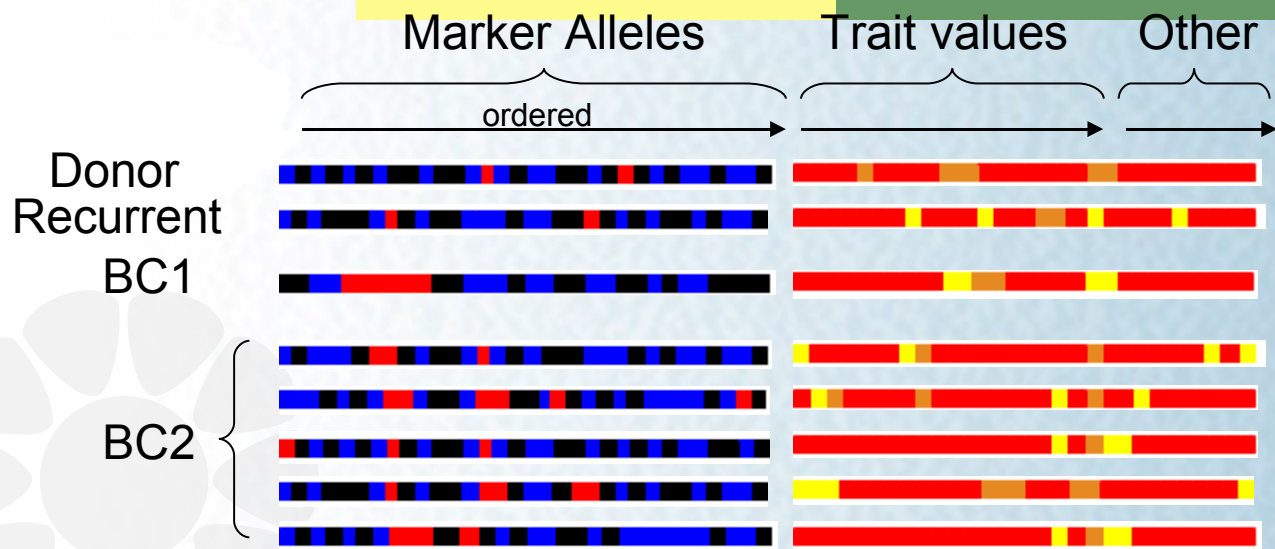
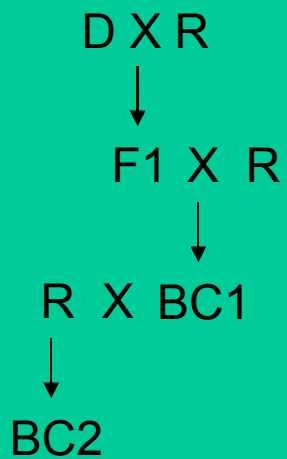
Input

- Test germplasm will be specified through a germplasm list.
- Genotyping and, optionally, phenotype data for test lines will be supplied through data studies.
 - Genotyping data consists of molecular genotypes for each test line at a number of loci.
 - Phenotype data consist of one or more phenotypic measurements on each line
- Ancestors are determined from the test germplasm by tracing pedigrees according to neighborhoods
- Founders are the earliest (genealogically), genotyped ancestors from which all the other ancestors and the test lines have been produced.
- Genetic information about the loci may be provided through a map data set.

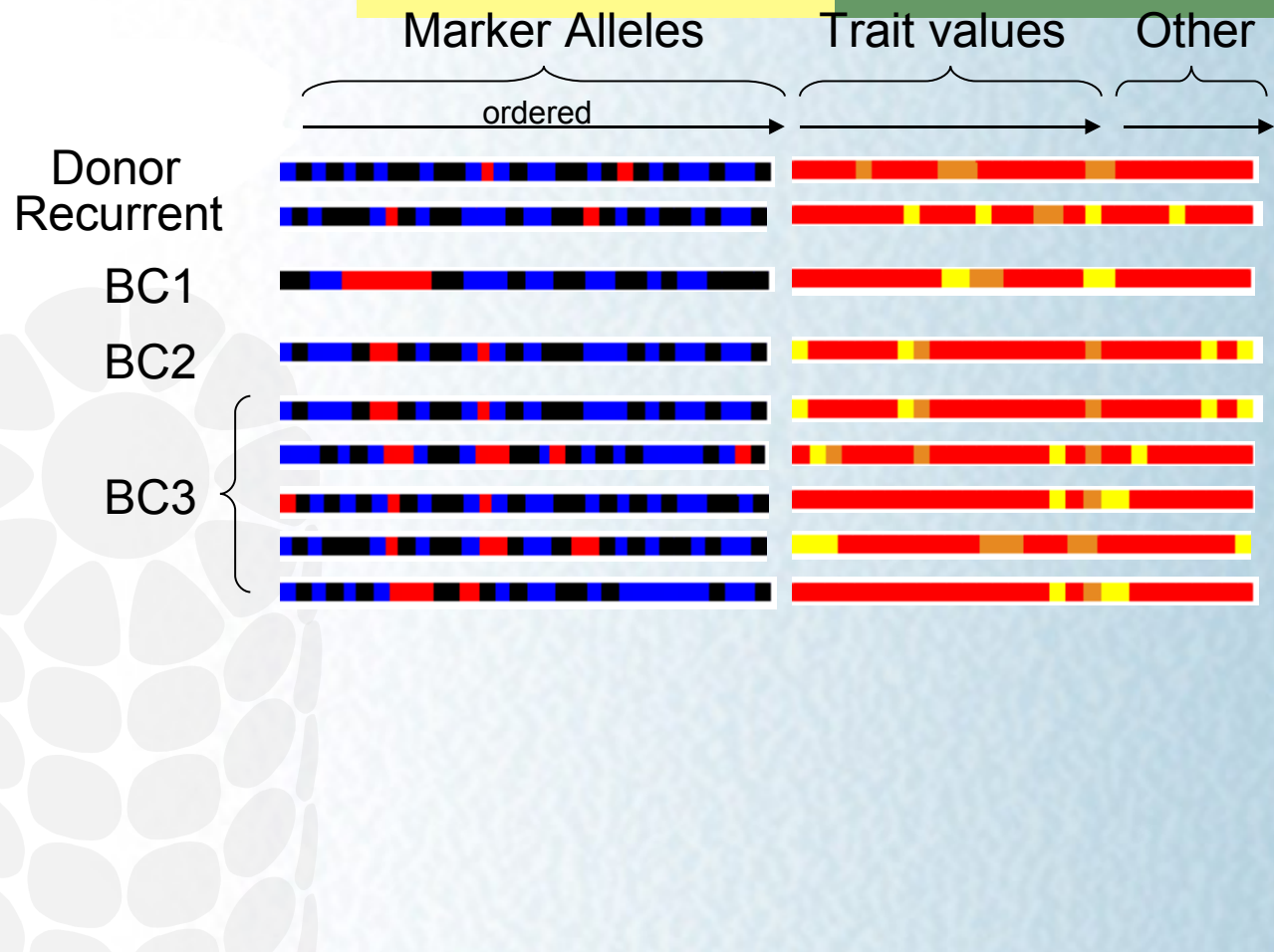
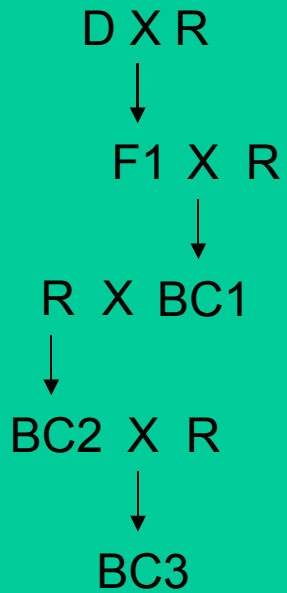
Breeding Strategy



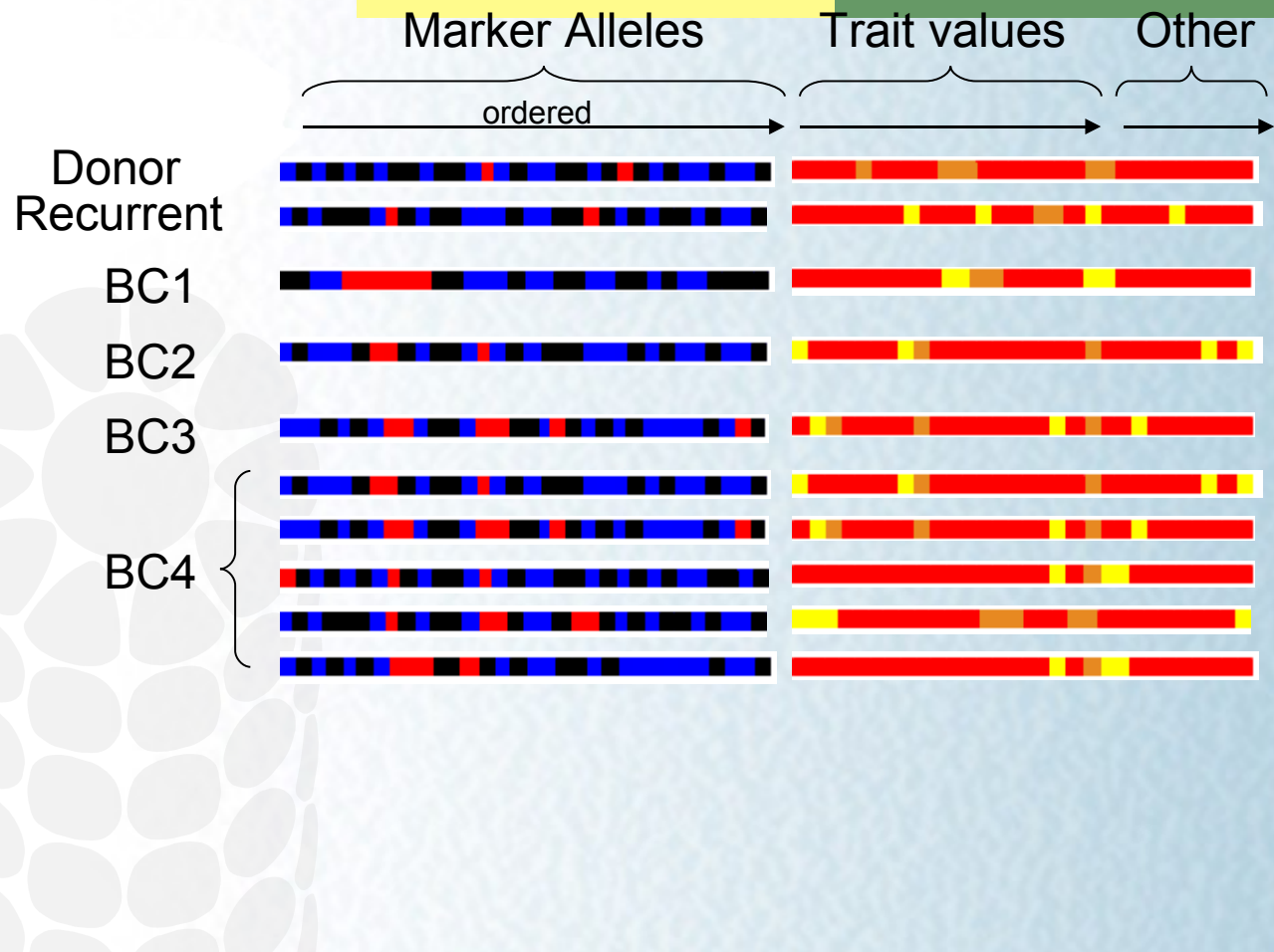
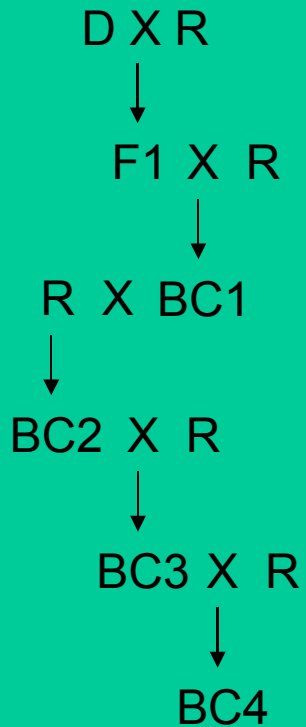
Breeding Strategy



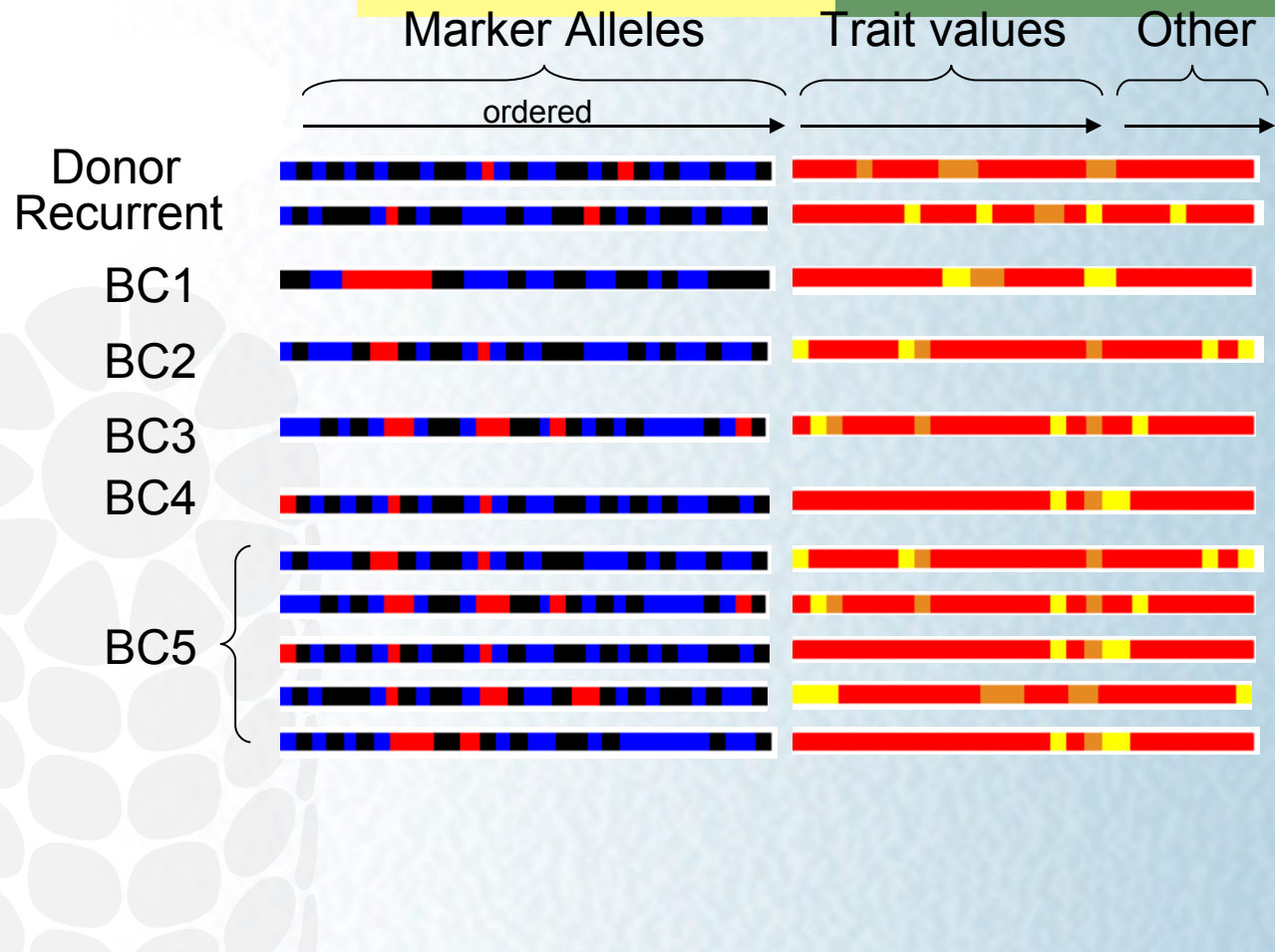
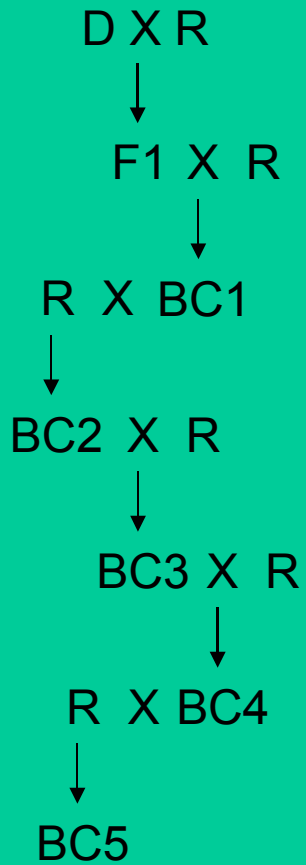
Breeding Strategy



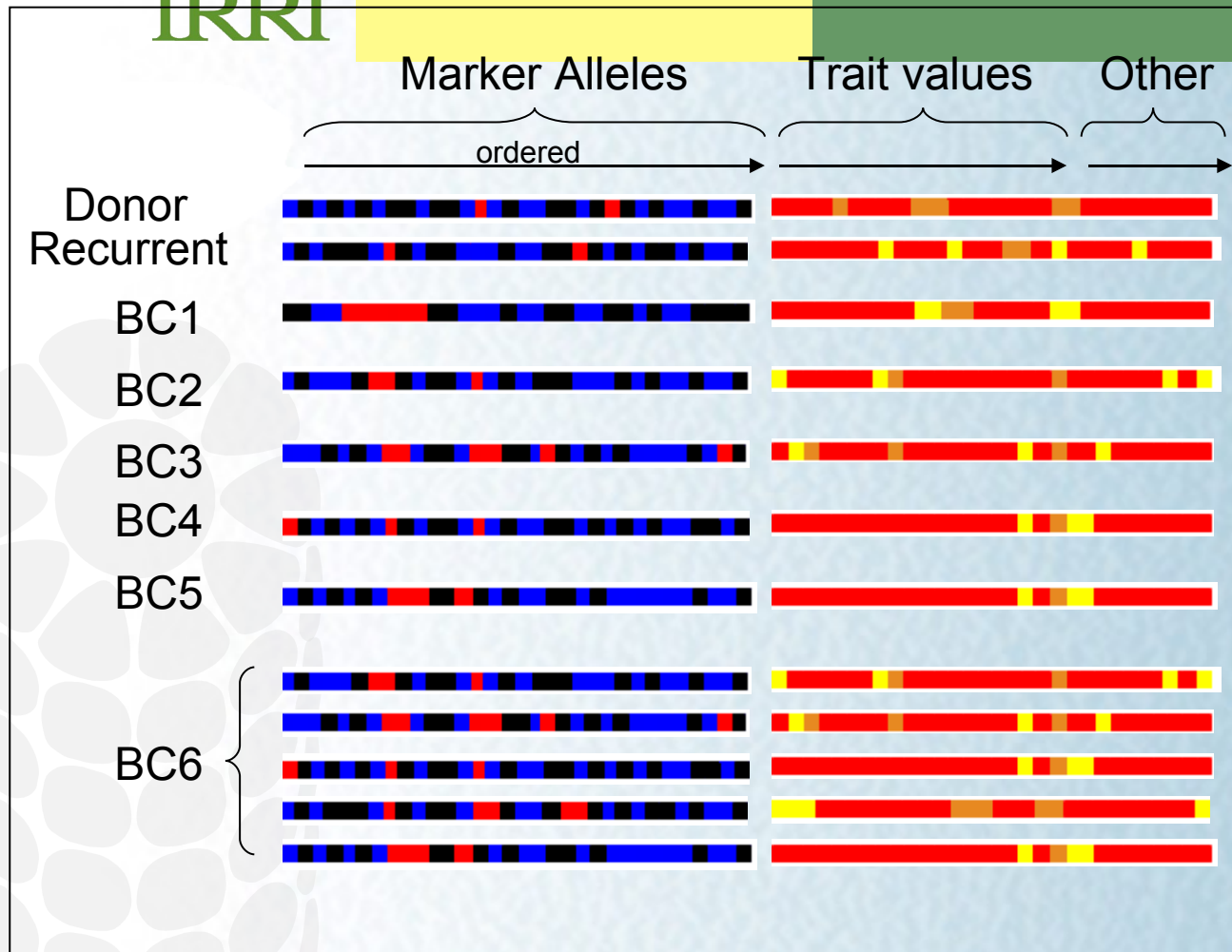
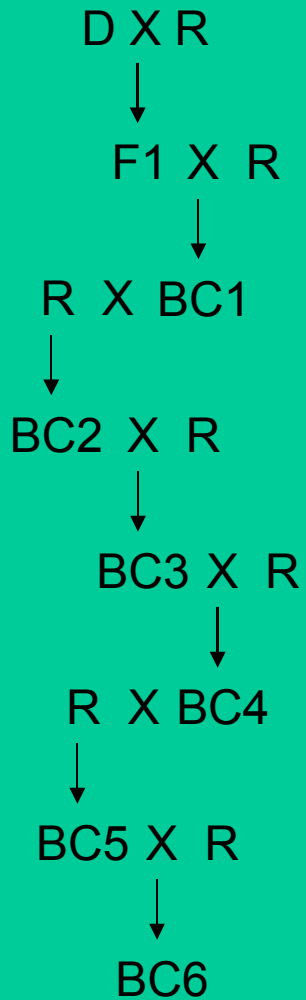
Breeding Strategy



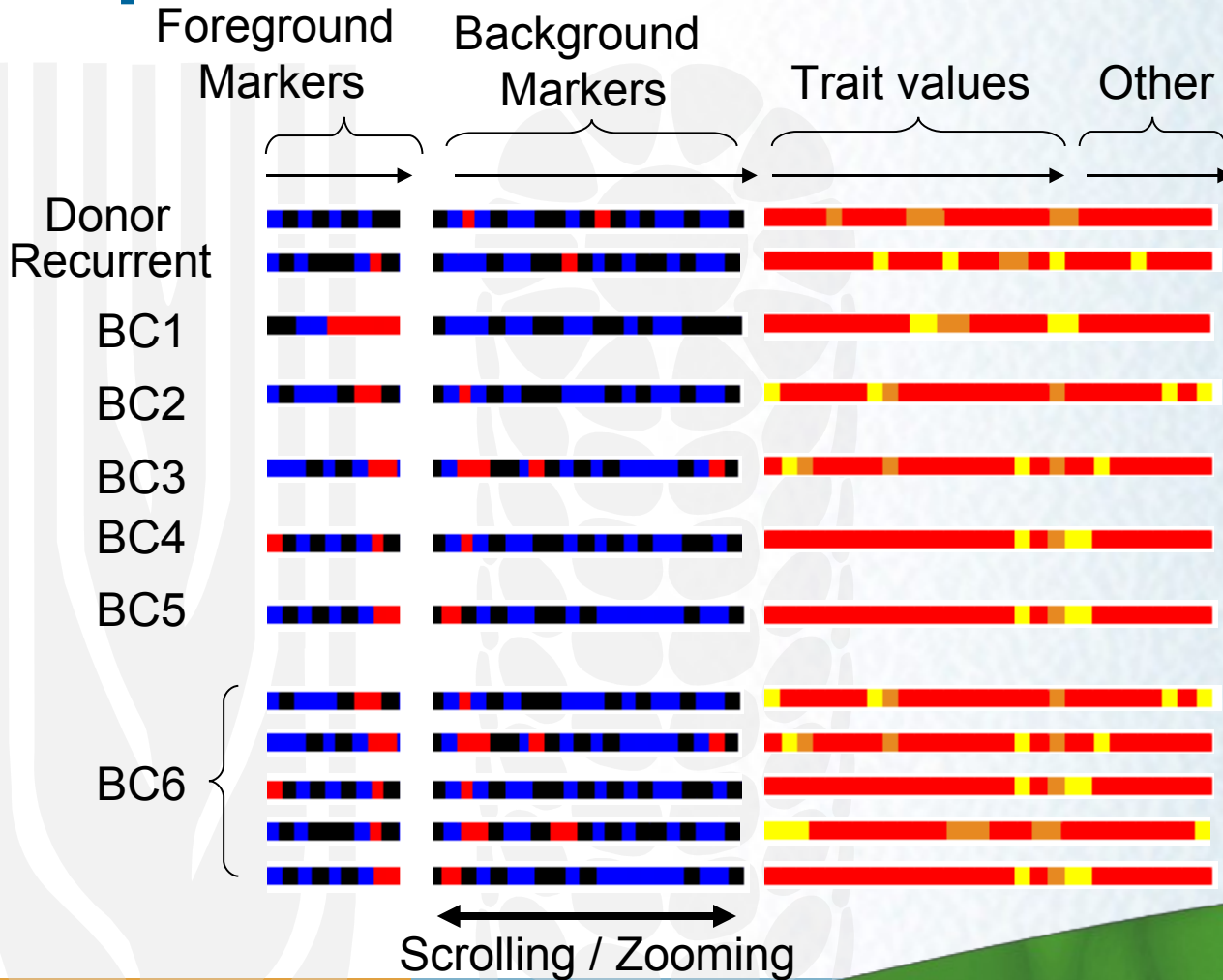
Breeding Strategy



Breeding Strategy



Options



- Hide or show
- Scrolling / Zooming
- Ordering
- Linkage to other data

Genotypes

- For any line and locus, the genotype consists of frequencies of one or more alleles and the probability that the allele comes from each founder.
- The genotype of line i ($i = 1, 2 \dots I$) at locus j ($j = 1, 2 \dots J$) consists of the frequencies X_{ijk} of allele k ($k = 1, 2 \dots K_j$) and the probabilities, p_{ijkl} that allele k came from founder l , ($l = 1, 2 \dots L$).
- For example, for germplasm i , locus j with two alleles and two founders, the genotype is:
 - AlleleID 1 =A 2=B
 - Frequency X_{ij1} X_{ij2}
 - Founder 1 P_{ij11} P_{ij12}
 - Founder 2 P_{ij21} P_{ij22}

Data Source Parameters

- The identification of the list of test germplasm (ICIS list ID)
- The identification of the test line genotyping data (ICIS study)
- The identification of the test line phenotype data (ICIS study)
- Ancestor genotyping data (ICIS study and effect)
- Ancestor phenotyping data (ICIS study and effect)
- Set of founder germplasm IDs (if the founders are amongst the ancestors) or identification of the list of founder germplasm (ICIS germplasm list ID)
- Founder genotype data source
- Founder phenotype data source
- Map data source (ICIS map study)

Requirements

- The data needs to be held in a GCP compliant data source that can hold phenotype, genotype and pedigree data
 - ICIS is most suitable database for this
 - Training on using ICIS to manage a breeding program is required
- The database must updated each time new data generated usually during each crossing cycle
 - This must be relatively easy to do

Prerequisites

- Initial curation of ICIS data
 - with initial donor and recurrent lines.
 - IRIS and IWIS are potential starting points. IWIS not yet available
- Provide training in the use of ICIS
 - How to upload phenotype and genotype data in ICIS
- ICIS Workbook for uploading data
 - Solved for most data, except SNP, DArT (coming soon)
- Development staff in position and trained

Prerequisites (2)

- ICIS implementation for Data Source API
 - Near completion for pedigree data
 - Further work is required for phenotype and genotype data
- Optional requirement for additional genomic data
 - Data source implementation for Gramene, MaizeGDB and GrainGenes