

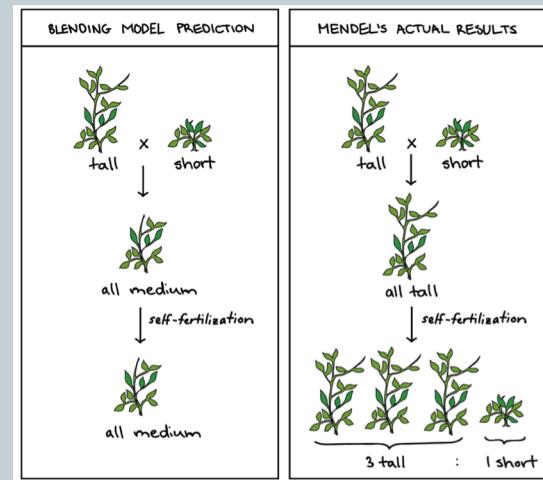
# **LINKAGE MAPPING**

**Sujan Mamidi**

## Mendel's laws of Inheritance

### I. Law of dominance

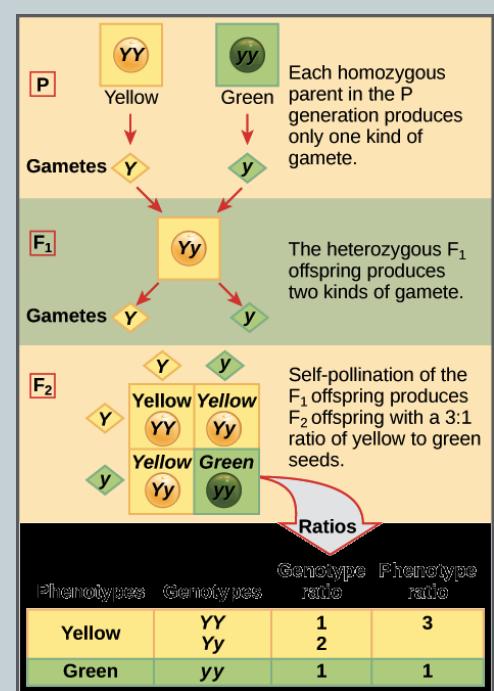
When two homozygous individuals with one (TT & tt) or more (RRYY & rryy) sets of contrasting traits are crossed, the visible characters in the F<sub>1</sub> hybrids are **dominant** and those which do not appear are **recessive**.



Khan academy

## 2. Law of segregation / Purity of gametes

- Only one of the two copies ( $YY$  &  $yy$ ) of the factors or genes enters each gamete (egg or sperm cell) and this allocation is random.
- During fertilization, gametes unite to produce  $F_1$ .  $F_1$ s are heterozygotes with one gene of  $Y$  and  $y$ .
- When  $F_1$  undergo fertilization ( $Yy \times Yy$ ), two types of gametes are formed with each gene having 50 %. All possible combinations occur in  $F_2$  when gametes unite at random.

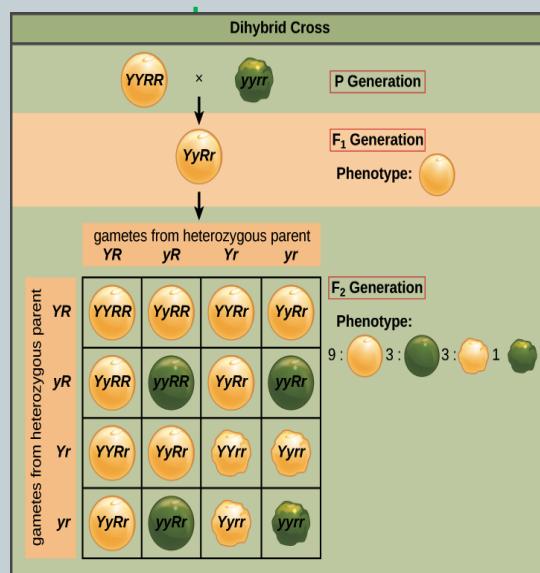


Khan academy

### 3. Law of Independent assortment

- Alleles of two or more different genes for each pair of traits/characters assort out independently into gametes of the other pairs.
- Dihybrid cross ratio – 9:3:3:1 (phenotypic ratio)

Dihybrid cross is an



Khan academy

## **Alleles may not assort independently at different loci**

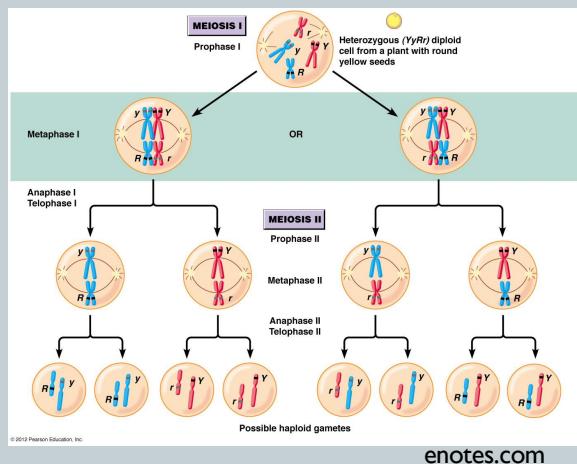
### **Reasons:**

1. Either by **chance/probability**
2. **Linkage** – genes/alleles at different loci on the same chromosomes are very close together

Law of independent assortment was framed before we knew “**meiosis**”.

## What happens during meiosis?

- Haploid gametes are formed from diploid cells.
- Independent Assortment of genetic loci**
1. Takes place on different chromosomes
  2. Also takes place on the same chromosome if the genes/loci are far apart via **crossover/recombination**.

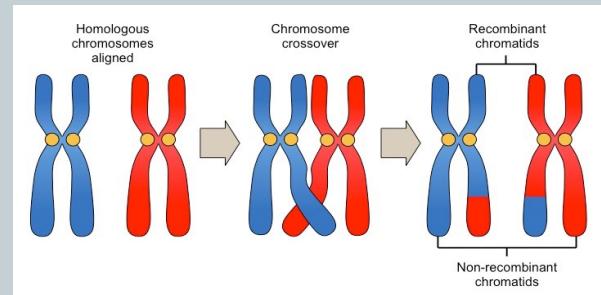


enotes.com

# Linkage and Crossing over

## Linkage

- ✓ Genes on the same chromosome are linked and they fail to assort independently
- ✓ Such genes segregate together in a population.
- ✓ Linkage is based on the frequency of crossing over between the two genes.



## Crossing Over/recombination

- ✓ Reciprocal exchange of chromosome segments between homologs (non-sister chromatids)

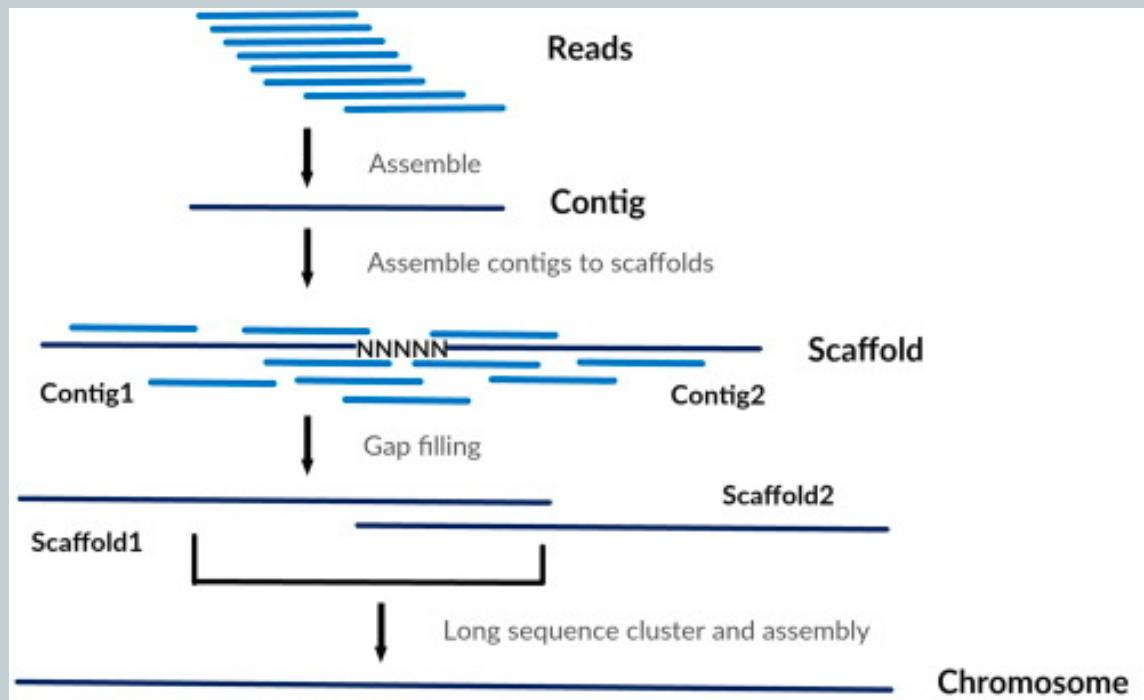
## **Linkage map**

- Called genetic map or chromosome map too.
- **Linkage map** – A map based on recombination frequencies between genetic markers during a crossover/recombination of homologous chromosomes during meiosis.
- **Recombination frequency/crossover rate/genetic distance (cM)** is a measure of CO or recombination between two genes that indicates how much recombination is observed in a particular experiment.
- The greater the frequency of recombination between two markers, the farther they are apart.
- **The map distance or genetic distance (cM)** between two genes equals one half the average number of crossovers in that region per meiotic cell

## **Applications/Uses of Linkage Maps:**

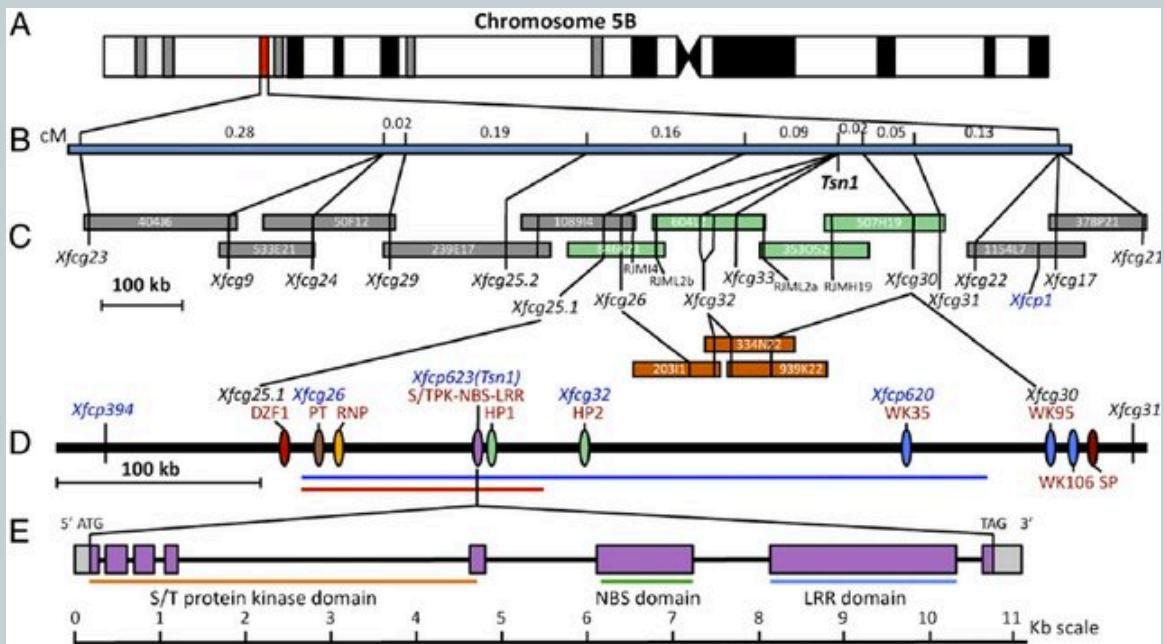
- Genome Assembly
- Map based cloning
- Studying genome structure, organization and evolution – Synteny.
- QTL analysis
  - Identify genes responsible for traits of interest
  - Estimation of gene effects of important agronomic traits
  - Tagging genes of interest to facilitate marker assisted selection (MAS) programs.

## Genome assembly



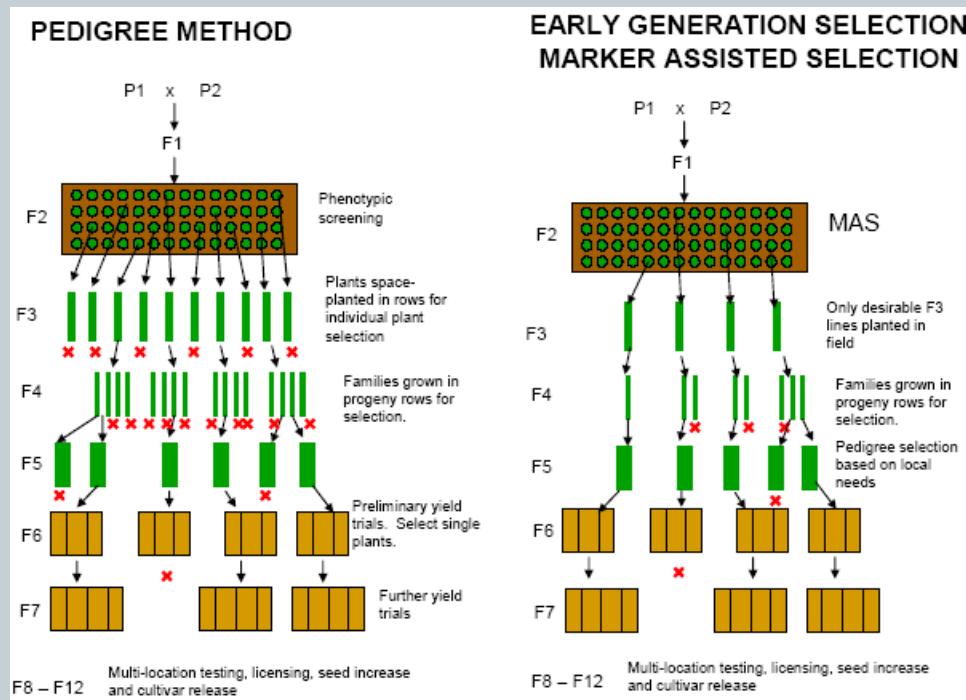
Guo2017: Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis

## Map based cloning :



Faris et.al 2010: A unique wheat disease resistance-like gene confers toxin-induced susceptibility to necrotrophic pathogens

# Marker assisted selection

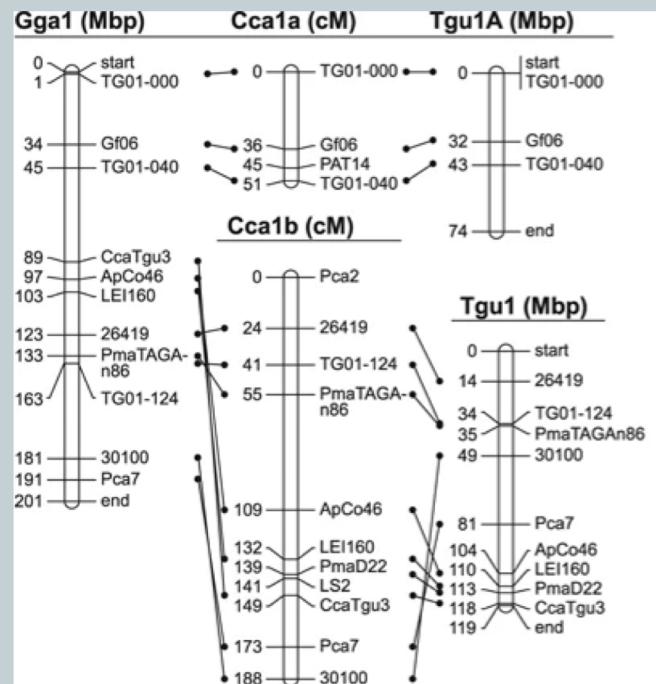


[http://www.knowledgebank.irri.org/ricebreedingcourse/Marker\\_assisted\\_breeding.htm](http://www.knowledgebank.irri.org/ricebreedingcourse/Marker_assisted_breeding.htm)

# Genome evolution

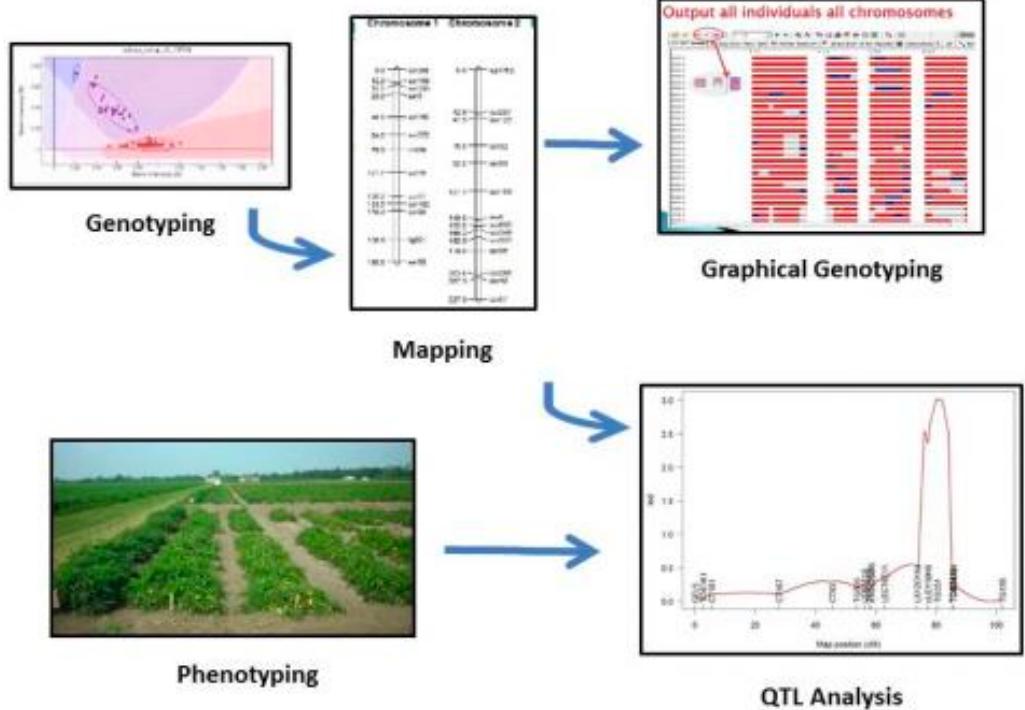
Blue tit (*Cyanistes caeruleus*; *Cca*) in cM

Orthologs loci  
Chicken *Gga* ) and zebra finch genomes  
(*Tgu* in Mbp).



Hannson et al. 2010. Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*)

# QTL analysis



<https://plant-breeding-genomics.extension.org/genetic-mapping-and-qtl-analysis/>

## **Steps in Genetic Linkage Mapping**

- Development of The Mapping Population
- Genotyping of Mapping Population (Molecular Markers)
- Map Construction (Linkage Map)

# I) Mapping Populations

Parents:

- Contrasting for trait of interest.

Develop a Segregating population

For highly heterozygous crops, we can map in **F<sub>1</sub> populations**

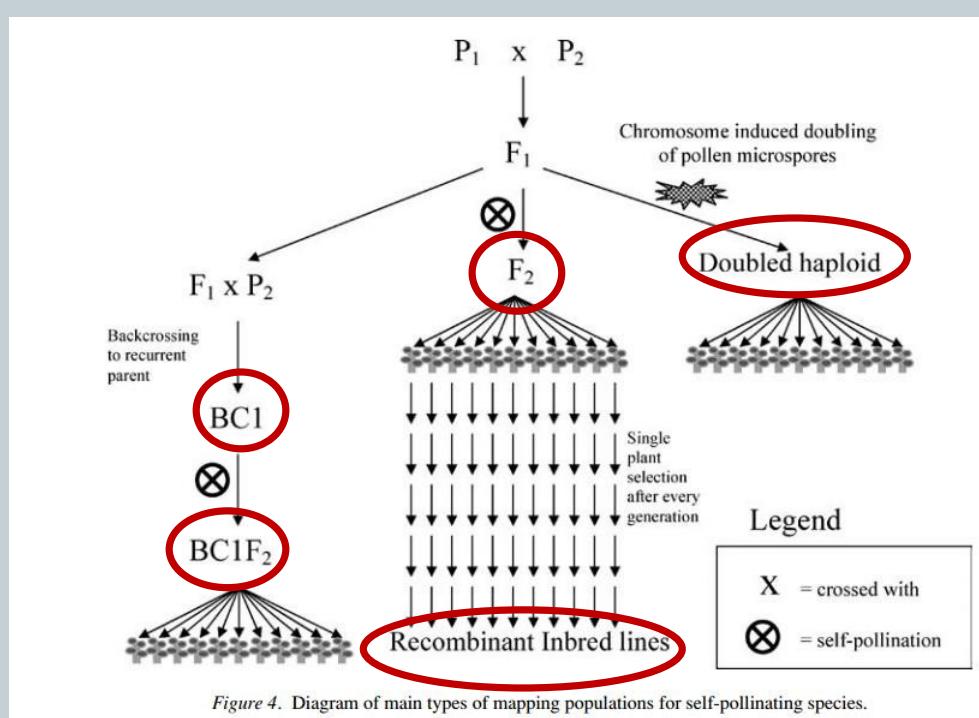


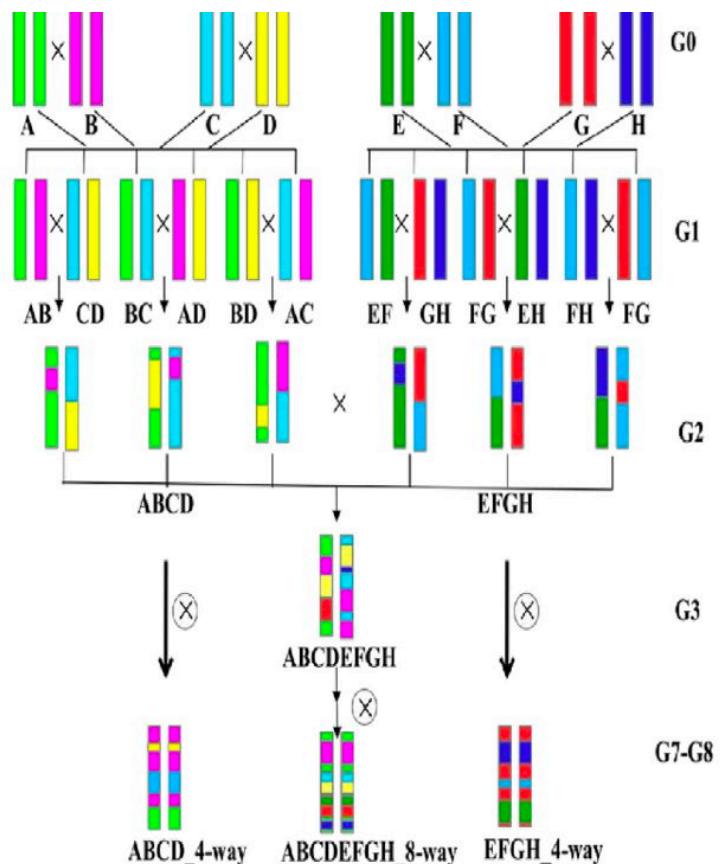
Figure 4. Diagram of main types of mapping populations for self-pollinating species.

Collard et al. 2004: An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts

# MAGIC Populations

4-way and 8-way populations

Meng et al. 2016 - Characterization of Three Rice Multiparent Advanced Generation Intercross (MAGIC) Populations for Quantitative Trait Loci Identification



## POLYPLOIDS

- Possess more than two complete sets of chromosome
- **Auto polyploids** - Arise within a species
  - Homozygous at every locus in the genome
- **Allo polyploids**- Arise due to the hybridization of two distinct species
  - Varying degrees of heterozygosity
  - Depends on the divergence of the parental genomes
- **Advantages**
  - Hybrid vigor and better adaptation
  - Gene redundancy

- Canola
- Grape
- Switchgrass
- Wheat
- Tobacco
- Banana
- Strawberry
- Sugarcane
- Alfa-alfa
- Potato

<https://www.nature.com/scitable/topicpage/polyploidy-1552814/>

- Parents (or at least one of them) carry several alleles
- segregate in the F<sub>1</sub> progeny.

Parent genotypes	Expected gamete ratios		Expected progeny phenotype ratios*	
	Disomic-digenic	Tetrasomic	Disomic-digenic	Tetrasomic
a <sub>1</sub> a <sub>1</sub> a <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> a <sub>1</sub> a <sub>2</sub> a <sub>2</sub>	(aa) × (1Aa:1aa)	(aa) × (1Aa:1aa)	1:1	1:1
a <sub>1</sub> a <sub>1</sub> a <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> A <sub>1</sub> a <sub>2</sub> a <sub>2</sub>	(aa) × (Aa)	(aa) × (1AA:4Aa:1aa)	1:0	5:1
a <sub>1</sub> a <sub>1</sub> a <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> a <sub>1</sub> A <sub>2</sub> a <sub>2</sub>	(aa) × (1AA:2Aa:1aa)	(aa) × (1AA:4Aa:1aa)	3:1	5:1
a <sub>1</sub> a <sub>1</sub> a <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> a <sub>1</sub> A <sub>2</sub> A <sub>2</sub>	(aa) × (1Aa:1Aa)	(aa) × (1Aa:1aa)	1:1	1:1
a <sub>1</sub> a <sub>1</sub> a <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> A <sub>1</sub> A <sub>2</sub> A <sub>2</sub>	(aa) × (AA)	(aa) × (AA)	1:0	1:0
a <sub>1</sub> a <sub>1</sub> A <sub>2</sub> a <sub>2</sub> × a <sub>1</sub> a <sub>1</sub> A <sub>2</sub> a <sub>2</sub>	(1aa:1Aa) × (1aa:1Aa)	(1aa:1Aa) × (1aa:1Aa)	3:1	3:1
a <sub>1</sub> a <sub>1</sub> A <sub>2</sub> A <sub>2</sub> × a <sub>1</sub> a <sub>1</sub> A <sub>2</sub> A <sub>2</sub>	(Aa) × (Aa)	(1AA:4Aa:1aa) × (1A:4Aa:1aa)	1:0	35:1
A <sub>1</sub> a <sub>1</sub> A <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> a <sub>1</sub> A <sub>2</sub> a <sub>2</sub>	(1AA:2Aa:1aa) × (1AA:2Aa:1aa)	(1AA:4Aa:1aa) × (1AA:4Aa:1aa)	15:1	35:1
A <sub>1</sub> A <sub>1</sub> A <sub>2</sub> a <sub>2</sub> × A <sub>1</sub> A <sub>1</sub> A <sub>2</sub> a <sub>2</sub>	(1AA:1Aa) × (1AA:1Aa)	(1AA:1Aa) × (1AA:1aa)	3:1	3:1

\* Assuming complete dominance of allele A.

## Disomic in Allotetraploids

## Tetrasomic in Auto-tetraploids

Ward 2014: Allotetraploid segregation for single-gene morphological characters in quinoa (*Chenopodium quinoa* Willd.)

## 2) Markers

Dominant – 2 states

Co-dominant – 3 states

- SSR
- SNP
  - Illumina chips
  - GBS
  - NGS
- Indels
- kmers

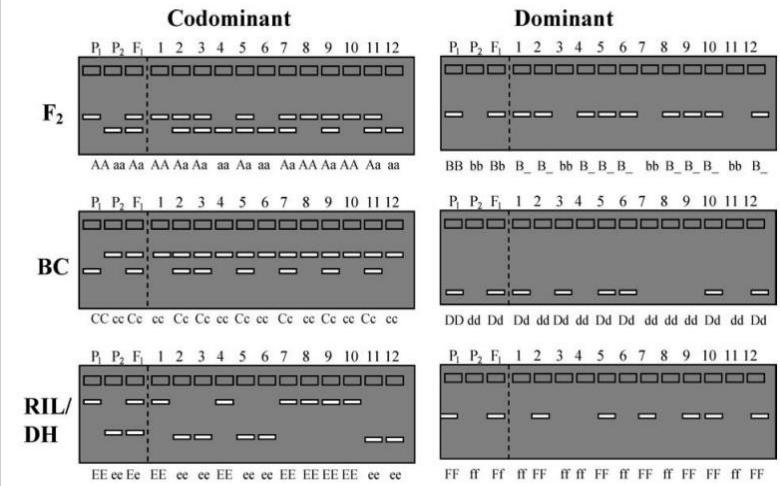


Figure 5. Hypothetical gel photos representing segregating codominant markers (left-hand side) and dominant markers (right-hand side) for typical mapping populations. Codominant markers indicate the complete genotype of a plant. Note that dominant markers cannot discriminate between heterozygotes and one homozygote genotype in F<sub>2</sub> populations. The segregation ratios of markers can be easily understood by using Punnett squares to derive population genotypes.

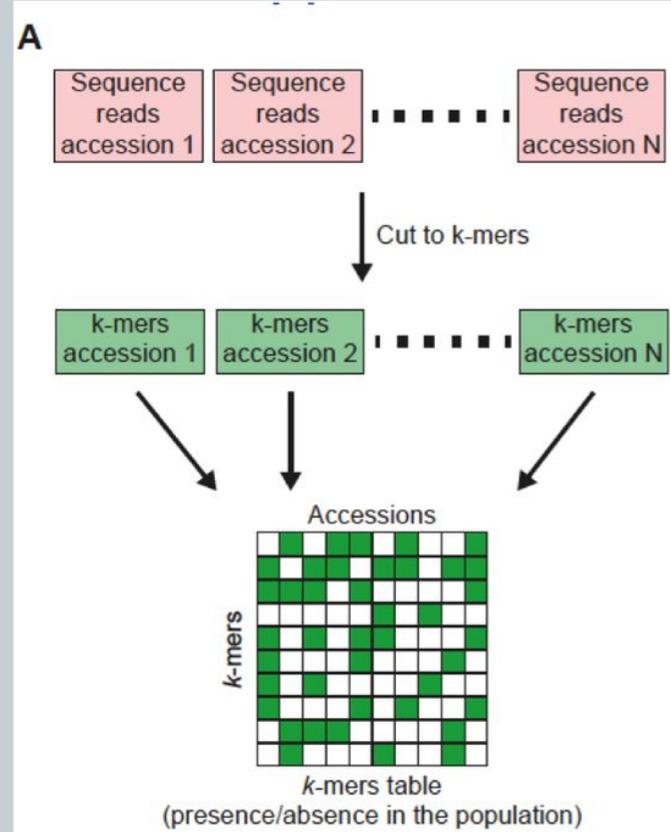
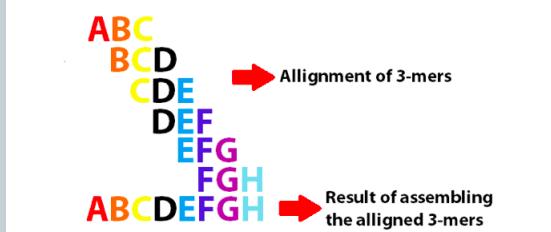
### Segregating Ratios

I:1 RILs

I:2:I – F2

# Kmers

Short fragment of sequences



<https://www.biorxiv.org/content/10.1101/818096v1.full>

## Polyploids

Type	Parent 1	Parent 2
Null (N)	AAAA	BBBB
Simplex (S)	AAAA	AAAB
Simplex (S)	ABBB	BBBB
Triplex (T)	AAAA	ABBB
Triplex (T)	AAAB	BBBB
Duplex (D)	AAAA	AABB
Duplex (D)	AABB	BBBB
Double-simplex (SS)	AAAB	AAAB
Double-simplex (SS)	ABBB	ABBB
X-double-simplex (XSS)	AAAB	ABBB
Simplex-duplex (SD)	AAAB	AABB
Duplex-simplex (DS)	AABB	ABBB
Double-duplex (DD)	AABB	AABB
Total		

Genotype classes ; [Ferreira](#) et al. 2019. Genetic Mapping With Allele Dosage Information in Tetraploid *Urochloa decumbens* (Stapf) R. D. Webster Reveals Insights Into Spittlebug (*Notozulia enteriana* Berg) Resistance

## Software for polyploid linkage map

- polymapR
- TetraploidSNPMap
- netgwas

## Autotetraploid Segregation

Cross	Nulliplex (aaaa)				
Nulliplex (aaaa)	All N	Simplex (Aaaa)			
Simplex (Aaaa)	1S:1N	1D:2S:1N	Duplex (AAaa)		
Duplex (AAaa)	1D:4S:1N	1T:5D:5S: 1N	1Q:8T:18D: 8S:1N	Triplex (AAAa)	
Triplex (AAAa)	1D:1S	1T:2D:1S	1Q:5T:5D: 1N	1Q:2T:1D	Quad'plex (AAAA)
Quad'plex (AAAA)	All D	1T:1D	1Q:4T:1D	1Q:1T	All Q

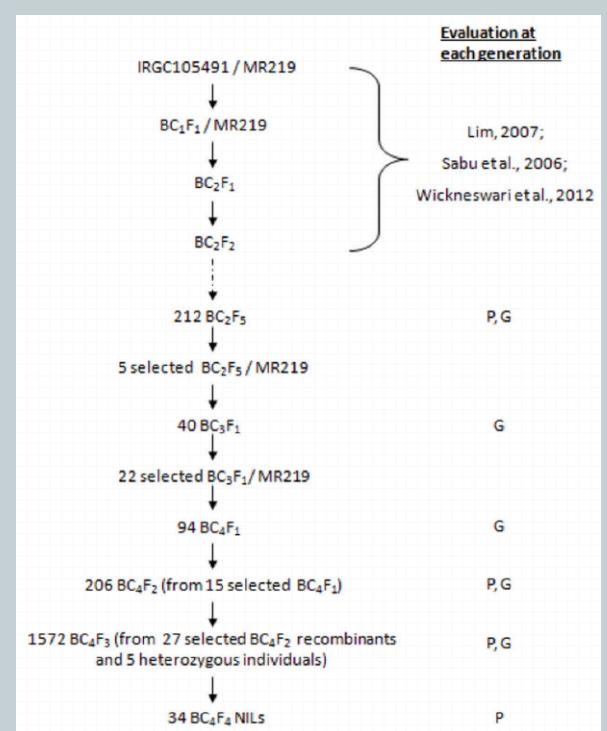
[https://www.webpages.uidaho.edu/jbrown/plsc546/class\\_notes/class-15-Qualitative%20Genetics%20III-2008.pdf](https://www.webpages.uidaho.edu/jbrown/plsc546/class_notes/class-15-Qualitative%20Genetics%20III-2008.pdf)

Bourke tal . 2018: Tools for Genetic Studies in Experimental Populations of Polyploids

## Map saturation

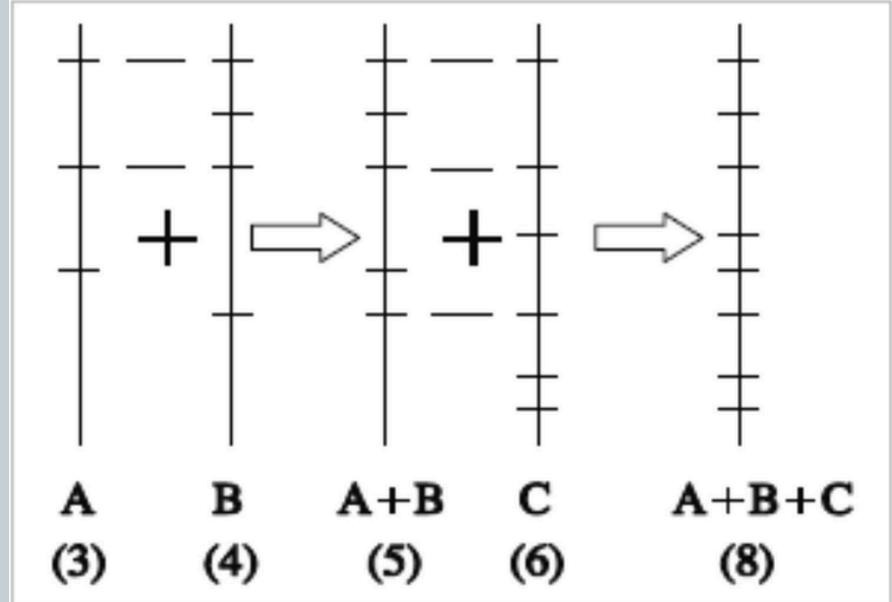
Ex: A region of interest is about 5cM with about 200 genes. To narrow the interval

- i. **Nearly isogenic lines (NIL)**: for targeting regions with a gene of interest
- a) Backcrossing a particular gene into desirable cultivar e.g., Williams, - cultivar in soybean
- Williams 79, and Williams 82 soybeans-- phytophthora genes introgressed)
- b) NILs have homogeneous genetic background, except in the location of the gene.



## Map Merging

- Lpmerge
- ASMap – R
- MergeMap



Boopathi 2020. Linkage Map Construction [in Genetic Mapping and Marker Assisted Selection](#) pp 179-227

## **PRACTICAL ASPECTS**

## Data cleaning

- Parents:
  - Polymorphic homozygous
  - Parent 1 is A allele, Parent 2 is B allele (always)
- Remove Duplicate Marker Names
- Remove Missing
- Check for ratios – Chi square test
- No parents data for mapping software

## **INPUT – CARTHAGENE/MAPMAKER FORMAT**

The very first line of your raw data file

**data type xxxx**

where xxxx is one of the allowed data types, either:

f2 intercross

f2 backcross

f3 self

ri self

ri sib

The second line of the raw file should contain a **list of three numbers, separated by spaces**, such as:

**46 362 0**

**The first value** - The number of progeny (46) for which data are included in the file

**The second value** - Number of genetic loci for which data are supplied (362).

**The third** indicates the number of quantitative traits in the data set (0)

Additional information may be optionally supplied at the end of second line

Ex: **46 362 2 symbols I=A 2=B 0=-**

After the first two header lines,

For each locus, you list

- (1) the name of the locus, preceded by an asterisk ("\*");
- (2) one or more spaces (or tabs); and
- (3) the genotypic data for all individuals, in order.

example:

**\*m234 AB-ABBABABABABA---BABABABA**

## **Carthagene useful commands**

help - to get all commands

Type '<command-name> -H' for more help with a particular command.

To change the defaults, edit the '.carthagenerc'

### **Manipulate data sets :**

dsload Load a Biological Data Set.

dsinfo Summarize the current data sets.

dsmergen Merge two Biological Data Sets.

dsget Get a list of the current data sets description.

### **Manipulate groups :**

group Identify linkage groups.

groupget Get a group by markers ID.

- **Manipulate loci :**

- mrkdoubleget Retrieve a list of pairs of markers with compatible typing.
- mrklod2p Print the two points LOD matrix.
- mrkselget Return the loci selection into a list.
- mrkselset Set the marker list(by Id).
- mrkdist2p Print the two points distance matrix.
- mrkmerge Merges two compatible markers.
- mrkdouble Identifies pairs of markers with compatible typing.
- mrkmerget Retrieve a list of merged markers.

## **Search for good maps :**

- **greedy** Find good maps using the greedy algorithm.
- **buildfw** Build a framework map.
- **nicemapd** Provide quickly a (nice) map, using the 2-points distances.
- **annealing** Find good maps using the annealing algorithm.
- **build** Build maps with two-points information
- **Verify maps**
  - **flips** Try to improve the best map by flipping.
  - **polish** Try to improve the best map by polishing.
- **Manipulate the heap :**
  - **heaprintd** Display the heap in detail, sorted.
  - **heapprint** Display the heap sorted.
  - **bestprint** Print the best map in the heap.
  - **bestprintd** Print the best map in the heap, in detail.

### **To manipulate a map :**

- **maprint** Print a map.
- **maprintdr** Print a map reverse, in detail.
- **mapget** Get a map, in a list.
- **maprintd** Print a map, in detail.

### **• Graphical commands :**

- **maprintg** Draw a map into a graphical display.

### **• Miscellaneous :**

- **cgresstart** Reset the application.
- **cgsstop** To stop a running command, type ctrl-c.
- **cgsave** Save the state of the current session.
- **clexport** Save the state of the current session.

## **Carthagene Main steps**

Load - select the file

Data/info - To get information on data file

Marker merge - optional

Loci/config

Loci/Identify Groups

**For each linkage group (>3 markers) - REPEAT for each group**

Loci/select a group

Loci/Dist2pt (if duplicates- merge 2 locus)

Buildfw & Detail

\* Other algorithms - ALGORITHM & Detail

flips & Detail

polish & Detail

**“bestprint”; “bestprintd”**

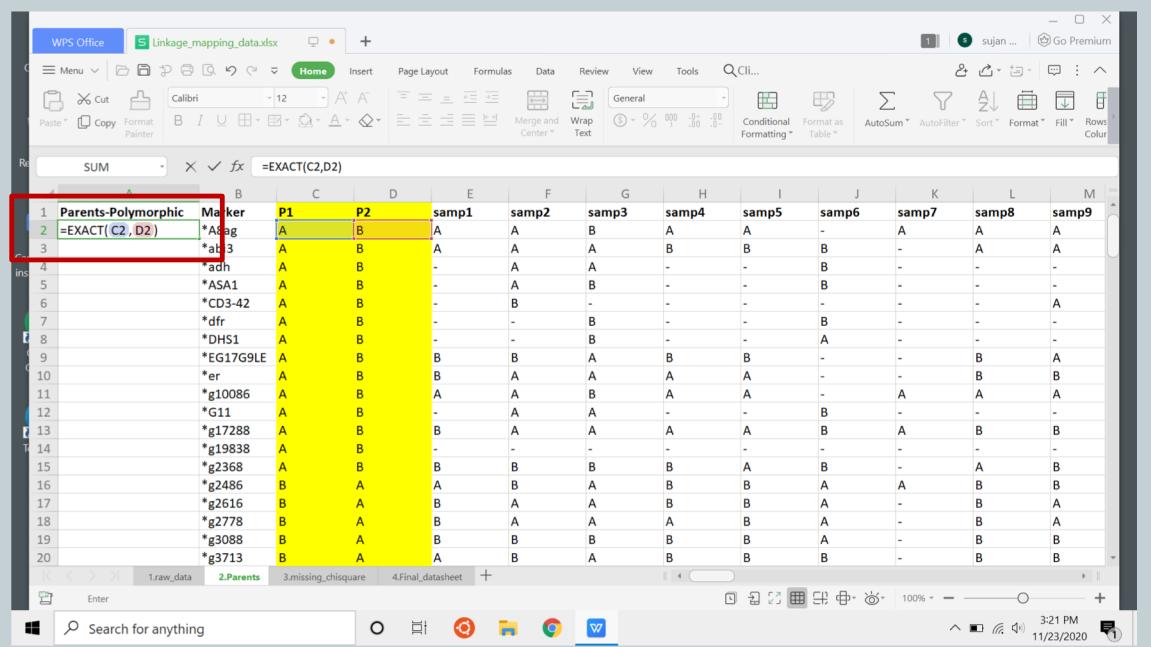
**Copy into a file**

## Parents highlight

# DATA CLEANING

Marker	raw_data	P1	samp1	samp2	samp3	samp4	samp5	samp6	samp7	samp8	samp9	samp10	samp11
1 *A8ag	A	B	A	A	B	A	A	-	A	A	A	A	B
2 *abi3	A	B	A	A	A	B	B	B	-	A	A	A	B
3 *adhl	A	B	-	A	A	-	-	B	-	-	-	-	-
4 *ASA1	A	B	-	A	B	-	-	B	-	-	-	-	-
5 *CD3-42	A	B	-	B	-	-	-	-	-	-	-	-	-
6 *dfr	A	B	-	-	B	-	-	B	-	-	-	-	B
7 *DHS1	A	B	-	-	B	-	-	A	-	-	-	-	-
8 *EG17G9LB	A	B	B	A	B	B	-	-	B	A	B	A	A
9 *er	A	B	B	A	A	A	A	-	B	B	B	A	B
10 *g10086	A	B	A	A	B	A	A	-	A	A	A	A	B
11 *G11	A	B	-	A	A	-	-	B	-	-	-	-	-
12 *g17288	A	B	B	A	A	A	A	B	-	-	-	-	-
13 *g19838	A	B	-	-	-	-	-	-	-	-	-	-	-
14 *g2368	A	B	B	B	B	B	A	B	-	A	B	A	B
15 *g2486	B	A	A	B	A	B	B	A	A	B	B	A	A
16 *g2616	B	A	B	A	A	B	B	A	-	B	A	B	A
17 *g2778	B	A	B	A	A	A	B	A	-	B	A	B	A
18 *g3088	B	A	B	B	B	B	B	A	-	B	B	A	A
19 *g3713	B	A	A	B	A	B	B	B	-	B	B	A	A

## Parents - Polymorphic



Parents-Polymorphic	Marker	P1	P2	samp1	samp2	samp3	samp4	samp5	samp6	samp7	samp8	samp9
1		A	B	A	A	B	A	A	-	A	A	A
2	=EXACT(C2,D2)	*A\$tag	A	B	A	A	A	B	B	-	A	A
3	*ab3	A	B	A	A	A	-	-	B	-	-	-
4	*adh	A	B	-	A	A	-	-	B	-	-	-
5	*ASA1	A	B	-	A	B	-	-	B	-	-	-
6	*CD3-42	A	B	-	B	-	-	-	-	-	-	A
7	*dfr	A	B	-	-	B	-	-	B	-	-	-
8	*DHS1	A	B	-	-	B	-	-	A	-	-	-
9	*EG17G9LE	A	B	B	B	A	B	B	-	-	B	A
10	*er	A	B	B	A	A	A	A	-	-	B	B
11	*g10086	A	B	A	A	B	A	A	-	A	A	A
12	*G11	A	B	-	A	A	-	-	B	-	-	-
13	*g17288	A	B	B	A	A	A	A	B	A	B	B
14	*g19838	A	B	-	-	-	-	-	-	-	-	-
15	*g2368	A	B	B	B	B	A	B	-	A	B	B
16	*g2486	B	A	A	B	A	B	B	A	A	B	B
17	*g2616	B	A	B	A	A	B	B	A	-	B	A
18	*g2778	B	A	B	A	A	A	B	A	-	B	A
19	*g3088	B	A	B	B	B	B	B	A	-	B	B
20	*g3713	B	A	A	B	A	B	B	B	-	B	B

## **Parents - PI - A; P2 - B**

The screenshot shows a WPS Office spreadsheet application window. The title bar reads "WPS Office" and "Linkage\_mapping\_data.xlsx". The ribbon menu includes Home, Insert, Page Layout, Formulas, Data, Review, View, Tools, and a search bar. The Home tab is selected, displaying various formatting tools like Calibri, font size 12, bold, italic, etc., and a toolbar with Paste, Cut, Copy, Format Painter, Merge & Center, Wrap Text, Conditional Formatting, AutoSum, AutoFilter, Sort, Format, Fill, and Rows Colur.

The main worksheet area contains a large table with columns labeled A through Z and rows numbered 111 to 145. A red box highlights a specific range of cells from row 111 to 120, columns C to Z. The data consists of mostly 'A' characters, with some variations in the first few columns. Below the table, there are tabs for "1.raw\_data", "2.Parents", "3.missing\_chisquare", and "4.Final\_datasheet".

At the bottom, a status bar shows "Sum=0 Average=0 Count=312", system icons, and the time "3:32 PM" and date "11/23/2020".

# Duplicate markers?

The screenshot shows a Microsoft Excel spreadsheet titled "Linkage\_mapping\_data.xlsx". The formula bar at the top displays the formula `=EXACT(B2,B1)`. A red box highlights this formula. The spreadsheet contains a large dataset with columns labeled P1 through R and rows numbered 1 to 79. The first few rows of data are as follows:

	A	B	P1	P2	samp1	samp2	samp3	samp4	samp5	samp6	samp7	samp8	samp9	samp10	samp11	samp12	samp13	samp14
1	Marker			B	A	A	B	A	-	A	A	A	A	B	B	B	A	
2				B	A	A	A	B	B	-	A	A	A	B	B	B	A	
3				B	-	A	A	-	-	B	-	-	-	-	-	-	-	
4				B	-	A	B	-	-	B	-	-	-	-	-	-	-	
5				B	-	B	-	-	-	-	-	-	A	-	B	-	B	
6				B	-	B	-	-	-	B	-	-	-	-	-	-	B	
7				B	-	-	B	-	-	B	-	-	-	-	-	-	-	
8				B	-	-	B	-	-	B	-	-	-	-	-	-	-	
9				B	B	B	A	B	A	B	-	B	A	B	A	B	A	
10				B	B	A	A	A	A	A	-	B	A	B	A	B	B	
11				B	A	A	B	A	A	A	-	A	A	A	B	B	B	
12				B	-	A	A	-	-	B	-	-	-	-	-	-	-	
13				B	B	A	A	A	A	B	A	B	B	A	B	A	A	
14				B	-	-	-	-	-	-	-	-	-	-	-	-	-	
15				B	B	B	B	B	A	B	-	A	B	A	B	B	B	
16				B	B	A	B	A	A	B	-	A	B	A	B	B	B	
17				B	A	B	B	A	B	B	-	A	B	A	B	B	-	
18				B	A	B	B	B	A	B	-	A	B	A	B	B	A	
19				B	A	A	A	A	A	B	-	A	A	B	B	A	A	
20				B	B	B	B	B	A	A	-	A	A	B	B	A	A	
21				B	A	A	B	A	B	B	-	B	B	A	B	A	B	
22				B	B	B	B	B	B	A	-	B	A	B	A	B	B	
23				B	B	B	B	B	B	A	-	B	A	B	A	B	B	
24				B	B	B	B	B	B	A	-	B	B	A	B	A	B	
25				B	A	A	B	B	B	B	-	B	B	A	B	A	B	
26				B	A	A	B	A	A	B	-	A	B	B	A	B	B	
27				B	B	-	-	-	-	-	-	-	-	-	-	-	-	
28				B	B	A	B	-	-	-	-	B	A	B	B	B	A	
29				B	R	A	A	A	A	R	R	R	R	A	A	R	A	

## Counts - Alleles, Missing

The screenshot shows a WPS Office spreadsheet window. The title bar indicates the file is 'Linkage\_mapping\_data.xlsx'. The formula bar shows the formula `=FLOOR(((E2+F2)/2),1)` entered into cell C2. The spreadsheet contains several columns: Notes, Chi-sq, expected-A, expected-B, A-allele, B-Allele, Missing-alleles, Parents-miss, Marker, P1, P2, samp1, samp2, samp3, samp4, samp5, samp6, samp7. The first row has formulas in the last four columns. The second row has values: 47, 54, 40, 7. The rest of the rows contain marker names and sample data. The bottom of the screen shows the Windows taskbar with the search bar containing 'Search for anything' and the date/time '11/23/2020 3:43 PM'.

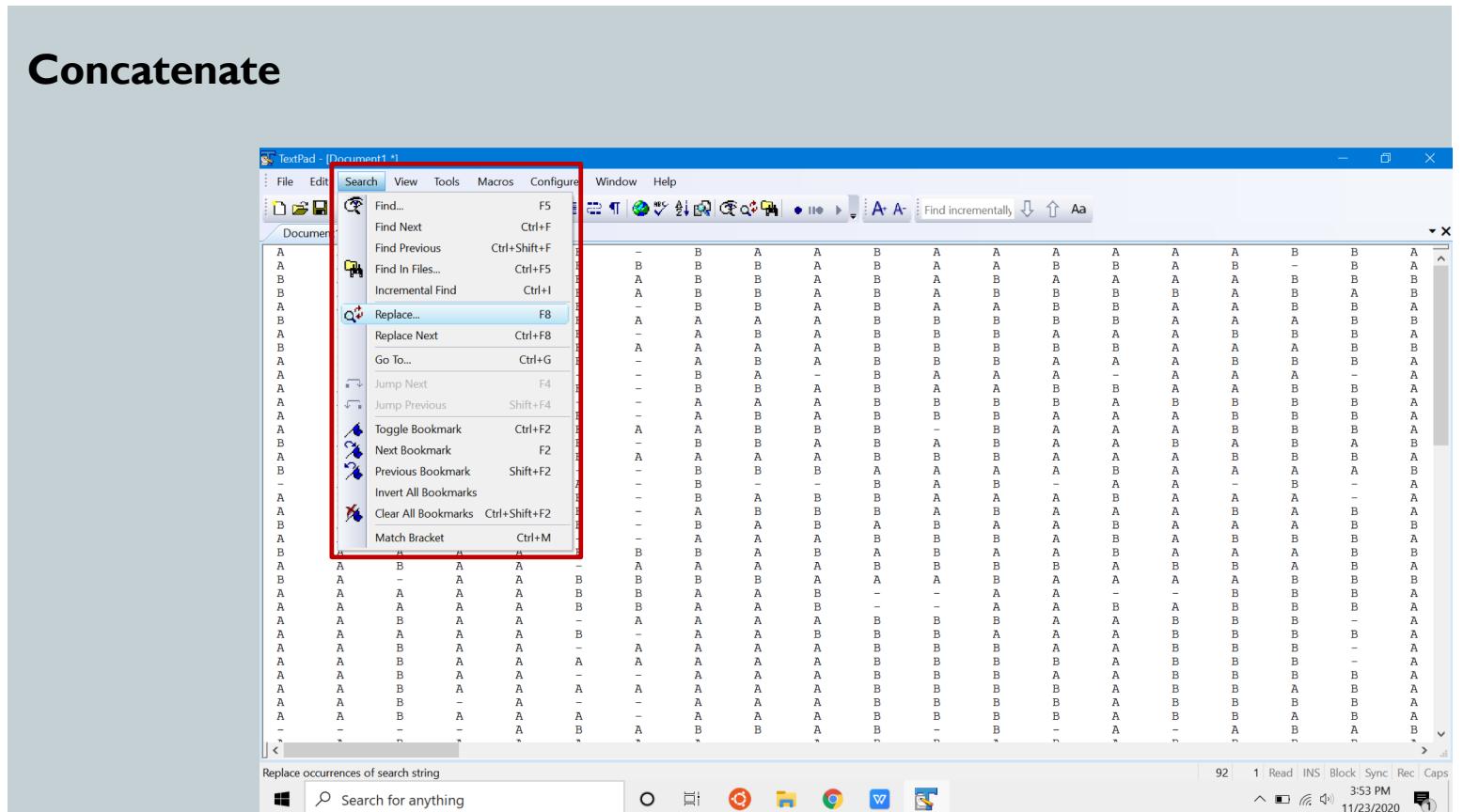
Notes	Chi-sq	expected-A	expected-B	A-allele	B-Allele	Missing-alleles	Parents-miss	Marker	P1	P2	samp1	samp2	samp3	samp4	samp5	samp6	samp7
				54	40	7		0 *A8ag	A	B	A	B	A	-	-	A	-
								0 *abi3	A	B	A	A	B	B	B	B	-
								0 *adn	A	B	-	A	A	-	-	B	-
								0 *ASA1	A	B	-	A	B	-	-	B	-
								0 *CD3-42	A	B	-	B	-	-	-	-	-
								0 *dfr	A	B	-	-	B	-	-	B	-
								0 *DHS1	A	B	-	-	B	-	-	A	-
								0 *EG17G9LE	A	B	B	A	B	B	-	-	-
								0 *er	A	B	B	A	A	A	-	-	-
								0 *g10086	A	B	A	A	B	A	A	-	A
								0 *G11	A	B	-	A	A	-	-	B	-
								0 *g17288	A	B	B	A	A	A	A	B	A
								0 *g19838	A	B	-	-	-	-	-	-	-
								0 *g2368	A	B	B	B	B	A	B	-	-
								0 *g2486	A	B	B	A	B	A	A	B	B
								0 *g2616	A	B	A	B	B	A	A	B	-
								0 *g2778	A	B	A	B	B	B	A	B	-
								0 *g3088	A	B	A	A	A	A	A	B	-
								0 *g3713	A	B	B	A	B	A	A	A	-
								0 *g3715	A	B	A	A	B	A	B	B	-
								0 *g3786	A	B	B	B	B	B	B	A	-
								0 *g3829	A	B	B	B	B	B	B	A	-
								0 *g3837	A	B	A	A	B	B	B	B	-
								0 *g3843	A	B	A	B	B	A	A	B	-
								0 *g3845	A	B	A	A	B	A	A	B	-
								0 *g3883	A	B	-	-	-	-	-	-	-
								0 *g4014	A	B	A	B	-	-	-	B	A
								0 *g4076	A	B	A	A	A	B	R	R	R

## Chisq test

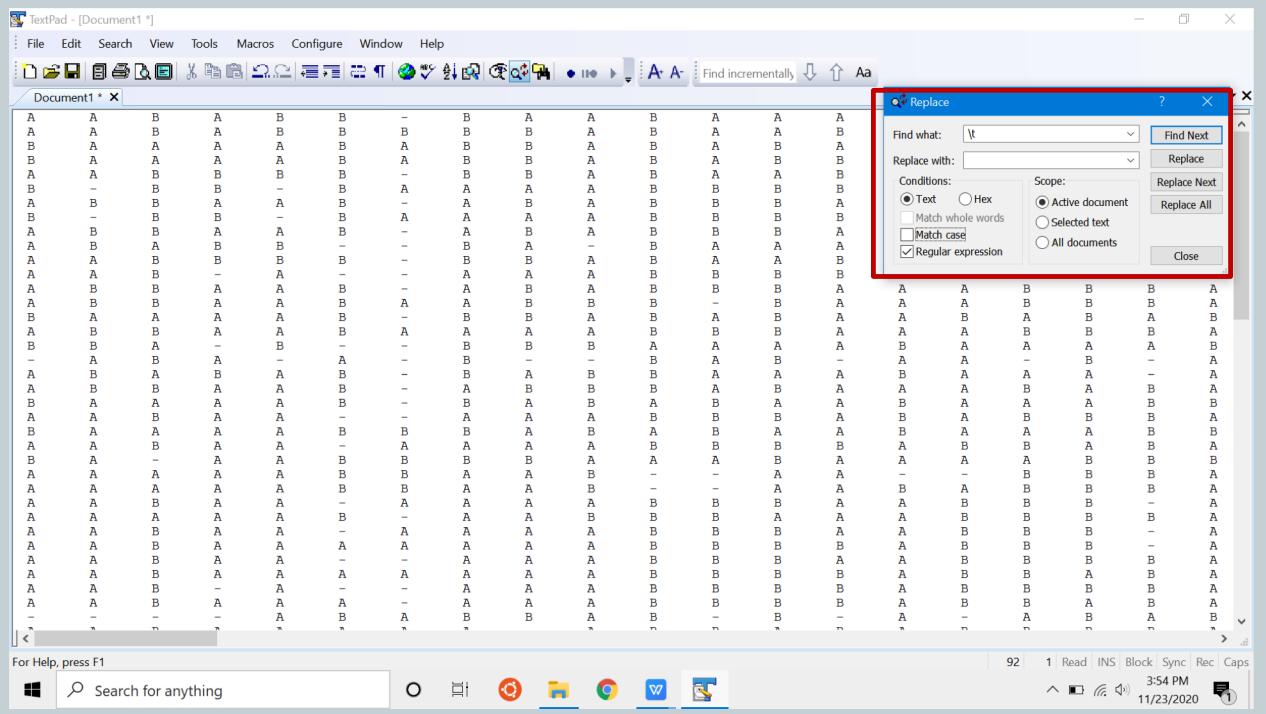
The screenshot shows a WPS Office spreadsheet titled "Linkage\_mapping\_data.xlsx". The formula bar at the top displays the formula `=CHITEST(E2:F2,c2:d2)`. A red box highlights the range E2:F2, which contains the value 47. The formula bar also shows the full formula `=CHITEST(E2:F2,c2:d2)`.

Notes	Chi-sq	expected-A	expected-B	A-allele	B-allele	Missing-alleles	Parents-miss	Marker	P1	P2	samp1	samp2	samp3	samp4	samp5	samp6	samp7
							O	*A8ag	A	B	A	B	A	A	-	-	
							O	*abi3	A	B	A	A	B	B	B	B	-
							O	*adh	A	B	-	A	-	-	-	-	
							O	*ASA1	A	B	-	A	B	-	-	-	
							O	*CD3-42	A	B	-	B	-	-	-	-	
							O	*dfr	A	B	-	-	B	-	-	-	
							O	*DHS1	A	B	-	-	B	-	-	-	
							O	*EG17G0LE	A	B	B	A	B	B	-	-	
							O	*er	A	B	B	A	A	A	-	-	
							O	*g10086	A	B	A	A	B	A	-	-	A
							O	*G11	A	B	-	A	A	-	-	-	
							O	*g17288	A	B	B	A	A	A	B	A	
							O	*g19838	A	B	-	-	-	-	-	-	
							O	*g2368	A	B	B	B	B	A	B	-	
							O	*g2486	A	B	B	A	A	A	B	B	
							O	*g2616	A	B	A	B	B	A	B	-	
							O	*g2778	A	B	A	B	B	A	B	-	
							O	*g3088	A	B	A	A	A	A	B	-	
							O	*g3713	A	B	B	A	B	A	A	-	
							O	*g3715	A	B	A	A	B	A	B	-	
							O	*g3786	A	B	B	B	B	B	A	-	
							O	*g3829	A	B	B	B	B	B	A	-	
							O	*g3837	A	B	A	B	B	B	B	-	
							O	*g3843	A	B	A	B	A	A	B	-	
							O	*g3845	A	B	A	A	B	A	A	A	
							O	*g3883	A	B	-	-	-	-	-	-	
							O	*g4014	A	B	A	B	-	-	-	B	A
							O	*s4026	A	R	A	A	A	R	R	R	

# Concatenate



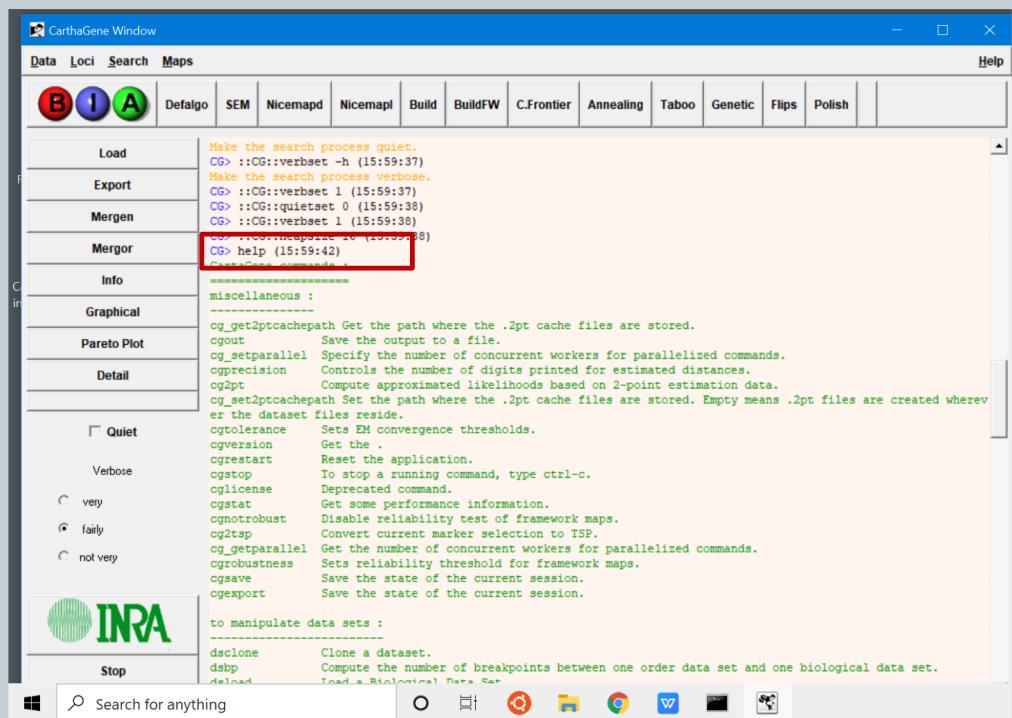
## Concatenate



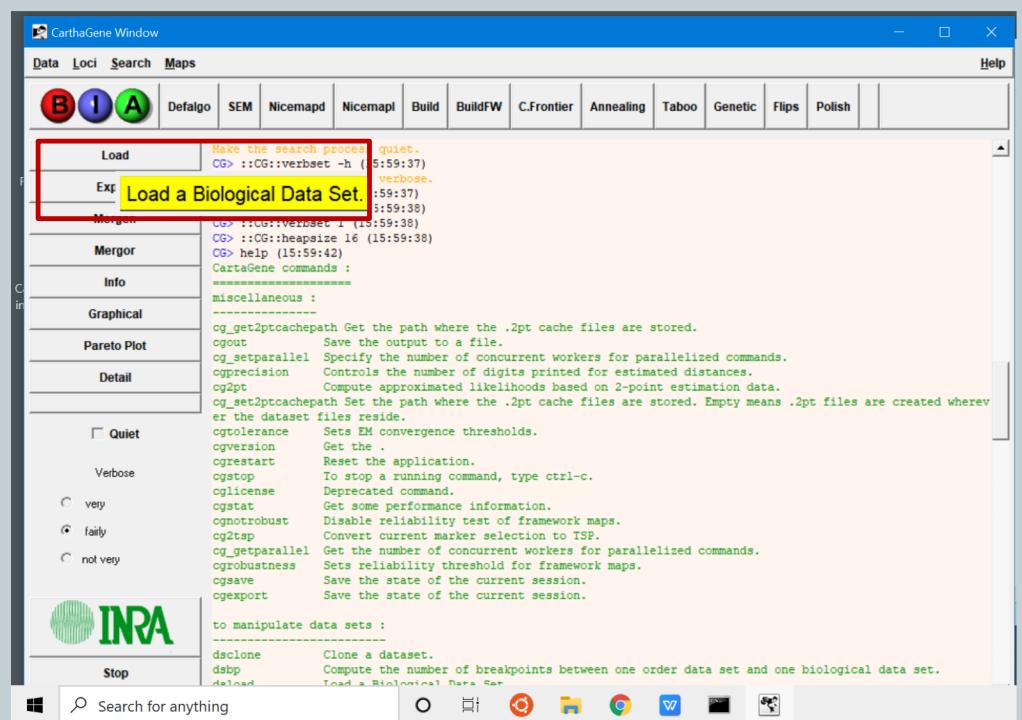
## Final Input file

# Carthagene

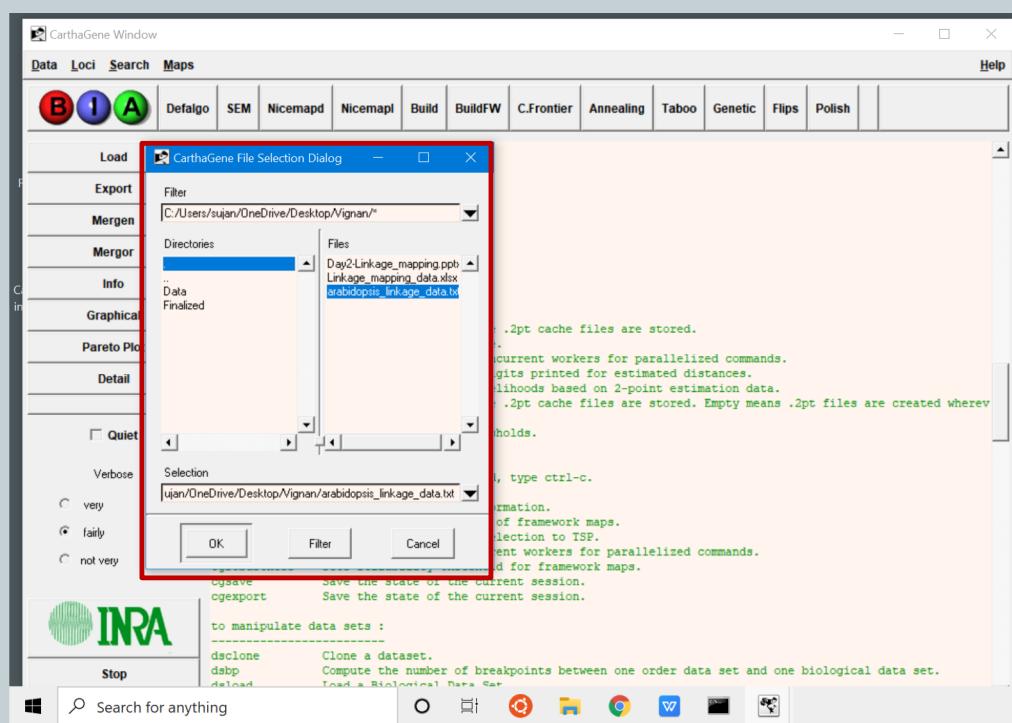
## Help command



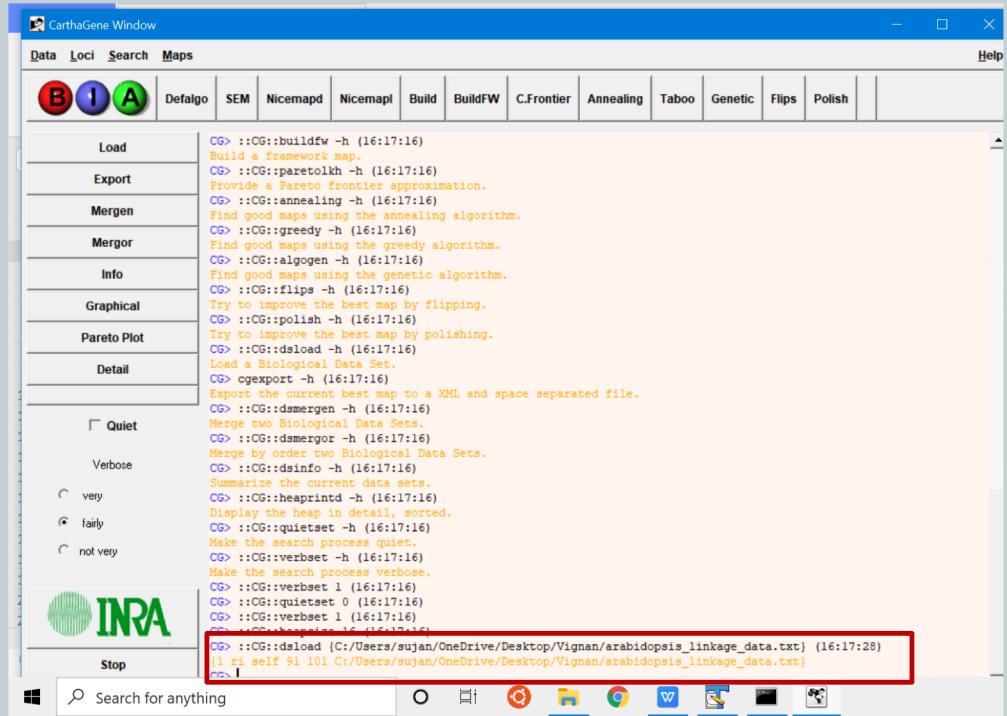
## Load Button



## select file - Path



## File loaded successfully



## Data/Info

The screenshot shows a WPS Office window titled "Linkage\_mapping\_data.xlsx". The main area displays a terminal-like interface for the CarthaGene software. The terminal output includes:

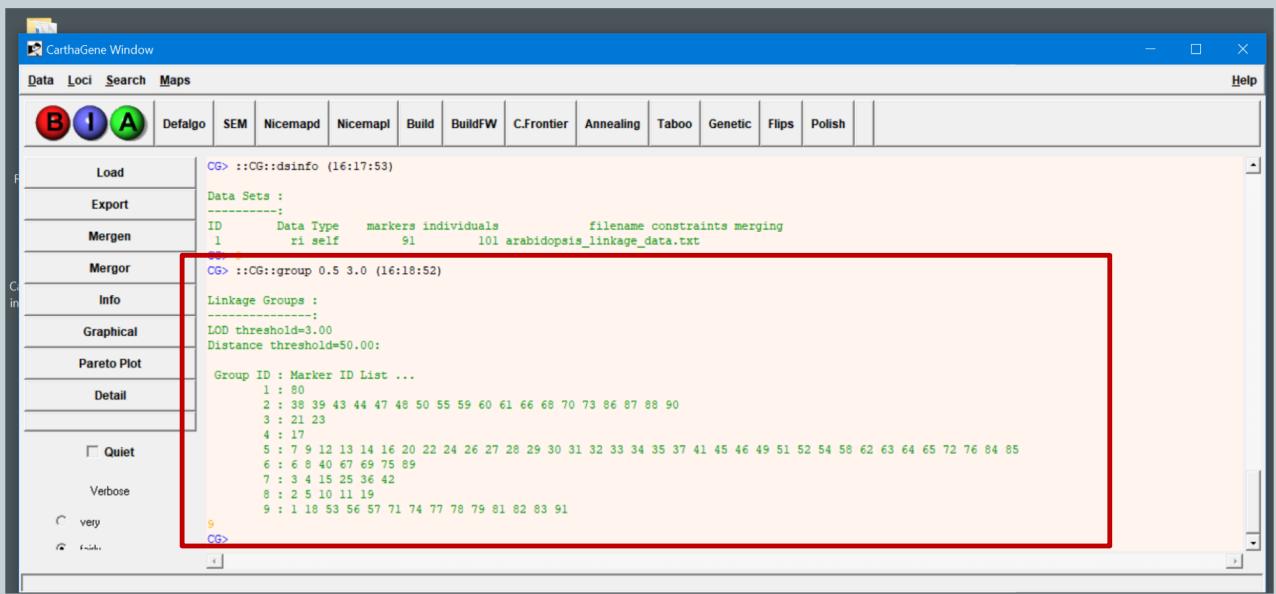
```
Export the current best map to a XML and space separated file.  
CG> ::CG::dmergen -h (16:17:16)  
Merge two Biological Data Sets.  
CG> ::CG::dmergor -h (16:17:16)  
Merge by order two Biological Data Sets.  
CG> ::CG::dsinfo -h (16:17:16)  
Summarize the current data sets.  
CG> ::CG::heaprintd -h (16:17:16)  
Display the heap in detail, sorted.  
CG> ::CG::quietset -h (16:17:16)  
Make the search process quiet.  
CG> ::CG::verbose -h (16:17:16)  
Make the search process verbose.  
CG> ::CG::verbose 1 (16:17:16)  
CG> ::CG::quietset 0 (16:17:16)  
CG> ::CG::verbose 1 (16:17:16)  
CG> ::CG::heapszie 16 (16:17:16)  
CG> ::CG::dsload [C:/Users/sujan/OneDrive/Desktop/Vignan/arabidopsis_linkage_data.txt] (16:17:28)  
[1..ri self 91 101 C:/Users/sujan/OneDrive/Desktop/Vignan/arabidopsis_linkage_data.txt]  
CG> ::CG::dsinfo (16:17:53)
```

The terminal interface also shows a "Data Sets :" section with a table:

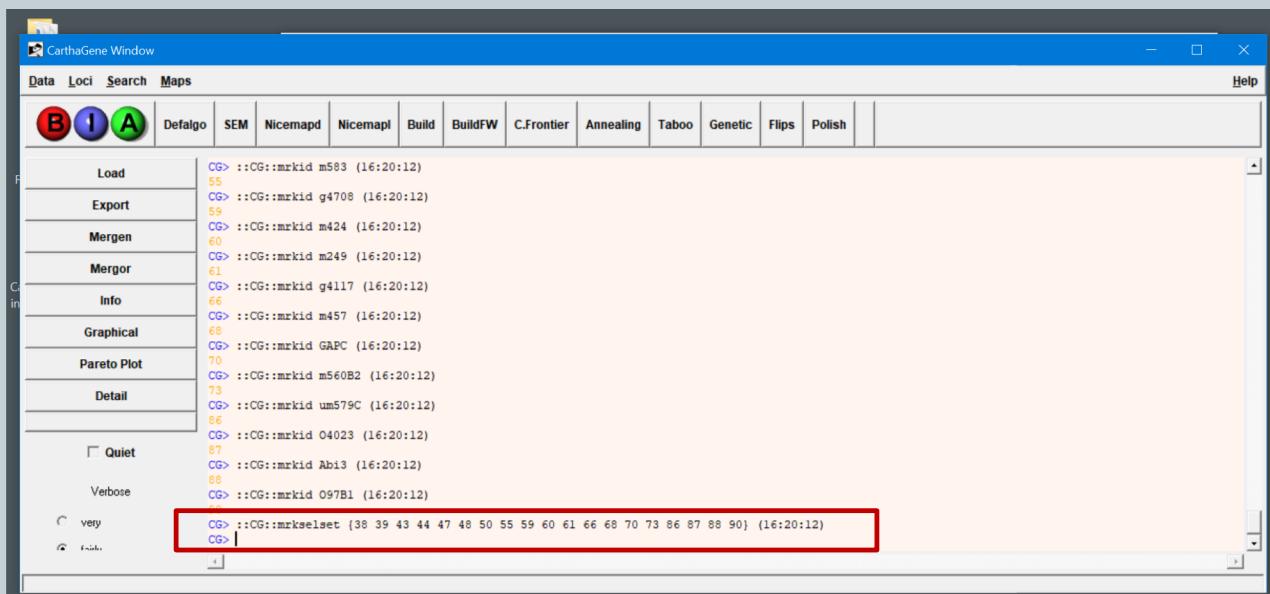
ID	Data Type	markers	individuals	filename	constraints	merging
1	ri self	91	101	arabidopsis_linkage_data.txt		

The WPS Office ribbon at the top includes tabs for Data, Loci, Search, and Maps. The status bar at the bottom shows the date and time: 4:18 PM 11/23/2020.

## Loci/Identify groups



## Loci/select a group



## Loci/Dist2pt

The screenshot shows the CarthaGene Window software interface. The menu bar includes Data, Loci, Search, Maps, and Help. The main window has a toolbar with icons for Identify groups, Select a group, Config, select by locus, merge two loci, Info, Lod2pt, Dist2pt (which is highlighted with a red box), Fr2pt, Graphical, Pareto Plot, and Detail. Below the toolbar is a status bar with checkboxes for Quiet, Verbose, and very (radio button selected). A text area displays the command "Print two points distance matrices of the loci selection :". The data set number is 1. The output table is as follows:

	g4014	m228	g2778	m105	g4564	m2	g4523	m583	g4708	m424	m249	g4117	m457	GAPC	m560B	um579	O4023	Abi13	O97B1
g4014	41.5	7.0	34.1	2.6	2.6	44.8	43.4	41.5	11.0	5.9	5.8	3.7	69.8	35.0	22.2	28.2	30.6	8.1	
m228	41.5	66.6	4.3	38.0	38.0	6.4	6.4	0.0	84.2	29.8	29.8	72.0	8.0	7.3	20.7	7.0	12.4	44.8	
g2778	7.0	66.6	48.6	12.2	12.2	72.5	85.3	57.8	4.1	11.5	11.5	5.4	172.7	43.9	24.8	38.3	35.2	12.3	
m105	34.1	4.3	48.6	25.4	25.4	11.2	11.8	4.0	49.7	24.1	23.7	45.3	16.4	1.7	19.4	1.3	8.8	30.2	
g4564b	2.6	38.0	12.2	25.4	0.0	40.2	40.2	35.2	18.2	3.5	3.5	7.7	50.6	25.8	18.4	24.3	20.6	6.4	
m2	2.6	38.0	12.2	25.4	0.0	40.2	40.2	35.2	18.2	3.5	3.5	7.7	50.6	25.8	18.4	24.3	20.6	6.4	
g4523	44.8	6.4	72.5	11.2	40.2	40.2	0.5	6.7	172.7	37.7	37.7	65.2	4.1	15.7	32.0	14.8	20.5	50.0	
m583	43.4	6.4	85.3	11.8	40.2	40.2	0.5	7.3	172.7	39.4	38.3	65.7	5.1	16.3	33.7	14.6	21.7	46.1	
g4708	41.5	0.0	57.8	4.0	35.2	35.2	6.7	7.3	68.8	27.5	27.5	52.0	7.5	6.2	21.0	6.8	13.1	37.1	
m424	11.0	84.2	4.1	49.7	18.2	18.2	172.7	172.7	68.8	-----	16.3	10.1	172.7	45.1	24.9	36.0	39.7	16.3	
m249	5.9	29.8	11.5	24.1	3.5	3.5	37.7	39.2	27.5	16.3	-----	0.0	10.7	47.7	24.4	13.9	22.3	17.6	5.8
g4117	5.8	29.8	11.5	23.7	3.5	3.5	37.7	38.3	27.5	16.3	0.0	-----	10.5	51.5	24.0	13.9	21.8	17.6	5.7
m457	3.7	72.0	5.4	45.3	7.7	7.7	65.2	65.7	52.0	10.1	10.7	10.5	-----	100.2	41.3	22.1	28.8	33.4	8.3
GAPC	69.8	8.0	172.7	16.4	50.6	50.6	4.1	5.1	7.5	172.7	47.7	51.5	100.2	-----	15.1	27.5	16.8	23.9	94.0
m560B	35.0	7.3	43.9	1.7	25.8	25.8	15.7	16.3	6.2	45.3	24.4	24.0	41.3	15.1	-----	19.1	0.6	12.8	30.2
um579c	22.2	20.7	24.8	19.4	18.4	18.4	32.0	33.7	21.0	24.9	13.9	13.9	22.1	27.5	19.1	-----	16.9	10.1	15.7
O4023	28.2	7.0	38.3	1.3	24.3	24.3	14.8	14.6	6.8	36.0	22.3	21.8	28.8	16.8	0.6	16.9	-----	11.7	26.8
Abi13	30.6	12.4	35.2	8.8	20.6	20.6	20.5	21.7	13.1	39.7	17.6	17.6	33.4	23.9	12.8	10.1	11.7	-----	22.6
O97B1	8.1	44.8	12.3	30.2	6.4	6.4	50.0	46.1	37.1	16.3	5.8	5.7	8.3	94.0	30.2	15.7	26.8	22.6	-----

Duplicates can be identified here

## “Nicemapd” button

The screenshot shows the CarthaGene Window software interface. At the top, there is a menu bar with options: Data, Loci, Search, Maps, Help, and a toolbar with icons for B, I, A, Defalgo, SEM, Nicemapd, Nicemapl, Build, BuildFW, C.Frontier, Annealing, Taboo, Genetic, Flips, Polish. The Nicemapd icon is highlighted with a red box.

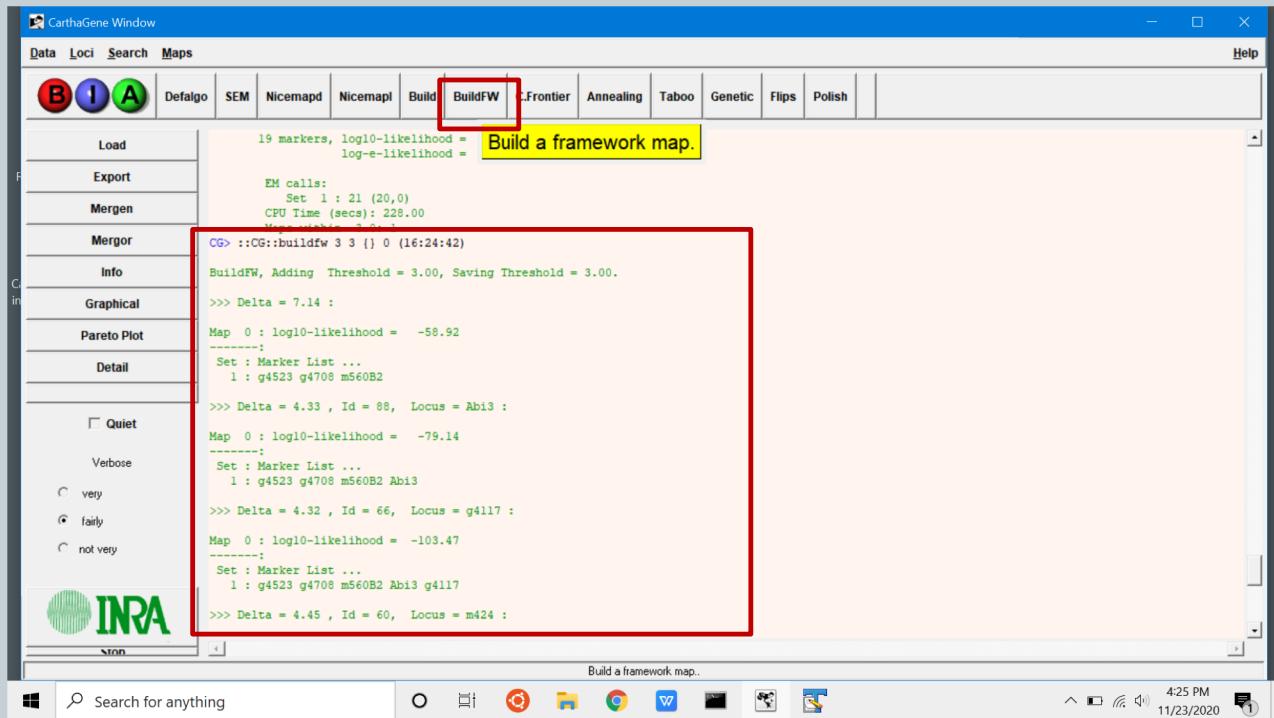
The main window displays a log file with the following content:

```
Map -1 : log10-likelihood = -205.95
-----
Set : Marker List ...
1 : m424 g2778 m457 g4014 g4564b m2 m249 g4117 097B1 um579C Abi3 m105 O4023 m560B2 g4708 m228 m583 g4523 GAC
27.5 16.3 0.0 ----- 10.5 51.5 24.0 13.9 21.8 17.6 5.7
m457 | 3.7 72.0 5.4 45.3 7.7 7.7 65.2 65.7 52.0 10.1 10.7 10.5 ----- 100.2 41.3 22.1 28.8 33.4 8.3
GAC | 69.8 8.0 172.7 16.4 50.6 50.6 4.1 5.1 7.5 172.7 47.7 51.5 100.2 ----- 15.1 27.5 16.8 23.9 94.0
m560B2 | 35.0 7.3 43.9 1.7 25.8 25.8 15.7 16.3 6.2 45.1 24.4 24.0 41.3 15.1 ----- 19.1 0.6 12.8 30.2
um579C | 22.2 20.7 24.8 19.4 18.4 18.4 32.0 33.7 21.0 24.9 13.9 13.9 22.1 27.5 19.1 ----- 16.9 10.1 15.7
O4023 | 28.2 7.0 38.3 1.3 24.3 24.3 14.8 14.6 6.8 36.0 22.3 21.8 28.8 16.8 0.6 16.9 ----- 11.7 26.8
Abi3 | 30.6 12.4 35.2 8.8 20.6 20.6 20.5 21.7 13.1 39.7 17.6 17.6 33.4 23.9 12.8 10.1 11.7 ----- 22.6
097B1 | 8.1 44.8 12.3 30.2 6.4 6.4 50.0 46.1 37.1 16.3 5.8 5.7 8.3 94.0 30.2 15.7 26.8 22.6 -----
Pareto Plot
Detail
Quiet
Verbose
very
fairly
not very
INRA
Stop
```

A red box highlights the "Detail" button under the Pareto Plot section. Another red box highlights the data table below:

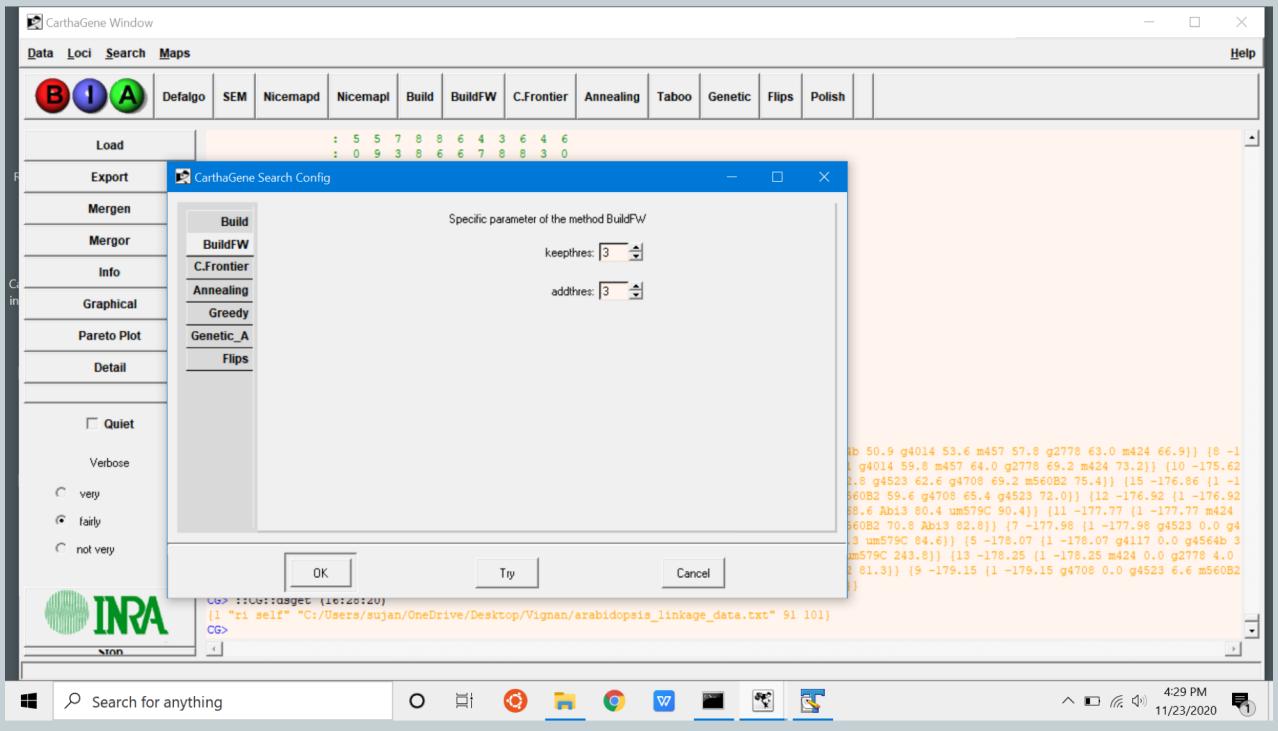
Pos	Id	name	Markers	Haldane	Cumulative Haldane	Distance Kosambi	Theta (%)age	2pt LOD
1	60	m424	4.1 cM	0.0 cM	4.0 cM	4.0 %	16.9	
2	43	g2778	5.5 cM	4.1 cM	5.2 cM	5.2 %	14.9	
3	68	m457	4.3 cM	9.6 cM	4.2 cM	4.2 %	16.7	
4	38	g4014	3.0 cM	13.9 cM	3.0 cM	3.0 %	17.7	
5	47	g4564b	0.0 cM	17.0 cM	0.0 cM	0.0 %	27.4	
6	48	m2	3.8 cM	17.0 cM	3.7 cM	3.7 %	17.8	
7	61	m249	0.0 cM	20.8 cM	0.0 cM	0.0 %	29.5	
8	66	g4117	6.8 cM	20.8 cM	6.4 cM	6.4 %	12.3	
9	90	097B1	22.6 cM	27.6 cM	19.0 cM	18.2 %	3.9	
10	86	um579C	11.3 cM	50.2 cM	10.3 cM	10.1 %	7.6	

## buildfw



Last map is the best

## Search/config



## Best map -

Last in heap (or)

bestprint; bestprintd commands

The screenshot shows the CarthaGene Window software interface. At the top, there is a menu bar with 'Data', 'Loci', 'Search', and 'Maps'. Below the menu is a toolbar with icons for 'Defago' (red), 'SEM' (blue), 'Nicemapd' (green), 'Nicemapl' (light green), 'Build' (orange), 'BuildFW' (yellow), 'C.Frontier' (purple), 'Annealing' (pink), 'Taboo' (light blue), 'Genetic' (light orange), 'Flips' (light pink), and 'Polish' (light yellow). A red box highlights the terminal window where the 'bestprint' command was run.

```
Maps within -3.0: 1
CG> bestmap (16:29:51)
invalid command name "baermap"
CG> bestprint (16:30:33)
Map 14 : log10-likelihood = -171.17
-----
Set : Marker List ...
1 : g4523 g4708 m560B2 Abi3 um579C g4117 g4564b g4014 m457 g2778 m424
14
CG> bestprintd (16:30:40)
Map 14 : log10-likelihood = -171.17, log-e-likelihood = -394.13
-----
Data Set Number 1 :
      Markers Distance Cumulative Distance Theta 2pt LOD
      Pos Id name Haldane Haldane Kosambi (%age) LOD
      Quiet
      Verbose
      very
 fairly
 not very
      1 50 g4523    7.0 cM   0.0 cM   6.6 % 13.6
      2 59 g4708    6.2 cM   7.0 cM   5.9 % 13.7
      3 73 m560B2   13.1 cM  13.2 cM  11.5 % 7.5
      4 85 Abi3     10.8 cM  23.9 cM  9.7 % 7.6
      5 86 um579C   15.2 cM  37.2 cM  13.1 % 5.7
      6 66 g4117    3.5 cM   52.4 cM  3.4 % 17.8
      7 47 g4564b   2.8 cM   55.9 cM  2.7 % 17.7
      8 38 g4014    4.4 cM   58.6 cM  4.2 % 16.7
      9 68 m457     5.5 cM   63.0 cM  5.2 % 14.9
      10 43 g2778   4.1 cM   68.5 cM  4.0 % 16.9
      11 60 m424    72.6 cM  66.9 cM

```

The terminal window displays command history and output from the 'bestprint' and 'bestprintd' commands. The output shows a list of markers and their positions, distances, and LOD values. The INRA logo is visible at the bottom of the software window.