Mapping with Record
When Joinmap can't handle your data

Presenter: Dr. Allen Van Deynze,
UC Davis

Host: Heather Merk

http://www.extension.org/plant_breeding_genomics

Slides available at http://www.extension.org/pages/63330
Sign up for PBG news at http://pbgworks.org

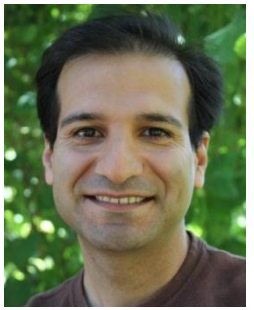
extension
more mind reach

Please fill out the survey evaluation! (You will be contacted via email)

Watch past webinars and sign up for future webinars!

http://www.extension.org/pages/60426

Mapping with RECORD
When Joinmap can't handle your data



Allen Van Deynze and Hamid Ashrafi

University of California, Davis

SOLCAP workshop April. 12, 2012,

Acknowledgements

- Richard Michelmore
- Alex Kozik
- Maria Truco
- Sebastian Reyes Chin-Wo
- Theresa Hill
- Kevin Stoffel

MadMapper: **Suite of Python scripts for quality control of genetic markers, group analysis and inference of linear order of markers on linkage groups**

Alexander Kozik, UC Davis Genome Center

http://cgpdb.ucdavis.edu/XLinkage/MadMapper/

Record: **Marker order and distances**

http://www.plantbreeding.wur.nl/UK/software_record.html

CheckMatrix: **Visualization and validation of genetic maps using two-dimensional heat-plots.**

MadMapper - group analysis and quality control of genetic markers

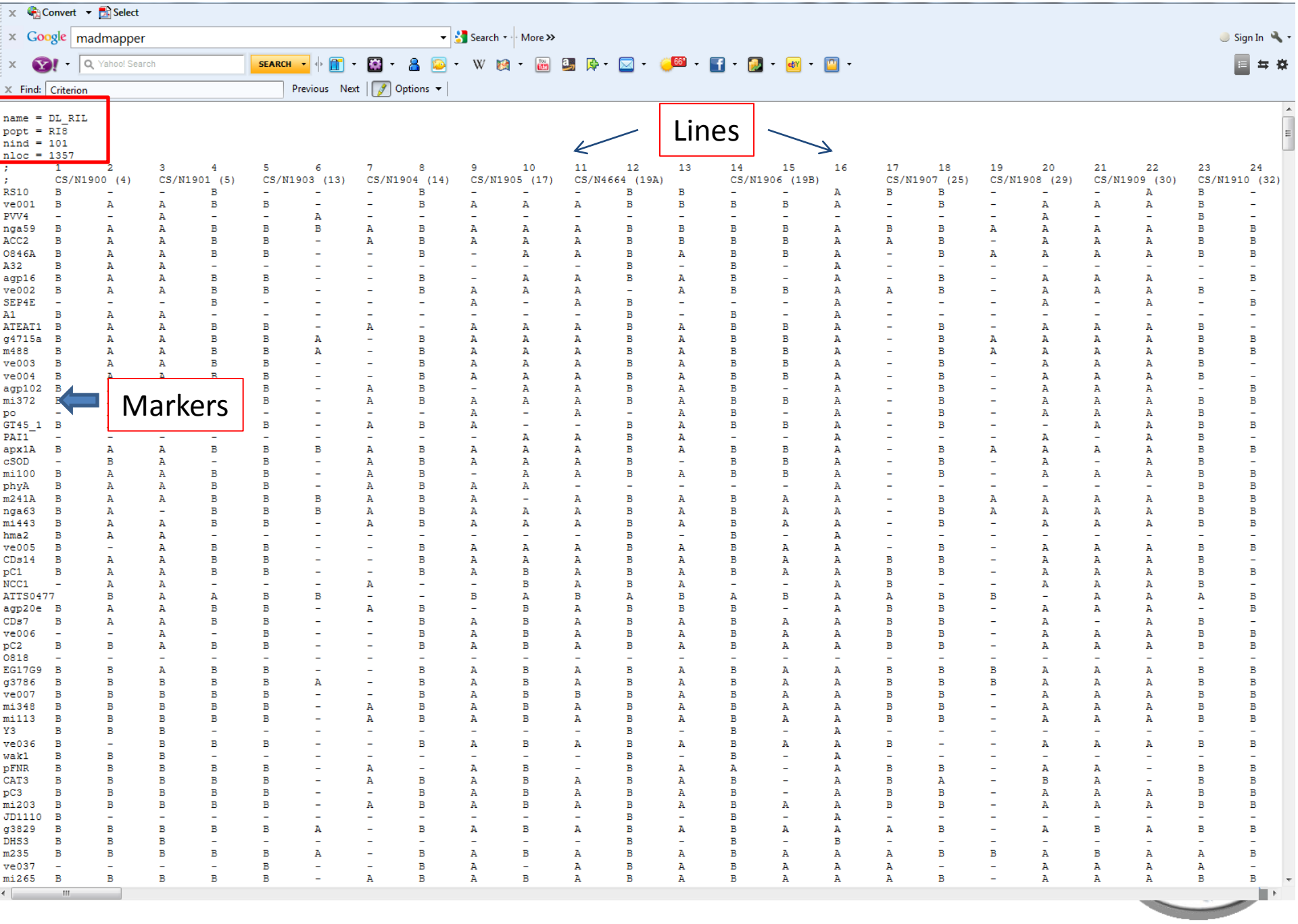
STEP 1

1. Group (cluster) all markers from locus file and assign markers to distinct linkage groups.

2. Select markers for linkage group 8 (as an example) and extract from that group reliable markers with high quality scores.

   LG8.loc locus file with raw marker scores.

3. Then we will go to Part 2 (Win Record) and will try to infer linear order of selected markers on Pepper linkage group 8.




```
name = DL_RIL
popt = RI8
nind = 101
nloc = 1357
```

;	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
;	CS/N1900 (4)		CS/N1901 (5)		CS/N1903 (13)		CS/N1904 (14)		CS/N1905 (17)		CS/N4664 (19A)			CS/N1906 (19B)			CS/N1907 (25)		CS/N1908 (29)		CS/N1909 (30)		CS/N1910 (32)	
RS10	B	-	-	B	-	-	-	B	-	-	-	B	B	-	-	A	B	B	-	-	-	A	B	-
ve001	B	A	A	B	B	-	-	B	A	A	A	B	B	B	B	A	-	B	-	A	A	A	B	-
PVV4	-	-	A	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	B	-
nga59	B	A	A	B	B	B	A	B	A	A	A	B	B	B	B	A	B	B	A	A	A	A	B	B
ACC2	B	A	A	B	B	-	A	B	A	A	A	B	B	B	B	A	A	B	-	A	A	A	B	B
0846A	B	A	A	B	B	-	-	B	-	A	A	B	A	B	B	A	-	B	A	A	A	A	B	B
A32	B	A	A	-	-	-	-	-	-	-	-	B	-	B	-	A	-	-	-	-	-	-	-	-
agp16	B	A	A	B	B	-	-	B	-	A	A	B	A	B	-	A	-	B	-	A	A	A	-	B
ve002	B	A	A	B	B	-	-	B	A	A	A	-	A	B	B	A	A	B	-	A	A	A	B	-
SEP4E	-	-	-	B	-	-	-	-	A	-	A	B	-	-	-	A	-	-	-	A	-	A	-	B
A1	B	A	A	-	-	-	-	-	-	-	-	B	-	B	-	A	-	-	-	-	-	-	-	-
ATEAT1	B	A	A	B	B	-	A	-	A	A	A	B	A	B	B	A	-	B	-	A	A	A	B	-
g4715a	B	A	A	B	B	A	-	B	A	A	A	B	A	B	B	A	-	B	A	A	A	A	B	B
m488	B	A	A	B	B	A	-	B	A	A	A	B	A	B	B	A	-	B	A	A	A	A	B	B
ve003	B	A	A	B	B	-	-	B	A	A	A	B	A	B	B	A	-	B	-	A	A	A	B	-
ve004	B	A	A	B	B	-	-	B	A	A	A	B	A	B	B	A	-	B	-	A	A	A	B	-
agp102	B	-	-	-	B	-	A	B	-	A	A	B	A	B	-	A	-	B	-	A	A	A	-	B
mi372	B	-	-	-	B	-	A	B	A	A	A	B	A	B	B	A	-	B	-	A	A	A	B	B
po	-	-	-	-	-	-	-	-	A	-	A	-	A	B	-	A	-	B	-	A	A	A	B	-
GT45_1	B	-	-	-	B	-	A	B	A	-	-	B	A	B	B	A	-	B	-	-	A	A	B	B
PAI1	-	-	-	-	-	-	-	-	-	A	A	B	A	-	-	A	-	-	-	A	-	A	B	-
apx1A	B	A	A	B	B	B	A	B	A	A	A	B	A	B	B	A	-	B	A	A	A	A	B	B
cSOD	-	B	A	-	B	-	A	B	A	A	A	B	-	B	B	A	-	B	-	A	-	A	B	-
mi100	B	A	A	B	B	-	A	B	-	A	A	B	A	B	B	A	-	B	-	A	A	A	B	B
phyA	B	A	A	B	B	-	A	B	A	A	-	-	-	-	-	A	-	-	-	-	-	-	B	B
m241A	B	A	A	B	B	B	A	B	A	-	A	B	A	B	A	A	-	B	A	A	A	A	B	B
nga63	B	A	-	B	B	B	A	B	A	A	A	B	A	B	A	A	-	B	A	A	A	A	B	B
mi443	B	A	A	B	B	-	A	B	A	A	A	B	A	B	A	A	-	B	-	A	A	A	B	B
hma2	B	A	A	-	-	-	-	-	-	-	-	B	-	B	-	A	-	-	-	-	-	-	-	-
ve005	B	-	A	B	B	-	-	B	A	A	A	B	A	B	A	A	-	B	-	A	A	A	B	B
CDs14	B	A	A	B	B	-	-	B	A	A	A	B	A	B	A	A	B	B	-	A	A	A	B	-
pC1	B	A	A	B	B	-	-	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B	B
NCC1	-	A	A	-	-	-	A	-	-	B	A	B	A	-	-	A	B	-	-	A	A	A	B	-
ATTS0477		B	A	A	B	B	-	-	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B
agp20e	B	A	A	B	B	-	A	B	-	B	A	B	B	B	-	A	B	B	-	A	A	A	-	B
CDs7	B	A	A	B	B	-	-	B	A	B	A	B	A	B	A	A	B	B	-	A	-	A	B	-
ve006	-	-	A	-	B	-	-	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B	B
pC2	B	B	A	B	B	-	-	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B	B
0818	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EG17G9	B	B	A	B	B	-	-	B	A	B	A	B	A	B	A	A	B	B	B	A	A	A	B	B
g3786	B	B	B	B	B	A	-	B	A	B	A	B	A	B	A	A	B	B	B	A	A	A	B	B
ve007	B	B	B	B	B	-	-	B	A	B	B	B	A	B	A	A	B	B	-	A	A	A	B	B
mi348	B	B	B	B	B	-	A	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B	B
mi113	B	B	B	B	B	-	A	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B	B
Y3	B	B	B	-	-	-	-	-	-	-	-	B	-	B	-	A	-	-	-	-	-	-	-	-
ve036	B	-	B	B	B	-	-	B	A	B	A	B	A	B	A	A	B	-	-	A	A	A	B	B
wak1	B	B	B	-	-	-	-	-	-	-	-	B	-	B	-	A	-	-	-	-	-	-	-	-
pFNR	B	B	B	B	B	-	A	-	A	B	-	B	A	A	-	A	B	B	-	A	A	-	B	B
CAT3	B	B	B	B	B	-	A	B	A	B	A	B	A	B	-	A	B	A	-	B	A	-	B	B
pC3	B	B	B	B	B	-	-	B	A	B	A	B	A	B	-	A	B	B	-	A	A	A	B	B
mi203	B	B	B	B	B	-	A	B	A	B	A	B	A	B	A	A	B	B	-	A	A	A	B	B
JD1110	B	-	-	-	-	-	-	-	-	-	-	B	-	B	-	A	-	-	-	-	-	-	-	-
g3829	B	B	B	B	B	A	-	B	A	B	A	B	A	B	A	A	A	B	-	A	B	A	B	B
DHS3	B	B	B	-	-	-	-	-	-	-	-	B	-	B	-	B	-	-	-	-	-	-	-	-
m235	B	B	B	B	B	A	-	B	A	B	A	B	A	B	A	A	A	B	B	A	B	A	A	B
ve037	-	-	-	B	-	-	-	B	A	-	A	B	A	B	A	A	A	-	-	A	A	A	A	-
mi265	B	B	B	B	B	-	A	B	A	B	A	B	A	B	A	A	A	B	-	A	A	A	B	B

First Step - Group Analysis - We will run the script with the following options/arguments:

Prompt> **python** *Python_MadMapper_V248_RECBIT_012.py*
Testdata.loc Testdataout 0.2 100 25 frame.IDs 0.33 50 NOTRIO 3

Prompt> python
Python_MadMapper_V248_RECBIT_012.py

PROGRAM USAGE:
MAD MAPPER TAKES 10 ARGUMENTS/OPTIONS IN THE FOLLOWING ORDER:
(1)input_file[LOC_DATA/MARKER SCORES] (2)output_file[NAME]
(3)rec_cut[0.2] (4)bit_cut[100] (5)data_cut[25]
(6)group_file[OPTIONAL] (7)allele_dist[0.33] (8)missing_data[50]
(9)trio_analysis[TRIO/NOTRIO] (10)double_cross[3]

Results of Step 1

Output will be represented by 76 files
[Note, that at this time the script was running with "NOTRIO" option]

At this step we are interested in TestDataout.x_tree_clust file only. This file has the information about marker clustering/grouping. From this file we will select markers belonging to the linkage group 8.

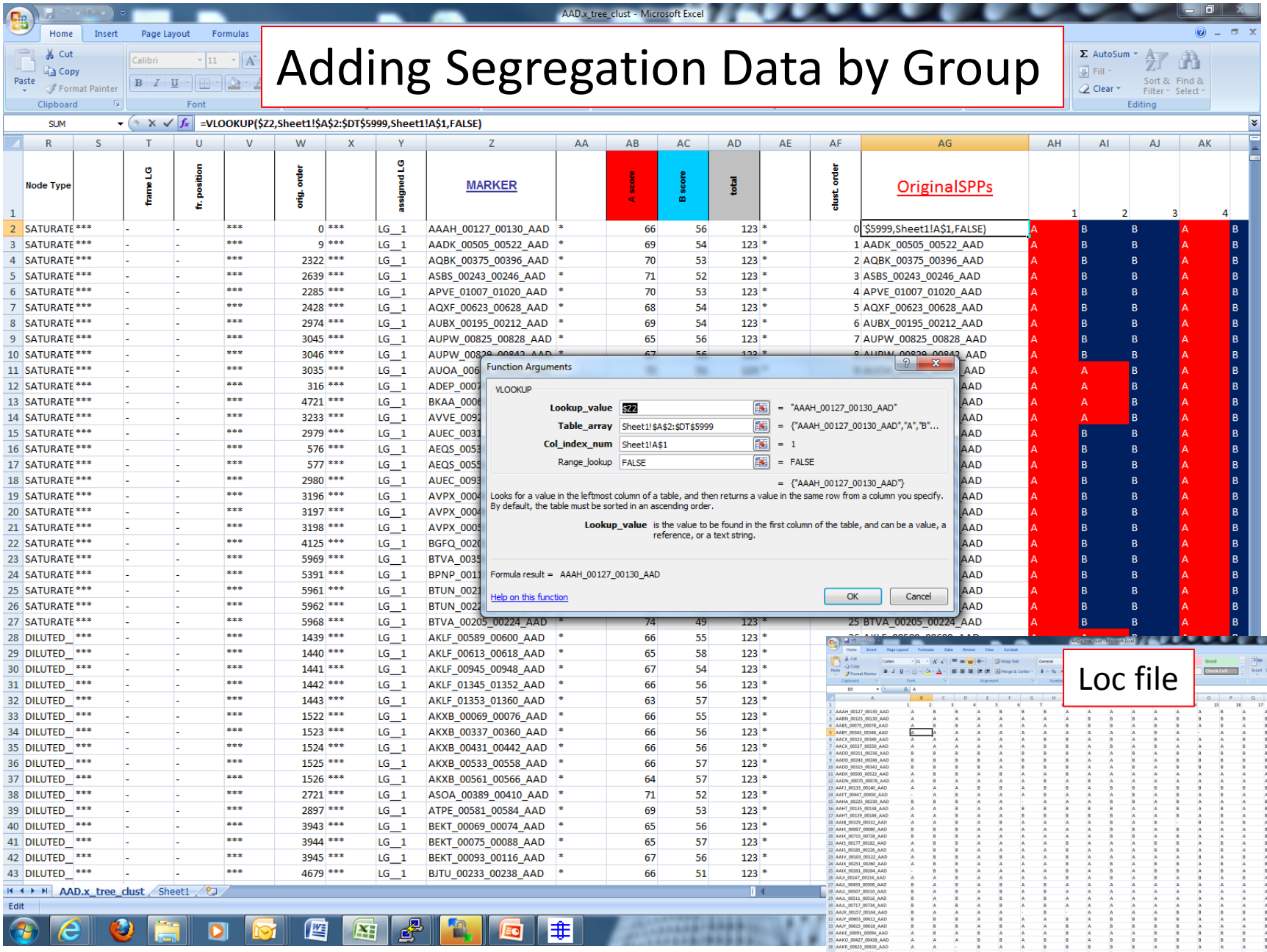
We will analyze TestDataout.x_tree_clust file using MS Excel to find and highlight (select) a set of markers belonging to the linkage group 8.
TestDataout.x_tree_clust.xls - tree clustering file in MS Excel format.

Z47 | ACOH_00171_00174_AAD

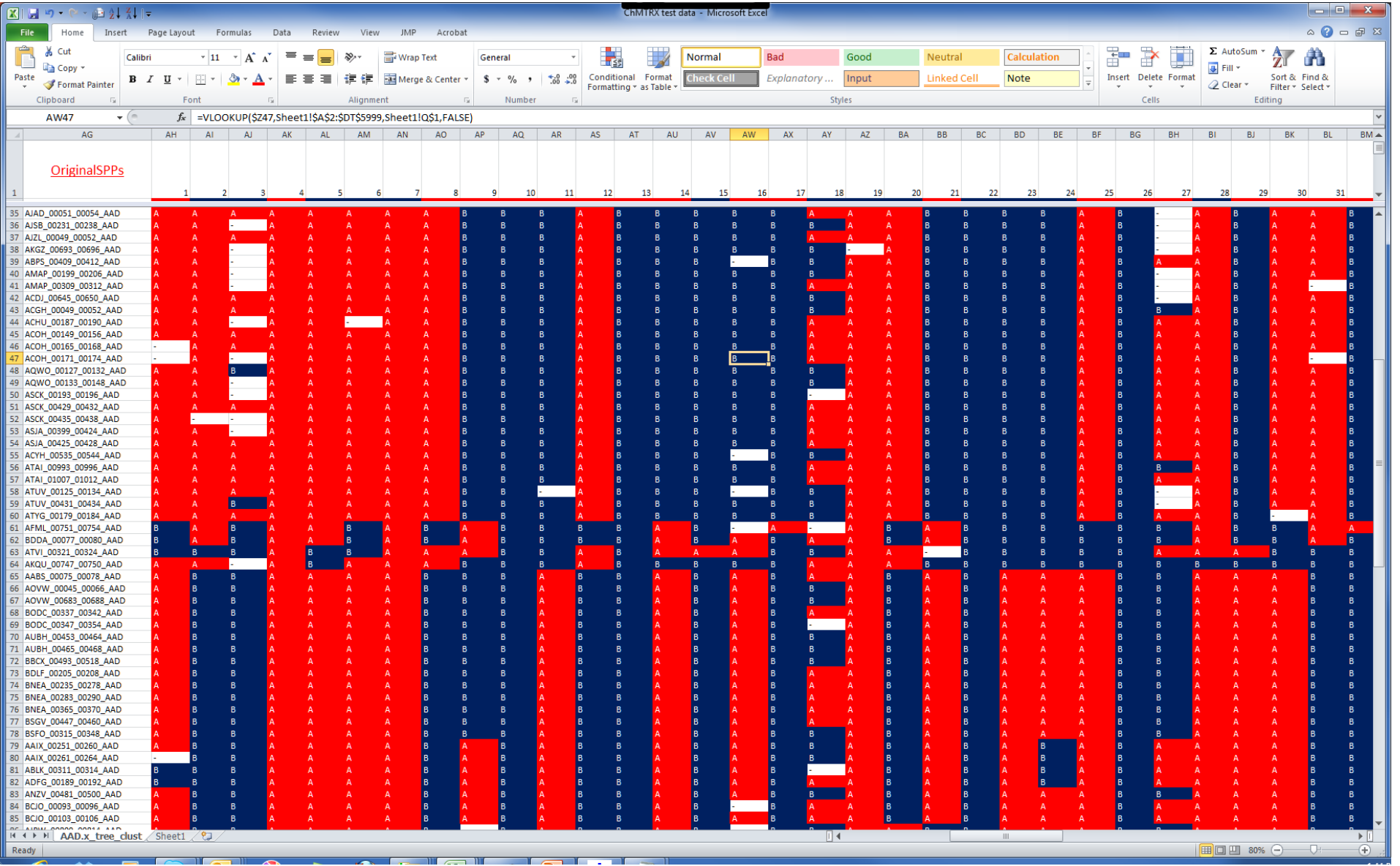
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	0.20	0.18	0.16	0.14	0.12	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Graph Class	Node Type		frame LG	fr. position		orig. order		assigned LG	MARKER		A score	B score	total
26	1	1	1	1	1	11	27	32	37	65	93	143	206	272	387	845	LINKED___GROUP_00845	DILUTED___NODE	***	-	-	***	4138	***	LG__1	BGIF_00075_00078_AAD	*	67	54	123
27	1	1	1	1	1	11	27	32	37	65	93	143	206	272	387	845	LINKED___GROUP_00845	SATURATED_NODE	***	LG1	12	***	1851	***	LG__1	ANII_00243_00246_AAD	*	68	54	123
28	1	1	1	1	1	11	27	32	37	65	93	143	206	272	387	1342	SINGLE____NODE_01342	SATURATED_NODE	***	-	-	***	3379	***	LG__1	AXYW_00167_00188_AAD	*	68	54	123
29	1	1	1	1	1	11	27	32	37	65	93	143	206	272	387	1368	SINGLE____NODE_01368	SATURATED_NODE	***	-	-	***	3475	***	LG__1	AYXK_00389_00392_AAD	*	69	49	123
30	1	1	1	1	1	11	27	32	37	65	93	143	206	272	763	1357	SINGLE____NODE_01357	SATURATED_NODE	***	-	-	***	3443	***	LG__1	AYQZ_00151_00180_AAD	*	70	53	123
31	1	1	17	18	23	32	40	50	63	88	136	244	389	599	967	1826	COMPLETE_GRAPH_01826	SATURATED_NODE	***	LG1	20	***	5222	***	LG__1	BOFS_00123_00126_AAD	*	95	27	123
32	1	1	17	18	23	32	40	50	63	88	136	244	389	599	967	1826	COMPLETE_GRAPH_01826	SATURATED_NODE	***	-	-	***	5223	***	LG__1	BOFS_00127_00142_AAD	*	96	27	123
33	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1	***	LG__2	AABN_00123_00130_AAD	*	63	58	123
34	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1131	***	LG__2	AIDH_00393_00400_AAD	*	60	60	123
35	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1251	***	LG__2	AJAD_00051_00054_AAD	*	62	59	123
36	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1344	***	LG__2	AJSB_00231_00238_AAD	*	59	60	123
37	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1387	***	LG__2	AJZL_00049_00052_AAD	*	60	59	123
38	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1421	***	LG__2	AKGZ_00693_00696_AAD	*	59	60	123
39	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	144	***	LG__2	ABPS_00409_00412_AAD	*	61	59	123
40	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1658	***	LG__2	AMAP_00199_00206_AAD	*	59	59	123
41	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	1659	***	LG__2	AMAP_00309_00312_AAD	*	58	59	123
42	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	189	***	LG__2	ACDJ_00645_00650_AAD	*	61	60	123
43	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	205	***	LG__2	ACGH_00049_00052_AAD	*	59	61	123
44	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	211	***	LG__2	ACHU_00187_00190_AAD	*	62	59	123
45	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	239	***	LG__2	ACOH_00149_00156_AAD	*	63	59	123
46	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	240	***	LG__2	ACOH_00165_00168_AAD	*	62	59	123
47	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	241	***	LG__2	ACOH_00171_00174_AAD	*	57	59	123
48	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2422	***	LG__2	AQWO_00127_00132_AAD	*	61	61	123
49	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2423	***	LG__2	AQWO_00133_00148_AAD	*	62	60	123
50	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2647	***	LG__2	ASCK_00193_00196_AAD	*	62	59	123
51	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2648	***	LG__2	ASCK_00429_00432_AAD	*	63	58	123
52	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2649	***	LG__2	ASCK_00435_00438_AAD	*	61	59	123
53	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2697	***	LG__2	ASJA_00399_00424_AAD	*	63	59	123
54	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2698	***	LG__2	ASJA_00425_00428_AAD	*	64	58	123
55	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	277	***	LG__2	ACYH_00535_00544_AAD	*	61	58	123
56	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2812	***	LG__2	ATAI_00993_00996_AAD	*	63	60	123
57	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2813	***	LG__2	ATAI_01007_01012_AAD	*	64	58	123
58	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2939	***	LG__2	ATUV_00125_00134_AAD	*	61	57	123
59	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2940	***	LG__2	ATUV_00431_00434_AAD	*	57	61	123
60	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	LINKED___GROUP_00002	DILUTED___NODE	***	-	-	***	2958	***	LG__2	ATYG_00179_00184_AAD	*	62	60	123
61	2	2	2	2	2	9	9	9	9	10	11	109	154	195	264	365	LINKED___GROUP_00365	SATURATED_NODE	***	-	-	***	705	***	LG__2	AFML_00751_00754_AAD	*	50	71	123
62	2	2	2	2	2	9	9	9	9	10	11	109	154	195	824	1490	SINGLE____NODE_01490	SATURATED_NODE	***	-	-	***	3824	***	LG__2	BDDA_00077_00080_AAD	*	50	72	123
63	2	2	2	2	2	9	9	9	9	10	116	202	310	446	698	1204	SINGLE____NODE_01204	SATURATED_NODE	***	-	-	***	2942	***	LG__2	ATVI_00321_00324_AAD	*	58	60	123
64	2	13	14	14	18	26	32	40	51	71	101	160	228	315	456	692	SINGLE____NODE_00692	SATURATED_NODE	***	-	-	***	1479	***	LG__2	AKQU_00747_00750_AAD	*	34	88	123
65	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	SINGLE____NODE_00003	SATURATED_NODE	***	-	-	***	2	***	LG__3	AABS_00075_00078_AAD	*	66	56	123
66	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	934	LINKED___GROUP_00934	DILUTED___NODE	***	-	-	***	2125	***	LG__3	AOVW_00045_00066_AAD	*	65	58	123
67	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	934	LINKED___GROUP_00934	DILUTED___NODE	***	-	-	***	2126	***	LG__3	AOVW_00683_00688_AAD	*	64	58	123
68	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	934	LINKED___GROUP_00934	DILUTED___NODE	***	-	-	***	5205	***	LG__3	BODC_00337_00342_AAD	*	66	57	123
69	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	934	LINKED___GROUP_00934	SATURATED_NODE	***	-	-	***	5206	***	LG__3	BODC_00347_00354_AAD	*	63	57	123
70	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1211	COMPLETE_GRAPH_01211	SATURATED_NODE	***	-	-	***	2968	***	LG__3	AUBH_00453_00464_AAD	*	61	61	123
71	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1211	COMPLETE_GRAPH_01211	SATURATED_NODE	***	-	-	***	2969	***	LG__3	AUBH_00465_00468_AAD	*	61	58	123
72	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1431	SINGLE____NODE_01431	SATURATED_NODE	***	-	-	***	3645	***	LG__3	BBCX_00493_00518_AAD	*	64	59	123
73	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1502	SINGLE____NODE_01502	SATURATED_NODE	***	-	-	***	3873	***	LG__3	BDLF_00205_00208_AAD	*	63	60	123
74	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1787	COMPLETE_GRAPH_01787	SATURATED_NODE	***	-	-	***	5086	***	LG__3	BNEA_00235_00278_AAD	*	64	59	123
75	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1787	COMPLETE_GRAPH_01787	SATURATED_NODE	***	-	-	***	5087	***	LG__3	BNEA_00283_00290_AAD	*	62	59	123
76	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1787	COMPLETE_GRAPH_01787	SATURATED_NODE	***	-	-	***	5088	***	LG__3	BNEA_00365_00370_AAD	*	61	59	123
77	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1787	COMPLETE_GRAPH_01787	SATURATED_NODE	***			***	5787	***	LG__3	BSCV_00447_00460_AAD	*	63		123

AAD x_type_dupl | Sheet1

Adding Segregation Data by Group



Save for each linkage group



Further processing the result files

From this group info file we will extract marker IDs belonging to the linkage group 8 LG8.tab and create new locus file LG8.loc which contains marker scores for linkage group 4 only. It has raw scores for 247 markers. We **vlookup** function in Excel.

=VLOOKUP($Z5223,Sheet1!$A$2:$DT$5999,Sheet1!A$1,FALSE)

OPTIONAL ANALYSIS

Now we will run Python_MadMapper_V248_RECBIT_012.py again on LG8.loc file with "**TRIO**" option:

$python Python_MadMapper_V248_RECBIT_012.py LG8.loc LG8.out 0.2 100 25 X 0.33 50 TRIO 3

83 files will be generated. We are interested in the Marker Summary file: LG8.out_marker_sum which contains information about loss of data and allele distortion for each marker, as well as useful information derived from "TRIO" analysis. From this file we will extract marker IDs which have "GOOD__MARKER" label (grade) only. It means we will use high-quality markers for further analysis. These markers have low fraction of data loss and meaningful ratio of "A"/"B" scores (allele distortion). There are 171 'good' markers in the dataset.

So, from here we are ready to jump to the **Part 2** to infer linear order of markers on linkage group 8.

Pairwise matrices- We will run the script with the following options/arguments for each linkage group

Prompt> **python** *Python_MadMapper_V248_RECBIT_012.py* LG_8.loc LG_8out 0.2 100 25 X 0.33 50 NOTRIO 3

See CheckMatrix analyses

Prompt> python

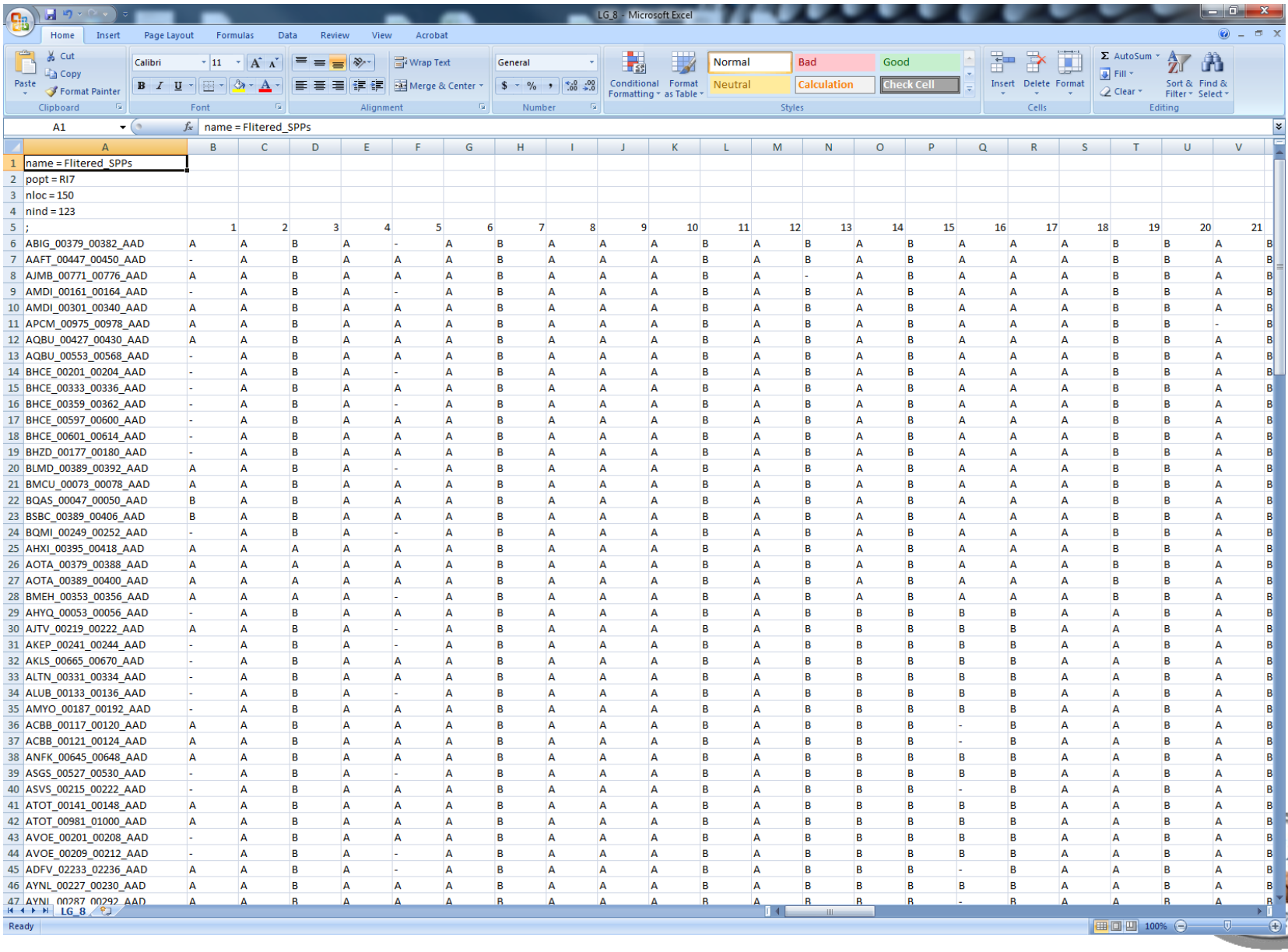
Python_MadMapper_V248_RECBIT_012.py

PROGRAM USAGE:

MAD MAPPER TAKES 10 ARGUMENTS/OPTIONS IN THE FOLLOWING ORDER:

(1)input_file[LOC_DATA/MARKER SCORES] (2)output_file[NAME]
(3)rec_cut[0.2] (4)bit_cut[100] (5)data_cut[25]
(6)group_file[OPTIONAL] (7)allele_dist[0.33] (8)missing_data[50]
(9)trio_analysis[TRIO/NOTRIO] (10)double_cross[3]

Linkage group 8



LG_8 - Microsoft Excel

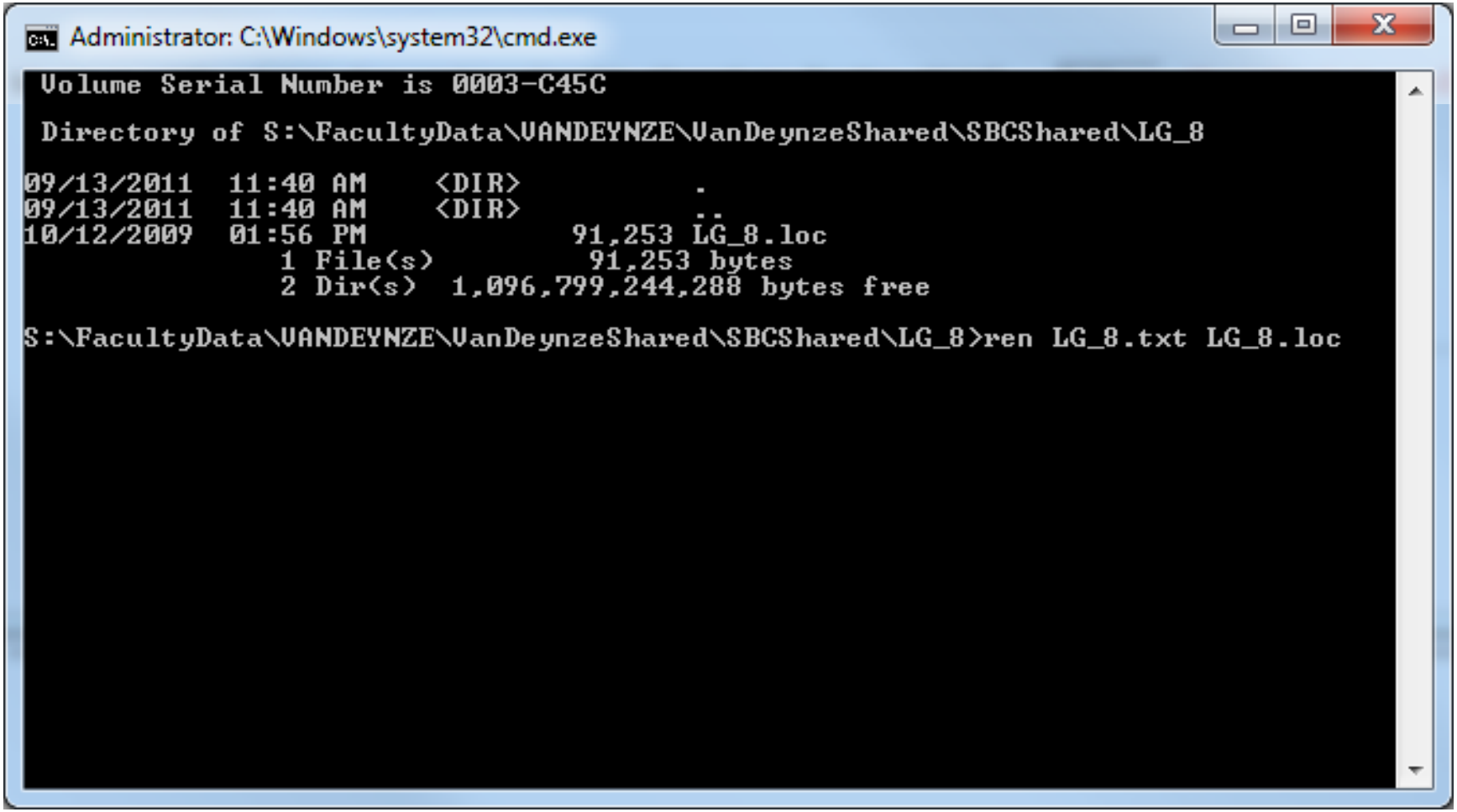
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	name = Flitered_SPPs																					
2	popt = RI7																					
3	nloc = 150																					
4	nind = 123																					
5	;	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
6	ABIG_00379_00382_AAD	A	A	B	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
7	AAFT_00447_00450_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
8	AJMB_00771_00776_AAD	A	A	B	A	A	A	B	A	A	A	B	A	-	A	B	A	A	A	B	B	A
9	AMDI_00161_00164_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
10	AMDI_00301_00340_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
11	APCM_00975_00978_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	-
12	AQBU_00427_00430_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
13	AQBU_00553_00568_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
14	BHCE_00201_00204_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
15	BHCE_00333_00336_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
16	BHCE_00359_00362_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
17	BHCE_00597_00600_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
18	BHCE_00601_00614_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
19	BHZD_00177_00180_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
20	BLMD_00389_00392_AAD	A	A	B	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
21	BMCU_00073_00078_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
22	BQAS_00047_00050_AAD	B	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
23	BSBC_00389_00406_AAD	B	A	B	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
24	BQMI_00249_00252_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
25	AHXI_00395_00418_AAD	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
26	AOTA_00379_00388_AAD	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
27	AOTA_00389_00400_AAD	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
28	BMEH_00353_00356_AAD	A	A	A	A	-	A	B	A	A	A	B	A	B	A	B	A	A	A	B	B	A
29	AHYQ_00053_00056_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
30	AJTV_00219_00222_AAD	A	A	B	A	-	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
31	AKEP_00241_00244_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
32	AKLS_00665_00670_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
33	ALTN_00331_00334_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
34	ALUB_00133_00136_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
35	AMYO_00187_00192_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
36	ACBB_00117_00120_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	-	B	A	A	B	A
37	ACBB_00121_00124_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	-	B	A	A	B	A
38	ANFK_00645_00648_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
39	ASGS_00527_00530_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
40	ASVS_00215_00222_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	B	B	-	B	A	A	B	A
41	ATOT_00141_00148_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
42	ATOT_00981_01000_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
43	AVOE_00201_00208_AAD	-	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
44	AVOE_00209_00212_AAD	-	A	B	A	-	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
45	ADFV_02233_02236_AAD	A	A	B	A	-	A	B	A	A	A	B	A	B	B	B	-	B	A	A	B	A
46	AYNL_00227_00230_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	B	B	A	A	B	A
47	AYNL_00287_00292_AAD	A	A	B	A	A	A	B	A	A	A	B	A	B	B	B	-	B	A	A	B	A

LG_8

Ready

100%

Rename txt file to loc file in command line of DOS



Inference of linear order of markers on linkage groups using WinRecord

STEP 2

We use Record software to order the markers

http://www.plantbreeding.wur.nl/UK/software_record.html

Welcome

Welcome to RECORD

Tue, 10 Apr 2012, 12:21

Freeware distributed by the Laboratory of PLant Breeding of Wageningen University

Copyright Piet Stam 2006

What does RECORD do?	Fast mapping of markers on a single linkage group
What population types can RECORD deal with?	BC1, F2 and RILs of any generation (selfed)
Which object function is used by RECORD?	RECORD minimises the total number of recombination events in the data

RECORD can be used as an additional tool to the JoinMap package , distributed by Kyazma company, Wageningen (www. kyazama.nl)

Input:

Raw genotypic data in JoinMap format

Output:

1. Best order of markers and their positions on the map. (Duplicate markers are removed before calculation starts.)
2. Semi-graphical output that displayss a set of 'plausible' alternative maps

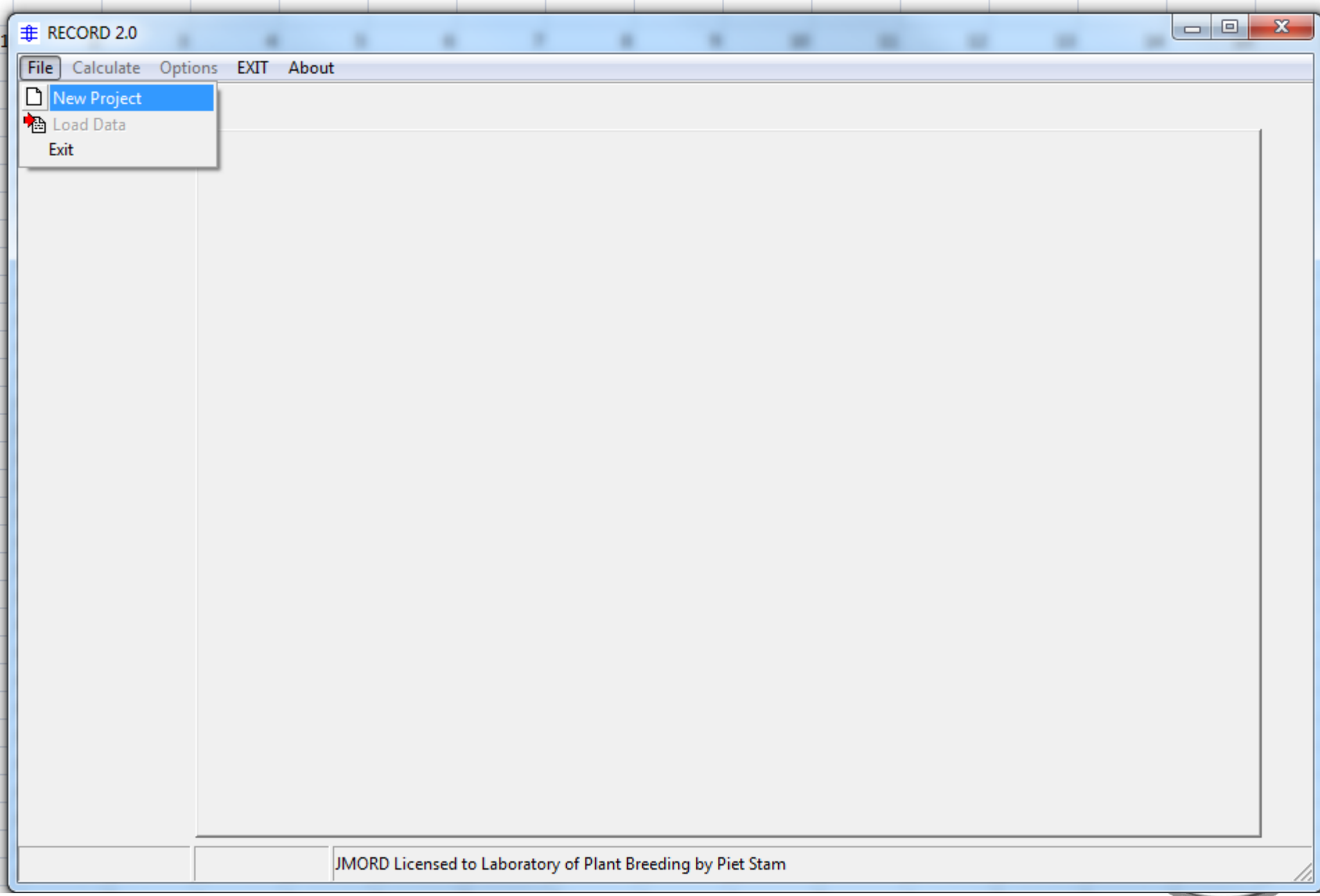
Output is stored in files that reside in a 'project directory' , to be created by the user at run time

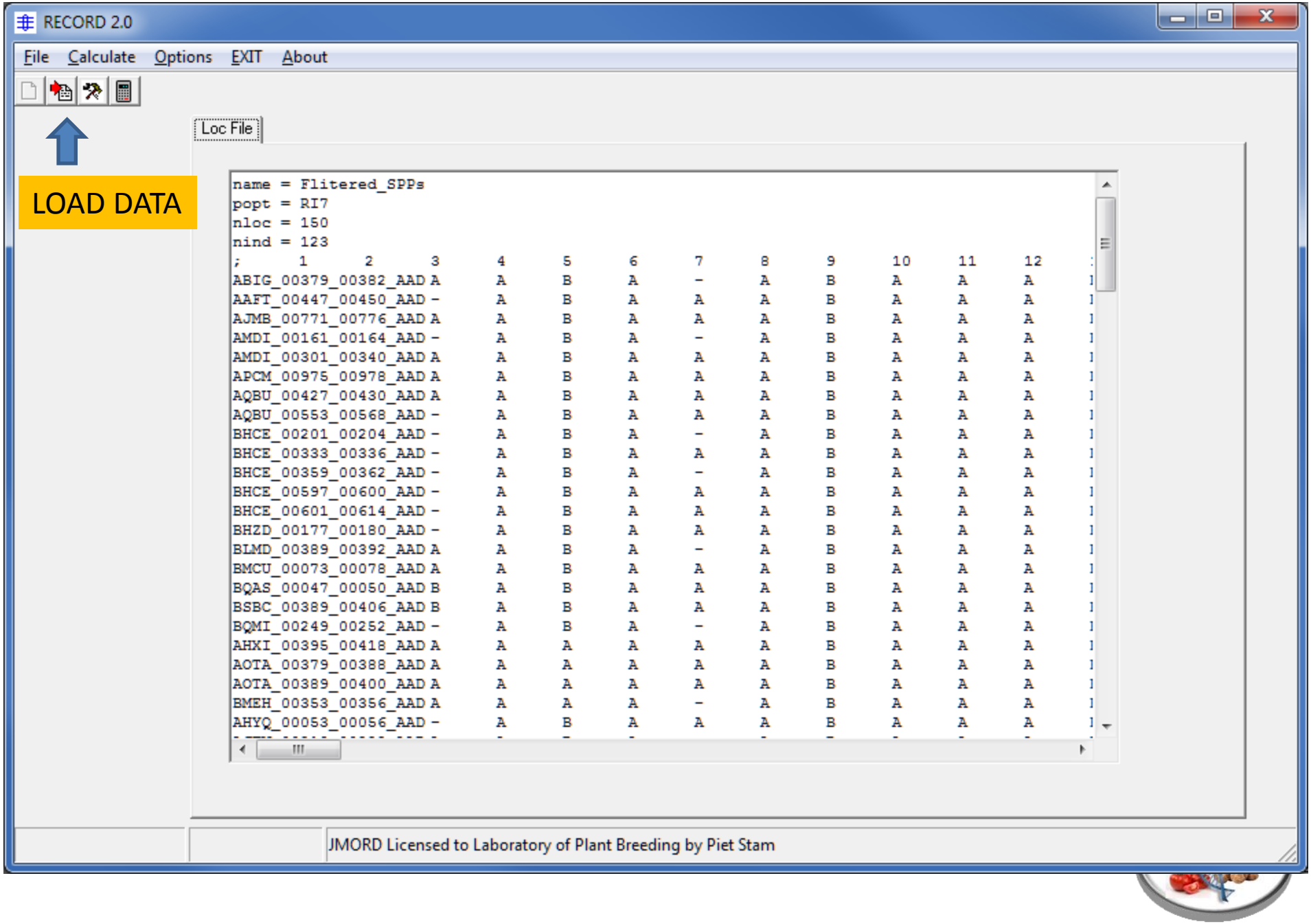
Reference: In publications that are (partially) based on this software refer to:

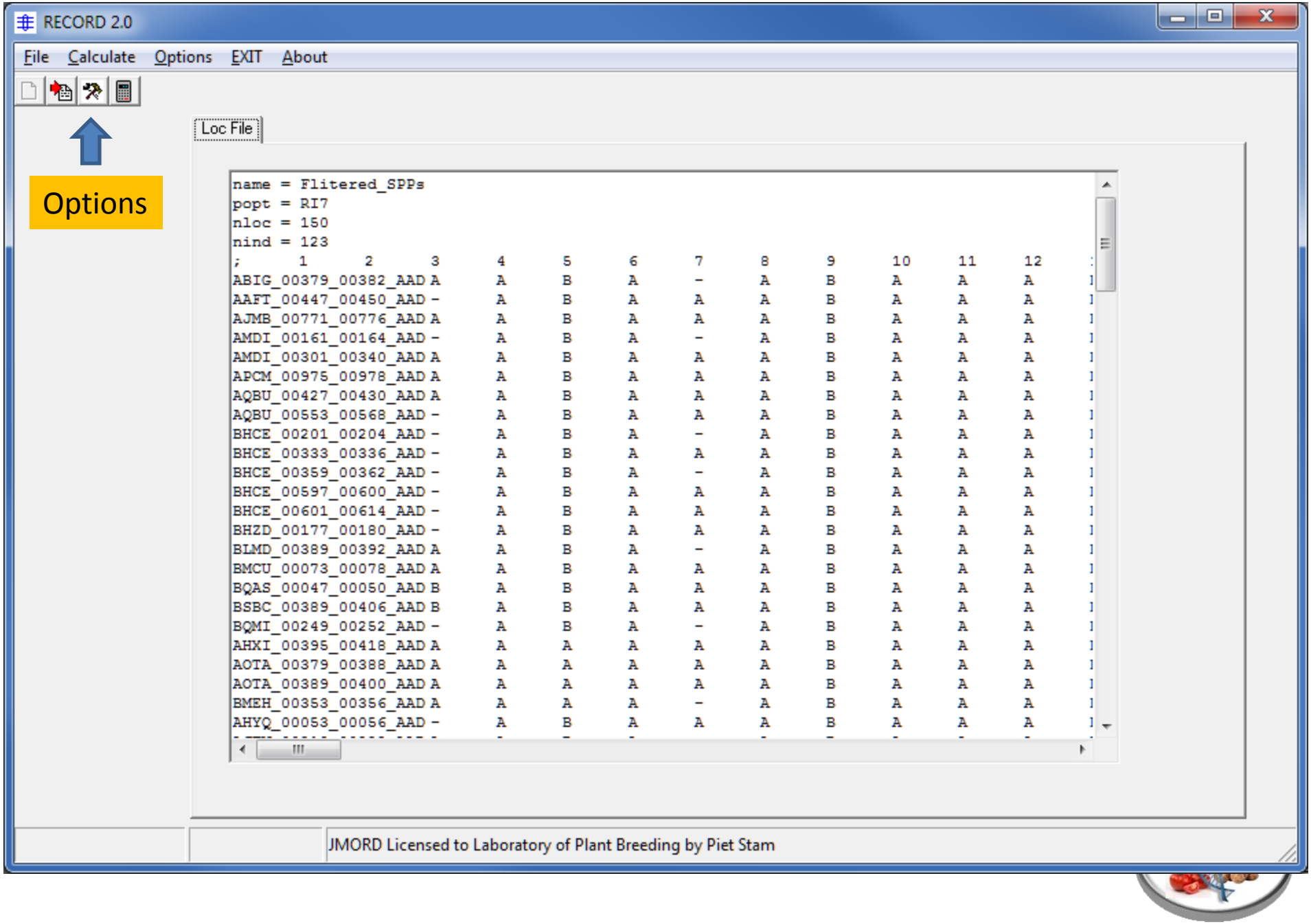
Os, van, H, P Stam, R G F Visser and H J van Eck (2005) : RECORD, a novel method for ordering loci on a genetic linkage map Theoretical and Applied Genetics 112:30-40

Click to continue

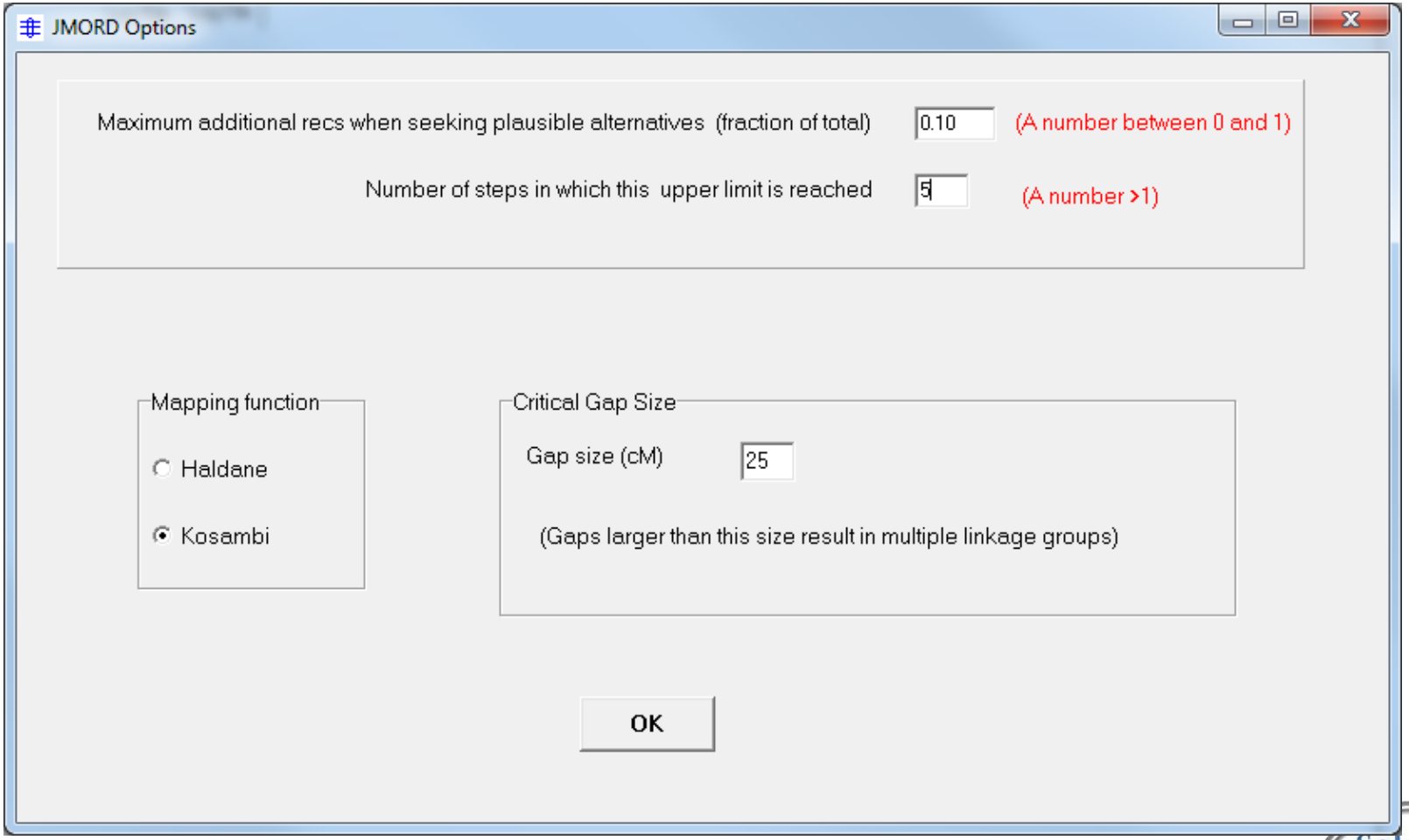
Sol CAP

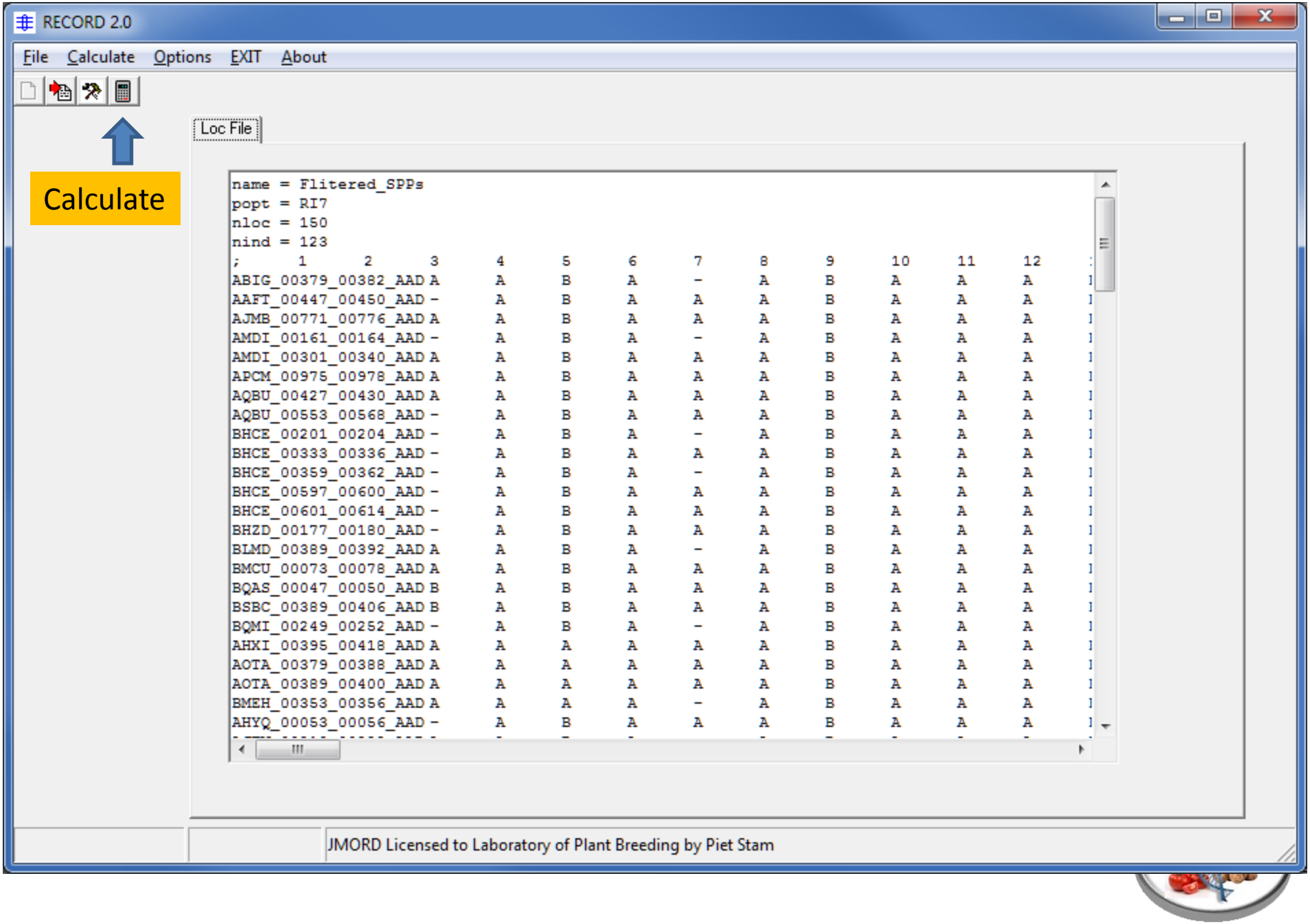

RECORD 2.0
File
Calculate
Options
EXIT
About
New Project
Load Data
Exit
JMORD Licensed to Laboratory of Plant Breeding by Piet Stam

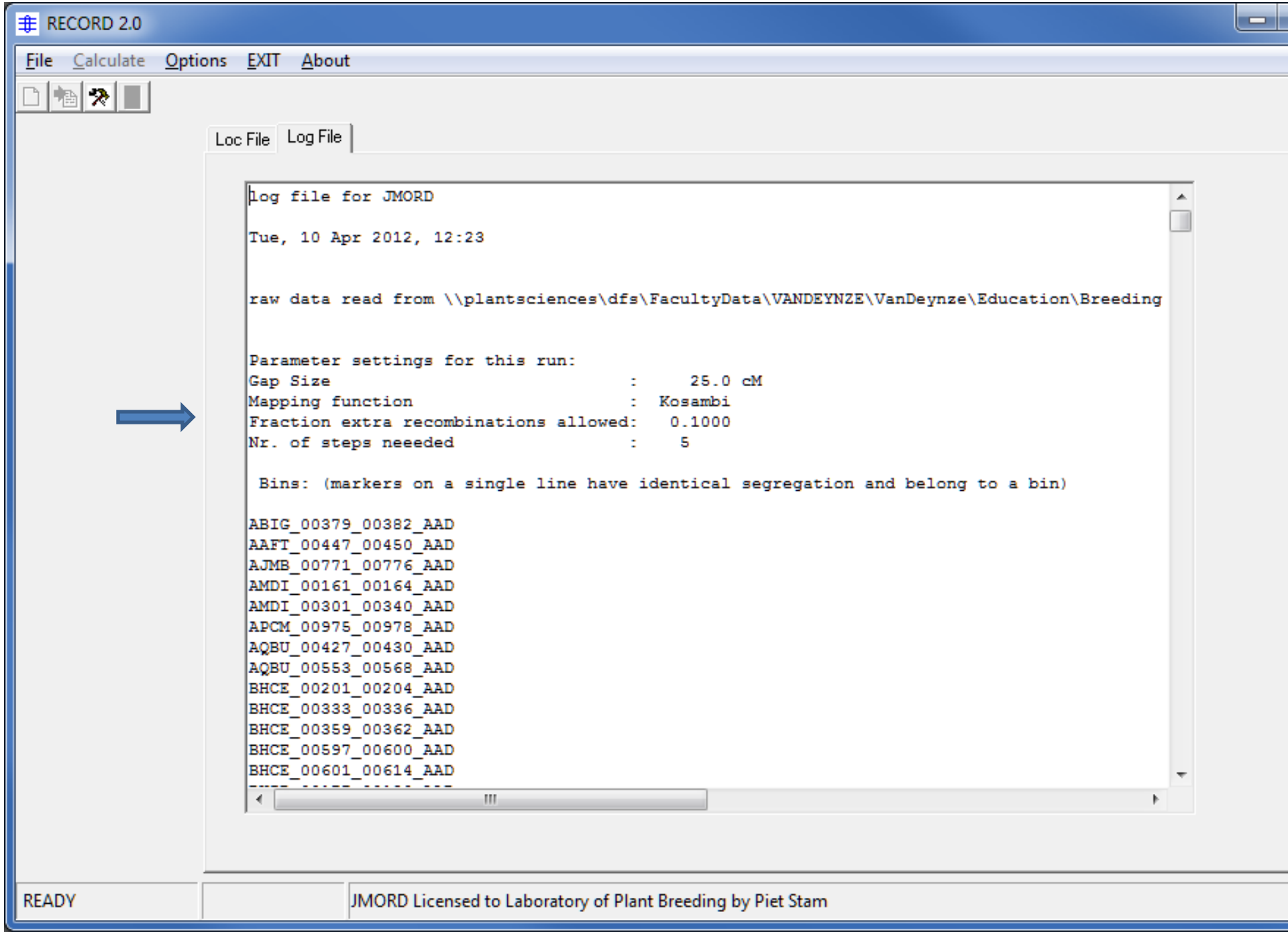

RECORD 2.0
File Calculate Options EXIT About
LOAD DATA
Loc File
name = Flitered_SPPs
popt = RI7
nloc = 150
nind = 123
; 1 2 3 4 5 6 7 8 9 10 11 12
ABIG_00379_00382_AAD A A B A - A B A A A
AAFT_00447_00450_AAD - A B A A A B A A A
AJMB_00771_00776_AAD A A B A A A B A A A
AMDI_00161_00164_AAD - A B A - A B A A A
AMDI_00301_00340_AAD A A B A A A B A A A
APCM_00975_00978_AAD A A B A A A B A A A
AQBU_00427_00430_AAD A A B A A A B A A A
AQBU_00553_00568_AAD - A B A A A B A A A
BHCE_00201_00204_AAD - A B A - A B A A A
BHCE_00333_00336_AAD - A B A A A B A A A
BHCE_00359_00362_AAD - A B A - A B A A A
BHCE_00597_00600_AAD - A B A A A B A A A
BHCE_00601_00614_AAD - A B A A A B A A A
BHZD_00177_00180_AAD - A B A A A B A A A
BLMD_00389_00392_AAD A A B A - A B A A A
BMCU_00073_00078_AAD A A B A A A B A A A
BQAS_00047_00050_AAD B A B A A A B A A A
BSBC_00389_00406_AAD B A B A A A B A A A
BQMI_00249_00252_AAD - A B A - A B A A A
AHXI_00395_00418_AAD A A A A A A B A A A
AOTA_00379_00388_AAD A A A A A A B A A A
AOTA_00389_00400_AAD A A A A A A B A A A
BMEH_00353_00356_AAD A A A A - A B A A A
AHYQ_00053_00056_AAD - A B A A A B A A A
JMORD Licensed to Laboratory of Plant Breeding by Piet Stam

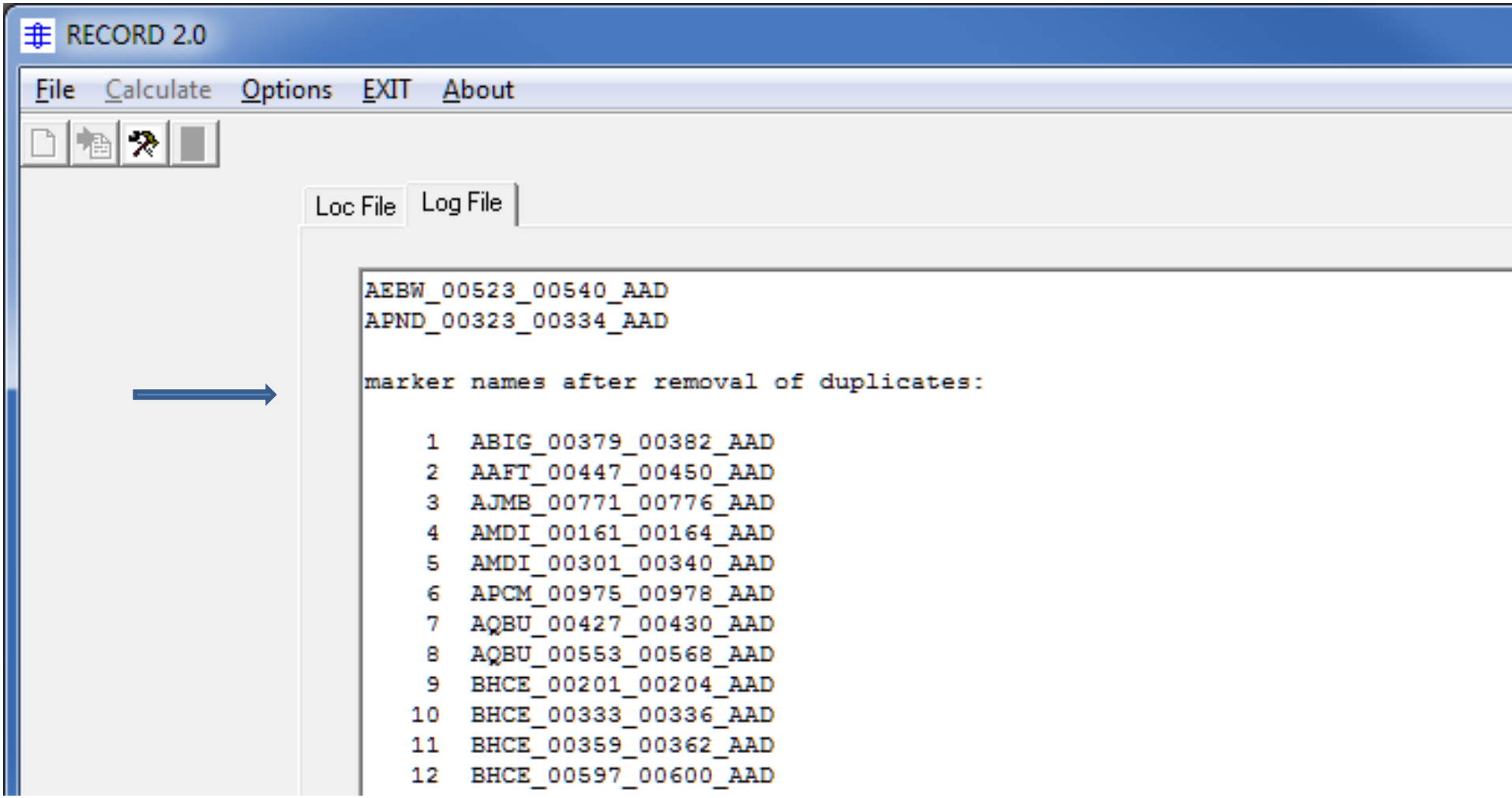

RECORD 2.0
File Calculate Options EXIT About
Options
Loc File
name = Flitered_SPPs
popt = RI7
nloc = 150
nind = 123
; 1 2 3 4 5 6 7 8 9 10 11 12
ABIG_00379_00382_AAD A A B A - A B A A A
AAFT_00447_00450_AAD - A B A A A B A A A
AJMB_00771_00776_AAD A A B A A A B A A A
AMDI_00161_00164_AAD - A B A - A B A A A
AMDI_00301_00340_AAD A A B A A A B A A A
APCM_00975_00978_AAD A A B A A A B A A A
AQBU_00427_00430_AAD A A B A A A B A A A
AQBU_00553_00568_AAD - A B A A A B A A A
BHCE_00201_00204_AAD - A B A - A B A A A
BHCE_00333_00336_AAD - A B A A A B A A A
BHCE_00359_00362_AAD - A B A - A B A A A
BHCE_00597_00600_AAD - A B A A A B A A A
BHCE_00601_00614_AAD - A B A A A B A A A
BHZD_00177_00180_AAD - A B A A A B A A A
BLMD_00389_00392_AAD A A B A - A B A A A
BMCU_00073_00078_AAD A A B A A A B A A A
BQAS_00047_00050_AAD B A B A A A B A A A
BSBC_00389_00406_AAD B A B A A A B A A A
BQMI_00249_00252_AAD - A B A - A B A A A
AHXI_00395_00418_AAD A A A A A A B A A A
AOTA_00379_00388_AAD A A A A A A B A A A
AOTA_00389_00400_AAD A A A A A A B A A A
BMEH_00353_00356_AAD A A A A - A B A A A
AHYQ_00053_00056_AAD - A B A A A B A A A
JMORD Licensed to Laboratory of Plant Breeding by Piet Stam

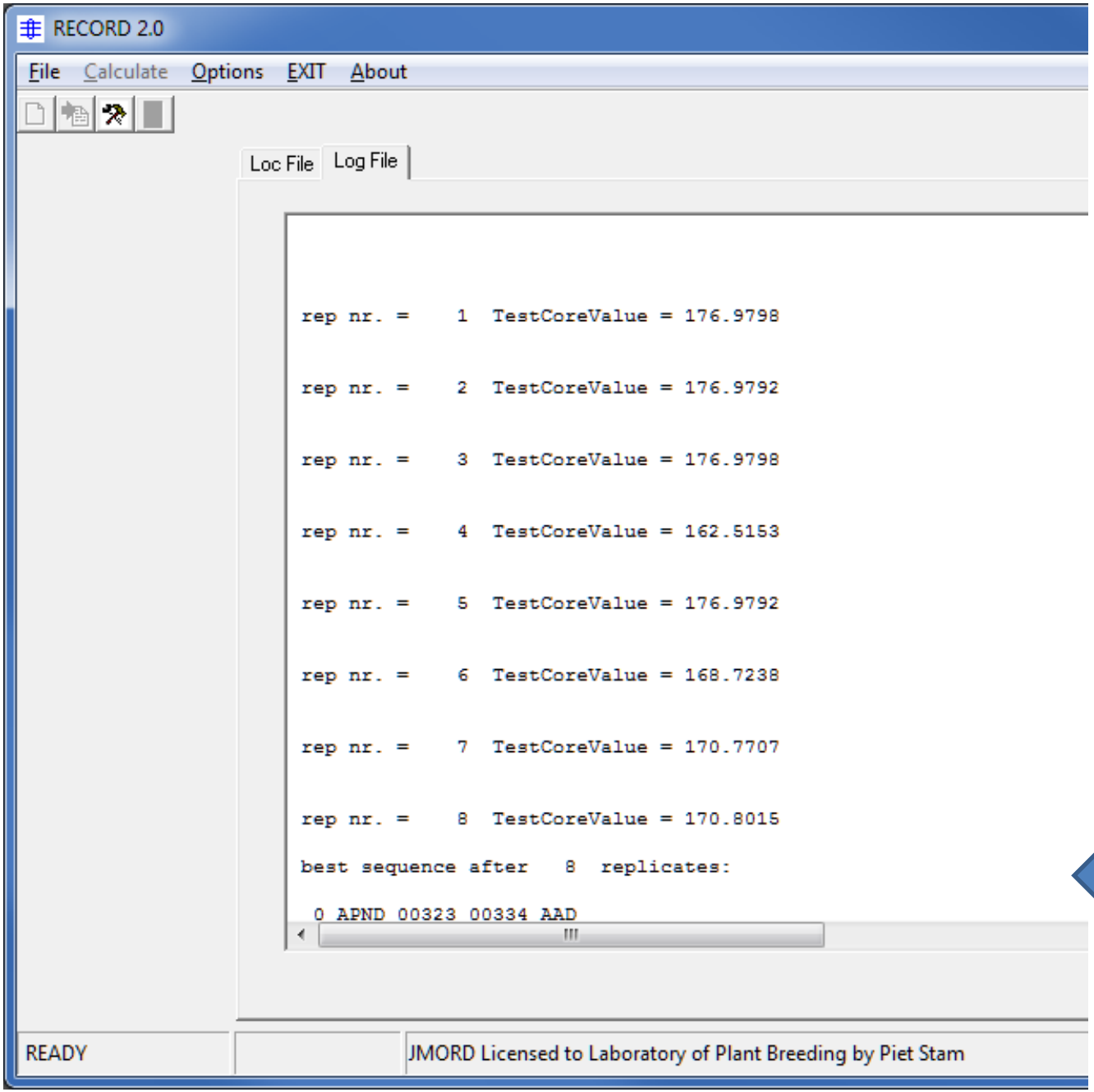
Options in WinRecord




RECORD 2.0
File Calculate Options EXIT About
Calculate
Loc File
name = Flitered_SPPs
popt = RI7
nloc = 150
nind = 123
; 1 2 3 4 5 6 7 8 9 10 11 12
ABIG_00379_00382_AAD A A B A - A B A A A
AAFT_00447_00450_AAD - A B A A A B A A A
AJMB_00771_00776_AAD A A B A A A B A A A
AMDI_00161_00164_AAD - A B A - A B A A A
AMDI_00301_00340_AAD A A B A A A B A A A
APCM_00975_00978_AAD A A B A A A B A A A
AQBU_00427_00430_AAD A A B A A A B A A A
AQBU_00553_00568_AAD - A B A A A B A A A
BHCE_00201_00204_AAD - A B A - A B A A A
BHCE_00333_00336_AAD - A B A A A B A A A
BHCE_00359_00362_AAD - A B A - A B A A A
BHCE_00597_00600_AAD - A B A A A B A A A
BHCE_00601_00614_AAD - A B A A A B A A A
BHZD_00177_00180_AAD - A B A A A B A A A
BLMD_00389_00392_AAD A A B A - A B A A A
BMCU_00073_00078_AAD A A B A A A B A A A
BQAS_00047_00050_AAD B A B A A A B A A A
BSBC_00389_00406_AAD B A B A A A B A A A
BQMI_00249_00252_AAD - A B A - A B A A A
AHXI_00395_00418_AAD A A A A A A B A A A
AOTA_00379_00388_AAD A A A A A A B A A A
AOTA_00389_00400_AAD A A A A A A B A A A
BMEH_00353_00356_AAD A A A A - A B A A A
AHYQ_00053_00056_AAD - A B A A A B A A A
JMORD Licensed to Laboratory of Plant Breeding by Piet Stam


RECORD 2.0
File
Calculate
Options
EXIT
About
Loc File
Log File
log file for JMORD
Tue, 10 Apr 2012, 12:23
raw data read from \\plantsciences\dfs\FacultyData\VANDEYNZE\VanDeynze\Education\Breeding
Parameter settings for this run:
Gap Size : 25.0 cM
Mapping function : Kosambi
Fraction extra recombinations allowed: 0.1000
Nr. of steps neeeded : 5
Bins: (markers on a single line have identical segregation and belong to a bin)
ABIG_00379_00382_AAD
AAFT_00447_00450_AAD
AJMB_00771_00776_AAD
AMDI_00161_00164_AAD
AMDI_00301_00340_AAD
APCM_00975_00978_AAD
AQBU_00427_00430_AAD
AQBU_00553_00568_AAD
BHCE_00201_00204_AAD
BHCE_00333_00336_AAD
BHCE_00359_00362_AAD
BHCE_00597_00600_AAD
BHCE_00601_00614_AAD
READY
JMORD Licensed to Laboratory of Plant Breeding by Piet Stam


RECORD 2.0
File
Calculate
Options
EXIT
About
Loc File
Log File
AEBW_00523_00540_AAD
APND_00323_00334_AAD
marker names after removal of duplicates:
1 ABIG_00379_00382_AAD
2 AAFT_00447_00450_AAD
3 AJMB_00771_00776_AAD
4 AMDI_00161_00164_AAD
5 AMDI_00301_00340_AAD
6 APCM_00975_00978_AAD
7 AQBU_00427_00430_AAD
8 AQBU_00553_00568_AAD
9 BHCE_00201_00204_AAD
10 BHCE_00333_00336_AAD
11 BHCE_00359_00362_AAD
12 BHCE_00597_00600_AAD


RECORD 2.0
File
Calculate
Options
EXIT
About
Loc File
Log File
rep nr. = 1 TestCoreValue = 176.9798
rep nr. = 2 TestCoreValue = 176.9792
rep nr. = 3 TestCoreValue = 176.9798
rep nr. = 4 TestCoreValue = 162.5153
rep nr. = 5 TestCoreValue = 176.9792
rep nr. = 6 TestCoreValue = 168.7238
rep nr. = 7 TestCoreValue = 170.7707
rep nr. = 8 TestCoreValue = 170.8015
best sequence after 8 replicates:
0 APND 00323 00334 AAD
READY
JMORD Licensed to Laboratory of Plant Breeding by Piet Stam

Sol CAP

jmord - Notepad

File Edit Format View Help

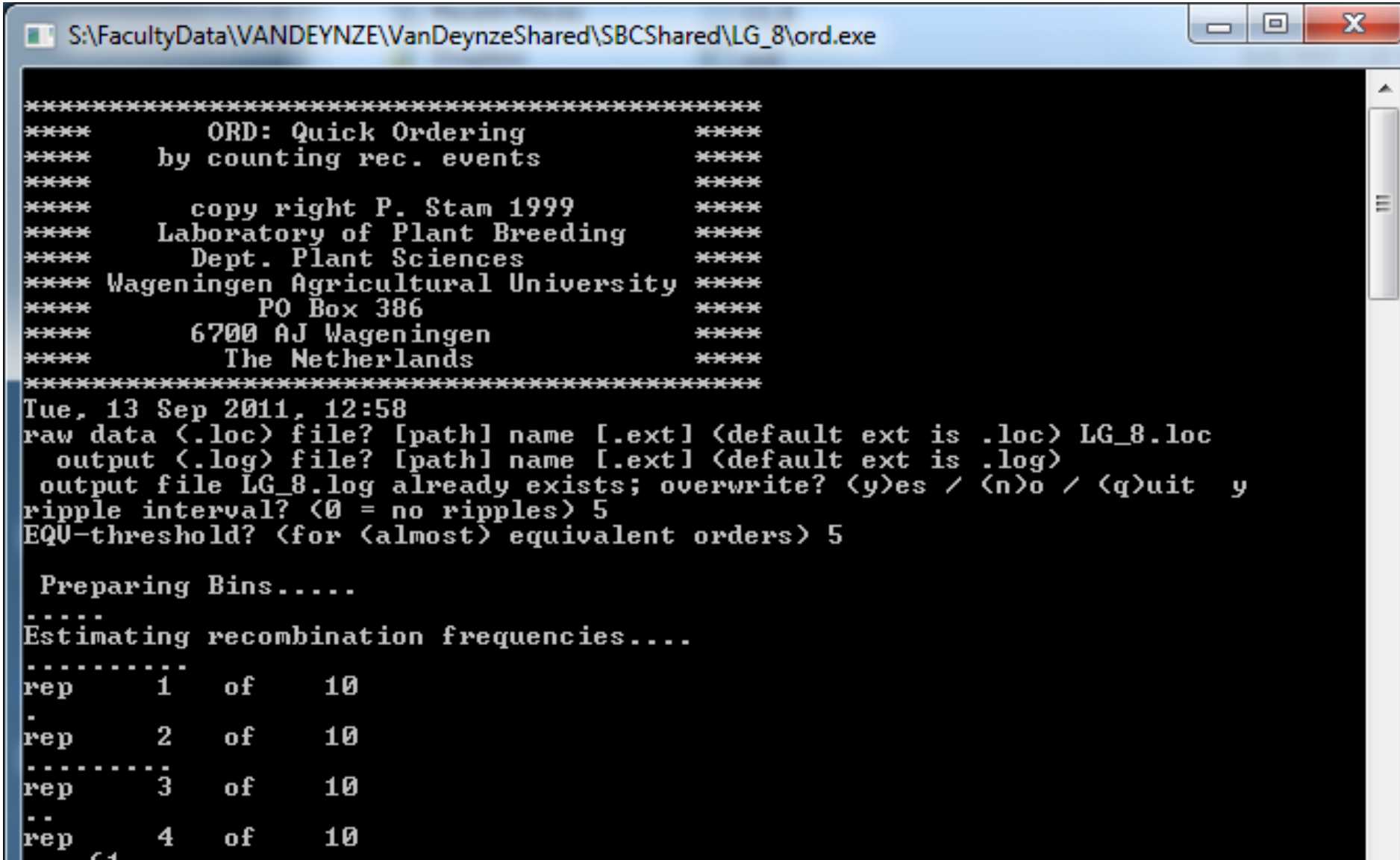
```
best sequence after   8  replicates:

 0 APND_00323_00334_AAD
 1 ACCM_00365_00374_AAD
 2 BGNG_00275_00282_AAD
 3 AEBW_00523_00540_AAD
 4 ACCM_00375_00382_AAD
 5 BGNG_00261_00274_AAD
 6 BQHV_00391_00414_AAD
 7 ABWO_00273_00276_AAD
 8 BQHV_00415_00420_AAD
 9 ABWO_00067_00070_AAD
10 AIPG_00121_00124_AAD
11 BDBZ_00121_00142_AAD
12 AONS_00509_00512_AAD
13 AFWK_00477_00480_AAD
14 AQRR_00085_00098_AAD
15 AQRR_00377_00384_AAD
```

Order

```
141 AOTA_00389_00400_AAD
142 BMCU_00073_00078_AAD
143 APCM_00975_00978_AAD
144 AMDI_00161_00164_AAD
145 BQMI_00249_00252_AAD
146 BHCE_00201_00204_AAD
147 BHZD_00177_00180_AAD
148 BSBC_00389_00406_AAD
149 BIFE_00379_00382_AAD

count = 162.5153


map in JoinMap format (up & down)

group UP
APND_00323_00334_AAD    0.000
ACCM_00365_00374_AAD    0.417
BGNG_00275_00282_AAD    0.417
AEBW_00523_00540_AAD    0.417
ACCM_00375_00382_AAD    0.417
BGNG_00261_00274_AAD    0.418
BQHV_00391_00414_AAD    3.461
ABWO_00273_00276_AAD    3.461
```

Distance

DOS version the same but faster
(No distances though)



Checkmatrix: Requirements

CheckMatrix **py_matrix_2D_V248_RECBIT.py** takes as input three files:

1. Pairwise Distance Matrix File madmapper_test_small.out.pairs_all
pairwise distance matrix file can be generated by Python MadMapper from **Locus File**

```
.....................
GM01    GM07    0.36
GM01    GM08    0.40
GM01    GM09    0.48
GM01    GM10    0.52
GM01    GM11    0.60
GM01    GM12    0.68
GM02    GM01    0.04
GM02    GM02    0.00
GM02    GM03    0.08
GM02    GM04    0.16
GM02    GM05    0.20
GM02    GM06    0.24
.....................
```

Pairwise Distance matrices from MadMapper

- ***Alexander Kozik, UC Davis Genome Center***
- http://cgpdb.ucdavis.edu/XLinkage/MadMapper/

2. Genetic Map File madmapper_test_small.map.right
(on this example last column reflects the order markers)

```
LG      GM01     0
LG      GM02     1
LG      GM03     2
LG      GM04     3
LG      GM05     4
LG      GM06     5
LG      GM07     6
LG      GM08     7
LG      GM09     8
LG      GM10     9
LG      GM11    10
LG      GM12    11
```

From Record (reformatted)

3. Locus File (Raw Marker Scores) madmapper_test_small.loc

```
        1               10                  20        25
        |               |                   |         |
GM01    A A A A A A A A A A A A A A A B B B B B B B B B
GM02    A A A A A A A A A A A A A A A B B B B B B B B B
GM03    A A A A A A A A A A A A A A B B B B B B B B B B
GM04    A A A A A A A A A A A A B B B B B B B B B B B B
GM05    A A A A A A A A A A B B B B B B B B B B B B B B
GM06    A A A A A A A A A B B B B B B B B B B B B B B B
GM07    A A A A A A A A A B B B B B B B B B B B B B A A
GM08    A A A A A A A A A B B B B B B B B B B B B A A A
GM09    A A A A A A A A A B B B B B B B B B B A A A A A
GM10    B A A A A A A A A B B B B B B B B B B A A A A A
GM11    B B A A A A A A A B B B B B B B B A A A A A A A
GM12    B B B A A A A A A A B B B B B B A A A A A A A A
```

Loc File (input for Record)

Visualization and validation of genetic maps using 2-dimensional heat plots using Check Matrix program

http://cgpdb.ucdavis.edu/XLinkage/Genetic_Map_Matrix_Plot_Art.html

Run from MadMapper on LG_8 to get this file

PROGRAM USAGE

Prompt> python py_matrix_2D_V248_RECBIT.py LG_8.out.pairs_all LG_8ord LG_8ord.out X Y LG_8.loc REC NOGRAPH 0.9 SMALL RIL

Right Order Map from WinRecord

Output file name

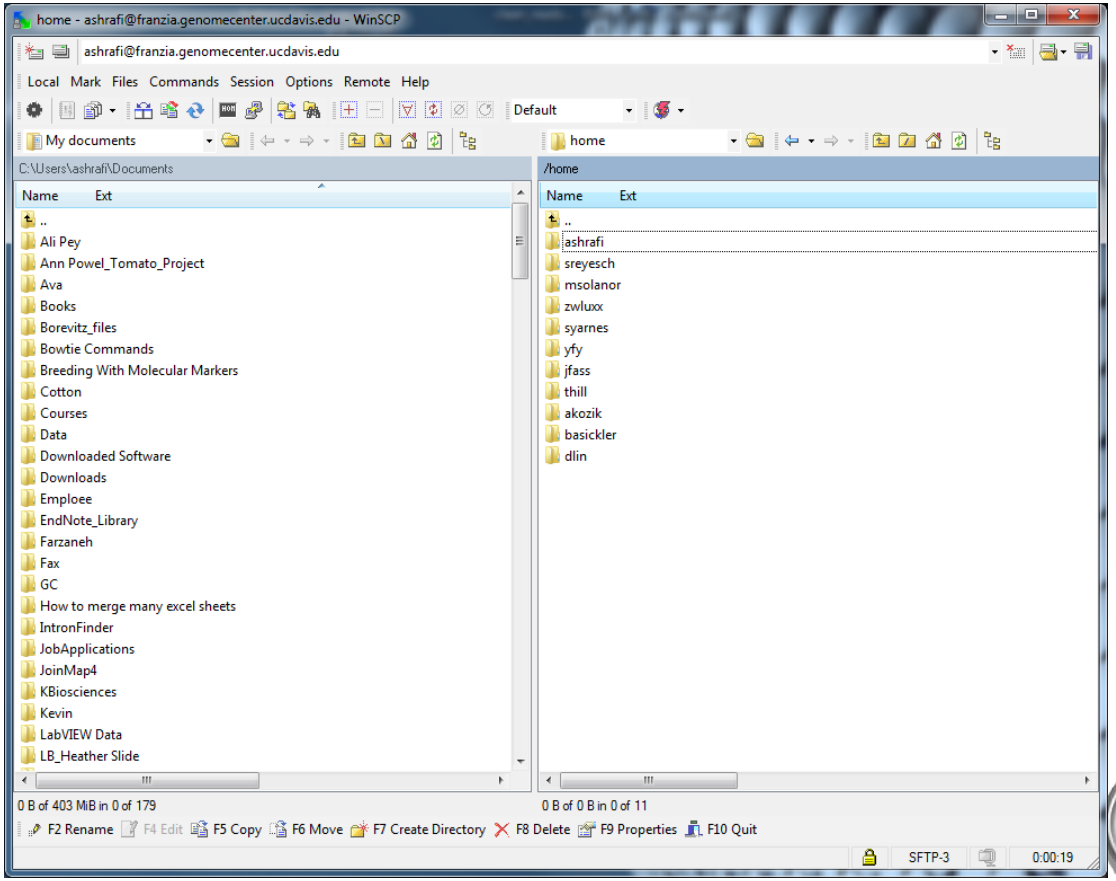
py_matrix_2D_V248_RECBIT.py takes as input **matrix file**, **map file** and three optional files: **framework markers list**, **list of IDs to highlight in red** and ***.loc file with recombination data** **(raw marker scores)**.

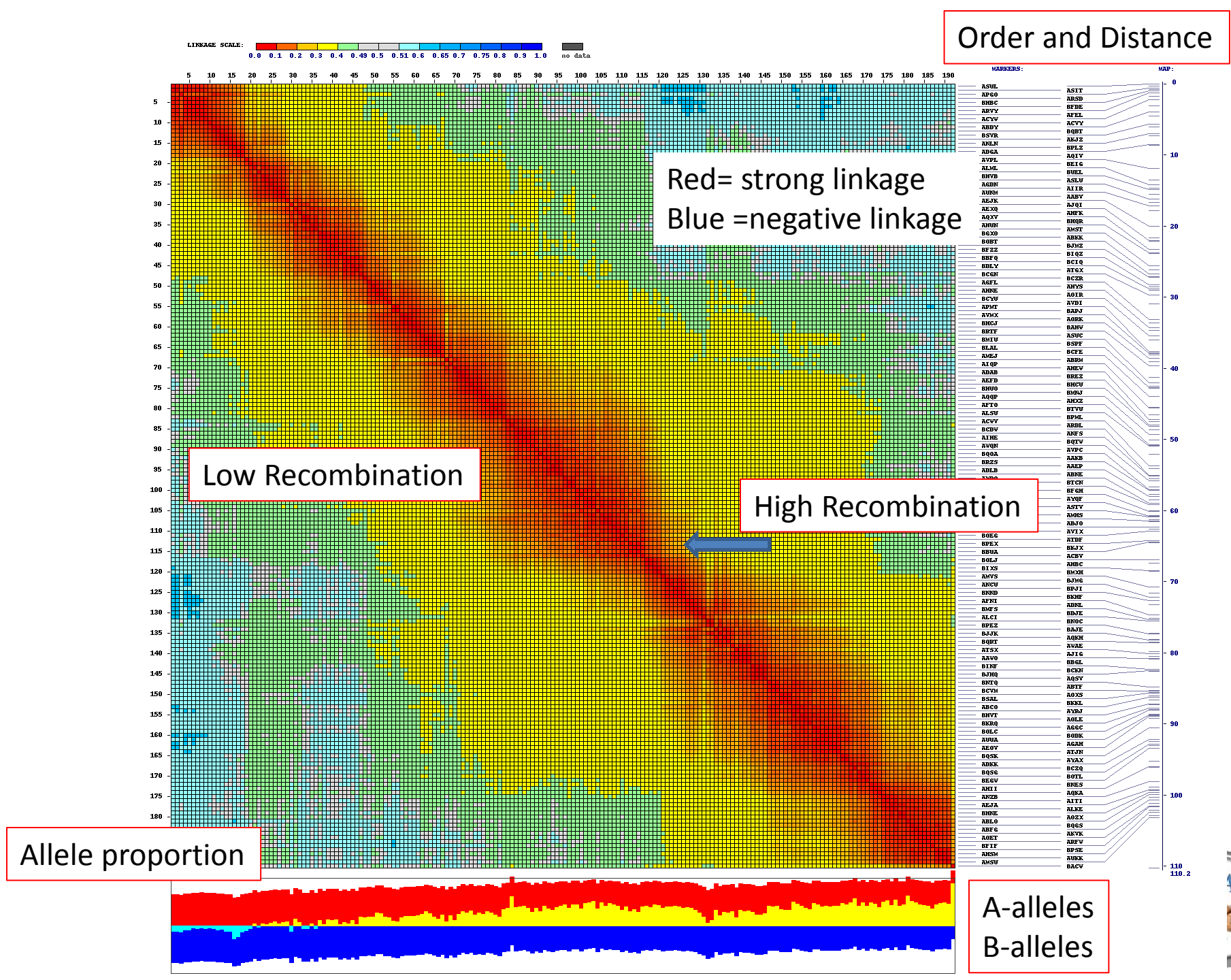
[matrix_file] [map_file] [output_file] [frame_marker_list] [red_list] [loc_file] [REC/BIT/LOD] [GRAPH_OPT] [LINK_CUT] [LARGE/SMALL] [Popn_TYPE]

Transfer PNG files from Unix to windows or Mac OS system.

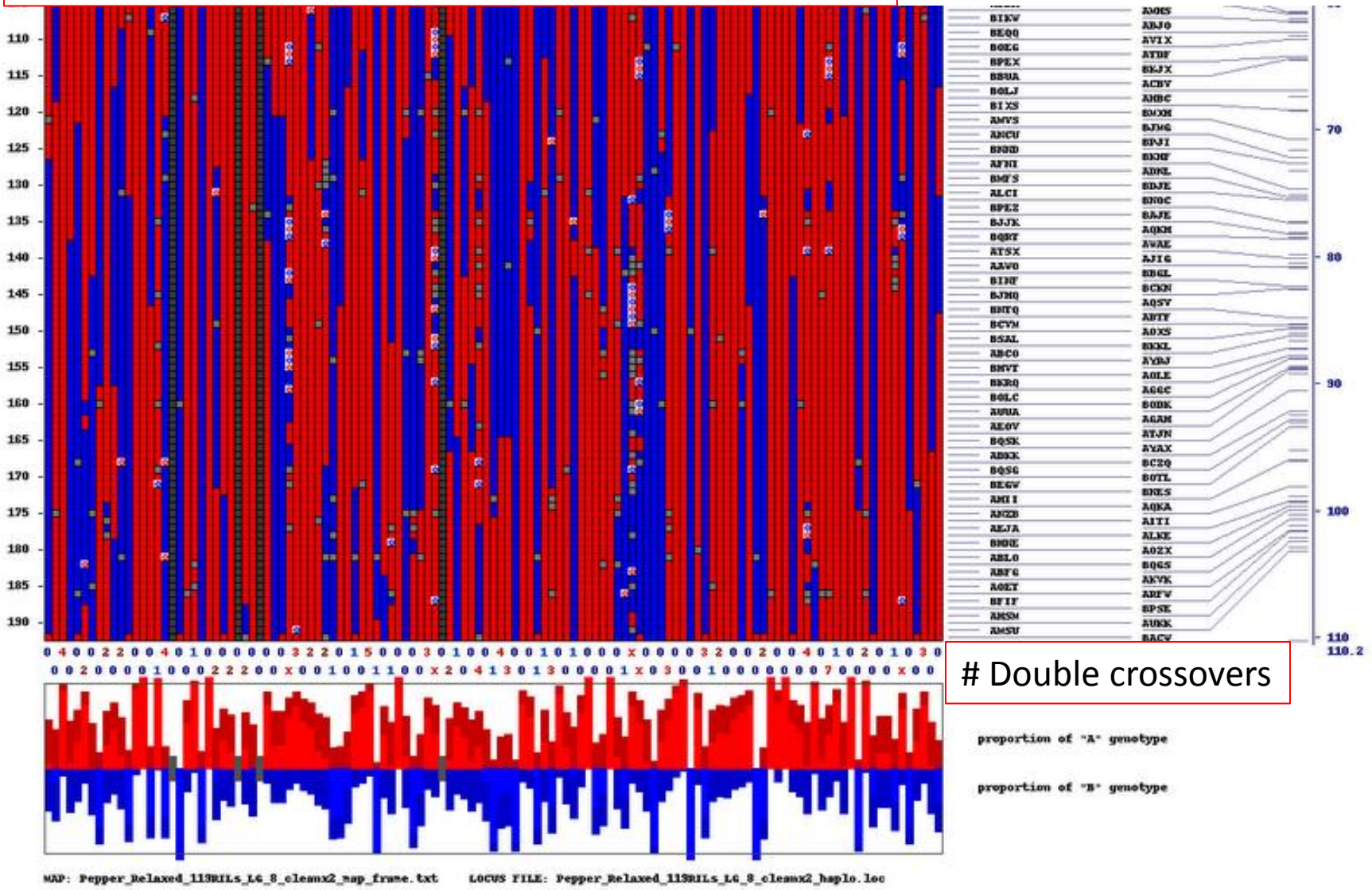
- WinSCP connects the unix server to windows environment.

http://winscp.net/eng/download.php




Order and Distance
LINKAGE SCALE:
0.0 0.1 0.2 0.3 0.4 0.49 0.5 0.51 0.6 0.65 0.7 0.75 0.8 0.9 1.0
no data
MARKERS:
MAP:
Red= strong linkage
Blue =negative linkage
Low Recombination
High Recombination
Allele proportion
A-alleles
B-alleles

Double crosser- White, Grey Missing data



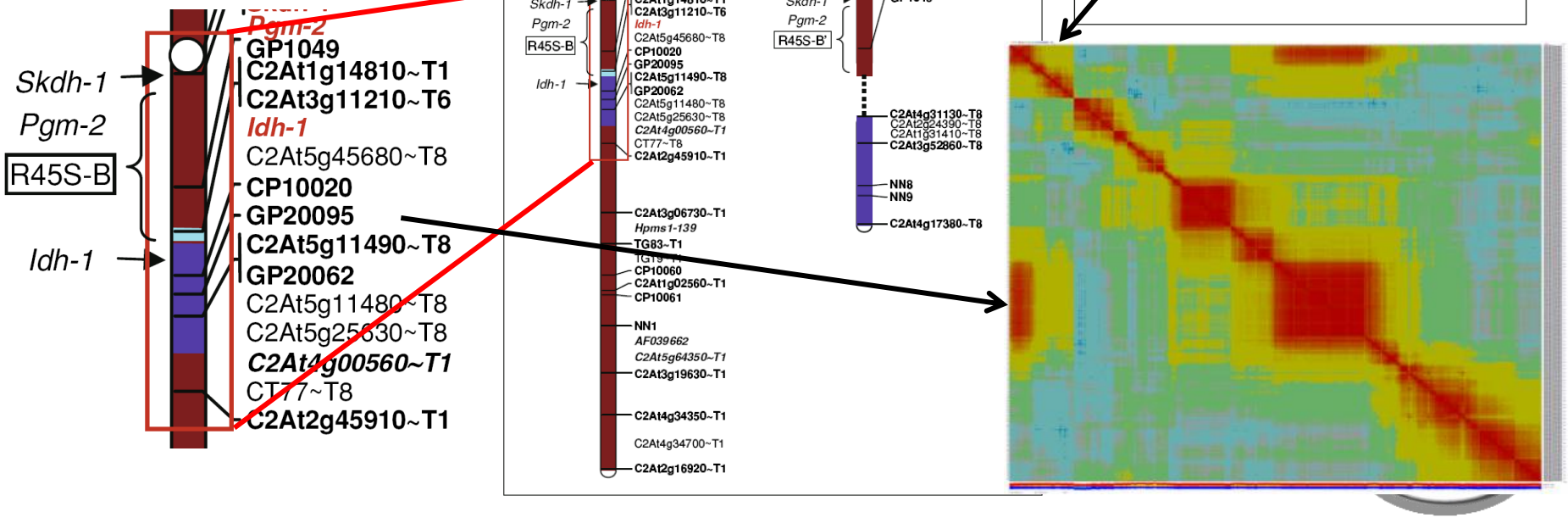
Double crossovers

Wu et al 2009 COSII F2 map

"Pseudolinkage of loci near breakpoints in reciprocal translocation."

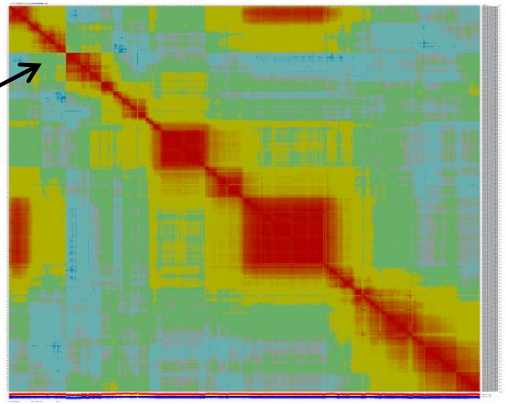
Markers linked to P1-wild have weak linkage to P8-wild

P8-wild stands as a separate linkage group.

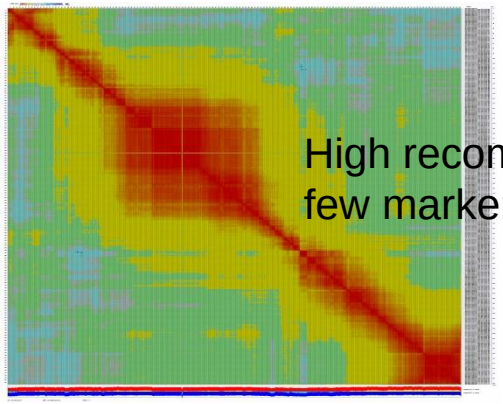


Pepper Interspecific map

LG1 = 1239 SPPs

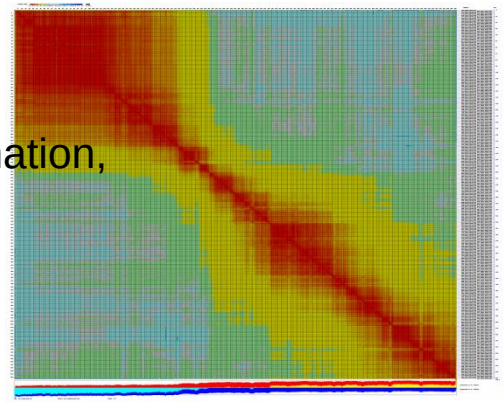


LG2 = 753 SPPs

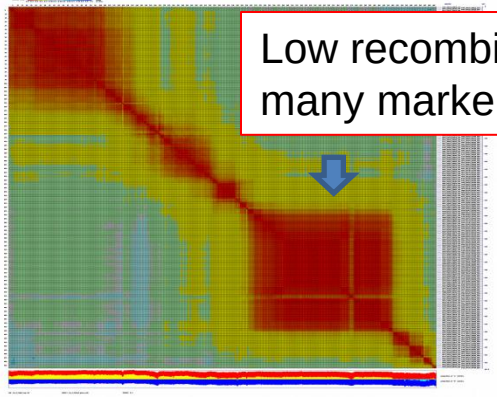


High recombination,
few markers

LG3 = 592 SPPs

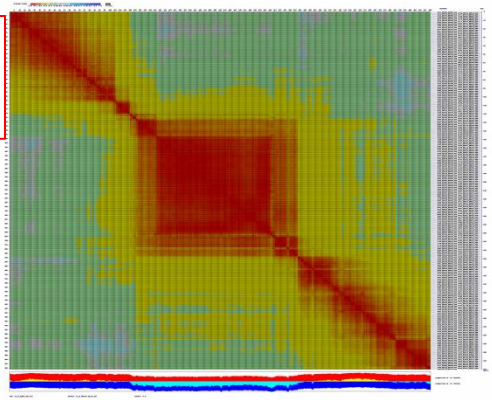


LG4 = 468 SPPs

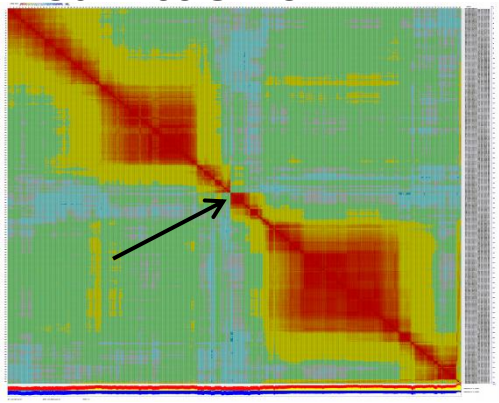


Low recombination,
many markers

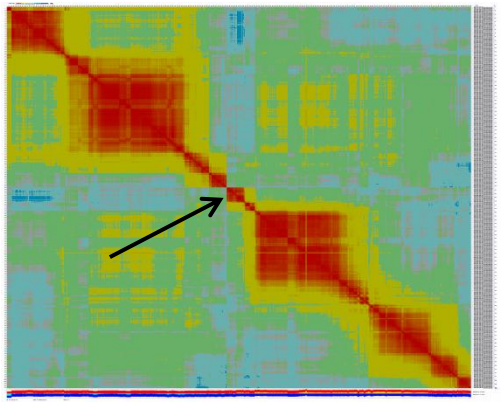
LG5 = 421 SPPs



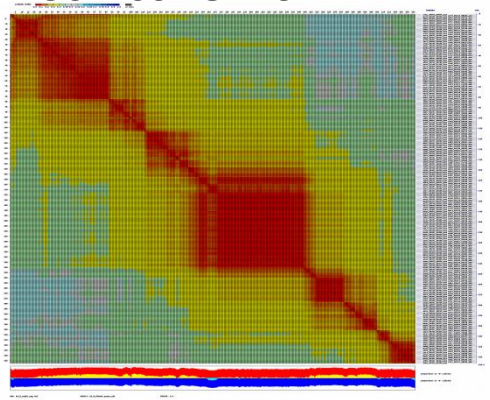
LG6 = 753 SPPs



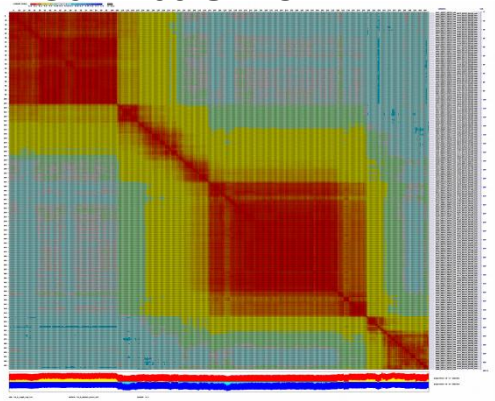
LG7 = 1022 SPPs

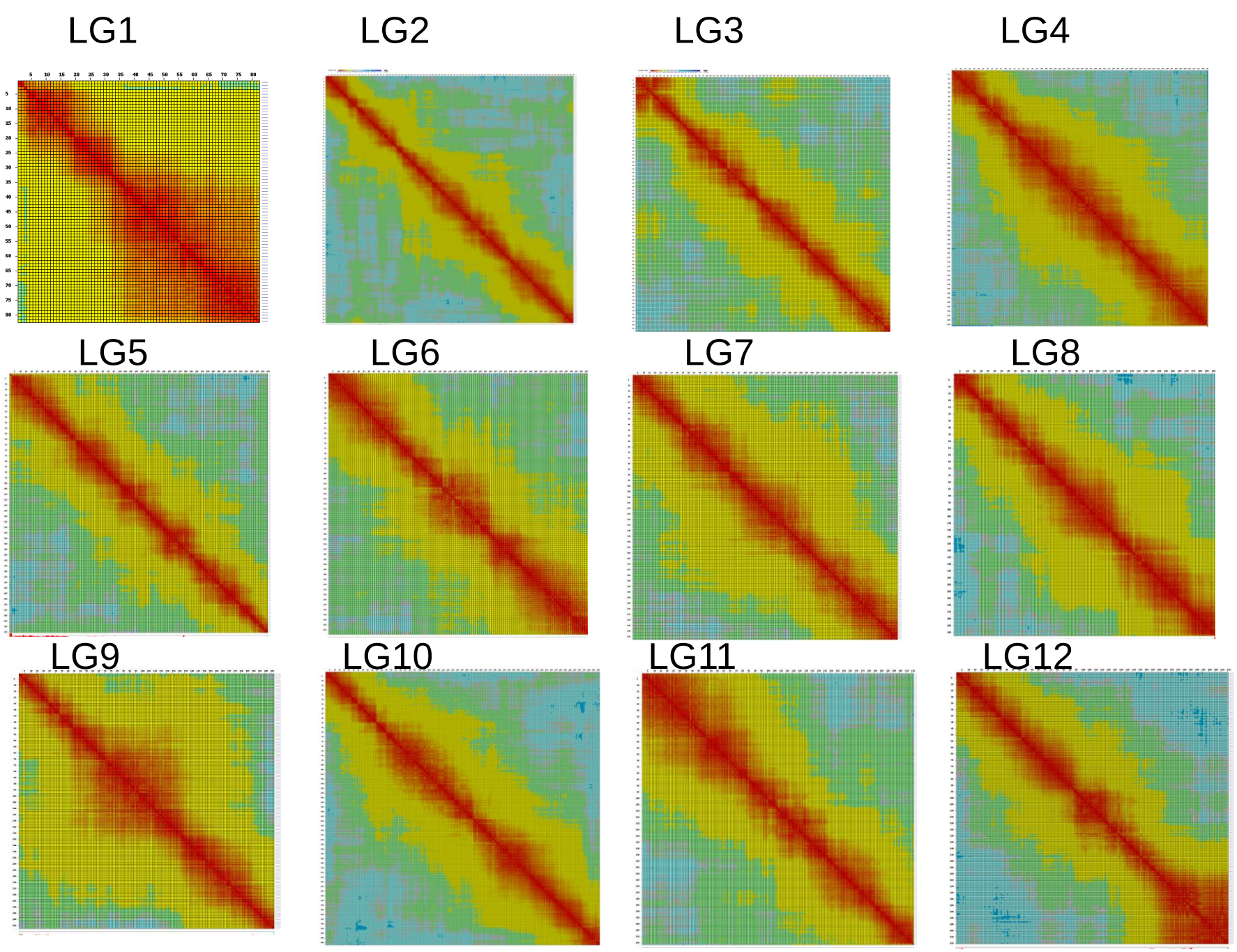


LG8 = 337 SPPs



LG9 = 409 SPPs





C. Frutescens x C. annuum. 12 Linkage Groups / 12 Chromosomes, 2,886 bins

Summary

- Group with MadMapper
- Define each linkage groups-Vlookup
- Create paired matrices file
- Determine order with RECORD
- Visualize with Checkmatrix

Thank you

Please fill out the survey evaluation! (You will be contacted via email)

Today's presentation available at:
http://www.extension.org/pages/63330

Sign up for PBG News
http://pbgworks.org

Watch past webinars and sign up for future webinars!
http://www.extension.org/pages/60426