Software for 50K SNP high density linkage map construction in Maize

From the article "Genome-wide landscape of recombination reveals intergenic recombination contributes to interspecific recombination variation"

Qingchun Pan, Yan's lab

2014/12/22

Correspondence should be addressed to (panqingchunmice@163.com),

J.Y. (yjianbing@mail.hzau.edu.cn)

Download from: www.maizego.org/Resources.html

1. To install prerequisite softwares

The high density linkage map construction need two softwares. One is the perl environment (version 5.01 or upper) and another is *carthagene* (version 1.2 or upper) are needed to be installed properly in linux system.

The carthagene software for linux system could be downloaded from the website http://www7.inra.fr/mia/T/CarthaGene/.

After installed the former 2 softwares "perl" and "*carthagene"*, you can download the perl script file "map.zip" from the webset http://www.maizego.org/Resources.html and decompression the file "map.zip " in your linux system.

2. To prepare two input files

For example:

One file is "KBY_final.csv". "KBY" is the pop name (the details was shown in 4), the first column of this file is the marker name of 50k data, the second column is the parent-one genotype, the third column is the parent-two genotype,the fourth and following columns are the genotypes of segregating family lines. This file is the standard file which was exported from the illumina goldgate 50k SNP chip software Beadstudio. The sample file format is as follows:

	12B480-M	12B481-M	KBY1	KBY2	KBY3	KBY4
abph1.15	AA	AA	AA	AA	AA	AA
abph1.22	AA	AA	AA	AA	AA	AA
ae1.4	GG	GG	GG	GG	GG	GG
ae1.8	GG	GG	GG	GG	GG	GG
an1.3	CC	CC	CC	CC	CC	CC
ba1.5	AA	AA	AA	AA	AA	AA
ba1.6	GG	GG	GG	GG	GG	GG
ba1.7	GG	GG	GG	GG	GG	GG
bt-000002(	AA	GG	GG	GG	AA	GG
bt2.2	GG	AA	AA	AA	GG	AA
bt2.4	GG	AA	AA	AA	GG	AA
bt2.5	GG	GG	GG	GG	GG	GG
bt2.8	AA	AA	AA	AA	AA	AA
csu1138.3	AA	GG	GG	AA	GG	AA
csu1138.4	GG	AA	AA	GG	AA	GG
csu1171.2	AA	GG	AG	GG	AA	GG
Fea2.2	GG	GG	GG	GG	GG	GG
fea2.3	AA	AA	AA	AA	AA	AA
Fea2.5	AA	AA	AA	AA	AA	AA
lg2.12	GG	GG	GG	GG	GG	GG

Another file is "chr_info.csv" which contains maize 50k chip chromosome and physical position information . It has five columns, first column is the SNP id, second column is the marker name, the third column is the chromosome information, the fourth column is the physical position and the fifth is the marker in gene annotation information.

SNP	Name	Best_hit	SNP_posi.	Annotation
SNP54	PHM14519.8	chr0	0	
SNP79	PHM2306.39	chr0	0	
SNP93	PHM3630.8	chr0	0	
SNP107	PHM4955.12	chr0	0	
SNP144	PUT-163a-101389210-16	chr0	0	
SNP147	PUT-163a-101392844-22	chr0	0	
SNP243	PUT-163a-13515860-232	chr0	0	
SNP244	PUT-163a-13515860-233	chr0	0	
SNP245	PUT-163a-13515860-234	chr0	0	
SNP246	PUT-163a-13515860-235	chr0	0	
SNP252	PUT-163a-13516046-244	chr0	0	
SNP297	PUT-163a-148928654-351	chr0	0	
SNP299	PUT-163a-148929070-368	chr0	0	
SNP313	PUT-163a-148937637-425	chr0	0	
SNP334	PUT-163a-148942702-462	chr0	0	
SNP340	PUT-163a-148943685-476	chr0	0	
SNP445	PUT-163a-148991135-703	chr0	0	
SNP586	PUT-163a-16919993-1041	chr0	0	
SNP669	PUT-163a-18163661-1261	chr0	0	
SNP670	PUT-163a-18163661-1262	chr0	0	
SNP687	PUT-163a-18167719-1296	chr0	0	
SNP718	PUT-163a-18172808-1387	chr0	0	
SNP752	PUT-163a-18179900-1492	chr0	0	
SNP754	PUT-163a-18180613-1496	chr0	0	
SNP814	PUT-163a-28982443-1665	chr0	0	

The two files are delimited by a comma, rather than the tab.

3.To run the software

Now, assuming that you save the two input files "KBY_final.csv" and "chr_info.csv" was put in the " fileone " documentary folder. Infact, you could download the "KBY_final.csv" and "chr_info.csv" from our website http://www.maizego.org/Resources.html and test the data.

First, you need to decompress the map.zip file in the directory whereever you like, and enter into the decompress "map_file" documentary folder by linux systerm terminal,after that you run the command as follow:

"perl new_perl/meta.pl ~/fileone/KBY_final.csv KBY ~/fileone/chr_info.csv".

This command will take about ten hours to finish all the computation. When running is done, you should enter the documentary folder "map_file\new_perl\zhenghe" by linux terminal.

and run the another command "perl meta_zhenghe.pl ~/fileone/ KBY". Two result files will be generated.

You will get the final map data, one is "KBY_map_all.csv", the other is "KBY_map_all_all.csv" in the"fileone" documentary folder. The “KBY_map_all.csv” is the bin map of the genetic map, while the “KBY_map_all_all.csv” contains all markers of the genetic map.

4. Our 12 populations data explanation

BB, BK, KC, KD, KUI3, MO17, SK, ZONG3, BYD, BYK, KBY and KSC are delegate the populations

B73/BY804, YU87-1/BK, K22/CI7, DAN340/K22, KUI3/B77, MO17/X26-4, ZHENG58/SK, ZONG3/YU87-1,

DE3/BY815, K22/BY815, KUI3/SC55 and BY815/KUI3.

5 A simple explanation of the perl script

step1.pm is a script to exclude the low quality lines, like the high missing, high heterogezosity.

Step2.pm a script to get the parents polymorphic information markers and exclude the family lines low quality markers like high missing, high heterogezosity, and exclude the no two parents source familylines. The method was introduced in the article methods and the method reference was Pan et al. 2012.

step3.pm is script to group the markers and merge the same markers and extract one marker of the cosegregation bin.

Step4.pm is script to random extract a marker and put the oraginal map for three times

Step5.pm is script to merge the three times result .

Step6.pm is try other markers which was not linkage in Step4.pm to the step4 map

Step7.pm is script to use the flip command and windows 7 markers to details justify the marker order.

Reference:

Pan Qingchun Ali Farhan, Yang Xiaohong, Li Jiansheng, Yan Jianbing.(2012) Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. e. PLoS ONE 7(12): e52777.