Computer Note

GACD: Integrated Software for Genetic Analysis in Clonal F1 and Double Cross Populations

Luyan Zhang, Lei Meng, Wencheng Wu, and Jiankang Wang

From the National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China; Institute of Crop Science, and CIMMYT China Office, Chinese Academy of Agricultural Sciences, No. 12 Zhongguancun South Street, Beijing 100081, China.

Address correspondence to J. Wang at the address above, or e-mail: wangjiankang@caas.cn

Received May 14, 2015; First decision July 1, 2015; Accepted September 14, 2015.

Corresponding Editor: Kenneth Olsen

Abstract

Clonal species are common among plants. Clonal F1 progenies are derived from the hybridization between 2 heterozygous clones. In self- and cross-pollinated species, double crosses can be made from 4 inbred lines. A clonal F1 population can be viewed as a double cross population when the linkage phase is determined. The software package GACD (Genetic Analysis of Clonal F1 and Double cross) is freely available public software, capable of building high-density linkage maps and mapping quantitative trait loci (QTL) in clonal F1 and double cross populations. Three functionalities are integrated in GACD version 1.0: binning of redundant markers (BIN); linkage map construction (CDM); and QTL mapping (CDQ). Output of BIN can be directly used as input of CDM. After adding the phenotypic data, the output of CDM can be used as input of CDQ. Thus, GACD acts as a pipeline for genetic analysis. GACD and example datasets are freely available from www.isbreeding.net.

Subject areas: Genomics and gene mapping; Bioinformatics and computational genetics

Key words: clonal species, double cross design, linkage map construction, QTL mapping

Clonal F1 and double cross populations have more alleles at each locus than bi-parental populations derived from 2 inbred parents, resulting in greater genetic variation (Allard 1999; Kreher et al. 2000). Many studies treat clonal F1 and double cross populations as pseudo-backcrosses or pseudo-testcrosses, and then used MapMaker (Lander et al. 1987), MapManager QTX (Manly et al. 2001) to build linkage maps. The CP model (cross pollinators) in software JoinMap (van Ooijen 2006, 2011) and R package OneMap (Margarido et al. 2007) are developed to build linkage maps for clonal F1 populations. The most frequently used software for quantitative trait loci (QTL) mapping in clonal F1 populations at present is MapQTL (van Ooijen 2009) using the CP model. An R/qtl package is capable of linkage map construction and QTL mapping in phase-known double cross populations (Broman et al. 2003).

We have proposed algorithms for linkage analysis in clonal F1 and double cross populations (Zhang et al. 2015a), which could build more accurate linkage maps in shorter time than software packages JoinMap4.1, OneMap and R/qtl. The QTL mapping method of inclusive composite interval mapping (ICIM) (Li et al. 2007; Zhang et al. 2008) was extended for QTL detection in clonal F1 and double cross populations. The efficiency of ICIM was demonstrated by extensive simulations and by comparisons with MAPQTL 6.0 and R/qtl (Zhang et al. 2015b). Most software was developed either for linkage map construction or for QTL analysis. Based on these theoretical studies, we developed the integrated software called GACD (Genetic Analysis of Clonal F1 and Double cross populations) for linkage analysis, map construction and QTL mapping in clonal F1 and double cross populations, freely available from www.isbreeding.net.

Implementation

Marker Categories and Valid Marker Types in GACD

Assume A and B are 2 alleles from the female parent at the marker locus, and C and D are 2 alleles from the male parent at the same locus. Based on the actual number of identifiable alleles in parents...
and the actual number of identifiable genotypes in the F₁ progenies, 4 marker categories can be assigned in clonal F₁ populations (Table 1). In Category I or ABCD, 4 genotypes can be identified in the clonal F₁ progenies, represented by AC, AD, BC, and BD. In Category II or A = B, 1 allele can be identified in the female parent, and 2 alleles can be identified in the male parent. In the clonal F₁ progenies, only 2 genotypes can be identified, represented by XC and XD, where X can be either A or B. In Category III or C = D, 2 alleles can be identified in the female parent, and 1 allele can be identified in the male parent. The 2 identifiable genotypes in the clonal F₁ progenies are represented by AX and BX, where X can be either C or D. In Category IV or AB = CD, both clonal parents show the same heterozygous genotype. The 2 alleles in both parents are represented by A and B, and 3 genotypes in clonal F₁ progenies are represented by AA, AB, and BB.

Five marker categories can be differentiated in the double cross populations (Table 1). Categories I–III are similar to those in clonal F₁ populations. When the female and male hybrids show the same heterozygous genotype, the 2 alleles in parents are represented by A and B, and 3 genotypes in double cross progenies are represented by AA, AB, and BB. In this case, 2 categories can be differentiated. In Category IV (or A = CB = D), allele A is the same as allele C, and allele B is the same as allele D. In Category V (or A = DB = C), allele A is the same as allele D, and allele B is the same as allele C. Missing marker types are coded as XX for any category. Here “missing” means the markers have no genotype calls. Null alleles are not considered in GACD, as it is difficult to determine whether 1 parent carries 2 identical alleles or carries 1 allele and 1 null allele in practice (Zhang et al. 2013a).

Development of the Integrated GACD Software
Core modules for recombination frequency estimation and QTL mapping are written in FORTRAN 90/95, those for building linkage maps are written in C#, and the interface is also written in C#. The software runs on Windows XP/Vista/7/8, with Microsoft.NET Framework 2.0/3.5. GACD is project-based software. All operations and results will be properly saved when the software is closed.

In GACD version 1.0, 3 functionalities can be used in a project: 1) BIN, binning of redundant markers; 2) CDM, construction of genetic linkage maps in clonal F₁ and double cross populations; 3) CDQ, mapping of additive and dominant loci in clonal F₁ and double cross populations. The input file of each functionality can be arranged in 3 formats, that is, pure text, MS Excel 2003, and MS Excel 2007. Several examples are provided for each functionality in the example folder. The main features of GACD are described here, and more information is available in a manual distributed with the software package.

BIN Functionality in GACD
The BIN functionality in GACD is similar to the BIN functionality in QTL IciMapping which is integrated software for genetic linkage map construction and QTL mapping in bi-parental populations derived from inbred parents (Meng et al. 2015). Format of input files is the same as that of the CDM functionality, and the output can be directly used as input of the CDM.

CDM Functionality in GACD
Figure 1 shows the interface of the CDM functionality. Three steps are involved in linkage map construction: grouping, ordering, and rippling. For grouping, markers having the same anchor number will be assigned to the same group first. Anchor information can come from the physical map of the species, or from previous genetic studies. Markers without any anchor information will be assigned to existing groups or a new group based on the specified threshold value of LOD score, or recombination frequency, or map distance. Ordering algorithm in GACD is called mnTwoOpt, where the nearest neighbor is used for tour construction, and two-opt is used for tour improvement (Lin and Kernighan 1973). Two rippling criteria are 1) SARF (sum of adjacent recombination frequencies), and 2) SAD (sum of adjacent distances). The window size used in rippling can be from 5 to 10 markers.

Clicking the buttons of “Grouping,” “Ordering,” “Rippling,” and “Outputting” in turn founds the pipeline of linkage map construction. Users can also modify any steps of linkage map construction in GACD, for example, to modify the anchor information, change the order of 2 chromosomes, and rename the chromosomes and so on. Three sets of linkage maps can be built, that is, combined maps, female and male parental maps. General information of these linkage maps and the input file for the CDQ functionality will be outputted, as well as the 4 haplotypes of the female and male parents. The GACD software also provides the function for drawing and editing linkage maps after running the CDM functionality.

Table 1. Marker categories and valid marker types in the GACD software

<table>
<thead>
<tr>
<th>Functionality</th>
<th>Population type</th>
<th>Marker category</th>
<th>Valid marker type</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIN and CDM</td>
<td>Clonal F₁</td>
<td>I or “ABCD”</td>
<td>AC, AD, BC, BD, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II or “A = B”</td>
<td>XC, XD, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III or “C = D”</td>
<td>AX, BX, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV or “AB = CD”</td>
<td>AA, AB, BB, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I or “ABCD”</td>
<td>AC, AD, BC, BD, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II or “A = B”</td>
<td>XC, XD, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III or “C = D”</td>
<td>AX, BX, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV or “A = CB = D”</td>
<td>AA, AB, BB, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V or “A = DB = C”</td>
<td>AA, AB, BB, XX</td>
</tr>
<tr>
<td>CDQ</td>
<td>Clonal F₁</td>
<td>I or “ABCD”</td>
<td>AC, AD, BC, BD, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II or “A = B”</td>
<td>XC, XD, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III or “C = D”</td>
<td>AX, BX, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV or “A = CB = D”</td>
<td>AA, AB, BB, XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V or “A = DB = C”</td>
<td>AA, AB, BB, XX</td>
</tr>
</tbody>
</table>
CDQ Functionality in GACD

Figure 2 shows the interface of the CDQ functionality. Three methods are available in CDQ functionality: 1) SMA: single marker analysis (Sax 1923; Soller et al. 1976); 2) IM: the conventional interval mapping of additive and dominant QTL (Lander and Botstein 1989); 3) ICIM: inclusive composite interval mapping of...
additive and dominant QTL (Li et al. 2007; Zhang et al. 2008, 2015b).

Genetic information at all scanned positions and at QTL positions identified by the selected QTL mapping methods will be outputted. Software GACD provides line graphs of LOD scores as well as estimated genetic effects from QTL analysis at all scanned positions on 1 chromosome, or on all chromosomes, or on selected chromosomes. The software also provides figures to combine LOD score with the linkage map, and identified QTL with linkage map for IM and ICIM.

Features and Capacity

Current genetic studies always depend on large genetic populations with huge amount of genotypic and phenotypic data. Genetic analysis of these populations requires extensive computing capacities. Therefore, many computer software packages have been developed. For example, MapMaker, MapManager QTX, and JoinMap are developed for linkage map construction. OneMAP is a software package in R used for linkage analysis in clonal F1 populations. MapQTL was developed for QTL mapping in clonal F1 populations. Package R/qtl can be used for linkage analysis and QTL mapping in phase-known double cross. However, users should use command lines to run R language instead of a user friendly interface. In addition, R/qtl is time-consuming and fails to construct dense maps. Integrated software with a simple interface for both linkage analysis and QTL mapping is particularly lacking. Software GACD integrates 3 functionalities, that is, BIN, CDM, and CDQ. One result from BIN can be directly used for CDM, and 1 result from CDM can be readily used as an input file for CDQ. The integration of GACD provides a seamless pipeline from treatment of redundant markers, to estimation of recombination frequency, to building of linkage map and haplotypes, to identification of unknown linkage phases, and finally to mapping of QTL. The current version of GACD can handle genetic populations consisting of 10,000 markers and 500 individuals in map construction and QTL mapping.

Applications and Expectations

Selfing of a clone (where inbreeding depression is not an issue) can be viewed as a special clonal F1, where female and male parents are from the same clone population. In self- and cross-pollinated species, an F1 population is the selfing generation of 1 F1 hybrid between 2 inbred parents. The GACD software can also be used in the above 2 populations. Haplotype building in selfed progenies of a clone can help in determining the 2 haplotypes in the clonal parent. Haplotype building in F1 may help in identifying and correcting markers which are wrongly classified for the 2 inbred parents. Moreover, genetic analysis in an F2 population can still be done in the software GACD, even when there is no genotypic data on its 2 parents or on its F1 ancestry. More broadly, genetic analysis in populations derived from any 2 heterozygotes in animals and plants can be done in GACD. At present, GACD can only handle diploid species. Genetic studies on autoploidy species are much more complex. In the future, we will continue to improve the graphs of linkage maps and QTL mapping results, and may consider epistatic QTL mapping and genetic analysis on polyploid species.

Availability

GACD is freely available from http://www.isbreeding.net. The download package contains also a manual and sample datasets.

Funding

The National Natural Science Foundation of China (31200917); the Generation Challenge Program (GCP, G8009.10); HarvestPlus Challenge Program of CGIAR (5317).

References

van Ooijen JW. 2006. JoinMap 4.0: software for the calculation of genetic linkage maps in experimental populations. Wageningen (the Netherlands): Kyazma BV.