SPECIAL TOPIC
Quantitative Genetics

July 2012 Vol.57 No.21: 2721–2726
doi: 10.1007/s11434-012-5108-0

Quantitative genetic analysis station for the genetic analysis of complex traits

CHEN GuoBo[†], ZHU ZhiXiang, ZHANG FuTao & ZHU Jun[*]

Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, China

Received November 14, 2011; accepted February 24, 2012; published online April 23, 2012

The Quantitative Genetic Analysis Station (QGAStation) is a software package that has been developed to perform statistical analysis for complex traits. It consists of five domains for handling data from diallel crosses, regional trials, core germplasm collections, QTL mapping, and microarray experiments. The first domain contains genetic models for diallel cross analysis, in which genetic variance components and genetic-by-environment interactions can be estimated, and genetic effects can be predicted. The second domain evaluates the performance of varieties in regional trials by implementing a general statistical method that outperforms ANOVA in tackling unbalanced data that arises frequently in trials across multiple locations and over a number of years. The third domain, using predicted genotypic values as proxy, constructs core germplasm collections covering sufficient genetic diversity with lower redundancy. The fourth domain manages genotypic and phenotypic data for QTL mapping. Linkage maps can be constructed and genetic distances can be estimated; the statistical methods that have been implemented apply to both chiasmatic and achiasmatic organisms. Another part of this domain can filter systematic noises in phenotypic data. The fifth domain focuses on the cDNA expression data that is generated by microarray experiments. A two-step strategy has been implemented to detect differentially expressed genes and to estimate their effects. Except in the fourth domain, the major statistical methods that have been used are mixed linear model approaches that have been implemented in the C language. Computational efficiency is further boosted for computers that are equipped with graphics processing units (GPUs). A user friendly graphic interface is provided for Microsoft Windows and Apple Mac operating systems. QGAStation is available at http: //ibi.zju.edu.cn/software/qga/.

QGAStation, genetic software, complex traits, GPU

Citation: Chen G B, Zhu Z X, Zhang F T, et al. Quantitative genetic analysis station for the genetic analysis of complex traits. Chin Sci Bull, 2012, 57: 2721–2726, doi: 10.1007/s11434-012-5108-0

For nearly a century, quantitative genetics has been synchronized to both statistics and genetics and, more recently, to computer science too. The discovery of Mendel's laws introduced the concept of genes to biology, and suggested that the resemblance between individuals can be ascribed to genes that are inherited along generations. Whether discrete genes could address the continuous distributions of phenotypes was in doubt until quantitative genetics answered this question in the affirmative; it is now accepted that Mendel's factors can explain why one individual differs from another in degree rather than in kind [1]. In the first 30 years of the last century, when biotechnology was in its infancy, quantitative genetics used mathematical theory that was largely laid down by pioneers like Ronald Fisher, Sewall Wright and J.B.S. Haldane. In the past two decades, developments in biotechnology and computer science have contributed to the rapid advance of quantitative genetics.

Quantitative genetics uses cutting-edge biotechnology that requires sophisticated experimental designs, such as factorial design, nested design, and Latin square design, to analyze and understand the results of complex genetic studies. The diallel cross method [2,3] is a well-established

[†]Current address: Queensland Brain Institute, University of Queensland, St Lucia, QLD 4072, Australia
[*]Corresponding author (email: jzhu@zju.edu.cn)

mating scheme used by plant and animal breeders, as well as by geneticists, to investigate the genetic underpinnings of quantitative traits. Through diallel cross analyses, it can partition phenotypic variations into various genetic components in terms of their genetic origins like additive, dominance, and epistasis [4]. The development of molecular markers in the 1980's supported the emergence of quantitative trait locus (QTL) mapping [5]; however, QTL mapping rarely pinpoints the exact functional genetic variants because QTLs are often quite wide and cover hundreds of genes. If the expressions of the genes under a QTL can be monitored, then further clues to the identity of the genetic variants responsible for the trait could be identified. Microarrays can be used to monitor the expressions of genes which help to narrow down the candidate genes for a QTL [6].

Most, if not all, phenotypic outcomes of interest are quantitative and complex traits, while studying the genetic architecture of complex traits poses a major challenge in elucidating the hierarchical metabolic processes that may perturb phenotypic outcomes [7]. Properly designed experiments can translate biological processes into reliable data encoding the mechanisms of complex traits, and this data can only be decoded by careful data processing. Experimental design is a Bayesian process in which clues obtained from previous experience are applied to current experimentation [8]. Single data from one individual experiment is rarely enough to address complex traits well, but integrated data from multiple biological layers can strengthen the analysis and, if this is still insufficient, a likelihood function can be used to dissect complex traits. The large amounts of data that are generated from different experiments with different designs and platforms are hard to manage. Further, biological data is often blunted by chaos making analysis difficult.

Although analyzing the dataset separately with different statistical packages remains an option, an integrated platform delivers more convenience. To meet the rising demands in data management and analysis, we have developed the Quantitative Genetic Analysis Station (QGAStation), which is a comprehensive genetic analysis package for the statistical analysis of complex traits. QGAStation enables geneticists and breeders to manage and analyze data collected from different experiments, such as diallel crosses, regional trials, core germplasm collections, QTL mapping, and microarrays. It has a user-friendly graphic interface and can run on Microsoft Windows and Mac OS X operating systems. The statistical methods and the genetic models in QGAStation were tested comprehensively by Monte Carlo simulations and real datasets, and thus should be highly reliable. To accelerate the computation process, the computationally intensive modules have been thoroughly implemented and optimized in C language, and are boosted by graphic processing unit (GPU). When the user's computer is equipped with one or more GPUs, QGAStation's computational efficiency can be further enhanced.

1 Features of QGAStation

1.1 An overview of statistical methodologies implemented in QGAStation

The statistical models implemented in QGAStation are mainly in the form of mixed linear models that are useful in handling complicated genetic models even for unbalanced data. The variance and covariance components of genetic models included in QGAStation are estimated using either the restricted maximum likelihood (REML) method or the minimum norm quadratic unbiased estimation (MINQUE) method [9], which is theoretically as precise as REML but requires less computation. For REML, the stopping rule for convergence is automated by QGAStation. When the use of REML creates an overwhelming computational burden, MINQUE can be used as an alternative. In general, MINQUE(1) with the prior values set to 1, is used to estimate variance components and MINQUE(0/1), with the prior values of covariances set to 0 and the prior values of variances set to 1, is used to estimate covariance and variance components simultaneously [10]. In the mixed linear models, best linear unbiased prediction (BLUP) is used to predict random effects if the variance components are estimated with REML, whereas linear unbiased prediction (LUP) or adjusted unbiased prediction (AUP) is the options in QGAStation if the variance components are estimated by MINQUE [11]. The empirical variances of parameters are estimated by jackknife resampling [12], and subsequently *t*-tests are constructed to test the null hypotheses.

In the simplest case, a set of genetic experimental data would contain observations of a single trait, with one observation per individual. However, most datasets are usually much more complicated than that. Some datasets may contain multiple observations per individual of a single trait across, for example, different growth stages; while other datasets may contain observations of multiple traits that may, or may not, be related. QGAStation implements a general conditioning method to take care of both types of datasets [13,14]. For the first case, a developmental trajectory of genes of the complex trait can be profiled. For the second case, causal inference among traits can be implemented. A conditioning method is embedded into the diallel cross analysis and the QTL mapping domains in QGAStation.

1.2 Description of the five domains of functions in QGAStation

The five domains of functions included in QGAStation are summarized in Table 1, and described briefly below. For full details of each model, please refer to the online user manual.

1.2.1 Diallel crosses

This domain contains statistical methods implementing diallel cross analyses for agronomic traits, seed traits, and animal traits. These statistical methods are implemented in

Table 1 Genetic and statistical models for complex traits that are included in QGAStation

Session[a)]	Model	Parameters
Diallel crosses		
Agronomic models	A	$A\|E$[b)], also called $G\|E$
	AD	$A+D\|E$
	ADAA	$A+D+AA\|E$
	ADM	$A+D+M\|E$
	ADPM	$A+D+P+M\|E$
	AMC	$A+M+C\|E$, a haploid model
Seed models	ADM	$A+D+M\|E$
	GoGe	$Ao+Do+Ae+De\|E$
	GoCGm (diploid organisms)	$Ao+Do+C+Am+Dm\|E$
	GoCGm (triploid organisms)	$Ao+Do+C+Am+Dm\|E$
	GoGeGm	$Ao+Do+Ae+De+Am+Dm\|E$
	GoGeCGm	$Ao+Do+Ae+De+C+Am+Dm\|E$
Animal models	Sex model	$A+D+M+L\|E$
	SexM model	$A+D+Am+Dm+L\|E$
Regional trials	Test model	$V+L+Y\|.$[c)]
	Treat model	$V+T+L+Y\|.$
Germplasm core collection	Stepwise clustering	Constructs an optimal core collection.
QTL mapping		
Linkage map	Achiasmatic model	Constructs linkage maps.
Preliminary analysis		Removes systematic effects and conditional analysis.
Microarray experiments	A two step strategy-based method	Selects differentially expressed genes, and estimates their effect.

a) The main sessions are in bold font. b) "$|E$" implies that the model includes interactions between environment (E) and every term preceding "|". c) "|." implies that the model includes all pair-wise interactions of terms preceding "|".

three sets of genetic models (described below) constructed under the general theory established by Cockerham [15]. The selection of the required minimum number of parental lines depends on the subjects being studied. For some organisms such as rice and cotton, six is considered as a good initial number to balance the capture of genetic diversity and its cost. If, for some organisms, the F_1 population is hard to retrieve, the F_2 population can be used as a substitute. When an experiment is conducted over multiple environment systems, randomize block design should be considered whenever possible [16].

(i) Agronomic models. In general, the total phenotypic variance of an agronomic trait is ascribed to genotypic and environmental factors [17]. In the simplest case, phenotypic outcomes of pure lines can be fitted by an additive (A) effect only model, the **A** model in QGAStation. However, most agronomic traits of interest are complex traits that are also controlled by other effects; thus, the $\boldsymbol{A}$ effect can be extended by including a dominance ($\boldsymbol{D}$) genetic variance component. For these complex cases, the **AD** model in QGAStation can be used. Epistasis [18] can be taken into account using the additive × additive ($\boldsymbol{AA}$) interaction term and users can select the **ADAA** model [19] to analyze additive, dominance, and additive × additive effects.

When the reciprocal effects, maternal ($\boldsymbol{M}$) and cytoplasmic ($\boldsymbol{C}$), are of interest, the **ADM** model [20] can be used. The **ADMP** model [21] can be used to estimate the maternal and paternal effect ($\boldsymbol{M}$, $\boldsymbol{P}$) contribution to the phenotypic outcomes [22,23]. The **AMC** model, which can be applied to haploid-based experiments, such as anther culture studies [24] should also be mentioned. Thus, using this model, the total genetic variance can be ascribed to its additive, maternal and cytoplasm origins.

(ii) Seed models. There are two kinds of quantitative traits for seeds, diploid or triploid, in which genetic variances are determined based on whether the plants are of dicotyledoneae or monotyledoneae origin. In addition to the direct embryo genetic effects ($\boldsymbol{G_o}$) of diploid seeds, the traits are also influenced by maternal nuclear genetic ($\boldsymbol{G_m}$) and cytoplasm ($\boldsymbol{C}$) effects. Consequently, three genetic components exist and they can be fitted by the $\mathbf{G_oCG_m}$ model [10] in QGAStation. In contrast, most cereal crops have triploid endosperms and thus the endosperm effect ($\boldsymbol{G_e}$) has to be considered. Therefore, there can be up to four distinct genetic components for cereal seed traits, summarized as $\mathbf{G_oG_eCG_m}$ model [25]. When the cytoplasmic effect is tiny or ignorable, this model can be reduced to the $\mathbf{G_oG_eG_m}$ model. In much simpler cases, the $\mathbf{G_eG_o}$ model may be sufficient to analyze the contribution of embryo and endosperm to QTLs in, for example, barley seeds [26].

(iii) Animal models. It is known that the genetic architectures of quantitative traits in animals are often substantially characterized by sex. There are three sex-determination systems, the XY system (XY for male, XX for female), ZW system (ZZ for male, ZW for female) and X0 system (haploid X for male, XX for female). The XY system is mainly found in mammals, while ZW is found in birds and insects, such as chicken and silk worm. For these two systems, the X (or Z) chromosome is dosage compensation when Y (or W) inserts to chromosome forming a XY (or ZW) cell, implying that a certain proportion of genes are sex-specific and have sex-linked ($\boldsymbol{L}$) effects. In addition, differences in maternal feeding and nursing patterns can also influence the phenotypes of offspring. Thus, for offspring, besides their own additive and dominance effects, it is essential to build into a genetic model both sex linked and maternal effects [11]. To fit these two genetic sources of variations, the **Sex** model is implemented in QGAStation. The maternal effect can be further partitioned into maternal additive (***Am***) and maternal dominance (***Dm***) effects and so the **SexM** model [27] has been implemented in QGAStation to include them.

For the three genetic models that can be used for diallel cross analyses, some parameters are automatically estimated by QGAStation depending upon the model that has been selected. The narrow sense and broad sense heritabilities can always be estimated, and, in addition, a generalized approach is implemented to predict heterosis effects specific to each environment [28]. For experiments that are conducted in multiple environment systems, genetic term × environment (*GE*) can also be estimated.

1.2.2 Regional trials

Regional trials are used to assess the performance of different varieties (***V***). A trial often takes place across multiple locations (***L***) and over a number of years (***Y***), so things like catastrophes and faults are hard to prevent. For these reasons, the data from such trials often contains missing values, which can eventually lead to imbalances in the data. For balanced data, ANOVA is the most widespread statistical method that has been used for analysis. However, the power of ANOVA appears to be low for unbalanced data. Because mixed linear model approaches outperform ANOVA in handling unbalance data, QGAStation uses mixed linear models to analyze regional trial data regardless of whether the data is balanced or unbalanced [29]. In QGAStation, the performance of varieties is assessed by linear contrasts of their effects with the control; other remaining terms that can take into account the ***L***, ***Y***, ***V*** × ***L***, ***V*** × ***Y***, and ***L*** × ***Y*** effects, can be estimated or predicted depending on whether they are fixed or random effects in a model. For qualities weighted over multiple traits, QGAStation applies a customizable vector, each element of which scores a trait to produce an overall evaluation [30].

1.2.3 Germplasm core collections

Collecting germplasm resource can be cumbersome and time-consuming, but it reserves genetic diversity that may benefit future breeding programs. After the concept of core collection was proposed [31], it was realized that a well-selected subset of the original resource can fulfill the breeding requirement but with a reduced size. If the core collection of an original germplasm resource includes conventionally weighted phenotypic values, then the genetic consistency can often become rather redundant. To avoid this, QGAStation uses predicted genotypic values to sample a population and to build its core collection [32]. This method keeps a substantial genetic polymorphism in the reference population. Two kinds of genetic distances (Euclidean and Mahalanobis distances), three sampling strategies, and seven linkage rules are available to draw an optimal core collection [33].

1.2.4 QTL mapping

QTL mapping is one of the most promising tools that has been used to assist breeding programs [34]. It often consists of two preliminary steps, construction of a linkage map using genetic markers, and processing of the phenotypic data.

(i) Linkage map construction. The software packages commonly used for constructing genetic linkage maps, for example MapMaker and JoinMap [35,36], are built on chiasmatic models that assume the occurrence of chiasmate in both male and female gametogenesis. However, model organisms such as fruit fly and silk worm, are achiasmatic. If the data for such organisms are processed using chiasmatic models, the genetic distances between pairs of markers will be underestimated. Because QGAStation has integrated an achiasmatic model [37], it groups and orders markers, and estimates their genetic distances more accurately than the existing software.

(ii) Phenotypic data processing. Theoretically, a mapping population grows in heterogeneous environments (differing either in degree or in kind) that may bring in substantial systematic noise (or bias) and reduce the mapping precision. However, if the environments are well documented, QGAStation can be used to filter out the noise to calibrate the phenotypes. This step significantly improves the accuracy and statistical power for the subsequent QTL mapping.

As mentioned in section 1.1, the QTL mapping domain of QGAStation supported a conditioning method to deal with complicated cases of trait data. If multiple trait observations are recorded for each individual in a mapping population, then the observations often can be sorted into two generic categories: longitudinal data of a single trait over its development, and, more loosely, a set of multiple relevant traits. For longitudinal data of a single trait with respect to a specific time point, conditional QTL analysis can generate vivid profiles as reported in a previous study [38,39]. For a set of multiple relevant traits, after conditioning on some trait(s), QTL mapping may give insightful results such as cause-result relationships between the different traits. For example, in a rapeseed QTL mapping study [40] of oil content and protein content traits that were previously believed to suppress each other, it was demonstrated that these traits were reciprocal in some of the detected QTLs.

1.2.5 Microarray experiments

Microarray technology is characterized by its unique strength in monitoring the expressions of thousands of genes simultaneously and it has been widely used to identify novel genes or pathways in a large number of organisms. However, gene expression is very sensitive to environmental factors and the data that is produced is often highly noisy. QGAStation offers a seamless two-step strategy to identify differentially expressed genes in cDNA microarray experiments [41]. In step one, genes expressed differentially in multiple environments are chosen using a loose criterion; the selected genes are then imported to the second step for a more stringent scrutiny. These two steps can be used to confirm the real differentially expressed genes and to estimate some quantities of interest, such as gene × treatment interactions. This strategy outperforms the *t*-test method and promises high power with a controlled false discovery rate.

2 Boosting the performance of QGAStation with GPU programming

By adopting the Compute Unified Device Architecture (CUDA) programming model, the variance and covariance component estimations in QGAStation was built to be compatible with the GPU framework. Although to lower the computational burden, users can choose to estimate the variance and covariance components by MINQUE or REML method, the intensive matrix calculations, especially matrix inversion and pseudo-inversion are inevitable and can exhaust computational resource when the number or the size of the matrices increase considerably. QGAStation software handles this problem using the massive parallel computation technique brought by the GPU. Proper numerical algorithms are chosen for all kinds of matrix calculations and thoroughly implemented to take advantage of the many core and stream processing features of the GPU. Especially for matrix inversions, the LU decomposition is used in the GPU mode instead of using the widespread and numerically cheap Sherman-Morrison formula which contains strong loop dependency and is hard to parallelize. The computational advantage rooted in the GPU framework is substantial. When hardware supporting CUDA, for example the Nvidia Tesla C2070 display card on which the software was tested, is available, QGAStation running in the GPU mode can achieve as much as a 150-fold speedup compared with when it is run in the CPU mode on a single Intel® Xeon® X5680 CPU (3.33 GHz). Because QGAStation software can detect the availability of the GPU automatically, the GPU mode is implicitly used whenever it is available.

3 Discussion

In quantitative genetics, when studying the genetic factors that underlie a trait of interest, laboratory work and statistical analyses are naturally complementary. It has been noted that, the further back we trace the history of the biological sciences, the more time and effort scientists have spent in generating and collecting data. Recently, especially after marker genotyping and gene expression profiling became routine laboratory techniques, the economic feasibility of understanding complex traits became a reality. Various methods have been used, such as diallel crosses, regional trials, core collection constructions, QTL mapping, and microarray analysis. Automated assaying platforms have produced a surge of data on an unprecedented scale, and this has led to a split between bench work and computational analysis; a shift that probably will become even more obviously in the future [42].

In this report, we have described QGAStation, an integrated platform that can be used by geneticists and breeders to manage and analyze genetic datasets collected from different biological layers. At its most basic level, QGAStation consists of a series of genetic models and statistical algorithms that can be found in the bin folder of the package. Now thoroughly streamlined, QGAStation only needs the user to identify the models that best fit the problems they want to address; thus, sparing the user the need to carry out detailed statistical modeling and parameter estimation. Advanced users, however, can develop their own scripts using Perl or Python to customize the workflow of the modules (in the bin folder) as required. To ensure that QGAStation remains synchronized with the frontier of quantitative genetics, it is updated frequently. Here we have presented a concise introduction to QGAStation, users can refer to the manual for more details on usage and on the data formats that can be used with this software package.

Two issues should be mentioned here. First, the choice of a genetic model for a dataset is dependent on the users' experience and for some traits this can be difficult and often requires heuristic skills. For example, it may be nontrivial to determine the appropriate genetic model of the ratio of length and width of a rice grain. Because the grain is formed before the endosperm fills, it could be an agronomic trait; however, as the organism grows, the endosperm will develop, making the seed model a potentially better choice. To use QGAStation effectively, it is not essential that there is firm agreement as to which models should be selected. The second issue for the genetic models for agronomic traits, seed traits, and animal traits, is that the number of generations required in a dataset should increase quasi-proportionally with the number of estimated parameters. There are two main reasons for this requirement. First, when a term is added to the model, the precision of estimates may decline because of the enlarged sampling variance caused by collinearity. Although all the models have been tested intensively via Monte Carlo simulations, their robustness might still be challenged by extreme cases that are as yet discovered. Second, each of the models was built with a certain design of experiments [43] in mind, and so the number of generations for a genetic model should be satisfied whenever possible. It is always true that a well-planned experiment translates to a reliable conclusion.

QGAStation is currently available as desktop versions to fit the computer environments used by most geneticists and breeders. Particularly a Mac OS X version of QGAStation has been developed because of the popularity of Apple computers. Current desktop versions of QGAStation usually offer sufficient computational power to address most of the genetic problems described in this report. Because genetic data is being generated at an increasingly rapid pace, it is not unlikely that this package will lag behind some of the analysis task in the coming years. For example, while QGAStation can handle an inverse matrix with hundreds of rows and columns, it will fail if presented with a matrix of thousands of rows and columns. Fortunately, information technology is already producing high-performance computing clusters, GPU-boosted computation, and cloud computing. As shown by QGAStation, GPU technology can dramatically enhance computational efficiency. Some or all of these platforms will help advance statistical computation and data storage in the near future. Integrating these technologies requires and deserves more attention, since they will contribute to a better understanding of genetics.

We thank all attendees at the Workshop on the Analysis of Quantitative Traits and QTLs that has been held annually in Zhejiang University since 2000. Without their comments and incentives, QGAStation would not have been what it is. We thank Drs. Haimin Xu, Yan Lu and Miss. Li-Yuan Zhou for their helps in the development of the QGAStation. This work was supported by the National Basic Research Program of China (2011CB109306), the "111" Project of China (B06014), the CNTC (110200701023), the YNTC (08A05) and Microsoft Research Asia.

1 Fisher A R. The correlation between relatives on the supposition of Mendelianin-heritance. Proc Roy Soc Edin, 1918, 52: 399–433
2 Hayman B I. The theory and analysis of diallel crosses. Genetics, 1954, 39: 789–809
3 Kempthorne O. The theory of the diallel cross. Genetics, 1956, 41: 451–459
4 Cockerham C C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics, 1954, 39: 859–882
5 Lander E S, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 1989, 121: 185–199
6 Jansen R C, Nap J P. Genetical genomics: The added value from segregation. Trends Genet, 2001, 17: 388–391
7 Mackay T F, Stone E A, Ayroles J F. The genetics of quantitative traits: Challenges and prospects. Nat Rev Genet, 2009, 10: 565–577
8 Hinklemann K, Kempthorne O. Design and Analysis of Experiments, Advanced Experimental Design. New York: Wiley-Interscience, 2005
9 Rao C. Estimation of variance and covariance components MINQUE theory. J Multivar Anal, 1971, 1: 257–275
10 Zhu J, Weir B S. Analysis of cytoplasmic and maternal effects, I. A genetic model for diploid plant seeds and animals. Theor Appl Genet, 1994, 89: 151–159
11 Zhu J, Weir B S. Diallel analysis for sex-linked and maternal effects. Theor Appl Genet, 1996, 92: 1–9
12 Miller R G. The jackknife: A review. Biometrika, 1974, 61: 1–15
13 Atchley W R, Zhu J. Developmental quantitative genetics, conditional epigenetic variability and growth in mice. Genetics, 1997, 147: 765–776
14 Zhu J. Analysis of conditional genetic effects and variance components in developmental genetics. Genetics, 1995, 141: 1633–1639
15 Cockerham C C. Random and fixed effects in plant genetics. Theor Appl Genet, 1980, 56: 119–131
16 Kang M S. Handbook of Formulas and Software for Plant Geneticists and Breeders. Binghanton, NY: The Haworth Press, 2003
17 Lynch M, Walsh B. Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer Associates, 1998
18 Phillips P C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet, 2008, 9: 855–867
19 Xu Z C, Zhu J. An ADAA model and its analysis method for agronomic traits based on the double-cross matting design (in Chinese). Acta Genet Sin, 2000, 27: 257–269
20 Abderrahmane A, Zhu J. Simulation studies for comparing genetics models with additive-dominance-maternal effects and GE interaction effects. J Biomathematics, 2002, 17: 208–214
21 Cockerham C C, Weir B S. Quadritic analysis of reciprocal crosses. Biometrics, 1977, 33: 187–203
22 Zhu J, Weir B S. Mixed model approaches for diallel analysis based on a bio-model. Genet Res, 1996, 68: 233–240
23 Lou X Y, Yang M C. Estimating effects of a single gene and polygenes on quantitative traits from a diallel design. Genetica, 2006, 128: 471–484
24 Yan J Q, Xue Q Z, Zhu J. Genetic studies of anther culture ability in rice (*Oryza sativa* L.). Plant Cell Tissue Organ Cult, 1996, 45: 253–258
25 Zhu J, Weir B S. Analysis of cytoplasmic and maternal effects, II. Genetic models for triploid endosperms. Theor Appl Genet, 1994, 89: 160–166
26 Yan X F, Xu S Y, Xu Y H, et al. Genetic investigation of contributions of embryo and endosperm genes to malt kolbach index, alpha-amylase activity and wort nitrogen content in barley. Theor Appl Genet, 1998, 96: 709–715
27 Zhu J, Duan J L. Genetic models and analysis methods for sex-linked and maternal gene effects. J Biomath, 1994, 9: 1–9
28 Xu Z C, Zhu J. An approach for predicting heterosis based on an additive, dominance and additive x additive model with environment interaction. Heredity, 1999, 82: 510–517
29 Zhu J, Xu F H, Lai M G. Analysis methods for unbalanced data from regional trials of crop variety, analysis for single trait (in Chinese). J Zhejiang Agri Univ, 1993, 19: 7–13
30 Zhu J, Lai M G, Xu F H. Analysis method for unbalanced data from regional trial of crop variety: Analysis for multiple traits (in Chinese). J Zhejiang Agri Univ, 1993, 19: 241–247
31 Frankel O H, Brown A H D. Current plant genetic resources-a critical appraisal. In: Genetics New Frontiers IV. New Delhi: Oxford & IBH Publishing Co., 1984. 1–11
32 Hu J, Zhu J, Xu H M. Methods of constructing core collections by step-wise clustering with three sampling strategies based on the genotypic values of crops. Theor Appl Genet, 2000, 101: 264–268
33 Xu H M, Hu J, Zhu J. An efficient method of sampling core collection from crop germplasm (in Chinese). Acta Agron Sin, 2000, 26: 157–162
34 Lande R, Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 1990, 124: 743–756
35 Lander E S, Green P. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA, 1987, 84: 2363–2367
36 Stam P. Construction of intergrated genetic linkage maps by means of a new computer package: JoinMap. Plant J, 1993, 3: 739–744
37 Wu J X, Zhu J, Jenkins J N, et al. Constructing linkage maps with achiasmatic gametogenesis. Acta Genet Sin, 2005, 32: 608–615
38 Yan J Q, Zhu J, He C X, et al. Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). Genetics, 1998, 150: 1257–1265
39 Yan J Q, Zhu J, He C X, et al.Quantitative trait loci analysis for developmental behavior of tiller number in rice (*Oryza sativa* L.). Theor Appl Genet, 1998, 97: 267–274
40 Zhao J, Becker H C, Zhang D, et al. Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor Appl Genet, 2006, 113: 33–38
41 Lu Y, Zhu J, Liu P. A two-step strategy for detecting differential gene expression in cDNA microarray data. Curr Genet, 2005, 47: 121–131
42 Schadt E, Linderman M, Sorenson J, et al. Computational solutions to large-scale data management and analysis. Nat Rev Genet, 2010, 11: 647–657
43 Zhu J. Analysis Methods for Genetic Models (in Chinese). Beijing: China Agriculture Press, 1997