Development of Novel Theory and Methods for QTL Analysis and Inferring Crossover Interference in Autotetraploids

by

JING CHEN

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

The University of Birmingham

September 2015

ABSTRACT

Chapter I-1

Quantitative genetics model of autotetraploid species is crucial for functional and evolutionary genomic analyses. However, compared with diploids, quantitative genetics study of autotetraploids lags far behind. I used orthogonal contrast scales to construct a genetics model for studying epistasis between genes in autotetraploid species, one very important statistical part to link the genotype of quantitative trait loci (QTL) to the corresponding phenotype. Here I established models for both one locus and two loci followed by a variety of allelic frequency distributions. I illustrated this genetics model for analysing QTL in a F_2 family of autotetraploid population under autotetrasomic inheritance and in a random mating equilibrium population. I also established a method for estimating genetic effects in linkage disequilibrium autotetraploid population. The simulation study showed the feasibility of a practical implementation of this method, detailed the procedure of the analysis, demonstrated the reliability in the parameter estimation, and discussed its utility and potential problems.

Chapter I-2

Insights into the relationship between phenotypic variation and genetic variation for the quantitative traits are helpful for improving selective breeding programmes in agriculturally and economically important plants and animals. To increase the speed of breeding in the world's

third most important crop, cultivated potato, a likelihood-based method of QTL interval mapping is developed for autotetraploid species in a full-sib family, which considers multivalent meiotic pairing of homologous chromosomes. Here I considered all the observed genotypes of genetic markers at both sides of the interval by using a Markov chain model, which would effectively improve the QTL mapping precision and resolution. The simulation study showed the reliability of this method as a practical implementation in analysis for autotetraploids with both bivalent pairing and quadrivalent pairing during meiosis.

Chapter II-1

Both theoretical and experimental evidence have suggested that recombination frequency would be increased in autotetraploids compared with their parental diploids. In almost all organisms, crossover interference is likely to play an important role in determining the frequency and patterns of recombination along chromosomes. To investigate into the underlying process of crossover in autotetraploids, a Chi-square model and novel statistical method is developed to explore crossover interference with properly accounts for the essential features of segregation and recombination under tetrasomic inheritance. The simulation studies were performed to confirm the accuracy of the maximum likelihood estimates of the model parameters, and a small data set of autotetraploid yeast was presented to apply the method. A significant decrease in the strength of crossover interference was found on one chromosome among the tested three chromosomes after polyploidization, suggesting a new hypothesis worthy of further investigation that the increase of recombination frequency after polyploidization would partly due to the decrease in the strength of crossover interference.

Chapter II-2

Taking advantages of the technology of next generation sequencing, it is possible to obtain dense genetic marker data from products of meiosis. To estimate crossover rate from the observed marker genotypic data in autotetraploids, a statistical method is proposed using genotype data called from the intensely distributed SNP markers. Here a yeast data set was presented to apply this method and an overall increase in the crossover rate was found after polyploidization.

ACKNOWLEDGMENTS

First of all I would like to thank Professor Zewei Luo for his guidance and help throughout my study. I can hardly make this work without his exceptional insight and patience and I will always be grateful for this. I feel really lucky to be under supervision by Professor Luo, who leads me into the field of statistical genetics and teaches me in both the academic research and daily life. Thank you so much for inspiring me and for encouraging me.

Thank you to Dr Lindsey Leach who has always been really nice and helpful during my time at Birmingham. I do really appreciate her encouragement and advice on my work. Thank you also to all of my colleagues in Birmingham and Fudan University for their helpful discussion and support, including Dr Minghui Wang, Dr Ning Jiang, Dr Chenqi Lu, Dr Xiaohua hu, Dr Xiaoping Wu, Dr Ou Fang, Anushree Choudhary. I am really honoured to work with all of you.

Thank you to all my friends who supported me a lot in the last four years in the school. I will never forget the relaxed time we spent together, sharing food, hiking, travelling…

Finally, a heartfelt thank you to my parents who are always understanding and supportive. I would like to say “爸爸妈妈，谢谢你们的支持，我爱你们！” to them. A special thank you to Dr Tian Li whose love, support and advice has been incredibly important to me and will always be.

Table of Contents

Overall Introduction: Genetic architecture of autotetraploid species

Part I: Theory and methods for QTL analysis in autotetraploids

Chapter I-1: Orthogonal contrast based models for quantitative genetic analysis in autotetraploid species

Chapter I-2: Interval mapping of QTL in autotetraploid species

List of Tables

Chapter II-2: Predicting meiotic crossover rate in *Saccharomyces cerevisiae* based on the whole genome wide sequencing data analysis for autotetraploids

Overall Discussion: Progress in the theoretical basis for statistical genetic analysis in autotetraploids

List of Tables

List of Tables

Chapter I-2

Part II

Chapter II-1

List of Tables

Chapter II-2

List of Figures

List of figures

Part II

Chapter II-1

Chapter II-2

List of abbreviations

ANOVA	Analysis of variance
CIM	Composite interval mapping
CO interference	Crossover interference
EM algorithm	Expectation and maximization algorithm
M-step	Maximum step
E-step	Expectation step
LOD score	Logarithm of the odds favouring linkage score
MIM	Multiple interval mapping
MLE	Maximum likelihood estimate
QTL	Quantitative trait loci
RI	Recombination interference
SNP	Single nucleotide polymorphism
SSE	Error sum of square
SST	Treatment sum of square
TSS	Total sum of square

OVERALL

INTRODUCTION

GENETIC ARCHITECTURE

OF

AUTOTETRAPLOID SPECIES

1.1. Over all introduction to polyploids

Polyploidization, the simultaneously duplication of the whole genome, widely occurs in the evolution of eukaryotes, especially for flowering plant species. It was estimated that all angiosperms have experienced at least once a state of polyploidy during the evolutionary history (Jiao 2011). Exploring the evolutionary significance of polyploidy remains a mystery and challenging job. It was thought that the tempo of evolution for a trait may not only be determined by the rate of environmental change but also depended on the form and extent of genetic variability present within the population (Fisher R.A. 1930) and the reason for widespread polyploidy may be due to polyploidization and the subsequent evolution are an extremely dynamic process, leading to faster evolution or in more novel directions than related diploid species (Soltis 2000). Generally, polyploidization has been recognized as an important driving force in the evolutionary history of plants (Otto and Whitton 2000; Soltis and Soltis 2000; Blanc and Wolfe 2004; Chen 2007; Otto S.P. 2007; Christian 2010). Under these views, we may consider that polyploidization of a genome could have profound long-term effects on genetic diversity and promotes adaptive evolutionary change. However, the direct effect on evolutionary success of polyploidy is still insufficiently known (Clausen 1945; Stebbins 1971; Grant 1981; Levin 2002; Comai 2005; Soltis and Soltis 2009). To address this fundamental question, mechanism underpinning the genetic changes should be explained by further investigation.

Besides playing as evolutionary important role in many species, polyploidy presents in several economically important species, like crops and aquacultural animals. For example, cultivated potato, an autotetraploid crop, is now known as the world's third most important food crop,

ranking just after rice and wheat. It is anticipated that the world's population will increase from current 6.9 billion to 9.6 billion, requiring about 50% increase in agricultural production by 2050 (United Nation 2012). This raises tremendous and serious challenges to the existing food resources. Given natural land and water resource are already used intensely for agriculture production, while potato requires much less land to grow for substantially higher production yield compared to other major crop, it is thus recognized as the food for future (Bovell-Benjamin 2007). Improving potato production and its potential in human food system requires multidisciplinary, integrated research and activities. Most agronomic trait including yield, quality, abiotic and biotic resistance targeted in crop breeding programs are quantitative traits whose phenotypic variation shares common features, polygenic control and environmental modification. Understanding polygenic architecture underlying quantitative traits is essential to improve efficiency of any breeding program of these traits.

In contrast to diploid species, progress in statistical genetics analysis in polyploid species has been hampered far behind due to the much more complicated inheritance which indicates the inappropriateness of applying the theory and methods of quantitative genetics analysis for diploids directly in the analysis for polyploids. In the work presented here, I attempt to accelerate progress in this challenging area in autotetraploid species not only because that autotetraploid, as a simple polyploidy form, is a good starting point to investigate into to study the evolutionary role played by polyploidization, but also for the reason that economically important cultivated potato is autotetraploids.

1.1.1. Allopolyploids and autopolyploids

Polyploid species are those holding more than two complete sets of chromosomes, prevalently recognized as three categories: allopolyploids, autopolyploids and segmental allopolyploids (Stebbins 1947). Allopolyploids, with chromosomes derived from hybridization between genetically distinct parents, is similar to diploids in terms of chromosome pairing and segregation pattern during meiosis. For example, triticale, the first successful man-made cereal grass crop, is allohexaploids, four sets of chromosomes from wheat (*Triticum turgidum*) and two sets of chromosomes from rye (*Secale cereale*) (Mergoum 2009). Thus allopolyploids predominantly form bivalents of paired chromosomes during meiosis (Jackson 1982; Ramsey 2002). Linkage analysis in strict allopolyploids can be more successful due to disomic inheritance by directly applying principle and method for diploids into allopolyploids. By contrast, autopolyploids, arising from doubling genome or fusion of two unreduced gametes, display much more complicated polysomic inheritance in which each homolog can pair with any other homolog during meiosis, like cultivated potato (Consortium The Potato Genome Sequencing 2011). Segmental allopolyploids consist of more than two partially differentiated genomes and much cytogenetic evidence indicates that homologous chromosomes may segregate due to a mixture of bivalent and quadrivalent pairing. It was observed that in autotetraploid *Saccharomyces cerevsiae*, homologs were mostly formed as bivalent pairing at pachytene, the stage when chromosomal crossover occurs, and sometimes formed quadrivalents by switching pairing partners (Loidl 1995). However, trivalents, quadrivalents and univalents have also been observed in potato with low frequencies (Swaminathan 1953). It should be stressed that autopolyploids forming only bivalents is different from behaviour in allopolyploids,

which have more than two sets of homologous chromosomes to be randomly paired during meiosis. Polysomic inheritance is regularly considered as a diagnostic trait to differentiate autopolyploids from allopolyploids (Soltis 1993; Jackson 1996; Landergott 2006; Stift 2008).

For a long time, researchers have paid little attention to the evolutionary advantages of autopolyploids. Compared with allopolyploids, autopolyploids were conventionally believed to be evolutionary disadvantages due to multivalent formation during meiosis. It was suggested that multivalent formation may lead to meiotic irregularities and reduced fertility (Clausen 1945; Stebbins 1971). However, it was estimated by Ramsey (1998) that the frequency of autopolyploid formation is higher than that of allopolyploids. The frequently occurrence discovered in natural autopolyploids suggest that genome multiplication may play a significant role in the evolutionary history (Soltis 2000). Christian (2010) proposed that the evolutionary advantages of autopolyploids mainly rely on two traits, namely genic redundancy and polysomic inheritance. For example, although the speed of selection would be slowed down in autopolyploids, effective population size would be increased and inbreeding depression would be reduced in the short-term evolution. It has been found that genic redundancy and polysomic inheritance seem to decline in long term, as only the transient evolutionary stage in the lifespan. Autopolyploids would restore disomic inheritance and accumulate adaptive genetic variation under genic redundancy (Christian 2010). Therefore, genetic analysis, taking the key features of polysomic inheritance into account, is essential for solving many open questions remained with respect to the evolution of autopolyploids.

1.1.2. Dynamic changes caused by polyploidization

Polyploidization is one of the most dramatic mutations by adding a complete set of chromosomes to the genome (Otto 2007). Because of the significant role played by polyploidization in the plant evolution, polyploidy has been the focus of great interest and studied lasting more than eighty years. Some evidence shows that polyploidy genomes would experience highly dynamic restructuring and reorganization of gene expression (Doyle 2008; Leitch 2008). Diverse aspects of polyploidy have been investigated, from external phenotypic effects of polyploids to genetic consequences of polyploidy evolution, facilitated by the dramatically development of molecular techniques. It was reported that polyploidization would commonly and universally increase cell size, which may affect the rate of metabolic process by altering the surface to volume ratio (Cavalier-Smith T. 1978, Levin DA 1983). Consequently, growth rate, overall size and shape could also be changed accordingly. Additionally, polyploidization has some effects on gene expression and organ structure and function, such as reproduction systems (Stebbins GL. 1980), and ecological and physiology tolerances of polyploids was found to be broader than their related diploids (Levin DA 1983; Lokki J. and Saura A. 1980). What's more, genomic investigations have revealed extensive genetic and epigenetic changes associated with polyploidy and the potential selective advantages to the polyploidy state (Otto SP and Whitton J 2000; Stupar 2007). All these indicate that genome structure and function of polyploids may differ markedly from that of their diploid relatives. Thus, it is inappropriate to roughly approximate genetic analysis of a polyploidy from that of its diploid relatives.

1.1.3. The challenges of modelling tetrasomic inheritance

Studies that facilitated dissection of the genetic architecture underlying genetic variation of complex and quantitative traits have been routine analysis in almost all important diploid species. In sharp contrast, corresponding studies in polyploids are far behind this level of progress for the much more complicated pattern of gene segregation and recombination than that in diploids. Tetrasomic linkage analysis has been one of the most challenging topics in theoretical and applied genetics since the pioneering works of quantitative geneticists including J.B.S. Haldane (1930), K. Mather (1935, 1936) and R. Fisher (1947).

Firstly, autotetraploids undergo tetrasomic inheritance in which each homologous chromosome can pair with each any other homologous chromosomes during meiosis. Much cytogenetic evidence demonstrated that homologous chromosomes would segregate either in bivalent pairing, quadrivalent pairing or a mixture of the two during meiosis. The essential features of bivalent chromosome pairing and quadrivalent chromosome pairing will be discussed in the next session (1.1.4). Due to the multivalent pairing, the up limit value of recombination frequency between two loci is 0.75 in autotetraploid species (Luo et al, 2006).

Secondly, multivalent pairing of homologous chromosomes during meiosis may cause sister chromatids to enter into the same gamete, the well-known phenomenon of double reduction in tetrasomic inheritance, resulting in systematic segregation distortion. Study by Luo et al (2006) showed that the coefficient of double reduction can reach to 0.25 at most.

Finally, multiple alleles at a locus of polyploids cause a substantially wider range of genotypic segregation. For example, consider one locus in diploids, at most two gametes with different

genotypes can be generated by an individual. However, in autotetraploids, at most 10 gametes with different genotypes can be generated by an individual if all the alleles are distinct with each other. As the number of loci increase, the distance between maximum number of gamete genotypes between diploids and autotetraploids would increase exponentially. Due to the existence of multiple alleles, there are no fully informative genetic markers in autotetraploids. A simple one-to-one relationship between observed genotyping data and genotypes in autotetraploids is usually difficult to obtain. For example, when considering two alleles (A_1 and A_2) segregate at locus in an autotetraploid population, the relationship between marker phenotypes (i.e. "marker phenotype" refers to the observed genotype throughout this thesis) and genotypes can be shown in Table 1.1 (Modified from Luo et al 2000). Taking into account the

Table 1.1. Relationship between marker phenotypes and genotypes at a single locus.

Phenotypic record		Corresponding genotypes
Gel-band 1	Gel-band 2	
1	0	(A_1 O O O), (A_1 A_1 O O), (A_1 A_1 A_1 O), (A_1 A_1 A_1 A_1)
0	1	(A_2 O O O), (A_2 A_2 O O), (A_2 A_2 A_2 O), (A_2 A_2 A_2 A_2)
1	1	(A_1 A_2 A_2 A_2), (A_1 A_2 A_2 O), (A_1 A_2 O O), (A_1 A_1 A_2 A_2), (A_1 A_1 A_2 O), (A_1 A_1 A_1 A_2)

Alleles A_1 and A_2 are revealed as the presence of PCR products indicated by gel-band 1 and gel-band 2 respectively. 1 or 0 is used to indicate the presence or absence of corresponding gel-band. O denotes a null allele.

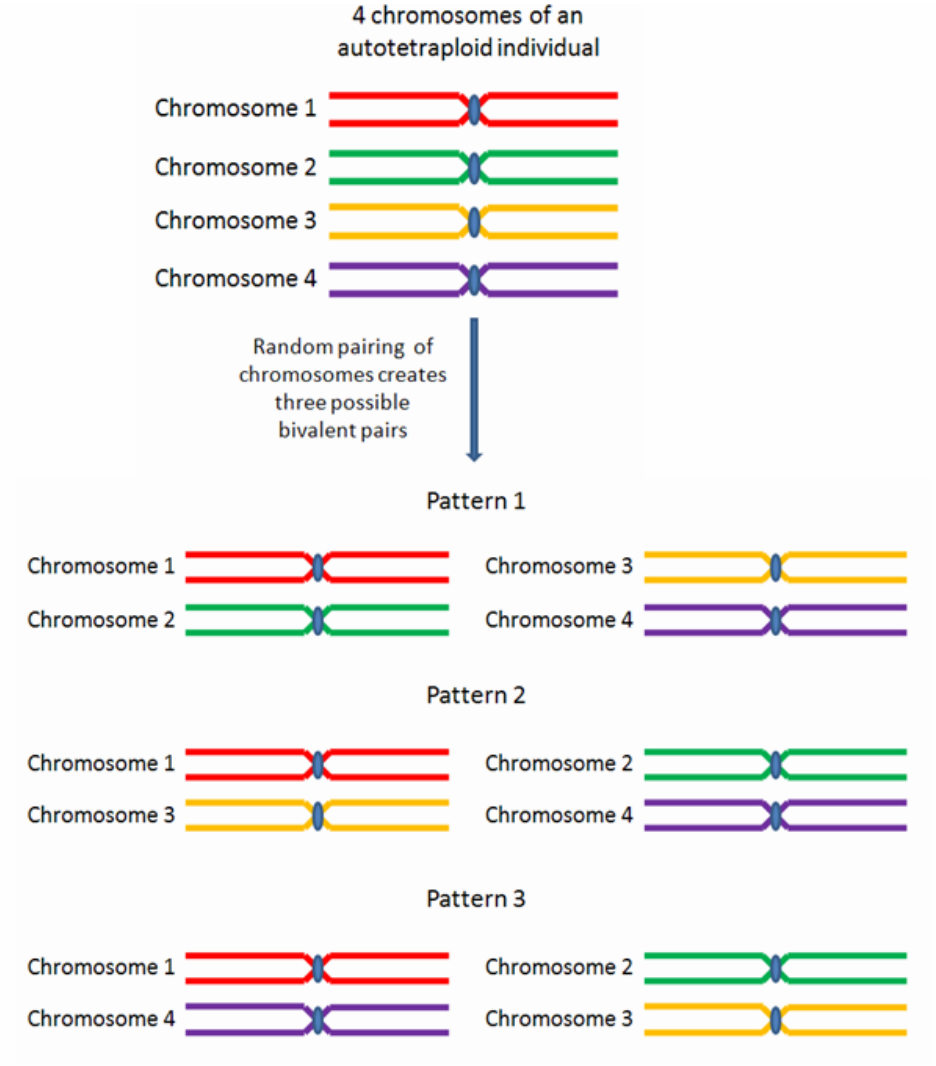
possibility of null alleles (indicated by failure of the PCR primers to anneal to the relevant DNA templates), it can be seen from Table 1.1 that there may be 4, 4 or 6 corresponding genotypes in three different phenotype categories.

1.1.4. Bivalent pairing and quadrivalent pairing

Autotetraploids undergo tetrasomic inheritance in which each homologous chromosome can pair with any other homologous chromosomes during meiosis and segregate either in bivalent, quadrivalent pairing or a mixture of the two.

The simplest model of chromosome pairing and segregation may have the full complement of bivalents as shown in Figure 1.1. Homologous chromosomes are randomly paired to create bivalent pairs and recombination only occurs between the two chromosomes of each bivalent pair which is similar to diploids. Then one recombined chromosome from each pair enters into the gamete. Autotetraploids with bivalent pairing therefore share some common features with diploids. For example, the upper limit value of recombination frequency is 0.5, sister chromatid will not enter into the same gamete and the relationship between recombination frequency and genetic distance is the same as that in diploids.

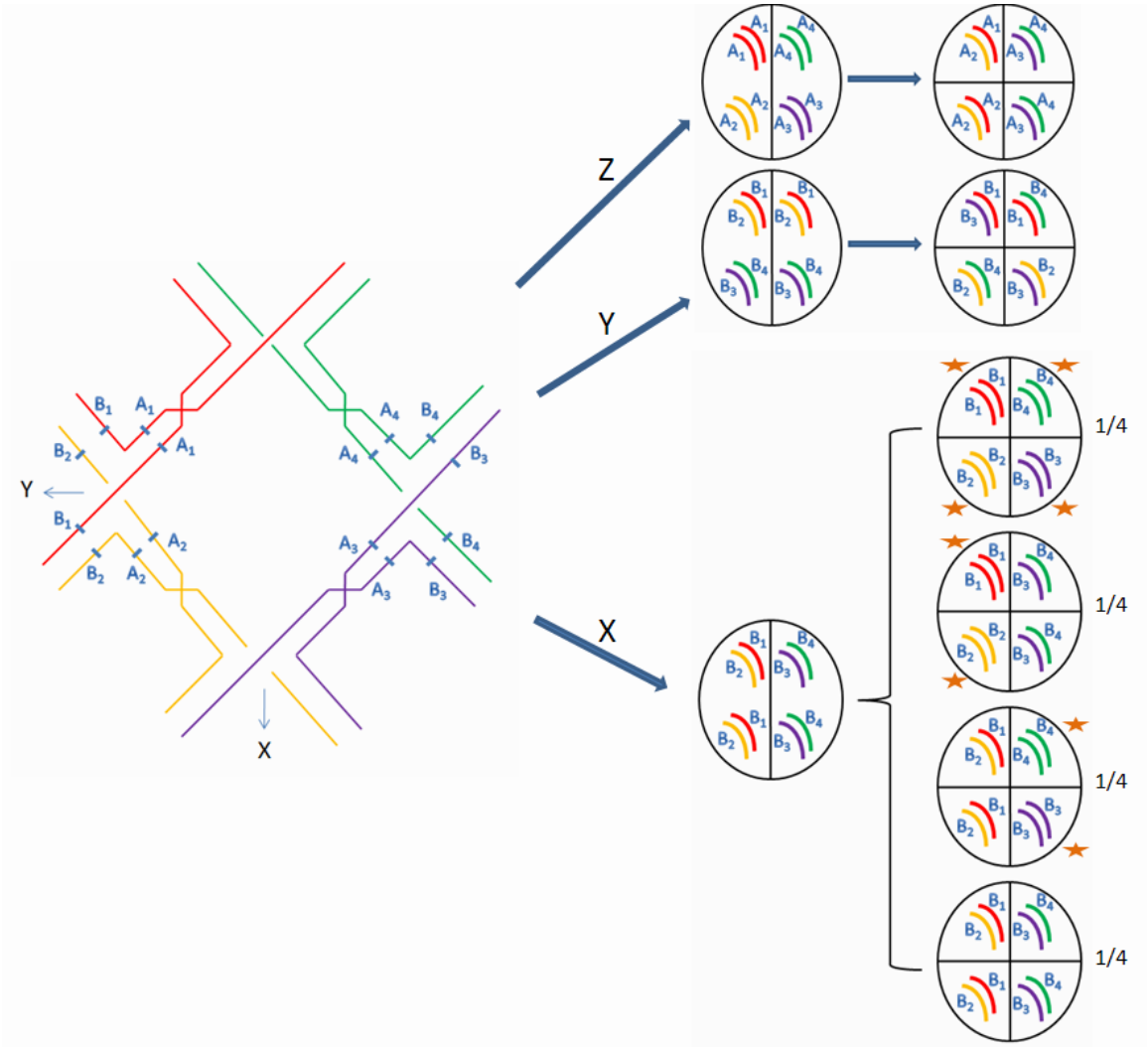
Figure 1.1. Chromosome segregation during bivalent meiosis of an autotetraploid species.



Here four colours (red, green, yellow and purple) represent four different sets of homologous chromosomes of an autotetraploid individual. Chromosomes are randomly paired to create three possible bivalent pairs patterns and recombination occurs only between the two chromosomes of each bivalent pair. Consequently, one recombined chromosome from each of the two pairing bundle enters and forms the diploid gamete.

One of the most important features of autotetrasomic inheritance is the phenomenon of double reduction due to quadrivalent pairing, in which sister chromatids can migrate to the same gametes during meiosis. Double reduction plays a significant role in the evolution of autotetraploid genomes. It was demonstrated that double reduction could enhance the ability to eliminate deleterious alleles even at low levels and contribute more to inbreeding depression (Butruille and Boiteus 2000). The probability of this meiotic event is defined as the coefficient of double reduction, which depends on the recombination frequency between the locus and its centromere and on the frequency of multivalent formation. Historially (Mather 1935; Bailey 1961; Ronfort 1998; Butruille and Boiteus 2000), a maximum value of 1/6 was cited for the coefficient of double reduction. However, it has been shown in Luo et al (2005) that the upper limit of the coefficient of double reduction is, in fact ¼, a value that is reached when recombination frequency takes its upper bound value of ¾. The limiting recombination frequency under quadrivalent pairing during meiosis was demonstrated to be ¾ by Sved (Sved 1964). Figure 1.2 illustrates how double reduction takes place during quadrivalent meiosis in an autotetraploid species.

Figure 1. 2. Segregation patterns of loci A and B during quadrivalent meiosis of an autotetraploid species.



Here four colours (red, green, purple, yellow) represent the four homologous chromosomes of an autotetraploid individual. Locus A has no crossover with the centromere and undergoes path Z with no double reduction. Crossover occurs between locus B and the centromere. It is equally likely for chromosomes to undergo path Y with no double reduction or path Z with double reduction. Gametes derived from a double reduction event are marked with an asterisk. The figure was modified from Wu et al (2001).

1.1.5. Statistical framework for autotetrasomic linkage analysis

As a theoretically challenging topic in the history of statistical genetics, the study of genetic linkage analysis in autotetraploids was pioneered by the quantitative geneticist K. Mather in the year of 1935 (Mather, 1935). However, the majority of research work was built on the assumptions of bivalent pairing of homologous chromosomes during tetrasomic meiosis. Although these methods could significantly reduce challenges in modelling of linkage analysis and decrease the degrees of complexity of the data analysis in autotetraploids, they ignored some key features of tetrasomic inheritance and were impractical in experimental data analysis. Some studies even use the corresponding related diploid as an approximation to the polyploidy case (Bonierbale et al. 1988, Gebhardt et al. 1989). Another commonly used strategy to construct linkage maps in polyploids has been the use of single-dose dominant marker, such as AFLPs and RAPDs (Wu et al., 1992; da Silva et al., 1993; Grattapaglia and Sederoff, 1994; Ukoskit and Thompson, 1997; Brouwer and Osborn, 1999; Hoarau et al., 2001; Barcaccia et al., 2003; Ghislain et al., 2004; Cervantes-Flores et al., 2008). Single-dose dominant markers are present in only one parent in a single copy (*i.e.* parental genotypes AOOO × OOOO), so only half of the gametes will contain the marker during gametogenesis. Recombination can therefore be analysed as for a diploid species if two such markers are present in a coupling phase.

To establish general theory for linkage analysis with polysomic inheritance, it is necessary to predict parental genotypes from phenotypes which are essential for distinguishing recombinant and parental genotypic classes. In this respect, Luo et al (2000) developed a method for predicting the genotypes of two parental individuals at a co-dominant or dominant marker locus

on the basis of the parental and offspring phenotypes scored at that locus in segregation autotetraploid populations.

This was followed by the development of methodology for constructing genetic linkage maps of co-dominant or dominant markers under the assumption of random bivalent pairing of four homologous chromosomes during meiosis (Luo et al., 2001). In order to integrate multivalent pairing and double reduction events into linkage analysis in autotetraploids, Wu et al (2001) proposed a statistical method by assuming that all four alleles segregating at each of two loci are all different in both parents. In this case, they could directly resolve both double reduction and recombination from the offspring populations. However, this assumption is unrealistic in practice since parental lines that match these requirements are extremely rare. Subsequently, this method did not properly solve major problems in statistical modelling of real data.

A well- developed theoretical basis for tetrasomic linkage analysis was later developed by Luo et al (2004) that took account of the major complexities of autotetrasomic inheritance including mixed bivalent and quadrivalent pairing in meiosis and the phenomena of double reduction. For the first time, this study was successfully in working out the distribution of two-locus genotypes in outbred population, in terms of the coefficient of double reduction and recombination frequency. This statistical method provided the analytical tools for predicting the maximum likelihood estimates (MLEs) of parameters of both double reduction and recombination frequency and testing their significance. Subsequently, Luo et al. elaborated the tetrasomic linkage analysis and demonstrated its efficacy through the construction of genetic marker linkage maps from an outbred segregation population of autotetraploid potato (*Solanum tuberosum*) (Luo et al., 2006).

To improve precision and accuracy of constructing linkage map of autotetraploid species, Leach et al (2010) proposed a hidden Markov chain based approach for multilocus linkage analysis in autotetraploids. This method significantly improved the accuracy and precision of parameters estimation, compared with the two-locus linkage analysis method proposed by Luo et al (2004). In addition, this multi-locus method provided a way to directly calculate the likelihood for any given linkage map and mapping population in autotetraploids, which made significant propress in statistical inference of linkage order and linkage phase (Leach et al, 2010).

1.2. Aims of the project

The project aims to address a challenging task to provide the novel theory and methods that enable QTL mapping in autotetraploids to be carried out on a rigorous theoretical basis and establish statistical methods for inferring recombination interference in both diploids and autotetraploids. The thesis reports four sub-projects which I conducted in the past four years.

Part I-1

Quantitative genetics model of autotetraploid species is crucial for functional and evolutionary analyses of quantitative genetic variation. I implemented orthogonal contrast scales to construct a genetics model for quantifying various genetic effects of QTL under tetrasomic inheritance and decomposing quantitative genetic variation into orthogonal variance components. I established models for both one locus and two loci followed by a variety of allelic frequency distributions. I illustrated these genetics models by quantifying various genetic effects from a F_2

family of autotetraploid population and in a random mating equilibrium population under autotetrasomic inheritance. I also established a method for estimating genetic effects in linkage disequilibrium autotetraploid populations. The simulation studies were presented to show the feasibility of a practical implementation of this method, detailed the procedure of the analysis, demonstrated the reliability in the parameter estimation, and discussed its utility and potential problems.

Part I-2

A starting point to unveil the genetic mechanisms controlling quantitative traits is to map the genes affecting the traits (i.e. mapping quantitative trait loci or QTL). Theory and methods have been well established for mapping QTL in diploid species and QTL analysis has become a routine practice in all important diploid plant and animal species as well as in humans. However, the study of statistical genetics of autotetraploid species is still in its infancy largely because the inheritance of polyploids, especially autopolyploids, is much more complicated in comparison to diploids. Based on the quantitative genetics model of autotetraploid species, I developed a method to map the genes affecting the traits onto the genome (i.e. mapping QTL). I developed an interval QTL mapping method for a segregation population derived from crossing two outbred autotetraploid parents, in which we considered all the possible parental QTL genotypes and linkage phases. Since genetic markers in autotetraploids are usually not fully informative, I considered use of multi-locus marker information in a Hidden Markov Chain model to improve the efficiency of parameter estimation. To investigate properties of this statistical analysis, a simulation study was performed to assess this approach and to investigate the effects of population size, parental genotypes and locations of QTL on the chromosome, and showed the

reliability of this method as a practical implementation in analysis for autotetraploids with both bivalent pairing and quadrivalent pairing during meiosis.

Part II-1

Recombination interference (RI) refers to the phenomenon in genetic recombination that simultaneous recombination in closely nearby chromosomal intervals occurs much less frequently than would be expected under independence of the recombination events across the intervals. Based on the work of inferring crossover interference in diploids by McPeek and Zhao (1995), I proposed a novel statistical method for inferring crossover interference in autotetraploid species which taking proper account for tetrasomic inheritance. I demonstrated the model's statistical properties by simulation studies and illustrate application of our model using phenotype datasets of three linked fluorescent marker loci scored from a large segregating population of autotetraploid budding yeast *S. cerevisiae*. A significant decrease in the strength of crossover interference was found on one chromosome among the tested three chromosomes after polyploidization.

Part II-2

To investigate into the change in crossover events after genome duplication in *Saccharomyces cerevisiae*, a new statistical method was proposed to predict crossover rate based on whole genome sequencing data from tetrads of autotetraploid meiosis. In addition, we demonstrated

utility of the method by implementing it to analyse genotype dataset collected from the intensely distributed SNP markers based on next generation sequencing approach. Statistical comparison was made to all four spores derived from meiosis between diploid and its related autotetraploid *Saccharomyces cerevisiae*. An increase in the overall crossover rate was found after polyploidization in the real data analysis

1.3. References

Bailey, N.T. (1961) **Introduction of mathematical theory of genetic linkage**. Oxford University Press, Lodon.

Bonierbale, M.W., et al. (1988) RFLP maps based on common set of clones reveal modes of chromosomal evolution in potato and tomato. **Genetics,** 120: 1095-1103.

Butruille, D.V. and Boiteux, L.S. (2000) Selection-mutation balance in polysomic tetraploids: Impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. **PNAS,** 97(12): 6608-6613.

Blanc, G., et al. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. **Plant Cell,** 16: 1667-1678.

Bovell-Benjamin, A.C. (2007) Sweet potato: a review of its past, present, and future role in human nutrition. **Adv. Food Nutr. Res.**, 52: 1-59.

Clausen J., et al. (1945) II. **Plant evolution through amphiploidy and autoploidy, with examples from the Madiinae.** Washington DC, USA: Carnegie Institution of Washington.

Cavalier-Smith, T. (1978) The evolutionary origin and phylogeny of microtubules, mitotic spindles and eukaryote flagella. **Biosystems,** 10(1-2): 93-114.

Comai, L. (2005) The advantages and disadvantages of being polyploidy. **Nature Reviews Genetics,** 6: 836-846.

Chen, Z. (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. **Annu Rev Plant Biol**, 58: 377-406.

Christian, P., et al. (2010) Evolutionary consequences of autopolyploidy. **New Phytologist,** 186 (1): 5-17.

Doyle, J.J., et al. (2008) Evolutionary genetics of genome merger and doubling in plants. **Annual Review of Genetics,** 42: 443-461.

Fisher, R.A. (1930) **The genetical theory of natural selection.** Oxford, UK: Oxford Univ. Press.

Fisher, R.A. (1947) The theory of linkage in polysomic inheritance. **Phil. Trans. R. Soc. Lond. B.,** 233: 55-87.

Grant, V. (1981) Genetic-ecological investigations in *Bissutella levigata* I. **Vero¨ffentlichungen des Geobotanischen Institutes ETH**, 86: 7–86.

Gebhardt, C., et al. (1989) RFLP maps of potato and their alignment with the homologous tomato genome. **Theor. Appl. Genet.,** 83: 49-57.

Haldane, J.B.S. (1930) The theoretical genetics of autopolyploids. **J. Genet.,** 22: 359-372.

Jackson, R.C. (1982) Polyploidy and diploidy: new perspectives on chromosome pairing and its evolutionary implications. **American Journal of Botany**, 69: 1512-1523.

Jackson, R.C. and Jackson, J.W. (1996) Gene segregation in autotetraploids: prediction from meiotic configurations. **American Journal of Botany,** 83: 673-678.

Jiao, Y., et al. (2011) Ancestral polyploidy in seed plants and angiosperms. **Nature**, 473: 97-100.

Levin, D.A. (1983) Polyploidy and novelty in flowering plants. **The American Naturalist,** 122(1): 1-25.

Loidl, J. (1995) Meiotic chromosome pairing in triploid and tetraploid Saccharomyces cerevisiae. **Genetics,** 139:1511-1520.

Levin, D.A. (2002) **The role of chromosomal change in plant evolution.** New York, NY, USA: Oxford University Press.

Luo, Z.W., et al. (2004) Theoretical basis for genetic linkage analysis in autotetraploid species. **PNAS,** 101(18): 7040-7045.

Landergott, U., et al. (2006) Allelic configuration and polysomic inheritance of highly variable microsatellites in tetraploid gynodioecious Thymus praccoxagg. **Theoretical and Applied Genetics,** 113: 453-465.

Leitch, A.R. and Leitch, I.J. (2008) Genomic plasticity and the diversity of polyploidy plants. **Science,** 320: 481-465.

Mather, K. (1935) Reductional and equational separation of the chromosomes in bivalents and tetravalents. **J. Genet.,** 30: 53-78.

Mather, K. (1936) Segregation and linkage in autotetraploids. **J. Genet.,** 32: 287-314.

Mergoum, M., et al. (2009) Triticale: A "new" crop with old chanllenges. **Handbook of plant breeding**, 3: 267-287.

Otto, S.P. and Whitton, J. (2000) Polyploid incidence and evolution. **Annu. Rev. Genet.,** 34: 401-37.

Otto, S.P. (2007) The evolutionary consequences of polyploidy. **Cell,** 131: 452-462.

Ramsey, J. and Schemske, D.W. (1998) Pathways, mechanisms, and rates of polyploidy formation in flowering plants. **Annual Review Ecology and Systematic**, 29: 467-501.

Ronfort, J. (1998) Analysis of population structure in autotetraploi species. **Genetics,** 150(2): 921-930.

Ramsey, J., et al. (2002) Neopolyploidy in flowering plants. **Annual Review Ecology and Systematic**, 33: 589-639.

Stebbins, G.L. (1947) Types of polyploids: their classification and significance. **Adv. Genet.,** 1: 403-429.

Stebbins, G.L. (1971) **Chromosomal evolution in higher plants.** London, UK: Edward Amold.

Swaminathan, M.S. (1953) **The cytology and genetics of the potato (Solanum tuberosum) and related species.** Martinus Nijhoff.

Sved, J.A. (1964) The relationship between diploid and tetraploid recombination frequencies. **Heredity**, 19: 585-596.

Soltis, D.E. and Soltis, P.S. (1993) Molecular-data and the dynamic nature of polyploidy. **Critical Reviews in Plant Sciences**, 12: 243-273.

Soltis, P.S. and Soltis, D.E. (2000) The role of genetic and genomic attributes in the success of polyploids. **PNAS,** 97(13): 7051-7067.

Stift, M., et al. (2008) Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (yellow cress) microsatellite data. **Genetics**, 179: 2113-2123.

Soltis, D.E. and Soltis, P.S. (2009) The role of hybridization in plant speciation. **Annual Review of Plant Biology**, 60: 561-588.

United Nation, Department of Economic and Social Affairs. **World Population Prospects: 2012 Revision.**

PART I

THEORY AND METHODS FOR QTL ANALYSIS IN AUTOTETRAPLOIDS

Chapter I-1: Orthogonal contrast based models for quantitative genetic analysis in autotetraploid species

1.1. Overview

The quantitative genetic model which links genetic effects of genes or genotypes at quantitative trait loci to phenotype of quantitative traits is the basis for any quantitative genetic analysis. The theory and methods for modelling and analysing quantitative genetic effects have been well established and routinely practised in diploid species (Falconer 1996), whilst such study in polyploids, like autotetraploids, is still in its infancy mainly because of the complexity of polysomic inheritance aforementioned.

Kempthorne (1955) was probably the first in proposing statistical models for quantitative genetic effects in tetraploids, and formulated effects of tetraploid genotypes in randomly mating populations by simply extending the quantitative genetic model in diploids. In particular, the model involves a total of 15 parameters for genetic effects of segregating alleles and their successively higher orders of interaction at single locus (Kempthorne 1955, 1957). To simplify the Kempthorne's model, Li (1957) worked on a bi-allelic model and proposed regression of genetic values of genotypes onto the corresponding frequencies in a random mating tetraploid population (Li 1957). The successive linear regression model allows genetic variance at a single locus to be presented by four major components. Mather (1982) extended their concept of additive and dominance effects of diploid quantitative genetics to define these effects in tetraploid populations. Obviously, any quantitative genetic model and its analysis rely on

distribution of QTL genotypes in the population under study, and the genotypic distribution in an autotetraploid population replies on the coefficient of double reduction. Having noted these, Kilick (1971) explored the influence of double reduction on the Mather tetraploid additive and dominance model through its influence on frequencies of genotypes in segregating populations from crossing two inbred tetraploid parents. By integrating the additive and dominance model proposed by Mather and Jinks and the idea of Li, Wright (1979) presented the genetic value at single locus as a polynomial of order *n*, which is the level of ploidy, and the genetic variances were presented as the so called differential coefficients, which was in fact derivatives of the polynomial with respect to the allelic frequency (Wright 1979).

All of the models reviewed above are directly or indirectly extended from the classical diploid of their diploid counterparts. In contrast to the classical quantitative genetic model, Cockerham (1954) pioneered in attempting desirable statistical properties to quantitative genetic models in diploids and in turn, to the model based quantitative genetic analyses (Cockerham 1954). He developed the quantitative genetic model using the principle of orthogonal linear comparison, which enables phenotypic variation of a quantitative trait to be partitioned into independent components (Cockerham 1954). Zeng et al conducted comprehensive exploitation of statistical properties of the orthogonal contrast model in several major quantitative genetic analyses including estimation of additive, dominance and epistatic effects, partition of genetic variance components and mapping QTL (Kao 2002; Zeng 2005; Wang 2006). Although involving much more sophisticated algebraic formulation, the orthogonal model shows several key advantageous properties over the standard quantitative genetic models including the Kempthorne's and Mather and Jink's models. The orthogonality of the model ensures that estimate of one model parameter is independent of estimates of other parameters in the model and that the variance-covariance

matrix of quantitative genetic effects is diagonal, i.e. there is no genetic covariance between the different genetic effect parameters. These statistical properties confer quantitative genetic analyses the robustness to different settings of model parameters.

This sub-project presents novel tetrasomic quantitative genetic models for nature and artificially designed populations of autotetraploid species. The models account properly for the essential features of tetrasomic inheritance at the quantitative trait loci and their statistical properties and utilities in modelling real datasets are exploited by intensive computer simulations.

1.2. Introduction to orthogonal contrasts

ANOVA (Analysis of variance) is well known as a powerful tool and widely used to compare several treatment means in biological research field. In such statistical tests, the null hypothesis is that the T true means are all equal ($H_0 : \mu_1 = \mu_2 = \cdots = \mu_T$). Thus we would accept the alternative hypothesis if the F test is significant, where merely reflect at least one mean is different from the others. However, it could not tell which mean(s) is/are different. Further comparisons can be carried out by further decomposing the treatment sum of square (SST) to provide additional statistical test to answer planned questions. For this purpose, the orthogonal contrast has been historically proposed to answer specific research questions of interest and to compare mixed effects, both fixed effects and random effects (Scheffé 1959; Winer 1971; Steel and Torrie 1981; Mead 1988; Hinkelmann and Kempthorne 1994; Kuehl 2000). With a planned test, the orthogonal contrast scales are designed based on a priori knowledge, either on

biological considerations or on the results of preliminary investigations. It is not only a simple and efficient way to analyse experimental data, but also an alternative way to do statistical analysis on data without a definite structure (Nogueira 2004).

In statistics, a contrast is a linear combination of two or more variables (genetic effects here) whose coefficients add up to zero (Casella 2008). For instance, let $x_1,\ldots,x_t$ be a set of variables and $a_1,\ldots,a_t$ be known constants. The expression $\sum_{i=1}^{t} a_i x_i$ is a linear combination and it is called contrast if $\sum_{i=1}^{t} a_i = 0$. Furthermore, two contrasts, $\sum_{i=1}^{t} a_i x_i$ and $\sum_{i=1}^{t} b_i x_i$, are orthogonal if $\sum_{i=1}^{t} a_i b_i = 0$. A set of linear combinations must satisfy the following two mathematical constrains so to be in orthogonal contrasts:

(1). The sum of the coefficients in each linear contrast must sum to zero, and

(2). The sum of the products of the corresponding coefficients in any two contrasts must be equal to zero.

In ANOVA, the total sum of square (TSS) could be perfectly divided into two parts (the treatment SS (SST) and the error SS (SSE)) due to the fact that they are mathematically orthogonal to each other. Their relationship can be expressed as

$$TSS = SST + SSE \qquad \text{(I-1.1)}$$

In an analogous way, orthogonal contrasts provide a way to partition SST into as many independent comparisons as the degree of freedom for treatments in the ANOVA, each having

one degree of freedom. The pair-wise orthogonality as defined above ensures that the variance of the contrasts, equal to the weighted sum of the variances, will be uncorrelated, proceeding to minimize the Type I error rate (Howell 2010).

1.2.1. Orthogonal contrast based genetic model in diploids

It was first discovered by Mendel early in the year of 1865 that epistasis existed among genes controlling quantitative traits. Epistasis has been historically difficult to discern and insufficiently discussed in many theoretical and statistical issue (Brim 1961; Lee 1968; Stuber 1971; Stuber 1992; Cheverud 1995; Doebley 1995; Cockerham 1996; Kao 1999; Goodnight 2000; Zeng 2000). Following Fisher (1918), Cockerham introduced the principle of orthogonal contrasts to partition the epistatic variance into components in diploids (Cockerham 1954). He partitioned the genetic variance contributed by two genes into eight independent components by orthogonal contrast scales. Anderson and Kempthorne proposed a specific simplified model for a F_2 population based on orthogonal partitioning, called the F_2-metric model (Anderson 1954). In this model, additive effect was defined as half of the difference between the two homozygote genotypic values which is the same as that in the traditional method (Falconer 1996). However, dominance effect, defined as the difference between the mean of homozygote genotypic values and the heterozygote genotypic value, was scaled to zero for allelic frequency (Anderson 1954). Therefore, this F_2-metric model is no more defined only based on genotypic values but also based on allelic frequencies. To provide a better way to model QTL in a segregating population, Zeng et al made a justification for the general two-allele model based on orthogonal contrast

scales (Zeng 2005). Zeng et al also conducted a comprehensive exploitation of statistical property of the general orthogonal model in mapping epistatic genes in both linkage equilibrium and disequilibrium population (Kao 2002; Zeng 2005; Wang 2006). Orthogonality ensures consistency in the definition of genetic effects with multiple loci and independence between different effects and variance components, which make Cockerham's model outperform all the others in modelling and mapping QTLs (Zeng 2005). In the following, I present development of quantitative genetic models for autotetraploid species following the principles of the orthogonal contrast linear model.

1.3. Theoretical models and analysis for autotetraploids

1.3.1. One locus model

I first consider segregation of two alleles (*A* and *a*) at a single locus in an autotetraploid population. Frequency of the allele *A* in the population is denoted by *p* and the coefficient of double reduction at the locus by α. There are a total of 5 possible genotypes at the bi-allelic locus. Frequency of the *i*th genotype, $A_i a_{4-i}$, is denoted by f_i with $i = 0, 1, \ldots, 4$, indicating the number of *A* allele involved in the genotype.

I define here the phenotypic effect (i.e. trait value) for an individual through a regression model of allelic effects

$$P = G + \varepsilon = \mu + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \varepsilon \quad \text{(I-1.2)}$$

where *P* is the phenotypic effect, *G* is the genotypic effect, μ is the population mean, and θ_i (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL, and x_i (i=1, …, 4) are the corresponding genetic-effect design variables, and ε is a random variable following a normal distribution with zero mean.

In a natural autotetraploid population, genotypic frequencies vary across different loci in the genome and are usually not in Hardy-Weinberg equilibrium (Luo 2000). According to the basic principle of orthogonal comparison of linear statistical models (Wang 2006, Zeng 2005), I propose here general orthogonal scales $W_i = \{\omega_{ij}\}$ (i=1,…,4; j=0,…,4) listed in Table I-1.1 for the genetic effects (summarised in Table I-1.2) in the model (I-1.2). In the orthogonal scale vectors, ω_{ij} must satisfy

1. $\omega_{1j}\ (j = 0,1,...,4)$

$$\begin{cases} \omega_{1j} = u + j & (j = 0,1,...,4) \\ u = -4f_4 - 3f_3 - 2f_2 - f_1 \end{cases}$$

2. $\omega_{2j}\ (j = 0,1,...,4)$

$$\begin{cases} \omega_{2,j+1} - 2\omega_{2j} + \omega_{2,j-1} = 1 & (j = 1,2,3) \\ \sum_{j=0}^{4} \omega_{2j} f_j = 0 \\ \sum_{j=0}^{4} \omega_{2j}\omega_{1j} f_j = 0 \end{cases}$$

3. $\omega_{3j}\,(j=0,1,...,4)$

$$\begin{cases} \omega_{3,j+1}-3\omega_{3j}+3\omega_{3,j-1}-\omega_{3,j-2}=1 & (j=2,3) \\ \sum_{j=0}^{4}\omega_{3j}f_j=0 \\ \sum_{j=0}^{4}\omega_{3j}\omega_{1j}f_j=0 \\ \sum_{j=0}^{4}\omega_{3j}\omega_{2j}f_j=0 \end{cases}$$

4. $\omega_{4j}\,(j=0,1,...,4)$

$$\begin{cases} \omega_{44}-4\omega_{43}+6\omega_{42}-4\omega_{41}+\omega_{40}=1 \\ \sum_{j=0}^{4}\omega_{4j}f_j=0 \\ \sum_{j=0}^{4}\omega_{4j}\omega_{1j}f_j=0 \\ \sum_{j=0}^{4}\omega_{4j}\omega_{2j}f_j=0 \\ \sum_{j=0}^{4}\omega_{4j}\omega_{3j}f_j=0 \end{cases}$$

Incorporating them into the model (I-1.2) and replacing the genetic-effect design variables x_i by

$$x_i=\begin{cases} \omega_{i0} & \textit{if } G \textit{ is } aaaa \\ \omega_{i1} & \textit{if } G \textit{ is } Aaaa \\ \omega_{i2} & \textit{if } G \textit{ is } AAaa \\ \omega_{i3} & \textit{if } G \textit{ is } AAAa \\ \omega_{i4} & \textit{if } G \textit{ is } AAAA \end{cases} \qquad (i=1,2,\ldots,4)$$

I derive a matrix form of the orthogonal model for the QTL effects in form of

$$G_A = \begin{bmatrix} G_0 \\ G_1 \\ G_2 \\ G_3 \\ G_4 \end{bmatrix} = \begin{bmatrix} 1 & \omega_{10} & \omega_{20} & \omega_{30} & \omega_{40} \\ 1 & \omega_{11} & \omega_{21} & \omega_{31} & \omega_{41} \\ 1 & \omega_{12} & \omega_{22} & \omega_{32} & \omega_{42} \\ 1 & \omega_{13} & \omega_{23} & \omega_{33} & \omega_{43} \\ 1 & \omega_{14} & \omega_{24} & \omega_{34} & \omega_{44} \end{bmatrix} \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = S_A E_A \quad \text{(I-1.3)}$$

The genetic effects of the QTL genotypes can be calculated from

$$E_A = S_A^{-1} G_A \quad \text{(I-1.4)}$$

Table I-1.1. The orthogonal contrast scales for one biallelic locus (general model)

	Genotype	*aaaa*	*Aaaa*	*AAaa*	*AAAa*	*AAAA*
	Frequency	f_0	f_1	f_2	f_3	f_4
	G	G_0	G_1	G_2	G_3	G_4
θ_1	W_1	ω_{10}	ω_{11}	ω_{12}	ω_{13}	ω_{14}
θ_2	W_2	ω_{20}	ω_{21}	ω_{22}	ω_{23}	ω_{24}
θ_3	W_3	ω_{30}	ω_{31}	ω_{32}	ω_{33}	ω_{34}
θ_4	W_4	ω_{40}	ω_{41}	ω_{42}	ω_{43}	ω_{44}

G's and *f*'s denote the genotypic values and genotypic frequencies for the five genotypes. θ_i (*i*=1,2,…,4) are the monogenic, digenic, trigenic and quandrigenic genetic effects respectively. $\omega_{ij} \in W_j$ (*i*=1,…,4; *j*=0,1,…,4) is the scale component of genotype *i* for the *j* contrast.

Table I-1.2. Definition of genetic parameters for two loci model

Scales	Parameter definition	Notation	Scales	Parameter definition	Notation
W_0	Mean	μ	W_{13}	Digenic × monogenic effect of loci A and B	$I_{\theta_2\zeta_1}$
W_1	Monogenic effect of locus A	θ_1	W_{14}	Digenic × digenic effect of loci A and B	$I_{\theta_2\zeta_2}$
W_2	Digenic effect of locus A	θ_2	W_{15}	Digenic × trigenic effect of loci A and B	$I_{\theta_2\zeta_3}$
W_3	Trigenic effect of locus A	θ_3	W_{16}	Digenic × quadrigenic effect of loci A and B	$I_{\theta_2\zeta_4}$
W_4	Quadrigenic effect of locus A	θ_4	W_{17}	Trigenic × monogenic effect of loci A and B	$I_{\theta_3\zeta_1}$
W_5	Monogenic effect of locus B	ζ_1	W_{18}	Trigenic × digenic effect of loci A and B	$I_{\theta_3\zeta_2}$
W_6	Digenic effect of locus B	ζ_2	W_{19}	Trigenic × trigenic effect of loci A and B	$I_{\theta_3\zeta_3}$
W_7	Trigenic effect of locus B	ζ_3	W_{20}	Trigenic × quadrigenic effect of loci A and B	$I_{\theta_3\zeta_4}$
W_8	Quadrigenic effect of locus B	ζ_4	W_{21}	Quadrigenic × monogenic effect of loci A and B	$I_{\theta_4\zeta_1}$
W_9	Monogenic × monogenic effect of loci A and B	$I_{\theta_1\zeta_1}$	W_{22}	Quadrigenic × digenic effect of loci A and B	$I_{\theta_4\zeta_2}$
W_{10}	Monogenic × digenic effect of loci A and B	$I_{\theta_1\zeta_2}$	W_{23}	Quadrigenic × trigenic effect of loci A and B	$I_{\theta_4\zeta_3}$
W_{11}	Monogenic × trigenic effect of loci A and B	$I_{\theta_1\zeta_3}$	W_{24}	Quadrigenic × quadrigenic effect of loci A and B	$I_{\theta_4\zeta_4}$
W_{12}	Monogenic × quadrigenic effect of loci A and B	$I_{\theta_1\zeta_4}$			

Accordingly total genetic variance V_G, contributed by segregation of alleles at the QTL, can be partitioned into four independent components. Each variance component is contributed by its own relevant genetic parameters as

$$\sigma_t^2 = \frac{\left(\sum_{j=0}^{4} f_j G_j \omega_{tj}\right)^2}{\left(\sum_{j=0}^{4} f_j \omega_{tj}^2\right)} \qquad (t = 1, 2, ..., 4) \tag{I-1.5}$$

In the following section, we characterize the model (I-1.2) in two populations with two specified genetic structures.

1.3.1.1. Model for F_2 populations

In an F_2 population created from crossing two parental lines with genotypes *AAAA* and *aaaa*, frequencies of the offspring genotypes can be expressed in term of α, the coefficient of double reduction at the QTL, as, $f_0 = (1+2\alpha)^2/36$, $f_1 = 2(1-\alpha)(1+2\alpha)/9$, $f_2 = [3-4\alpha(1-\alpha)]/6$, $f_3 = 2(1-\alpha)(1+2\alpha)/9$ and $f_4 = (1+2\alpha)^2/36$. With these, the genotypic values $G_A = (G_0\ G_1\ G_2\ G_3\ G_4)^T$ can be presented in a matrix form of

$$G_A = S_A E_A = \begin{bmatrix} 1 & -2 & (5-2\alpha)/3 & -2(1-\alpha)/3 & (1-\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(12(2+\alpha)\right) \\ 1 & -1 & (1-4\alpha)/6 & (1+2\alpha)/6 & -(1+2\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(24(2+\alpha)\right) \\ 1 & 0 & -(1+2\alpha)/3 & 0 & (1-\alpha)(2\alpha+1)^2/\left(12(2+\alpha)\right) \\ 1 & 1 & (1-4\alpha)/6 & -(1+2\alpha)/6 & -(1+2\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(24(2+\alpha)\right) \\ 1 & 2 & (5-2\alpha)/3 & 2(1-\alpha)/3 & (1-\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(12(2+\alpha)\right) \end{bmatrix} \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

The genetic values can be calculated from $E_A = S_A^{-1} G_A$ where

$$S_A^{-1} = \begin{bmatrix} (1+2\alpha)^2/36 & 2(1+2\alpha)(1-\alpha)/9 & (4\alpha^2-4\alpha+3)/6 & 2(1+2\alpha)(1-\alpha)/9 & (1+2\alpha)^2/36 \\ -(1+2\alpha)/12 & (\alpha-1)/3 & 0 & (1-\alpha)/3 & (1+2\alpha)/12 \\ \frac{(1+2\alpha)(5-2\alpha)}{12(2+\alpha)} & \frac{(\alpha-1)(4\alpha-1)}{3(2+\alpha)} & -\frac{(4\alpha^2-4\alpha+3)}{2(2+\alpha)} & \frac{(\alpha-1)(4\alpha-1)}{3(2+\alpha)} & \frac{(1+2\alpha)(5-2\alpha)}{12(2+\alpha)} \\ -1/2 & 1 & 0 & -1 & 1/2 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

1.3.1.2. Model for randomly matting populations

We once worked out the equilibrium distribution of genotypes at a multi-allelic locus in randomly mating autotetraploid populations (Luo 2006). In the present context, the probability distribution of genotypes at a biallelic locus in randomly matting populations is presented by

$$f_0 = \frac{(1-p)^2}{(2+\alpha)^2}\left[9\alpha^2 + 12\alpha(1-\alpha)(1-p) + 4(1-\alpha)^2(1-p)^2\right]$$

$$f_1 = \frac{4(1-\alpha)p(1-p)^2}{(2+\alpha)^2}\left[6\alpha + 4(1-\alpha)(1-p)\right]$$

$$f_2 = \frac{6p(1-p)}{(2+\alpha)^2}\left[3\alpha^2 + 2\alpha(1-\alpha) + 4(1-\alpha)^2 p(1-p)\right]$$

$$f_3 = \frac{4(1-\alpha)p^2(1-p)}{(2+\alpha)^2}\left[6\alpha + 4(1-\alpha)p\right]$$

$$f_4 = \frac{p^2}{(2+\alpha)^2}\left[9\alpha^2 + 12\alpha(1-\alpha)p + 4(1-\alpha)^2 p^2\right]$$

Details of deriving this distribution can be found in our previous study (Luo 2006). It can be shown that $\sum_{j=0}^{4} f_j = 1$. In the above, allele frequency of *A* is denoted by p and the coefficient of double reduction at the locus is denoted by α. For simplicity but without loss of generality, the difference in frequency between alleles *A* and *a* is denoted as *s* ($s = p-(1-p) = 2p-1$). If the populations are in the Hardy- Weinberg equilibrium, I worked out the orthogonal contrast scales for the biallelic quantitative genetic model in the random-mating equilibrium population as listed below

1. For monogenic effects

$$x_1 = \begin{cases} \omega_{10} = -2-2s \\ \omega_{11} = -1-2s \\ \omega_{12} = -2s \\ \omega_{13} = 1-2s \\ \omega_{14} = 2-2s \end{cases}$$

2. For digenic effects

$$x_2=\begin{cases}\omega_{20}=\dfrac{3(s+1)}{m_2}\left[(s+1)+\alpha(s+2)+\alpha^2 s\right]\\ \omega_{21}=\dfrac{3}{2m_2}\left[2s(s+1)+\alpha\left(2s^2+3s-1\right)+\alpha^2\left(2s^2+s-2\right)\right]\\ \omega_{22}=\dfrac{1}{m_2}\left[3s^2-1+\left(\alpha+\alpha^2\right)\left(3s^2-4\right)\right]\\ \omega_{23}=\dfrac{3}{2m_2}\left[2s(s-1)+\alpha\left(2s^2-3s-1\right)+\alpha^2\left(2s^2-s-2\right)\right]\\ \omega_{24}=\dfrac{3(s-1)}{m_2}\left[(s-1)+\alpha(s-2)+\alpha^2 s\right]\end{cases}$$

where $m_2=2+5\alpha+2\alpha^2$

3. For trigenic effects

$$x_3=\begin{cases}\omega_{30}=\dfrac{(\alpha-1)(s+1)^2}{m_3}\left[\begin{array}{l}\alpha^3\left(2s^3-7s^2+8s-4\right)+\alpha^2\left(-3s^3+3s^2+4s-8\right)\\+\alpha\left(3s^2-2s-5\right)+(s-1)(s+1)^2\end{array}\right]\\ \omega_{31}=\dfrac{(s-1)}{2m_3}\left[\begin{array}{l}\alpha^4\left(4s^4-7s^3-10s^2+20s-8\right)-\alpha^3\left(10s^4-10s^3+3s^2-20s+20\right)\\+\alpha^2\left(6s^4+3s^2+9s-18\right)+\alpha\left(2s^4-2s^3+7s^2+4s-7\right)-\left(2s^4+s^3-3s^2-s+1\right)\end{array}\right]\\ \omega_{32}=\dfrac{s\left(s^2-1\right)\left(2\alpha^2-\alpha-1\right)}{m_3}\left[\alpha^2\left(s^2-4\right)-2\alpha\left(s^2+2\right)+\left(s^2-1\right)\right]\\ \omega_{33}=\dfrac{(s-1)}{2m_3}\left[\begin{array}{l}\alpha^4\left(4s^4+7s^3-10s^2-20s-8\right)-\alpha^3\left(10s^4+10s^3+3s^2+20s+20\right)\\+\alpha^2\left(6s^4+3s^2-9s-18\right)+\alpha\left(2s^4+2s^3+7s^2-4s-7\right)-\left(2s^4-s^3-3s^2+s+1\right)\end{array}\right]\\ \omega_{34}=\dfrac{(\alpha-1)(s-1)^2}{m_3}\left[\begin{array}{l}\alpha^3\left(2s^3+7s^2+8s+4\right)+\alpha^2\left(-3s^3-3s^2+4s+8\right)\\+\alpha\left(-3s^2-2s+5\right)+(s-1)^2(s+1)\end{array}\right]\end{cases}$$

where $m_3=(2+\alpha)\left[\alpha^3\left(5s^2-4\right)+\alpha^2\left(9s^2-8\right)+\alpha\left(3s^2-5\right)+s^2-1\right]$

4. For quadrigenic effects

$$x_4=\begin{cases}\omega_{40}=\dfrac{1}{m_4}\left\{(1-\alpha)(1+s)^2\left[\alpha(2-s)+1+s\right]^2\left[\alpha^2(2-s^2)+2\alpha s^2+1-s^2\right]\right\}\\ \omega_{41}=\dfrac{1}{m_4}\left\{(s+1)\left[\alpha(2-s)+1+s\right]^2\left[\begin{array}{l}\alpha^3(s^3+2s^2-2s-4)-\alpha^2(3s^3+3s^2-2s+2)\\+\alpha(3s^3-s-2)-(1+s)(1-s)^2\end{array}\right]\right\}\\ \omega_{42}=\dfrac{1}{m_4}\left\{(1-\alpha)(1-s^2)\left[\alpha^2(s^2-4)-2\alpha(s^2+2)+s^2-1\right]^2\right\}\\ \omega_{43}=\dfrac{1}{m_4}\left\{(s-1)\left[\alpha(2+s)+1-s\right]^2\left[\begin{array}{l}\alpha^3(s^3-2s^2-2s+4)-\alpha^2(3s^3-3s^2-2s-2)\\+\alpha(3s^3-s+2)+(1-s)(1+s)^2\end{array}\right]\right\}\\ \omega_{44}=\dfrac{1}{m_4}\left\{(1-\alpha)(1-s)^2\left[\alpha(2+s)+1-s\right]^2\left[\alpha^2(2-s^2)+2\alpha s^2+1-s^2\right]\right\}\end{cases}$$

where $m_4=4(2+\alpha)^2\left[\alpha^3(s^2+4)+\alpha^2(3s^2+8)-\alpha(3s^2-5)-(s^2-1)\right]$

In the same formulation, the quantitative genetic model is fully characterized as $G_A=S_AE_A$ and $E_A=S_A^{-1}G_A$. S_A and S_A^{-1} are detailed as.

$$S_A = \begin{bmatrix}
1 & -2-2s & \frac{3(s+1)}{m_2}\left[(s+1)+\alpha(s+2)+\alpha^2 s\right] & \frac{(\alpha-1)(s+1)^2}{m_3}\begin{bmatrix}\alpha^3(2s^3-7s^2+8s-4)+\alpha^2(-3s^3+3s^2+4s-8)\\ +\alpha(3s^2-2s-5)+(s-1)(s+1)^2\end{bmatrix} & \frac{(1-\alpha)(1+s)^2}{m_4}\left[\alpha(2-s)+1+s\right]^2\left[\alpha^2(2-s^2)+2\alpha s^2+1-s^2\right] \\
1 & -1-2s & \frac{3}{2m_2}\begin{bmatrix}2s(s+1)+\alpha(2s^2+3s-1)\\ +\alpha^2(2s^2+s-2)\end{bmatrix} & \frac{(s-1)}{2m_3}\begin{bmatrix}\alpha^4(4s^4-7s^3-10s^2+20s-8)\\ -\alpha^3(10s^4-10s^3+3s^2-20s+20)\\ +\alpha^2(6s^4+3s^2+9s-18)\\ +\alpha(2s^4-2s^3+7s^2+4s-7)\\ -(2s^4+s^3-3s^2-s+1)\end{bmatrix} & \frac{(s+1)}{m_4}\left[\alpha(2-s)+1+s\right]^2\begin{bmatrix}\alpha^3(s^3+2s^2-2s-4)-\\ \alpha^2(3s^3+3s^2-2s+2)+\\ \alpha(3s^3-s-2)-(1+s)(1-s)^2\end{bmatrix} \\
1 & -2s & \frac{1}{m_2}\left[3s^2-1+(\alpha+\alpha^2)(3s^2-4)\right] & \frac{s(s^2-1)(2\alpha^2-\alpha-1)}{m_3}\left[\alpha^2(s^2-4)-2\alpha(s^2+2)+(s^2-1)\right] & \frac{(1-\alpha)(1-s^2)}{m_4}\left[\alpha^2(s^2-4)-2\alpha(2+s^2)+s^2-1\right]^2 \\
1 & 1-2s & \frac{3}{2m_2}\begin{bmatrix}2s(s-1)+\alpha(2s^2-3s-1)\\ +\alpha^2(2s^2-s-2)\end{bmatrix} & \frac{(s-1)}{2m_3}\begin{bmatrix}\alpha^4(4s^4+7s^3-10s^2-20s-8)\\ -\alpha^3(10s^4+10s^3+3s^2+20s+20)\\ +\alpha^2(6s^4+3s^2-9s-18)\\ +\alpha(2s^4+2s^3+7s^2-4s-7)\\ -(2s^4-s^3-3s^2+s+1)\end{bmatrix} & \frac{(s-1)}{m_4}\left[\alpha(2+s)+1-s\right]^2\begin{bmatrix}\alpha^3(s^3-2s^2-2s+4)-\\ \alpha^2(3s^3-3s^2-2s-2)+\\ \alpha(3s^3-s+2)+(1-s)(1+s)^2\end{bmatrix} \\
1 & 2-2s & \frac{3(s-1)}{m_2}\left[(s-1)+\alpha(s-2)+\alpha^2 s\right] & \frac{(\alpha-1)(s-1)^2}{m_3}\begin{bmatrix}\alpha^3(2s^3+7s^2+8s+4)+\alpha^2(-3s^3-3s^2+4s+8)\\ +\alpha(-3s^2-2s+5)+(s-1)^2(s+1)\end{bmatrix} & \frac{(1-\alpha)(1-s)^2}{m_4}\left[\alpha(2+s)+1-s\right]^2\left[\alpha^2(2-s^2)+2\alpha s^2+1-s^2\right]
\end{bmatrix}$$

$$S_A^{-1} = \begin{bmatrix}
\frac{(s-1)^2}{4m_1}\left[1-s+\alpha(2+s)\right]^2 & \frac{(1-\alpha)(1-s)^2(1+s)}{m_1}\begin{bmatrix}1-s\\+\alpha(2+s)\end{bmatrix} & \frac{3(s^2-1)}{2m_1}\begin{bmatrix}s^2-1-2\alpha s^2\\+\alpha^2(s^2-2)\end{bmatrix} & \frac{(1-\alpha)(1-s)(1+s)^2}{m_1}\begin{bmatrix}1+s\\+\alpha(2-s)\end{bmatrix} & \frac{(s+1)^2}{4m_1}\left[1+s+\alpha(2-s)\right]^2 \\
\frac{(s-1)}{4m_2}\left[1-s+\alpha(2+s)\right]^2 & \frac{(1-\alpha)(2s^2-s-1)}{2m_2}\begin{bmatrix}1-s\\+\alpha(2+s)\end{bmatrix} & \frac{3s}{2m_2}\begin{bmatrix}s^2-1-2\alpha s^2\\+\alpha^2(s^2-2)\end{bmatrix} & \frac{(\alpha-1)(2s^2+s-1)}{2m_2}\begin{bmatrix}1+s\\+\alpha(2-s)\end{bmatrix} & \frac{(s+1)}{4m_2}\left[1+s+\alpha(2-s)\right]^2 \\
\frac{(s-1)}{2m_3}\left\{\begin{matrix}\left[1-s+\alpha(2+s)\right]^2\times\\\left[s+1+\alpha(s+2)+\alpha^2 s\right]\end{matrix}\right\} & \frac{(\alpha-1)(1-s)}{m_3}\begin{bmatrix}2s(1-s^2)\\-\alpha(1-8s-5s^2)\\-\alpha^2(4-8s-8s^2)\\-\alpha^3(4-5s^2-2s^3)\end{bmatrix} & \frac{1}{m_3}\begin{bmatrix}1-4s^2+3s^4\\+\alpha(4-5s^2-3s^4)\\+6\alpha^2(1-s^2)\\+\alpha^3(8-2s^2-3s^4)\\+\alpha^4(8-10s^2+3s^4)\end{bmatrix} & \frac{(1-\alpha)(1+s)}{m_3}\begin{bmatrix}2s(1-s^2)\\+\alpha(1+8s-5s^2)\\+\alpha^2(4+8s-8s^2)\\+\alpha^3(4-5s^2+2s^3)\end{bmatrix} & \frac{(s+1)}{2m_3}\left\{\begin{matrix}\left[1+s+\alpha(2-s)\right]^2\times\\\left[s-1+\alpha(s-2)+\alpha^2 s\right]\end{matrix}\right\} \\
\frac{2(2+\alpha)^2}{m_4}\begin{bmatrix}-(1-s)^2(1+s)\\-\alpha(5-2s-3s^2)\\-\alpha^2(8-2s+3s^2-3s^3)\\-\alpha^3(4-4s+s^2+2s^3)\end{bmatrix} & \frac{4(2+\alpha)^2}{m_4}\begin{bmatrix}(2s-1)(s^2-1)\\+\alpha(5-4s-3s^2)\\+\alpha^2(8-4s+3s^2-6s^3)\\+\alpha^3(4-8s+s^2+4s^3)\end{bmatrix} & \frac{4(2+\alpha)^2}{m_4}\left[3(1+2\alpha)s\begin{pmatrix}1-s^2\\+2\alpha s\\+\alpha^2(2-s^2)\end{pmatrix}\right] & \frac{4(2+\alpha)^2}{m_4}\begin{bmatrix}(2s+1)(s^2-1)\\-\alpha(5+4s-3s^2)\\-\alpha^2(8+4s+3s^2+6s^3)\\-\alpha^3(4+8s+s^2-4s^3)\end{bmatrix} & \frac{2(2+\alpha)^2}{m_4}\begin{bmatrix}(1-s)(1+s)^2\\+\alpha(5+2s-3s^2)\\+\alpha^2(8+2s+3s^2+3s^3)\\+\alpha^3(4+4s+s^2-2s^3)\end{bmatrix} \\
1 & -4 & 6 & -4 & 1
\end{bmatrix}$$

Here $m_1=(2+\alpha)^2$, $m_2=2\alpha^2+5\alpha+2$, $m_3=(2+\alpha)\left[s^2-1+\alpha(3s^2-5)+\alpha^2(9s^2-8)+\alpha^3(5s^2-4)\right]$ and $m_4=4(2+\alpha)^2$ $\left[\alpha^3(s^2+4)+\alpha^2(3s^2+8)-\alpha(3s^2-5)-(s^2-1)\right]$. s is denoted as $p-q$, with p, q as the frequency of allele *A* and allele *a* at the locus respectively. α is the coefficient of double reduction of locus A.

1.3.2. Two loci model

In this session, the one locus method described above is extended to two bi-allelic loci, A and B, in an autotetraploid population with a specified genetic structure. There will be twenty-five possible genotypes at the two loci (without accounting for linkage phase). A general form of a two-locus tetraploid genotype may be presented as $A_i a_{(4-i)} B_j b_{(4-j)}$ with $i = 0,1,...,4$ for the number of allele A and $j = 0,1,...,4$ for the number of allele B in the genotype. The genotypic value and frequency of the genotype are denoted by G_{ij} and f_{ij}, and $f_{i.}$ and $f_{.i}$ ($i = 0,1,...,4$) are the marginal frequency of genotypes at locus A and locus B. Frequencies of the allele *A* and B in the population are denoted by p_A and p_B, and the coefficients of double reduction at locus A and locus B are denoted by α and β respectively. A liner model for the phenotypic value is comprised of genic effects at each of the two loci, epistatic effects of genes at the two loci and a random variable, and is fully characterized by a total of twenty-five parameters in form of

$$\begin{aligned} P_{ij} = G_{ij} + \varepsilon = \mu + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \zeta_1 y_1 + \zeta_2 y_2 + \zeta_3 y_3 + \zeta_4 y_4 + I_{\theta_1\zeta_1} w_{\theta_1\zeta_1} + I_{\theta_1\zeta_2} w_{\theta_1\zeta_2} + \\ I_{\theta_1\zeta_3} w_{\theta_1\zeta_3} + I_{\theta_1\zeta_4} w_{\theta_1\zeta_4} + I_{\theta_2\zeta_1} w_{\theta_2\zeta_1} + I_{\theta_2\zeta_2} w_{\theta_2\zeta_2} + I_{\theta_2\zeta_3} w_{\theta_2\zeta_3} + I_{\theta_2\zeta_4} w_{\theta_2\zeta_4} + I_{\theta_3\zeta_1} w_{\theta_3\zeta_1} + \\ I_{\theta_3\zeta_2} w_{\theta_3\zeta_2} + I_{\theta_3\zeta_3} w_{\theta_3\zeta_3} + I_{\theta_3\zeta_4} w_{\theta_3\zeta_4} + I_{\theta_4\zeta_1} w_{\theta_4\zeta_1} + I_{\theta_4\zeta_2} w_{\theta_4\zeta_2} + I_{\theta_4\zeta_3} w_{\theta_4\zeta_3} + I_{\theta_4\zeta_4} w_{\theta_4\zeta_4} + \varepsilon \end{aligned}$$

(I-1.6)

where P_{ij} is the phenotypic effect, G_{ij} is the genotypic effect, μ is the population mean, θ_i (or ζ_i) (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects at locus A (or B), and x_i (or y_i) (i=1, …, 4) are design variables for the corresponding genetic-

effects. $I_{\theta_i\zeta_j}$ are epistasis between the effects θ_i and ζ_j (i=1,…,4; j=1,…,4). Table I-1.2 (On page 31) lists detailed descriptions of the parameters. ε is a normal residual variable.

In a similar but algebraically more tedious way, we derived the orthogonal contrast scales for the two-locus tetrasomic model under two different scenarios of mutual dependency of genotypic distribution at the two loci: linkage equilibrium and linkage disequilibrium.

1.3.2.1. Model for linkage equilibrium population

When alleles at the two loci model are in linkage equilibrium in the population under question, the probability of a joint genotype at the two loci equals the product of probabilities of genotypes at each locus, i.e. $f_{ij} = f_{i.} \times f_{.j}$ $(i, j = 0,1,...,4)$, the design variables for the genetic-effects in the Equation (I-1.6) can be written as

$$x_i = \begin{cases} \omega_{i0} & if\ G_A\ is\ aaaa \\ \omega_{i1} & if\ G_A\ is\ Aaaa \\ \omega_{i2} & if\ G_A\ is\ AAaa \\ \omega_{i3} & if\ G_A\ is\ AAAa \\ \omega_{i4} & if\ G_A\ is\ AAAA \end{cases} \quad \text{and} \quad y_i = \begin{cases} \upsilon_{i0} & if\ G_B\ is\ bbbb \\ \upsilon_{i1} & if\ G_B\ is\ Bbbb \\ \upsilon_{i2} & if\ G_B\ is\ BBbb \\ \upsilon_{i3} & if\ G_B\ is\ BBBb \\ \upsilon_{i4} & if\ G_B\ is\ BBBB \end{cases} \quad i = 1,2,3,4$$

$$w_{\theta_i\zeta_j} = x_i \times y_j \ (i, j = 1,2,3,4)$$

Here ω_{ij} and υ_{ij} (i =0,1,…,4; j=1,2,…,4) are the orthogonal contrast scales calculated separately according to the way for the bi-allelic one locus model as listed in Table I-1.3.

Table I-1.3. The orthogonal contrast scales for locus A and locus B

	Genotype	*aaaa*	*Aaaa*	*AAaa*	*AAAa*	*AAAA*
Locus A	Frequency	$f_{0.}$	$f_{1.}$	$f_{2.}$	$f_{3.}$	$f_{4.}$
	G	$G_{0.}$	$G_{1.}$	$G_{2.}$	$G_{3.}$	$G_{4.}$
θ_1	W_1	ω_{10}	ω_{11}	ω_{12}	ω_{13}	ω_{14}
θ_2	W_2	ω_{20}	ω_{21}	ω_{22}	ω_{23}	ω_{24}
θ_3	W_3	ω_{30}	ω_{31}	ω_{32}	ω_{33}	ω_{34}
θ_4	W_4	ω_{40}	ω_{41}	ω_{42}	ω_{43}	ω_{44}

	Genotype	*bbbb*	*Bbbb*	*BBbb*	*BBBb*	*BBBB*
Locus B	Frequency	$f_{.0}$	$f_{.1}$	$f_{.2}$	$f_{.3}$	$f_{.4}$
	G	$G_{.0}$	$G_{.1}$	$G_{.2}$	$G_{.3}$	$G_{.4}$
ζ_1	V_1	υ_{10}	υ_{11}	υ_{12}	υ_{13}	υ_{14}
ζ_2	V_2	υ_{20}	υ_{21}	υ_{22}	υ_{23}	υ_{24}
ζ_3	V_3	υ_{30}	υ_{31}	υ_{32}	υ_{33}	υ_{34}
ζ_4	V_4	υ_{40}	υ_{41}	υ_{42}	υ_{43}	υ_{44}

$G_{i.}$ ($G_{.i}$) and $f_{i.}$ ($f_{.i}$) (i = 0, 1,…, 4) denote the genotypic values and genotypic frequencies for the five genotypes of locus A (locus B). θ_i (ζ_i) (i=1, 2,…, 4) are the monogenic, digenic, trigenic and quadrigenic effects for locus A (locus B) respectively. Here $W_i = \{\omega_{ij}\}$ and $V_i = \{\upsilon_{ij}\}$ (i=1,…,4; j=0, …,4), the orthogonal contrast scales calculated separately via the general bi-allelic one locus model, are the scale component of genotype i for the j contrast.

A matrix form of the model can thus be written as

$$G_{AB} = S_{AB}E_{AB} = \left[S_A \otimes S_B\right]E_{AB} \tag{I-1.7}$$

$$E_{AB} = S_{AB}^{-1}G_{AB} = \left[\left(S_A^{-1}\right)\otimes\left(S_B^{-1}\right)\right]G_{AB} \tag{I-1.8}$$

Here $G_{AB} = (G_{00}\cdots G_{04}\cdots G_{44})^T$, $E_{AB} = (\mu\ \theta_1\cdots\theta_4\ \zeta_1\cdots\zeta_4\ I_{\theta_1\zeta_1}\cdots I_{\theta_4\zeta_4})^T$ and ' $\otimes$ ' stands for the Kronecker product. The product $S_A \otimes S_B$ needs to have some columns to be rearranged to match change in dimension of E_{AB}. S_{AB}^{-1} is equal to the Kronecker product of $\left(S_A^{-1}\right)$ and $\left(S_B^{-1}\right)$ with some rows that need to be rearranged correspondingly. Then the total genetic variance, V_G, contributed by the two loci, can be partitioned into twenty-four independent variance components. Each of the variance components is involved only with its own genetic parameter. The *t*-th component can be written as

$$\sigma_t^2 = \frac{\left(\sum_{i,j} f_{ij}G_{ij}w_{tij}\right)^2}{\left(\sum_{i,j} f_{ij}w_{tij}^2\right)}, \qquad t = 1,2,...,24 \tag{I-1.9}$$

where $w_{tij} \in W_t$ (t=1,2,…,8) is the scale component of genotype *ij* for the *t*th contrast. W_1, W_2, W_3 and W_4 (W_5, W_6, W_7 and W_8) are the orthogonal contrast for monogenic, digenic, trigenic and quadrigenic effects of locus A (locus B). W_t (*t*=9, …, 24) are the orthogonal contrast for the interaction effect between loci A and B, as shown in Table I-1.2 (On page 31). For simplicity but without loss of generality, we illustrate the model and analysis with a specific F_2 population, in which double reduction rate are set to be zero at both the loci for algebraic simplicity below

1.3.2.1.1. The orthogonal contrasts for a bi-allelic two loci model in an F_2 population.

Here $\alpha = 0, \beta = 0$, assuming linkage equilibrium between locus A and locus B. The genetic-effect design matrix for two loci, S_{AB}, is a Kronecker product ($\otimes$) of two one-locus design matrices, S_A and S_B, with some columns rearranged to conform the order in E_{AB}. S_{AB}^{-1} is a Kronecker product of $\left(S_A^{-1}\right)$ and $\left(S_B^{-1}\right)$ with some rows rearranged correspondingly

$$G_{AB} = \left[S_A \otimes S_B\right] E_{AB} = S_{AB} E_{AB}, \qquad S_A = S_B = \begin{bmatrix} 1 & -2 & 5/3 & -2/3 & 1/8 \\ 1 & -1 & 1/6 & 1/6 & -1/16 \\ 1 & 0 & -1/3 & 0 & 1/24 \\ 1 & 1 & 1/6 & -1/6 & -1/16 \\ 1 & 2 & 5/3 & 2/3 & 1/8 \end{bmatrix}$$

$$E_{AB} = S_{AB}^{-1} G_{AB} = \left[\left(S_A^{-1}\right) \otimes \left(S_B^{-1}\right)\right] G_{AB}, \qquad S_A^{-1} = S_B^{-1} = \begin{bmatrix} 1/36 & 2/9 & 1/2 & 2/9 & 1/36 \\ -1/12 & -1/3 & 0 & 1/3 & 1/12 \\ 5/24 & 1/6 & -3/4 & 1/6 & 5/24 \\ -1/2 & 1 & 0 & -1 & 1/2 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

with detailed as

$$
\begin{bmatrix} G_{00} \\ G_{01} \\ G_{02} \\ G_{03} \\ G_{04} \\ G_{10} \\ G_{11} \\ G_{12} \\ G_{13} \\ G_{14} \\ G_{20} \\ G_{21} \\ G_{22} \\ G_{23} \\ G_{24} \\ G_{30} \\ G_{31} \\ G_{32} \\ G_{33} \\ G_{34} \\ G_{40} \\ G_{41} \\ G_{42} \\ G_{43} \\ G_{44} \end{bmatrix} =
\begin{bmatrix}
1 & -2 & 5/3 & -2/3 & 1/8 & -2 & 5/3 & -2/3 & 1/8 & 4 & -10/3 & 4/3 & -1/4 & -10/3 & 25/9 & -10/9 & 5/24 & 4/3 & -10/9 & 4/9 & -1/12 & -1/4 & 5/24 & -1/12 & 1/64 \\
1 & -2 & 5/3 & -2/3 & 1/8 & -1 & 1/6 & 1/6 & -1/16 & 2 & -1/3 & -1/3 & 1/8 & -5/3 & 5/18 & 5/18 & -5/48 & 2/3 & -1/9 & -1/9 & 1/24 & -1/8 & 1/48 & 1/48 & -1/128 \\
1 & -2 & 5/3 & -2/3 & 1/8 & 0 & -1/3 & 0 & 1/24 & 0 & 2/3 & 0 & -1/12 & 0 & -5/9 & 0 & 5/72 & 0 & 2/9 & 0 & -1/36 & 0 & -1/24 & 0 & 1/192 \\
1 & -2 & 5/3 & -2/3 & 1/8 & 1 & 1/6 & -1/6 & -1/16 & -2 & -1/3 & 1/3 & 1/8 & 5/3 & 5/18 & -5/18 & -5/48 & -2/3 & -1/9 & 1/9 & 1/24 & 1/8 & 1/48 & -1/48 & -1/128 \\
1 & -2 & 5/3 & -2/3 & 1/8 & 2 & 5/3 & 2/3 & 1/8 & -4 & -10/3 & -4/3 & -1/4 & 10/3 & 25/9 & 10/9 & 5/24 & -4/3 & -10/9 & -4/9 & -1/12 & 1/4 & 5/24 & 1/12 & 1/64 \\
1 & -1 & 1/6 & 1/6 & -1/16 & -2 & 5/3 & -2/3 & 1/8 & 2 & -5/3 & 2/3 & -1/8 & -1/3 & 5/18 & -1/9 & 1/48 & -1/3 & 5/18 & -1/9 & 1/48 & 1/8 & -5/48 & 1/24 & -1/128 \\
1 & -1 & 1/6 & 1/6 & -1/16 & -1 & 1/6 & 1/6 & -1/16 & 1 & -1/6 & -1/6 & 1/16 & -1/6 & 1/36 & 1/36 & -1/96 & -1/6 & 1/36 & 1/36 & -1/96 & 1/16 & -1/96 & -1/96 & 1/256 \\
1 & -1 & 1/6 & 1/6 & -1/16 & 0 & -1/3 & 0 & 1/24 & 0 & 1/3 & 0 & -1/24 & 0 & -1/18 & 0 & 1/144 & 0 & -1/18 & 0 & 1/144 & 0 & 1/48 & 0 & -1/384 \\
1 & -1 & 1/6 & 1/6 & -1/16 & 1 & 1/6 & -1/6 & -1/16 & -1 & -1/6 & 1/6 & 1/16 & 1/6 & 1/36 & -1/36 & -1/96 & 1/6 & 1/36 & -1/36 & -1/96 & -1/16 & -1/96 & 1/96 & 1/256 \\
1 & -1 & 1/6 & 1/6 & -1/16 & 2 & 5/3 & 2/3 & 1/8 & -2 & -5/3 & -2/3 & -1/8 & 1/3 & 5/18 & 1/9 & 1/48 & 1/3 & 5/18 & 1/9 & 1/48 & -1/8 & -5/48 & -1/24 & -1/128 \\
1 & 0 & -1/3 & 0 & 1/24 & -2 & 5/3 & -2/3 & 1/8 & 0 & 0 & 0 & 0 & 2/3 & -5/9 & 2/9 & -1/24 & 0 & 0 & 0 & 0 & -1/12 & 5/72 & -1/36 & 1/192 \\
1 & 0 & -1/3 & 0 & 1/24 & -1 & 1/6 & 1/6 & -1/16 & 0 & 0 & 0 & 0 & 1/3 & -1/18 & -1/18 & 1/48 & 0 & 0 & 0 & 0 & -1/24 & 1/144 & 1/144 & -1/384 \\
1 & 0 & -1/3 & 0 & 1/24 & 0 & -1/3 & 0 & 1/24 & 0 & 0 & 0 & 0 & 0 & 1/9 & 0 & -1/72 & 0 & 0 & 0 & 0 & 0 & -1/72 & 0 & 1/576 \\
1 & 0 & -1/3 & 0 & 1/24 & 1 & 1/6 & -1/6 & -1/16 & 0 & 0 & 0 & 0 & -1/3 & -1/18 & 1/18 & 1/48 & 0 & 0 & 0 & 0 & 1/24 & 1/144 & -1/144 & -1/384 \\
1 & 0 & -1/3 & 0 & 1/24 & 2 & 5/3 & 2/3 & 1/8 & 0 & 0 & 0 & 0 & -2/3 & -5/9 & -2/9 & -1/24 & 0 & 0 & 0 & 0 & 1/12 & 5/72 & 1/36 & 1/192 \\
1 & 1 & 1/6 & -1/6 & -1/16 & -2 & 5/3 & -2/3 & 1/8 & -2 & 5/3 & -2/3 & 1/8 & -1/3 & 5/18 & -1/9 & 1/48 & 1/3 & -5/18 & 1/9 & -1/48 & 1/8 & -5/48 & 1/24 & -1/128 \\
1 & 1 & 1/6 & -1/6 & -1/16 & -1 & 1/6 & 1/6 & -1/16 & -1 & 1/6 & 1/6 & -1/16 & -1/6 & 1/36 & 1/36 & -1/96 & 1/6 & -1/36 & -1/36 & 1/96 & 1/16 & -1/96 & -1/96 & 1/256 \\
1 & 1 & 1/6 & -1/6 & -1/16 & 0 & -1/3 & 0 & 1/24 & 0 & -1/3 & 0 & 1/24 & 0 & -1/18 & 0 & 1/144 & 0 & 1/18 & 0 & -1/144 & 0 & 1/48 & 0 & -1/384 \\
1 & 1 & 1/6 & -1/6 & -1/16 & 1 & 1/6 & -1/6 & -1/16 & 1 & 1/6 & -1/6 & -1/16 & 1/6 & 1/36 & -1/36 & -1/96 & -1/6 & -1/36 & 1/36 & 1/96 & -1/16 & -1/96 & 1/96 & 1/256 \\
1 & 1 & 1/6 & -1/6 & -1/16 & 2 & 5/3 & 2/3 & 1/8 & 2 & 5/3 & 2/3 & 1/8 & 1/3 & 5/18 & 1/9 & 1/48 & -1/3 & -5/18 & -1/9 & -1/48 & -1/8 & -5/48 & -1/24 & -1/128 \\
1 & 2 & 5/3 & 2/3 & 1/8 & -2 & 5/3 & -2/3 & 1/8 & -4 & 10/3 & -4/3 & 1/4 & -10/3 & 25/9 & -10/9 & 5/24 & -4/3 & 10/9 & -4/9 & 1/12 & -1/4 & 5/24 & -1/12 & 1/64 \\
1 & 2 & 5/3 & 2/3 & 1/8 & -1 & 1/6 & 1/6 & -1/16 & -2 & 1/3 & 1/3 & -1/8 & -5/3 & 5/18 & 5/18 & -5/48 & -2/3 & 1/9 & 1/9 & -1/24 & -1/8 & 1/48 & 1/48 & -1/128 \\
1 & 2 & 5/3 & 2/3 & 1/8 & 0 & -1/3 & 0 & 1/24 & 0 & -2/3 & 0 & 1/12 & 0 & -5/9 & 0 & 5/72 & 0 & -2/9 & 0 & 1/36 & 0 & -1/24 & 0 & 1/192 \\
1 & 2 & 5/3 & 2/3 & 1/8 & 1 & 1/6 & -1/6 & -1/16 & 2 & 1/3 & -1/3 & -1/8 & 5/3 & 5/18 & -5/18 & -5/48 & 2/3 & 1/9 & -1/9 & -1/24 & 1/8 & 1/48 & -1/48 & -1/128 \\
1 & 2 & 5/3 & 2/3 & 1/8 & 2 & 5/3 & 2/3 & 1/8 & 4 & 10/3 & 4/3 & 1/4 & 10/3 & 25/9 & 10/9 & 5/24 & 4/3 & 10/9 & 4/9 & 1/12 & 1/4 & 5/24 & 1/12 & 1/64
\end{bmatrix}
\times
\begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ I_{\theta_1\zeta_1} \\ I_{\theta_1\zeta_2} \\ I_{\theta_1\zeta_3} \\ I_{\theta_1\zeta_4} \\ I_{\theta_2\zeta_1} \\ I_{\theta_2\zeta_2} \\ I_{\theta_2\zeta_3} \\ I_{\theta_2\zeta_4} \\ I_{\theta_3\zeta_1} \\ I_{\theta_3\zeta_2} \\ I_{\theta_3\zeta_3} \\ I_{\theta_3\zeta_4} \\ I_{\theta_4\zeta_1} \\ I_{\theta_4\zeta_2} \\ I_{\theta_4\zeta_3} \\ I_{\theta_4\zeta_4} \end{bmatrix}
$$

$$
\begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ I_{\theta_1\zeta_1} \\ I_{\theta_1\zeta_2} \\ I_{\theta_1\zeta_3} \\ I_{\theta_1\zeta_4} \\ I_{\theta_2\zeta_1} \\ I_{\theta_2\zeta_2} \\ I_{\theta_2\zeta_3} \\ I_{\theta_2\zeta_4} \\ I_{\theta_3\zeta_1} \\ I_{\theta_3\zeta_2} \\ I_{\theta_3\zeta_3} \\ I_{\theta_3\zeta_4} \\ I_{\theta_4\zeta_1} \\ I_{\theta_4\zeta_2} \\ I_{\theta_4\zeta_3} \\ I_{\theta_4\zeta_4} \end{bmatrix} = \begin{bmatrix}
1/1296 & 1/162 & 1/72 & 1/162 & 1/1296 & 1/162 & 4/81 & 1/9 & 4/81 & 1/162 & 1/72 & 1/9 & 1/4 & 1/9 & 1/72 & 1/162 & 4/81 & 1/9 & 4/81 & 1/162 & 1/1296 & 1/162 & 1/72 & 1/162 & 1/1296 \\
-1/432 & -1/54 & -1/24 & -1/54 & -1/432 & -1/108 & -2/27 & -1/6 & -2/27 & -1/108 & 0 & 0 & 0 & 0 & 0 & 1/108 & 2/27 & 1/6 & 2/27 & 1/108 & 1/432 & 1/54 & 1/24 & 1/54 & 1/432 \\
5/864 & 5/108 & 5/48 & 5/108 & 5/864 & 1/216 & 1/27 & 1/12 & 1/27 & 1/216 & -1/48 & -1/6 & -3/8 & -1/6 & -1/48 & 1/216 & 1/27 & 1/12 & 1/27 & 1/216 & 5/864 & 5/108 & 5/48 & 5/108 & 5/864 \\
-1/72 & -1/9 & -1/4 & -1/9 & -1/72 & 1/36 & 2/9 & 1/2 & 2/9 & 1/36 & 0 & 0 & 0 & 0 & 0 & -1/36 & -2/9 & -1/2 & -2/9 & -1/36 & 1/72 & 1/9 & 1/4 & 1/9 & 1/72 \\
1/36 & 2/9 & 1/2 & 2/9 & 1/36 & -1/9 & -8/9 & -2 & -8/9 & -1/9 & 1/6 & 4/3 & 3 & 4/3 & 1/6 & -1/9 & -8/9 & -2 & -8/9 & -1/9 & 1/36 & 2/9 & 1/2 & 2/9 & 1/36 \\
-1/432 & -1/108 & 0 & 1/108 & 1/432 & -1/54 & -2/27 & 0 & 2/27 & 1/54 & -1/24 & -1/6 & 0 & 1/6 & 1/24 & -1/54 & -2/27 & 0 & 2/27 & 1/54 & -1/432 & -1/108 & 0 & 1/108 & 1/432 \\
5/864 & 1/216 & -1/48 & 1/216 & 5/864 & 5/108 & 1/27 & -1/6 & 1/27 & 5/108 & 5/48 & 1/12 & -3/8 & 1/12 & 5/48 & 5/108 & 1/27 & -1/6 & 1/27 & 5/108 & 5/864 & 1/216 & -1/48 & 1/216 & 5/864 \\
-1/72 & 1/36 & 0 & -1/36 & 1/72 & -1/9 & 2/9 & 0 & -2/9 & 1/9 & -1/4 & 1/2 & 0 & -1/2 & 1/4 & -1/9 & 2/9 & 0 & -2/9 & 1/9 & -1/72 & 1/36 & 0 & -1/36 & 1/72 \\
1/36 & -1/9 & 1/6 & -1/9 & 1/36 & 2/9 & -8/9 & 4/3 & -8/9 & 2/9 & 1/2 & -2 & 3 & -2 & 1/2 & 2/9 & -8/9 & 4/3 & -8/9 & 2/9 & 1/36 & -1/9 & 1/6 & -1/9 & 1/36 \\
1/144 & 1/36 & 0 & -1/36 & -1/144 & 1/36 & 1/9 & 0 & -1/9 & -1/36 & 0 & 0 & 0 & 0 & 0 & -1/36 & -1/9 & 0 & 1/9 & 1/36 & -1/144 & -1/36 & 0 & 1/36 & 1/144 \\
-5/288 & -1/72 & 1/16 & -1/72 & -5/288 & -5/72 & -1/18 & 1/4 & -1/18 & -5/72 & 0 & 0 & 0 & 0 & 0 & 5/72 & 1/18 & -1/4 & 1/18 & 5/72 & 5/288 & 1/72 & -1/16 & 1/72 & 5/288 \\
1/24 & -1/12 & 0 & 1/12 & -1/24 & 1/6 & -1/3 & 0 & 1/3 & -1/6 & 0 & 0 & 0 & 0 & 0 & -1/6 & 1/3 & 0 & -1/3 & 1/6 & -1/24 & 1/12 & 0 & -1/12 & 1/24 \\
-1/12 & 1/3 & -1/2 & 1/3 & -1/12 & -1/3 & 4/3 & -2 & 4/3 & -1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & -4/3 & 2 & -4/3 & 1/3 & 1/12 & -1/3 & 1/2 & -1/3 & 1/12 \\
-5/288 & -5/72 & 0 & 5/72 & 5/288 & -1/72 & -1/18 & 0 & 1/18 & 1/72 & 1/16 & 1/4 & 0 & -1/4 & -1/16 & -1/72 & -1/18 & 0 & 1/18 & 1/72 & -5/288 & -5/72 & 0 & 5/72 & 5/288 \\
25/576 & 5/144 & -5/32 & 5/144 & 25/576 & 5/144 & 1/36 & -1/8 & 1/36 & 5/144 & -5/32 & -1/8 & 9/16 & -1/8 & -5/32 & 5/144 & 1/36 & -1/8 & 1/36 & 5/144 & 25/576 & 5/144 & -5/32 & 5/144 & 25/576 \\
-5/48 & 5/24 & 0 & -5/24 & 5/48 & -1/12 & 1/6 & 0 & -1/6 & 1/12 & 3/8 & -3/4 & 0 & 3/4 & -3/8 & -1/12 & 1/6 & 0 & -1/6 & 1/12 & -5/48 & 5/24 & 0 & -5/24 & 5/48 \\
5/24 & -5/6 & 5/4 & -5/6 & 5/24 & 1/6 & -2/3 & 1 & -2/3 & 1/6 & -3/4 & 3 & -9/2 & 3 & -3/4 & 1/6 & -2/3 & 1 & -2/3 & 1/6 & 5/24 & -5/6 & 5/4 & -5/6 & 5/24 \\
1/24 & 1/6 & 0 & -1/6 & -1/24 & -1/12 & -1/3 & 0 & 1/3 & 1/12 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/3 & 0 & -1/3 & -1/12 & -1/24 & -1/6 & 0 & 1/6 & 1/24 \\
-5/48 & -1/12 & 3/8 & -1/12 & -5/48 & 5/24 & 1/6 & -3/4 & 1/6 & 5/24 & 0 & 0 & 0 & 0 & 0 & -5/24 & -1/6 & 3/4 & -1/6 & -5/24 & 5/48 & 1/12 & -3/8 & 1/12 & 5/48 \\
1/4 & -1/2 & 0 & 1/2 & -1/4 & -1/2 & 1 & 0 & -1 & 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 & -1 & 0 & 1 & -1/2 & -1/4 & 1/2 & 0 & -1/2 & 1/4 \\
-1/2 & 2 & -3 & 2 & -1/2 & 1 & -4 & 6 & -4 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 4 & -6 & 4 & -1 & 1/2 & -2 & 3 & -2 & 1/2 \\
-1/12 & -1/3 & 0 & 1/3 & 1/12 & 1/3 & 4/3 & 0 & -4/3 & -1/3 & -1/2 & -2 & 0 & 2 & 1/2 & 1/3 & 4/3 & 0 & -4/3 & -1/3 & -1/12 & -1/3 & 0 & 1/3 & 1/12 \\
5/24 & 1/6 & -3/4 & 1/6 & 5/24 & -5/6 & -2/3 & 3 & -2/3 & -5/6 & 5/4 & 1 & -9/2 & 1 & 5/4 & -5/6 & -2/3 & 3 & -2/3 & -5/6 & 5/24 & 1/6 & -3/4 & 1/6 & 5/24 \\
-1/2 & 1 & 0 & -1 & 1/2 & 2 & -4 & 0 & 4 & -2 & -3 & 6 & 0 & -6 & 3 & 2 & -4 & 0 & 4 & -2 & -1/2 & 1 & 0 & -1 & 1/2 \\
1 & -4 & 6 & -4 & 1 & -4 & 16 & -24 & 16 & -4 & 6 & -24 & 36 & -24 & 6 & -4 & 16 & -24 & 16 & -4 & 1 & -4 & 6 & -4 & 1
\end{bmatrix} \times \begin{bmatrix} G_{00} \\ G_{01} \\ G_{02} \\ G_{03} \\ G_{04} \\ G_{10} \\ G_{11} \\ G_{12} \\ G_{13} \\ G_{14} \\ G_{20} \\ G_{21} \\ G_{22} \\ G_{23} \\ G_{24} \\ G_{30} \\ G_{31} \\ G_{32} \\ G_{33} \\ G_{34} \\ G_{40} \\ G_{41} \\ G_{42} \\ G_{43} \\ G_{44} \end{bmatrix}
$$

In fact, this can be done when *a* and *b* not equal to zero. But the form will be much tedious.

1.3.2.2. Model for linkage disequilibrium population

When alleles at the two loci are segregating in linkage disequilibrium in the population under study, co-variation rises between genetic effects in Equation (I-1.6). Let α and β be the coefficients of double reduction at loci A and B (locus A is assumed to be closer to the centromere than locus B), and *r* be recombination frequency between the loci. It has been demonstrated in Luo et al (2004) that the three genetic parameters are related each other in form of

$$\beta = \left[\alpha(3-4r)^2 + 2r(3-2r)\right]/9 \qquad \text{(I-1.10)}$$

The genetic effects, which are not independent to each other, are partial regression coefficients in the regression model (I-1.6). So according to Equation (I-1.6), the covariance between genotypic value and genetic-effect design variables (taking x_1 for example) can be calculated as follows

$$\begin{aligned}\operatorname{cov}(G, x_1) &= \operatorname{cov}(\mu, x_1) + \operatorname{cov}(\theta_1 x_1, x_1) + \operatorname{cov}(\theta_2 x_2, x_1) + \ldots + \operatorname{cov}\left(I_{\theta_4\zeta_4} x_4 y_4, x_1\right) \\ &= 0 + \theta_1 \operatorname{var}(x_1) + \theta_2 \operatorname{cov}(x_2, x_1) + \ldots + I_{\theta_4\zeta_4} \operatorname{cov}(x_4 y_4, x_1)\end{aligned} \qquad \text{(I-1.11)}$$

Similarly, all the other genetic-effect design variables can be written in the form as Equation (I-1.11).

It can be shown that these covariances between genotypic value and genetic-effect design variables can be written into matrix notation for estimation of genetic effects, $[P_{AB}]_{24\times1} = [V]^{-1}_{24\times24}[COV]_{24\times1}$, where P_{AB} is an vector of the twenty-four different genetic effects. V is an asymptotic variance-covariance matrix with variances appearing along the diagonal and covariance appearing in the off-diagonal elements. COV is a matrix of covariances between genetic value and the twenty-four genetic-effect design variable. I worked out these matrices in an explicit form as follows.

Firstly we calculated orthogonal contrast scales, A_{ij} and B_{ij} (i=0,1,…,4; j=1,2,…,4), for loci A and B separately via the general bi-allelic one locus model as shown in the table below.

	Genotype	*aaaa*	*Aaaa*	*AAaa*	*AAAa*	*AAAA*
Locus A	Frequency	$f_{0.}$	$f_{1.}$	$f_{2.}$	$f_{3.}$	$f_{4.}$
	G	$G_{0.}$	$G_{1.}$	$G_{2.}$	$G_{3.}$	$G_{4.}$
θ_1	W_{A1}	A_{41}	A_{31}	A_{21}	A_{11}	A_{01}
θ_2	W_{A2}	A_{42}	A_{32}	A_{22}	A_{12}	A_{02}
θ_3	W_{A3}	A_{43}	A_{33}	A_{23}	A_{13}	A_{03}
θ_4	W_{A4}	A_{44}	A_{34}	A_{24}	A_{14}	A_{04}

	Genotype	*bbbb*	*Bbbb*	*BBbb*	*BBBb*	*BBBB*
Locus B	Frequency	$f_{.0}$	$f_{.1}$	$f_{.2}$	$f_{.3}$	$f_{.4}$
	G	$G_{.0}$	$G_{.1}$	$G_{.2}$	$G_{.3}$	$G_{.4}$
ζ_1	W_{B1}	B_{41}	B_{31}	B_{21}	B_{11}	B_{01}
ζ_2	W_{B2}	B_{42}	B_{32}	B_{22}	B_{12}	B_{02}

ζ_3	W_{B3}	B_{43}	B_{33}	B_{23}	B_{13}	B_{03}
ζ_4	W_{B4}	B_{44}	B_{34}	B_{24}	B_{14}	B_{04}

Then the genetic-effect design variables x_i and y_i (i = 1,…, 4) are calculated as

$$x_i = \begin{cases} A_{4i} & if\ A\ is\ AAAA \\ A_{3i} & if\ A\ is\ AAAa \\ A_{2i} & if\ A\ is\ AAaa \\ A_{1i} & if\ A\ is\ Aaaa \\ A_{0i} & if\ A\ is\ aaaa \end{cases} \quad \text{and} \quad y_i = \begin{cases} B_{4i} & if\ B\ is\ BBBB \\ B_{3i} & if\ B\ is\ BBBb \\ B_{2i} & if\ B\ is\ BBbb \\ B_{1i} & if\ B\ is\ Bbbb \\ B_{0i} & if\ B\ is\ bbbb \end{cases} \quad i = 1, 2, 3, 4$$

The genetic effects can be estimated as follows

$$\left[P_{AB}\right]_{24\times1} = \left[V\right]^{-1}_{24\times24} \times \left[COV\right]_{24\times1}$$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ I_{\theta_1\zeta_1} \\ I_{\theta_1\zeta_2} \\ I_{\theta_1\zeta_3} \\ I_{\theta_1\zeta_4} \\ I_{\theta_2\zeta_1} \\ I_{\theta_2\zeta_2} \\ I_{\theta_2\zeta_3} \\ I_{\theta_2\zeta_4} \\ I_{\theta_3\zeta_1} \\ I_{\theta_3\zeta_2} \\ I_{\theta_3\zeta_3} \\ I_{\theta_3\zeta_4} \\ I_{\theta_4\zeta_1} \\ I_{\theta_4\zeta_2} \\ I_{\theta_4\zeta_3} \\ I_{\theta_4\zeta_4} \end{bmatrix} = \begin{bmatrix} \mathrm{var}(x_1) & \mathrm{cov}(x_1,x_2) & \dots & \mathrm{cov}(x_1,x_4) & \mathrm{cov}(x_1,y_1) & \dots & \dots & \mathrm{cov}(x_1,y_4) & \mathrm{cov}(x_1,x_1y_1) & \dots & \dots & \mathrm{cov}(x_1,x_1y_4) & \mathrm{cov}(x_1,x_2y_1) & \dots & \mathrm{cov}(x_1,x_3y_1) & \dots & \mathrm{cov}(x_1,x_4y_1) & \dots \\ & \dots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & \mathrm{cov}(x_3,x_4) & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & \mathrm{var}(x_4) & \mathrm{cov}(x_4,y_1) & \dots & \dots & \mathrm{cov}(x_4,y_4) & \mathrm{cov}(x_4,x_1y_1) & \dots & \dots & \mathrm{cov}(x_4,x_1y_4) & \mathrm{cov}(x_4,x_2y_1) & \dots & \mathrm{cov}(x_4,x_3y_1) & \dots & \mathrm{cov}(x_4,x_4y_1) & \dots \\ & & & & \mathrm{var}(y_1) & \mathrm{cov}(y_1,y_1) & \dots & \mathrm{cov}(y_1,y_4) & \mathrm{cov}(y_1,x_1y_1) & \dots & \dots & \mathrm{cov}(y_1,x_1y_4) & \mathrm{cov}(y_1,x_2y_1) & \dots & \mathrm{cov}(y_1,x_3y_1) & \dots & \mathrm{cov}(y_1,x_4y_1) & \dots \\ & & & & & \dots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & \dots & \mathrm{cov}(y_3,y_4) & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & & \mathrm{var}(y_4) & \mathrm{cov}(y_4,x_1y_1) & \dots & \dots & \mathrm{cov}(y_4,x_1y_4) & \mathrm{cov}(y_4,x_2y_1) & \dots & \mathrm{cov}(y_4,x_3y_1) & \dots & \mathrm{cov}(y_4,x_4y_1) & \dots \\ & & & & & & & & \mathrm{var}(x_1y_1) & \mathrm{cov}(x_1y_1,x_1y_2) & \dots & \mathrm{cov}(x_1y_1,x_1y_4) & \mathrm{cov}(x_1y_1,x_2y_1) & \dots & \mathrm{cov}(x_1y_1,x_3y_1) & \dots & \mathrm{cov}(x_1y_1,x_4y_1) & \dots \\ & & & & & & & & & \mathrm{var}(x_1y_2) & \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & & & & & \dots & \mathrm{cov}(x_1y_3,x_1y_4) & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & & & & & & \mathrm{var}(x_1y_4) & \mathrm{cov}(x_1y_4,x_2y_1) & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & & & & & & & \mathrm{var}(x_2y_1) & \dots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & & & & & & & & \dots & \dots & \vdots & \vdots & \vdots \\ & & & & & & & & & & & & & & \dots & \dots & \vdots & \vdots \\ & & & & & & & & & & & & & & & \dots & \dots & \vdots \\ & & & & & & & & & & & & & & & \dots & \mathrm{cov}(x_2y_4,x_3y_1) & \vdots \\ & & & & & & & & & & & & & & & \mathrm{var}(x_3y_1) & \dots & \vdots \\ & & & & & & & & & & & & & & & & \dots & \dots \\ & & & & & & & & & & & & & & & & \dots & \vdots \\ & & & & & & & & & & & & & & & & \dots & \dots \\ & & & & & & & & & & & & & & & & \dots & \mathrm{cov}(x_3y_4,x_4y_1) \\ & & & & & & & & & & & & & & & & \mathrm{var}(x_4y_1) & \dots \\ & & & & & & & & & & & & & & & & & \dots \\ & & & & & & & & & & & & & & & & & \dots \\ & & & & & & & & & & & & & & & & & \dots \end{bmatrix}^{-1} \times \begin{bmatrix} \mathrm{cov}(G,x_1) \\ \mathrm{cov}(G,x_2) \\ \mathrm{cov}(G,x_3) \\ \mathrm{cov}(G,x_4) \\ \mathrm{cov}(G,y_1) \\ \mathrm{cov}(G,y_2) \\ \mathrm{cov}(G,y_3) \\ \mathrm{cov}(G,y_4) \\ \mathrm{cov}(G,x_1y_1) \\ \mathrm{cov}(G,x_1y_2) \\ \mathrm{cov}(G,x_1y_3) \\ \mathrm{cov}(G,x_1y_4) \\ \mathrm{cov}(G,x_2y_1) \\ \mathrm{cov}(G,x_2y_2) \\ \mathrm{cov}(G,x_2y_3) \\ \mathrm{cov}(G,x_2y_4) \\ \mathrm{cov}(G,x_3y_1) \\ \mathrm{cov}(G,x_3y_2) \\ \mathrm{cov}(G,x_3y_3) \\ \mathrm{cov}(G,x_3y_4) \\ \mathrm{cov}(G,x_4y_1) \\ \mathrm{cov}(G,x_4y_2) \\ \mathrm{cov}(G,x_4y_3) \\ \mathrm{cov}(G,x_4y_4) \end{bmatrix}$$

V is a 24×24 asymptotic variance- covariance matrix of the variance of the genetic-effect design variables which appear along the diagonal and the covariance appear in the off-diagonal elements. The inverse of this matrix, V^{-1}, is known as the concentration matrix or precision matrix. Details of the calculation of these variances and covariance are shown as follows

$$\mathrm{var}\left(x_i\right) = E\left(x_i^2\right) - \left[E\left(x_i\right)\right]^2 = \sum_{j=0}^{4} A_{ji}^2 f_{j.} \qquad i \in \{1,2,3,4\}$$

$$\mathrm{var}\left(y_i\right) = E\left(y_i^2\right) - \left[E\left(y_i\right)\right]^2 = \sum_{j=0}^{4} B_{ji}^2 f_{.j} \qquad i \in \{1,2,3,4\}$$

$$\mathrm{var}\left(x_i y_j\right) = E\left(x_i^2 y_j^2\right) - \left[E\left(x_i y_j\right)\right]^2 = \sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi}^2 B_{nj}^2 f_{mn} - \left[\sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} B_{nj} f_{mn}\right]^2 \qquad i, j \in \{1,2,3,4\}$$

$$\mathrm{cov}\left(x_i, x_j\right) = \mathrm{cov}\left(y_i, y_j\right) = 0 \qquad i < j \in \{1,2,3,4\}$$

$$\mathrm{cov}\left(x_i, y_j\right) = E\left(x_i y_j\right) - E\left(x_i\right)E\left(y_j\right) = \sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} B_{nj} f_{mn} \qquad i, j \in \{1,2,3,4\}$$

$$\mathrm{cov}\left(x_i, x_i y_j\right) = E\left(x_i^2 y_j\right) - E\left(x_i\right)E\left(x_i y_j\right) = \sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi}^2 B_{nj} f_{mn} \qquad i, j \in \{1,2,3,4\}$$

$$\mathrm{cov}\left(y_i, x_j y_i\right) = E\left(x_j y_i^2\right) - E\left(y_i\right)E\left(x_j y_i\right) = \sum_{m=0}^{4}\sum_{n=0}^{4} A_{mj} B_{ni}^2 f_{mn} \qquad i, j \in \{1,2,3,4\}$$

$$\mathrm{cov}\left(x_i, x_j y_k\right) = E\left(x_i x_j y_k\right) - E\left(x_i\right)E\left(x_j y_k\right) = \sum_{m=0}^{4}\sum_{n=0}^{4}\sum_{s=0}^{4} A_{si} A_{mj} B_{nk} f_{mn} f_{s.} \qquad i, j, k \in \{1,2,3,4\} \text{ and } i \neq j$$

$$\mathrm{cov}\left(y_i, x_j y_k\right) = E\left(y_i x_j y_k\right) - E\left(y_i\right)E\left(x_j y_k\right) = \sum_{m=0}^{4}\sum_{n=0}^{4}\sum_{s=0}^{4} B_{si} A_{mj} B_{nk} f_{mn} f_{.s} \qquad i, j, k \in \{1,2,3,4\} \text{ and } i \neq k$$

$$\mathrm{cov}\left(x_i y_j, x_k y_l\right) = E\left(x_i y_j x_k y_l\right) - E\left(x_i y_j\right)E\left(x_k y_l\right)$$
$$= \sum_{m=0}^{4}\sum_{n=0}^{4}\sum_{s=0}^{4}\sum_{t=0}^{4} A_{mi} B_{nj} A_{sk} B_{tl} f_{mn} f_{st} - \left(\sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} B_{nj} f_{mn}\right)\left(\sum_{m=0}^{4}\sum_{n=0}^{4} A_{mk} B_{nl} f_{mn}\right) \qquad i, j, k, l \in \{1,2,3,4\} \text{ and } ij \neq kl$$

$$\operatorname{cov}(G, x_i) = E(Gx_i) - E(G)E(x_i) = \sum_{m=0}^{4} E(x_i = A_{mi})E(G \mid x_i = A_{mi})$$

$$= \sum_{m=0}^{4} f_{m.} A_{mi} \frac{\sum_{n=0}^{4} f_{mn} G_{mn}}{f_{m.}} = \sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} f_{mn} G_{mn} \qquad i \in \{1,2,3,4\}$$

$$\operatorname{cov}(G, y_i) = E(Gy_i) - E(G)E(y_i) = \sum_{n=0}^{4} E(y_i = B_{ni})E(G \mid y_i = B_{ni})$$

$$= \sum_{n=0}^{4} f_{.n} B_{ni} \frac{\sum_{m=0}^{4} f_{mn} G_{mn}}{f_{.n}} = \sum_{m=0}^{4}\sum_{n=0}^{4} B_{ni} f_{mn} G_{mn} \qquad i \in \{1,2,3,4\}$$

$$\operatorname{cov}(G, x_i y_j) = E(Gx_i y_j) - E(G)E(x_i y_j)$$

$$= \sum_{m=0}^{4}\sum_{n=0}^{4} E(x_i = A_{mi}, y_j = B_{nj})E(G \mid x_i = A_{mi}, y_j = B_{nj}) - \left(\sum_{m=0}^{4}\sum_{n=0}^{4} f_{mn} G_{mn}\right)\left(\sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} B_{nj} f_{mn}\right)$$

$$= \sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} B_{nj} f_{mn} G_{mn} - \left(\sum_{m=0}^{4}\sum_{n=0}^{4} f_{mn} G_{mn}\right)\left(\sum_{m=0}^{4}\sum_{n=0}^{4} A_{mi} B_{nj} f_{mn}\right) \qquad i, j \in \{1,2,3,4\}$$

1.3.2.3. Estimation of genetic parameters in a reduced model

Because of limited sample size, it is probably in practice that not all the genotypes would appear in sufficient counts in the sample, especially when a two locus model is considered. Statistically, this will lead to singularity of the linear model and thus not all the genetic parameters are estimable. To tackle this problem, I proposed to estimate the parameters in a reduced model by being focused on main effects in the models. Here main effects refer to genic effects at each of the two loci, excluding the epistasis effects.

No matter the population is in linkage equilibrium or not, the genotypic value can be converted into the following matrix notation,

$$G=\begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{25} \end{bmatrix}_{25\times 1} = X \bullet P = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,25} \\ x_{21} & x_{22} & \cdots & x_{2,25} \\ \vdots & \vdots & \vdots & \vdots \\ x_{25,1} & x_{25,2} & \cdots & x_{25,25} \end{bmatrix}_{25\times 25} \bullet \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{25} \end{bmatrix}_{25\times 1} \tag{I-1.12}$$

Here $G=[G_{00},G_{01},\cdots,G_{43},G_{44}]^T$, $P=\left[\mu,\theta_1,\theta_2,\cdots I_{\theta_4\zeta_3},I_{\theta_4\zeta_4}\right]^T$

However, according to the theoretical genotypic frequency distribution, some genotype may unlikely exist in offspring population with limited population size. So we should estimate genetic parameters in a reduced model. Assuming there would be only *r* genotypes existing in offspring, the model can be reduced into

$$G^{'}=\begin{bmatrix} G_1^{'} \\ G_2^{'} \\ \vdots \\ G_r^{'} \end{bmatrix}_{r\times 1} = X^{'} \bullet P = \begin{bmatrix} x_{11}^{'} & x_{12}^{'} & \cdots & x_{1,25}^{'} \\ x_{21}^{'} & x_{22}^{'} & \cdots & x_{2,25}^{'} \\ \vdots & \vdots & \vdots & \vdots \\ x_{r,1}^{'} & x_{r,1}^{'} & \cdots & x_{r,25}^{'} \end{bmatrix}_{r\times 25} \bullet \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{25} \end{bmatrix}_{25\times 1} \qquad r<25 \tag{I-1.13}$$

This linear model is not full rank and not all the parameters p_i are estimatable. So we reduce the model into one of full rank. Firstly, $X^{'}$ can be divided into two parts, $X_1^{'}$ and $X_2^{'}$. Let $X_1^{'}$ consist of *r* linearly independent columns from $X^{'}$ and let $X_2^{'}$ consist of the remaining columns.

$$X^{'}=\left[X_1^{'}\middle|X_2^{'}\right]=\left[\begin{array}{cccc|cccc} x_{11}^{'} & x_{12}^{'} & \cdots & x_{1r}^{'} & x_{1,r+1}^{'} & x_{1,r+2}^{'} & \cdots & x_{1,25}^{'} \\ x_{21}^{'} & x_{22}^{'} & \cdots & x_{2r}^{'} & x_{2,r+1}^{'} & x_{2,r+2}^{'} & \cdots & x_{2,25}^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{r1}^{'} & x_{r2}^{'} & \cdots & x_{rr}^{'} & x_{r,r+1}^{'} & x_{r,r+2}^{'} & \cdots & x_{25,25}^{'} \end{array}\right]_{r\times 25} \quad \text{(I-1.14)}$$

Let $X_2^{'}=X_1^{'}A$, $X_2^{'}$ being linearly dependent on $X_1^{'}$, then Equation (I-1.13) can be written as

$$\begin{aligned} G^{'} &= X^{'}\bullet P=\left[X_1^{'}\right]_{r\times r}\bullet\left(I_{r\times r},A\right)\bullet P \\ &=\left[X_1^{'}\right]_{r\times r}\bullet\left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & a_{1,r+1} & a_{1,r+2} & \cdots & a_{1,25} \\ 0 & 1 & \cdots & 0 & a_{2,r+1} & a_{2,r+2} & \cdots & a_{2,25} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & a_{r,r+1} & a_{r,r+2} & \cdots & a_{r,25} \end{array}\right]_{r\times 25}\bullet\left[P\right]_{25\times 1} \end{aligned} \quad \text{(I-1.15)}$$

with $A_{r\times(25-r)}=X_1^{'-1}\bullet X_2^{'}$.

By calculating $\left(I_{r\times r},A\right)\bullet P$ first, Equation (I-1.15) can be rewritten as

$$\left[G^{'}\right]_{r\times 1}=\left[X_1^{'}\right]_{r\times r}\bullet\left(I_{r\times r},A\right)\bullet P=\left[X_1^{'}\right]_{r\times r}\bullet\left[\begin{array}{c} p_1+\sum_{i=r+1}^{25}a_{1i}p_i \\ p_2+\sum_{i=r+1}^{25}a_{2i}p_i \\ \vdots \\ p_r+\sum_{i=r+1}^{25}a_{ri}p_i \end{array}\right]_{r\times 1} \quad \text{(I-1.16)}$$

And then

$$\begin{bmatrix} p_1 + \sum_{i=r+1}^{25} a_{1i} p_i \\ p_2 + \sum_{i=r+1}^{25} a_{2i} p_i \\ \vdots \\ p_r + \sum_{i=r+1}^{25} a_{ri} p_i \end{bmatrix}_{r\times 1} = \left[X_1^{'}\right]_{r\times r}^{-1} \bullet \left[G^{'}\right]_{r\times 1} \qquad \text{(I-1.17)}$$

So the genetic parameters can be estimated jointly in the form of Equation (I-1.17).

1.4. Simulation study and analysis

I carried out an intensive simulation study to test reliability of the theoretical models presented above to model phenotype of quantitative traits in autotetraploid populations and to explore statistical properties of the statistical methods developed here for estimating the model parameters. The simulation program mimics gametogensis of an autotetraploid genotype and generation of a zygote. Segregation and recombination of alleles at the loci of interest were simulated under a strict tetrasomic inheritance model. Although the simulation program was flexible to simulate any numbers of linked or unlinked loci for any given values of the coefficient of double reduction and recombination frequency, we considered here F_2 populations from crossing two parents which were divergent at a single or two loci for a demonstration purpose. As long as an offspring genotype at the simulated locus or loci was generated, phenotype of the offspring was determined as sum of genotypic value calculated from the

correspondingly simulated genetic model (I-1.2) or (I-1.6), depending on the number of loci considered, and a random variable sampled from a normal distribution $N(0,\sigma^2)$. The residual variance was defined by a prior given phenotypic variance of the trait in question and heritability of the QTL. Thus, phenotype of the offspring population can be modelled as a mixed normal distribution with m = 5 or 25 component distributions, each corresponding to a genotype at the QTL, as given by

$$F\left(x;m,\Omega_m\right)=\sum_{i=1}^{m} f_i g_i\left(x;G_i,\sigma^2\right) \qquad \text{(I-1.18)}$$

Where the form of model parameter vector, $\Omega_m=(f_1\cdots f_m G_1\ldots G_m\sigma^2)$ and genotypic frequency, f_i (i=1,…,m) which depends on double reduction and/or recombination parameters. $g_i\left(x;\mu_i,\sigma^2\right)$ stands for the probability density function of normal distribution with mean μ_i and variance σ^2.

To calculate estimates of the genetic effect parameters, we first calculated the mean for each QTL genotype from the offspring population. This is equivalent to estimating means of a finite mixture of component distributions. We considered here the scenario that QTL genotypes of the offspring individuals were unknown but the coefficient of double reduction at the QTL was known or can be estimated from other source of information, for instance the data of genotypes of genetic markers at the nearby QTL region using the methods we developed before (Luo 2000). Thus, the parameter estimation can be formulated as analysis of a finite mixture of normal distributions with known proportions, $f_1,\cdots, f_m$, through implementing the EM (Expectation

and Maximization) algorithm (Dempster 1977). The EM algorithm involves iterating the E-step that calculates the conditional probability of the i^{th} individual having the j^{th} QTL genotype, i.e.

$$\omega_{ij} = \frac{f_j g_j\left(x_i; G_j, \sigma_j^2\right)}{\sum_{k=1}^{m} f_k g_k\left(x_i; G_k, \sigma_k^2\right)} \tag{I-1.19}$$

and the M-step that calculates the maximum likelihood estimates (MLEs) of the model parameters given the conditional probabilities from the above E step from the following formula

$$\hat{G}_j = \sum_{i=1}^{n} \omega_{ij} x_i \Big/ \sum_{i=1}^{n} \omega_{ij} \tag{I-1.20}$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} \omega_{ij} (x_i - \hat{G}_j)^2 \Big/ n \tag{I-1.21}$$

1.4.1. One-locus analysis

This simulation model considered a bi-allelic quantitative trait locus, Q_A, with two alleles, segregating in an F_2 population from crossing a pair of parental autotetraploids with genotypes *AAAA* and *aaaa*. In simulating allele segregation at the QTL, I set the coefficient of double reduction to be 0.0 or 0.15. Segregation at the simulated QTL contributed 10% of phenotypic variance of the trait in the population. All genetic effects at the QTL were set to be 1.0 and residual variance was determined accordingly. To better estimate genetic parameters from such mixture normal distributions, I simulated some genetic markers along the QTL which is

assumed to be closer to centromere to make f_i more informative. The detailed set of simulated values was listed in Table I-1.4.

Table I-1.4. Simulation parameters of the coefficient of double reduction at QTL and recombination frequencies between QTL and 4 linked marker loci and the corresponding parental genotypes and genetic effects used to simulate the populations

Locus	r	Parental genotype		Simulated parameters			Theoretical value	
		P_1	P_2				α=0.00	α=0.15
						σ	2.896	3.252
QTL	0.00	*AAaa*	*AAaa*	μ	1.000	G_4	5.458	5.215
L_1	0.05	$M_1M_2M_3M_3$	$M_1M_5M_6M_7$	θ_1	1.000	G_3	1.938	1.787
L_2	0.10	$M_1M_2M_2M_4$	$M_5M_2M_2M_6$	θ_2	1.000	G_2	0.708	0.622
L_3	0.15	$M_1M_1M_3M_4$	$M_3M_5M_6M_7$	θ_3	1.000	G_1	0.271	0.221
L_4	0.20	$M_4M_2M_3M_4$	$M_5M_6M_6M_8$	θ_4	1.000	G_0	0.125	0.082

Markers were located on the same side of the QTL which is closer to centromere. Offspring population were generated under autotetrasomic inheritance with double reduction rate α equal to 0.00 or 0.15. Offspring population size was 500 and heritability was assumed to be 0.1. Alleles listed in the same column had the same linkage phase.

I generated offspring individuals under autotetrasomic inheritance (the coefficient of double reduction was denoted as a = 0.0 or 0.15) for each case, assuming its meiosis was all quadrivalent pairing of homologous chromosomes. The orthogonal contrast scales can be obtained according to the theoretical genotypic frequencies distribution for given parental genotypes and the coefficient of double reduction by using the one-locus F_2 population model as developed in Section 1.3.1.1. To explore the effects caused by double reduction on the estimation of the genetic parameters, the estimation procedure was carried out in two different ways: taking double reduction into account when calculating the orthogonal contrasts of linear model, and ignoring double reduction when doing the parameters estimation. The simulation results are shown in Table I-1.5.

Table I-1.5 tabulates the means and standard errors of the estimated genetic parameters compared with the true genetic parameters based on 100 repeated simulations. This simulation was designed to investigate the effects of double reduction, which is the most important feature of autotetrasomic inheritance, on the reliability of the model to estimate genetic parameters. It can be seen from Table I-1.5 that the genetic parameters were predicted adequately with low heritability when taking double reduction into account, while estimation of the genetic effects without consideration of double reduction performed poorer comparatively. From the estimates of heritability in Table I-1.5, it can be found that overestimate of the coefficient of double reduction would result in overestimating in genetic variance and vice versa. We can conclude that taking account of double reduction and accurate estimation of the coefficient of double reduction is important in estimation of genetic parameters in autotetraploids.

Table I-1.5. Means and standard errors of the parameter estimates based on 100 repeated simulations of a single QTL model.

Offspring data generated with double reduction rate α=0.00

True values				Estimates [a]		Estimates [b]		Estimates [c]		Estimates [d]	
μ	1.000			0.997 (0.013)		1.014 (0.013)		1.037 (0.013)		1.061 (0.013)	
θ_1	1.000	V_1	0.667	1.024 (0.017)	0.699	1.003 (0.017)	0.738	0.997 (0.017)	0.795	0.997 (0.017)	0.861
θ_2	1.000	V_2	0.222	1.057 (0.033)	0.248	1.033 (0.033)	0.267	1.023 (0.033)	0.293	1.018 (0.033)	0.322
θ_3	1.000	V_3	0.037	0.809 (0.087)	0.024	0.789 (0.085)	0.027	0.786 (0.085)	0.030	0.786 (0.085)	0.033
θ_4	1.000	V_4	0.003	0.663 (0.251)	0.002	0.570 (0.265)	0.001	0.553 (0.272)	0.001	0.545 (0.279)	0.001
σ	2.892			2.875 (0.009)		2.875 (0.009)		2.875 (0.009)		2.875 (0.009)	
h^2	0.100			0.105		0.111		0.119		0.128	

Offspring data generated with double reduction rate α=0.15

Simulated Parameters				Estimates [d]		Estimates [a]		Estimates [c]		Estimates [e]	
μ	1.000			1.008 (0.017)		0.977 (0.017)		0.986 (0.016)		1.031 (0.017)	
θ_1	1.000	V_1	0.867	0.992 (0.018)	0.853	1.063 (0.023)	0.753	0.996 (0.019)	0.794	0.994 (0.018)	0.922
θ_2	1.000	V_2	0.311	1.031 (0.027)	0.330	1.114 (0.032)	0.276	1.041 (0.028)	0.303	1.027 (0.026)	0.361
θ_3	1.000	V_3	0.053	1.077 (0.081)	0.062	1.074 (0.097)	0.043	1.077 (0.083)	0.056	1.078 (0.080)	0.067
θ_4	1.000	V_4	0.004	1.329 (0.364)	0.007	1.540 (0.404)	0.008	1.303 (0.371)	0.006	1.355 (0.361)	0.007
σ	3.333			3.304 (0.010)		3.308 (0.010)		3.303 (0.010)		3.304 (0.010)	
h^2	0.100			0.103		0.090		0.096		0.111	

Here μ is the population mean and and θ_i (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL. σ is a random variable and h^2 is the heritability. V_1, V_2, V_3 and V_4 represent monogenic, digenic,trigenic and quadrigenic genetic variance components, respectively. The estimation procedure was carried out in ways as follows:

a – estimates obtained when α=0.00; b – estimates obtained when α=0.05; c – estimates obtained when α=0.10;

d – estimates obtained when α=0.15; e – estimates obtained when α=0.20

1.4.2. Two-locus analysis

We simulated a F_2 population of 300 individuals generated from crossing two autotetraploid parental genotypes, *AAAA/BBBB* and *aaaa/bbbb*, respectively. A simulated quantitative trait was controlled by two linked QTLs with a recombination frequency r = 0.2, and the coefficient of double reduction was a = 0.10 at the QTL A, implying that the coefficient of double reduction at the QTL B was $\beta = \left[\alpha(3-4r)^2 + 2r(3-2r)\right]/9$ = 0.1693 (Luo 2004). All the genetic parameters as listed in Table I-1.2 (on page 31) are set to be 1.0. The simulation mimicked gametogenesis of an autotetraploid individual with all quadrivalent pairing of homologous chromosomes during meiosis, which meant that recombination can occur between two non-sister chromatids. QTL A and B were obviously in linkage disequilibrium in the F_2 population, thus the genetic effects can be calculated as developed in Section 1.3.2.2. Phenotype of an offspring individual was generated as sum of the corresponding genotypic value and a randomly generated number from a normal distribution with mean zero and variance, which was adjusted for a simulated heritability of 0.1 for the quantitative trait.

Theoretically, there are twenty-five different genotypes (without consideration of linkage phase) in the current simulated offspring populations. However, due to the limited population, some genotypes at the QTLs may not appear with sufficient counts in the segregation population. In this case, not all the genetic parameters are estimable. I first calculated the theoretical genotypic frequencies of the twenty-five different genotypes by a computer-based algorithm (Luo et al, 2004). Based on the theoretical genotype distribution, it can be anticipated that some of these genotypes may be present in a small number (e.g. less than 5) among the total 300 offspring

individuals. Of all the possible genotypes, thirteen were expected to be present in the number greater than 5, namely G_{00}, G_{01}, G_{10}, G_{11}, G_{12}, G_{21}, G_{22}, G_{23}, G_{32}, G_{33}, G_{34}, G_{43} and G_{44}. By using method in Section 1.3.2.3, the twenty-five genetic parameters can be reduced into thirteen as shown in part (i) of Table I-1.6. The newly defined parameters $p_i\,(i=0,1,\ldots,12)$ can be expressed linearly in terms of original genetic parameters, with p_0 highlighted the effect of population mean, $p_1 \ldots p_4$ highlighted monogentic, digenic, trigenic and quadrigenic genetic effects at locus A, $p_5 \ldots p_8$ highlighted monogentic, digenic, trigenic and quadrigenic genetic effects at locus B and $p_9 \ldots p_{12}$ highlighted epistasis between the effects θ_1 and $\varsigma_i\left(i=1,\ldots,4\right)$. Especially, when the residual epistatic effects were small enough, the parameters p_i can well reflect the main genetic effects which we interest in this way. Table I-1.6 also summarized the means and stand error of the estimated parameter, $p_i\,(i=0,1,\ldots,12)$ compared with the true value based on 100 repeated simulations. It can be seen from Table I-1.6 that the parameters were adequately estimated with small population size of 300 and low heritability of 0.1. In the current study, most of the main genetic effects could be detected theoretically. However, some of them (e.g. monogenic effect at Locus A reflected by p_1, epistatic effects reflected by p_{10} and p_{11}) could not be well detected due to large residual epistatic effects and corresponding weighted value.

Table I-1.6. Estimation of genetic parameters in linkage disequilibrium population by the bi-allelic two loci F_2 population model

(i). the genetic parameters in a reduced model

$$\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \end{bmatrix} = \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ I_{\theta_1\zeta_1} \\ I_{\theta_1\zeta_2} \\ I_{\theta_1\zeta_3} \\ I_{\theta_1\zeta_4} \end{bmatrix} + \begin{bmatrix} 0 & 0.228 & 0 & -0.004 & -0.189 & 0 & 0.115 & 0 & 0 & -0.012 & 0 & 0.005 \\ -2.208 & 0 & 0.813 & 0 & 0 & 0.265 & 0 & 0.021 & 0.234 & 0 & -0.035 & 0 \\ 0 & -5.591 & 0 & 0.620 & 7.481 & 0 & -0.982 & 0 & 0 & 1.115 & 0 & -0.151 \\ -0.485 & 0 & -0.395 & 0 & 0 & 0.433 & 0 & -0.108 & 0.840 & 0 & -0.237 & 0 \\ 0 & -2.132 & 0 & 0.235 & 9.090 & 0 & -1.642 & 0 & 0 & 0.688 & 0 & -0.120 \\ 2.136 & 0 & -0.236 & 0 & 0 & 0.235 & 0 & 0.022 & -0.312 & 0 & 0.098 & 0 \\ 0 & 6.493 & 0 & -0.587 & -6.783 & 0 & 1.117 & 0 & 0 & -1.033 & 0 & 0.151 \\ 0.824 & 0 & 3.701 & 0 & 0 & 2.748 & 0 & 0.376 & -1.318 & 0 & 0.630 & 0 \\ 0 & 8.027 & 0 & -0.812 & -3.894 & 0 & 0.451 & 0 & 0 & -1.032 & 0 & 0.142 \\ 0 & -0.141 & 0 & -0.014 & 0.063 & 0 & -0.247 & 0 & 0 & -0.020 & 0 & -0.007 \\ 0.938 & 0 & -0.968 & 0 & 0 & -0.880 & 0 & -0.085 & 0.169 & 0 & -0.127 & 0 \\ 0 & -0.262 & 0 & 0.070 & -0.670 & 0 & -0.076 & 0 & 0 & -0.023 & 0 & -0.007 \\ -1.675 & 0 & -6.650 & 0 & 0 & -6.045 & 0 & -1.022 & 2.775 & 0 & -1.482 & 0 \end{bmatrix} \bullet \begin{bmatrix} I_{\theta_2\zeta_1} \\ I_{\theta_2\zeta_2} \\ I_{\theta_2\zeta_3} \\ I_{\theta_2\zeta_4} \\ I_{\theta_3\zeta_1} \\ I_{\theta_3\zeta_2} \\ I_{\theta_3\zeta_3} \\ I_{\theta_3\zeta_4} \\ I_{\theta_4\zeta_1} \\ I_{\theta_4\zeta_2} \\ I_{\theta_4\zeta_3} \\ I_{\theta_4\zeta_4} \end{bmatrix}$$

(ii). estimation result by the bi-allelic two loci model

	p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}
True	1.142	0.090	3.492	1.049	7.119	2.943	0.357	7.960	3.882	0.634	0.047	0.030	-13.100
Estimates	1.074	0.092	3.388	1.015	6.964	2.914	0.418	7.931	3.872	0.618	0.030	0.029	-13.104
s.e.	0.321	0.003	0.069	0.025	0.095	0.023	0.062	0.083	0.047	0.011	0.016	0.006	0.174

All the original genetic parameters were assumed to be 1.0, including the mean. The F_2 population of size of 300 were simulated under tetrasomic inheritance with the coefficient of double reduction locus A being set to 0.1 and recombination frequency between locus A and locus B being 0.2. The heritability here was 0.1 and simulated replicates were 100.

1.5. Discussion

Considerable advancement has been made recently in the genetic linkage analysis with autotetrasomic inheritance, providing theory and tools for genetic map construction (Luo et al. 2000, 2004, 2006; Leach et al. 2010). These methods also provided ways to calculate the conditional probability distribution of genotypes at any location along the chromosome given the parental and offspring marker phenotypes at its linked genetic markers, achieving a step forward for QTL analysis. However, as another key part of QTL analysis, progress in the study of quantitative genetics of autotetraploids is emergent in the era of genomic genetics.

In this chapter, I have succeeded in extending Cockerham's orthogonal contrast based quantitative model for diploid species to autotetraploid species and establishing a quantitative genetics model for analysing QTL effects and epistasis by using orthogonal contrast scales with two alleles in autotetraploid species. Although less comprehensive in their representation of natural populations, two allele models do have the advantage that they can be expressed in terms of relatively few parameters which deal with genetic phenomena at the level of gene rather than the populations, thus allowing more fundamental analysis. This quantitative genetic model has taken the existence of epistasis into account and decomposed the genotypic value at the QTLs into monogenic, digenic, trigenic, quadrigenic and epistatic effects. Under the assumption that there is only two-locus epistasis for pairs of loci, the property of orthogonality ensures that the monogenic, digenic, trigenic, quadrigenic and epistatic effects can be estimated independently for any number of loci. In practice, a quantitative trait is usually controlled by more than one QTL and the number of QTLs seems to be always incorrectly identified in QTL mapping. Thus

it is very important to keep consistency in QTL effects estimation in a multi-locus setting, which is essential for the QTL analysis to be multi-locus comparable. This orthogonal contrast scales based quantitative genetics model ensures that genetic effects and genetic variance components are consistently estimated no matter how many QTLs get involved and can benefit the study of QTL mapping in autotetraploids.

This quantitative genetic model is generally used for populations with different structures, like randomly-mating natural population or artificial designed F_2 population illustrated here. Parameters estimation can be carried out under different levels of recombination frequency or coefficient of double reduction. This two-allele model largely decreases number of parameters used to describe the monogenic, digenic, trigenic, quadrigenic and epistatic genetic effects for QTL from 9408 by Kempthorne (1957) to 24 for two loci analysis. I also proposed methods for parameters estimation if the population was in linkage disequilibrium. This method provides a way to divide genetic variances into components explained by genetic effects and covariance between them in a linkage disequilibrium population. The variance components of genetic effects would correspond to those components in the linkage equilibrium population, while the components of covariance between genetic effects indicate the degree of disequilibrium in the population. However, there was over parameterization problem because of the small population size. Since the genetic effects could not be estimated in the full model in this case, I developed a reduced model to select a subset of statistically interested genetic effects.

Simulation examples demonstrated the feasibility of estimation of genetic parameters using this orthogonal model and validated the adequacy of parameters estimation under various situations. This quantitative genetic model for genetic analysis based no orthogonal comparison of

genotypic values at one or two QTL can also extend to more loci via the similar procedures if not considering about three or more-locus epistasis but just two-locus epistasis for pairs of loci, making the theoretical and methodological foundation for genetic analysis in these evolutionary and agriculturally important species.

In QTL mapping experiments, finite sample size would cause a practical problem in estimating the genetic effects, especially for autotetraploids. Some genotypes involving two or more loci would be observed rarely or not at all. Thus I selected a subset of statistically significant genetic effects in a reduced model and they could also be adequately estimated as shown in Table I-1.6.

1.6. References

Anderson, V.L. and Kempthorne, O. (1954) A model for the study of quantitative inheritance. **Genetics,** 39: 883-898.

Brim, C.A. and Cockerham, C.C. (1961) Inheritance of quantitative characters in soybeans. **Crop Sci.,** 1: 187-190.

Cockerham, C.C. (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. **Genetics**, 39: 859-882.

Cheverud, J.M. and Routman, E.J. (1995) Epistasis and its contribtion to genetic variance components. Genetics **139:** 1455–1461.

Cockerham, C.C. and Zeng, Z.B. (1996) Design III with marker loci. **Genetics,** 143: 1437-1456.

Casella, G. (2008) **Statistical design**. Springer. ISBN 978-0-387-75965-4.

Dempster, A.P. (1977) Maximum likelihood from incomplete data via the EM algorithm. **JR Statist Soc B,** 39: 1-22.

Doebley, J.A. and Gustus, C. (1995) teosinte branched l and origin of maize: evidence for epistasis and the evolution of dominance. **Genetics,** 141: 333-346.

Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. **Trans. Roy. Soc. Edinburgh,** 52: 399-433.

Falconer, D.S. and Mackay, T.F.C. (1996) **Introduction to quantitative genetics (Ed. 4).** Longman, Harlow, UK.

Goodnight, C.J. (2000) Quantitative trait loci and gene interaction: the quantitative genetics of metapopulation. **Heredity**, 84: 587-598.

Hinkelmann, K. and Kempthore, O. (1994) **Design and analysis of experiments.** New York: Wiley-Interscience.

Howell, D.C. (2010) **Statistical methods for psychology (7th ed.).** Belmont, CA: Thomson Wadsworth.

Kempthorne, O. (1955) The correlation between relatives in a simple autotetraploid population. **Genetics,** 40: 168-174.

Kempthorne, O. (1957) **An Introduction to Genetic Statistics.** New York: John Wiley & Sons.

Kao, C.H. and Zeng, Z.B. (1997) General formulas for obtaining the MLE and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. **Biometrics,** 53: 359-371.

Kuehl Robert, O. (2000) **Design of experiments: statistical principles of research design and analysis (2nd ed.).** Pacific Grove, CA: Duxbury/Thomson Learning.

Kao, C.H., et al. (2002) Modelling epistasis of quantitative trait loci using Cockerham's model. **Genetics,** 160: 1243-1261.

Li, C.C. (1957) The genetic variance of autotetraploids with two alleles. **Genetics,** 42: 583-592.

Lee, J.A., et al. (1968) The inheritance of gossypol level in gossypium. I: additive, dominance, epistatic, and maternal effects associated with seed gossypol in two varieties of Gossypium hirsulum L. **Genetics,** 59: 285-298.

Luo, Z.W., et al. (2000) Predicting parental genotypes and gene segregation for tetrasomic inheritance. **Theor Appl Genet,** 100: 1067-1073.

Luo, Z.W., et al. (2004) Theoretical basis for genetic linkage analysis in autotetraploid species. **PNAS,** 101: 7040-7045.

Luo, Z.W., et al. (2006) Modeling population genetic data in autotetraploid species. **Genetics,** 172: 639-646.

Leach, J.L., et al. (2010) Multilocus tetrasomic linkage analysis using hidden Markov chain model. **PNAS**, 107: 4270-4274.

Mather, K. (1982) **Biometrical Genetics.** London: Chapman and Hall.

Mead, R. (1988) **The design of experiment: statistical principles for practical application.** Cambridge University Press.

Nogueria, M. C. S. (2004) Orthogonal contrasts: definitions and concepts. **Sci. agric.,** 61(1): 118-124.

Scheffé, H. (1959) **Analysis of variance.** London: John Wiley & Sons.

Stuber, C.W. and Moll, R.H. (1971) Epistasis in maize (Zea mays L.) II. Comparison of selected with unselected populations. Genetics, 67: 137-149.

Stell, R.G.D. and Torrie, J.H. (1981) **Principles and procedures of statistics: A biometrical approach (2nd ed.).** New York MacGraw-Hill Book Company.

Stuber, C.W., et al. (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular marker. **Genetics,** 132: 832-839.

Soltis, D.E. (1995) The dynamic nature of polyploid genome. **PNAS**, 92: 8089-8091.

Song, K., et al. (1995) Rapid genome change in synthetic polyploids of Brassica and its imiplications for polyploid evolution. **PNAS,** 92: 7719-7723.

Winer, B.J. (1971) **Statistical principles in experimental design (2nd ed.).** New York MacGraw-Hill Book Company.

Wright, A.J. (1979) The use of differential coefficients in the development and interpretation of quantitative genetic models. **Heredity**, 43: 1-8.

Wang, J., et al. (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. **Genetics**, 167: 1961-1973.

Wang, T., et al. (2006) Models and partition of variance for quantitative trait loci with epistasis and linkage disequilibriuim. **BMC Genetics**, 7.

Zeng, Z.B., et al. (2000) Genetic architecture of a morphological shape difference between two drosophila species. **Genetics,** 154: 299-310.

Zeng, Z.B., et al. (2005) Modeling quantitative trait loci and interpretation of models. **Genetics,** 169: 1711-1725.

Chapter I-2: Interval mapping of QTL in autotetraploid species

2.1. Overview

Phenotypic variation in morphology, behaviour, physiology is widely exists in natural populations and is partly due to underlying genetic variation from segregation of alleles at multiple interacting loci and sensitivity to the external environmental condition (Mackay 2009). To understand the relationship between such variation in phenotypes and genetic variation for quantitative traits is a great challenge in the modern genomics. Insights into this question are helpful for predicting disease susceptibility and providing individual therapeutic treatments, for increasing the speed of selective breeding programmes in agriculturally and economically important plants and animals. The principles of mapping QTLs that affect phenotypic variation by linked to polymorphic molecular marker loci with Mendelian segregation have been known since the early twentieth century. With the rapid development in discovering of abundant molecular marker and effective genotyping method, the field of mapping QTLs has been revolutionized since the landmark work by Lander and Botstein (Lander 1989).

QTL mapping methods are well developed and widely used in diploid species nowadays. However, the corresponding methods for autotetraploids are still in the infancy stage for the much more complicated inheritance. Several reasons exist why theory and methods for QTL mapping in autotetraploids are far behind that in diploids. First, meiotic process in autotetraploids is quite different from that in diploids. Crossovers may occur between any two non-sister chromatids due to the formation of multivalent pairing of chromosomes during

meiosis. One of the most important features of autotetrasomic inheritance caused by quadrivalent pairing, double reduction, would result in segregation distortion in autotetrasomic linkage analysis. This feature implies that it is not appropriate to do QTL analysis in autotetraploids by remarkably reducing challenges in autotetrasomic inheritance under the assumption of bivalent pairing of homologous chromosomes during meiosis. This finding suggests a requirement to properly take account of the key features of gene segregation of autotetrasomic inheritance during QTL analysis. Second, the existence of multiplex alleles in autotetraploids would cause a substantially wider range of genotypic segregation. Since fully informative genetic markers are not available in autotetraploids, a smiple one-to-one relationhip between genotyping data and genotypes is usually difficult to obtain. The ideal marker loci selected for QTL mapping in auotetraploids should be highly polymorphic so as to be more informative. Thus theory and methods for QTL analysis in autotetraploids requires properly modelling the inheritance of multiplex alleles of the autotetraploids. Finally, by contrast with diploids, QTL parental genotypes and linkage phase between markers are unknown in practice in the autotetraploid mapping population. The general methods for identifying and mapping QTLs by linkage with markers are carried out based on crosses between lines that differ for the trait of interest. To achieve the maximum linkage disequilibrium between the loci in F_1 population, it is preferred that phenotypic increasing alleles should be homozygous in one parental line and phenotypic decreasing alleles should be homozygous in the other parental line. Homozygosity of QTLs in parental populations is likely to be met by divergent artificial selection for the trait of interest and subsequently inbreeding. In diploids, the parental inbred lines are commonly crossed to generate the F_1 population and then either backcrossed to one parent, or inbred to produce the F_2 population. Thus, QTL parental genotypes are known in

either the backcross design or the F_2 design. However, in autotetraploids, QTL parental genotypes are hardly to be known for two reasons: firstly, homozygosity of QTL in parental lines is difficult to be achieved; secondly, no sharing of common marker alleles between two parents is practically impossible. As a consequence, autotetraploid breeding has remained empirical and genetically unsophisticated, particularly for those quantitative traits.

Any QTL analysis can be divided into two statistically independent parts, namely genetics model and linkage analysis. For the first part, I have proposed an orthogonal based model for autotetraploids and was discussed in part I. This quantitative genetics model divided the genotypic value of an individual at the bi-allelic loci into monogenic, digenic, trigenic, quadrigenic and epistatic effects independently, which takes account properly the key features of autotetrasomic inheritance at the QTLs. For the second part, much effort has been made for linkage analysis and QTL mapping in autotetraploids facilitated by newly developed genotyping technologies, such as RAD sequencing (Baird 2008) and genotyping by sequencing (Elshire 2011), and advances in development of the theory and statistical method (Hackett 1998; Hackett 2001; Cao 2005; Hackett 2013). However, all these works did not properly take account of tetrasomic inheritance, typically assuming only bivalent pairing during meiosis. To improve modelling of data from autotetraploid populations, Luo et al developed the theoretical basis for linkage analysis with genetic marker data from autotetraploid segregation population (Luo et al 2000, 2004, 2006; Leach et al 2010). The theory takes properly account of several essential features of tetrasomic inheritance, under either bivalent pairing or quadrivalent paring, which has been used to construct markers linkage maps. This provides a useful start for QTL linkage analysis. However, from linkage analysis between markers to linkage analysis between markers and QTL, it is certainly not just a trivial extension. Besides the need to take properly account of

tetrasomic inheritance, this work would still face two main difficulties. First, the parental genotype of QTL is not observable, especially in the outbreeding tetraploid population. Second, linkage phase between marker alleles and QTL alleles is unknown in parents of most outbreeding auto tetraploid populations.

In early studies, some authors used regression models to compare mean of the trait for different phenotypes at a single marker (Meyer 1998, Sills 1995). But all these method gave little insight into QTL individual effects and interaction. Hackett (2001) established an interval mapping method to test for the existence of a QTL at positions between markers. To combat the challenges of the incomplete parental configuration information, Cao (2005) proposed model selection-based interval-mapping method under an autopolyploid bivalent pairing framework. Later, with the development of new sequencing and genotyping technologies, researchers are able to obtain high density SNP genotype data for mapping population. Based on these advances, Hackett used a new type of genome information, namely the allele dosage, inferred from allele intensity ratios by using of SNP arrays, to further improved the interval mapping method using SNP dosage information for autotetraploids (Hackett 2013; Hackett 2014). However, these methods did not take tetrasomic inheritance into consideration by assuming chromosomes were paired at random to give two bivalents and crossing over was restricted to within each bivalent, which did not account properly for the essential features of tetrasomic in heritance. Thus these methods are not appropriate for linkage analysis and QTL analysis with real datasets.

In this project, I proposed a likelihood-based method for mapping bi-allelic QTL in an outbred segregating population of an autotetraploid species. The method considers multivalent meiotic pairing of homologous chromosomes. I proposed an EM algorithm to estimate the model

parameters, including chromosomal location of QTL, QTL genetic effects, parental QTL genotypes and linkage phase between QTL and markers in parents. Simulation studies were carried out to investigate the performance of the method.

2.2. Introduction to QTL mapping in diploids

The principle of detecting and localizing QTLs is based on linkage disequilibrium between alleles at the QTL and alleles at the linked marker loci. If a QTL in closely linked to a marker locus, linkage disequilibrium will result in different mean values of the quantitative trait for individuals among different genotypes at marker locus. Thus the basic information required for mapping QTLs includes a linkage map of polymorphic marker loci that intensively distribute throughout the whole genome and variation for the quantitative trait within the study populations (Falconer 1996). With the rapid development of molecular technology, a variety of molecular markers can be chosen to construct the linkage map for QTL analysis, including simple sequence repeats (SSR), polymorphic insertions or deletions (indels) and single nucleotide polymorphisms (SNPs).

2.2.1. Analysis of variance

Studies of genetic mapping in diploids were pioneered early in twentieth century and succeeded to detect genetic linkage to the putative QTLs occasionally (Sax 1923; Rasmusson 1933; Thoday 1961; T anksley 1982; Edwards 1987). The simplest method of QTL mapping is

analysis of variance (ANOVA) at the marker loci. At each genotype marker, progeny may be classified into different groups according to their genotypes and compare mean value of phenotypes of these groups by F-statistic. This method is quite simple and even does not require a linkage map of markers. QTL location is roughly denoted by the marker showing the largest difference between marker genotype groups (the largest F statistic). However, QTL effects would be usually underestimated because of the recombination between the marker and the QTL and the power of detection QTL would be greatly decreased when markers are sparsely mapped. Thus mapping of QTLs in such a way will not be systematic nor accurate because genome wide distributed genetic markers were not available at that time and the statistical methods used were limited in both power and accuracy.

2.2.2. Interval mapping

As restriction fragment length polymorphisms (RFLPs) were detected as genetic marker (Botstein 1980), several statistical methods had been proposed to systematically map major QTLs in various populations (Soller 1976; Weller 1986; Lander 1989). The interval mapping method proposed by Lander and Botstein was the landmark work and become the most popular method for QTL analysis. They improved the method from exploiting single genetic markers one-at-a-time to test the presence of QTL within a pair of markers interval, which significantly increased the power of detecting QTL and effectively decreased the number of individuals to be genotyped to obtain a reasonable power (Lander 1989). The basic idea to construct likelihood function for interval mapping lies in calculation of a putative QTL genotype probability distribution given genotypes at the nearest flanking markers and recombination frequencies

between them. Given a putative QTL at location *z* at a chromosome, likelihood of the model parameters given the marker and trait phenotype data was evaluated in term of likelihood ratio statistic given below:

$$LOD(z) = \log\left(likelihood\ ratio\ comparing\ the\ hypothesis\ of\ a\ QTL\ at\ position\ z\ versus\ that\ of\ no\ QTL\right)$$
$$= \log\left(\frac{\Pr\left(observed\ data \middle| QTL\ at\ position\ z, \hat{\mu}_z, \hat{\sigma}_z^2\right)}{\Pr\left\{observed\ data \middle| no\ QTL, \hat{\mu}_0, \hat{\sigma}_0^2\right\}}\right)$$

Here $\hat{\mu}_z$ and $\hat{\sigma}_z^2$ are the MLEs of means and variance of QTL genotypes, assuming a single QTL at position *z*. In the no QTL model, the phenotypes are assumed to be independent and identically normally distributed with mean μ_0 and variances σ_0^2. The observed data includes the numbers of individuals and their phenotypes in each marker class. The LOD score indicates statistical significance for the presence of a QTL at the location *z*. The LOD score is calculated for varying locations of QTL along the chromosome and MLEs of z, μ and σ^2 are values for which the LOD score is maximized.

As the most popular method for single QTL analysis, interval mapping outperforms others mainly in three aspects. First, the profile of LOD score along the chromosome may provide information for the most likely position of QTL in the marker linkage map and allow inference of QTLs to be located between markers. Second, compared with ANOVA method, it largely enables estimation of QTL effects. Finally, interval mapping allows incomplete marker genotype data by changing location of the next available flanking marker (Lincoln 1992).

2.2.3. Multiple QTLs methods

If multiple QTLs exist on a chromosome, detecting QTLs by the method of interval mapping would be seriously biased (Knott 1992; Martinez 1992). To increase reliability of QTL mapping for multiple linked QTLs, Zeng (1994) established a method named composite interval mapping (CIM), in which the test statistic is not affected by QTLs located outside a given test interval. This method utilized the property of partial regression analysis to improve resolution and accuracy of detecting QTLs. However, the key problem with CIM is the choice of suitable markers, which in the closest position to the true QTLs, to serve as covariates. Apparently, if we could find these, the QTL mapping problem had already been solved anyway. Another interesting method proposed is multiple interval mapping (MIM), which is an extension of interval mapping to multiple QTLs and consider interaction between QTLs (Kao 1999; Zeng 1999). In the MIM model, genetic parameters are interpreted based on Cockerham's model which introduced orthogonal contrast scales into genetics model, facilitating readily analysis and evaluation in genetic effects of individuals and epistasis between QTLs. This method uses multiple marker intervals to simultaneously search for QTLs, which largely improves QTL mapping power and precision. Besides linkage mapping studies, several other methods, including Bayesian approach and genetic algorithm, have been introduced into QTL mapping by genetic researchers (Hoeschele 1993; Satagopan 1996; Uimari 1997; Sillanpaa 1999; Carlborg 2000).

2.3. A statistical framework of linkage analysis for autotetraploids

Linkage analysis is a prerequisite for QTL mapping. Luo and his group have worked on the topic of linkage analysis for autotetraploids for more than ten years since 2000 (Luo 2000). The first step made by Luo et al was development of method for prediction of parental genotypes at a codominant or dominant marker locus based on phenotype score of parents and progenies at that locus (Luo 2000). This progress properly takes account of all aspects of the complexities and provides solid basis for further linkage analysis.

Later Luo et al proposed a method for constructing linkage maps of molecular markers by assuming bivalent pairing of homologous chromosomes during meiosis of autotetraploids (Luo 2001). To better model tetrasomic inheritance, a more matured statistical method was proposed by Luo et al who considered the essential features of autotetrasomic inheritance such as the occurrence of double reduction and mixed bivalent and quadrivalent pairing during meiosis (Luo 2004). The general theory proposed a method to estimate the coefficient of double reduction and recombination frequency between two marker loci with tetrasomic inheritance. This work was able to calculate the probability distribution for the gamete modes in terms of the recombination frequency between the two marker loci and the coefficient of double reduction at the first marker locus.

Subsequently, Luo et al elaborated the linkage analysis for autotetraploids and demonstrated its reliability by constructing genetic marker linkage maps from a mapping population of cultivated autotetraploid potato (*Solanum tuberosum*) with dominant and codominant markers (Luo 2006). However, these algorithms searched for optimal markers order and genetic distance between markers based on pairwise linkage analysis. The strategy to construct genetic linkage maps is to

calculate the recombination frequencies and LOD scores for all possible linkage phase for each pair of markers, and then use the MLE of recombination frequency and LOD scores for the most likely linkage phase to assemble the genetic markers into linkage map using the JoinMap software (Stam, 1993). This strategy has been demonstrated to have high level of reliability for populations of size 150 or more (Hackett 1998).

To improve the reliability and efficiency of linkage map construction in autotetraploids, Leach et al (2010) proposed a statistical method for multi-locus linkage analysis using hidden Markov chain model that simultaneously utilized information from all partially informative markers along a chromosome. This study provided the way to calculate the conditional probability distribution of genotypes at any paired markers interval given the phenotypes of their linked genetic makers, facilitating further QTL analysis in autotetraploid species.

2.4. Methods of interval mapping for autotetraploids

2.4.1. Model notations

The QTL mapping approach is developed for the first generation of segregation population derived by crossing two autotetraploid parents, P_1 and P_2, composed of *n* offspring individuals. The parents can have up to eight distinct alleles at each marker locus: they are represented by alleles 1, 2, 3, 4 from parent P_1 and alleles 5, 6, 7, 8 from parent P_2.

With a given order of *m* molecular markers, $M_1, M_2, \ldots, M_m$, we denote matrix $p_1 = (p_1^j)_{j=1,\ldots,m}$ and $p_2 = (p_2^j)_{j=1,\ldots,m}$ as two parental marker phenotype data for P_1 and P_2, and denote matrix $o_i = (o_i^j)_{j=1,\ldots,m}$ as marker phenotype data for the i^{th} offspring individual. Here j (j=1, … , m) indicates the j^{th} marker locus. Similarly, a marker genotype is denoted as $g_1 = \left(g_1^j\right)_{j=1,\ldots,m}$, $g_2 = \left(g_2^j\right)_{j=1,\ldots,m}$ and $z_i = \left(z_i^j\right)_{j=1,\ldots,m}$, for the *m* markers for parents P_1, P_2 and the i^{th} offspring individual, respectively. Let $r_j \in r$ (j=1, 2, …, m-1) be the recombination frequency in the jth marker interval flanked by marker M_j and M_{j+1}, and $\alpha_j \in \alpha$ (j=1,2,…, m) be the coefficient of double reduction at the jth marker locus. Phenotypes p_1^j, p_2^j and o_i^j are denoted by 1×8 vectors (i.e. each locus may have up to eight different alleles coming from two parents) and we use 1/0 to indicate the presence/absence of one particular allele. Genotypes g_1^j, g_2^j and z_i^j are denoted by 1×4 vectors and alleles appearring in the same vector column for different loci indicates that these alleles are linked on the same chromosome. In practice, we can only observe marker phenotype data directly from experiments. However, we can estimate the parameters of recombination frequencies and coefficient of double reduction and infer the most-likely parental genotypes and linkage phase between alleles at marker loci (Luo 2000, 2006).

For a quantitative trait Y, we observe phenotypic value y_i for the i^{th} individual. Here we consider two alleles existing on the QTL, namely increasing phenotype allele, *Q*, and decreasing phenotype allele, *q*. Thus there are five different QTL genotypes as $qqqq$, $Qqqq$, $QQqq$, $QQQq$ and $QQQQ$. The genotypic values are denoted by $G_i\,(i = 0,1,\ldots,4)$, with i representing the

number of increasing phenotype alleles. The QTL genotypes of two parents P_1 and P_2 are denoted by q_{P_1} and q_{P_2}.

To showing logical relationship between these model notations, all the model notations mentioned above are summarized in Table I-2.1. The notations highlighted with pink colour are the observed data in this study. The notations highlighted with yellow colours are unobservable in practice but can be inferred and estimated based on previous work (Luo et al 2000, 2004, 2006; Leach et al 2010), so notations in yellow area are assumed to be known. Notations highlighted with both pink and yellow are input data in this model. The notations highlighted with green colour are not observable but will be used to identify and map QTL. Finally, the notations highlighted with blue colour are the parameters to be outputted in the model.

Table I-2.1. Notations used in interval mapping model for autotetraploids.

		Markers							QTL		
		M_1	M_2	…	M_j	…	M_{m-1}	M_m	Phenotypic value	Genotypes	Genotypic value
Parent P1	Phenotype p_1	p_1^1	p_1^2	…	p_1^j	…	p_1^{m-1}	p_1^m	T_1	$qqqq$	G_0
	Genotype g_1	g_1^1	g_1^2	…	g_1^j	…	g_1^{m-1}	g_1^m		$Qqqq$	G_1
Parent P2	Phenotype p_2	p_2^1	p_2^2	…	p_2^j	…	p_2^{m-1}	p_2^m	T_2	$QQqq$	G_2
	Genotype g_2	g_2^1	g_2^2	…	g_2^j	…	g_2^{m-1}	g_2^m		$QQQq$	G_3
The i^{th} offspring individual	Phenotype o_i	o_i^1	o_i^2	…	o_i^j	…	o_i^{m-1}	o_i^m	y_i	$QQQQ$	G_4
	Genotype z_i	z_i^1	z_i^2	…	z_i^j	…	z_i^{m-1}	z_i^m			
Recombination		r_1	r_2	…	r_j	…	r_{m-1}		Parental genotypes		q_{P_1} and q_{P_2}
Coefficient of double reduction (α)		α_1	α_2	…	α_j	…	α_{m-1}	α_m	Chromosomal location (cM)		L_{QTL}

2.4.2. Modelling for a quantitative trait

For simplicity but without loss of biological basis, we assume that there are two different alleles existing on the QTL, namely increasing phenotype allele, *Q*, and decreasing phenotype allele, *q*. As discussed in Chapter I-1. The phenotypic value of an autotetraploid individual *i* can be expressed as

$$y_i = G_{QTL} + \varepsilon \qquad \text{(I-2.1)}$$

where G_{QTL} is the genotypic value of QTL genotype and ε is a random variable following a normal distribution with mean of zero and variance of σ^2. G_{QTL} is given by

$$G_{QTL} = G_j = \mu + \theta_1\omega_{1j} + \theta_2\omega_{2j} + \theta_3\omega_{3j} + \theta_4\omega_{4j} \quad \textit{if the genotype of QTL is } Q_jq_{(4-j)}\,(j = 0,1,...,4) \qquad \text{(I-2.2)}$$

where μ is the population mean, and θ_i (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL, and $\omega_{ij}\,(i = 1,...,4; j = 0,1,...,4)$ are the corresponding orthogonal contrast scales. ω_{ij} variables are relevant to genotype frequencies of QTL as detailed in Chapter I-1. Genotype frequencies of QTL are unknown in practice, but can be expressed in terms of recombination frequencies between QTL and flanking markers given marker data and parental QTL genotypes, which I will discussed later in Section 2.4.4.

2.4.3. Interval mapping for autotetraploids

Here we develop an interval mapping approach for fitting a single QTL in a full-sib family, considering one chromosome at a time. Since markers of autotetraploids are not fully informative, we need to use the information from all markers on a chromosome to calculate the conditional probabilities of genotypes at QTL given parental QTL/markers genotypes and linkage phase between them and markers phenotypes of offspring. Let q_k (k=0, 1, … , 4) be the set of possible QTL genotypes for the ith individual (i.e. QTL is assumed to be biallelic here, so there are five possible genotypes at QTL, as *qqqq, Qqqq, QQqq, QQQq, QQQQ,* for all individuals) and k represent the number of increasing phenotype allele. Let $c_i^{jl} \in C_i^j \left(l = 1, \ldots, L_j\right)$ be the l^{th} chromosome configuration of interval j flanked by markers M_j and M_{j+1} for the i^{th} offspring individual. Here chromosome configuration refers to flanking marker genotypes, flanking marker linkage phase and parental chromosomes from which marker alleles come. The conditional probability distribution of c_i^{jl} can be calculated from three sources of information (Further discussed in Section 2.4.4): (1). Parental marker genotype data of the offspring i, g_1 and g_2, and the parental chromosomes from which the marker alleles come; (2). z_i^j and z_i^{j+1}, the i^{th} individual's genotypes and phases at the flanking markers M_j and M_{j+1}; (3). Phenotype of markers on the same chromosome but excluding the flankging markers of the i^{th} offspring individual. Then we can fit a QTL at location z in the jth interval and maximizing the likelihood for each location as a function of the QTL parameters $\Omega = \{\Omega_1, \Omega_2, \Omega_3\} = \{(\alpha, r, g_1, g_2), (r_j, r_{j1}, g_1^j, g_2^j, g_1^{j+1}, g_2^{j+1}, q_{P_1}, q_{P_2}), (\mu, \theta_1, \theta_2, \theta_3, \theta_4, \sigma^2)\}$. Ω_1 is a set parameters assumed to be known (i.e. can be estimated from marker data by Luo et al. 2000, 2004, 2006), including the

coefficient of double reuduction at marker loci, the recombination frequencies between markers and parental genotypes on the markers. Ω_2 is a set of known parameters relevant to the j^{th} marker interval and parental QTL genotypes. r_j indicates the recombination frequency in the j^{th} marker interval and r_{j1} indicates the recombination frequency between the marker M_j and the putative QTL in the j^{th} interval. g_1^j, g_2^j, g_1^{j+1} and g_2^{j+1} are parental marker genotypes at two flanking markers of the j^{th} interval. Ω_3 is a set of parameters of genetic effects, with μ is the population mean, and θ_i (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL, and σ^2 indicate residual variance.

The likelihood function of the trait phenotypic values, $y_i \in Y$, and the offspring marker data, $o_i \in O$ $(i = 1,...,n)$, of the n offspring individuals is

$$\begin{aligned} L(\Omega|O,Y) &= \prod_{i=1}^{n} \Pr\{o_i, y_i|\Omega_1,\Omega_2,\Omega_3\} = \prod_{i=1}^{n} \Pr\{y_i|o_i,\Omega_1,\Omega_2,\Omega_3\} \Pr\{o_i|\Omega_1,\Omega_2,\Omega_3\} \\ &\propto \prod_{i=1}^{n} \Pr\{y_i|o_i,\Omega_1,\Omega_2,\Omega_3\} \end{aligned} \quad \text{(I-2.3)}$$

Here,

$$\begin{aligned} \Pr\{y_i|o_i,\Omega_1,\Omega_2,\Omega_3\} &= \sum_{c_i^{jl}\in C_i^j} \sum_{k=0}^{4} \Pr\{y_i, c_i^{jl}, q_k|o_i,\Omega_1,\Omega_2,\Omega_3\} \\ &= \sum_{c_i^{jl}\in C_i^j} \sum_{k=o}^{4} f(y_i|q_k,\Omega_1,\Omega_2,\Omega_3) \Pr\{q_k|c_i^{jl},\Omega_1,\Omega_2,\Omega_3\} \Pr\{c_i^{jl}|o_i,\Omega_1,\Omega_2,\Omega_3\} \end{aligned}$$

(I-2.4)

We can simplified Equation (I-2.4) by considering three aspects as follows: Fisrt, the trait phenotype values, y_i , are only relevant to QTL genotypes, q_k , and parameters of genetic effects μ , θ_1 , θ_2 , θ_3 , θ_4 and residual variance σ^2 ; Second, QTL genotype distribution given two flanking markers is independent of genotypes at the other markers on the chromosome; Third, c_i^{jl} , configuration of chromosome interval flanked by markers M_j and M_{j+1}, is determined by the coefficient of double reduction, recombination frequencies, parental marker genotypes and offspring marker phenotypes. Thus Equation (I-2.4) can be rewritten as:

$$\Pr\left\{y_i \middle| o_i, \Omega_1, \Omega_2, \Omega_3\right\} = \sum_{c_i^{jl} \in C_i^j} \sum_{k=o}^{4} f\left(y_i \middle| q_k, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\} \Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\} \quad \text{(I-2.5)}$$

Thus the likelihood function can be calculated as

$$L\left(\Omega \middle| O, Y\right) \propto \prod_{i=1}^{n} \sum_{c_i^{jl} \in C_i^j} \sum_{k=o}^{4} f\left(y_i \middle| q_k, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\} \Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\} \quad \text{(I-2.6)}$$

The calculation of conditional probability $\Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\}$ and $\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\}$ is discussed in the next Section 2.4.4. The trait phenotypic value, y_i, is assumed to follow normal distribution, with G_k and σ^2 as the mean and variance for the k^{th} QTL genotype. G_k can be expressed as a linear function of genetic parameters (μ , θ_1 , θ_2 , θ_3 , θ_4) in Equation (I-2.2). Then

$$f\left(y_i \middle| q_k, \Omega_3\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - G_k\right)}{2\sigma^2}\right) \quad \text{(I-2.7)}$$

The maximum likelihood estimate of parameters G_k and σ^2 can be obtained by using EM algorithm (Dempster 1977). The E-step calculates the probability of the *i*th individual having the *k*th genotype on quantitative trait locus in the *j*th marker interval at the scanned location, ω_{i,q_k}^{j}

$$\begin{aligned}\omega_{i,q_k}^{j} &= \Pr\left\{q_k \middle| y_i, o_i, \Omega_1, \Omega_2, \Omega_3\right\} \\ &= \sum_{c_i^{jl} \in C_i^j} \Pr\left\{q_k \middle| y_i, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\} \Pr\left\{c_i^{jl} \middle| o_i, \Omega_1, \Omega_2, \Omega_3\right\} \\ &= \sum_{c_i^{jl} \in C_i^j} \frac{f\left(y_i \middle| q_k, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\} \Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\}}{\sum_{k=0}^{4} f\left(y_i \middle| q_k, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\}}\end{aligned} \tag{I-2.8}$$

According to Bayes theorem, $\Pr\left\{q_k \middle| y_i, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\}$ can be calculated as

$$\Pr\left\{q_k \middle| y_i, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\} = \frac{f\left(y_i \middle| q_k, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\}}{\sum_{k=0}^{4} f\left(y_i \middle| q_k, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\}} \tag{I-2.9}$$

Substituting by Equation (I-2.9), Equation (I-2.8) can be expressed by

$$\omega_{i,q_k}^{j} = \sum_{c_i^{jl} \in C_i^j} \frac{f\left(y_i \middle| q_k, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\} \Pr\left\{c_i^{jl} \middle| o_i, \Omega_1, \Omega_2, \Omega_3\right\}}{\sum_{k=0}^{4} f\left(y_i \middle| q_k, c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_1, \Omega_2, \Omega_3\right\}} \tag{I-2.10}$$

Since y_i is only relevant to q_k and Ω_3, q_k is only relevant to c_i^{jl} and Ω_2, and c_i^{jl} is only relevant to o_i and Ω_1, Equation (I-2.10) can be simplified as:

$$\omega_{i,q_k}^{j} = \sum_{c_i^{jl} \in C_i^j} \frac{f\left(y_i \middle| q_k, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\} \Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\}}{\sum_{k=0}^{4} f\left(y_i \middle| q_k, \Omega_3\right) \Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\}} \tag{I-2.11}$$

The M-step updates the estimates of the genetic parameters from

$$G_k = \sum_{i=1}^{n} \omega_{i,q_k}^{j} y_i \Big/ \sum_{i=1}^{n} \omega_{i,q_k}^{j} \tag{I-2.12}$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \sum_{k=0}^{4} \omega_{i,q_k}^{j} \left(y_i - G_k\right)^2 \Big/ n \tag{I-2.13}$$

As the E and M steps are repeated iteratively following equation (I-2.11), (I-2.12) and (I-2.13), the likelihood function will increase and the parameters will converge to the MLEs, G_k^* and σ^{*2}. Then the likelihood ratio of locating the QTL at the site of r_{j1} recombination frequency away from its left marker locus M_j, is

$$LOD\left(r_{j1}\right) = \log\left[\frac{L\left(G_k^*, \sigma^{*2} \middle| O, Y\right)}{L\left(G^{**}, \sigma^{**2} \middle| O, Y\right)}\right] \tag{I-2.14}$$

where

$$G^{**} = \sum_{i=1}^{n} y_i \Big/ n$$

$$\sigma^{**2} = \sum_{i=1}^{n} \left(y_i - G^{**}\right)^2 \Big/ n$$

G^{**} and σ^{**2} are estimates of mean of all genotypes and residual variance under no QTL model. In the no QTL model, the phenotypes are assumed to be independent and identically normally distributed with mean G and variances σ^2. The LOD score indicates statistical significance for the presence of a QTL at the location *z* with the j^{th} marker interval, with recombination frequency of r_{j1} between the QTL and the j^{th} marker. If every interval of each chromosome of the genome is scanned for the presence of QTL at all possible locations by the methods discussed above, a curve of LOD scores against the searched chromosomal locations will be obtained for every chromosome.

2.4.4. Calculation of QTL genotype probabilities distribution

In the previous section, I demonstrate the way to construct likelihood function in this model and establish an EM algorithm to estimate model parameters. To calculate the likelihood function and conditional probabilities of QTL genotypes, ω_{i,q_k}^{j}, two terms as mentioned in the previous Section 2.4.3, $\Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\}$ and $\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\}$, need to be calculated first. In this section, I will discuss the methods to carry out these calculations in details as follows.

2.4.4.1. The conditional probability of the chromosome configuration c_i^{jl} of interval *j* flanked by markers M_j and M_{j+1} for the *i*th offspring individual, given the marker phenotypes of the *i*th offspring individual and parental marker genotypes and linkage phase between them, $\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\}$

Given o_i^j and o_i^{j+1}, the *i*th individual's phenotypes at two flanking markers M_j and M_{j+1}, we can work out all the possible genotype and linkage phase configuration at these two marker loci, M_j and M_{j+1}, taking origins of marker alleles into consideration (i.e. consider parental chromosomes from which marker alleles come). We denote L_i^j as the total number of possible chromosome configuration of marker interval *j* for the offspring individual *i*. For clarity, $c_i^{jl}\left(l=1,...,L_i^j\right)$ can be expressed into a gamete configuration $a_i^{jl_a} / b_i^{jl_b}\ \left(l_a=1,...,L_{ai}^{\ j};\ l_b=1,...,L_{bi}^{\ j}\right)$ where $a_i^{jl_a}$ and $b_i^{jl_b}$ indicate gametes that make up of the offspring zygote. The c_i^{jl} expressed in terms of $a_i^{jl_a}$ and $b_i^{jl_b}$ is detailed as $c_i^{j1}=a_i^{j1}/b_i^{j1}$, $c_i^{j2}=a_i^{j1}/b_i^{j2}$, …, $c_i^{jL_{bi}^{\ j}}=a_i^{j1}/b_i^{jL_{bi}^{\ j}}$, $c_i^{j(L_{bi}^{\ j}+1)}=a_i^{j2}/b_i^{j1}$, …, $c_i^{jL_i^j}=a_i^{jL_{ai}^{\ j}}/b_i^{jL_{bi}^{\ j}}$. Thus $l=L_{bi}^{\ j}\left(l_a-1\right)+l_b$ and $L_i^j=L_{ai}^{\ j}L_{bi}^{\ j}$. Accordingly, the *i*th individual's genotypes of chromosome configuration $c_i^{jl}\left(l=1,...,L_i^j\right)$ at two flanking markers of interval *j*, $z_i^{j,l}z_i^{j+1,l}\left(l=1,...,L_i^j\right)$, can be expressed by gametes genotypes that make up the offspring zygote as $z_{ai}^{\ j,l_a}z_{ai}^{\ j+1,l_a}/z_{bi}^{\ j,l_b}z_{bi}^{\ j+1,l_b}\ \left(l_a=1,...,L_{ai}^{\ j};\ l_b=1,...,L_{bi}^{\ j}\right)$, respectively. Here $z_{ai}^{\ j,l_a}z_{ai}^{\ j+1,l_a}$ are gamete genotypes of marker configuration $a_i^{jl_a}$ at two flanking markers of interval *j*, and

$z_{bi}^{j,l_b} z_{bi}^{j+1,l_b}$ are another gamete genotypes of marker configuration $b_i^{jl_b}$ at two flanking markers of interval *j*. Then the conditional probability $\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\} (l = 1, \ldots L_j)$ can be expressed as

$$\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\} = \Pr\left\{a_i^{jl_a} / b_i^{jl_b} \middle| o_i, \Omega_1\right\} = \Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} / z_{bi}^{j,l_b} z_{bi}^{j+1,l_b} \middle| o_i, \Omega_1\right\} \quad \text{(I-2.15)}$$

where $\Omega_1 = \{\alpha, r, g_1, g_2\}$, α are the coefficient of double reduction at marker loci, r are recombination frequencies between markers, and g_1, g_2 are the parental genotypes at the marker loci. Since gametes are randomly unioned to generate offspring, the probability of zygote can be calculated as products of the probabilities of two gamtes as

$$\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\} = \Pr\left\{a_i^{jl_a} \middle| o_i, \Omega_1\right\} \Pr\left\{b_i^{jl_b} \middle| o_i, \Omega_1\right\} = \Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| o_i, \Omega_1\right\} \Pr\left\{z_{bi}^{j,l_b} z_{bi}^{j+1,l_b} \middle| o_i, \Omega_1\right\} \quad \text{(I-2.16)}$$

According to Bayes theorem, the conditional probability distribution of gamete genotypes at makers M_j and M_{j+1} can be calculated as

$$\Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| o_i, \Omega_1\right\} = \frac{\Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^m \middle| z_{ai}^{j,l_a} z_{ai}^{j+1,l_a}, \Omega_1\right\}}{\sum_{l_a=1}^{L_{ai}^j} \Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^m \middle| z_{ai}^{j,l_a} z_{ai}^{j+1,l_a}, \Omega_1\right\}} \quad \text{(I-2.17)}$$

Since the genotype distributions of markers at the left side of marker M_j given marker genotype of M_j is independent of marker genotype of M_{j+1} and vice versa, Equation (I-2.17) can be calculated by

$$\Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| o_i, \Omega_1\right\}$$
$$= \frac{\Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{ai}^{j,l_a}, \Omega_1\right\} \Pr\left\{o_i^{j+1} o_i^{j+2} \ldots o_i^m \middle| z_{ai}^{j+1,l_a}, \Omega_1\right\}}{\sum_{l_a=1}^{L_{ai}^j} \Pr\left\{z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{ai}^{j,l_a}, \Omega_1\right\} \Pr\left\{o_i^{j+1} o_i^{j+2} \ldots o_i^m \middle| z_{ai}^{j+1,l_a}, \Omega_1\right\}} \quad \text{(I-2.18)}$$

To calculate Equation (I-2.18), we have first to carry out the probability distribution of gamete genotypes at two linked loci from a bivalent or quadrivalent meiosis of autotetraploid species. Under bivalent pairing, the element $\Pr\left\{z_{ai}^{j,l_a} / z_{ai}^{j+1,l_a} \middle| \Omega_1\right\}$ in Equation (I-2.18) can be calculated by classifying the gametes generated from an bivalent meiosis into four modes of gamete formation according to the occurrence of recombination events. For simplicity, but without loss of generality, a general presentation for an autotetraploid genotype at two linked marker loci can be $A_1B_1 / A_2B_2 / A_3B_3 / A_4B_4$, indicating that A_i and B_i (i=1,2,3,4) are linked on the same chromosome. Let r be the recombination frequency between these two loci. During bivalent meiosis, four homologous chromosomes of an autotetraploid species is assumed to randomly form two pairs of bivalents with an equal chance of 1/3, and then recombination event occurs only within each paired bivalent. Table I-2.2 summarizes the probability distribution of two-locus gamete genotypes from a bivalent meiosis of autotetraploid species. Then $\Pr\left\{z_{ai}^{j,l_a} / z_{ai}^{j+1,l_a} \middle| \Omega_1\right\}$ is worked out in the last column of Table I-2.2, depending on which gamete mode $z_{ai}^{j,l_a} / z_{ai}^{j+1,l_a}$ belong to.

Table I-2.2. Probability distribution of the gamete genotypes at two linked loci from a bivalent meiosis of autotetraploid species

Gamete mode $(1 \le i, j, k, l \le 4)$	Frequency	Recombination events	Probabilities	
			Modes	Gametes
A_iB_i/A_jB_j	4	0	$(1-r)^2/3$	$(1-r)^2/12$
A_iB_i/A_jB_l	4	1	$(1-r)r/3$	$(1-r)r/12$
A_iB_k/A_jB_j	4	1	$r(1-r)/3$	$r(1-r)/12$
A_iB_k/A_jB_l	4	2	$r^2/3$	$r^2/12$

The number in the third column denotes the number of recombinant chromosomes in the gametes.

Under quadrivalent pairing, Luo et al. (2004) has classified the two-loucs gametes generated from a quadrivalent meiosis into eleven modes according to the occurrence of double reduction and recombination events during the process of gamete formation and carried out the corresponding probability distribution of gamete genotypes at two linked loci as detailed in Table I-2.3 (Table and annotation reproduced from Luo et al 2004). $\Pr\left\{z_{ai}^{j,l_a} / z_{ai}^{j+1,l_a} \middle| \Omega_1\right\}$ is worked out in terms of the coefficient of double reduction at locus M_j and the recombination frequency between marker M_j and M_{j+1} as detailed in the last column of Table I-2.3, depending on which gamete mode $z_{ai}^{j,l_a} / z_{ai}^{j+1,l_a}$ belong to.

Table I-2.3. Probability distribution of the modes of gamete formation and gamete genotypes at two linked loci from a quadrivalent meiosis of autotetraploid species (Table and annotation reproduced from Luo et al 2004)

Gamete mode $(1 \le i, j, k, l \le 4)$	Frequency	Double reduction and recombination events	Probabilities	
			Modes	Gametes
A_iB_i/A_iB_i	4	A and B (0)	$\alpha(1-r)^2$	$27\alpha(1-r)^2/108$
A_iB_j/A_iB_j	12	A and B (12)	$\alpha r^2/3$	$3\alpha r^2/108$
A_iB_i/A_iB_j	12	A (1)	$2\alpha r(1-r)$	$18\alpha r(1-r)/108$
A_iB_j/A_iB_k	12	A (2)	$2\alpha r^2/3$	$6\alpha r^2/108$
A_iB_i/A_jB_i	12	B (1)	$2(1-\alpha)r(1-r)/3$	$6(1-\alpha)r(1-r)/108$
A_iB_j/A_kB_j	12	B (2)	$2(1-\alpha)r^2/9$	$2(1-\alpha)r^2/108$
A_iB_i/A_jB_j	6	— (0)	$(1-\alpha)(1-r)^2$	$18(1-\alpha)(1-r)^2/108$
A_iB_i/A_jB_k	24	— (1)	$4(1-\alpha)r(1-r)/3$	$6(1-\alpha)r(1-r)/108$
A_iB_j/A_jB_i	6	— (2)	$(1-\alpha)r^2/9$	$2(1-\alpha)r^2/108$
A_iB_j/A_jB_k	24	— (2)	$4(1-\alpha)r^2/9$	$2(1-\alpha)r^2/108$
A_iB_j/A_kB_l	12	— (2)	$2(1-\alpha)r^2/9$	$2(1-\alpha)r^2/108$

The number in parentheses denotes the number of recombinant chromosomes in the gametes; —means that neither loci A nor B involves double reduction; α and r represent the coefficient of double reduction at locus A and the recombination frequency between locus A and B.

The next step to calculate the right, $\Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{ai}^{\ j+1,l_a}, \Omega_1\right\}$ and the left, $\Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{ai}^{\ j,l_a}, \Omega_1\right\}$ conditional probabilities defined in Equation (I-2.18) is readily solved by Leach et al (2010). Leach et al. (2010) proposed a method using "the Markov property of genotype distribution at linked loci, i.e., genotype of an individual at marker M_k given its genotype at M_{k-1} or M_{k+1} is independent of genotype at any other marker loci".

Similarly, the conditional probability distribution of another gamete genotypes at makers M_j and M_{j+1} can be calculated as

$$\begin{aligned}&\Pr\left\{z_{bi}^{\ j,l_b} z_{bi}^{\ j+1,l_b} \middle| o_i, \Omega_1\right\}\\ &= \frac{\Pr\left\{z_{bi}^{\ j,l_b} z_{bi}^{\ j+1,l_b} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{bi}^{\ j,l_b}, \Omega_1\right\} \Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{bi}^{\ j+1,l_b}, \Omega_1\right\}}{\sum_{l_b=1}^{L_{bi}^{\ j}} \Pr\left\{z_{bi}^{\ j,l_b} z_{bi}^{\ j+1,l_b} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{bi}^{\ j,l_b}, \Omega_1\right\} \Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{bi}^{\ j+1,l_b}, \Omega_1\right\}}\end{aligned} \quad \text{(I-2.19)}$$

Finally, substituting by Equation (I-2.18) and (I-2.19), the conditional probability $\Pr\left\{c_i^{\ jl} \middle| o_i, \Omega_1\right\}$ in Equation (I-2.16) can be calculated as

$$\begin{aligned}&\Pr\left\{c_i^{jl} \middle| o_i, \Omega_1\right\} = \Pr\left\{z_{ai}^{\ j,l_a} z_{ai}^{\ j+1,l_a} \middle| o_i, \Omega_1\right\} \Pr\left\{z_{bi}^{\ j,l_b} z_{bi}^{\ j+1,l_b} \middle| o_i, \Omega_1\right\}\\ &= \frac{\Pr\left\{z_{ai}^{\ j,l_a} z_{ai}^{\ j+1,l_a} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{ai}^{\ j,l_a}, \Omega_1\right\} \Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{ai}^{\ j+1,l_a}, \Omega_1\right\}}{\sum_{l_a=1}^{L_{ai}^{\ j}} \Pr\left\{z_{ai}^{\ j,l_a} z_{ai}^{\ j+1,l_a} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{ai}^{\ j,l_a}, \Omega_1\right\} \Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{ai}^{\ j+1,l_a}, \Omega_1\right\}} \bullet\\ &\frac{\Pr\left\{z_{bi}^{\ j,l_b} z_{bi}^{\ j+1,l_b} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{bi}^{\ j,l_b}, \Omega_1\right\} \Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{bi}^{\ j+1,l_b}, \Omega_1\right\}}{\sum_{l_b=1}^{L_{bi}^{\ j}} \Pr\left\{z_{bi}^{\ j,l_b} z_{bi}^{\ j+1,l_b} \middle| \Omega_1\right\} \Pr\left\{o_i^1 o_i^2 \ldots o_i^j \middle| z_{bi}^{\ j,l_b}, \Omega_1\right\} \Pr\left\{o_i^{j+1}o_i^{j+2}\ldots o_i^m \middle| z_{bi}^{\ j+1,l_b}, \Omega_1\right\}}\end{aligned} \quad \text{(I-2.20)}$$

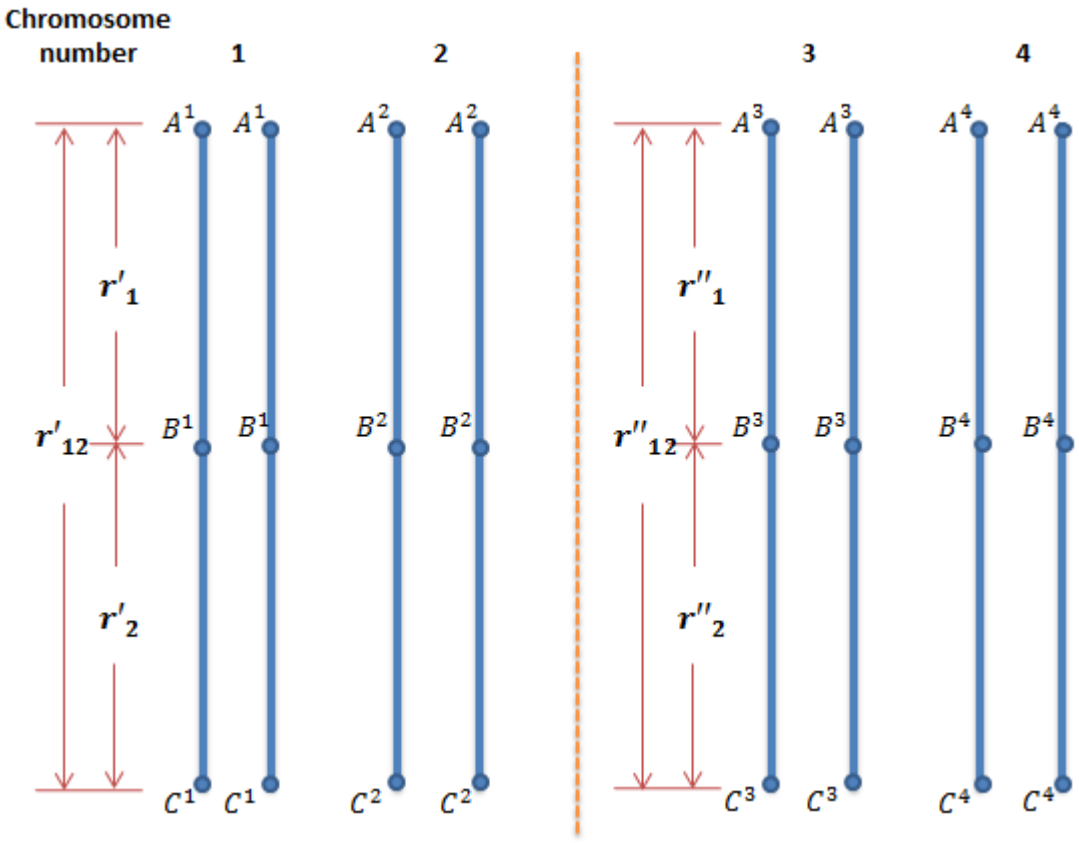
where $1 \le l = L_{bi}^{\ j}\left(l_a - 1\right) + l_b \le L$ and $L = L_{ai}^{\ j} \cdot L_{bi}^{\ j}$.

2.4.4.2. The conditional probability of the QTL genotype, given the chromosome configuration, c_i^{jl}, $\Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\}$

Given the chromosome configuration c_i^{jl}, we can identify possible QTL genotype and calculate their probability distribution for a putative QTL location along the chromosome. If a QTL, denoted as *Q*, locates in the j^{th} interval with flanking marker M_j and M_{j+1}, recombination frequencies between M_j and *Q*, M_{j+1} and *Q*, M_j and M_{j+1} are denoted as r_{j1}, r_{j2} and r_j, respectively, which satisfy $r_j = r_{j1} + r_{j2} - 2r_{j1}r_{j2}$ for bivalent pairing and $r_j = r_{j1} + r_{j2} - 4r_{j1}r_{j2}/3$ for quadrivalent pairing under the assumption of absence of recombination interference. The demonstration of relationship between r_{j1}, r_{j2} and r_j for bivalent pairing or quadrivalent pairing can be verified as follows:

Under bivalent pairing, we consider three linked loci *A*, *B* and *C* in autotetraploid species shown in Figure I-2.1. For simplicity, but without loss of generalilty, four duplicated homologous chromosomes are paired to create two bivalent pairs as: chromosome 1 is paired with chromosome 2 and chromosome 3 is paired with chromosome 4. Recombination event may occur only within each bivalent pair. We denote r'_1, r'_2 and r'_{12} as recombination frequencies in interval *AB*, *BC* and *AC* for the first bivalent pair, and r''_1, r''_2 and r''_{12} as recombination frequencies in interval *AB*, *BC* and *AC* for the second bivalent pair. The average recombination frequencies in interval *AB*, *BC* and *AC* are represented by r_1, r_2 and r_{12}. In each bivalent pair, behavious of crossover events is the same as that in diploids, thus recombination frequencies satisfy the relationship as $r'_{12} = r'_1 + r'_2 - 2r'_1 r'_2$ and $r''_{12} = r''_1 + r''_2 - 2r''_1 r''_2$. Then the average recombination frequency between marker *A* and *C*, r_{12} , can be calculated by

Figure I-2.1. Diagrammatic illustration of recombination events in the three-locus linkage model for autotetraploid species under bivalent pairing



Here the eight blue lines represent duplicated homologous chromosome in autotetraploids. Under bivalent pairing during meiosis, chromosomes are paired to create two bivalent pairs (i.e. here chromosome 1 is paired with chromosome 2 and chromosome 3 is paired with chromosome 4).

$$\begin{aligned} r_{12} &= \frac{1}{2}\left(r'_{12}+r''_{12}\right)=\frac{1}{2}\left(r'_1+r'_2-2r'_1r'_2+r''_1+r''_2-2r''_1r''_2\right) \\ &= \frac{1}{2}\left(r'_1+r''_1\right)+\frac{1}{2}\left(r'_2+r''_2\right)-2\left(\frac{1}{2}r'_1r'_2+\frac{1}{2}r''_1r''_2\right) \end{aligned} \tag{I-2.21}$$

Since it is assumed that recombination frequencies are the same between two bivalent pairs (i.e., $r'_1=r''_1,\ r'_2=r''_2$ and $r'_{12}=r''_{12}$), we have

$$r_1=\frac{1}{2}\left(r'_1+r''_1\right) \tag{I-2.22}$$

$$r_2=\frac{1}{2}\left(r'_2+r''_2\right) \tag{I-2.23}$$

$$\begin{aligned} 2r_1r_2 &= 2\cdot\frac{1}{2}\left(r'_1+r''_1\right)\cdot\frac{1}{2}\left(r'_2+r''_2\right) \\ &= 2\left(\frac{1}{4}r'_1r'_2+\frac{1}{4}r'_1r''_2+\frac{1}{4}r''_1r'_2+\frac{1}{4}r''_1r''_2\right) \\ &= 2\left(\frac{1}{4}r'_1r'_2+\frac{1}{4}r'_1r''_2+\frac{1}{4}r''_1r''_2+\frac{1}{4}r''_1r''_2\right) \\ &= 2\left(\frac{1}{2}r'_1r'_2+\frac{1}{2}r''_1r''_2\right) \end{aligned} \tag{I-2.24}$$

Substituting by Equation (I-2.22), (I-2.23) and (I-2.24), Equation (I-2.21) can calculated by

$$r_{12}=r_1+r_2-2r_1r_2 \tag{I-2.25}$$

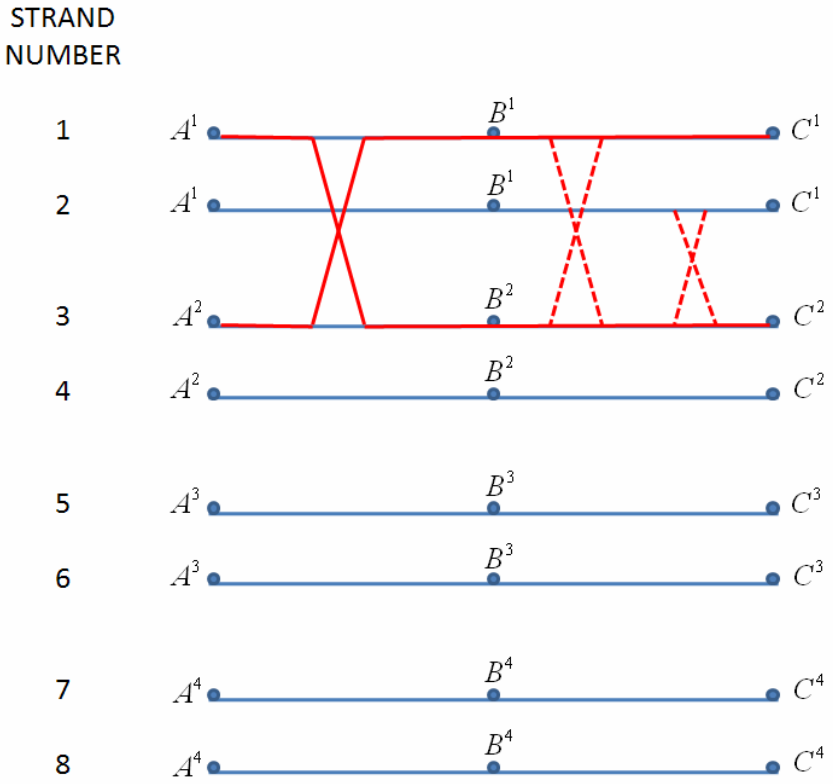
Under quadrivalent pairing, we consider three linked loci *A*, *B* and *C* in autotetraploid species shown in Figure I-2.1. We denote r_1, r_2 and r_{12} as recombination frequencies in interval *AB*, *BC* and *AC*, respectively. Assuming that crossovers take place randomly between four chromosomes and there is no recombination interference, we can deduce the relationship between r_1, r_2 and r_{12} as follows. In Figure I-2.1, the eight strands represent eight chromatids of autotetraploids with three loci *A*, *B* and *C*. A^i (B^i, C^i) (i=1, … , 4) indicate four different alleles on the four chromosomes, respectively. To calculate the recombination frequency, r_{12}, is equivalent to calculate the probability that allele A^1 on strand 1 will not be on the same chromosome of allele C^1 after recombination for simplified interpretation. If there is no recombination between allele A^1 and allele B^1 on strand 1, then recombination event between allele A^1 and C^1 would happen only if B^1 recombine to C^2, C^3 or C^4 on the remaining six strands. In this situation, the probability that allele A^1 will not be on the same chromosome of C^1 on strand 1 is $(1-r_1)r_2$. If recombination happens between allele A^1 and allele B^1 on strand 1, for example A^1 on strand 1 is connected to B^2 on strand 3 in Figure I-2.1, to ensure recombination event occuring between A^1 and C^1 on strand 1, two of the possible six recombination events for strand 1 in the second interval *BC* should not happen as shown as dotted red line in Figure I-2.1, or it will restore the connection between A^1 and C^1 on strand 1. In this situation, the probability that A^1 on strand 1 will not link to C^1 is $r_1(1-r_2/3)$. So we can calculate the probability that A^1 on strand 1 will not link to C^1 after recombination as $r_1(1-r_2/3)+(1-r_1)r_2$. Then we have the relationship between r_{12}, r_1 and r_2 as follows,

$$r_{12} = r_1(1-r_2/3) + (1-r_1)r_2 = r_1 + r_2 - 4r_1r_2/3 \quad \text{(I-2.21)}$$

where r_2 can be solved for a known r_j and a given r_1 as

$$r_2 = \frac{r_{12} - r_1}{1 - 4r_1/3} \quad \text{(I-2.22)}$$

Figure I-2.2. Diagrammatic illustration of recombination events in the three-locus linkage model for autotetraploid species under quadrivalent pairing



Here the eight blue lines represent duplicated homologous chromosomes in autotetraploids, with markers *A*, *B* and *C* locating along the chromosome. The red lines indicate recombination events occurring between non-homologous chromsomes.

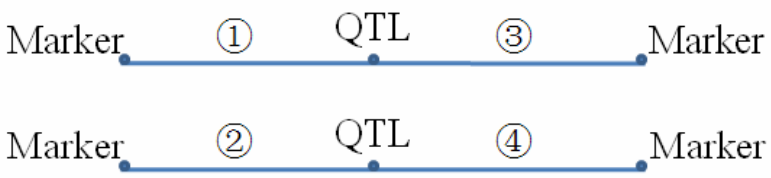
Given the relationship of recombination frequencies along three loci, we can proceed to calculate conditional probability distribution of QTL genotypes, q_k, given chromosome configuration, c_i^{jl}. Since the conditional probability distribution of QTL genotypes, q_k, is independent of other marker genotypes given the genotypes at its flanking markers, we have

$$\Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\} = \Pr\left\{q_k \middle| z_{ai}^{j,l_a} z_{ai}^{j+1,l_a} / z_{bi}^{j,l_b} z_{bi}^{j+1,l_b}, \Omega_2\right\} \quad \text{(I-2.23)}$$

where $\Omega_2 = \{ r_j, r_{j1}, g_1^j, g_2^j, g_1^{j+1}, g_2^{j+1}, q_{P_1}, q_{P_2} \}$, r_j is the recombination frequency in marker interval j, r_{j1} is the recombination frequency between the marker M_j and the QTL, g_1^j, g_2^j, g_1^{j+1}, g_2^{j+1} are parental genotypes at the marker M_j and M_{j+1}.

The gamete genotypes at two flanking markers of interval j can be classified into 4 or 11 modes according to the occurrence of double reduction and recombination events under bivalent pairing or quadrivalent pairing, as detailed in Table I-2.2 or Table I-2.3. For a QTL within the j^{th} interval, I classified the marker-QTL-marker configuration of each mode of gamete into 5 modes according to the recombination events. There may be up to 4 recombinaiton events occuring between the marker and QTL alleles in the diploid gamete of an autotetraploid individual. Figure I-2.3 shows the four possible positions of recombination events and lists all the possible combinations of recombination events. For example, if there is two recombination events occurring within the marker-QTL-marker configuration, there will be six possible combinations of recombination events occurring, detailed as: recombination events take place at position ① and ②, position ① and ③, position ① and ④, position ② and ④, position ② and ③, or postion ③ and ④.

Figure I-2.3. Recombination events in marker-QTL-marker configuration in a gamete for autotetraploids



No. of recombination events	No. of combinations	Possible combination of recombination events
0	1	No recombination
1	4	①; ②; ③; ④
2	6	①②; ①③; ①④; ②④; ②③; ③④
3	4	①②③; ①②④; ①③④; ②③④
4	1	①②③④

Here the two blue lines represent two chromosomes in a gamete of autotetraploids with a QTL locating between two flanking markers. Recombination events may occur within interval ①, ②, ③ and ④.

Based on the classification of two-locus gametes as detailed in Table I-2.2 and Table I-2.3, I further divided marker-QTL-marker configuration of gametes according to the number of recombination events in Table I-2.4 and Table I-2.5 under bivalent pairing or quadrivalent pairing. For example, we consider the first two-locus gamete mode, A_iB_i/A_iB_i (with probability of $\alpha(1-r_{12})^2$), in Table I-2.3 under quadrivalent pairing. Let α be the coefficient of double reduction at locus *A*, r_1, r_2 and r_{12} be the recombination frequencies between locus *A* and QTL, QTL and locus *B*, and locus *A* and locus *B*. For a QTL within the interval *AB*, I worked out the possible marker-QTL-marker configuration according to recombination events one by one: Fisrt, considering no recombination events occur within the interval, there will be only one possible marker-QTL-marker configuration, $A_iQ_iB_i/A_iQ_iB_i$, with probability of $\alpha(1-r_1)^2(1-r_2)^2$. And the conditional probability of the marker-QTL-marker configuration given the flanking marker genotypes is calculated by $\alpha(1-r_1)^2(1-r_2)^2/\alpha(1-r_{12})^2=(1-r_1)^2(1-r_2)^2/(1-r_{12})^2$. Given a particular two-locus gamete A_iB_i/A_iB_i, the frequency of three-locus gametes with configuration of $A_iQ_iB_i/A_iQ_iB_i$ is 1 and the conditional probability of a particular three-locus gamete with configuration of $A_iQ_iB_i/A_iQ_iB_i$ given flanking marker genotypes is $(1-r_1)^2(1-r_2)^2/(1-r_{12})^2$; Second, considering only one recombination event occurring, there will be no such marker-QTL-marker configurion; Third, considering two recombination events taking place, there is one possible marker-QTL-marker configuration (i.e., two recombination events take place at positions ①, ③ or ②, ④ and the outcome configurations are the same), $A_iQ_iB_i/A_iQ_jB_i$, with probability of $2\alpha r_1(1-r_1)r_2(1-r_2)/3$. And the conditional probability of the marker-QTL-marker configuration given the flanking marker genotypes is $2r_1(1-r_1)r_2(1-r_2)/3(1-r_{12})^2$.

Given a particular two-locus gamete A_iB_i/A_iB_i, the frequency of three-locus gametes with configuration $A_iQ_iB_i/A_iQ_jB_i$ is 6 (i.e., Q_j may present any of the two chromosome with any of the other three alleles) and the conditional probability of a particular three-locus gamete with configuration of $A_iQ_iB_i/A_iQ_jB_i$ given flanking marker genotypes is equal to $r_1(1-r_1)r_2(1-r_2)/9(1-r_{12})^2$; Fourth, considering three recombination events occurring, there will be no such marker-QTL-marker configurion; Fifth, considering four recombination events occurring, there are two possible marker-QTL-marker configurations, namely $A_iQ_jB_i/A_iQ_jB_i$ and $A_iQ_jB_i/A_iQ_kB_i$, with probability of $\alpha r_1^2 r_2^2/27$ and $2\alpha r_1^2 r_2^2/27$, respectively. Given a particular two-locus gamete A_iB_i/A_iB_i, the frequencies of three-locus gametes with configuration of $A_iQ_jB_i/A_iQ_jB_i$ and $A_iQ_jB_i/A_iQ_kB_i$ are 3 and 6. Accordingly, the conditional probability of a particular three-locus gamete with configuration of $A_iQ_jB_i/A_iQ_jB_i$ or $A_iQ_jB_i/A_iQ_kB_i$ given flanking marker genotypes is $r_1^2 r_2^2/81(1-r_{12})^2$. Similary, I worked out all the marker-QTL-marker configurations in this way and calculated the conditional probability distributions of gamete genotype at QTL given flanking marker genotypes as listed in the last column of Table I-2.4 or Table I-2.5 under bivalent pairing or quadrivalent pairing.

Let $z_{ai}^{j,l_a} q_a z_{ai}^{j+1,l_a}$ or $z_{bi}^{j,l_b} q_b z_{bi}^{j+1,l_b}$ denote gamete genotypes at marker M_j, QTL and marker M_{j+1}. The conditional probabilities, $\Pr\left\{q_a \middle| z_{ai}^{j,l_a} z_{ai}^{j+1,l_a}, \Omega_2\right\}$ and $\Pr\left\{q_b \middle| z_{bi}^{j,l_b} z_{bi}^{j+1,l_b}, \Omega_2\right\}$, have been worked out in the last column of Table I-2.4 or Table I-2.5 under bivalent pairing or quadrivalent pairing, as explained above. By assuming random union of gametes to generate

offspring zygote, the conditional probability distribution of QTL genotypes, q_k, given flanking marker genotypes in Equation (I-2.23), can be calculated by

$$\Pr\left\{q_k \middle| c_i^{jl}, \Omega_2\right\} = \sum_{q_a / q_b \in q_k} \Pr\left\{q_a \middle| z_{ai}^{j,l_a} z_{ai}^{j+1,l_a}, \Omega_2\right\} \Pr\left\{q_b \middle| z_{bi}^{j,l_b} z_{bi}^{j+1,l_b}, \Omega_2\right\} \quad \text{(I-2.24)}$$

Here $q_a / q_b \in q_k$ represents the QTL zygote consisted of gametes q_a and q_b has the genotype q_k.

Table I-2.4. Conditional probability distribution of marker-QTL-marker gamete modes and gametic genotypes given two flanking genotypes from a bivalent meiosis of autotetraploid species

Gamete modes of flanking markers $(1 \le i, j, k, l \le 4)$	Recombination events	Marker-QTL-marker configuration	Probability	Frequency	Conditional probability of marker-QTL-marker gametes
$\frac{A_iB_i}{A_jB_j}$	0	$\frac{A_iQ_iB_i}{A_jQ_jB_j}$	$(1-r_1)^2(1-r_2)^2/3$	1	$(1-r_1)^2(1-r_2)^2/(1-r_{12})^2$
	1	-	-	-	-
	2	$\frac{A_iQ_kB_i}{A_jQ_jB_j}$	$r_1(1-r_1)r_2(1-r_2)/3$	1	$r_1(1-r_1)r_2(1-r_2)/(1-r_{12})^2$
		$\frac{A_iQ_iB_i}{A_jQ_lB_j}$	$r_1(1-r_1)r_2(1-r_2)/3$	1	$r_1(1-r_1)r_2(1-r_2)/(1-r_{12})^2$
	3	-	-	-	-
	4	$\frac{A_iQ_kB_i}{A_jQ_lB_j}$	$r_1^2r_2^2/3$	1	$r_1^2r_2^2/(1-r_{12})^2$
Total			$(1-r_{12})^2/3$		

$\frac{A_iB_i}{A_jB_l}$	0	-	-	-	-
	1	$\frac{A_iQ_iB_i}{A_jQ_lB_l}$	$r_1(1-r_1)(1-r_2)^2/3$	1	$r_1(1-r_1)(1-r_2)^2/r_{12}(1-r_{12})$
		$\frac{A_iQ_iB_i}{A_jQ_jB_l}$	$(1-r_1)^2 r_2(1-r_2)/3$	1	$(1-r_1)^2 r_2(1-r_2)/r_{12}(1-r_{12})$
	2	-	-	-	-
	3	$\frac{A_iQ_kB_i}{A_jQ_jB_l}$	$r_1(1-r_1)r_2^2/3$	1	$r_1(1-r_1)r_2^2/r_{12}(1-r_{12})$
		$\frac{A_iQ_kB_i}{A_jQ_lB_l}$	$r_1^2 r_2(1-r_2)/3$	1	$r_1^2 r_2(1-r_2)/r_{12}(1-r_{12})$
	4	-	-	-	-
Total			$r_{12}(1-r_{12})/3$		

$\frac{A_iB_k}{A_jB_j}$	0	-	-	-	-
	1	$\frac{A_iQ_kB_k}{A_jQ_jB_j}$	$r_1(1-r_1)(1-r_2)^2/3$	1	$r_1(1-r_1)(1-r_2)^2/r_{12}(1-r_{12})$

		$\frac{A_i Q_i B_k}{A_j Q_j B_j}$	$(1-r_1)^2 r_2 (1-r_2)/3$	1	$(1-r_1)^2 r_2 (1-r_2)/r_{12}(1-r_{12})$
	2	-	-	-	-
	3	$\frac{A_i Q_k B_k}{A_j Q_l B_j}$	$r_1^2 r_2 (1-r_2)/3$	1	$r_1^2 r_2 (1-r_2)/r_{12}(1-r_{12})$
		$\frac{A_i Q_i B_k}{A_j Q_l B_j}$	$r_1 (1-r_1) r_2^2/3$	1	$r_1 (1-r_1) r_2^2/r_{12}(1-r_{12})$
	4	-	-	-	-
Total			$r_{12}(1-r_{12})/3$		

$\frac{A_i B_k}{A_j B_l}$	0	-	-	-	-
	1	-	-	-	-
	2	$\frac{A_i Q_k B_k}{A_j Q_l B_l}$	$r_1^2 (1-r_2)^2/3$	1	$r_1^2 (1-r_2)^2/r_{12}^2$
		$\frac{A_i Q_k B_k}{A_j Q_j B_l}$	$r_1 (1-r_1) r_2 (1-r_2)/3$	1	$r_1 (1-r_1) r_2 (1-r_2)/r_{12}^2$

		$\frac{A_i Q_i B_k}{A_j Q_l B_l}$	$r_1(1-r_1)r_2(1-r_2)/3$	1	$r_1(1-r_1)r_2(1-r_2)/r_{12}^2$
		$\frac{A_i Q_i B_k}{A_j Q_j B_l}$	$(1-r_1)^2 r_2^{\,2}/3$	1	$(1-r_1)^2 r_2^{\,2}/r_{12}^2$
	3	-	-	-	-
	4	-	-	-	-
Total	$r_{12}^{\,2}/3$				

Table I-2.5. Conditional probability distribution of marker-QTL-marker gamete modes and gametic genotypes given two flanking marker genotypes from a quadrivalent meiosis of autotetraploid species

Gamete modes of flanking markers $1 \le i, j, k, l \le 4$	Recombination events	Possible marker-QTL-marker configuration	Probability	Frequency	Conditional probability of marker-QTL-marker gametes	
					Modes	Gametes
$\frac{A_iB_i}{A_iB_i}$	0	$\frac{A_iQ_iB_i}{A_iQ_iB_i}$	$\alpha(1-r_1)^2(1-r_2)^2$	1	$\frac{(1-r_1)^2(1-r_2)^2}{(1-r_{12})^2}$	$\frac{(1-r_1)^2(1-r_2)^2}{(1-r_{12})^2}$
	1	-	-	-	-	-
	2	$\frac{A_iQ_iB_i}{A_iQ_jB_i}$	$\frac{2}{3}\alpha r_1(1-r_1)r_2(1-r_2)$	6	$\frac{2r_1(1-r_1)r_2(1-r_2)}{3(1-r_{12})^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{9(1-r_{12})^2}$
	3	-	-	-	-	-
	4	$\frac{A_iQ_jB_i}{A_iQ_jB_i}, \frac{A_iQ_jB_i}{A_iQ_kB_i}$	$\alpha r_1^2 r_2^2/9$	3, 6	$r_1^2 r_2^2/9(1-r_{12})^2$	$r_1^2 r_2^2/81(1-r_{12})^2$
Total			$\alpha(1-r_{12})^2$			
$\frac{A_iB_j}{A_iB_j}$	0	-	-	-	-	-

	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_j}{A_iQ_jB_j}$	$\alpha r_1^2(1-r_2)^2/3$	1	$r_1^2(1-r_2)^2/r_{12}^2$	$r_1^2(1-r_2)^2/r_{12}^2$
		$\frac{A_iQ_iB_j}{A_iQ_jB_j}$	$\frac{2}{3}\alpha r_1(1-r_1)r_2(1-r_2)$	2	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$
		$\frac{A_iQ_iB_j}{A_iQ_iB_j}$	$\alpha(1-r_1)^2r_2^2/3$	1	$(1-r_1)^2r_2^2/r_{12}^2$	$(1-r_1)^2r_2^2/r_{12}^2$
	3	$\frac{A_iQ_iB_j}{A_iQ_kB_j}$	$4\alpha r_1(1-r_1)r_2^2/9$	4	$4r_1(1-r_1)r_2^2/3r_{12}^2$	$r_1(1-r_1)r_2^2/3r_{12}^2$
		$\frac{A_iQ_jB_j}{A_iQ_kB_j}$	$4\alpha r_1^2r_2(1-r_2)/9$	4	$4r_1^2r_2(1-r_2)/3r_{12}^2$	$r_1^2r_2(1-r_2)/3r_{12}^2$
	4	$\frac{A_iQ_kB_j}{A_iQ_kB_j}, \frac{A_iQ_kB_j}{A_iQ_lB_j}$	$4\alpha r_1^2r_2^2/27$	2, 2	$4r_1^2r_2^2/9r_{12}^2$	$r_1^2r_2^2/9r_{12}^2$
Total			$\alpha r_{12}^2/3$			
$\frac{A_iB_i}{A_iB_j}$	0	-	-	-	-	-

	1	$\frac{A_iQ_iB_i}{A_iQ_jB_j}$	$2\alpha r_1(1-r_1)(1-r_2)^2$	1	$\frac{r_1(1-r_1)(1-r_2)^2}{r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)(1-r_2)^2}{r_{12}(1-r_{12})}$
		$\frac{A_iQ_iB_i}{A_iQ_iB_j}$	$2\alpha(1-r_1)^2 r_2(1-r_2)$	1	$\frac{(1-r_1)^2 r_2(1-r_2)}{r_{12}(1-r_{12})}$	$\frac{(1-r_1)^2 r_2(1-r_2)}{r_{12}(1-r_{12})}$
	2	$\frac{A_iQ_iB_i}{A_iQ_kB_j}$	$\frac{4}{3}\alpha r_1(1-r_1) r_2(1-r_2)$	2	$\frac{2r_1(1-r_1) r_2(1-r_2)}{3r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1) r_2(1-r_2)}{3r_{12}(1-r_{12})}$
	3	$\frac{A_iQ_jB_i}{A_iQ_jB_j}, \frac{A_iQ_kB_i}{A_iQ_jB_j}$	$2\alpha r_1^2 r_2(1-r_2)/3$	1, 2	$\frac{r_1^2 r_2(1-r_2)}{3r_{12}(1-r_{12})}$	$\frac{r_1^2 r_2(1-r_2)}{9r_{12}(1-r_{12})}$
		$\frac{A_iQ_jB_i}{A_iQ_iB_j}, \frac{A_iQ_kB_i}{A_iQ_iB_j}$	$2\alpha r_1(1-r_1) r_2^2/3$	1, 2	$\frac{r_1(1-r_1) r_2^2}{3r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1) r_2^2}{9r_{12}(1-r_{12})}$
	4	$\frac{A_iQ_kB_i}{A_iQ_kB_j}, \frac{A_iQ_jB_i}{A_iQ_kB_j}, \frac{A_iQ_lB_i}{A_iQ_kB_j}$	$4\alpha r_1^2 r_2^2/9$	2, 2, 2	$2r_1^2 r_2^2/9r_{12}(1-r_{12})$	$r_1^2 r_2^2/27r_{12}(1-r_{12})$
Total			$2\alpha r_{12}(1-r_{12})$			
	0	-	-	-	-	-

$\frac{A_iB_j}{A_iB_k}$	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_j}{A_iQ_kB_k}$	$2\alpha r_1^2(1-r_2)^2/3$	1	$r_1^2(1-r_2)^2/r_{12}^2$	$r_1^2(1-r_2)^2/r_{12}^2$
		$\frac{A_iQ_jB_j}{A_iQ_iB_k}, \frac{A_iQ_iB_j}{A_iQ_kB_k}$	$\frac{4}{3}\alpha r_1(1-r_1)r_2(1-r_2)$	1, 1	$2r_1(1-r_1)r_2(1-r_2)/r_{12}^2$	$r_1(1-r_1)r_2(1-r_2)/r_{12}^2$
		$\frac{A_iQ_iB_j}{A_iQ_iB_k}$	$2\alpha(1-r_1)^2r_2^2/3$	1	$(1-r_1)^2r_2^2/r_{12}^2$	$(1-r_1)^2r_2^2/r_{12}^2$
	3	$\frac{A_iQ_kB_j}{A_iQ_kB_k}, \frac{A_iQ_lB_j}{A_iQ_kB_k}, \frac{A_iQ_jB_j}{A_iQ_jB_k}, \frac{A_iQ_jB_j}{A_iQ_lB_k}$	$8\alpha r_1^2r_2(1-r_2)/9$	1, 1, 1, 1	$4r_1^2r_2(1-r_2)/3r_{12}^2$	$r_1^2r_2(1-r_2)/3r_{12}^2$
		$\frac{A_iQ_kB_j}{A_iQ_iB_k}, \frac{A_iQ_lB_j}{A_iQ_iB_k}, \frac{A_iQ_iB_j}{A_iQ_jB_k}, \frac{A_iQ_iB_j}{A_iQ_lB_k}$	$8\alpha r_1(1-r_1)r_2^2/9$	1, 1, 1, 1	$4r_1(1-r_1)r_2^2/3r_{12}^2$	$r_1(1-r_1)r_2^2/3r_{12}^2$
	4	$\frac{A_iQ_kB_j}{A_iQ_jB_k}, \frac{A_iQ_kB_j}{A_iQ_lB_k}$,	$8\alpha r_1^2r_2^2/27$	1, 1, 1, 1	$4r_1^2r_2^2/9r_{12}^2$	$r_1^2r_2^2/9r_{12}^2$

		$\frac{A_iQ_lB_j}{A_iQ_jB_k}, \frac{A_iQ_lB_j}{A_iQ_lB_k}$				
Total	$2\alpha r_{12}^2/3$					

$\frac{A_iB_i}{A_jB_i}$	0	-	-	-	-	-
	1	$\frac{A_iQ_iB_i}{A_jQ_iB_i}$	$\frac{2}{3}(1-\alpha)r_1(1-r_1)(1-r_2)^2$	1	$\frac{r_1(1-r_1)(1-r_2)^2}{r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)(1-r_2)^2}{r_{12}(1-r_{12})}$
		$\frac{A_iQ_iB_i}{A_jQ_jB_i}$	$\frac{2}{3}(1-\alpha)(1-r_1)^2 r_2(1-r_2)$	1	$\frac{(1-r_1)^2 r_2(1-r_2)}{r_{12}(1-r_{12})}$	$\frac{(1-r_1)^2 r_2(1-r_2)}{r_{12}(1-r_{12})}$
	2	$\frac{A_iQ_iB_i}{A_jQ_kB_i}$	$\frac{4}{9}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	2	$\frac{2r_1(1-r_1)r_2(1-r_2)}{3r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{3r_{12}(1-r_{12})}$
	3	$\frac{A_iQ_jB_i}{A_jQ_iB_i}, \frac{A_iQ_kB_i}{A_jQ_iB_i}$	$2(1-\alpha)r_1^2r_2(1-r_2)/9$	1, 2	$\frac{r_1^2r_2(1-r_2)}{3r_{12}(1-r_{12})}$	$\frac{r_1^2r_2(1-r_2)}{9r_{12}(1-r_{12})}$
		$\frac{A_iQ_jB_i}{A_jQ_jB_i}, \frac{A_iQ_kB_i}{A_jQ_jB_i}$	$2(1-\alpha)r_1(1-r_1)r_2^2/9$	1,2	$\frac{r_1(1-r_1)r_2^2}{3r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)r_2^2}{9r_{12}(1-r_{12})}$
	4	$\frac{A_iQ_jB_i}{A_jQ_kB_i}, \frac{A_iQ_kB_i}{A_jQ_kB_i},$	$4(1-\alpha)r_1^2r_2^2/27$	2, 2, 2	$2r_1^2r_2^2/9r_{12}(1-r_{12})$	$r_1^2r_2^2/27r_{12}(1-r_{12})$

		$\frac{A_iQ_lB_i}{A_jQ_kB_i}$				
Total			$2(1-\alpha)r_{12}(1-r_{12})/3$			

$\frac{A_iB_j}{A_kB_j}$	0	-	-	-	-	-
	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_j}{A_kQ_jB_j}$	$2(1-\alpha)r_1^2(1-r_2)^2/9$	1	$r_1^2(1-r_2)^2/r_{12}^2$	$r_1^2(1-r_2)^2/r_{12}^2$
		$\frac{A_iQ_jB_j}{A_kQ_kB_j}, \frac{A_iQ_iB_j}{A_kQ_jB_j}$	$\frac{4}{9}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	1, 1	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$
		$\frac{A_iQ_iB_j}{A_kQ_kB_j}$	$2(1-\alpha)(1-r_1)^2r_2^2/9$	1	$(1-r_1)^2r_2^2/r_{12}^2$	$(1-r_1)^2r_2^2/r_{12}^2$
	3	$\frac{A_iQ_kB_j}{A_kQ_jB_j}, \frac{A_iQ_lB_j}{A_kQ_jB_j}, \frac{A_iQ_jB_j}{A_kQ_iB_j}, \frac{A_iQ_jB_j}{A_kQ_lB_j}$	$8(1-\alpha)r_1^2r_2(1-r_2)/27$	1, 1, 1, 1	$4r_1^2r_2(1-r_2)/3r_{12}^2$	$r_1^2r_2(1-r_2)/3r_{12}^2$
		$\frac{A_iQ_kB_j}{A_kQ_kB_j}$,	$8(1-\alpha)r_1(1-r_1)r_2^2/27$	1, 1, 1, 1	$4r_1(1-r_1)r_2^2/3r_{12}^2$	$r_1(1-r_1)r_2^2/3r_{12}^2$

		$\frac{A_iQ_lB_j}{A_kQ_kB_j}$, $\frac{A_iQ_iB_j}{A_kQ_iB_j}$, $\frac{A_iQ_iB_j}{A_kQ_lB_j}$				
	4	$\frac{A_iQ_kB_j}{A_kQ_lB_j}$, $\frac{A_iQ_lB_j}{A_kQ_iB_j}$, $\frac{A_iQ_kB_j}{A_kQ_iB_j}$, $\frac{A_iQ_lB_j}{A_kQ_lB_j}$	$8(1-\alpha)r_1^2r_2^2/81$	1, 1, 1, 1	$4r_1^2r_2^2/9r_{12}^2$	$r_1^2r_2^2/9r_{12}^2$
Total		$2(1-\alpha)r_{12}^2/9$				

$\frac{A_iB_i}{A_jB_j}$	0	$\frac{A_iQ_iB_i}{A_jQ_jB_j}$	$(1-\alpha)r_1^2(1-r_2)^2$	1	$r_1^2(1-r_2)^2/(1-r_{12})^2$	$r_1^2(1-r_2)^2/(1-r_{12})^2$
	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_i}{A_jQ_jB_j}$, $\frac{A_iQ_kB_i}{A_jQ_jB_j}$, $\frac{A_iQ_iB_i}{A_jQ_iB_j}$, $\frac{A_iQ_iB_i}{A_jQ_kB_j}$	$\frac{2}{3}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	1, 2, 1, 2	$\frac{2r_1(1-r_1)r_2(1-r_2)}{3(1-r_{12})^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{9(1-r_{12})^2}$
	3	-	-	-	-	-

	4	$\frac{A_iQ_jB_i}{A_jQ_iB_j}, \frac{A_iQ_jB_i}{A_jQ_kB_j}, \frac{A_iQ_kB_i}{A_jQ_iB_j}, \frac{A_iQ_kB_i}{A_jQ_kB_j}, \frac{A_iQ_kB_i}{A_jQ_lB_j}$	$(1-\alpha)r_1^2r_2^2/9$	1, 2, 2, 2, 2	$r_1^2r_2^2/9(1-r_{12})^2$	$r_1^2r_2^2/81(1-r_{12})^2$
Total		$(1-\alpha)(1-r_{12})^2$				

$\frac{A_iB_i}{A_jB_k}$	0	-	-	-	-	-
	1	$\frac{A_iQ_iB_i}{A_jQ_kB_k}$	$\frac{4}{3}(1-\alpha)r_1(1-r_1)(1-r_2)^2$	1	$\frac{r_1(1-r_1)(1-r_2)^2}{r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)(1-r_2)^2}{r_{12}(1-r_{12})}$
		$\frac{A_iQ_iB_i}{A_jQ_jB_k}$	$\frac{4}{3}(1-\alpha)(1-r_1)^2r_2(1-r_2)$	1	$\frac{(1-r_1)^2r_2(1-r_2)}{r_{12}(1-r_{12})}$	$\frac{(1-r_1)^2r_2(1-r_2)}{r_{12}(1-r_{12})}$
	2	$\frac{A_iQ_iB_i}{A_jQ_iB_k}, \frac{A_iQ_iB_i}{A_jQ_lB_k}$	$\frac{8}{9}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	1, 1	$\frac{2r_1(1-r_1)r_2(1-r_2)}{3r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{3r_{12}(1-r_{12})}$
	3	$\frac{A_iQ_kB_i}{A_jQ_kB_k}$,	$4(1-\alpha)r_1^2r_2(1-r_2)/9$	1, 1, 1	$\frac{r_1^2r_2(1-r_2)}{3r_{12}(1-r_{12})}$	$\frac{r_1^2r_2(1-r_2)}{9r_{12}(1-r_{12})}$

		$\frac{A_iQ_jB_i}{A_jQ_kB_k}, \frac{A_iQ_lB_i}{A_jQ_kB_k}$				
		$\frac{A_iQ_jB_i}{A_jQ_jB_k}, \frac{A_iQ_kB_i}{A_jQ_jB_k}, \frac{A_iQ_lB_i}{A_jQ_jB_k}$	$4(1-\alpha)r_1(1-r_1)r_2^2/9$	1, 1, 1	$\frac{r_1(1-r_1)r_2^2}{3r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)r_2^2}{9r_{12}(1-r_{12})}$
	4	$\frac{A_iQ_jB_i}{A_jQ_iB_k}, \frac{A_iQ_kB_i}{A_jQ_iB_k}, \frac{A_iQ_lB_i}{A_jQ_iB_k}, \frac{A_iQ_jB_i}{A_jQ_lB_k}, \frac{A_iQ_kB_i}{A_jQ_lB_k}, \frac{A_iQ_lB_i}{A_jQ_lB_k}$	$8(1-\alpha)r_1^2r_2^2/27$	1, 1, 1, 1, 1, 1	$2r_1^2r_2^2/9r_{12}(1-r_{12})$	$r_1^2r_2^2/27r_{12}(1-r_{12})$
Total			$4(1-\alpha)r_{12}(1-r_{12})/3$			

$\frac{A_iB_j}{A_jB_i}$	0	-	-	-	-	-
	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_j}{A_jQ_iB_i}$	$(1-\alpha)r_1^2(1-r_2)^2/9$	1	$r_1^2(1-r_2)^2/r_{12}^2$	$r_1^2(1-r_2)^2/r_{12}^2$

		$\frac{A_iQ_jB_j}{A_jQ_jB_i}, \frac{A_iQ_iB_j}{A_jQ_iB_i}$	$\frac{2}{9}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	1, 1	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$
		$\frac{A_iQ_iB_j}{A_jQ_jB_i}$	$(1-\alpha)(1-r_1)^2 r_2^2/9$	1	$(1-r_1)^2 r_2^2/r_{12}^2$	$(1-r_1)^2 r_2^2/r_{12}^2$
	3	$\frac{A_iQ_kB_j}{A_jQ_iB_i}, \frac{A_iQ_jB_j}{A_jQ_kB_i}$	$4(1-\alpha)r_1^2r_2(1-r_2)/27$	2, 2	$4r_1^2r_2(1-r_2)/3r_{12}^2$	$r_1^2r_2(1-r_2)/3r_{12}^2$
		$\frac{A_iQ_kB_j}{A_jQ_jB_i}, \frac{A_iQ_iB_j}{A_jQ_kB_i}$	$4(1-\alpha)r_1(1-r_1)r_2^2/27$	2, 2	$4r_1(1-r_1)r_2^2/3r_{12}^2$	$r_1(1-r_1)r_2^2/3r_{12}^2$
	4	$\frac{A_iQ_kB_j}{A_jQ_kB_i}, \frac{A_iQ_kB_j}{A_jQ_lB_i}$	$4(1-\alpha)r_1^2r_2^2/81$	2, 2	$4r_1^2r_2^2/9r_{12}^2$	$r_1^2r_2^2/9r_{12}^2$
Total		$(1-\alpha)r_{12}^2/9$				

$\frac{A_iB_j}{A_jB_k}$						
	0	-	-	-	-	-
	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_j}{A_jQ_kB_k}$	$4(1-\alpha)r_1^2(1-r_2)^2/9$	1	$r_1^2(1-r_2)^2/r_{12}^2$	$r_1^2(1-r_2)^2/r_{12}^2$

		$\frac{A_iQ_jB_j}{A_jQ_jB_k}$, $\frac{A_iQ_iB_j}{A_jQ_kB_k}$	$\frac{8}{9}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	1, 1	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$
		$\frac{A_iQ_iB_j}{A_jQ_jB_k}$	$4(1-\alpha)(1-r_1)^2r_2^2/9$	1	$(1-r_1)^2r_2^2/r_{12}^2$	$(1-r_1)^2r_2^2/r_{12}^2$
	3	$\frac{A_iQ_kB_j}{A_jQ_kB_k}$, $\frac{A_iQ_lB_j}{A_jQ_kB_k}$, $\frac{A_iQ_jB_j}{A_jQ_iB_k}$, $\frac{A_iQ_jB_j}{A_jQ_lB_k}$	$16(1-\alpha)r_1^2r_2(1-r_2)/27$	1, 1, 1, 1	$4r_1^2r_2(1-r_2)/3r_{12}^2$	$r_1^2r_2(1-r_2)/3r_{12}^2$
		$\frac{A_iQ_kB_j}{A_jQ_jB_k}$, $\frac{A_iQ_lB_j}{A_jQ_jB_k}$, $\frac{A_iQ_iB_j}{A_jQ_iB_k}$, $\frac{A_iQ_iB_j}{A_jQ_lB_k}$	$16(1-\alpha)r_1(1-r_1)r_2^2/27$	1, 1, 1, 1	$4r_1(1-r_1)r_2^2/3r_{12}^2$	$r_1(1-r_1)r_2^2/3r_{12}^2$
	4	$\frac{A_iQ_kB_j}{A_jQ_iB_k}$, $\frac{A_iQ_kB_j}{A_jQ_lB_k}$,	$16(1-\alpha)r_1^2r_2^2/81$	1, 1, 1, 1	$4r_1^2r_2^2/9r_{12}^2$	$r_1^2r_2^2/9r_{12}^2$

		$\frac{A_iQ_lB_j}{A_jQ_iB_k},\frac{A_iQ_lB_j}{A_jQ_lB_k}$				
Total	$4(1-\alpha)r_{12}^2/9$					

$\frac{A_iB_j}{A_kB_l}$	0	-	-	-	-	-
	1	-	-	-	-	-
	2	$\frac{A_iQ_jB_j}{A_kQ_lB_l}$	$2(1-\alpha)r_1^2(1-r_2)^2/9$	1	$r_1^2(1-r_2)^2/r_{12}^2$	$r_1^2(1-r_2)^2/r_{12}^2$
		$\frac{A_iQ_jB_j}{A_kQ_kB_l},\frac{A_iQ_iB_j}{A_kQ_lB_l}$	$\frac{4}{9}(1-\alpha)r_1(1-r_1)r_2(1-r_2)$	1, 1	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$
		$\frac{A_iQ_iB_j}{A_kQ_kB_l}$	$2(1-\alpha)(1-r_1)^2r_2^2/9$	1	$(1-r_1)^2r_2^2/r_{12}^2$	$(1-r_1)^2r_2^2/r_{12}^2$
	3	$\frac{A_iQ_kB_j}{A_kQ_lB_l},\frac{A_iQ_lB_j}{A_kQ_lB_l},$ $\frac{A_iQ_jB_j}{A_kQ_iB_l},\frac{A_iQ_jB_j}{A_kQ_jB_l}$	$8(1-\alpha)r_1^2r_2(1-r_2)/27$	1, 1, 1, 1	$4r_1^2r_2(1-r_2)/3r_{12}^2$	$r_1^2r_2(1-r_2)/3r_{12}^2$
		$\frac{A_iQ_lB_j}{A_kQ_kB_l},\frac{A_iQ_kB_j}{A_kQ_kB_l},$	$8(1-\alpha)r_1(1-r_1)r_2^2/27$	1, 1, 1, 1	$4r_1(1-r_1)r_2^2/3r_{12}^2$	$r_1(1-r_1)r_2^2/3r_{12}^2$

		$\frac{A_iQ_iB_j}{A_kQ_jB_l}, \frac{A_iQ_iB_j}{A_kQ_iB_l}$				
	4	$\frac{A_iQ_kB_j}{A_kQ_iB_l}, \frac{A_iQ_kB_j}{A_kQ_jB_l}, \frac{A_iQ_lB_j}{A_kQ_iB_l}, \frac{A_iQ_lB_j}{A_kQ_jB_l}$	$8(1-\alpha)r_1^2r_2^2/81$	1, 1, 1, 1	$4r_1^2r_2^2/9r_{12}^2$	$r_1^2r_2^2/9r_{12}^2$
Total		$2(1-\alpha)r_{12}^2/9$				

2.4.5. Method with unknown parental genotypes on QTL

In practice, it is usually that the parental genotypes of QTL and the linkage phase between a putative QTL and markers can't be observed directly. For simplicity, but without loss of generality, QTL genotype of an autotetraploid individual can be presented by $Q_1Q_2Q_3Q_4$. In the biallelic model here, QTL on each chromosmome may carry allele *Q* or *q*. Since there are four homologous chromosomes in an autotetraploid individual, there are up to 2^4 QTL genotypes for each autotetraploid parent, taking linkage phases into consideration (i.e., $Q_i(i=1,2,3,4)$ may be *Q* or *q*). By crosses between two autotetraploid parents, P_1 and P_2, there will be up to 256 ($2^4\times2^4$) possible combinations of parental genotypes at QTL. Since QTL genotypes are expected to segregated in the first generation of population derived by crossing two autotetraploid parents, it is impossible to be homozygotes at QTL for both parents. Thus four crosses between two parents should be excluded, detailed as: $QQQQ\times QQQQ$, $QQQQ\times qqqq$, $qqqq\times QQQQ$ and $qqqq\times qqqq$. Among the remaining 252 possible parental QTL genotype combinations, I use computer-intensive search method to find the most likely parental genotypes with the maximum LOD score at location z on the chromosome, by repeating the interval mapping method developed in Section 2.4.3. The maxium LOD score with predicted parental genotypes will be chosen as the final LOD score at location z.

2.4.6. Estimation of parameters of genetic effects

In Section 2.4.3 I have discussed the method to obtain the MLEs of genotypic values of QTL. In this section, I will discuss how to estimate genetic effects to give some insight into QTL. For the one locus biallic model, the genotypic value can be expressed as

$$G_j = \mu + \theta_1\omega_{1j} + \theta_2\omega_{2j} + \theta_3\omega_{3j} + \theta_4\omega_{4j} \qquad (j = 0,1,...,4) \tag{I-2.25}$$

where μ is the population mean, and θ_i (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL. ω_{ij} are the correspding orthogonal scales, which are determined by the theoretical probability distribution of QTL genotypes given parental genotypes.

Let $f_k\left(k = 0,1,...,4\right)$ denote frequency of QTL with genotype $Q_k q_{4-k}$. A general presentation for an autotetraploid genotype at QTL can be $Q_1Q_2Q_3Q_4$, with allele Q_i $\left(i = 1,...,4\right)$ can be *Q* or *q*.

Under bivalent pairing, there are six different gametes produced by the autotetraploid parent and each gamete $Q_iQ_j\left(1 \le i, j \le 4; i \ne j\right)$ is generated with equal probability of 1/6. By random union between gametes, there are 36 different zygote with equal probability of 1/36. Then f_k can be calculated by sorting these zygotes according to their genotypes and summing over the probabilities of zygote with the same genotype.

Under quadrivalent pairing, there are ten different gametes produced by the autotetraploid parent, including four double reduction gametes, $Q_iQ_i(1 \le i \le 4)$, and six non-double reduction gametes, $Q_iQ_j(1 \le i, j \le 4; i \ne j)$. The coefficient of double reduction rate at QTL, α_{QTL}, can be expressed in term of a function of the coefficient of double reduction at the flanking marker which is nearer to centromere, α, and recombination frequency between QTL and the marker as (Luo et al. 2004):

$$\alpha_{QTL} = \left[\alpha(3-4r)^2 + 2r(3-2r)\right]/9 \qquad \text{(I-2.26)}$$

Thus the probability of each double reducntion gamete equals to $\alpha_{QTL}/4$ and the probability of each non-double reduction gamete equals to $(1-\alpha_{QTL})/6$. Assuming random union of these gametes, the 100 zygotes can be sorted into three different categories according to the number of double reduction gametes involved and their probabilities are given by

$$\begin{cases} (1-\alpha_{QTL})^2/36 & \textit{if no double reduction gametes involved} \\ \alpha_{QTL}(1-\alpha_{QTL})/24 & \textit{if no double reduction gametes involved} \\ {\alpha_{QTL}}^2/16 & \textit{if no double reduction gametes involved} \end{cases}$$

Then f_k can be calculated by sorting these zygotes according to their genotypes and summing over the probabilities of zygote with the same genotype.

Then orthogonal scales ω_{ij} can be worked out follow the method developed in Section 1.3.1 of Chapter I-1. The genetic effects can be then calculated by

$$\begin{bmatrix} \mu^* \\ \theta_1^* \\ \theta_2^* \\ \theta_3^* \\ \theta_4^* \end{bmatrix} = \begin{bmatrix} 1 & \omega_{10} & \omega_{20} & \omega_{30} & \omega_{40} \\ 1 & \omega_{11} & \omega_{21} & \omega_{31} & \omega_{41} \\ 1 & \omega_{12} & \omega_{22} & \omega_{32} & \omega_{42} \\ 1 & \omega_{13} & \omega_{23} & \omega_{33} & \omega_{43} \\ 1 & \omega_{14} & \omega_{24} & \omega_{34} & \omega_{44} \end{bmatrix}^{-1} \begin{bmatrix} G_0^* \\ G_1^* \\ G_2^* \\ G_3^* \\ G_4^* \end{bmatrix} \quad \text{(I-2.27)}$$

where $G_k^*\left(k=0,1,...,4\right)$ are MLEs of genotypic values of QTL, μ^* is MLE of the population mean, and θ_i^* (i = 1, …, 4) are accordingly MLEs of monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL.

2.5. Simulation study

A simulation study was implemented to investigate this approach of QTL mapping for autotetraploid species. The present simulation study considered a linkage group of 14 marker loci and a QTL located on a simulated chromosome. The simulation programs produced marker phenotypes and trait value from a full-sib family of individuals generated by crossing two genetically unrelated parental autotetraploids under tetrasomic inheritance with bivalent pairing and quadrivalent pairing, respectively. The simulated parental markers and QTL are listed in Table I-2.6.

Table I-2.6. Simulation parameters of the coefficient of double reduction at and recombination frequencies between 15 linked marker loci and QTL and parental genotypes used to simulate the mapping populations

Locus	α quadrivalent	r	$d_{bivalent}$ (cM)	$d_{quadrivalent}$ (cM)	Parental genotype	
					P_1	P_2
L_1	0.050	0.00	0	0	$M_3M_1M_2M_2$	$M_4M_3M_5M_0$
L_2	0.100	0.10	11.16	10.73	$M_3M_1M_2M_1$	$M_2M_3M_3M_1$
L_3	0.137	0.10	22.31	21.47	$M_2M_3M_1M_5$	$M_1M_3M_1M_2$
L_4(QTL)	0.152	0.05	27.58	26.64	$q\ q\ Q\ Q$	$q\ q\ Q\ Q$
L_5	0.164	0.05	32.85	31.81	$M_0M_2M_3M_1$	$M_1M_1M_2M_4$
L_6	0.175	0.05	38.12	36.99	$M_2M_1M_1M_0$	$M_3M_4M_3M_1$
L_7	0.185	0.05	43.39	42.16	$M_4M_0M_1M_5$	$M_1M_2M_1M_2$
L_8	0.201	0.10	54.54	52.90	$M_2M_0M_1M_1$	$M_4M_1M_2M_2$
L_9	0.207	0.05	59.81	58.07	$M_2M_2M_4M_2$	$M_1M_2M_1M_4$
L_{10}	0.218	0.10	70.97	68.80	$M_4M_4M_2M_5$	$M_2M_2M_1M_4$
L_{11}	0.222	0.05	76.24	73.98	$M_1M_5M_4M_5$	$M_4M_1M_3M_1$
L_{12}	0.229	0.10	87.39	84.71	$M_3M_4M_5M_5$	$M_5M_2M_3M_3$
L_{13}	0.234	0.10	98.55	95.44	$M_2M_1M_3M_3$	$M_5M_4M_1M_1$
L_{14}	0.236	0.05	103.82	100.62	$M_4M_1M_4M_2$	$M_2M_3M_3M_1$
L_{15}	0.240	0.10	114.98	111.35	$M_3M_5M_5M_1$	$M_1M_1M_2M_4$

Simulation parameters of the coefficient of double reduction (α for quadrivalent pairing) at and recombination frequencies (r) between 15 linked marker loci and QTL and parental genotypes at markers and QTL. M_i(i=1,…,5) represent five distinct alleles from two parent and M_0 represent the null allele. Alleles listed in the same column are on the same chromosome.

For the QTL, the genetic parameters of genetic mean (μ), monogenic (θ_1), digenic (θ_2), trigenic (θ_3) and quadrigenic effects (θ_4) are assumed to be 500, 100, 60, 30 and 10, respectively. The mapping population size is 300 (mapping population size usually generated in practice) and heritability is 0.1 (with a low heritability to test the reliability of this method). Under the model parameters setting in Table I-2.6, the genetics model developed in Chapter I-1 can be carried out as fallows.

In the full-sib family created from crossing two parental lines with genotypes *QQqq* and *QQqq*. For bivalent pairing, frequencies of the offspring genotypes *qqqq*, *Qqqq*, *QQqq*, *QQQq* and *QQQQ* are 1/36, 2/9, 1/2, 2/9, and 1/36, respectively. For quadrivalent pairing, frequencies of the offspring genotypes can be expressed in term of α, the coefficient of double reduction at the QTL, as $f_0 = (1+2\alpha)^2/36$, $f_1 = 2(1-\alpha)(1+2\alpha)/9$, $f_2 = [3-4\alpha(1-\alpha)]/6$, $f_3 = 2(1-\alpha)(1+2\alpha)/9$ and $f_4 = (1+2\alpha)^2/36$. With these, the genotypic values $G_{bivalent} = (G_4\ G_3\ G_2\ G_1\ G_0)^T$ and $G_{quadrivalent} = (G_4\ G_3\ G_2\ G_1\ G_0)^T$ can be presented in a matrix form of

$$G_{bivalent} = S_b E_b = \begin{bmatrix} 1 & 2 & 5/3 & 2/3 & 1/8 \\ 1 & 1 & 1/6 & -1/6 & -1/16 \\ 1 & 0 & -1/3 & 0 & 1/24 \\ 1 & -1 & 1/6 & 1/6 & -1/16 \\ 1 & -2 & 5/3 & -2/3 & 1/8 \end{bmatrix} \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \quad \text{(I-2.28)}$$

and

$$G_{quadrivalent} = S_q E_q = \begin{bmatrix} 1 & 2 & (5-2\alpha)/3 & 2(1-\alpha)/3 & (1-\alpha)(4\alpha^2-4\alpha+3)/(12(2+\alpha)) \\ 1 & 1 & (1-4\alpha)/6 & -(1+2\alpha)/6 & -(1+2\alpha)(4\alpha^2-4\alpha+3)/(24(2+\alpha)) \\ 1 & 0 & -(1+2\alpha)/3 & 0 & (1-\alpha)(2\alpha+1)^2/(12(2+\alpha)) \\ 1 & -1 & (1-4\alpha)/6 & (1+2\alpha)/6 & -(1+2\alpha)(4\alpha^2-4\alpha+3)/(24(2+\alpha)) \\ 1 & -2 & (5-2\alpha)/3 & -2(1-\alpha)/3 & (1-\alpha)(4\alpha^2-4\alpha+3)/(12(2+\alpha)) \end{bmatrix} \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

(I-2.29)

The genetic values can be calculated from $E_b = S_b^{-1} G_{bivalent}$ and $E_q = S_q^{-1} G_{bivalent}$ where

$$S_b^{-1} = \begin{bmatrix} 1/36 & 2/9 & 1/2 & 2/9 & 1/36 \\ 1/12 & 1/3 & 0 & -1/3 & -1/12 \\ \frac{5}{24} & \frac{1}{6} & -\frac{3}{4} & \frac{1}{6} & \frac{5}{24} \\ 1/2 & -1 & 0 & 1 & -1/2 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

(I-2.30)

and

$$S_q^{-1} = \begin{bmatrix} (1+2\alpha)^2/36 & 2(1+2\alpha)(1-\alpha)/9 & (4\alpha^2-4\alpha+3)/6 & 2(1+2\alpha)(1-\alpha)/9 & (1+2\alpha)^2/36 \\ (1+2\alpha)/12 & (1-\alpha)/3 & 0 & (\alpha-1)/3 & -(1+2\alpha)/12 \\ \frac{(1+2\alpha)(5-2\alpha)}{12(2+\alpha)} & \frac{(\alpha-1)(4\alpha-1)}{3(2+\alpha)} & -\frac{(4\alpha^2-4\alpha+3)}{2(2+\alpha)} & \frac{(\alpha-1)(4\alpha-1)}{3(2+\alpha)} & \frac{(1+2\alpha)(5-2\alpha)}{12(2+\alpha)} \\ 1/2 & -1 & 0 & 1 & -1/2 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

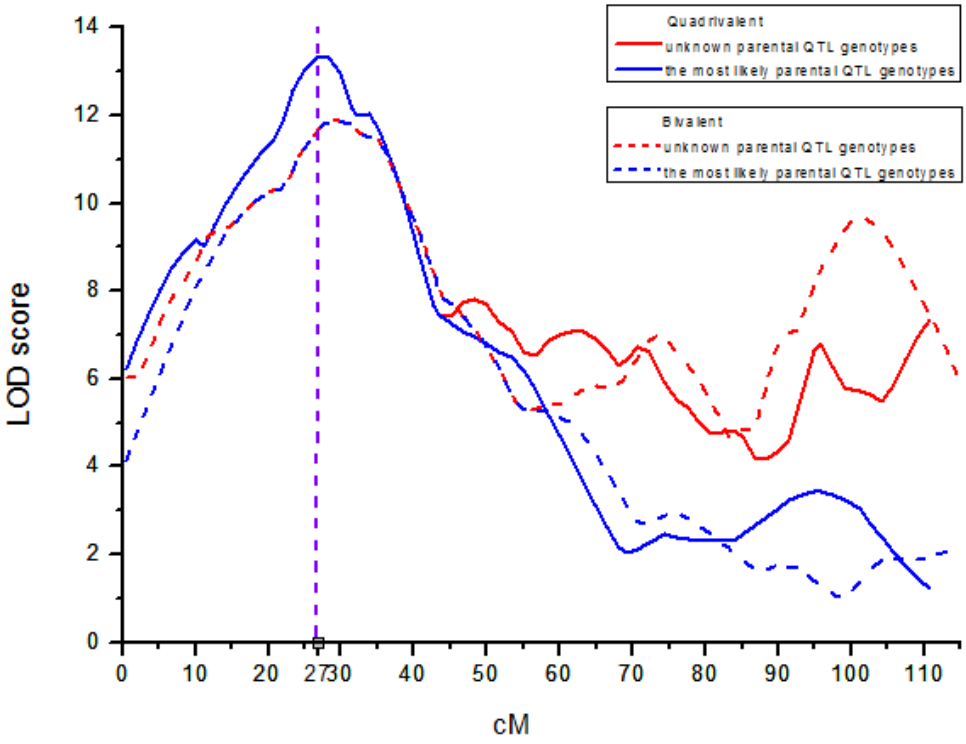
(I-2.31)

for bivalent pairing and quadrivalent pairing, respectively.

Then the trait values for offspring individuals were carried out as genotypic values plus an environmental effect sampled from a normal distribution of $N\left(0,\sigma^2\right)$. The value of the residual variance σ^2 was calculated by $\sigma^2=\sigma_G^2/h^2-\sigma_G^2$, where h^2 is the desired heritability and σ_G^2 is the genetic variance. Based on the simulated parameters setting in Table I-2.6, we obtained a profile of LOD scores along the chromosome for two pairing pattern. Figure I-2.4 shows such a profile with population size of 300 and a heritability of 10%. The higher LOD score indicates the more likely QTL existing on the position along the scanned chromosome. The solid lines and dotted lines represent the profile of LOD scores estimated under bivalent and quadrivalent pairing, respectively. The blue lines indicate that statistical estimation is carried out without know of parental QTL genotypes and the red lines are obtained by re-estimating with the most likely parental QTL genotypes. We took the maximum of LOD scores as the most likely location of the QTL. When LOD scores were estimated with all possible parental QTL genotypes, we can see from Figure I-2.4 that there are other peaks (except the peak indicating the existing of true QTL) along the curves of LOD scores (indicated by red lines). However, it can be seen that the existence of the ghost QTL could be removed after the profile of LOD scores were re-estimated with the most likely parental QTL genotypes (indicated by bule lines).

To detect a QTL, we need first to set a threshold for LOD scores, above which we declare the presence of a QTL. Simulations were run with a similar pattern to simulation setting in Table I-2.6, but the phenotypic values of the trait were randomly shuffled among the offspring individuals. In this way, we can see how the LOD scores distribute under the null hypothesis of no QTL existing on the chromosome. We would declare that a QTL exist on the scanned location significantly (e.g. significance level is 0.05) when the corresponding observed LOD

Figure I-2.4. Profile of LOD scores along the chromosome.



The red curve is the profile of LOD scores estimated with unknown parental QTL genotypes and the blue curve is the profile of LOD scores re-estimated with the most likely parental QTL genotypes. The solid lines represent quadrivalent pairing and the dotted lines represent bivalent pairing. The purple straight line indicates the position of QTL. Population size is 300 and heritability is 0.1.

scores exceeds the 95% point of the LOD score distribution under the null hypothesis. Under the simulation model in Table I-2.6, the threshold for LOD scores to declare a QTL is 3.79 (SE 0.15) and 3.50 (SE 0.16) under bivalent pairing and quadrivalent pairing based on 100 replicates.

Table I-2.7 summarizes the estimates of genotypic values and residual vriance under both bivalent pairing and quadrivalent pairing with the parameters setting in Table I-2.6. For each study, the mean and standard errors and presented over 100 replicates. For a heritability, all the data set with or without known of parental QTL genotypes had LOD scores greater than the threshold of 3.79 under bivalent pairing and 3.50 under quadrivalent pairing. Thus the power of QTL detection was 1.00 in current simulation studies under both bivalent pairing model and quadrivalent pairing model. The row labelled $Q_{genotype}$ is the proportion of correct prediction of the parental QTL genotypes. We can see that the parental QTL genotypes and and linkage phases have been correctly predicted in nearly half of these simulations under bivalent and quadrivalent pairing. The proportion was a little bit higher in the bivalent pairing model than that in quadrivalent pairing model. To investigate the mapping accuracy, I calculated the mean distance between the estimated most likely QTL location and the true QTL location in the row labled "Accuracy (CM)" in Table I-2.7. From the result, we can see both bivalent and quadrivalent methods predicted QTL location adequately with or without knowing parental QTL genotypes. Comparing with analysis with known parental QTL genotypes, the mapping accuracy and parameters estimation that without known of parental QTL genotypes are comparatively poorer but still in an acceptable range (i.e. average estimated location is within 5 cM away from the true QTL location). In addition, the proportions of the simulations indicating a QTL within 10 cM away from the true QTL location are also shown in Table I-2.7. From these simulation studies, around 80% of predicted QTL positions were located within 10 cM away from the true QTL position with a small population size of 300 and a low heritability of 10%. This proportion in analysis with unknown parental QTL genotypes did not significantly decreased compared with analysis with known parental QTL genotypes. Again, this

Table I-2.7. Results of simulation studies for QTL mapping in autotetraploids

Parameters	True value	Bivalent pairing		True value	Quadrivalent pairing	
		Known parental QTL genotypes	Unknown parental QTL genotypes		Known parental QTL genotypes	Unknown parental QTL genotypes
G_1	381.25	374.33 (7.96)	525.88 (22.14)	377.78	382.80 (8.70)	485.24 (23.68)
G_2	414.38	423.32 (3.75)	491.82 (11.36)	409.82	417.32 (3.82)	466.82 (11.52)
G_3	480.42	480.73 (2.18)	480.73 (2.89)	474.49	475.85 (2.31)	484.60 (4.54)
G_4	604.38	600.16 (3.45)	529.88 (8.93)	596.78	590.77 (3.82)	537.86 (8.84)
G_5	821.25	797.80 (13.19)	646.90 (23.58)	811.71	782.01 (8.00)	656.94 (24.04)
σ	259.81	259.27 (1.14)	258.89 (1.13)	297.89	296.72 (1.23)	295.76 (1.25)
Detection power	-	1.00	1.00	-	1.00	1.00
$Q_{genotype}$	-	-	0.48	-	-	0.42
Accuracy (cM)	-	1.21 (0.94)	3.89 (1.50)	-	2.11(1.46)	4.48 (1.83)
Proportion in (±10 cM)	-	0.82	0.80	-	0.79	0.75

The heritability is 10% for simulation. Sample size of each mapping population is 300. G_i (i=1,…,5) represents genotypic value for QTL with genotype $Q_{(i-1)}q_{(5-i)}$. σ is the normal residual variant. The simulation replicates is 100.

proportion was found slightly higher in bivalent pairing model than that in quadrivalent pairing model.

From Table I-2.7, we can see that QTL mapping performance was comparatively better in bivalent pairing model than that in quadrivalent pairing model. However, in practice autotetraploids undergo tetrasomic inheritance in which homologous chromosomes segregate either in bivalent, quadrivalent pairing or a mixture of the two during meiosis, which means neither bivalent pairing method nor quadrivalent pairing method probably may not perform well in fitting experimental data. To investigate the robustness of different models, namely bivalent pairing model and quadrivalent pairing model, for fitting data generated under different pairing patterns, I simulated data with both bivalent and quadrivalent pairing and each was analysed using both bivalent method and quadrivalent method with known of parental QTL genotypes. For bivalent pairing data, the coefficient of double reduction was assumed to be zero on all the marker loci when using quadrivalent method. For quadrivalent pairing data, bivalent method was applied after screening data that were compatible with bivalent pairing, which would reduce sample size to about one half in current data sets as shown in Table I-2.8. Table I-2.8 summarizes the estimates of genotypic values, residual variance, mapping accuracy. From the result we can see that performance of quadrivalent method was still good when analysing bivalent data, with estimation result similar to that analysed with bivalent method. On the contrary, bivalent method did poorly in modelling quadrivalent data. First, bivalent method has to discard individuals uncompatible with bivalent pairing due to the occurrence of double reduction under quadrivalent pairing. And then QTL detection power and mapping accuracy were significantly decreased in quadrivalent datasets using bivalent pairing method.

Table I-2.8. Comparison of two pairing pattern methods with two different datasets

	True value		Bivalent method Estimates	Quadrivalent method Estimates
Bivalent data	G_1	381.25	374.33 (7.96)	377.32 (7.37)
	G_2	414.38	423.32 (3.75)	425.19 (3.80)
	G_3	480.42	480.73 (2.18)	480.72 (2.12)
	G_4	604.38	600.16 (3.45)	596.72 (3.36)
	G_5	821.25	797.80 (13.19)	789.73 (12.17)
	σ	259.81	259.27 (1.14)	259.78 (1.12)
	Detection power		1.00	1.00
	Accuracy (cM)		1.21 (0.94)	1.80 (0.97)
	Proportion in (±10 cM)		0.82	0.79
Quadrivalent data	G_1	377.78	356.14 (18.28)	382.80 (8.70)
	G_2	409.82	414.11 (6.99)	417.32 (3.82)
	G_3	474.49	484.11 (4.56)	475.85 (2.31)
	G_4	596.78	606.36 (8.23)	590.77 (3.82)
	G_5	811.71	761.31 (24.18)	782.01 (8.00)
	σ	297.89	289.85 (1.94)	296.72 (1.23)
	Proportion of discarded (%)		56.99 (0.26)	-
	Detection power		0.91	1.00
	Accuracy (cM)		10.44 (2.61)	2.11(1.46)
	Proportion in (±10 cM)		0.45	0.79

The heritability is 10% for simulation. Sample size of each mapping population is 300. G_i (i=1,…,5) represents genotypic value for QTL with genotype $Q_{(i-1)}q_{(5-i)}$. σ is the normal residual variant. The simulation replicates is 100. The row labled with proportion of discarded shows the mean and s.e. of proportion of data has been discarded when ‘bivalent method’ was used to analyse ‘quadrivalent data’.

To further investigate the reliability of this method in practical implementation, I simulated twelve linkage groups (i.e. for the twelve sets of chromosomes) of marker loci and two QTLs in an autotetraploid potato genome. The simulation programs produced marker phenotypes and trait value from a full-sib family of individuals generated by crossing two genetically unrelated parental autotetraploids under tetrasomic inheritance with quadrivalent pairing. The simulated parental markers in the twelve linkage groups and two QTLs (i.e. located on Chromosome 1 and 3) are listed in Table I-2.9.

Table I-2.9. Simulation parameters of the coefficient of double reduction at and recombination frequencies between the linked marker loci within the twelve linkage groups and two QTLs and parental genotypes used to simulate the whole autotetraploid potato genome for the mapping populations

Chromosome 1					
Locus	α quadrivalent	r	$d_{quadrivalent}$ (cM)	Parental genotype	
				P_1	P_2
L_1	0.050	0.00	0	$M_3M_1M_2M_2$	$M_4M_3M_5M_0$
L_2	0.100	0.10	10.73	$M_3M_1M_2M_1$	$M_2M_3M_3M_1$
L_3	0.137	0.10	21.47	$M_2M_3M_1M_5$	$M_1M_3M_1M_2$
L_4(QTL)	0.152	0.05	26.64	$q\ q\ Q\ Q$	$q\ q\ Q\ Q$
L_5	0.164	0.05	31.81	$M_0M_2M_3M_1$	$M_1M_1M_2M_4$
L_6	0.175	0.05	36.99	$M_2M_1M_1M_0$	$M_3M_4M_3M_1$
L_7	0.185	0.05	42.16	$M_4M_0M_1M_5$	$M_1M_2M_1M_2$
L_8	0.201	0.10	52.90	$M_2M_0M_1M_1$	$M_4M_1M_2M_2$

L_9	0.207	0.05	58.07	$M_2M_2M_4M_2$	$M_1M_2M_1M_4$
L_{10}	0.218	0.10	68.80	$M_4M_4M_2M_5$	$M_2M_2M_1M_4$
L_{11}	0.222	0.05	73.98	$M_1M_5M_4M_5$	$M_4M_1M_3M_1$
L_{12}	0.229	0.10	84.71	$M_3M_4M_5M_5$	$M_5M_2M_3M_3$
L_{13}	0.234	0.10	95.44	$M_2M_1M_3M_3$	$M_5M_4M_1M_1$
L_{14}	0.236	0.05	100.62	$M_4M_1M_4M_2$	$M_2M_3M_3M_1$
L_{15}	0.240	0.10	111.35	$M_3M_5M_5M_1$	$M_1M_1M_2M_4$
Chromosome 2					
L_1	0.010	0.00	0.00	$M_4M_1M_2M_3$	$M_0M_5M_2M_2$
L_2	0.041	0.05	5.17	$M_5M_2M_0M_1$	$M_1M_2M_3M_4$
L_3	0.093	0.10	15.91	$M_2M_2M_1M_3$	$M_4M_5M_1M_1$
L_4	0.132	0.10	26.64	$M_2M_3M_1M_2$	$M_1M_2M_4M_4$
L_5	0.147	0.05	31.81	$M_4M_4M_3M_1$	$M_2M_2M_5M_6$
L_6	0.160	0.05	36.99	$M_1M_1M_2M_3$	$M_2M_1M_4M_4$
L_7	0.183	0.10	47.72	$M_4M_1M_1M_3$	$M_2M_2M_3M_1$
L_8	0.200	0.10	58.45	$M_1M_2M_2M_3$	$M_2M_3M_4M_5$
L_9	0.212	0.10	69.19	$M_1M_2M_4M_3$	$M_3M_3M_2M_1$
L_{10}	0.217	0.05	74.36	$M_5M_1M_1M_2$	$M_3M_4M_3M_4$
L_{11}	0.225	0.10	85.09	$M_2M_2M_4M_3$	$M_1M_1M_2M_4$
Chromosome 3					
L_1	0.020	0.00	0.00	$M_1M_1M_2M_3$	$M_2M_3M_4M_4$
L_2	0.050	0.05	5.17	$M_2M_2M_1M_3$	$M_1M_1M_2M_4$
L_3	0.100	0.10	15.91	$M_2M_3M_1M_1$	$M_1M_5M_2M_3$
L_4	0.137	0.10	26.64	$M_1M_5M_2M_2$	$M_3M_4M_1M_1$
L_5	0.152	0.05	31.81	$M_2M_4M_3M_3$	$M_1M_1M_2M_3$
L_6	0.176	0.10	42.55	$M_1M_2M_4M_4$	$M_3M_3M_2M_4$
L_7	0.194	0.10	53.28	$M_1M_3M_1M_2$	$M_2M_4M_5M_1$
L_8	0.202	0.05	58.45	$M_2M_2M_3M_4$	$M_1M_1M_2M_5$

L_9	0.208	0.05	63.63	$M_1M_2M_1M_3$	$M_2M_3M_4M_4$
L_{10}	0.218	0.10	74.36	$M_3M_3M_2M_1$	$M_3M_4M_5M_1$
L_{11}	0.226	0.10	85.09	$M_1M_2M_5M_6$	$M_2M_3M_4M_4$
L_{12}	0.232	0.10	95.83	$M_2M_3M_4M_4$	$M_4M_1M_2M_2$
L_{13}(QTL)	0.234	0.05	101.00	$q\ q\ Q\ Q$	$q\ q\ Q\ Q$
L_{14}	0.236	0.05	106.17	$M_1M_3M_1M_2$	$M_2M_4M_5M_1$
L_{15}	0.240	0.10	116.91	$M_2M_2M_3M_4$	$M_1M_1M_2M_5$
L_{16}	0.242	0.10	127.64	$M_1M_2M_1M_3$	$M_2M_3M_4M_4$
L_{17}	0.243	0.05	132.81	$M_3M_3M_2M_1$	$M_3M_4M_5M_1$
L_{18}	0.245	0.10	143.55	$M_1M_2M_5M_6$	$M_2M_3M_4M_4$
L_{19}	0.246	0.05	148.72	$M_2M_3M_4M_4$	$M_4M_1M_2M_2$
Chromosome 4					
L_1	0.040	0.00	0.00	$M_1M_1M_2M_3$	$M_2M_3M_4M_1$
L_2	0.067	0.05	5.17	$M_2M_3M_1M_1$	$M_3M_2M_2M_4$
L_3	0.113	0.10	15.91	$M_2M_4M_5M_5$	$M_1M_2M_3M_3$
L_4	0.147	0.10	26.64	$M_1M_3M_1M_2$	$M_1M_3M_1M_4$
L_5	0.160	0.05	31.81	$M_5M_2M_3M_0$	$M_1M_2M_4M_4$
L_6	0.172	0.05	36.99	$M_3M_2M_2M_1$	$M_2M_3M_3M_4$
L_7	0.191	0.10	47.72	$M_1M_2M_5M_5$	$M_1M_3M_3M_4$
L_8	0.206	0.10	58.45	$M_1M_1M_2M_3$	$M_2M_2M_3M_4$
L_9	0.217	0.10	69.19	$M_2M_3M_4M_5$	$M_1M_1M_2M_4$
L_{10}	0.221	0.05	74.36	$M_1M_1M_3M_4$	$M_2M_3M_3M_{40.05}$
Chromosome 5					
L_1	0.050	0.00	0.00	$M_1M_1M_2M_3$	$M_2M_3M_2M_4$
L_2	0.076	0.05	5.17	$M_2M_4M_1M_1$	$M_2M_3M_4M_5$
L_3	0.119	0.10	15.91	$M_1M_1M_1M_2$	$M_3M_2M_2M_4$
L_4	0.152	0.10	26.64	$M_1M_2M_2M_3$	$M_2M_4M_4M_3$
L_5	0.176	0.10	37.37	$M_1M_5M_1M_2$	$M_2M_3M_3M_4$

L_6	0.186	0.05	42.55	$M_1M_2M_1M_5$	$M_2M_3M_2M_4$
L_7	0.194	0.05	47.72	$M_2M_3M_3M_1$	$M_1M_4M_2M_5$
L_8	0.208	0.10	58.45	$M_3M_2M_2M_1$	$M_1M_2M_4M_4$
L_9	0.213	0.05	63.63	$M_4M_5M_1M_1$	$M_2M_3M_3M_3$
L_{10}	0.222	0.10	74.36	$M_1M_2M_3M_3$	$M_2M_3M_4M_4$
L_{11}	0.226	0.05	79.36	$M_2M_1M_2M_5$	$M_1M_3M_3M_4$
L_{12}	0.232	0.10	90.27	$M_1M_2M_1M_3$	$M_1M_1M_3M_4$
Chromosome 6					
L_1	0.030	0.00	0.00	$M_2M_2M_1M_3$	$M_0M_1M_3M_4$
L_2	0.085	0.10	10.73	$M_1M_2M_3M_3$	$M_2M_3M_1M_4$
L_3	0.126	0.10	21.47	$M_1M_1M_3M_2$	$M_2M_2M_4M_5$
L_4	0.142	0.05	26.64	$M_2M_3M_3M_1$	$M_1M_2M_1M_3$
L_5	0.169	0.10	37.37	$M_1M_5M_1M_2$	$M_2M_3M_3M_4$
L_6	0.189	0.10	48.10	$M_1M_2M_2M_3$	$M_1M_3M_4M_4$
L_7	0.204	0.10	58.84	$M_1M_2M_1M_2$	$M_3M_4M_1M_2$
L_8	0.210	0.05	64.01	$M_1M_1M_2M_3$	$M_4M_5M_5M_1$
L_9	0.220	0.10	74.74	$M_1M_3M_3M_2$	$M_3M_4M_2M_2$
L_{10}	0.227	0.10	85.48	$M_1M_1M_2M_0$	$M_3M_4M_4M_5$
Chromosome 7					
L_1	0.000	0.00	0.00	$M_3M_4M_4M_5$	$M_0M_1M_1M_2$
L_2	0.032	0.05	5.17	$M_2M_3M_3M_5$	$M_1M_4M_4M_2$
L_3	0.086	0.10	15.91	$M_1M_0M_3M_4$	$M_2M_2M_5M_1$
L_4	0.108	0.05	21.08	$M_3M_3M_2M_1$	$M_1M_1M_4M_5$
L_5	0.126	0.05	26.26	$M_2M_4M_4M_3$	$M_3M_1M_1M_5$
L_6	0.157	0.10	36.99	$M_1M_1M_2M_3$	$M_3M_4M_5M_5$
L_7	0.180	0.10	47.72	$M_1M_3M_4M_5$	$M_2M_2M_1M_3$
L_8	0.189	0.05	52.90	$M_1M_2M_3M_3$	$M_2M_2M_4M_5$
L_9	0.204	0.10	63.63	$M_0M_1M_1M_3$	$M_2M_4M_2M_3$

L_{10}	0.216	0.10	74.36	$M_2M_2M_3M_4$	$M_2M_1M_5M_4$
L_{11}	0.220	0.05	79.54	$M_1M_1M_3M_5$	$M_2M_4M_2M_5$
L_{12}	0.224	0.05	84.71	$M_1M_3M_5M_5$	$M_2M_2M_4M_5$
L_{13}	0.227	0.05	89.88	$M_2M_3M_4M_4$	$M_1M_2M_1M_5$
L_{14}	0.233	0.10	100.62	$M_1M_4M_3M_2$	$M_5M_1M_1M_2$
L_{15}	0.237	0.10	111.35	$M_0M_4M_4M_2$	$M_1M_3M_1M_5$
Chromosome 8					
L_1	0.010	0.00	0.00	$M_3M_1M_3M_2$	$M_4M_5M_1M_2$
L_2	0.041	0.05	5.17	$M_1M_1M_0M_3$	$M_2M_4M_4M_5$
L_3	0.068	0.05	10.35	$M_2M_3M_1M_4$	$M_2M_5M_5M_1$
L_4	0.091	0.05	15.52	$M_0M_1M_2M_2$	$M_2M_3M_5M_4$
L_5	0.131	0.10	26.26	$M_2M_5M_4M_4$	$M_1M_3M_3M_0$
L_6	0.160	0.10	36.99	$M_2M_3M_4M_1$	$M_5M_0M_1M_2$
L_7	0.172	0.05	42.16	$M_3M_3M_4M_5$	$M_2M_1M_1M_5$
L_8	0.182	0.05	47.34	$M_2M_1M_3M_1$	$M_4M_5M_5M_3$
L_9	0.199	0.10	58.07	$M_2M_2M_3M_5$	$M_4M_0M_1M_5$
L_{10}	0.206	0.05	63.24	$M_3M_4M_4M_5$	$M_1M_2M_2M_3$
L_{11}	0.211	0.05	68.42	$M_4M_0M_1M_2$	$M_3M_3M_5M_1$
Chromosome 9					
L_1	0.020	0.00	0.00	$M_1M_1M_2M_3$	$M_4M_5M_2M_1$
L_2	0.077	0.10	10.73	$M_3M_5M_4M_3$	$M_1M_1M_2M_3$
L_3	0.100	0.05	15.91	$M_1M_3M_1M_5$	$M_4M_3M_2M_2$
L_4	0.137	0.10	26.64	$M_0M_5M_5M_2$	$M_1M_3M_4M_4$
L_5	0.165	0.10	37.37	$M_3M_2M_3M_4$	$M_1M_2M_1M_0$
L_6	0.186	0.10	48.10	$M_5M_4M_4M_2$	$M_3M_2M_3M_1$
L_7	0.194	0.05	53.28	$M_1M_1M_3M_4$	$M_1M_2M_2M_5$
L_8	0.208	0.10	64.10	$M_3M_2M_1M_4$	$M_5M_5M_1M_3$
Chromosome 10					

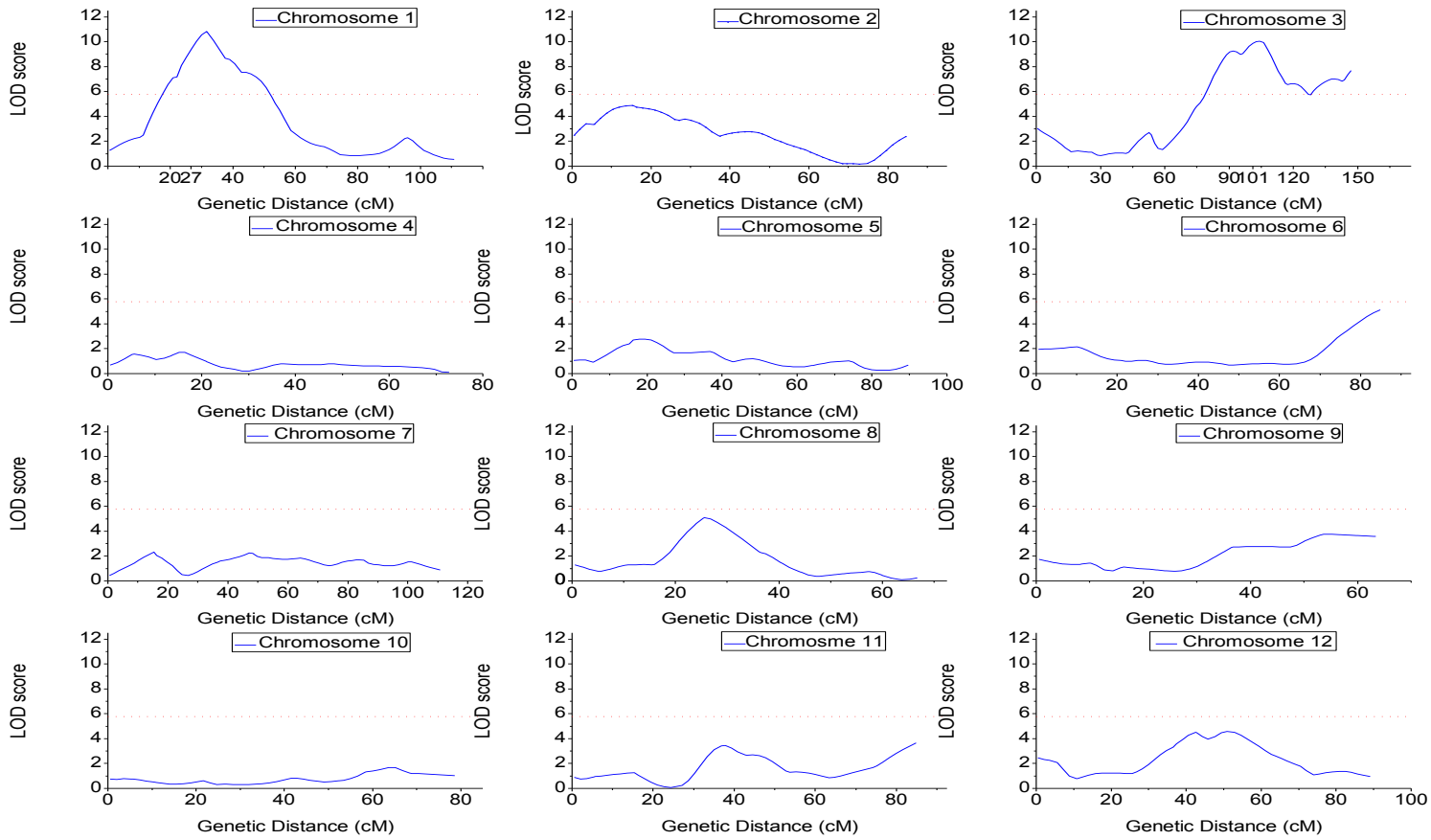
L_1	0.050	0.00	0.00	$M_2M_3M_3M_5$	$M_1M_1M_4M_2$
L_2	0.076	0.05	5.17	$M_1M_0M_3M_4$	$M_2M_2M_5M_1$
L_3	0.098	0.05	10.35	$M_3M_1M_3M_2$	$M_4M_1M_5M_2$
L_4	0.136	0.10	21.08	$M_1M_1M_4M_5$	$M_2M_3M_2M_1$
L_5	0.151	0.05	26.26	$M_0M_4M_1M_2$	$M_3M_5M_1M_2$
L_6	0.163	0.05	31.43	$M_1M_1M_3M_4$	$M_2M_5M_1M_2$
L_7	0.185	0.10	42.16	$M_3M_4M_1M_2$	$M_2M_2M_5M_1$
L_8	0.201	0.10	52.90	$M_3M_3M_4M_2$	$M_3M_5M_1M_2$
L_9	0.207	0.05	58.07	$M_1M_2M_2M_4$	$M_3M_3M_1M_5$
L_{10}	0.213	0.05	63.24	$M_1M_3M_1M_4$	$M_5M_2M_0M_1$
L_{11}	0.218	0.05	68.42	$M_2M_3M_2M_1$	$M_4M_4M_1M_0$
L_{12}	0.226	0.10	79.15	$M_1M_1M_3M_5$	$M_2M_2M_1M_4$
Chromosome 11					
L_1	0.010	0.00	0.00	$M_1M_2M_1M_3$	$M_3M_4M_2M_2$
L_2	0.041	0.05	5.17	$M_0M_2M_2M_1$	$M_3M_4M_5M_1$
L_3	0.093	0.10	15.91	$M_1M_1M_2M_5$	$M_3M_4M_4M_2$
L_4	0.132	0.10	26.64	$M_2M_1M_1M_3$	$M_3M_4M_5M_5$
L_5	0.161	0.10	37.37	$M_3M_3M_2M_5$	$M_0M_4M_1M_1$
L_6	0.173	0.05	42.55	$M_3M_4M_2M_2$	$M_1M_2M_1M_3$
L_7	0.192	0.10	53.28	$M_3M_5M_1M_1$	$M_0M_1M_4M_2$
L_8	0.206	0.10	64.01	$M_1M_3M_1M_4$	$M_5M_2M_0M_1$
L_9	0.217	0.10	74.74	$M_5M_2M_1M_0$	$M_3M_4M_4M_2$
L_{10}	0.225	0.10	85.48	$M_1M_2M_4M_4$	$M_3M_3M_5M_0$
Chromosome 12					
L_1	0.020	0.00	0.00	$M_2M_1M_3M_4$	$M_3M_3M_5M_0$
L_2	0.050	0.05	5.17	$M_4M_2M_5M_1$	$M_1M_2M_3M_3$
L_3	0.075	0.05	10.35	$M_1M_1M_2M_5$	$M_3M_4M_4M_1$
L_4	0.098	0.05	15.52	$M_3M_4M_3M_2$	$M_1M_1M_5M_0$

L_5	0.136	0.10	26.26	$M_0M_4M_1M_2$	$M_3M_5M_1M_1$
L_6	0.164	0.10	36.99	$M_1M_1M_4M_2$	$M_3M_2M_2M_1$
L_7	0.175	0.05	42.16	$M_4M_1M_1M_3$	$M_2M_5M_2M_1$
L_8	0.185	0.05	47.34	$M_5M_0M_1M_2$	$M_3M_4M_4M_1$
L_9	0.193	0.05	52.51	$M_4M_3M_3M_1$	$M_0M_2M_2M_5$
L_{10}	0.207	0.10	63.24	$M_1M_2M_2M_5$	$M_4M_3M_3M_1$
L_{11}	0.213	0.05	68.42	$M_1M_0M_5M_2$	$M_3M_4M_3M_1$
L_{12}	0.218	0.05	73.59	$M_5M_2M_1M_1$	$M_4M_3M_3M_1$
L_{13}	0.222	0.05	78.77	$M_1M_2M_2M_4$	$M_5M_3M_3M_4$
L_{14}	0.229	0.10	89.50	$M_0M_2M_1M_3$	$M_4M_4M_0M_5$

Simulation parameters of the coefficient of double reduction (α for quadrivalent pairing) at and recombination frequencies (r) between 15 linked marker loci and QTL and parental genotypes at markers and QTL. M_i(i=1,…,5) represent five distinct alleles from two parent and M_0 represent the null allele. Alleles listed in the same column are on the same chromosome.

For each QTL, the genetic parameters of genetic mean (μ), monogenic (θ_1), digenic (θ_2), trigenic (θ_3) and quadrigenic effects (θ_4) are assumed to be 500, 100, 60, 30 and 10, respectively. Interaction effects between genetic effects of two QTLs are all assumed to be 1. These two QTLs are locating on two different chromosomes, thus we can assume they are in linkage equilibrium in the mapping population. The mapping population size is 300 and heritability of the two QTLs is 0.2. Based on the simulated parameters setting in Table I-2.9, I obtained a profile of LOD scores along the twelve chromosomes as shown in Figure I-2.5. Two QTLs were detected on Chromosome 1 and Chromosome 3, whose LOD scores exceeded the threshold of 5.78. It can be seen from Figure I-2.5 that the locations of QTLs were adequately predicted around the true QTL locations (26.64 cM at Chromosome 1 and 101.00 cM at Chromosome 3).

Figure I-2.5. Profile of LOD scores along the twelve chromosomes in the autotetraploid potato genome.



The true QTL are located on Chromosome 1 (26.64 cM) and Chromosome 3 (101.00 cM). Population size is 300 and heritability is 0.2.

2.6. Discussion

In this chapter of Part I, I proposed a theoretical model for interval mapping of QTL in a full-sib family of autotetraploids, which properly taking account for tetrasomic inheritance, including the key feature of double reduction and multiplex allele segregation. This method allows modelling and analysing data generated under bivalent paring or quadrivalent pairing of homologous chromosomes during meiosis in an autotetraploid individual. In addition, this method was designed to be generally applied to all kinds of marker genotyping data without any further modification. However, the current method has not taken the information of allele dosage into consideration, which could be obtained from genotyping technology by next generation gene sequencing. The adequacy of the method in estimating the model parameters and in mapping QTL was demonstrated by extensive simulation studies under both bivalent pairing and quadrivalent pairing models.

I analysed extensive simulated datasets to show that the method of interval mapping in autotetraploids gives adequate estimates of model parameters and mapping accuracy under bivalent meiosis or quadrivalent meiosis. With a small population size of 300 and a low level of heritability of 0.1, both bivalent method and quadrivalent method were powerful to detect QTL with 100 percent over the 100 replicated simulations. As expected, the bivalent method had better performance in mapping accuracy in fitting the data generated under a bivalent pairing model compared with that of the quadrivalent method which introduces an additional parameter as the coefficient of double reduction. However, the quadrivalent method showed stronger robustness in fitting data generated under different pairing models. On the one hand, the

bivalent method would collapse in fitting data generated under quadrivalent pairing model. From the simulation studies, we can see more one half of data generated under quadrivalent pairng did not fit bivalent pairing pattern and we have no alternative but to discarded them. This strategy would definitely cause loss of statistical power and loss of accuracy in parameter estimation. On the other hand, the quadrivalent method analysed data generated under bivalent pairing model with acceptable result both in statistical power and estimation accuracy. In the quadrivalent pairing model, the allele segregation distribution would be very close to that in the bivalent pairing model when the coefficients of double reduction on all the markers equal to zero. This explains why quadrivalent pairing model can still work well in fitting bivalent pairing data. From this aspect, this investigation indicates that quadrivalent method would outperform the bivalent method in modelling and analysing experimental data collected for QTL analysis in autotetraploids. In practice, it seems to be always the case that most autotetraploid species would undergo a mixture of bivalent pairing and quadrivalent pairing of homologous chromosomes during meiosis. Thus the bivalent method would be not applicable in the real data analysis, which would probably cause large bias in the parameters estimation, while the performance of quadrivalent method could be still satisfied. The quadrivalent method developed here modelled quadrivalent pairing of homologous chromosomes during meiosis and introduced the parameter of the coefficient of double reduction in the QTL mapping of autotetraploid species for the first time, acheiveing a step forward for QTL analysis in real dataset from autotetraploids.

In the current simulation study, it is assumed that marker order and recombination frequencies between them were known without error, which may not be the situation for experimental data. However, the reconstruction methods based on autotetrasomic model can be used to construct genetic map with high consistency with the true marker genotype and order (Luo et al 2000,

2004, 2006; Leach et al 2010), which have provided a good basis for QTL analysis with experimental data. In addition, I integrated the Hidden Markov chain statistical method proposed by Leach et al (2010) to calculate the conditional probability distribution of QTL genotypes using all the genetic markers on the chromosome but not just two flanking markers, which would improve the informativeness in fitting a QTL. In practice, parental QTL genotypes and linkage phase between QTL and marker are usually unknown. We used a computer-intensive search method to find the most likely parental QTL genotypes among 252 possible parental genotypes and phases, which adequately in detecting and locating a QTL on the chrosome under bivalent pairing and quadrivalent pairing in autotetraploid species. Moreover, the statistical methodology described here for QTL mapping in autotetraploids can be fully extendible to any other experimental design and any types of genetic markers. The theoretical method developed in this thesis, which first taking quadrivalent pairing into consideration, would provide analytical tools to recent lauched genome projects in autotetraploids, such as cultivated potato and farmed salmon, and improve breeding efficiency for economically important autotetraploid species in both agriculture and aquaculture.

From the simulation study of whole potato genome analysis, we can see that QTLs can be well detected by using this method when no or small amount of interactions existing between QTLs. However, if there are large amount of interactions between QTLs or QTLs are closely linked on the same chromosome, statistical inference of QTLs would be seriously biased. Thus simultaneously estimation of multiple QTLs would be appreciated by further efforts.

2.7. References

Bailey, N.T.J. (1961) **Introduction to the mathematical theory of genetic linkage.** Clarendon, Oxford.

Botstein, D., et al. (1980) Construction of a genetic map in man using restriction fragment length polymorphisms. **Am. J. Hum. Genet.,** 32: 314-331.

Baird, N.A., et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PLoS ONE,** 3:e3376.

Carlborg, O., et al. (2000) The use of a genetic algorithm for simultaneous mapping of +multiple interacting quantitative trait loci. **Genetics,** 155: 2003-2010.

Cao, D.C., et al. (2005) A model selection-based interval-mapping method for autopolyploids. **Genetics**, 169: 2371-2382.

Dempster, A.P. (1977) Maximum likelihood from incomplete data via the EM algorithm. **JR Statist SocB**, 39: 1-22.

Edwards, M.D., et al. (1987) Molecular-marker-facilitated investigation of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. **Genetics,** 116: 113-125.

Elshire, R.J., et al. (2011) A robust, simple genotyping-by-sequencing approach for high diversity species. **PLoS ONE,** 6: e19379.

Falconer, D.S. and Mackay T.F.C. (1996) **Introduction to quantitative genetics (Ed. 4).** Longman Group Ltd, UK.

Hoeschele, I. and VanRaden, P.M. (1993) Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. **Theor. Appl. Genet.**, 85: 953-960.

Hackett, C.A., et al. (1998) Linkage analysis in tetraploid species: a simulation study. **Genet Res,** 71: 143-154.

Hackett, C.A., et al. (2001) Interval mapping of quantitative trait loci in autotetraploid species. **Genetics,** 159: 1819-1832.

Hackett, C.A., et al. (2013) Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. **PLoS ONE,** 8(5): e63939.

Hackett, C.A., et al (2014) QTL mapping in autotetraploids using SNP dosage information. **Theor Appl Genet,** 127: 1885-1904.

Knott, S.A. and Haley, C.S. (1992) Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. **Genet. Res.,** 60: 139-151.

Kao, C.H., et al. (1999) Multiple interval mapping for quantitative trait loci. **Genetics,** 152:1203-1216.

Lander, E.S. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics,** 121: 185-199.

Lincoln, S.E. and Lander, E.S. (1992) Systematic detection of errors in genetic linkage data. **Genomics**, 14: 604-610.

Luo, Z.W., et al. (2000) Predicting parental genotypes and gene segregation for tetrasomic inheritance. **Theor Appl Genet,** 100: 1067-1073.

Luo, Z.W., et al. (2001) Construction of a genetic linkage map in tetraploid species using molecular makrers. **Genetics,** 157: 1369-1385.

Luo, Z.W., et al. (2004) Theoretical basis for genetic linkage analysis in autotetraploid species. **PNAS**, 101: 7040-7045.

Luo, Z.W., et al. (2006) Constructing genetic linkage maps under a tetrasomic model. **Genetics,** 172: 2635-2645.

Leach, J.L., et al. (2010) Multilocus tetrasomic linkage analysis using hidden Markov chain model. **PNAS**, 107: 4270-4274.

Martinez, O. and Curnow, R.N. (1992) Estimation the locations and the sizes of the effects of quantitative trait loci using flanking markers. **Theor. Appl. Genet.,** 85: 480-488.

Meyer, R.C., et al. (1998) Linkage analysis tetraploid potato adn associations of markers with quantitative resistance to late blight. **Mol. Gen. Genet.,** 259: 415-419.

Mackay, T.F.C., et al. (2009) The genetics of quantitative traits: challenges and prospects. **Nature,** 10: 565-576.

Rasmusson, J.M. (1933) A contribution to the theory of quantitative character inheritance. **Hereditas,** 18: 245-261.

Sax, K. (1923) The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. **Genetics,** 8: 552-560.

Soller, M., et al. (1976) On the power of experimental design for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. **Theor. Appl. Genet.**, 47: 35-39.

Stam, P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. **The Plant Journal,** 3(5): 739-744.

Sills, G.R., et al. (1995) Genetic analysis of agronomic traits in a cross Between sugarcane (Saccharum officinarum L.) and its presumed progenitor (S. robustum Brandes and Jesw. ex Grassl). **Mol Breed**, 1: 355-363.

Satagopan, J.M., et al. (1996) A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. **Genetics,** 144: 805-816.

Sillanpaa, M.J. and Arjas, E. (1999) Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. **Genetics,** 151: 1605-1619.

Thoday, J.M. (1961) Location of polygnes. **Nature,** 191: 368-370.

Tanksley, S.D., et al (1982) Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. **Heredity,** 49: 11-25.

Uimari, P. and Hoeschele, I. (1997) Mapping-linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. **Genetics,** 146: 735-743.

Weller, J.I. (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. **Biometrics,** 42: 627-640.

Zeng, Z.B. (1994) Precision mapping of quantitative trait loci. **Genetics,** 136: 1457-1468.

Zeng, Z.B., et al. (1999) Estimating the genetic architecture of quantitative traits. **Genet. Res.,** 74: 279-289.

PART II

THEORY AND METHODS FOR ANALYSIS OF CROSSOVERS DURING MEIOSIS IN AUTOTETRAPLOIDS

Chapter II-1: Statistical inference of crossover interference in both diploids and autotetraploids

1.1. Overview

The world is currently facing what is arguably its most serious challenge yet, to meet the demand for a sustainable food supply for its rapidly expanding population. A crucial goal to address this crisis is to develop tools for breeding of both diploid and polyploidy crops that are designed to fully realize selection response through the release of genetic variation that is currently "locked up" in crop plant genomes. This "release" of genetic variation occurs during the process of recombination between paired homologous chromosomes during meiosis. Pairing and recombination is essential for ensuring balanced chromosome segregation and enables generation of new combinations of chromosomes segments or alleles at different genetic loci, boosting genome variability. The genetic variation so created forms the most important basis both for natural selection that drives the evolution of species and also artificial selection that enables target alleles to be integrated into an elite line or strain in genetic breeding programs of domesticated animals or agricultural crops.

Over the past decade it has emerged that chromosome pairing and synapsis, followed by recombination, as the key events in meiosis, are subject to highly stringent and complex control**.** It has been well established that a series of genes or proteins are involved in the process of synapsis and in the subsequent promotion or limitation of meiotic DNA double-strand breaks and crossovers (COs), thus influencing the frequency of meiotic recombination (Osman et al

2011). It has been established since the era of Thomas Morgan that recombination occurring at one chromosome site is not independent of the other recombination at nearby sites, which describes the well-known phenomenon of recombination interference (Sturtevant 1915, Muller 1916). In almost all organisms, CO interference is likely to play an important role in determining the frequency and patterns of recombination along chromosomes (Drouaud et al 2007). For example, extensive variation in recombination frequency between maize populations has been linked to variation in the strength of interference (Bauer et al 2013). However, remarkably little is known about the mechanisms of interference or about factors that affects its strength.

In diploid meiosis, crossing over, the cytological organization of homologous chromosomes prior to recombination, takes place along the bundle of four chromatids and therefore recombination interference (RI) may be attributed to two types of interference: First, chromatid interference, where different pair of non-sister chromatids are not equally likely to be involved in the formation of crossovers; second, chiasmata interference, where the occurrence of one crossover event at a given position along the chromosomal bundle affects the chances of an additional crossover occurring in a nearby region (Stam 1979). There is sparse evidence of chromatid interference in the literature and it can only be detected if all four products of a single meiosis can be recovered, such as in the fungi *Saccharomyces cerevisiae, Neurospora crassa* and *Asperqillus nidulans* (Lindegren 1942; Strickland 1958; Hawthorne 1960) or in specific mutants of *Arabidopsis* (Copenhaver 1998). Most work on RI has there focused on chiasma interference. RI can be measured in terms of the coefficient of coincidence in the form of $C = r_{11} / \left[(r_{10} + r_{11})(r_{01} + r_{11}) \right]$, which is the ratio of the observed frequency of simultaneous recombination in two disjoint chromosomal regions over the expected frequency of

recombination in these regions under independence of recombination between the two regions (Sturtevant 1915; Muller 1916). Key for calculating the interference parameter is to estimate the frequency of double recombination, which is usually infeasible, if not impossible, in practice. Thus, this kind of RI analysis has been limited to those species such as yeast where gamete genotyping is practically feasible (Malkova et al 2004), or to species where sperm or pollen typing is possible.

Statistically, the prediction of RI from zygotic genotype data requires modelling the mathematical relationship between the rate of crossover and frequency of recombination and hence the probabilistic distribution of crossover events along the chromatid bundle. Many different models have been proposed to model the crossover distribution in diploid species, including the count-location model proposed by Karlin and Liberman (1979), and the Poisson model proposed by Cox and Isham (1980), in which crossovers occur as a stationary Poisson point process. A hard core model proposed by Stoyan et al (1987) formulates crossover events as a stationary renewal process with renewal intervals distributed as a constant scaled exponential. By assuming the absence of chromatid interference, McPeek et al (1995) proposed a gamma model by generalizing the Poisson point process of crossovers to be a renewal process taking general gamma distributed intervals. In this the coincidence parameter was estimated together with other model parameters through a maximum likelihood method, which outperformed other rival methods in the fitting of multi-locus genotype data. To make the gamma model mathematically more tractable, Zhao et al (1995) proposed a Chi-square model in the form of $C_x(C_o)^m$, which can be interpreted the model as the occurrence of a crossover event (C_x) followed by a number *m* of non-crossover events (C_o). Parameter, *m* therefore measures the intensity of crossover interference, with larger value indicating stronger interference. The

Chi-square model has an inherent property of mathematical tractability while retaining a strong biological basis (Foss et al. 1993). The stochastic process of crossovers along paired chromosomes may be considered as a renewal process whereby the distance between adjacent crossover events are independently drawn from a probability distribution, originally proposed by Mather (1936a, 1937) and further elaborated extensively in Owen (1950) and Carter and Robertson (1952).

Based on the Chi-square model, I developed here a novel statistical method for recombination interference analysis with autotetraploid species. The method properly accounts for the essential features of segregation and recombination under tetrasomic inheritance. We tested reliability of the method and explored its statistical properties through an intensive computer simulation study. In addition, we demonstrated utility of the method by implementing it to model and analysed phenotype datasets of three linked fluorescent marker loci scored from a large segregating population of diploid and autotetraploid budding yeast *S. cerevisiae*.

1.2. Methods of inferring crossover interference for zygote in diploids

1.2.1. The Chi-square model

The Chi-square model for crossovers has been historically of interest (Mather 1936, 1937; Owen 1950; Carter and Robertson 1952) extensively. The Chi-square model can be represented in a

form of $C_x(C_o)^m$ (Zhao 1995a), explained as that occurrence of crossover events (C_x) along the paired chromosomes were separated by *m* consecutive non-crossover events (C_o), where *m* measured the intensity of crossover interference. Both C_x and C_o are termed as *C* events and are randomly distributed on the four-strand bundle of paired chromosomes in a diploid meiosis and the number of *C* events in any given chromosomal interval is assumed to follow a Poisson distribution. Secondly, the chi-square model assumes absence of chromatid interference, i.e. crossover may occur between any pair of non-sister chromatids with an equal chance. It is referred as "chi-square" model because the probability distribution of interexchange distances is a chi-square distribution with an even number of degrees of freedom (Lange *et al*. 1997).

The Chi-square model, which was originally developed for diploid species, has two key features. First, it models the distribution of crossover and non-crossover event along paired chromosomal bundles without any limitation on the number of homologous chromosomes involved. Second, RI is modelled in terms of the distance between two consecutive crossover events. These basic features are not inherently specific to diploid genomes where the bundle involves four homologous chromosomes. Thus, the model can in principle be extended to autotetraploid species where the bundle may involve eight chromosomes. However, such an extension is not trivial autotetraploids may undergo tetrasomic inheritance and thus shows a much more complicated pattern of gene segregation and recombination when compared to disomic inheritance in diploids. Firstly, in autopolyploids, multivalent pairing of homologous chromosomes during meiosis may result in the well-known phenomenon of double reduction, in which sister chromatids enter into the same gamete (Mather 1936b), resulting in systematic allelic segregation distortion in comparison to disomic gene segregation and recombination.

Secondly, multiple alleles at individual loci of polyploids cause a substantially wider spectrum of genotypic segregation (Luo et al 2004).

1.2.2. Assumption and notations

We consider three marker loci, A, B and C, along the chromosome of a diploid species. Three parameters are needed to specify the three-locus model, *m* indicating the number of non-crossovers (C_o) between two crossovers (C_x), and d_1 and d_2 being genetic distances of the first and second marker interval respectively. The genetic distances are defined as the expected number of crossovers occurring on a single chromatid within the given interval. In the Chi-square model aforementioned, *C* events (including both C_x and C_o) are assumed to be randomly distributed on the four-strand bundle, and the number of *C* events in any given chromosomal interval follows a Poisson distribution with a mean of *y*. Thus, the probability of *s C* events is equal to $e^{-y}y^s/s!$. In absence of chromatid interference, each strand made up of the four-strand bundle has a chance of ½ to be involved in every crossover. Let $p=m+1$ be the length of a complete set of *C* events, $C_x(C_o)^m$. Each strand will be involved in an expected number of $s/(2p)$ crossovers among *s C* events. According to the definition of genetic distance here, the average number of *C* events (y_1 or y_2) occurring within any given interval can be expressed in terms of genetic distance of the interval (d_1 or d_2), for example $y_1=2pd_1$ (or $y_2=2pd_2$). For three marker loci, there are up to four different recombination configurations for each chromatid. We denoted one of the recombination configurations by a vector $X=(i,j)$ and the

corresponding probability of $X=(i,j)$ by x_{ij}, where i (or j) takes a value of 0 standing for non-recombination or 1 for recombination that occurs in the first (or second) marker interval.

1.2.3. Prediction of probability distribution of crossovers occurring within marker intervals

In the following discussion, I suppose that markers are laid out from left to right, and the *C* events occur also from left to right. The chi-square model assumes that the *C* events resolve in sequence as $\cdots C_x C_0 C_0 \cdots C_0 C_x C_0 \cdots$ and that the process is stationary, so the first *C* event occurring in the first marker interval could be any one of the *m+1* elements in $C_x (C_0)^m$, each of which occurs with an equal probability of $1/p$.

In each marker interval, the occurrence of k ($k \geq 1$) crossovers (C_x) between two flanking markers might be the result of p^2 possible situations, depending on the number of C_0's between the first crossover C_x^1 and the left marker of the interval, *L*, and the number of C_0's between the last crossover C_x^k and the right marker of the interval, *R*, as illustrated in Figure 1. These p^2 possible *C* events can be expressed as a $p \times p$ matrix as follows:

$$M_k = \begin{bmatrix} (C_0)^m C_x^1 (C_0)^m \cdots C_x^k & (C_0)^m C_x^1 (C_0)^m \cdots C_x^k C_0 & \cdots\cdots & (C_0)^m C_x^1 (C_0)^m \cdots C_x^k (C_0)^m \\ (C_0)^{m-1} C_x^1 (C_0)^m \cdots C_x^k & (C_0)^{m-1} C_x^1 (C_0)^m \cdots C_x^k C_0 & \cdots\cdots & (C_0)^{m-1} C_x^1 (C_0)^m \cdots C_x^k (C_0)^m \\ \vdots & \vdots & \cdots\cdots & \vdots \\ C_0 C_x^1 (C_0)^m \cdots C_x^k & C_0 C_x^1 (C_0)^m \cdots C_x^k C_0 & \cdots\cdots & C_0 C_x^1 (C_0)^m \cdots C_x^k (C_0)^m \\ C_x^1 (C_0)^m \cdots C_x^k & C_x^1 (C_0)^m \cdots C_x^k C_0 & \cdots\cdots & C_x^1 (C_0)^m \cdots C_x^k (C_0)^m \end{bmatrix}_{p \times p}$$

(II-1.1)

The matrix lists all possible *C* events given *k* crossovers occurring in the chromosomal interval. Within the marker interval, there are *k*–1 sets of $C_x (C_0)^m$, C_x and a varying number of C_0 events between the left marker and the first C_x event (listed as column elements of the matrix) or between the last C_x event and the right marker (listed as row elements). Thus, each element in this $p \times p$ matrix represents one possible *C* event sequence occurring in a marker interval, i.e. the (*i*, *j*)th element in the matrix for (*p*-*i*) C_0 events occurring between the first C_x and the left marker and (*j*-1) C_0 events between the last C_x and the right marker. In the C_x^i stands for the *i*th (*i* = 1, 2, …, *k*) C_x in the chromosomal interval and $(C_0)^j$ for *j* consecutive C_0 events.

If there are no crossover events occurring in the marker interval, then the *C* events can only be C_0 events varying in number from zero to *m*. We represent the corresponding matrix M_k as M_0 for the special case of no crossovers in the form of:

$$M_0 = \begin{bmatrix} no\ C\ event & C_0 & \cdots\cdots & (C_0)^m \\ & no\ C\ event & \cdots\cdots & (C_0)^{m-1} \\ & & \ddots & \vdots \\ & & & C_0 \\ & & & no\ C\ event \end{bmatrix}_{p\times p} \qquad \text{(II-1.2)}$$

It is assumed that the *C* events are randomly distributed along the chromosome and the number of *C* events within the interval follows Poisson distribution with the mean parameter $y = 2pd$, thus the probabilities of *C* events listed in matrix M_k can be listed accordingly in matrix $D_k(y)$ ($k \geq 1$) in the form of:

$$D_k(y) = e^{-y}\begin{bmatrix} y^{pk+1-1}/(pk+1-1)! & \cdots & y^{pk+j-1}/(pk+j-1)! & \cdots & y^{pk+p-1}/(pk+p-1)! \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y^{pk+i-1}/(pk+1-i)! & \cdots & y^{pk+j-i}/(pk+j-i)! & \cdots & y^{pk+p-i}/(pk+p-i)! \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y^{pk+p-1}/(pk+1-p)! & \cdots & y^{pk+j-p}/(pk+j-p)! & \cdots & y^{pk+p-p}/(pk+p-p)! \end{bmatrix}_{p\times p} \triangleq \left(Q_y^{ijk}\right)_{p\times p}$$

(II-1.3)

When $k = 0$, the probability matrix of no crossovers between markers corresponding to the *C* events listed in M_0 can be expressed as:

$$D_0(y) = e^{-y}\begin{bmatrix} y^{1-1}/(1-1)! & \cdots & y^{i-1}/(i-1)! & \cdots & y^{p-1}/(p-1)! \\ 0 & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & y^{i-i}/(i-i)! & \cdots & y^{p-i}/(p-i)! \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & y^{p-p}/(p-p)! \end{bmatrix}_{p\times p} \triangleq \left(Q_y^{oij}\right)_{p\times p} \qquad \text{(II-1.4)}$$

Since the first *C* event in the first marker interval has an equal chance of being any one of the *m+1* elements of $C_x(C_0)^m$ with probability of $1/p$, then the probability of *k* crossovers between markers A and B can be computed by summing over probabilities of all the p^2 possibilities as

$$\begin{cases} \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p} Q_y^{ijk} & if\ k \geq 1 \\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p} Q_y^{oij} & if\ k = 0 \end{cases} \tag{II-1.5}$$

where $Q_y^{ijk} = e^{-y} y^{pk+j-i} / (pk + j - i)!$, which is the $(i, j)^{th}$ element in matrix $D_k(y)$ (Equation II-1.3) , and $Q_y^{oij} = \begin{cases} e^{-y} y^{j-i}/(j-i)! & if\ i \leq j \\ 0 & if\ i > j \end{cases}$, the $(i, j)^{th}$ element in matrix $D_0(y)$ (Equation II-1.4).

We can represent the probability of *k* crossovers between markers A and B in matrix notation as

$$\frac{1}{p} I \cdot D_{k_1}(y_1) \cdot I' \tag{II-1.6}$$

where $I = [1 \quad \cdots \quad 1]_{1\times p}$.

As demonstrated by Zhao (1995), the analysis formulated above can be extended from two markers to that with three markers A, B and C. We firstly noted that sum of elements in the j^{th} $(1 \leq j \leq p)$ column in the matrix M_k is the probability of the situation that *k* crossovers occurr between A and B followed by the (*j*-1) C_0 events ($1 \leq j \leq p$),

$p_k^j = \sum_{i=1}^{p} Q_y^{ijk}$ $(\text{for } k \geq 1)$ or $\sum_{i=1}^{p} Q_y^{oij}$ $(k = 0)$. Thus, the probability of k crossovers between the markers A and B can also be expressed as

$$\frac{1}{p} I \cdot D_k(y) \cdot I' = \frac{1}{p} \left(p_k^1 \quad p_k^2 \quad \ldots \quad p_k^{m+1} \right) \cdot I' \tag{II-1.7}$$

In the three marker model, we formulate the probability of k_1 and k_2 crossovers in the first and second chromosomal intervals respectively. Because the probability that there are l C_0 's between marker B and the first C_x in the marker interval flanked by B and C is equivalent to the probability that the last C event between markers A and B is the $(p\text{-}l\text{-}1)^{\text{th}}$ C event after the last C_x in the marker interval flanked by A and B, which is $p_{k_1}^{p-l}$. Thus, the probability of k_1 crossovers occurring between markers A and B, and k_2 crossovers between markers B and C is given by

$$\frac{1}{p} \sum_{l=1}^{p} p_{k_1}^{p-l} \sum_{j=1}^{p} Q_{y_2}^{ljk_2} \quad (k_2 \geq 1) \quad \text{or} \quad \frac{1}{p} \sum_{l=1}^{p} p_{k_1}^{p-l} \sum_{j=1}^{p} Q_{y_2}^{olj} \quad (k_2 = 0) \tag{II-1.8}$$

Equation (1.8) can be expressed in a matrix form as

$$\frac{1}{p} \left(p_{k_1}^1 \quad p_{k_1}^2 \quad \ldots \quad p_{k_1}^p \right) \cdot D_{k_2}(y_2) I' \tag{II-1.9}$$

or in a general form of

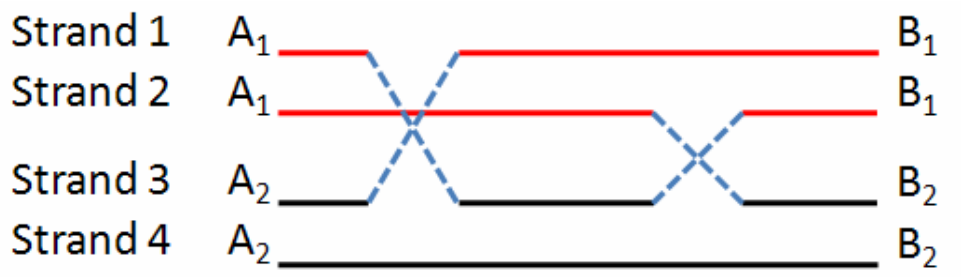
$$\frac{1}{p} I \cdot D_{k_1}(y_1) \cdot D_{k_2}(y_2) \cdot I' \tag{II-1.10}$$

1.2.4. Prediction of probability distribution of recombination along a chromatid

As mentioned before, if no crossover occurs between markers, no strand in the bundle will show any recombination in that interval, while the expected recombination frequency is ½ given $k \geq 1$ crossovers between two markers under the assumption of absence of chromatid interference as demonstrated below.

In Figure II-1.1, solid lines in red and black represent two replicated chromosomes in diploids and blue forks in dotted lines indicate crossovers occurring between marker loci A and B. A_1, A_2 and B_1, B_2 are marker alleles on loci A and B, respectively. Since there is no chromatid interference, chiasmata, the point where two homologous non-sister chromatids exchange genetic material during meiosis, is equally likely to involve in any two strands can lead to a crossover. Regarding only one of the A alleles, *e.g.* A_1 on strand 1, the expected recombination frequency after i crossovers between marker A and marker B is equal to the probability that allele A_1 is not linked to B_1 on the same strand after i crossovers. Here the number of crossovers within this chromosome region is denoted as x and the number of crossover event involving strand 1 is denoted as y. The probability distribution of y is listed in Table II-1.1.

Figure II-1.1. Diagrammatic representation of crossovers occurring between marker A and B on the chromosome in diploids, showing a typical double crossing-over on strand 1.



Here solid lines in red and black represent two replicated chromosomes in diploids and blue forks in dotted lines indicate crossovers occurring between marker loci A and B.

Table II-1.1. The probability distribution of *y* crossover events involving strand 1 with total *x* crossovers occurring in the chromosome region

Chiasmata (x)	Crossover (y)						
	0	1	2	3	…	i	…
0	1						
1	1/2	1/2					
2	1/4	1/2	1/4				
3	1/8	3/8	3/8	1/8			
⋮	⋮	⋮	⋮	⋮	⋮		
i	$\frac{1}{2^i}\binom{i}{0}$	$\frac{1}{2^i}\binom{i}{1}$	$\frac{1}{2^i}\binom{i}{2}$	$\frac{1}{2^i}\binom{i}{3}$	…	$\frac{1}{2^i}\binom{i}{i}$	
⋮							

As shown in Figure II-1.1, double crossover on strand 1 will restore the original relation of allele A_1 and B_1 and so will be non-recombination. Similarly triple crossover will give recombination, quadruple crossover no recombination and so on. Hence the expected frequency of recombination given i crossovers occurring between marker A and B, p, will be given by the summed frequencies of the single, triple, quintuple, etc. crossovers as,

$$p=\begin{cases}\dfrac{1}{2^i}\left[\binom{i}{1}+\binom{i}{3}+\cdots+\binom{i}{i-2}+\binom{i}{i}\right] & \textit{if } i \textit{ is odd}\\ \dfrac{1}{2^i}\left[\binom{i}{1}+\binom{i}{3}+\cdots+\binom{i}{i-3}+\binom{i}{i-1}\right] & \textit{if } i \textit{ is even}\end{cases} \tag{II-1.11}$$

When *i* odd, since

$$\begin{cases}\binom{i}{1}+\binom{i}{3}+\cdots+\binom{i}{i-2}+\binom{i}{i}=\binom{i}{i-1}+\binom{i}{i-3}+\cdots+\binom{i}{2}+\binom{i}{0}\\ \left[\binom{i}{1}+\binom{i}{3}+\cdots+\binom{i}{i-2}+\binom{i}{i}\right]+\left[\binom{i}{i-1}+\binom{i}{i-3}+\cdots+\binom{i}{2}+\binom{i}{0}\right]=2^i\end{cases} \tag{II-1.12}$$

it can be deduced that $\binom{i}{1}+\binom{i}{3}+\cdots+\binom{i}{i-2}+\binom{i}{i}=2^{i-1}$ and the expected recombination frequency in Equation (1.1) is $p=2^{i-1}/2^i=1/2$.

When *i* even, according to the property of binomial coefficient, $\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}$, the expected recombination frequency of Equation (1.1) can be rewrite as

$$\begin{aligned}p&=\frac{1}{2^i}\left[\binom{i-1}{0}+\binom{i-1}{1}+\binom{i-1}{2}+\binom{i-1}{2}+\cdots+\binom{i-1}{i-2}+\binom{i-1}{i-1}\right]\\ &=\frac{1}{2^i}\cdot 2^{i-1}=\frac{1}{2}\end{aligned} \tag{II-1.13}$$

So as long as *i* ($i\geq 1$) crossover occurring between two loci, the expected recombination frequencies are all equal to 1/2.

Let $o_0^{(k)}$ denote the probability of no recombination and $o_1^{(k)}$ denote the probability of recombination given *k* crossovers occurring within the marker interval. It is clear that $o_0^{(0)}=1$, $o_1^{(0)}=0$ and $o_0^{(k)}=o_1^{(k)}=1/2$ when $k\geq 1$. Let $X=(i_1 \quad i_2)$ represent recombination configuration of a chromatid, with i_1 = 1 or 0 indicating recombination or non-recombination between marker A and B along the chromatid and similarly i_2 = 1 or 0 indicating recombination or non-recombination between marker B and C. We denote $O_{(i_1 \quad i_2)}^{(k_1 \quad k_2)}$ for the probability of recombination configuration of a chromatid $X=(i_1 \quad i_2)$ given that there are k_1 and k_2 crossovers in the first and second interval, respectively. Then x_{ij}, the probability of observing recombination configuration of a chromatid $X=(i_1 \quad i_2)$, can be calculated as

$$\begin{aligned} x_{i_1 i_2} &= \sum_{k_1,k_2} O_{(i_1 \quad i_2)}^{(k_1 \quad k_2)} \\ &= \sum_{k_1}\sum_{k_2} o_{i_1}^{(k_1)} o_{i_2}^{(k_2)} \frac{1}{p} \times I \cdot D_{k_1}(y_1) \cdot D_{k_2}(y_2) \cdot I' \\ &= \frac{1}{p} I \cdot \left(\sum_{k_1}\sum_{k_2} o_{i_1}^{(k_1)} o_{i_2}^{(k_2)} \times D_{k_1}(y_1) \cdot D_{k_2}(y_2) \right) \cdot I' \\ &= \frac{1}{p} I \cdot \left(\sum_{k_1} o_{i_1}^{(k_1)} D_{k_1}(y_1) \sum_{k_2} o_{i_2}^{(k_2)} D_{k_2}(y_2) \right) \cdot I' \end{aligned} \qquad \text{(II-1.14)}$$

Define

$$M_j = \begin{cases} D_0(y_j) + 1/2 \sum_{s\geq 1} D_s(y_j) & \text{when } i_j = 0 \\ 1/2 \sum_{s\geq 1} D_s(y_j) & \text{when } i_j = 1 \end{cases}$$

Then the probability of recombination configuration of a chromatid $X=(i_1 \quad i_2)$ is

$$x_{i_1 i_2}=\frac{1}{p} I \cdot M_1 M_2 \cdot I' \tag{II-1.15}$$

where $I=[1 \quad \cdots \quad 1]_{1\times p}$.

1.2.5. Prediction of probability distribution of marker phenotypes at three loci

We consider a generic heterozygous genotype at three loci, i.e. $A_1B_1C_1/A_2B_2C_2$. Gametes to be generated from this individual can be divided into four categories according to the recombination events in the intervals. A general form for frequency of the gametic genotype *i* can be expressed as

$$\begin{aligned} G_i &= w_{i00}\cdot\frac{1}{2}x_{00}+w_{i01}\cdot\frac{1}{2}x_{01}+w_{i10}\cdot\frac{1}{2}x_{10}+w_{i11}\cdot\frac{1}{2}x_{11} \\ &= \frac{1}{2}\sum_{s=0}^{1}\sum_{t=0}^{1} w_{ist}x_{st} \end{aligned} \tag{II-1.16}$$

where w_{ist} is the number of gametes with recombination configuration $X=(s\ t)$ within the i^{th} gametic genotype category. By assuming random union between all possible gametes generated from two parents, a general form for the frequency of zygote genotype *i* can be carried out after sorting the zygotes according to their genotypes and be written as

$$H_i = \frac{1}{4}\sum_{s=0}^{1}\sum_{t=0}^{1}\sum_{s'=0}^{1}\sum_{t'=0}^{1} z_{i,st/s't'} x_{st} x_{s't'} \quad \text{(II-1.17)}$$

where $z_{i,st/s't'}$ indicates the number of zygotes made up with the two gametes with recombination configuration $X=(s\ t)$ and $X'=(s'\ t')$ within the i^{th} zygotic genotype class.

Probability of the i^{t} phenotype of the three markers among offspring can be readily derived by summing up the probabilities of those genotypes that are compatible to the same phenotype, and is expressed as

$$f_i = \sum_{g\in i} H_g = \frac{1}{4}\sum_{g\in i}\sum_{s=0}^{1}\sum_{t=0}^{1}\sum_{s'=0}^{1}\sum_{t'=0}^{1} z_{g,st/s't'} x_{st} x_{s't'} \quad \text{(II-1.18)}$$

where $z_{g,st/s't'}$ indicates the number of zygotes made up with the two gametes with recombination configuration $X=(s\ t)$ and $X'=(s'\ t')$ within the g^{th} zygotic genotype class.

1.2.6. The maximum likelihood estimates of the model parameters

In the model above, the unknown parameters are *m*, d_1 and d_2. The statistical analysis presented below predicts these model parameters based on the P_1 and P_2, the parental genotypes (i.e. parental genotypes could be predicted from phenotypes of parents and offspring), and $O=(o_1,o_2,...,o_n)$, the phenotype records of a random sample of *n* offspring individuals from the parental lines. Let *M* be the number of phenotype categories in the offspring. We assume that the

phenotype of offspring is randomly sampled from a multinomial distribution with probability parameters given by $f_i(i=1,2,...,M)$. Then the likelihood function of the parameters $\Omega=(m,d_1,d_2)$ has a form of

$$L(P_1,P_2,\Omega\,|\,O)\propto \Pr(O\,|\,P_1,P_2,\Omega)=\binom{n}{n_1 n_2\cdots n_M} f_1^{n_1} f_2^{n_2}\cdots f_M^{n_M} \tag{II-1.19}$$

Where $n_i(i=1,2,\ldots,M)$ is the number of individuals within the i^{th} phenotype class. Logarithm of the likelihood in

$$\log\left(L(P_1,P_2,\Omega\,|\,O)\right)\propto \sum_{i=1}^{M} n_i \ln(f_i)\propto \sum_{i=1}^{M} n_i \ln\left(\sum_{g\in i}\sum_{s=0}^{1}\sum_{t=0}^{1}\sum_{s'=0}^{1}\sum_{t'=0}^{1} z_{g,st/s't'}x_{st}x_{s't'}\right) \tag{II-1.20}$$

where $z_{i,st/s't'}$ indicates the number of zygotes made up with the two gametes with recombination configuration $X=(s\ t)$ and $X'=(s'\ t')$ within the i^{th} zygotic genotype class. x_{st} and $x_{s't'}$ can be calculated according to From equation (II-1.14) as follows

$$x_{00}=\frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{Q_{y_1}^{oj_1j_2}+\frac{1}{2}\sum_{k_1\geq 1}Q_{y_1}^{j_1j_2k_1}\right\}\left\{Q_{y_2}^{oj_2j_3}+\frac{1}{2}\sum_{k_2\geq 1}Q_{y_2}^{j_2j_3k_2}\right\}$$

$$x_{01}=\frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{Q_{y_1}^{oj_1j_2}+\frac{1}{2}\sum_{k_1\geq 1}Q_{y_1}^{j_1j_2k_1}\right\}\left\{\frac{1}{2}\sum_{k_2\geq 1}Q_{y_2}^{j_2j_3k_2}\right\}$$

$$x_{10}=\frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{\frac{1}{2}\sum_{k_1\geq 1}Q_{y_1}^{j_1j_2k_1}\right\}\left\{Q_{y_2}^{oj_2j_3}+\frac{1}{2}\sum_{k_2\geq 1}Q_{y_2}^{j_2j_3k_2}\right\}$$

$$x_{11}=\frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{\frac{1}{2}\sum_{k_1\geq 1}Q_{y_1}^{j_1j_2k_1}\right\}\left\{\frac{1}{2}\sum_{k_2\geq 1}Q_{y_2}^{j_2j_3k_2}\right\}$$

where $Q_y^{ijk} = e^{-y} y^{pk+j-i} / (pk+j-i)!$ and $Q_y^{oij} = \begin{cases} e^{-y} y^{j-i} / (j-i)! & if\ i \le j \\ 0 & if\ i > j \end{cases}$.

Let a_{st}^k denote the conditional probability of recombination configuration $X = (s\ t)$ along the three marker loci given *k* crossovers occurring in the marker interval A and B, thus when $k \ge 1$

$$a_{00}^k = a_{10}^k = \frac{1}{p} \sum_{j_3=1}^{p} \sum_{j_2=1}^{p} \sum_{j_1=1}^{p} \left\{ \frac{1}{2} Q_{y_1}^{j_1 j_2 k} \right\} \left\{ Q_{y_2}^{oj_2 j_3} + \frac{1}{2} \sum_{k_2 \ge 1} Q_{y_2}^{j_2 j_3 k_2} \right\}$$

$$a_{11}^k = a_{01}^k = \frac{1}{p} \sum_{j_3=1}^{p} \sum_{j_2=1}^{p} \sum_{j_1=1}^{p} \left\{ \frac{1}{2} Q_{y_1}^{j_1 j_2 k} \right\} \left\{ \frac{1}{2} \sum_{k_2 \ge 1} Q_{y_2}^{j_2 j_3 k_2} \right\}$$

when *k* = 0, it can only resolve in non-recombinant in the marker interval flanked by A and B, then

$$a_{00}^0 = \frac{1}{p} \sum_{j_3=1}^{p} \sum_{j_2=1}^{p} \sum_{j_1=1}^{p} \left\{ Q_{y_1}^{oj_1 j_2} \right\} \left\{ Q_{y_2}^{oj_2 j_3} + \frac{1}{2} \sum_{k_2 \ge 1} Q_{y_2}^{j_2 j_3 k_2} \right\}$$

$$a_{01}^0 = \frac{1}{p} \sum_{j_3=1}^{p} \sum_{j_2=1}^{p} \sum_{j_1=1}^{p} \left\{ Q_{y_1}^{oj_1 j_2} \right\} \left\{ \frac{1}{2} \sum_{k_2 \ge 1} Q_{y_2}^{j_2 j_3 k_2} \right\}$$

Similarly, let b_{st}^k denote the conditional probability of recombination configuration $X = (s\ t)$ along the three marker loci given *k* crossovers occurring in the marker interval B and C, thus when $k \ge 1$

$$b_{00}^{k} = b_{01}^{k} = \frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{Q_{y_1}^{oj_1j_2} + \frac{1}{2}\sum_{k_1\geq 1} Q_{y_1}^{j_1j_2k_1}\right\}\left\{\frac{1}{2}Q_{y_2}^{j_2j_3k}\right\}$$

$$b_{11}^{k} = b_{10}^{k} = \frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{\frac{1}{2}\sum_{k_1\geq 1} Q_{y_1}^{j_1j_2k_1}\right\}\left\{\frac{1}{2}Q_{y_2}^{j_2j_3k}\right\}$$

when $k = 0$, it can only resolve in non-recombinant in the marker interval flanked by B and C, then

$$b_{00}^{0} = \frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{Q_{y_1}^{oj_1j_2} + \frac{1}{2}\sum_{k_1\geq 1} Q_{y_1}^{j_1j_2k_1}\right\}\left\{Q_{y_2}^{oj_2j_3}\right\}$$

$$b_{10}^{0} = \frac{1}{p}\sum_{j_3=1}^{p}\sum_{j_2=1}^{p}\sum_{j_1=1}^{p}\left\{\frac{1}{2}\sum_{k_1\geq 1} Q_{y_1}^{j_1j_2k_1}\right\}\left\{Q_{y_2}^{oj_2j_3}\right\}$$

Because m takes non-negative integer values, we propose here an EM algorithm to calculate the MLE of the model parameters d_1 and d_2 at any given value of m, and determine the MLE of all three parameters by searching for the maximum of the likelihood profiles at every given value of m. In the EM algorithm, the expectation (E) step calculates the conditional probability of a total of k crossovers within the first marker interval in the offspring with the ith phenotype, γ_{ik}, given the model parameters, which can be formulated as

$$\gamma_{ik} = \sum_{k'=0}^{k}\left\{\frac{1}{4}\sum_{g\in i}\sum_{s=0}^{1}\sum_{t=0}^{1}\sum_{s'=0}^{1}\sum_{t'=0}^{1} z_{g,st/s't'}\, a_{st}^{k'} a_{s't'}^{k-k'} \Big/ f_i\right\} \qquad \text{(II-1.21)}$$

$$if\ s\neq 0, k'\ can\ not\ be\ 0\ \ and\ \ if\ s'\neq 0, k'\ can\ not\ be\ k$$

and also the conditional probability of a total of l crossovers within the second marker interval among the offspring with the i^{th} phenotype, ω_{il}, which is given by

$$\omega_{il} = \sum_{k'=0}^{l} \left\{ \frac{1}{4} \sum_{g \in i} \sum_{s=0}^{1} \sum_{t=0}^{1} \sum_{s'=0}^{1} \sum_{t'=0}^{1} z_{g,st/s't'} b_{st}^{k'} b_{s't'}^{k-k'} \Big/ f_i \right\} \quad \text{if } t \neq 0, k' \text{ can not be } 0 \text{ and if } t' \neq 0, k' \text{ can not be } k \tag{II-1.22}$$

Since each strand made up of the four-strand bundle has a chance of ½ to be involved in each crossover in diploids, the M step updates the estimates of the genetic distances, which are defined as the expected number of crossovers occurring on a single chromatid within that interval, from

$$d_1^{'} = \frac{1}{2} \left(\sum_{i=1}^{M} n_i \sum_{k \geq 1} k \gamma_{ik} \Big/ 2n \right) \tag{II-1.23}$$

$$d_2^{'} = \frac{1}{2} \left(\sum_{i=1}^{M} n_i \sum_{k \geq 1} k \omega_{ik} \Big/ 2n \right) \tag{II-1.24}$$

The likelihood function increases as the E step and M step repeated and the parameter estimates converge to the MLEs conditional for a given integer value of parameter, *m*. We then infer the most likely index of crossover interference, *m*, as that maximies the likelihood function.

1.3. Methods of inferring crossover interference for zygote in autotetraploids

The Chi-square model, which was originally developed for diploid species, has actually two key features. First, it models distribution of crossover and non-crossover events along paired

chromosomal bundles without any limitation on the number of homologous chromosomes pairing to constitute the bundles. Second, recombination interference is modelled in term of distance between two consecutive crossover events. Obviously, these basic features are not only inherently specific to diploids. Thus, we propose here to extend basic idea of the Chi-square model to autotetraploids by properly accounting for the essential features of gene segregation and recombination in meiosis of the complicated species.

In meiosis of autotetraploids, homologous chromosomes may pair into two different patterns as mentioned before, bivalent pairing and quadrivalent pairing. In the former, four chromosomes are randomly paired into two pairs, each making up of two homologous chromosomes, and crossovers occur only between the paired chromosomes. Thus, under bivalent pairing, tetraploid genes show the disomic inheritance like diploids. While homologous chromosomes pair together forming a quadrivalent pair, the crossover may occur between any pair of the homologous chromosomes. When recombination occurs between the centromere and a marker locus, duplicated sister chromatids at the marker locus may enter the same gamete during meiosis. This is the phenomenon of so called double reduction, which is the key feature of the tetrasomic inheritance of autotetraploids.

In the autotetraploid chi-square model, we consider three marker loci, A, B and C, along the chromosome of an autotetraploid species. Four parameters are needed to specify the three-loccus model, *m* indicating the number of non-crossovers (C_o) between two crossovers (C_x), α being the coefficient of double reduction, and d_1 and d_2 being genetic distances of the first and second marker interval respectively. To make the analysis of recombination interference comparable

between diploids and autotetraploids, genetic distance is also defined as the expected number of crossovers occurring on a single chromatid within the given interval.

1.3.1. Prediction of probability distribution of crossover occurring within marker intervals, $\Pr\{k_1, k_2 | m, d_1, d_2\}$

Firstly of all, it is necessary to find a proper probability distribution to describe crossover events occurring along the chromosomes. As mentioned before, the Chi-square model proposed by Zhao et al (1995) models distribution of crossover and non-crossover events along paired chromosomal bundles without any limitation on the number of homologous chromosomes pairing to constitute the bundles. Thus I can apply the basic principle of chi-square model to model crossover events distribution along chromosomes in auotetraploids as follows.

In the absence of chromatid interference, each strand made up of the four-strand bundle during meiosis has an equal chance of of 1/2 to be involved in any given crossover in autotetraploids with bivalent pairing during meiosis. Let $p = m+1$ be length of a complete set of C events, $C_x(C_0)^m$. Among s C events, each strand will be involved in an expected number of $s/(2p)$ crossovers. According to the definition of genetic distance here, the average number of C events (y_1 or y_2) occurring within any given interval can be expressed in terms of genetic distance of the interval (d_1 or d_2), for example $y_1 = 2pd_1$ (or $y_2 = 2pd_2$). However, in autotetraploids with quadrivalent pairing during meiosis, each strand made up of the eight-strand bundle during

meiosis has an equal chance of of 1/4 to be involved in any given crossover. Accordingly, the average number of *C* events (y_1 or y_2) occurring within any given interval can be expressed in terms of genetic distance of the interval (d_1 or d_2), for example $y_1 = 4pd_1$ (or $y_2 = 4pd_2$) in autotetraploids with quadrivalent pairing.

The occurrence of *k* crossovers (C_x) between two flanking markers A and B on a pairing bundle might be the result of p^2 possible situations as shown in Equation (1.1) and Equation (1.2) for $k \geq 1$ and $k = 0$ in Section 1.2.3. By assuming the *C* events are randomly distributed (Poisson distribution) along the chromosome and the first *C* event has an equal chance of being any of the *m*+1 elements of $C_x(C_0)^m$, the probability of *k* crossovers between markers A and B can be computed by summing over probabilities of all the p^2 possibilities as

$$\begin{cases} \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p} Q_{y_1}^{ijk} & if\ k\ \geq 1 \\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p} Q_{y_1}^{oij} & if\ k\ = 0 \end{cases} \qquad \text{(II-1.25)}$$

where $Q_{y_1}^{ijk} = e^{-y} y_1^{\,pk+j-i} / (pk+j-i)!$ and $Q_{y_1}^{oij} = \begin{cases} e^{-y} y_1^{\,j-i} / (j-i)! & if\ i \leq j \\ 0 & if\ i > j \end{cases}$. Here y_1 is mean parameter of Poisson distribution, with $y_1 = 2pd_1$ for bivalent pairing and $y_1 = 4pd_1$ for quadrivalent pairing.

Similarly, as introduced in Section 1.2.3, analysis formulated above for two markers can be extended to that with three markers A, B and C. Consequently, the probability of k_1 crossovers

occurring between markers A and B, and k_2 crossovers between markers B and C along one pairing bundle of chromosomes is given by

$$\frac{1}{p} I \cdot D_{k_1}(y_1) \cdot D_{k_2}(y_2) \cdot I' \tag{II-1.26}$$

where the details of matrix $D_{k_1}(y_1)$ and $D_{k_2}(y_2)$ can be found in Equation (II-1.3) and (II-1.4) in Section 1.2.3, with $y_1 = 2pd_1$ for bivalent pairing and $y_1 = 4pd_1$ for quadrivalent pairing.

In autotetraploids with bivalent pairing during meiosis, there are two sets of paired chromosomes, $\Pr\{k_1, k_2 | m, d_1, d_2\}$ can be calculated by noting independence of the crossover events between the created bivalents as

$$\frac{1}{p^2} \sum_{k_1'=0}^{k_1} \sum_{k_2'=0}^{k_2} \left[I \cdot D_{k_1'}(y_1) \cdot D_{k_2'}(y_2) \cdot I' \right] \left[I \cdot D_{k_1 - k_1'}(y_1) \cdot D_{k_2 - k_2'}(y_2) \cdot I' \right] \tag{II-1.27}$$

where $y_l = 2pd_l$ for bivalent pairing, k_l' indicates the number of crossovers occurring on the first chromatid among the k_l crossover events.

In autotetraploids with quadrivalent pairing during meiosis, all the chromosomes paired together to form one pairing bundle. The probability of k_1 crossovers occurring between markers A and B, and k_2 crossovers between markers B and C along chromosomes, $\Pr\{k_1, k_2 | m, d_1, d_2\}$, is given by:

$$\frac{1}{p} I \cdot D_{k_1}(y_1) \cdot D_{k_2}(y_2) \cdot I' \tag{II-1.28}$$

where $y_l = 4pd_l$.

1.3.2. Prediction of diploid gamete mode distribution, $C_{i_1 i_2 / j_1 j_2}^{I_\alpha, I_\beta, e_1, e_2}$

Given the probability distribution of crossover events along the three marker loci, we can proceed to calculate the probability distribution of gamete mode, $C_{i_1 i_2 / j_1 j_2}^{I_\alpha, I_\beta, e_1, e_2}$, under bivalent pairing or quadrivalent pairing during meiosis. Here I_α or I_β equals to 1 or 0, indicating the presence or absence of double reduction for the first marker of the first or second interval. e_1 and e_2 indicate the e_1^{th} and e_2^{th} two-locus gamete mode corresponding to the recombination configuration of the gamete $i_1 \quad i_2 / j_1 \quad j_2$. In autotetraploids, we denote a recombination configuration of a diploid gamete by $C = (i_1 \quad i_2 / j_1 \quad j_2)$, which is made up of recombination configuration $Y_1 = (i_1 / j_1)$ within the first marker interval of the two chromatids in the gamete and recombination configuration $Y_2 = (i_2 / j_2)$ within the second marker interval of the two chromatids in the gamete. Here, i_1 (or i_2) takes a value of 1 if recombination occurs in the first (or second) chromosomal interval on one strand, and zero otherwise. Similarly, j_1 (or j_2) indicates recombination events in the first (or second) chromosomal intervals on the other strand. Let $o_{i_1 / j_1, e_1}^{I_\alpha, k_1}$ denotes the probability of $G_{i_1 / j_1, e_1}^{I_\alpha}$, the e_1^{th} two-locus gamete mode with recombination configuration of $Y_1 = (i_1 / j_1)$ given k_1 crossovers occurring in the first marker interval. Similarly,

$o_{i_2/j_2,e_2}^{I_\beta,k_2}$ denotes the probability of $G_{i_2/j_2,e_2}^{I_\beta}$, the e_2^{th} two-locus gamete mode with recombination configuration of $Y_2=(i_2/j_2)$ given k_2 crossovers occurring in the second marker interval.

For bivalent pairing, there are four different two-locus gamete modes as summarized in Table II-1.2. It is clear to see if no crossovers occur, all marker intervals will be non-recombinants. Mather (1935) demonstrated that under the assumption of no chromatid interference, if there are $k\geq 1$ crossovers between two markers, then the probability for the two markers to be recombinant on any given single strand is expected to be ½ for bivalents. In autotetraploids with bivalent pairing, a total of *k* crossovers within a marker interval will be assigned across the two bivalents with *k'* crossovers on one bivalent and *k*-*k'* crossovers on the other bivalent, as shown in Equation (II-1.27). Thus, the frequency of two- locus gamete modes, $o_{i/j,e}^{I_\alpha,k}$, depends not only on *k* but also on *k',* as summarized in Table II-1.2.

Table II-1.2. Two-locus gamete modes under bivalent pairing in an autotetraploid meiosis

Recombination configuration (Y)	The e^{th} gamete mode	Gamete modes $(1 \le i, j, k, l \le 4)$	Frequency of two-locus gamete mode given k crossovers occurring			
			$k = 0$	$k \ge 1$		
				$k' = 0$	$k' = k$	$k' \ne 0$ and $k' \ne k$
0/0	1	A_iB_i / A_jB_j	$o_{0/0,1}^{0,k}$	1	1/2	1/2	1/4
0/1	1	A_iB_i / A_jB_l	$o_{0/1,1}^{0,k}$	0	1/2	0	1/4
1/0	1	A_iB_k / A_jB_j	$o_{1/0,1}^{0,k}$	0	0	1/2	1/4
1/1	1	A_iB_k / A_jB_l	$o_{1/1,1}^{0,k}$	0	0	0	1/4

Here k' indicates the number of crossovers occurring on the first chromatid among the k crossover events.

For quadrivalent pairing, there are five different two-locus gamete modes involving double-reduction and ten different two-locus gamete modes involving no-double-reduction, as summarized in Table II-1.3.

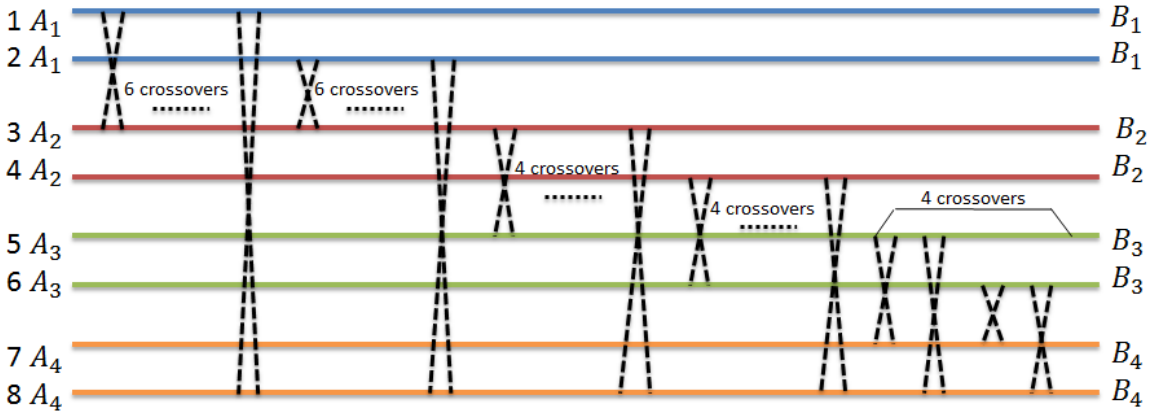
Table II-1.3. Two-locus gamete modes under quadrivalent pairing with or without double reduction in an autotetraploid meiosis

Double Reduction ($I_\alpha = 1$)			
Recombination configuration (Y)	The e^{th} gamete mode	Gamete modes ($1 \le i, j, k, l \le 4$)	Frequency of two-locus gamete mode given k crossovers occurring
0/0	1	A_iB_i / A_iB_i	$o_{0/0,1}^{1,k}$
0/1	1	A_iB_i / A_iB_j	$o_{0/1,1}^{1,k}$
1/0	1	A_iB_j / A_iB_i	$o_{1/0,1}^{1,k}$
1/1	1	A_iB_j / A_iB_j	$o_{1/1,1}^{1,k}$
	2	A_iB_j / A_iB_k	$o_{1/1,2}^{1,k}$
No Double Reduction ($I_\alpha = 0$)			
Recombination configuration (Y)	The e^{th} gamete mode	Gamete modes ($1 \le i, j, k, l \le 4$)	Frequency of two-locus gamete mode given k crossovers occurring
0/0	1	A_iB_i / A_jB_j	$o_{0/0,1}^{0,k}$
0/1	1	A_iB_i / A_jB_i	$o_{0/1,1}^{0,k}$
	2	A_iB_i / A_jB_k	$o_{0/1,2}^{0,k}$
1/0	1	A_iB_j / A_jB_j	$o_{1/0,1}^{0,k}$
	2	A_iB_k / A_jB_j	$o_{1/2,2}^{0,k}$
1/1	1	A_iB_j / A_jB_i	$o_{1/1,1}^{0,k}$
	2	A_iB_j / A_jB_k	$o_{1/1,2}^{0,k}$
	3	A_iB_k / A_jB_i	$o_{1/1,3}^{0,k}$
	4	A_iB_k / A_jB_k	$o_{1/1,4}^{0,k}$
	5	A_iB_k / A_jB_l	$o_{1/1,5}^{0,k}$

To calculate the probability distribution of two-locus gamete modes with *k* crossovers occurring under quadrivalent pairing, we have first to calculate the transition probability from gamete modes with *k* crossovers occurring within the interval to gamete modes with *k*-1 crossovers occurring with diagrammatic demonstration as follows.

During meiosis in autotetraploids under quadrivalent pairing, crossovers may occur between any two non-sister chromatids, which leads to twenty-four different crossovers as shown in Figure II-1.2.

Figure II-1.2. Twenty-four different crossovers between any two non-sister chromatids in autotetraploids with quadrivalent pairing



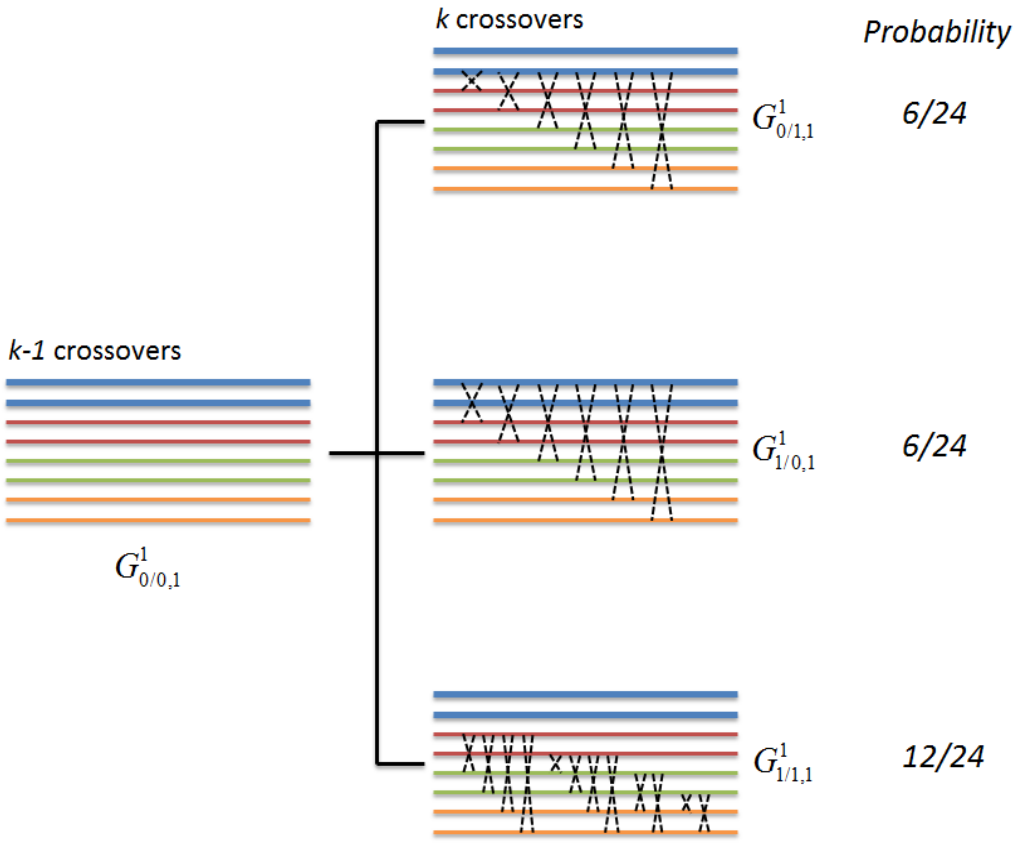
Here the eight solid lines with different colors represent four duplicated chromosomes in autotetraploids, flanking by markers A and B. The black forks in dotted lines indicate twenty-four different crossovers involving any two non-sister chromatids.

As listed in Table II-1.3, there are five different two-locus gamete modes involving double reduction and ten different two-locus gamete modes not involving any double reduction. The expected frequency of a particular gamete mode, $o_{i/j,t}^{I_\alpha,k}$, given *k* crossovers is difficult to calculate directly. However, a recurrent relationship can readily be found for $o_{i/j,t}^{I_\alpha,k}$ of gametes with or without any double reduction gametes by considering the fate of two chromatids that would form the gamete.

1. Double reduction occurs on the first locus of the interval (I_α)

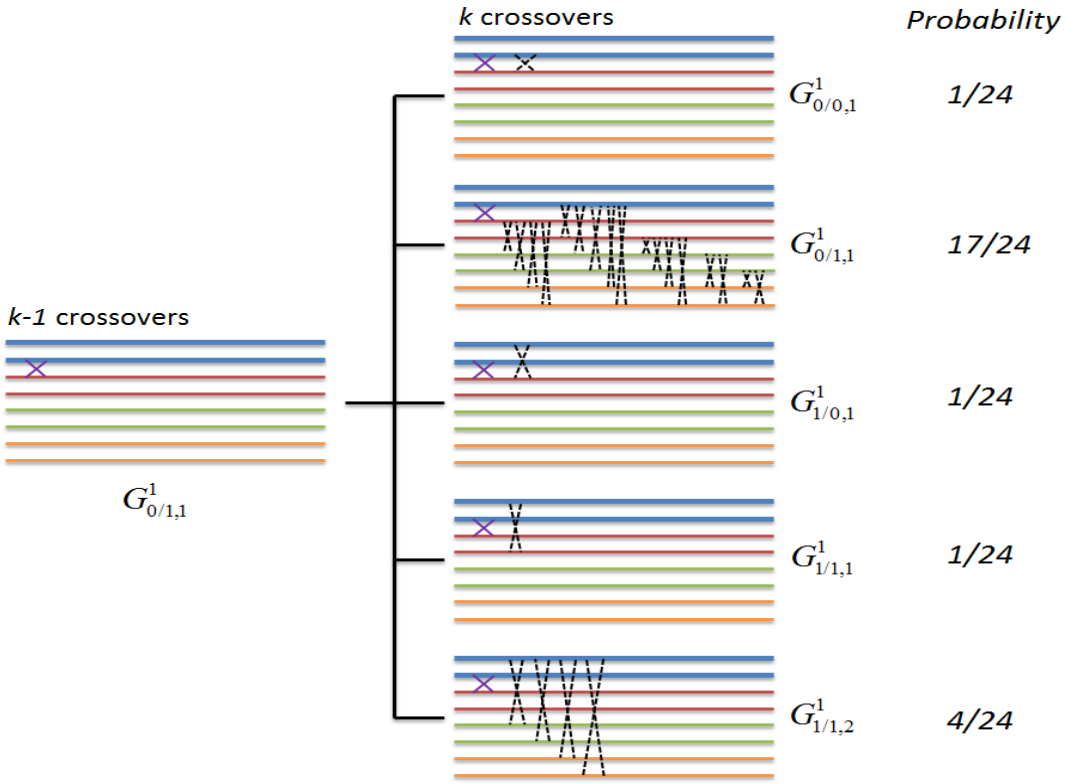
For illustration purposes but without loss of generality, regarding only one pair of A alleles on the sister chromatids, e.g. A_1 on strand 1 and strand 2 in Figure II-1.2, the expected frequency $o_{i/j,t}^{I_\alpha,k}$ is equivalent to the probability of the gamete with two alleles A_1 on strand 1 and strand 2 having the gamete mode $G_{i/j,t}^{1}$ after *k* crossovers occurring between two loci. To obtain the recurrent relationship, we consider the crossover configuration after *k*-1 crossovers between the two loci, and the effect of adding another crossover adjacent to these as follows. Here eight straight lines, indicated by four different colours (blue, red, green and yellow), represent eight duplicated chromosomes during meiosis. The chromosomes with the same colour are the sister chromatids, between which crossing over could not occur. The black dotted forks indicate the k^{th} crossover occurring during meiosis.

1.1.After *k*-1 crossovers, the gamete mode is $G^1_{0/0,1}$



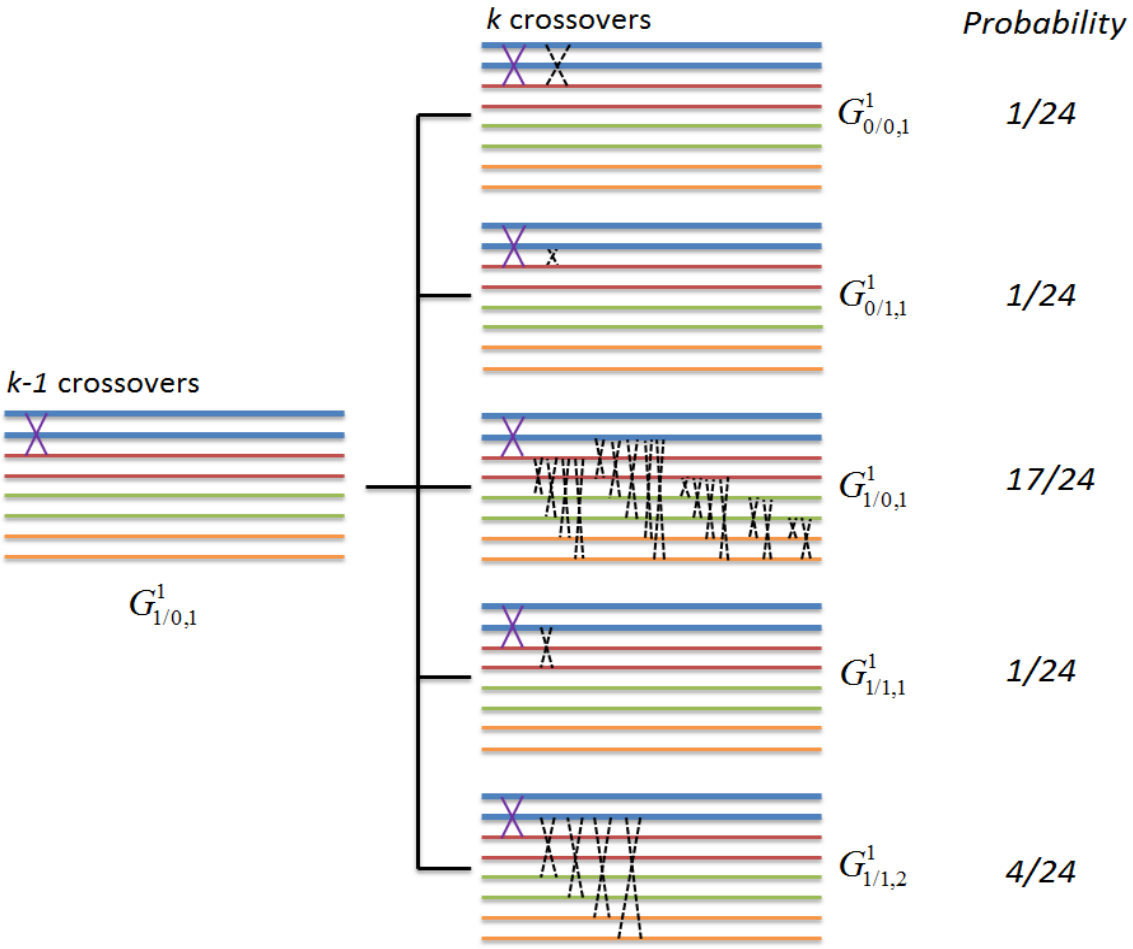
If the gamete mode is $G^1_{0/0,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in three different gamete modes as shown above.

1.2.After *k*-1 crossovers, the gamete mode is $G^1_{0/1,1}$



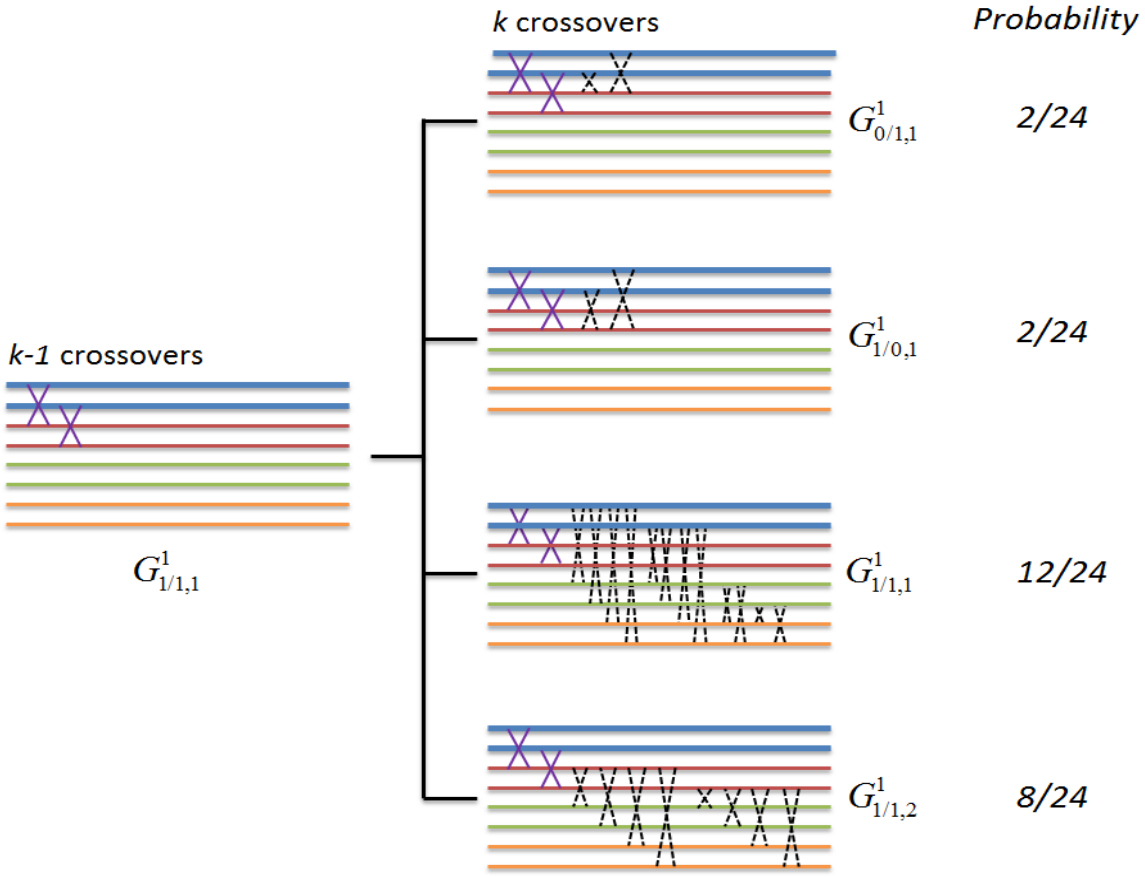
If the gamete mode is $G^1_{0/1,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in five different gamete modes as shown above.

1.3.After *k*-1 crossovers, the gamete mode is $G^1_{1/0,1}$



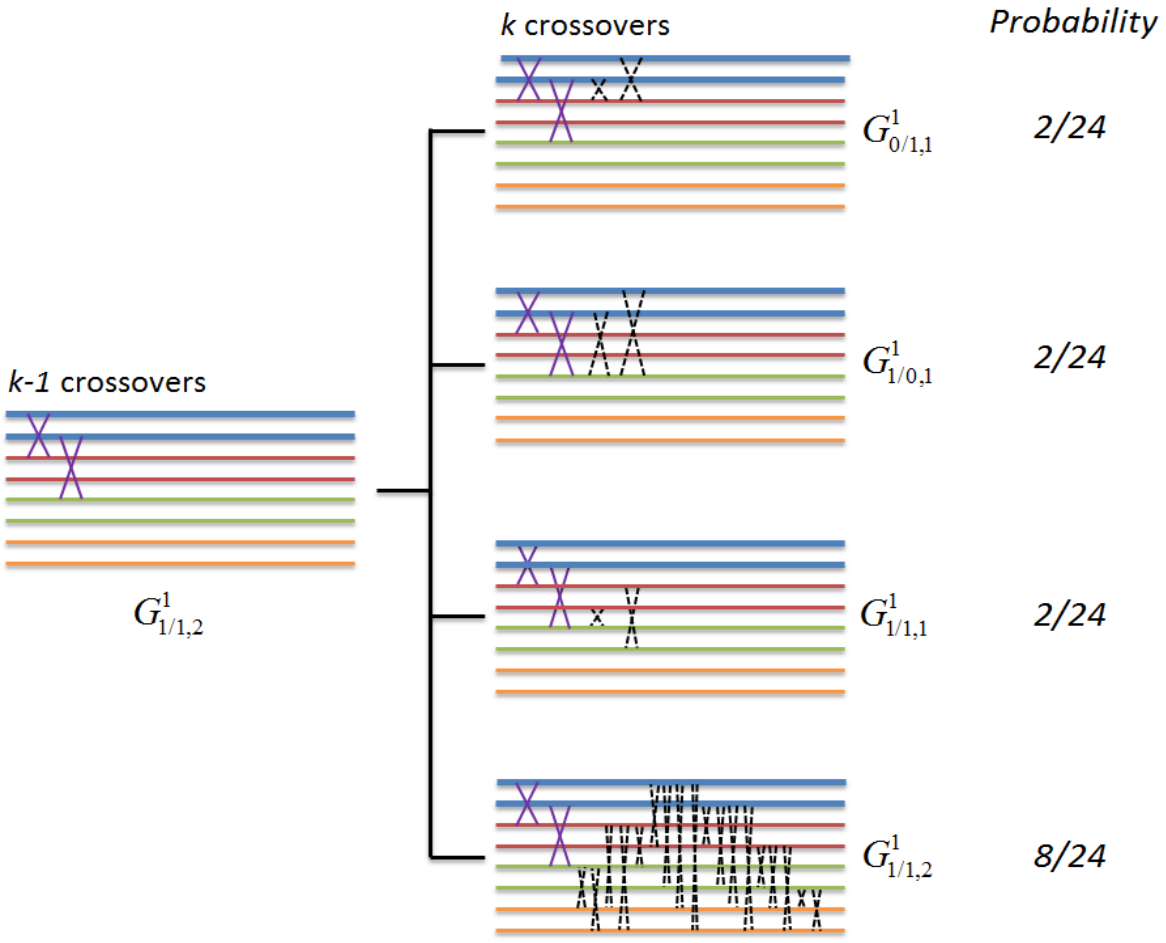
If the gamete mode is $G^1_{1/0,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in five different gamete modes as shown above.

1.4.After *k*-1 crossovers, the gamete mode is $G^1_{1/1,1}$



If the gamete mode is $G^1_{1/1,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in four different gamete modes as shown above.

1.5.After *k*-1 crossovers, the gamete mode is $G^1_{1/1,2}$

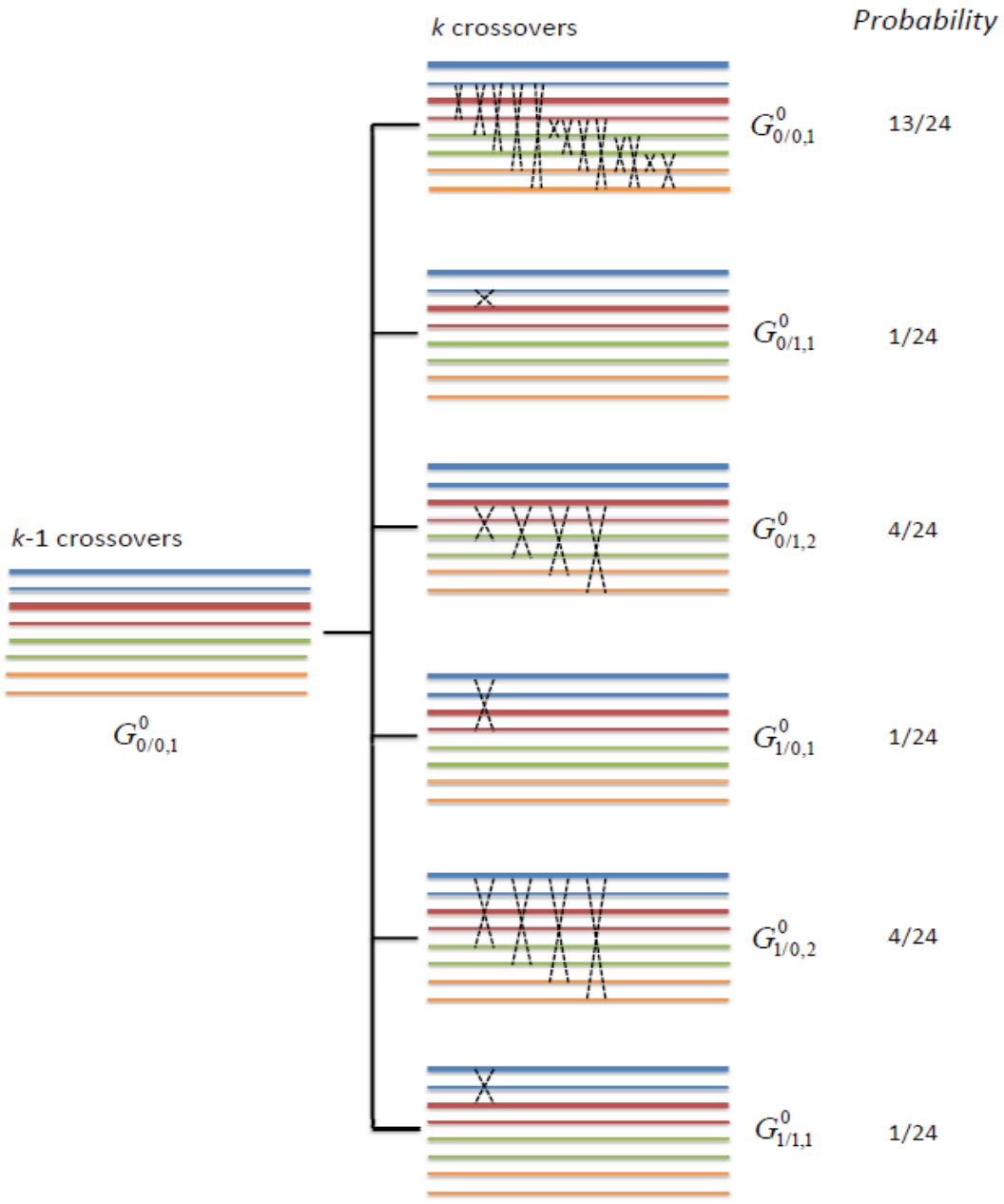


If the gamete mode is $G^1_{1/1,2}$ after *k*-1 crossovers, adding the k^{th} crossover would result in four different gamete modes as shown above.

2. No double reduction occurs on the first locus of the interval (I_α)

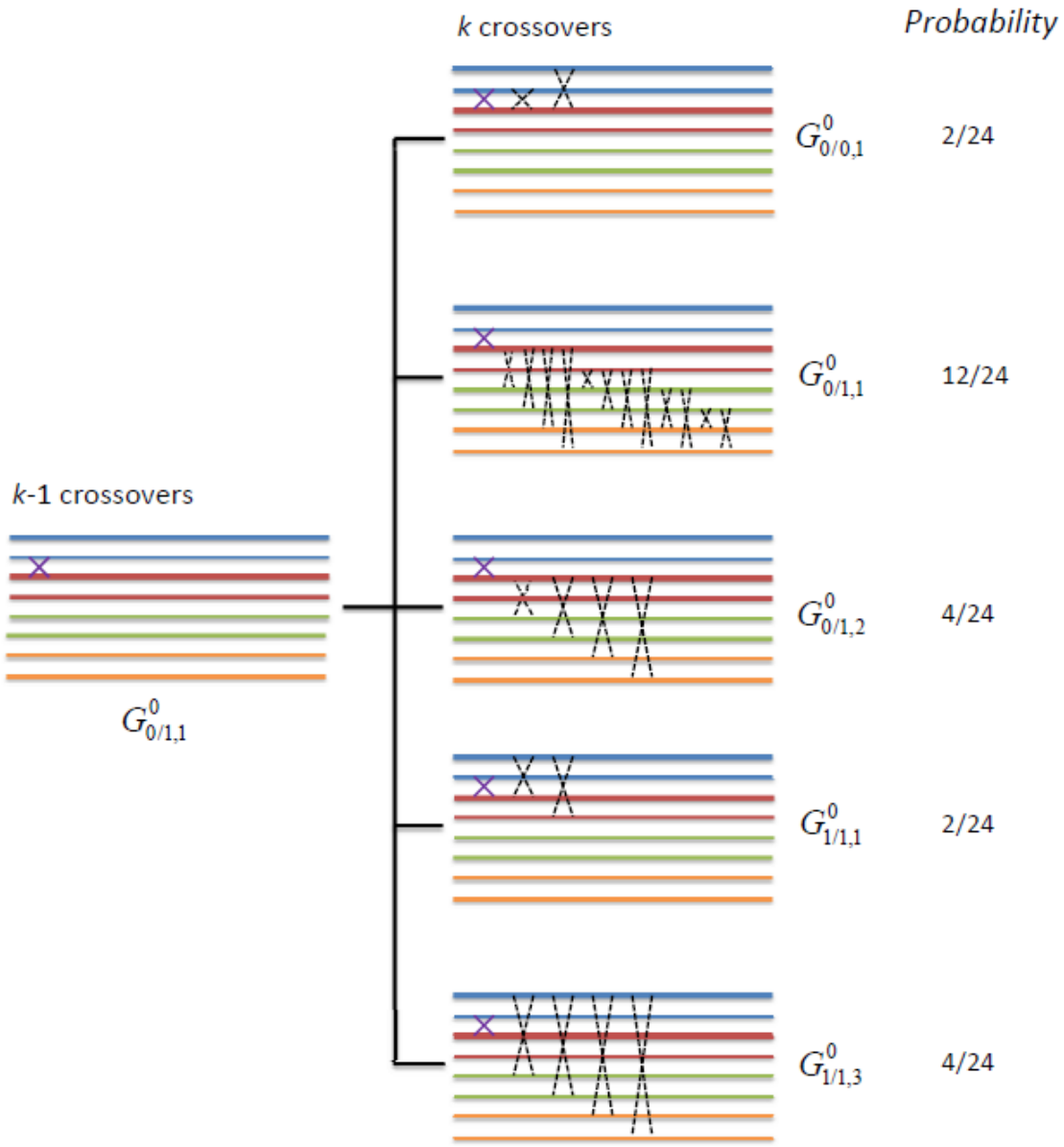
For illustration purposes but without loss of generality, regarding only one pair of alleles on the non-sister chromatids, e.g. A_1 and B_1 on strand 1 and strand 3 in Figure II-1.2, the expected frequency $o_{i/j,t}^{I_\alpha,k}$ is equivalent to the probability of the gamete with two alleles A_1 and B_1 on strand 1 and strand 3 having the gamete mode $G_{i/j,t}^{0}$ after k crossovers occurring between two loci. To obtain the recurrent relationship, we consider the crossover configuration after k-1 crossovers between the two loci, and the effect of adding another crossover adjacent to these as follows. Similarly, here eight straight lines, indicated by four different colours (blue, red, green and yellow), represent eight duplicated chromosomes during meiosis. The chromosomes with the same colour are the sister chromatids, between which crossing over could not occur. The black dotted forks indicate the k^{th} crossover occurring during meiosis.

2.1.After *k*-1 crossovers, the gamete mode is $G^0_{0/0,1}$



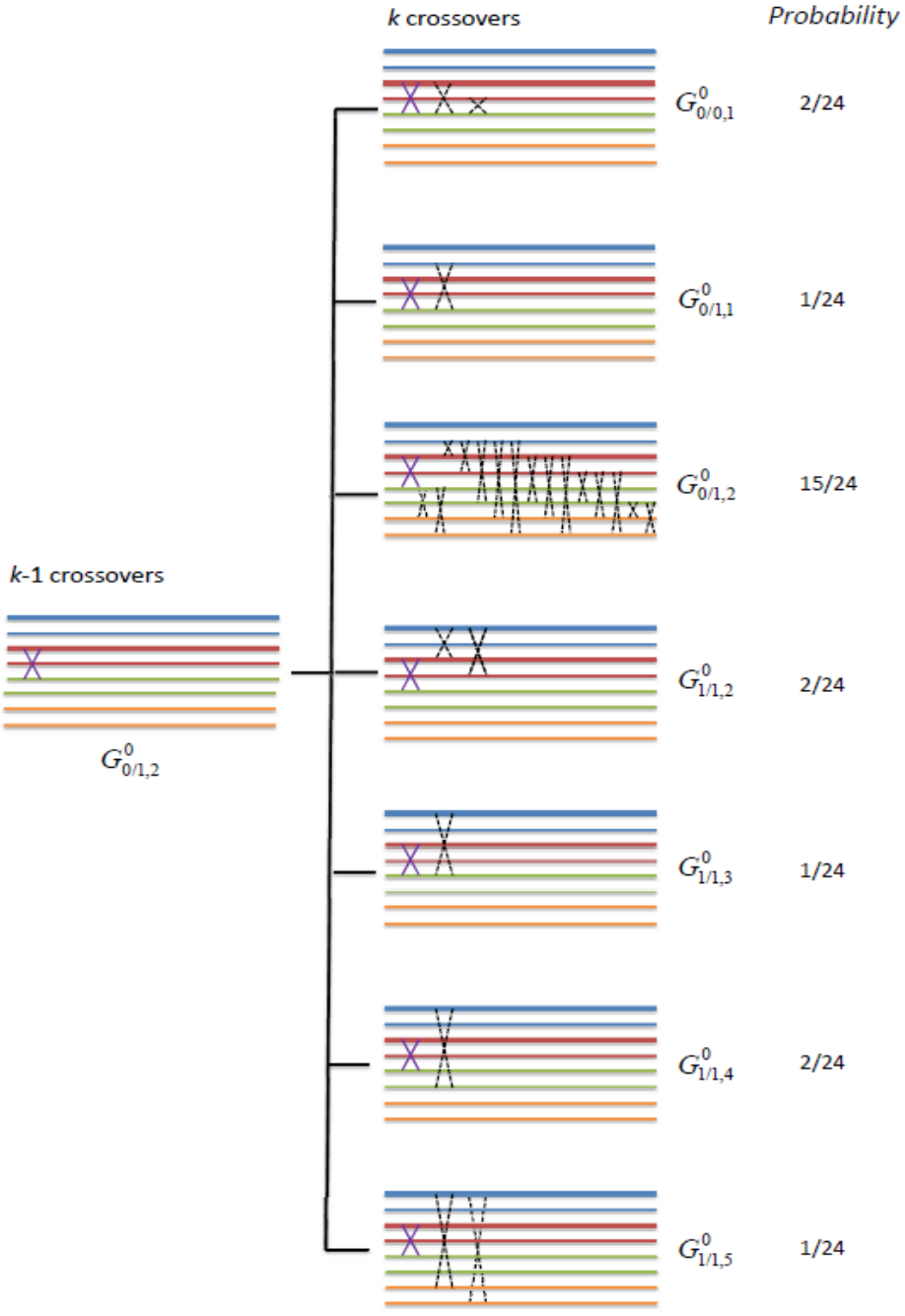
If the gamete mode is $G^0_{0/0,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in six different gamete modes as shown above.

2.2.After *k*-1 crossovers, the gamete mode is $G^0_{0/1,1}$



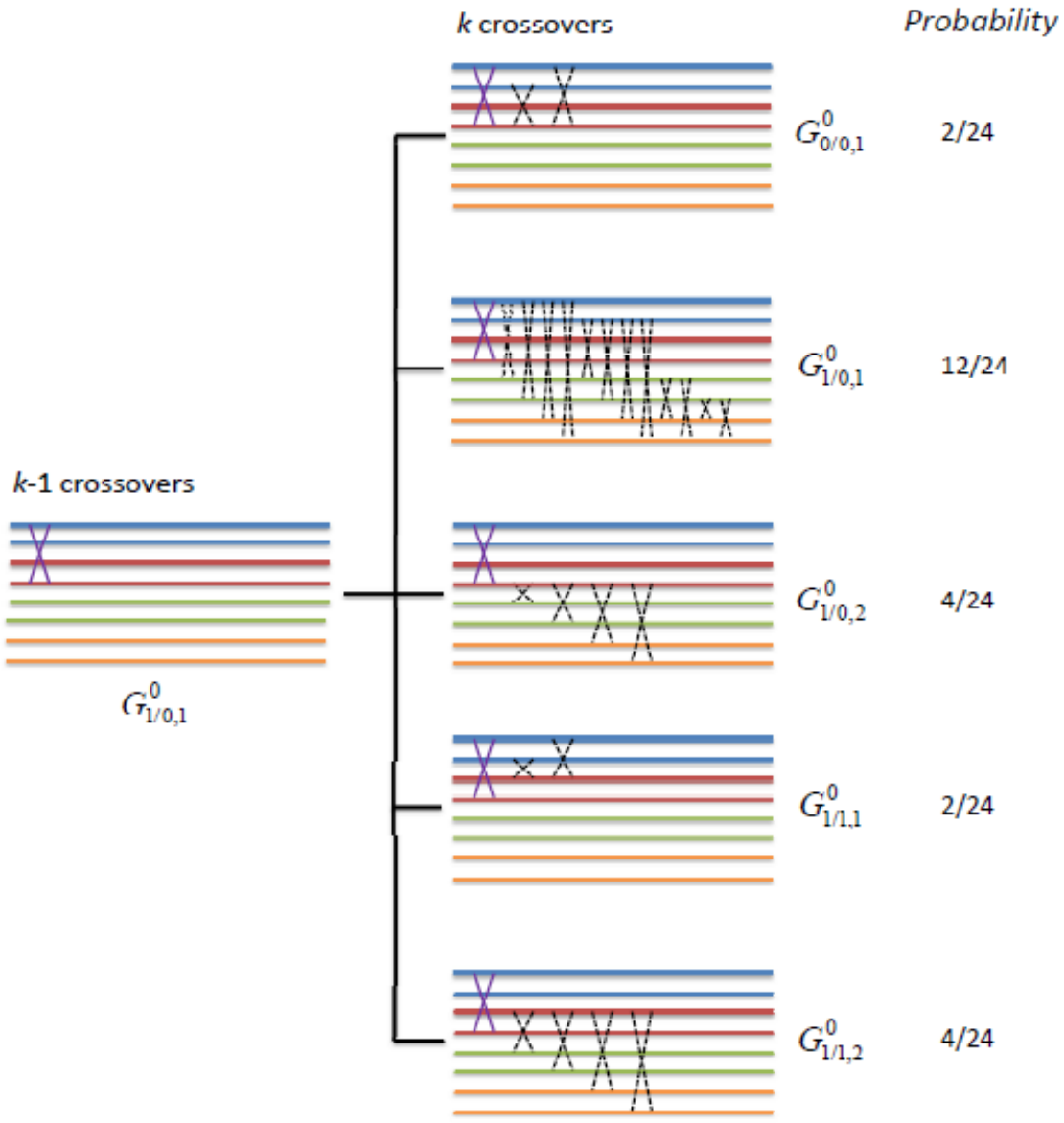
If the gamete mode is $G^0_{0/1,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in five different gamete modes as shown above.

2.3.After *k*-1 crossovers, the gamete mode is $G^0_{0/1,2}$



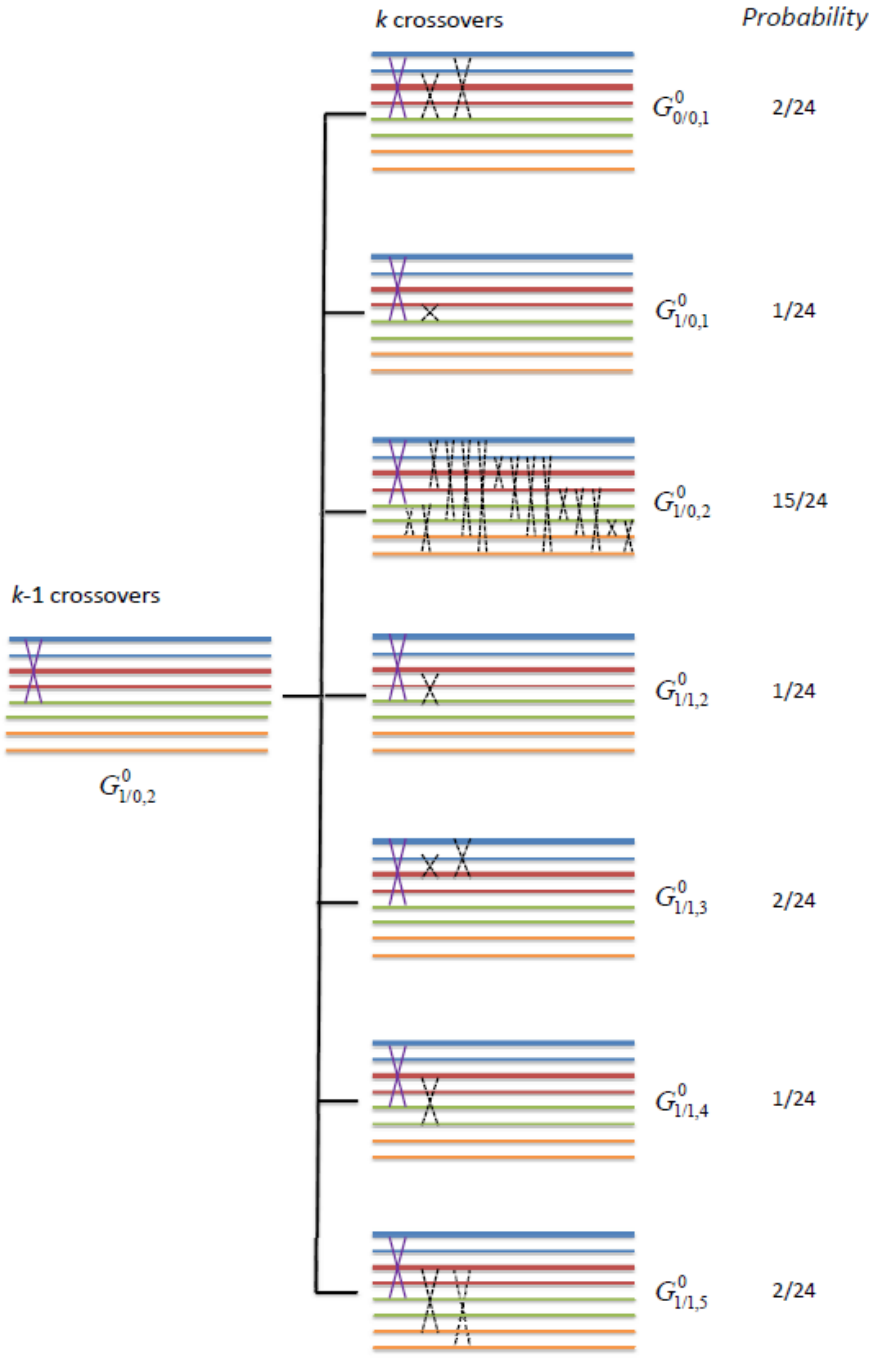
If the gamete mode is $G^0_{0/1,2}$ after *k*-1 crossovers, adding the k^{th} crossover would result in seven different gamete modes as shown above.

2.4.After *k*-1 crossovers, the gamete mode is $G^0_{1/0,1}$



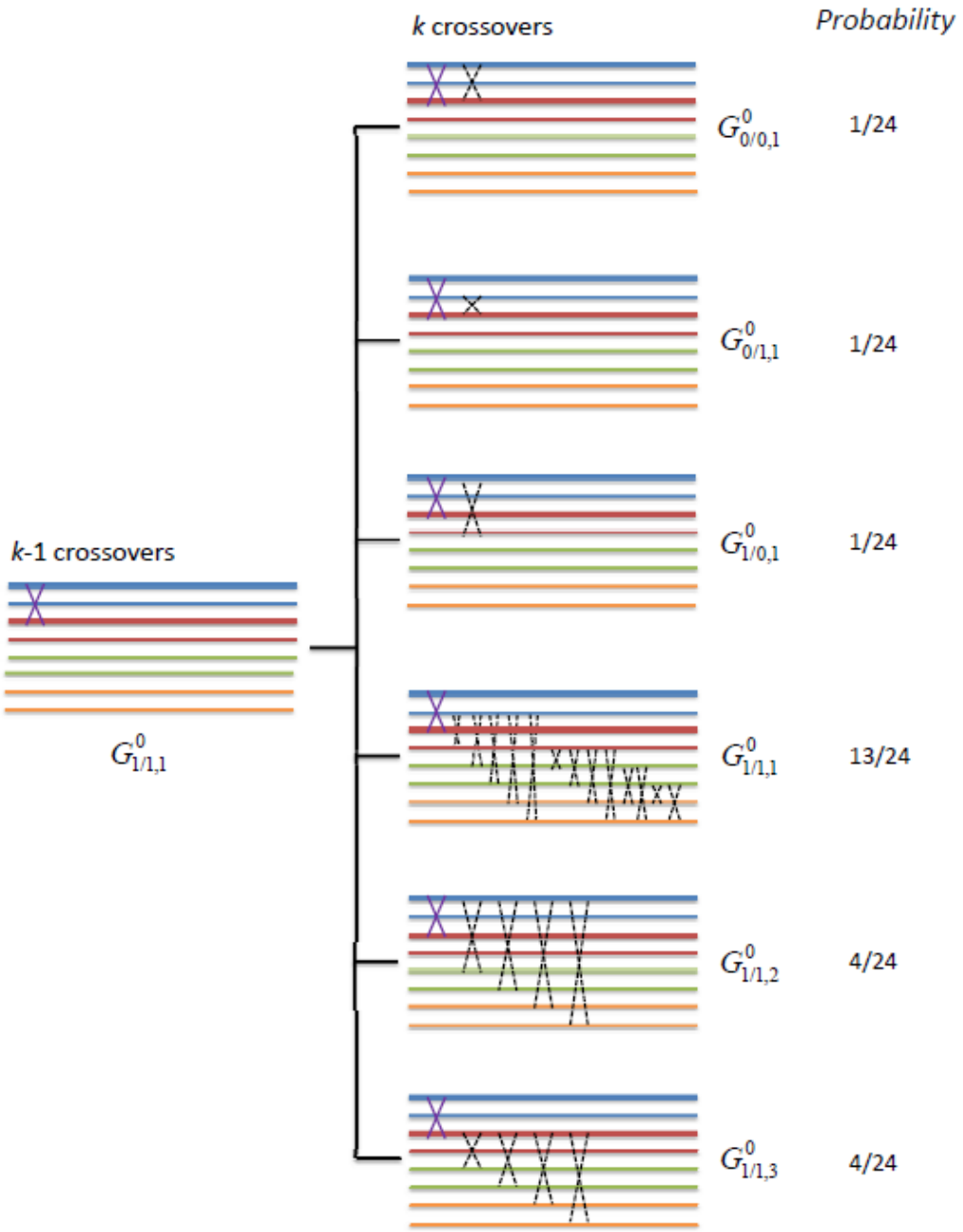
If the gamete mode is $G^0_{1/0,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in five different gamete modes as shown above.

2.5.After *k*-1 crossovers, the gamete mode is $G^0_{1/0,2}$



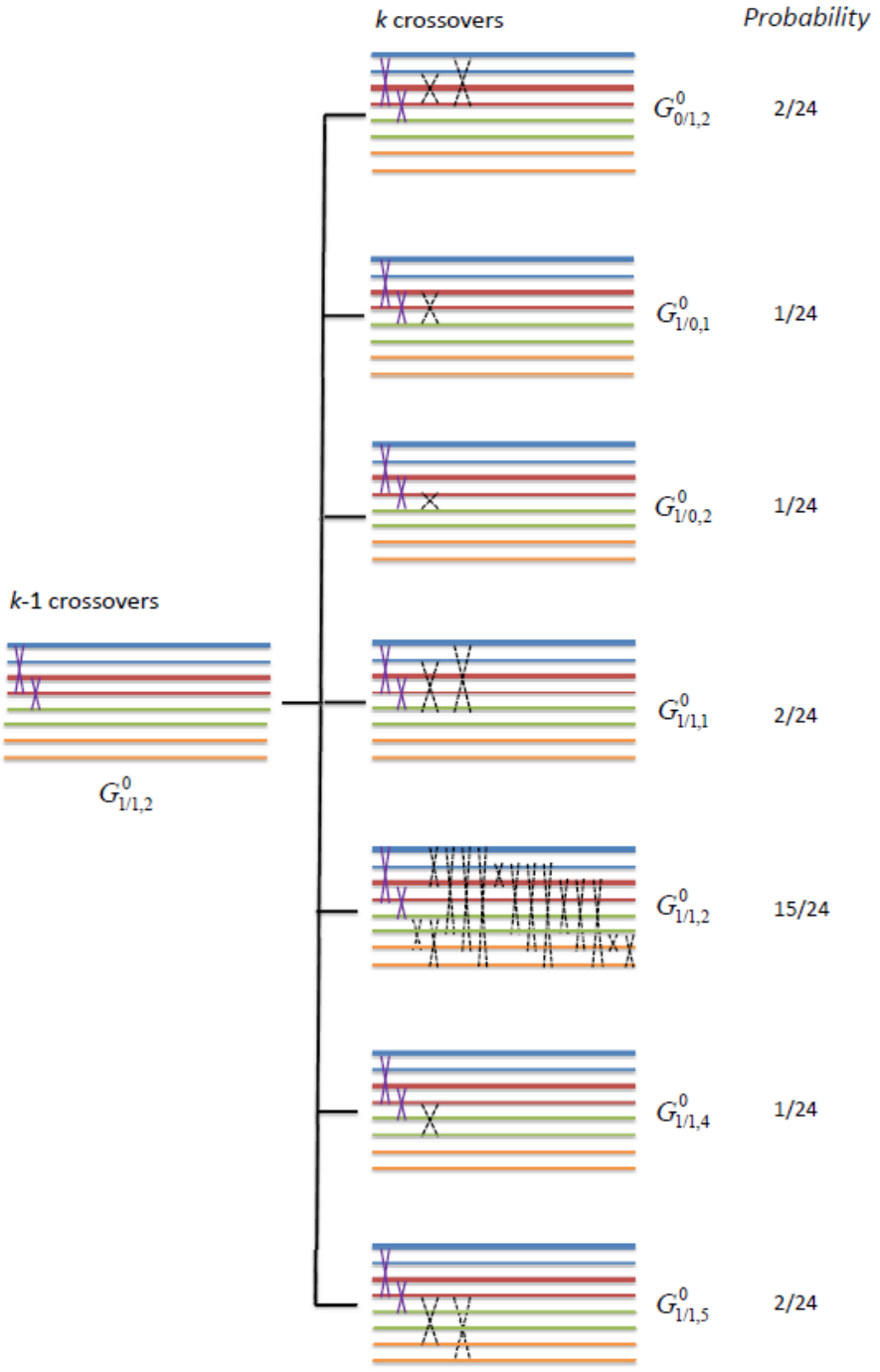
If the gamete mode is $G^0_{1/0,2}$ after *k*-1 crossovers, adding the k^{th} crossover would result in seven different gamete modes as shown above.

2.6.After *k*-1 crossovers, the gamete mode is $G^0_{1/1,1}$



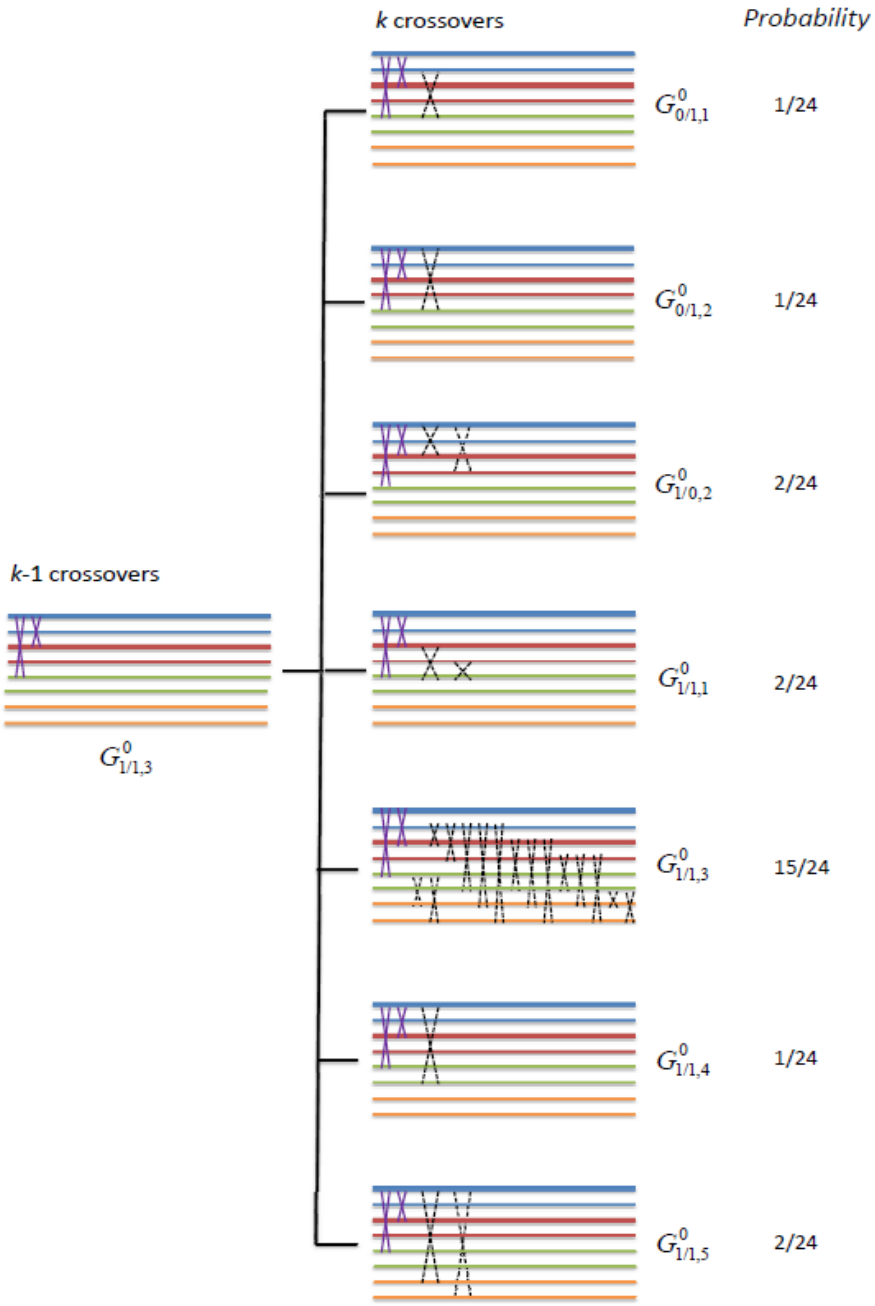
If the gamete mode is $G^0_{1/1,1}$ after *k*-1 crossovers, adding the k^{th} crossover would result in six different gamete modes as shown above.

2.7.After *k*-1 crossovers, the gamete mode is $G^0_{1/1,2}$



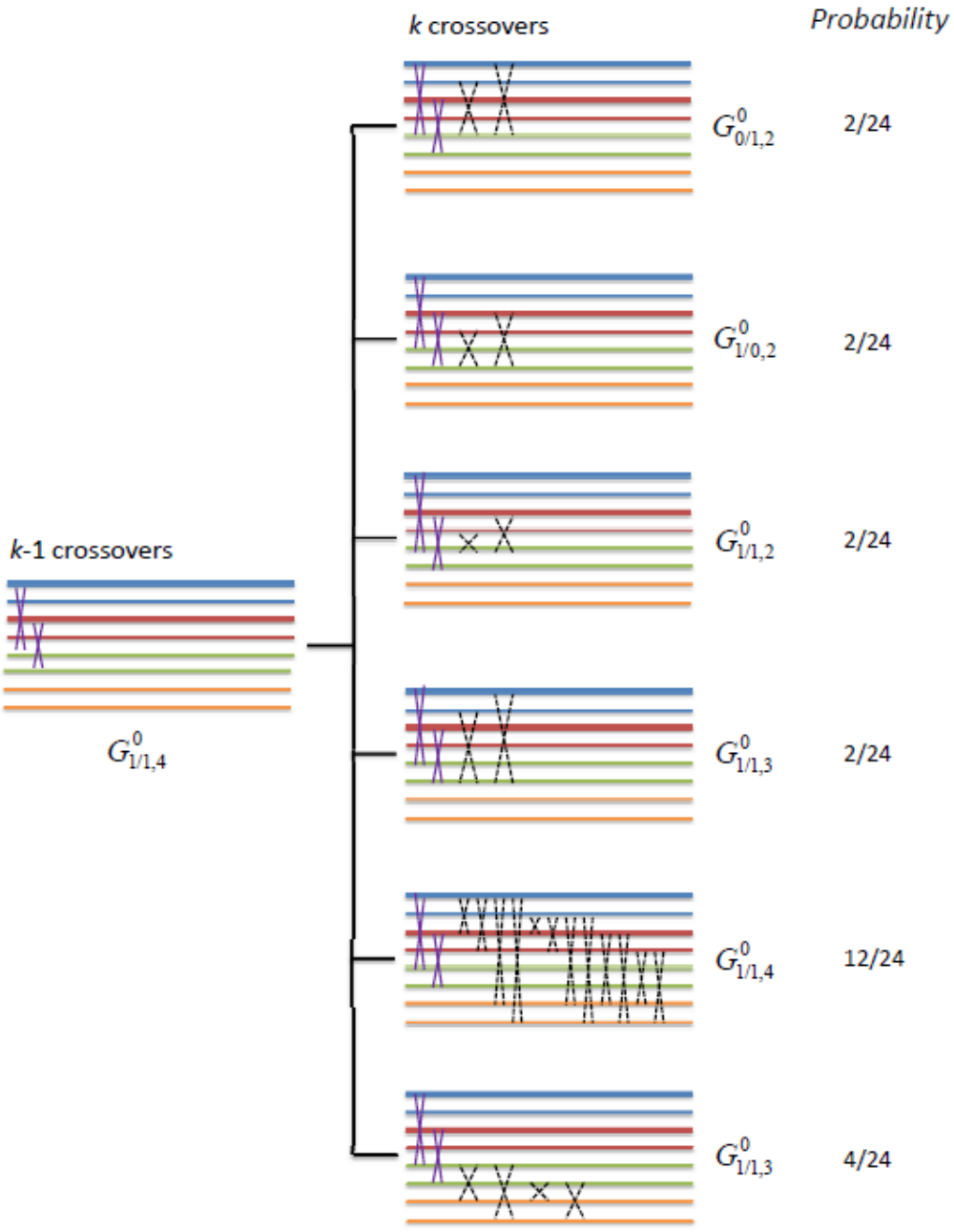
If the gamete mode is $G^0_{1/1,2}$ after *k*-1 crossovers, adding the k^{th} crossover would result in seven different gamete modes as shown above.

2.8.After *k*-1 crossovers, the gamete mode is $G^0_{1/1,3}$



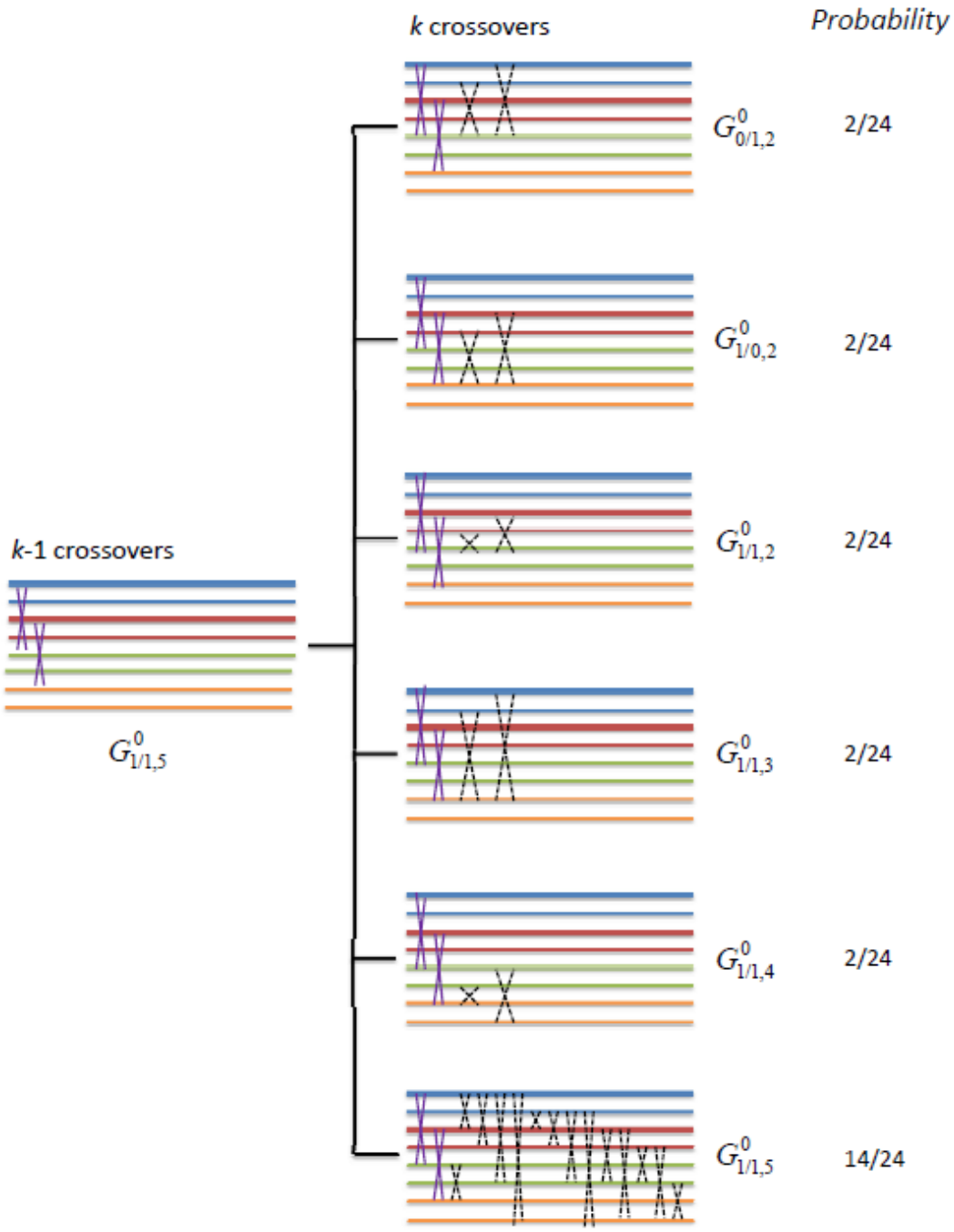
If the gamete mode is $G^0_{1/1,3}$ after *k*-1 crossovers, adding the k^{th} crossover would result in seven different gamete modes as shown above.

2.9.After *k*-1 crossovers, the gamete mode is $G^0_{1/1,4}$



If the gamete mode is $G^0_{1/1,4}$ after *k*-1 crossovers, adding the k^{th} crossover would result in six different gamete modes as shown above.

2.10. After *k*-1 crossovers, the gamete mode is $G^0_{1/1,5}$



If the gamete mode is $G^0_{1/1,5}$ after *k*-1 crossovers, adding the k^{th} crossover would result in six different gamete modes as shown above.

All these diagrams above (1.1-1.5, 2.1-2.10) illustrate the transition probabilities and are summarized in Table II-1.4. Let O_1^k and O_0^k represent the vector for probability distribution of gamete modes after *k* crossovers with or without double reduction on the first marker locus,

$$O_1^k = \begin{bmatrix} o_{0/0,1}^{1,k} & o_{0/1,1}^{1,k} & o_{1/0,1}^{1,k} & o_{1/1,1}^{1,k} & o_{1/1,2}^{1,k} \end{bmatrix}^T \quad \text{(II-1.29)}$$

and $$O_0^k = \begin{bmatrix} o_{0/0,1}^{0,k} & o_{0/1,1}^{0,k} & o_{0/1,2}^{0,k} & o_{1/0,1}^{0,k} & o_{1/0,2}^{0,k} & o_{1/1,1}^{0,k} & o_{1/1,2}^{0,k} & o_{1/1,3}^{0,k} & o_{1/1,4}^{0,k} & o_{1/1,5}^{0,k} \end{bmatrix}^T \quad \text{(II-1.30)}$$

Matrices T_1 and T_0 represent the transition probability from gamete modes with *k*-1 crossovers occurring to gamete modes with *k* crossovers occurring with or without double reduction on the first marker locus, detailed as:

$$T_1 = \begin{bmatrix} 1/2 & 1/24 & 1/24 & 0 & 0 \\ 1/4 & 17/24 & 1/24 & 1/12 & 1/12 \\ 1/4 & 1/24 & 17/24 & 1/12 & 1/12 \\ 0 & 1/24 & 1/24 & 1/2 & 1/12 \\ 0 & 1/6 & 1/6 & 1/3 & 3/4 \end{bmatrix}$$

and

$$T_0 = \begin{bmatrix} 13/24 & 1/12 & 1/12 & 1/12 & 1/12 & 1/24 & 0 & 0 & 0 & 0 \\ 1/24 & 1/2 & 1/24 & 0 & 0 & 1/24 & 0 & 1/24 & 0 & 0 \\ 1/6 & 1/6 & 5/8 & 0 & 0 & 0 & 1/12 & 1/24 & 1/12 & 1/12 \\ 1/24 & 0 & 0 & 1/2 & 1/24 & 1/24 & 1/24 & 0 & 0 & 0 \\ 1/6 & 0 & 0 & 1/6 & 5/8 & 0 & 1/24 & 1/12 & 1/12 & 1/12 \\ 1/24 & 1/12 & 0 & 1/12 & 0 & 13/24 & 1/12 & 1/12 & 0 & 0 \\ 0 & 0 & 1/12 & 1/6 & 1/24 & 1/6 & 5/8 & 0 & 1/12 & 1/12 \\ 0 & 1/6 & 1/24 & 0 & 1/12 & 1/6 & 0 & 5/8 & 1/12 & 1/12 \\ 0 & 0 & 1/24 & 0 & 1/24 & 0 & 1/24 & 1/24 & 1/2 & 1/12 \\ 0 & 0 & 1/12 & 0 & 1/12 & 0 & 1/12 & 1/12 & 1/6 & 7/12 \end{bmatrix}$$

Table II-1.4. Transition probability from gamete modes with *k-1* crossovers occurring to gamete modes with *k* crossovers occurring

Double reduction on the first marker locus

$k-1$ \ k		0/0 A_iB_i / A_iB_i	0/1 A_iB_i / A_iB_j	1/0 A_iB_j / A_iB_i	1/1 A_iB_j / A_iB_j	1/1 A_iB_j / A_iB_k
0/0	A_iB_i / A_iB_i	1/2	1/4	1/4	0	0
0/1	A_iB_i / A_iB_j	1/24	17/24	1/24	1/24	1/6
1/0	A_iB_j / A_iB_i	1/24	1/24	17/24	1/24	1/6
1/1	A_iB_j / A_iB_j	0	1/12	1/12	1/2	1/3
	A_iB_j / A_iB_k	0	1/12	1/12	1/12	3/4

No double reduction on the first marker locus

$k-1$ \ k		0/0 A_iB_i / A_jB_j	0/1 A_iB_i / A_jB_i	0/1 A_iB_i / A_jB_k	1/0 A_iB_j / A_jB_j	1/0 A_iB_k / A_jB_j	1/1 A_iB_j / A_jB_i	1/1 A_iB_j / A_jB_k	1/1 A_iB_k / A_jB_i	1/1 A_iB_k / A_jB_k	1/1 A_iB_k / A_jB_l
0/0	A_iB_i / A_jB_j	13/24	1/24	1/6	1/24	1/6	1/24	0	0	0	0
0/1	A_iB_i / A_jB_i	1/12	1/2	1/6	0	0	1/12	0	1/6	0	0
	A_iB_i / A_jB_k	1/12	1/24	5/8	0	0	0	1/12	1/24	1/24	1/12
1/0	A_iB_j / A_jB_j	1/12	0	0	1/2	1/6	1/12	1/6	0	0	0
	A_iB_k / A_jB_j	1/12	0	0	1/24	5/8	0	1/24	1/12	1/24	1/12
1/1	A_iB_j / A_jB_i	1/24	1/24	0	1/24	0	13/24	1/6	1/6	0	0
	A_iB_j / A_jB_k	0	0	1/12	1/24	1/24	1/12	5/8	0	1/24	1/12
	A_iB_k / A_jB_i	0	1/24	1/24	0	1/12	1/12	0	5/8	1/24	1/12
	A_iB_k / A_jB_k	0	0	1/12	0	1/12	0	1/12	1/12	1/2	1/6
	A_iB_k / A_jB_l	0	0	1/12	0	1/12	0	1/12	1/12	1/12	7/12

Then the probability distribution of gamete modes after *k* crossovers occurring within the marker interval under quadrivalent pairing in autotetraploids can be calculated as

$$O_1^k = \left(T_1\right)^k \cdot O_1^0 \tag{II-1.31}$$

and $$O_0^k = \left(T_0\right)^k \cdot O_0^0 \tag{II-1.32}$$

where $O_1^0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}^T$ and $O_0^0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$.

It can be seen from the demonstration above that the probability distribution of two-locus gamete modes is only related to the double reduction event on the first marker locus and the number of crossover events within the marker interval. Thus the probability distributions of two-locus gamete modes of any pairs of markers are independent with each other. Take three linked loci into consideration, the probability distribution of gamete modes on three loci, $G_{i_1 i_2 / j_1 j_2, e_1, e_2}^{I_\alpha, I_\beta}$, can be calculated as

$$C_{i_1 i_2 / j_1 j_2}^{I_\alpha, I_\beta, e_1, e_2} = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} o_{i_1 / j_1, e_1}^{I_\alpha, k_1} o_{i_2 / j_2, e_2}^{I_\beta, k_2} \Pr\{k_1, k_2 \,|\, m, d_1, d_2\} \tag{II-1.33}$$

where $\Pr\left\{k_1, k_2 \,|\, m, d_1, d_2\right\}$ can be obtained from Equation (II-1.27) for bivalent pairing and Equation (II-1.28) for quadrivalent pairing, $o_{i_1 / j_1, e_1}^{I_\alpha, k_1}$ and $o_{i_2 / j_2, e_2}^{I_\beta, k_2}$ can be obtained in Table II-1.2 for bivalent pairing and in Equation (II-1.31) and (II-1.32) for quadrivalent pairing.

1.3.3. Prediction of probability distribution of marker phenotypes at three loci, f_i

For simplicity but without loss of generality, a three-locus genotype of an autotetraploid parent can be represented by $A_1B_1C_1/A_2B_2C_2/A_3B_3C_3/A_4B_4C_4$. During gametogenesis, gametes can be divided into 16 or 125 categories according to different gamete modes for bivalent pairing (or quadrivalent pairing) as summarized Table II-1.5 (or Table II-1.6). A general formula for the frequency of these gametes can be expressed as:

$$g_k = \sigma_k \alpha^{I_{\alpha,k}} (1-\alpha)^{1-I_{\alpha,k}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} \quad \text{(II-1.34)}$$

where σ_k is a constant which equals to 1/12 for bivalent pairing and takes various values such as 36/144, 12/144, 12/144, …, 6/144 as shown in the last column of Table II-1.6 under quadrivalent pairing. α represents the coefficient of double reduction under quadrivalent pairing, while there is no such term for bivalent pairing. $I_{\alpha,k}$ takes the value of 1 or 0, indicating the presence or absence of double reduction on locus A. $C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k}$ can be calculated by Equation (II-1.33)

For any individual with a particular three-locus genotype rather than the generic genotype, the probability of a gamete genotype can be worked out on the basis of Table II-1.5 or Table II-1.6 through the following formulation

$$G_i = \sum_{k \in i} g_k = \sum_{k \in i} \sigma_k \alpha^{I_{\alpha,k}} (1-\alpha)^{1-I_{\alpha,k}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} \quad \text{(II-1.35)}$$

where $k \in i$ indicates the sum over the frequencies of all those gametes *k* that correspond to the same gametic genotype *i*.

Table II-1.5. Probability distribution of the modes of gamete formation and gamete genotypes at three linked loci of an autotetraploid species with bivalent pairing

In the case of bivalent pairing, three equally likely pairs of bivalents can be generated for parent $A_1B_1C_1/A_2B_2C_2/A_3B_3C_3/A_4B_4C_4$, as follows $A_1B_1C_1/A_2B_2C_2 \| A_3B_3C_3/A_4B_4C_4$, $A_1B_1C_1/A_3B_3C_3 \| A_2B_2C_2/A_4B_4C_4$, and $A_1B_1C_1/A_4B_4C_4 \| A_2B_2C_2/A_3B_3C_3$, where $\|$ is used to distinguish paired homologous chromosomes. For a given pairs of bivalents, gamete genotypes at three loci gametes can be classified into 16 categories as follows

Recombination configuration for the first interval (i_1 / j_1)	Recombination configuration for the second interval (i_2 / j_2)	Gametes (i,j,k,l =1,2,3,4)	Frequency	Probabilities	
				Modes	Gametes
0/0	0/0	$A_iB_iC_i/A_jB_jC_j$	4	$C_{00/00}^{0,0,1,1}$	$C_{00/00}^{0,0,1,1}/4$
	0/1	$A_iB_iC_i/A_jB_jC_l$	4	$C_{00/01}^{0,0,1,1}$	$C_{00/01}^{0,0,1,1}/4$
	1/0	$A_iB_iC_k/A_jB_jC_j$	4	$C_{01/00}^{0,0,1,1}$	$C_{01/00}^{0,0,1,1}/4$
	1/1	$A_iB_iC_k/A_jB_jC_l$	4	$C_{01/01}^{0,0,1,1}$	$C_{01/01}^{0,0,1,1}/4$
0/1	0/0	$A_iB_iC_i/A_jB_lC_l$	4	$C_{00/10}^{0,0,1,1}$	$C_{00/10}^{0,0,1,1}/4$
	0/1	$A_iB_iC_i/A_jB_lC_j$	4	$C_{00/11}^{0,0,1,1}$	$C_{00/11}^{0,0,1,1}/4$

	1/0	$A_iB_iC_k/A_jB_lC_l$	4	$C_{01/10}^{0,0,1,1}$	$C_{01/10}^{0,0,1,1}/4$
	1/1	$A_iB_iC_k/A_jB_lC_j$	4	$C_{01/11}^{0,0,1,1}$	$C_{01/11}^{0,0,1,1}/4$
1/0	0/0	$A_iB_kC_k/A_jB_jC_j$	4	$C_{10/00}^{0,0,1,1}$	$C_{10/00}^{0,0,1,1}/4$
	0/1	$A_iB_kC_k/A_jB_jC_l$	4	$C_{10/01}^{0,0,1,1}$	$C_{10/01}^{0,0,1,1}/4$
	1/0	$A_iB_kC_i/A_jB_jC_j$	4	$C_{11/00}^{0,0,1,1}$	$C_{11/00}^{0,0,1,1}/4$
	1/1	$A_iB_kC_i/A_jB_jC_l$	4	$C_{11/01}^{0,0,1,1}$	$C_{11/01}^{0,0,1,1}/4$
1/1	0/0	$A_iB_kC_k/A_jB_lC_l$	4	$C_{10/10}^{0,0,1,1}$	$C_{10/10}^{0,0,1,1}/4$
	0/1	$A_iB_kC_k/A_jB_lC_j$	4	$C_{10/11}^{0,0,1,1}$	$C_{10/11}^{0,0,1,1}/4$
	1/0	$A_iB_kC_i/A_jB_lC_l$	4	$C_{11/10}^{0,0,1,1}$	$C_{11/10}^{0,0,1,1}/4$
	1/1	$A_iB_kC_i/A_jB_lC_j$	4	$C_{11/11}^{0,0,1,1}$	$C_{11/11}^{0,0,1,1}/4$

Table II-1.6. Probability distribution of the modes of gamete formation and gamete genotypes at three linked loci of an autotetraploid species with quadrivalent pairing

Double reduction on locus A (I_α)	Double reduction on locus B (I_β)	Recombination configuration for the first interval (i_1 / j_1)	Recombination configuration for the second interval (i_2 / j_2)	Number		Gametes ($i, j, k, l = 1, 2, 3, 4$)	Frequency	Probability	
				e_1	e_2			Mode	Gametes
1	1	0/0	0/0	1	1	$A_iB_iC_i$ / $A_iB_iC_i$	4	$\alpha C_{00/00}^{1,1,1,1}$	$36\alpha C_{00/00}^{1,1,1,1}/144$
		0/0	0/1	1	1	$A_iB_iC_i$ / $A_iB_iC_j$	12	$\alpha C_{00/01}^{1,1,1,1}$	$12\alpha C_{00/01}^{1,1,1,1}/144$
		0/0	1/0	1	1	$A_iB_iC_j$ / $A_iB_iC_i$	12	$\alpha C_{01/00}^{1,1,1,1}$	$12\alpha C_{01/00}^{1,1,1,1}/144$
		0/0	1/1	1	1	$A_iB_iC_j$ / $A_iB_iC_j$	12	$\alpha C_{01/01}^{1,1,1,1}$	$12\alpha C_{01/01}^{1,1,1,1}/144$
				1	2	$A_iB_iC_j$ / $A_iB_iC_k$	24	$\alpha C_{01/01}^{1,1,1,2}$	$6\alpha C_{01/01}^{1,1,1,2}/144$
		1/1	0/0	1	1	$A_iB_jC_j$ / $A_iB_jC_j$	12	$\alpha C_{10/10}^{1,1,1,1}$	$12\alpha C_{10/10}^{1,1,1,1}/144$
		1/1	0/1	1	1	$A_iB_jC_j$ / $A_iB_jC_i$, $A_iB_jC_j$ / $A_iB_jC_k$	12, 24	$\alpha C_{10/11}^{1,1,1,1}$	$4\alpha C_{10/11}^{1,1,1,1}/144$
		1/1	1/0	1	1	$A_iB_jC_i$ / $A_iB_jC_j$, $A_iB_jC_k$ / $A_iB_jC_j$	12, 24	$\alpha C_{11/10}^{1,1,1,1}$	$4\alpha C_{11/10}^{1,1,1,1}/144$
		1/1	1/1	1	1	$A_iB_jC_i$ / $A_iB_jC_i$, $A_iB_jC_k$ / $A_iB_jC_k$	12, 24	$\alpha C_{11/11}^{1,1,1,1}$	$4\alpha C_{11/11}^{1,1,1,1}/144$

				1	2	$A_iB_jC_i / A_iB_jC_k, A_iB_jC_k / A_iB_jC_i$ $A_iB_jC_k / A_iB_jC_l$	24, 24, 24	$\alpha C_{11/11}^{1,1,1,2}$	$2\alpha C_{11/11}^{1,1,1,2}/144$
1	0	0/1	0/0	1	1	$A_iB_iC_i / A_iB_jC_j$	12	$\alpha C_{00/10}^{1,0,1,1}$	$12\alpha C_{00/10}^{1,0,1,1}/144$
		0/1	0/1	1	1	$A_iB_iC_i / A_iB_jC_i$	12	$\alpha C_{00/11}^{1,0,1,1}$	$12\alpha C_{00/11}^{1,0,1,1}/144$
				1	2	$A_iB_iC_i / A_iB_jC_k$	24	$\alpha C_{00/11}^{1,0,1,2}$	$6\alpha C_{00/11}^{1,0,1,2}/144$
		0/1	1/0	1	1	$A_iB_iC_j / A_iB_jC_j$	12	$\alpha C_{01/10}^{1,0,1,1}$	$12\alpha C_{01/10}^{1,0,1,1}/144$
				1	2	$A_iB_iC_k / A_iB_jC_j$	24	$\alpha C_{01/10}^{1,0,1,2}$	$6\alpha C_{01/10}^{1,0,1,2}/144$
		0/1	1/1	1	1	$A_iB_iC_j / A_iB_jC_i$	12	$\alpha C_{01/11}^{1,0,1,1}$	$12\alpha C_{01/11}^{1,0,1,1}/144$
				1	2	$A_iB_iC_j / A_iB_jC_k$	24	$\alpha C_{01/11}^{1,0,1,2}$	$6\alpha C_{01/11}^{1,0,1,2}/144$
				1	3	$A_iB_iC_k / A_iB_jC_i$	24	$\alpha C_{01/11}^{1,0,1,3}$	$6\alpha C_{01/11}^{1,0,1,3}/144$
				1	4	$A_iB_iC_k / A_iB_jC_k$	24	$\alpha C_{01/11}^{1,0,1,4}$	$6\alpha C_{01/11}^{1,0,1,4}/144$
				1	5	$A_iB_iC_k / A_iB_jC_l$	24	$\alpha C_{01/11}^{1,0,1,5}$	$6\alpha C_{01/11}^{1,0,1,5}/144$
		1/0	0/0	1	1	$A_iB_jC_j / A_iB_iC_i$	12	$\alpha C_{10/00}^{1,0,1,1}$	$12\alpha C_{10/00}^{1,0,1,1}/144$
		1/0	0/1	1	1	$A_iB_jC_j / A_iB_iC_j$	12	$\alpha C_{10/01}^{1,0,1,1}$	$12\alpha C_{10/01}^{1,0,1,1}/144$

				1	2	$A_iB_jC_j / A_iB_iC_k$	24	$\alpha C_{10/01}^{1,0,1,2}$	$6\alpha C_{10/01}^{1,0,1,2}/144$
		1/0	1/0	1	1	$A_iB_jC_i / A_iB_iC_i$	12	$\alpha C_{11/00}^{1,0,1,1}$	$12\alpha C_{11/00}^{1,0,1,1}/144$
				1	2	$A_iB_jC_k / A_iB_iC_i$	24	$\alpha C_{11/00}^{1,0,1,2}$	$6\alpha C_{11/00}^{1,0,1,2}/144$
		1/0	1/1	1	1	$A_iB_jC_i / A_iB_iC_j$	12	$\alpha C_{11/01}^{1,0,1,1}$	$12\alpha C_{11/01}^{1,0,1,1}/144$
				1	2	$A_iB_jC_i / A_iB_iC_k$	24	$\alpha C_{11/01}^{1,0,1,2}$	$6\alpha C_{11/01}^{1,0,1,2}/144$
				1	3	$A_iB_jC_k / A_iB_iC_j$	24	$\alpha C_{11/01}^{1,0,1,3}$	$6\alpha C_{11/01}^{1,0,1,3}/144$
				1	4	$A_iB_jC_k / A_iB_iC_k$	24	$\alpha C_{11/01}^{1,0,1,4}$	$6\alpha C_{11/01}^{1,0,1,4}/144$
				1	5	$A_iB_jC_l / A_iB_iC_k$	24	$\alpha C_{11/01}^{1,0,1,5}$	$6\alpha C_{11/01}^{1,0,1,5}/144$
		1/1	0/0	2	1	$A_iB_jC_j / A_iB_kC_k$	24	$\alpha C_{10/10}^{1,0,2,1}$	$6\alpha C_{10/10}^{1,0,2,1}/144$
		1/1	0/1	2	1	$A_iB_jC_j / A_iB_kC_j$	24	$\alpha C_{10/11}^{1,0,2,1}$	$6\alpha C_{10/11}^{1,0,2,1}/144$
				2	2	$A_iB_jC_j / A_iB_kC_i, A_iB_jC_j / A_iB_kC_l$	24, 24	$\alpha C_{10/11}^{1,0,2,2}$	$3\alpha C_{10/11}^{1,0,2,2}/144$
		1/1	1/0	2	1	$A_iB_jC_k / A_iB_kC_k$	24	$\alpha C_{11/10}^{1,0,2,1}$	$6\alpha C_{11/10}^{1,0,2,1}/144$
				2	2	$A_iB_jC_i / A_iB_kC_k, A_iB_jC_l / A_iB_kC_k$	24, 24	$\alpha C_{11/10}^{1,0,2,2}$	$3\alpha C_{11/10}^{1,0,2,1}/144$
		1/1	1/1	2	1	$A_iB_jC_k / A_iB_kC_j$	24	$\alpha C_{11/11}^{1,0,2,1}$	$6\alpha C_{11/11}^{1,0,2,1}/144$

				2	2	$A_iB_jC_k / A_iB_kC_i, A_iB_jC_k / A_iB_kC_l$	24, 24	$\alpha C_{11/11}^{1,0,2,2}$	$3\alpha C_{11/11}^{1,0,2,2}/144$
				2	3	$A_iB_jC_i / A_iB_kC_j, A_iB_jC_l / A_iB_kC_j$	24, 24	$\alpha C_{11/11}^{1,0,2,3}$	$3\alpha C_{11/11}^{1,0,2,3}/144$
				2	4	$A_iB_jC_i / A_iB_kC_i, A_iB_jC_l / A_iB_kC_l$	24, 24	$\alpha C_{11/11}^{1,0,2,4}$	$3\alpha C_{11/11}^{1,0,2,4}/144$
				2	5	$A_iB_jC_i / A_iB_kC_l, A_iB_jC_l / A_iB_kC_i$	24, 24	$\alpha C_{11/11}^{1,0,2,5}$	$3\alpha C_{11/11}^{1,0,2,5}/144$
0	1	0/1	0/0	1	1	$A_iB_iC_i / A_jB_iC_i$	12	$(1-\alpha)C_{00/10}^{0,1,1,1}$	$12(1-\alpha)C_{00/10}^{0,1,1,1}/144$
		0/1	0/1	1	1	$A_iB_iC_i / A_jB_iC_j, A_iB_iC_i / A_jB_iC_k$	12, 24	$(1-\alpha)C_{00/11}^{0,1,1,1}$	$4(1-\alpha)C_{00/11}^{0,1,1,1}/144$
		0/1	1/0	1	1	$A_iB_iC_j / A_jB_iC_i, A_iB_iC_k / A_jB_iC_i$	12, 24	$(1-\alpha)C_{01/10}^{0,1,1,1}$	$4(1-\alpha)C_{01/10}^{0,1,1,1}/144$
		0/1	1/1	1	1	$A_iB_iC_j / A_jB_iC_j, A_iB_iC_k / A_jB_iC_k$	12, 24	$(1-\alpha)C_{01/11}^{0,1,1,1}$	$4(1-\alpha)C_{01/11}^{0,1,1,1}/144$
				1	2	$A_iB_iC_j / A_jB_iC_k, A_iB_iC_k / A_jB_iC_j$ $A_iB_iC_k / A_jB_iC_l$	24, 24, 24	$(1-\alpha)C_{01/11}^{0,1,1,2}$	$2(1-\alpha)C_{01/11}^{0,1,1,2}/144$
		1/0	0/0	1	1	$A_iB_jC_j / A_jB_jC_j$	12	$(1-\alpha)C_{10/00}^{0,1,1,1}$	$12(1-\alpha)C_{10/00}^{0,1,1,1}/144$
		1/0	0/1	1	1	$A_iB_jC_j / A_jB_jC_i, A_iB_jC_j / A_jB_jC_k$	12, 24	$(1-\alpha)C_{10/01}^{0,1,1,1}$	$4(1-\alpha)C_{10/01}^{0,1,1,1}/144$
		1/0	1/0	1	1	$A_iB_jC_i / A_jB_jC_j, A_iB_jC_k / A_jB_jC_j$	12, 24	$(1-\alpha)C_{11/00}^{0,1,1,1}$	$4(1-\alpha)C_{11/00}^{0,1,1,1}/144$
		1/0	1/1	1	1	$A_iB_jC_i / A_jB_jC_i, A_iB_jC_k / A_jB_jC_k$	12, 24	$(1-\alpha)C_{11/01}^{0,1,1,1}$	$4(1-\alpha)C_{11/01}^{0,1,1,1}/144$

				1	2	$A_iB_jC_i / A_jB_jC_k$, $A_iB_jC_k / A_jB_jC_i$ $A_iB_jC_k / A_jB_jC_l$	24, 24, 24	$(1-\alpha)C_{11/01}^{0,1,1,2}$	$2(1-\alpha)C_{11/01}^{0,1,1,2}/144$
		1/1	0/0	4	1	$A_iB_kC_k / A_jB_kC_k$	12	$(1-\alpha)C_{10/10}^{0,1,4,1}$	$12(1-\alpha)C_{10/10}^{0,1,4,1}/144$
		1/1	0/1	4	1	$A_iB_kC_k / A_jB_kC_i$, $A_iB_kC_k / A_jB_kC_j$ $A_iB_kC_k / A_jB_kC_l$	24, 24, 24	$(1-\alpha)C_{10/11}^{0,1,4,1}$	$2(1-\alpha)C_{10/11}^{0,1,4,1}/144$
		1/1	0/1	4	1	$A_iB_kC_i / A_jB_kC_k$, $A_iB_kC_j / A_jB_kC_k$ $A_iB_kC_l / A_jB_kC_k$	24, 24, 24	$(1-\alpha)C_{11/10}^{0,1,4,1}$	$2(1-\alpha)C_{11/10}^{0,1,4,1}/144$
		1/1	1/1	4	1	$A_iB_kC_i / A_jB_kC_i$, $A_iB_kC_j / A_jB_kC_j$ $A_iB_kC_l / A_jB_kC_l$	12, 12, 12	$(1-\alpha)C_{11/11}^{0,1,4,1}$	$4(1-\alpha)C_{11/11}^{0,1,4,1}/144$
				4	2	$A_iB_kC_i / A_jB_kC_l$, $A_iB_kC_l / A_jB_kC_i$ $A_iB_kC_j / A_jB_kC_l$, $A_iB_kC_l / A_jB_kC_j$ $A_iB_kC_i / A_jB_kC_j$, $A_iB_kC_j / A_jB_kC_i$	12, 12, 12 12, 12, 12	$(1-\alpha)C_{11/11}^{0,1,4,2}$	$2(1-\alpha)C_{11/11}^{0,1,4,2}/144$
0	0	0/0	0/0	1	1	$A_iB_iC_i / A_jB_jC_j$	6	$(1-\alpha)C_{00/00}^{0,0,1,1}$	$24(1-\alpha)C_{00/00}^{0,0,1,1}/144$
		0/0	0/1	1	1	$A_iB_iC_i / A_jB_jC_i$	12	$(1-\alpha)C_{00/01}^{0,0,1,1}$	$12(1-\alpha)C_{00/01}^{0,0,1,1}/144$
				1	2	$A_iB_iC_i / A_jB_jC_k$	24	$(1-\alpha)C_{00/01}^{0,0,1,2}$	$6(1-\alpha)C_{00/01}^{0,0,1,2}/144$
		0/0	1/0	1	1	$A_iB_iC_j / A_jB_jC_j$	12	$(1-\alpha)C_{01/00}^{0,0,1,1}$	$12(1-\alpha)C_{01/00}^{0,0,1,1}/144$

				1	2	$A_iB_iC_k / A_jB_jC_j$	24	$(1-\alpha)C_{01/00}^{0,0,1,2}$	$6(1-\alpha)C_{01/00}^{0,0,1,2}/144$
		0/0	1/1	1	1	$A_iB_iC_j / A_jB_jC_i$	6	$(1-\alpha)C_{01/01}^{0,0,1,1}$	$24(1-\alpha)C_{01/01}^{0,0,1,1}/144$
				1	2	$A_iB_iC_j / A_jB_jC_k$	12	$(1-\alpha)C_{01/01}^{0,0,1,2}$	$12(1-\alpha)C_{01/01}^{0,0,1,2}/144$
				1	3	$A_iB_iC_k / A_jB_jC_i$	12	$(1-\alpha)C_{01/01}^{0,0,1,3}$	$12(1-\alpha)C_{01/01}^{0,0,1,3}/144$
				1	4	$A_iB_iC_k / A_jB_jC_k$	12	$(1-\alpha)C_{01/01}^{0,0,1,4}$	$12(1-\alpha)C_{01/01}^{0,0,1,4}/144$
				1	5	$A_iB_iC_k / A_jB_jC_l$	12	$(1-\alpha)C_{01/01}^{0,0,1,5}$	$12(1-\alpha)C_{01/01}^{0,0,1,5}/144$
		0/1	0/0	2	1	$A_iB_iC_i / A_jB_kC_k$	24	$(1-\alpha)C_{00/10}^{0,0,2,1}$	$6(1-\alpha)C_{00/10}^{0,0,2,1}/144$
		0/1	0/1	2	1	$A_iB_iC_i / A_jB_kC_i$	24	$(1-\alpha)C_{00/11}^{0,0,2,1}$	$6(1-\alpha)C_{00/11}^{0,0,2,1}/144$
				2	1	$A_iB_iC_i / A_jB_kC_j, A_iB_iC_i / A_jB_kC_l$	24, 24	$(1-\alpha)C_{00/11}^{0,0,2,2}$	$3(1-\alpha)C_{00/11}^{0,0,2,2}/144$
		0/1	1/0	2	1	$A_iB_iC_k / A_jB_kC_k$	24	$(1-\alpha)C_{01/10}^{0,0,2,1}$	$6(1-\alpha)C_{01/10}^{0,0,2,1}/144$
				2	2	$A_iB_iC_j / A_jB_kC_k, A_iB_iC_l / A_jB_kC_k$	24, 24	$(1-\alpha)C_{01/10}^{0,0,2,2}$	$3(1-\alpha)C_{01/10}^{0,0,2,2}/144$
		0/1	1/1	2	1	$A_iB_iC_k / A_jB_kC_i$	24	$(1-\alpha)C_{01/11}^{0,0,2,1}$	$6(1-\alpha)C_{01/11}^{0,0,2,1}/144$
				2	2	$A_iB_iC_k / A_jB_kC_j, A_iB_iC_k / A_jB_kC_l$	24, 24	$(1-\alpha)C_{01/11}^{0,0,2,2}$	$3(1-\alpha)C_{01/11}^{0,0,2,2}/144$

				2	3	$A_iB_iC_j / A_jB_kC_i, A_iB_iC_l / A_jB_kC_i$	24, 24	$(1-\alpha)C_{01/11}^{0,0,2,3}$	$3(1-\alpha)C_{01/11}^{0,0,2,3}/144$
				2	4	$A_iB_iC_j / A_jB_kC_j, A_iB_iC_l / A_jB_kC_l$	24, 24	$(1-\alpha)C_{01/11}^{0,0,2,4}$	$3(1-\alpha)C_{01/11}^{0,0,2,4}/144$
				2	5	$A_iB_iC_j / A_jB_kC_l, A_iB_iC_l / A_jB_kC_j$	24, 24	$(1-\alpha)C_{01/11}^{0,0,2,5}$	$3(1-\alpha)C_{01/11}^{0,0,2,5}/144$
		1/0	0/0	2	1	$A_iB_kC_k / A_jB_jC_j$	24	$(1-\alpha)C_{10/00}^{0,0,2,1}$	$6(1-\alpha)C_{10/00}^{0,0,2,1}/144$
		1/0	0/1	2	1	$A_iB_kC_k / A_jB_jC_k$	24	$(1-\alpha)C_{10/01}^{0,0,2,1}$	$6(1-\alpha)C_{10/01}^{0,0,2,1}/144$
				2	2	$A_iB_kC_k / A_jB_jC_i, A_iB_kC_k / A_jB_jC_l$	24, 24	$(1-\alpha)C_{10/01}^{0,0,2,2}$	$3(1-\alpha)C_{10/01}^{0,0,2,2}/144$
		1/0	1/0	2	1	$A_iB_kC_j / A_jB_jC_j$	24	$(1-\alpha)C_{11/00}^{0,0,2,1}$	$6(1-\alpha)C_{11/00}^{0,0,2,1}/144$
				2	2	$A_iB_kC_i / A_jB_jC_j, A_iB_kC_l / A_jB_jC_j$	24, 24	$(1-\alpha)C_{11/00}^{0,0,2,2}$	$3(1-\alpha)C_{11/00}^{0,0,2,2}/144$
		1/0	1/1	2	1	$A_iB_kC_j / A_jB_jC_k$	24	$(1-\alpha)C_{11/01}^{0,0,2,1}$	$6(1-\alpha)C_{11/01}^{0,0,2,1}/144$
				2	2	$A_iB_kC_j / A_jB_jC_i, A_iB_kC_j / A_jB_jC_l$	24, 24	$(1-\alpha)C_{11/01}^{0,0,2,2}$	$3(1-\alpha)C_{11/01}^{0,0,2,2}/144$
				2	3	$A_iB_kC_i / A_jB_jC_k, A_iB_kC_l / A_jB_jC_k$	24, 24	$(1-\alpha)C_{11/01}^{0,0,2,3}$	$3(1-\alpha)C_{11/01}^{0,0,2,3}/144$
				2	4	$A_iB_kC_i / A_jB_jC_i, A_iB_kC_l / A_jB_jC_l$	24, 24	$(1-\alpha)C_{11/01}^{0,0,2,4}$	$3(1-\alpha)C_{11/01}^{0,0,2,4}/144$
				2	5	$A_iB_kC_i / A_jB_jC_l, A_iB_kC_l / A_jB_jC_i$	24, 24	$(1-\alpha)C_{11/01}^{0,0,2,5}$	$3(1-\alpha)C_{11/01}^{0,0,2,5}/144$

		1/1	0/0	1	1	$A_iB_jC_j / A_jB_iC_i$	6	$(1-\alpha)C_{10/10}^{0,0,1,1}$	$24(1-\alpha)C_{10/10}^{0,0,1,1}/144$
				2	1	$A_iB_jC_j / A_jB_kC_k$	12	$(1-\alpha)C_{10/10}^{0,0,2,1}$	$12(1-\alpha)C_{10/10}^{0,0,2,1}/144$
				3	1	$A_iB_kC_k / A_jB_iC_i$	12	$(1-\alpha)C_{10/10}^{0,0,3,1}$	$12(1-\alpha)C_{10/10}^{0,0,3,1}/144$
				5	1	$A_iB_kC_k / A_jB_lC_l$	12	$(1-\alpha)C_{10/10}^{0,0,5,1}$	$12(1-\alpha)C_{10/10}^{0,0,5,1}/144$
		1/1	0/1	1	1	$A_iB_jC_j / A_jB_iC_j$	12	$(1-\alpha)C_{10/11}^{0,0,1,1}$	$12(1-\alpha)C_{10/11}^{0,0,1,1}/144$
				1	2	$A_iB_jC_j / A_jB_iC_k$	24	$(1-\alpha)C_{10/11}^{0,0,1,2}$	$6(1-\alpha)C_{10/11}^{0,0,1,2}/144$
				2	1	$A_iB_jC_j / A_jB_kC_j$	24	$(1-\alpha)C_{10/11}^{0,0,2,1}$	$6(1-\alpha)C_{10/11}^{0,0,2,1}/144$
				2	2	$A_iB_jC_j / A_jB_kC_i, A_iB_jC_j / A_jB_kC_l$	24, 24	$(1-\alpha)C_{10/11}^{0,0,2,2}$	$3(1-\alpha)C_{10/11}^{0,0,2,2}/144$
				3	1	$A_iB_kC_k / A_jB_iC_k$	24	$(1-\alpha)C_{10/11}^{0,0,3,1}$	$6(1-\alpha)C_{10/11}^{0,0,3,1}/144$
				3	2	$A_iB_kC_k / A_jB_iC_j, A_iB_kC_k / A_jB_iC_l$	24, 24	$(1-\alpha)C_{10/11}^{0,0,3,2}$	$3(1-\alpha)C_{10/11}^{0,0,3,2}/144$
				5	1	$A_iB_kC_k / A_jB_lC_k$	24	$(1-\alpha)C_{10/11}^{0,0,5,1}$	$6(1-\alpha)C_{10/11}^{0,0,5,1}/144$
				5	2	$A_iB_kC_k / A_jB_lC_i, A_iB_kC_k / A_jB_lC_j$	24, 24	$(1-\alpha)C_{10/11}^{0,0,5,2}$	$3(1-\alpha)C_{10/11}^{0,0,5,2}/144$
		1/1	1/0	1	1	$A_iB_jC_i / A_jB_iC_i$	12	$(1-\alpha)C_{11/10}^{0,0,1,1}$	$12(1-\alpha)C_{11/10}^{0,0,1,1}/144$

				1	2	$A_iB_jC_k / A_jB_iC_i$	24	$(1-\alpha)C_{11/10}^{0,0,1,2}$	$6(1-\alpha)C_{11/10}^{0,0,1,2}/144$
				2	1	$A_iB_jC_k / A_jB_kC_k$	24	$(1-\alpha)C_{11/10}^{0,0,2,1}$	$6(1-\alpha)C_{11/10}^{0,0,2,1}/144$
				2	2	$A_iB_jC_i / A_jB_kC_k, A_iB_jC_l / A_jB_kC_k$	24, 24	$(1-\alpha)C_{11/10}^{0,0,2,2}$	$3(1-\alpha)C_{11/10}^{0,0,2,2}/144$
				3	1	$A_iB_kC_i / A_jB_iC_i$	24	$(1-\alpha)C_{11/10}^{0,0,3,1}$	$6(1-\alpha)C_{11/10}^{0,0,3,1}/144$
				3	2	$A_iB_kC_j / A_jB_iC_i, A_iB_kC_l / A_jB_iC_i$	24, 24	$(1-\alpha)C_{11/10}^{0,0,3,2}$	$3(1-\alpha)C_{11/10}^{0,0,3,2}/144$
				5	1	$A_iB_kC_l / A_jB_lC_l$	24	$(1-\alpha)C_{11/10}^{0,0,5,1}$	$6(1-\alpha)C_{11/10}^{0,0,5,1}/144$
				5	2	$A_iB_kC_i / A_jB_lC_l, A_iB_kC_j / A_jB_lC_l$	24, 24	$(1-\alpha)C_{11/10}^{0,0,5,2}$	$3(1-\alpha)C_{11/10}^{0,0,5,2}/144$
		1/1	1/1	1	1	$A_iB_jC_i / A_jB_iC_j$	6	$(1-\alpha)C_{11/11}^{0,0,1,1}$	$24(1-\alpha)C_{11/11}^{0,0,1,1}/144$
				1	2	$A_iB_jC_i / A_jB_iC_k$	12	$(1-\alpha)C_{11/11}^{0,0,1,2}$	$12(1-\alpha)C_{11/11}^{0,0,1,2}/144$
				1	3	$A_iB_jC_k / A_jB_iC_j$	12	$(1-\alpha)C_{11/11}^{0,0,1,3}$	$12(1-\alpha)C_{11/11}^{0,0,1,3}/144$
				1	4	$A_iB_jC_k / A_jB_iC_k$	12	$(1-\alpha)C_{11/11}^{0,0,1,4}$	$12(1-\alpha)C_{11/11}^{0,0,1,4}/144$
				1	5	$A_iB_jC_k / A_jB_iC_l$	12	$(1-\alpha)C_{11/11}^{0,0,1,5}$	$12(1-\alpha)C_{11/11}^{0,0,1,5}/144$
				2	1	$A_iB_jC_k / A_jB_kC_j$	12	$(1-\alpha)C_{11/11}^{0,0,2,1}$	$12(1-\alpha)C_{11/11}^{0,0,2,1}/144$

				2	2	$A_iB_jC_k / A_jB_kC_i, A_iB_jC_k / A_jB_kC_l$	12, 12	$(1-\alpha)C_{11/11}^{0,0,2,2}$	$6(1-\alpha)C_{11/11}^{0,0,2,2}/144$
				2	3	$A_iB_jC_i / A_jB_kC_j, A_iB_jC_l / A_jB_kC_j$	12, 12	$(1-\alpha)C_{11/11}^{0,0,2,3}$	$6(1-\alpha)C_{11/11}^{0,0,2,3}/144$
				2	4	$A_iB_jC_i / A_jB_kC_i, A_iB_jC_l / A_jB_kC_l$	12, 12	$(1-\alpha)C_{11/11}^{0,0,2,4}$	$6(1-\alpha)C_{11/11}^{0,0,2,4}/144$
				2	5	$A_iB_jC_i / A_jB_kC_l, A_iB_jC_l / A_jB_kC_i$	12, 12	$(1-\alpha)C_{11/11}^{0,0,2,5}$	$6(1-\alpha)C_{11/11}^{0,0,2,5}/144$
				3	1	$A_iB_kC_i / A_jB_iC_k$	12	$(1-\alpha)C_{11/11}^{0,0,3,1}$	$12(1-\alpha)C_{11/11}^{0,0,3,1}/144$
				3	2	$A_iB_kC_i / A_jB_iC_j, A_iB_kC_i / A_jB_iC_l$	12, 12	$(1-\alpha)C_{11/11}^{0,0,3,2}$	$6(1-\alpha)C_{11/11}^{0,0,3,2}/144$
				3	3	$A_iB_kC_j / A_jB_iC_k, A_iB_kC_l / A_jB_iC_k$	12, 12	$(1-\alpha)C_{11/11}^{0,0,3,3}$	$6(1-\alpha)C_{11/11}^{0,0,3,3}/144$
				3	4	$A_iB_kC_j / A_jB_iC_j, A_iB_kC_l / A_jB_iC_l$	12, 12	$(1-\alpha)C_{11/11}^{0,0,3,4}$	$6(1-\alpha)C_{11/11}^{0,0,3,4}/144$
				3	5	$A_iB_kC_j / A_jB_iC_l, A_iB_kC_l / A_jB_iC_j$	12, 12	$(1-\alpha)C_{11/11}^{0,0,3,5}$	$6(1-\alpha)C_{11/11}^{0,0,3,5}/144$
				5	1	$A_iB_kC_l / A_jB_lC_k$	12	$(1-\alpha)C_{11/11}^{0,0,5,1}$	$12(1-\alpha)C_{11/11}^{0,0,5,1}/144$
				5	2	$A_iB_kC_l / A_jB_lC_i, A_iB_kC_l / A_jB_lC_j$	12, 12	$(1-\alpha)C_{11/11}^{0,0,5,2}$	$6(1-\alpha)C_{11/11}^{0,0,5,2}/144$
				5	3	$A_iB_kC_i / A_jB_lC_k, A_iB_kC_j / A_jB_lC_k$	12, 12	$(1-\alpha)C_{11/11}^{0,0,5,3}$	$6(1-\alpha)C_{11/11}^{0,0,5,3}/144$
				5	4	$A_iB_kC_i / A_jB_lC_i, A_iB_kC_j / A_jB_lC_j$	12, 12	$(1-\alpha)C_{11/11}^{0,0,5,4}$	$6(1-\alpha)C_{11/11}^{0,0,5,4}/144$

				5	5	$A_iB_kC_i / A_jB_lC_j, A_iB_kC_j / A_jB_lC_i$	12, 12	$(1-\alpha)C_{11/11}^{0,0,5,5}$	$6(1-\alpha)C_{11/11}^{0,0,5,5}/144$

Under the assumption of random union of gametes from two parents, a general form for the frequency of zygote *j*, which is composed of gametes *k* and *l* from the two parental genotypes, may be expressed as:

$$h_j = g_k g_l = \sigma_k \sigma_l \alpha^{I_{\alpha,k}+I_{\alpha,l}} (1-\alpha)^{2-I_{\alpha,k}-I_{\alpha,l}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} C^{I_{\alpha,l},I_{\beta,l},e_{1,l},e_{2,l}}_{s_l t_l / s'_l t'_l} \quad \text{(II-1.36)}$$

By sorting the zygotes according to their genotypes, a general formula for the frequency of zygotic genotype *i* may be written as:

$$\begin{aligned} H_i &= \sum_{k\&l\in i} g_k g_l \\ &= \sum_{k\&l\in i} \sigma_k \sigma_l \alpha^{I_{\alpha,k}+I_{\alpha,l}} (1-\alpha)^{2-I_{\alpha,k}-I_{\alpha,l}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} C^{I_{\alpha,l},I_{\beta,l},e_{1,l},e_{2,l}}_{s_l t_l / s'_l t'_l} \end{aligned} \quad \text{(II-1.37)}$$

where $k\&l \in i$ indicates the sum of the frequencies of all those zygotes made up of gamete *k* and gamete *l*, which correspond to the same zygotic genotype *i*.

The phenotypic distribution of offspring from two parental autotetraploids can be derived by summing up the probabilities of those genotypes that result in the same phenotypes. A general formula for the probability of zygote phenotype *i* takes the form of:

$$\begin{aligned} f_i &= \sum_{g\in i} H_g = \sum_{g\in i} \sum_{k\&l\in g} \sigma_k \sigma_l \alpha^{I_{\alpha,k}+I_{\alpha,l}} (1-\alpha)^{2-I_{\alpha,k}-I_{\alpha,l}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} C^{I_{\alpha,l},I_{\beta,l},e_{1,l},e_{2,l}}_{s_l t_l / s'_l t'_l} \\ &= \sum_{g\in i} \sum_{k\&l\in g} \theta_{g_{kl}} \alpha^{W_{g_{kl}}} (1-\alpha)^{2-W_{g_{kl}}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} C^{I_{\alpha,l},I_{\beta,l},e_{1,l},e_{2,l}}_{s_l t_l / s'_l t'_l} \end{aligned} \quad \text{(II-1.38)}$$

where $k\&l \in g$ indicates sum over the frequencies of all zygote combined with gametes *k* and *l* that corresponding to the same zygotic genotype *g* and $\sum_{g\in i} H_g$ indicates the sum over the

frequencies of all those zygotic genotypes *g* that correspond to the same phenotype *i*, $\theta_{g_{kl}} = \sigma_k \cdot \sigma_l$, $W_{g_{kl}} = I_{\alpha,k} + I_{\alpha,l}$.

1.3.4. Maximum likelihood estimates of the model parameters

The above statistical method predicts the unknown parameters *m*, α, d_1 and d_2 in the model with the information of the given parental phenotypes at the three marker loci, P_1 and P_2, and the marker phenotypes of a full-sib family of *n* segregating offspring individuals from a cross between the parental lines, $O = (o_1, o_2, ..., o_n)$. It has been demonstrated previously (Luo et al 2000) that genotypes of the two parental lines, G_1 and G_2, at each of the three loci, can be accurately estimated from P_1, P_2 and $O = (o_1, o_2, ..., o_n)$. The linkage phase of the parental genotypes can be predicted through searching all possible allelic combinations at the marker loci under a tetrasomic linkage analysis model as shown in (Luo et al 2004, 2006). To simplify formulation of the RI analysis, we focus here on G_1^* and G_2^*, the most likely estimated parental genotypes. When the offspring can be classified into *M* phenotype categories, each with n_i individuals, the likelihood function of the parameters $\Omega = (m, \alpha, d_1, d_2)$ can be written as follows:

$$L(\Omega \mid P_1, P_2, O) = \sum_{G_1 \subset P_1, G_2 \subset P_2} \Pr(G_1, G_2 \mid O, P_1, P_2) \Pr(O \mid G_1, G_2, \Omega)$$

$$\propto \Pr(O \mid G_1^*, G_2^*, \Omega) = \binom{n}{n_1 n_2 \cdots n_M} f_1^{n_1} f_2^{n_2} \cdots f_M^{n_M} \quad \text{(II-1.39)}$$

The logarithm of the likelihood is thus

$$\begin{aligned}\ln\left(L\left(P_1,P_2,\Omega\,|\,O\right)\right) &\propto \sum_{i=1}^{M} n_i \ln\left(f_i\right) \\ &\propto \sum_{i=1}^{M} n_i \ln\left(\sum_{g\in i}\sum_{k\&l\in g} \theta_{g_{kl}} \alpha^{W_{g_{kl}}} \left(1-\alpha\right)^{2-W_{g_{kl}}} C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} C^{I_{\alpha,l},I_{\beta,l},e_{1,l},e_{2,l}}_{s_l t_l / s'_l t'_l}\right)\end{aligned} \tag{II-1.40}$$

In which

$$C^{I_{\alpha,k},I_{\beta,k},e_{1,k},e_{2,k}}_{s_k t_k / s'_k t'_k} = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} o^{I_{\alpha,k},k_1}_{s_k / s'_k,e_{1,k}} o^{I_{\beta,k},k_2}_{t_k / t'_k,e_{2,k}} \Pr\{k_1,k_2\,|\,d_1,d_2\}$$

$$C^{I_{\alpha,l},I_{\beta,l},e_{1,l},e_{2,l}}_{s_l t_l / s'_l t'_l} = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} o^{I_{\alpha,l},k_1}_{s_l / s'_l,e_{1,l}} o^{I_{\beta,l},k_2}_{t_l / t'_l,e_{2,l}} \Pr\{k_1,k_2\,|\,d_1,d_2\}$$

To simplify the formulation below, we write

$$A^{I_\alpha,I_\beta,k}_{i_1 i_2 / j_1 j_2,e_1,e_2} = \sum_{k_2=0}^{\infty} o^{I_\alpha,k}_{i_1 / j_1,e_1} o^{I_\beta,k_2}_{i_2 / j_2,e_2} \Pr\{k,k_2\,|\,d_1,d_2\} \tag{II-1.41}$$

which is the frequency of a three locus gamete mode, $G^{I_\alpha,I_\beta}_{i_1 i_2 / j_1 j_2,e_1,e_2}$, on three loci given *k* crossovers occurring in the marker interval AB,

and

$$B^{I_\alpha,I_\beta,k}_{i_1 i_2 / j_1 j_2,e_1,e_2} = \sum_{k_1=0}^{\infty} o^{I_\alpha,k_1}_{i_1 / j_1,e_1} o^{I_\beta,k}_{i_2 / j_2,e_2} \Pr\{k_1,k\,|\,d_1,d_2\} \tag{II-1.42}$$

which represents the frequency of a three locus gamete mode, $G^{I_\alpha,I_\beta}_{i_1\,i_2/\,j_1\,j_2,e_1,e_2}$, given *k* crossovers occurring in the marker interval BC.

Since the interference coefficient *m* takes non-negative integer values, the maximum likelihood estimates of the other model parameters will be calculated based on a prior given *m*. We repeat the estimation procedure at different *m* values and determine the global MLEs of all parameters by comparing the likelihood function (Equation (II-1.39)) at these different parameter values. For a given value of *m*, we propose an EM algorithm to calculate the other model parameters. In particular, the expectation (E) step of the EM algorithm calculates the conditional probability that individuals of the i^{th} phenotype carry a total of *k* crossovers within the first marker interval, which are generated during meiosis of their two parents, γ_{ik}.

$$\gamma_{ik}=\sum_{k'=0}^{k}\left\{\sum_{g\in i}\sum_{u\&v\in g}\theta_{g_{uv}}\alpha^{W_{g_{uv}}}(1-\alpha)^{2-W_{g_{uv}}}A^{I_{\alpha,u},I_{\beta,u},k'}_{s_ut_u/s'_ut'_u,e_{1,u},e_{2,u}}A^{I_{\alpha,v},I_{\beta,v},k-k'}_{s_vt_v/s'_vt'_v,e_{1,v},e_{2,v}}\Big/f_i\right\} \tag{II-1.43}$$

Here *k'* indicates the number of crossover events occurring during the meiosis of one parent and *k*-*k'* indicates the number of crossover events occurring during meiosis of on the other parent.

The conditional probability of individuals of the i^{th} phenotype with a total of *k* crossovers from both parents within the second marker interval, ω_{ik}, is:

$$\omega_{ik}=\sum_{k'=0}^{k}\left\{\sum_{g\in i}\sum_{u\&v\in g}\theta_{g_{uv}}\alpha^{W_{g_{uv}}}(1-\alpha)^{2-W_{g_{uv}}}B^{I_{\alpha,u},I_{\beta,u},k'}_{s_ut_u/s'_ut'_u,e_{1,u},e_{2,u}}B^{I_{\alpha,v},I_{\beta,v},k-k'}_{s_vt_v/s'_vt'_v,e_{1,v},e_{2,v}}\Big/f_i\right\} \tag{II-1.44}$$

The conditional probability of individuals of the i^{th} phenotype with k double-reduction gametes, ξ_{ik}, is calculated as

$$\xi_{ik} = \sum_{g \in i} \sum_{u\&v \in g, w_{g_{uv}} = k} \theta_{g_{uv}} \alpha^{w_{g_{uv}}} C^{I_{\alpha,u}, I_{\beta,u}, e_{1,u}, e_{2,u}}_{s_u t_u / s'_u t'_u} C^{I_{\alpha,v}, I_{\beta,v}, e_{1,v}, e_{2,v}}_{s_v t_v / s'_v t'_v} \Big/ f_i \qquad \text{(II-1.45)}$$

where $\sum_{u\&v \in g, W_{g_{uv}} = k}$ indicates the sum of the frequencies of zygotes consisting of gametes u and gamete v that also have the same genotype g and the number of double reduction gametes in each zygote, $w_{g_{uv}}$, is equal to k.

Since each chromosomal strand has a chance of ¼ to be involved in each crossover under both the bivalent and quadrivalent pairing models, the M step updates the estimates of the parameters of genetic distances, which are defined as the expected number of crossovers occurring on a single chromatid within that interval, from:

$$d_1^{'} = \frac{1}{4}\left(\sum_{i=1}^{M} n_i \sum_{k=1}^{\infty} k\gamma_{ik} \Big/ 2n \right) \qquad \text{(II-1.46)}$$

$$d_2^{'} = \frac{1}{4}\left(\sum_{i=1}^{M} n_i \sum_{k=1}^{\infty} k\omega_{ik} \Big/ 2n \right) \qquad \text{(II-1.47)}$$

and estimates the coefficient of double reduction from

$$\alpha' = \sum_{i=1}^{M} n_i \sum_{k=0}^{2} k\xi_{ik} \Big/ 2n \qquad \text{(II-1.48)}$$

The likelihood function increases monotonically as the E step and M step repeat and the parameter estimates converge to the MLEs conditional on the given integer parameter, *m*. We can infer the most likely coefficient of crossover interference, *m*, by examining the likelihood.

1.4. Simulation studies

We conduct an intensive computer simulation study to test reliability of the theoretical analyses, the feasibility of implementing the methods developed here for data analysis, and to explore the statistical properties of the methods developed. The simulation model and programs mimic chromosome segregation and recombination during gametogenesis of an autotetraploid species under either bivalent or quadrivalent homologous chromosome pairing (Luo et al 2006).

To demonstrate the theory and method for inference of crossover interference developed here, I simulated a full-sib family of 1000 individuals from crossing two diploid genotypes *ABC/abc* and *ABC/abc*, and two autotetraploid genotypes *ABC/ABC/abc/abc* and *ABC/ABC/abc/abc* with both bivalent pairing and quadrivalent pairing during meiosis. Three sets of parameters were considered for each ploidy level and the means and standard errors (in brackets) of the MLEs based on 30 replicate simulations are shown in Table II-1.7.

To speed up the computation in autotetraploids with quadrivalent pairing, we first estimated the coefficient of double reduction for locus A independently according to the offspring phenotype data for locus A. The phenotypic probability distribution of offspring generated from parents $AAaa \times AAaa$ is $(1+2\alpha)^2/36$ for phenotype (1, 0), $1-(1+2\alpha)^2/18$ for phenotype (1, 1) and

$(1+2\alpha)^2/36$ for phenotype (0, 1). Here 1 or 0 in the first (or second) element of the phenotype vector indicates the presence or absence of allele *A* (or *a*). Then the likelihood function of offspring data can be calculated as

$$L=(n_1+n_3)\ln\frac{1}{36}(1+2\alpha)^2+n_2\ln\left[1-\frac{1}{18}(1+2\alpha)^2\right]+C \quad \text{(II-1.49)}$$

where n_1, n_2 or n_3 denotes the number of individuals in phenotype category (1, 0), (1, 1) or (0, 1) and *C* is a constant.

Setting the derivative of the likelihood with respect to α equal to 0, the likelihood function reaches to maximum at the most likely estimate of α as:

$$\alpha=\left[\sqrt{18n(n_1+n_3)}-n\right]/2n \quad \text{(II-1.50)}$$

Here $n=n_1+n_2+n_3$.

For a given estimated coefficient of double reduction, further estimation of other model parameters can be achieved more rapidly. The MLEs of the coefficient of interference, *m*, were searched from 0 to 5 based on 1000 offspring individuals. It can be seen from Table II-1.7 that all model parameters were predicted accurately and with reasonable precision.

Table II-1.7. Simulated parameters and means and standard errors (in brackets) of their MLEs

	m	α	d_1	d_2	$\hat{m}$	$\hat{\alpha}$	$\hat{d}_1$	$\hat{d}_2$
Diploid	0	-	0.15	0.15	0.1333 (0.0631)	-	0.1490 （0.0023）	0.1487 （0.0195）
	1	-	0.15	0.15	1.2000 (0.1213)	-	0.1520 （0.0019）	0.1497 （0.0021）
	2	-	0.15	0.15	2.2000 （0.2166）	-	0.1520 （0.0016）	0.1507 （0.0018）
Tetraploid[1]	0	-	0.15	0.15	0.2000 （0.0884）	-	0.1511 （0.0030）	0.1487 （0.0038）
	1	-	0.15	0.15	1.1667 （0.2095）	-	0.1511 （0.0027）	0.1534 （0.0029）
	2	-	0.15	0.15	2.1667 （0.2673）	-	0.1526 （0.0033）	0.1475 （0.0023）
Tetraploid[2]	0	0.1	0.15	0.15	0.3333 （0.0875）	0.1007 （0.0044）	0.1487 （0.0008）	0.1508 （0.0011）
	1	0.1	0.15	0.15	1.2333 (0.3096)	0.1007 （0.0044）	0.1510 (0.0021)	0.1528 (0.0013)
	2	0.1	0.15	0.15	2.0667 (0.3553)	0.1007 （0.0044）	0.1502 (0.0007)	0.1522 (0.0018)

m ($\hat{m}$), α ($\hat{\alpha}$), d_1 ($\hat{d}_1$) and d_2 ($\hat{d}_2$) are simulated value (or MLEs) of the coefficient of interference, the coefficient of double reduction and genetic distances for the two consecutive marker intervals. Here Tetraploid[1] represents autotetraploids with bivalent pairing during meiosis and Tetraploid[2] represents autotetraploids with quadrivalent pairing.

1.5. Real data analysis

In this section the model we have developed is implemented to fit three marker locus gamete data generated from both diploid and autotetraploid *Saccharomyces cerevisiae* on three different chromosomes 3, 6 and 8. Two haploid strains of budding yeast, YH1A and YL1C, were used to initiate creation of autotetraploid strains. YH1A is isogenic to the standard strain, s288c, and YL1C is a laboratory strain. The lithium acetate method was implemented to transform the three anti-biotic genes hphMX4 (anti-hygromycin B), natMX4 (anti-nourseothricin), kanMX4 (anti_G418), into a given chromosome at pre-designed locations in the haploid strain YH1A with MATa. The transformation of the anti-biotic genes was repeatedly and respectively done in yeast chromosomes III, VI and VIII. The genetically modified haploid strain was then used to construct diploid and autotetraploid strains. Right after completion of meiosis, tetrads generated from the diploid and autotetraploid strains were dissected by use of a micromanipulator (Singer, MSM300). Single-colony cultures were patched on a YPD plate added with hygromycin B, a YPD plate added with nourseothricin, a YPD plate added with G418 and a standard YPD plate. Genotype of each spore was confirmed from whether it could grow on a plate added the corresponding anti-biotic. This part of experimental work was done by our collaborator from Fudan University in Shanghai, China.

For each chromosome, parental genotypes for the three dominant marker loci are denoted by $Aa\,Bb\,Cc$ and $Aaaa\,Bbbb\,Cccc$ for both diploid and autotetraploid parents, where capital letter alleles are linked on the same chromosome. There are eight different phenotypes of gametes generated by the genotyping data as follows: $o_1 = (1 \quad 1 \quad 1)$, $o_2 = (1 \quad 1 \quad 0)$, $o_3 = (1 \quad 0 \quad 1)$,

$o_4 = (0 \quad 1 \quad 1)$, $o_5 = (1 \quad 0 \quad 0)$, $o_6 = (0 \quad 1 \quad 0)$, $o_7 = (0 \quad 0 \quad 1)$, $o_8 = (0 \quad 0 \quad 0)$. Here the three elements in the vector represent the phenotyping results for the marker loci A, B and C, respectively. A value of 1 (or 0) indicates the presence (or absence) of capital letter alleles in the gametes for the corresponding loci. The observed counts of the different gamete phenotypes for the three chromosomes in both diploids and autotetraploids are listed in Table II-1.8.

In the current study, the total *n* gametes generated by a diploid/autotetraploid parent can be classified into eight phenotype categories, each with n_i individuals, the likelihood function of the parameters $\Omega = (m, \alpha, d_1, d_2)$ can be written as follows:

$$L(\Omega | G, O) \propto \Pr(O | G, \Omega) = \binom{n}{n_1 n_2 \ldots n_8} f_1^{n_1} f_2^{n_2} \ldots f_8^{n_8} \qquad \text{(II-1.51)}$$

Table II-1.8. Observed counts of gametes from *S. cerevisiae* data

Phenotypes of gametes		o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	Total
		111	110	101	011	100	010	001	000	
Chromosome 3	Diploids	346	136	21	63	61	15	137	349	1128
	Autotetraploids	430	176	30	66	66	22	158	476	1424
Chromosome 6	Diploids	529	65	12	173	177	11	73	524	1564
	Autotetraploids	395	67	39	183	188	26	68	426	1392
Chromosome 8	Diploids	382	87	30	147	143	30	87	382	1288
	Autotetraploids	390	120	38	145	150	32	115	406	1396

Here *G* denotes the known parental genotype on three linked loci, with *Aa Bb Cc* for diploids and *Aaaa Bbbb Cccc* for autotetraploids. *O* represents the phenotype records of the *n* gametes on the three marker loci. f_i is the frequency of gamete with the i^{th} phenotype.

In analysis of diploids, the probability of the gamete with the i^{th} phenotype can be calculated as

$$f_i = \sum_{h \in i} G_h = \sum_{h \in i} \frac{1}{2} \sum_{s=0}^{1} \sum_{t=0}^{1} w_{hst} x_{st} \qquad \text{(II-1.52)}$$

where G_h is a general form for frequency of the gametic genotype *h* defined in Equation (II-1.16) of Section 1.2.5 and $\sum_{h \in i} G_h$ means summing up the probabilities of those genotypes that are compatible to the same phenotype *i*.

To calculate the MLE of the model parameters d_1 and d_2 at any given value of m, the E-step calculates the conditional probability of *k* crossovers within the first marker interval in the gamete with the i^{th} phenotype as

$$\gamma_{ik} = \frac{1}{2} \sum_{h \in i} \sum_{s=0}^{1} \sum_{t=0}^{1} w_{hst} a_{st}^{k} \Big/ f_i \qquad if\ s \neq 0, k\ can\ not\ be\ 0 \qquad \text{(II-1.53)}$$

and also the conditional probability of *l* crossovers within the second marker interval in the gamete with the i^{th} phenotype as

$$\omega_{il} = \frac{1}{2} \sum_{h \in i} \sum_{s=0}^{1} \sum_{t=0}^{1} w_{hst} b_{st}^{k} \Big/ f_i \qquad if\ t \neq 0, k\ can\ not\ be\ 0 \qquad \text{(II-1.54)}$$

where w_{ist} is the number of gametes with recombination configuration $X=(s\ t)$ within the i^{th} gametic genotype class. a_{st}^{k} and b_{st}^{k} are defined in Section 1.2.6 on page 19 to 21.

Since each strand made up of the four-strand bundle has a chance of ½ to be involved in each crossover in diploids, the M-step updates the estimates of the genetic distances from

$$d_1^{'}=\frac{1}{2}\left(\sum_{i=1}^{8}n_i\sum_{k\geq 1}k\gamma_{ik}\Big/n\right) \tag{II-1.55}$$

$$d_2^{'}=\frac{1}{2}\left(\sum_{i=1}^{8}n_i\sum_{k\geq 1}k\omega_{ik}\Big/n\right) \tag{II-1.56}$$

As the E-step and M-step repeated, the parameter estimates converge to the MLEs conditional for a given integer value of parameter, *m*.

In analysis of autotetraploids, the probability of the gamete with the i^{th} phenotype can be calculated as

$$f_i=\sum_{h\in i}G_h=\sum_{h\in i}\sum_{u\in h}g_u=\sum_{h\in i}\sum_{u\in h}\sigma_u\alpha^{I_{\alpha,u}}(1-\alpha)^{1-I_{\alpha,u}}C_{s_ut_u/s'_ut'_u}^{I_{\alpha,u},I_{\beta,u},e_{1,u},e_{2,u}} \tag{II-1.57}$$

where G_h is a general form for frequency of the gametic genotype *h* defined in Equation (II-1.35) of Section 1.3.3.

The E-step of the EM algorithm calculates the conditional probability that gametes of the i^{th} phenotype carry k crossovers generated during meiosis of parent within the first marker interval, γ_{ik}.

$$\gamma_{ik} = \sum_{h\in i}\sum_{u\in h} \sigma_h \alpha^{I_{\alpha,u}} (1-\alpha)^{1-I_{\alpha,u}} A^{I_{\alpha,u},I_{\beta,u},k}_{s_u t_u / s'_u t'_u, e_{1,u}, e_{2,u}} \Big/ f_i \quad \text{(II-1.58)}$$

Here σ_h is defined in Equation (II-1.34) and $A^{I_{\alpha,u},I_{\beta,u},k}_{s_u t_u / s'_u t'_u, e_{1,u}, e_{2,u}}$ can be calculated from Equation (II-1.41). α is the coefficient of double reduction and $I_{\alpha,u}$ takes value of 1 or 0 to represent presence or absence of double reduction on the first marker locus.

The conditional probability of individuals of the i^{th} phenotype with k crossovers from parent within the second marker interval, ω_{ik}, is:

$$\omega_{ik} = \sum_{h\in i}\sum_{u\in g} \sigma_h \alpha^{I_{\alpha,u}} (1-\alpha)^{1-I_{\alpha,u}} B^{I_{\alpha,u},I_{\beta,u},k}_{s_u t_u / s'_u t'_u, e_{1,u}, e_{2,u}} \Big/ f_i \quad \text{(II-1.59)}$$

where $B^{I_{\alpha,u},I_{\beta,u},k}_{s_u t_u / s'_u t'_u, e_{1,u}, e_{2,u}}$ can be calculated from Equation (II-1.42).

The conditional probability of gametes of the i^{th} phenotype with k double-reduction gametes, ξ_{ik}, is calculated as

$$\xi_{ik} = \sum_{h\in i} \sum_{u\in g, I_{\alpha,u}=k} \sigma_h \alpha^{I_{\alpha,u}} C^{I_{\alpha,u},I_{\beta,u},e_{1,u},e_{2,u}}_{s_u t_u / s'_u t'_u} \Big/ f_i \quad \text{(II-1.60)}$$

where $C^{I_{\alpha,u},I_{\beta,u},e_{1,u},e_{2,u}}_{s_u t_u / s'_u t'_u}$ is given in Equation (II-1.33).

Since each chromosomal strand has a chance of ¼ to be involved in each crossover under both the bivalent and quadrivalent pairing models, the M step updates the estimates of the parameters of genetic distances from:

$$d_1^{'} = \frac{1}{4}\left(\sum_{i=1}^{8} n_i \sum_{k=1}^{\infty} k\gamma_{ik} \Big/ n \right) \quad \text{(II-1.61)}$$

$$d_2^{'} = \frac{1}{4}\left(\sum_{i=1}^{8} n_i \sum_{k=1}^{\infty} k\omega_{ik} \Big/ n \right) \quad \text{(II-1.62)}$$

and estimates the coefficient of double reduction from

$$\alpha' = \sum_{i=1}^{8} n_i \sum_{k=0}^{1} k\xi_{ik} \Big/ n \quad \text{(II-1.63)}$$

The likelihood function increases monotonically as the E step and M step repeat and the parameter estimates converge to the MLEs conditional on the given integer parameter, *m*.

For both diploids and autotetraploids, I calculated MLEs of parameters given values of m from 0 to 5 and infer the most likely coefficient of crossover interference, *m*, by examining the likelihood. The results of statistical inference were shown in Table II-1.9 with estimated optimal *m*, genetic distances and coefficient of double reduction for autotetraploids with quadrivalent pairing. For autotetraploids, the analysis was carried out under the assumption of bivalent pairing (denoted as Tetraploids[1]) or quadrivalent pairing (denoted as Tetraploids[2]).

To test the goodness of fit under the proposed $C_x\left(C_0\right)^m$ model, the Pearson chi-squared statistic, χ^2, and corresponding P value were calculated (Table II-1.9). The test statistic is calculated as

$$\chi^2 = \sum_{i=1}^{8} \frac{\left(n_i - f_i \cdot n\right)^2}{f_i \cdot n} \tag{II-1.64}$$

Here n_i is the observed gamete counts of the i^{th} phenotype category and n is the total number of generated gametes. f_i is the expected frequency of gamete with the i^{th} phenotype under the model with MLEs of parameters and can be calculated according to Equation (II-1.52) for diploids and Equation (II-1.57) for autotetraploids. The degree of freedom is equal to 7.

It can be seen that the model fitted well to the data from diploid yeast (high P values). A mild degree of crossover interference was observed on chromosomes 3 and 6 for which the MLE of the integer shape parameter, *m*, was larger than 0. In the analysis of autotetraploid data, the model fitted the data reasonably under the assumption of quadrivalent pairing, but the goodness of fit was notably improved under the assumption of bivalent pairing. We infer that bivalent pairing of chromosomes was more likely than quadrivalent pairing in this autotetraploid yeast.

To test the significance of crossover interference along the three chromosomes when *m* exceeded 0 in either diploids or autotetraploids, the likelihood-ratio test statistic was calculated by:

$$R^2 = 2\left\{\ln\left[L\left(P_1, P_2, \hat{m}, \alpha, d_1, d_2 \mid O\right)\right] - \ln\left[L\left(P_1, P_2, m=0, \alpha, d_1, d_2 \mid O\right)\right]\right\} \tag{II-1.65}$$

Table II-1.9. Statistical inference of genetic parameters from data of diploid and autotetraploid *S.cerevisiae*

Chromosome 3						
	m	d_1	d_2	α	χ^2 (P_value)	Likelihood ratio test (P_value)
Diploids	1	0.148	0.323		0.006 (1.000)	2.318 (0.130)
Tetraploids[1]	0	0.150	0.391		0.142 (0.986)	-
Tetraploids[2]	0	0.203	0.547	0.028	3.413 (0.844)	-
Chromosome 6						
	m	d_1	d_2	α	χ^2 (P_value)	Likelihood ratio test (P_value)
Diploids	2	0.255	0.104		0.044 (0.998)	9.955 (0.002)
Tetraploids[1]	0	0.493	0.169		0.151 (0.985)	-
Tetraploids[2]	0	0.679	0.236	0.001	8.354 (0.302)	-
Chromosome 8						
	m	d_1	d_2	α	χ^2 (P_value)	Likelihood ratio test (P_value)
Diploids	0	0.392	0.226		0.340 (0.952)	-
Tetraploids[1]	1	0.304	0.243		0.077 (0.994)	2.029 (0.150)
Tetraploids[2]	0	0.504	0.400	0.000	9.278 (0.233)	-

Here d_1, d_2 indicate the genetic distances for the first and second interval, respectively. α is the coefficient of double reduction under quadrivalent pairing of autotetraploids. Tetraploids[1] represents the data set is analyzed under the assumption of bivalent pairing during meiosis, and tetraploids[2] represents the data set is analysed under the assumption of quadrivalent pairing.

As shown in Table II-1.9, significant crossover interference was found on chromosome 6 (P value 0.002) in diploid yeast with *m* equal to 2, but was absent in the corresponding autotetraploid with either bivalent or quadrivalent pairing assumptions (*m*=0). Correspondingly, we can see that compared with other chromosomes, the estimated genetic distances between markers on chromosome 6 showed a greater increase in the autotetraploids compared with diploids.

1.6. Discussion

Theoretical analysis of crossover interference has been a historically challenging area since 1915 and an important topic in genome research (Sturtevant, 1915; Muller, 1916). Although some progress has been made in this field, the biological nature of crossover interference is still not adequately understood (Haldane, 1931; McPeek and Speed, 1995). The present study addresses some key problems in statistical inference of crossover interference in both diploids and autotetraploids. In diploids, Zhao *et al* (1995) have already proposed a $C_x(C_o)^m$ model which mathematically formulate crossover process and fitted well to genetic data from various organisms. In the present study, I extended this model to analyse three-locus data for autotetraploids and proposed an EM algorithm to obtain the MLE of model parameters.

To address crossover interference in autotetraploids, I have developed a new model for the distribution of offspring genotypes from a cross between two parents at three linked loci in terms of the genetic distances of the two marker intervals, the coefficient of crossover interference and the coefficient of double reduction (where double reduction is present). This

model takes into account several key properties of tetrasomic inheritance and can be applied to the case of both bivalent pairing and quadrivalent pairing meiosis. These features include alleles with multiple dosages, allelic segregation distortion due to double reduction, the existence of null alleles, homologous chromosome pairing pattern during meiosis, and incomplete information of phenotype in regard to genotype. The EM algorithm was developed to calculate the MLEs of the model parameters. This work has therefore filled a longstanding theoretical and methodological gap in the genetic analysis of crossover interference in autotetraploid species. The feasibility of our new method in parameter estimation from genetic data of diploid or autotetraploid species was demonstrated through extensive simulation analysis and the analysis of real data from large populations of diploid and autotetraploid species of the yeast *Saccharomyces cerevisiae*.

In our model, the coefficient of crossover interference, *m*, has the same definition in both diploids and autotetraploids, which providing a convenient way to compare the degree of crossover interference between diploids species and their corresponding autotetraploid relatives. All flowering plants have experienced at least one polyploidization event during their evolutionary history (Jiao et al., 2011), and as such, polyploidization has been an important driving force in evolutionary of plants (Chen, 2007; Soltis and Soltis, 2009). Theoretical and experimental evidence suggests that recombination frequency is increased in autotetraploid plants compared with their parental diploids (Pecinka et al., 2011; Wang and Luo, 2012), but little is known about the underlying mechanism. The real data analysis in the present study described a direct comparison of crossover interference in diploids and their corresponding autotetraploid species. We found evidence for a decrease in the strength of crossover interference after polyploidization on one of the three chromosomes analysed, suggesting a new

hypothesis worthy of future exploration that an increase in recombination in autotetraploids compared with diploids could be explained by a corresponding decrease in the level of crossover interference.

At the same time, this theoretical development in autotetraploids provides a way to calculate the probability distribution of two/three-locus gametes/zygotes in terms of the genetic distances between loci, the coefficient of double reduction and the coefficient of crossover interference, which would be helpful to incorporate varying degrees of crossover interference into the linkage analysis in autotetraploid species.

In the current real data analysis, parental genotypes on three marker loci were very special and quite simple, which have only two different alleles and alleles denoted by capital letters are linked on the same chromosome. In this special case, the traditional method for inference of recombination interference could be applied to analyse the gamete data for autotetraploids under the assumption of bivalent pairing. Thus using the statistical method I proposed here to analyse this kind of data would cause another problem named over-paramerization. To have a better understanding of the change of crossover interference after polyploidization, it is better to have more general and larger dataset analysis in different autotetraploid species.

1.7. References

Blanc, G., et al. (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. **Curr Biol,** 21: 551-556.

Bauer, E., et al. (2013) Intraspecific variation in recombination rate in maize. **Genome Biol.** 14: R103.

Carter, T.C. and Robertson, A. (1952) A mathematical treatment of genetic recombination using a four strand model. **Proc.Royal Soc.** London, B 139: 410-426.

Chen, Z. (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. **Annu Rev Plant Biol,** 58: 377-406.

Dempster, A.P. (1977) Maximum likelihood from incomplete data via the EM algorithm. **JR Statist SocB,** 39: 1-22.

Drouaud, J., et al. (2007) Sex-specific crossover distribution and variations in interference level along *Arabidopsis thaliana* chromosome 4. **PLoS Genet.** 3: 1096-1107.

Foss, E., et al. (1993) Chiasma interference as a function of genetic distance. **Genetics**, 133: 681-691.

Haldane, J.B.S. (1931) The cytological basis of genetical interference. **Cytologia**, 3(1): 54-65

Hawthorne, D.C. and Mortimer, R.K. (1960) Chromosome mapping in Saccharomyces: centromere-linked genes. **Genetics**, 45: 1085-1110.

Jiao, Y., et al. (2011) Ancestral polyploidy in seed plants and angiosperms. **Nature**, 473: 97-100.

Karlin, S. and Liberman, U. (1979) A natural class of multilocus recombination processes and related measures of crossover interference. **Adv. Appl. Probab.,** 11: 479-501.

Lindegren, C.C. and Lindegren, G. (1942) Locally specific patterns of chromatid and chromosome interference in Neurospora. **Genetics,** 27: 1-24.

Lange, E., et al. (1997) The poisson-skip model of crossing-over. **The Annals of Applied Probability,** 7(2): 299-313.

Luo, Z.W., et al. (2000) Predicting parental genotypes and gene segregation for tetrasomic inheritance. **Theor Appl Genet**, 100: 1067-1073.

Luo, Z.W., et al. (2004) Theoretical basis for genetic linkage analysis in autotetraploid species. **PNAS,** 101: 7040-7045.

Luo, Z.W., et al. (2006) Constructing genetic linkage maps under a tetrasomic model. **Genetics,** 172: 2635-2645.

Lu, Y.F., et al. (2013) A multivalent three-point linkage analysis model of auotetraplois. **Briefing in bioinformatics,** 14(4): 460-468.

Muller, H.J. (1916) The mechanism of crossing-over**. Am. Nat.,** 50: 193-221, 284-305,421-434.

Mather, K. (1936) The determination of position in crossing over. **J. Genet.,** 33: 207-235.

Mather, K. (1937) The determination of position in crossing over II. The chromosome length-chiasma frequency relation. **Cytologia Jub**., 514-526.

McPeek, M. and Speed, T.P. (1995) Modelling interference in genetic recombination. **Genetics,** 139: 1031-44.

Malkova, A., et al. (2004) Gene conversion and crossing over along the 405-kb left arm of *Saccharomyces cerevisiae* chromosome VII. **Genetics** 168: 49-63.

Owen, A.R.G. (1950) The theory of genetical recombination. **Advan. Genet.,** 3: 117-157.

Osman, K., et al. (2011) Pathways to meiotic recombination in Arabidopsis thaliana. **New Phytologist,** 190(3): 523-544.

Pecinka, A., et al. (2011) Polyploidization increases meiotic recombination frequency in Arabidopsis. **BMC Biology,** 9:24.

Sturtevant, A.H. (1915) The behaviour of the chromosomes as studied through linkage. **Z. Indukt. Abstammungs. Vererbungsl.,** 13: 234-287.

Strickland, W.N. (1958) An analysis of interference in Aspergillus nidulans. **Proc. Roy. Soc. Lond. Ser. B**, 149: 82-101.

Stam, P. (1979) Interference in genetic crossing over and chromosome mapping. **Genetics,** 92: 573-594

Soltis, P., et al. (2009) The role of hybridization in plant speciation. **Ann Rev Plant Biol,** 60: 561-588.

Wang, L., et al. (2012) Polyploidization increases meiotic recombination frequenc y in Arabidopsis: a close look at statistical modelling and data analysis. **BMC biology**, 10:30.

Zhao, H., et al. (1995) Statistical analysis of crossover interference using the chi-square model. **Genetics,** 139(2): 1045-56.

Chapter II-2: Predicting meiotic crossover rate in *Saccharomyces cerevisiae* based on the whole genome wide sequencing data analysis for autotetraploids

2.1. Overview

Polyploidization plays a very important role in the evolutionary history of plants in nature and under domestication (Soltis 1995, Otto 2000, Comai 2005). More and more evidence has been found to support that polyploidization would increase meiotic recombination frequency and result rapid creation of genetic diversity (Pecinka 2011, Wang 2011). To investigate the underlying genetic mechanism, we would like to understand the dynamic change of crossover events during meiosis after polyploidization. One aspect is to compare the meiotic crossover rate, defined as the expected number of crossover events occurring on a chromatid, between diploids and autotetraploids.

Crossovers are essential for reciprocal exchange of genetic material during meiosis in most eukaryotes, which would result in the outcome recombination events. This process increases genetic diversity and is tightly regulated. In diploids, high-resolution mapping of crossover events can be achieved by monitoring recombination between closely located markers along a chromatid based on sequencing data (Mancera 2008). However, it is impractical to observe all the recombination events in autotetraploids due to the existence of multiplex alleles.

In the context here I first proposed a likelihood-based method to predict crossover rate in autotetraploids. Second, we applied next generation sequencing approach to all four spores

derived from meiosis of both diploid and its related autotetraploid *Saccharomyces cerevisiae* and obtained a set of genotype data called from the intensely distributed SNP markers. Using our statistical inference method, we found that crossover rate significantly increased in autotetraploid yeast than that in diploid yeast.

2.2. Methods

We assume SNP markers were intensely distributed on a chromosome and at most only one crossover event may occur within a marker interval due to the high resolution. To make it comparable between diploids and autotetraploids, the crossover rate, *p*, is defined as the probability of occurring one crossover on a chromatid within the marker interval. Consider a marker interval and we focused here gametogenesis of a diploid individual with genotype, *AB/ab*, and an autotetraploid individual with genotype, *AB/ab/ab/ab*, with *A* and *B* corresponding to s288c (SK1) alleles, and *a* and *b* to SK1 (or s288c) alleles in the autotetraploid strain s288c/SK1/SK1/SK1 (or SK1/s288c/s288c/s288c).

2.2.1. Counting crossover events in diploids

In diploids, it is directly to observe crossover events according to the SNP marker genotypes of all spores in the tetrad called from sequence data. There are only two two-locus genotype categories for the tetrad: $AAaa/BBbb$ for non-crossover and $AAaa/BbBb$ for crossover. For a particular chromatid, it can get involved into two different crossovers among the total four

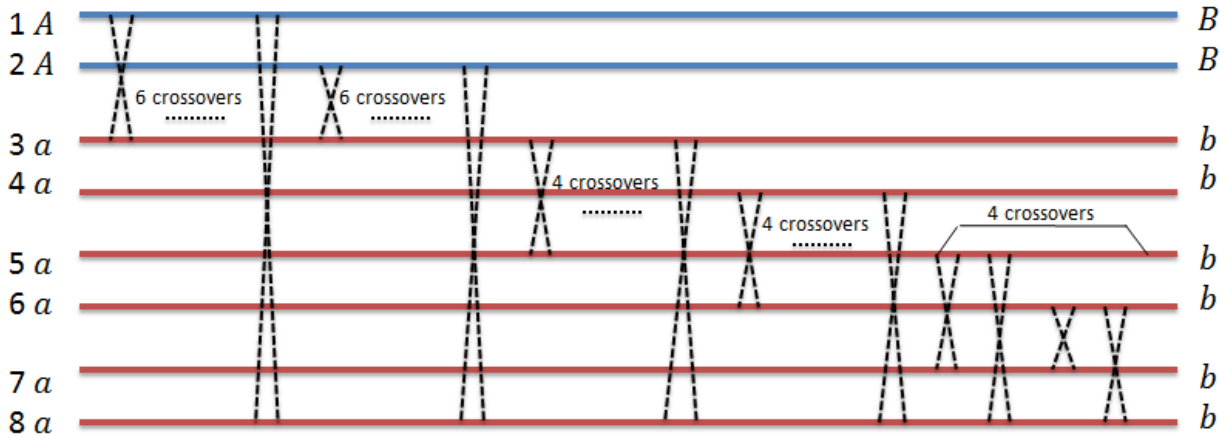
distinct crossovers. To make it comparable, the expected number of crossover events on a chromatid equals to half of the number of marker intervals with crossover occurring.

2.2.2. Predicting the average number of crossovers in autotetraploids

In autotetraploids, although we can observe the genotype of tetrad on each marker locus from sequence data, we do not know whether alleles on consecutive markers are linked in coupling or in repulsion in the diploid spore. Assume there are *N* marker intervals intensely distributed on a chromosome. The coefficient of double reduction at the flanking marker locus, which is nearer to the centromere, is denoted as α and the probability of occuring a crossover event within a marker interval is *p'*. We considered the crossover occurring between all possible non-sister chromatids and all possible configurations of diploid gamete generation under a tetrasomic model, and worked out distribution of possible tetrads at the two marker loci in terms of α and *p'* as follows:

During meiosis in autotetraploids under quadrivalent pairing, crossovers may occur between any two non-sister chromatids, which lead to twenty-four different crossovers as shown in Figure II-2.1.

Figure II-2.1. Twenty-four different crossovers between any two non-sister chromatids in autotetraploids with quadrivalent pairing



Here the eight solid lines with different colours represent four duplicated chromosomes in autotetraploids, flanking by markers A and B. The black forks in dotted lines indicate twenty-four different crossovers involving any two non-sister chromatids.

1. No crossover occurs within the marker interval

1.1. Double reduction occurs on locus A of the first chromosome

In the case of no crossover occurring within the marker interval and double reduction occurring on locus A of the first chromosome, the outcome tetrad would be

and the corresponding probability is $(1-p')\alpha$.

1.2. No double reduction occurs on locus A of the first chromosome

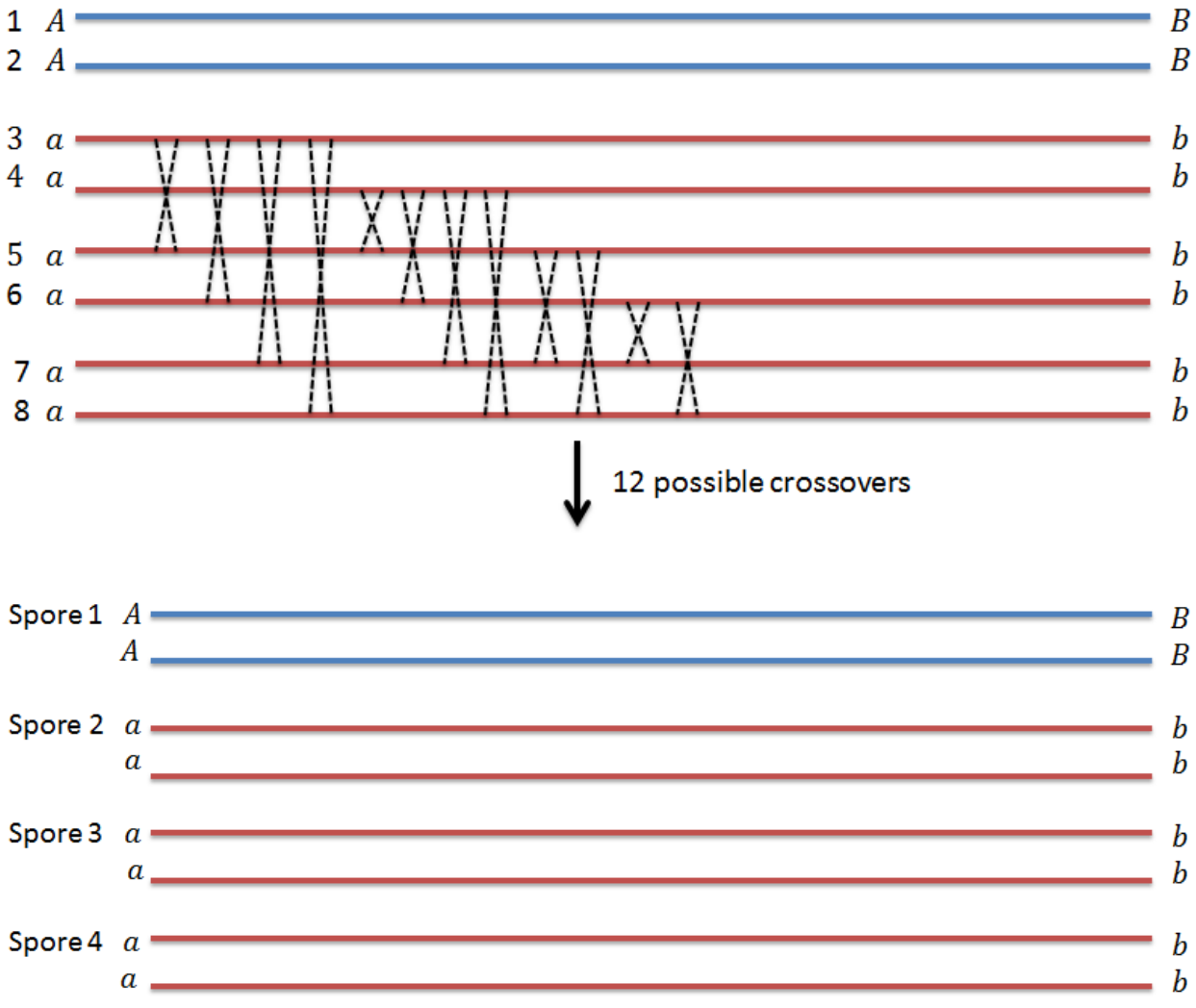
In the case of no crossover occurring within the marker interval and no double reduction occurring on locus A of the first chromosome, the outcome tetrad would be

and the corresponding probability is $(1-p')(1-\alpha)$.

2. One crossover occurs within the marker interval

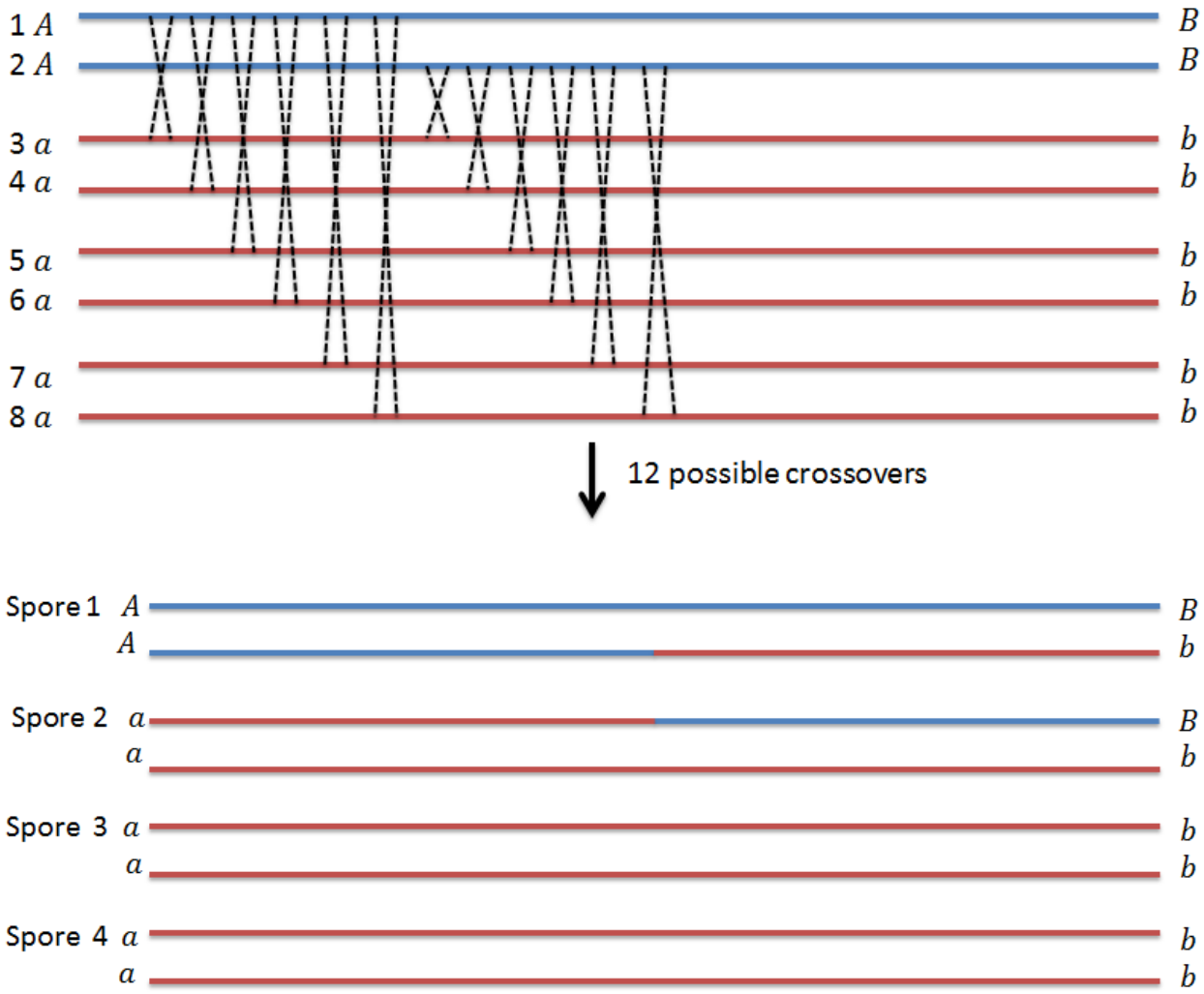
2.1. Double reduction occurs on locus A of the first chromosome

In the case of one crossover occurring within the marker interval and double reduction occurring on locus A of the first chromosome, there will be two different outcome tetrads depending on which two strands involved into the crossover. The first one is generated as



and the corresponding probability is $p'\alpha/2$.

The second one is generated as

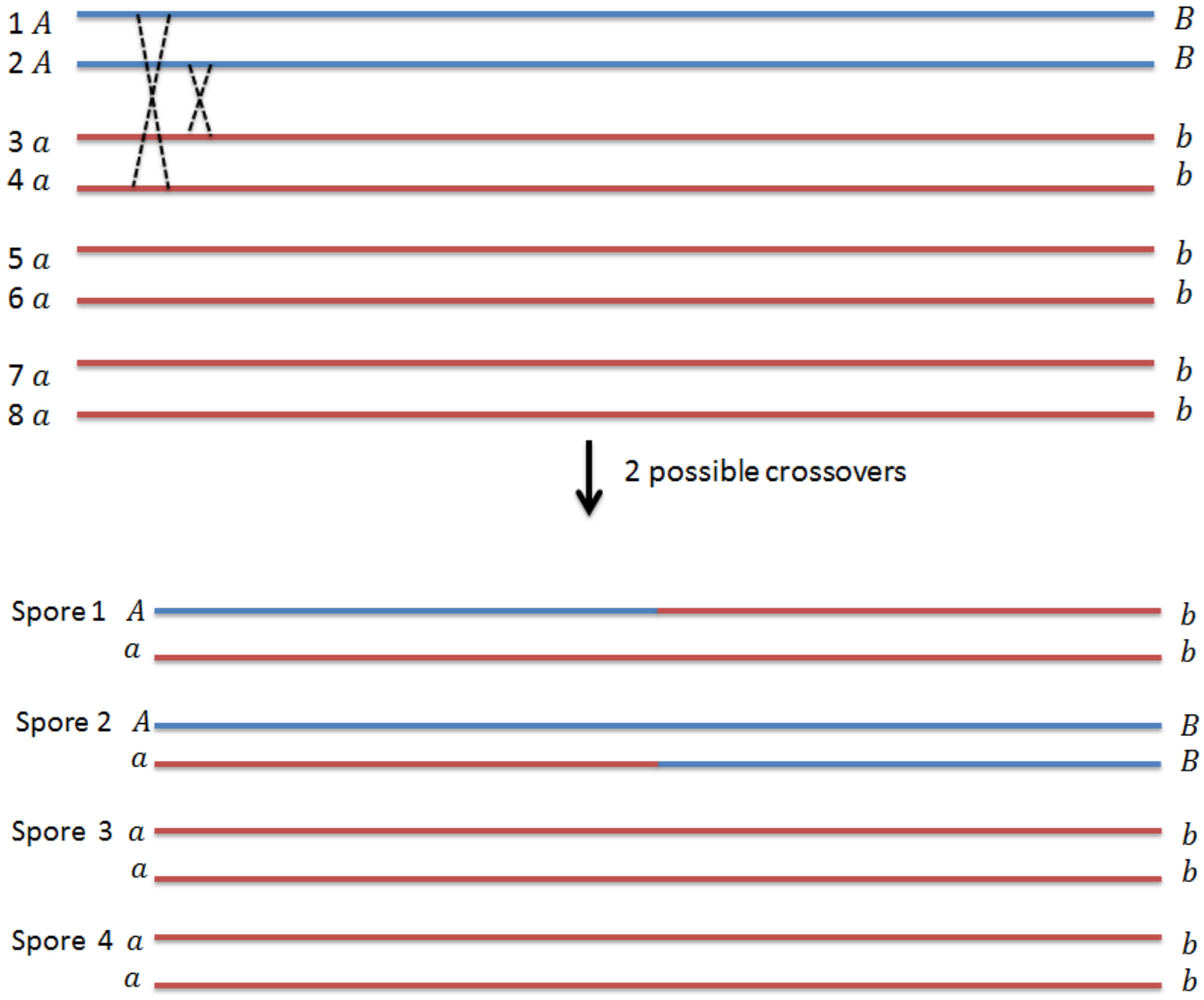


and the corresponding probability is $p'\alpha/2$.

2.2. No double reduction occurs on locus A of the first chromosome

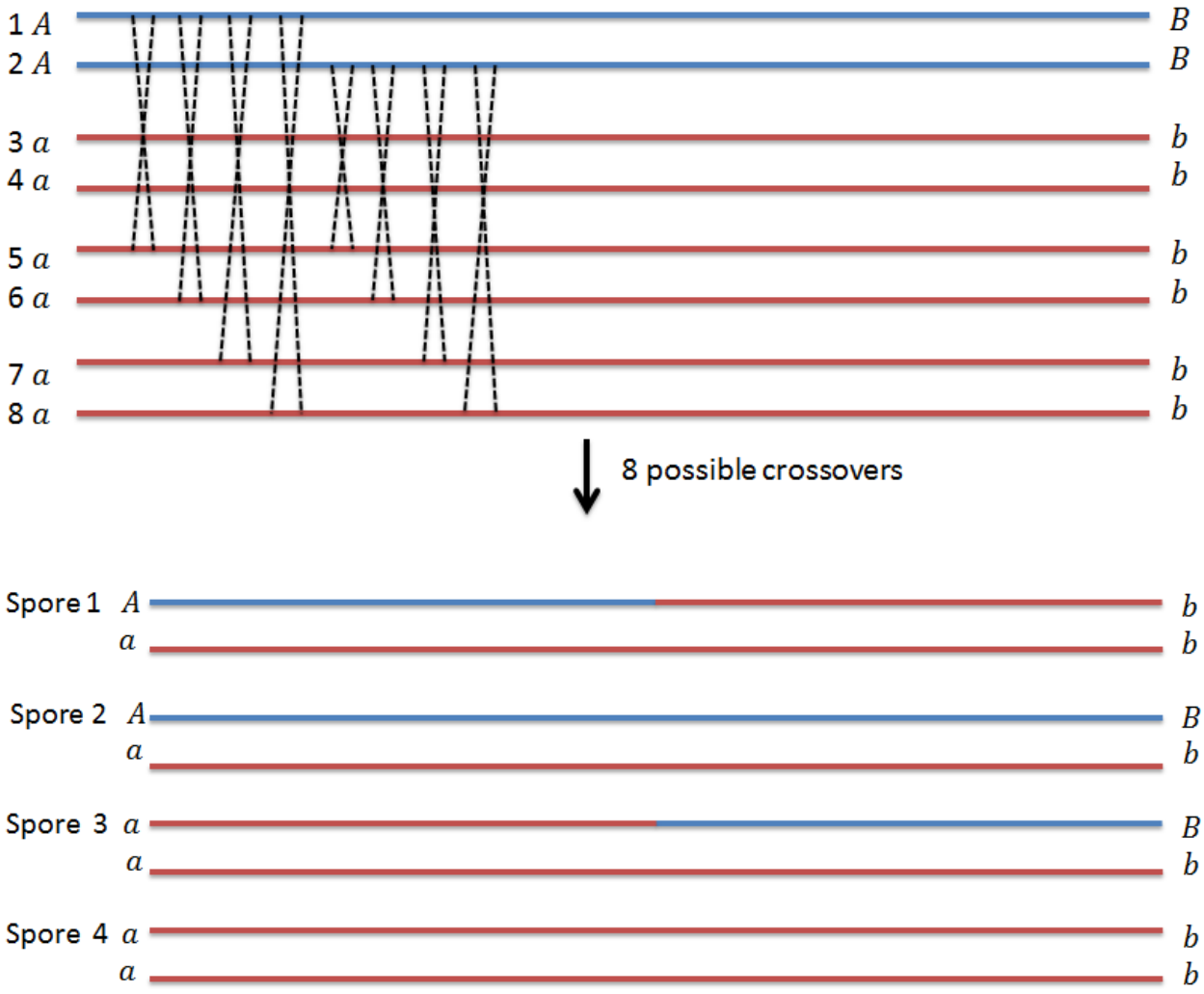
In the case of one crossover occurring within the marker interval and no double reduction occurring on locus A of the first chromosome, there will be four different outcome tetrads depending on which two strands involved into the crossover. For illustration purpose but without

loss of generality, we consider that strand 1 and strand 3 would enter into the same spore, and strand 2 and strand 4 would enter into the same spore. The first one is generated as



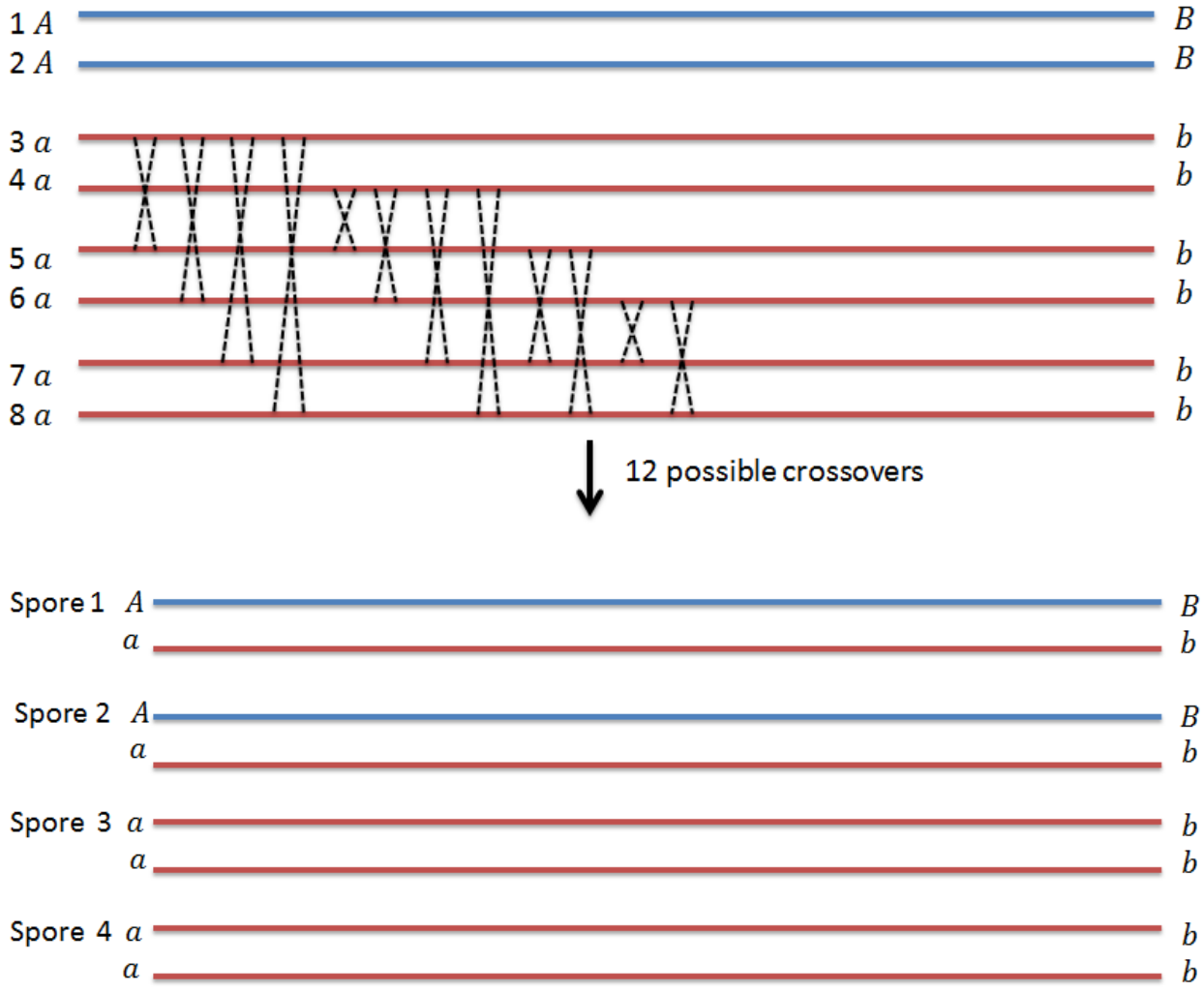
and the corresponding probability is $p'(1-\alpha)/12$.

The second one is generated as



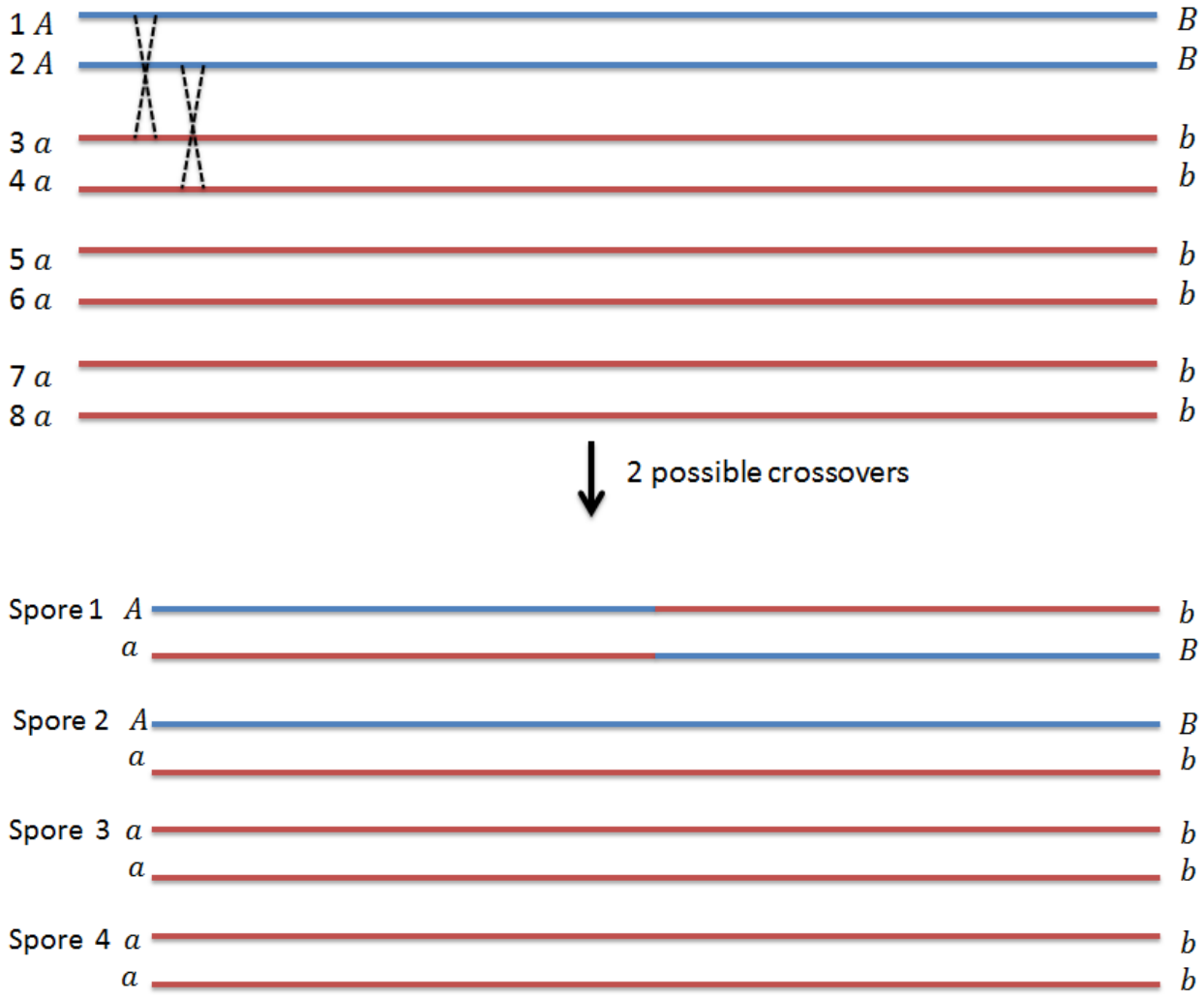
and the corresponding probability is $p'(1-\alpha)/3$.

The third one is generated as



and the corresponding probability is $p'(1-\alpha)/2$.

The fourth one is generated as



and the corresponding probability is $p'(1-\alpha)/12$.

Based all the above possible configurations of tetrad generation under a tetrasomic model, I worked out distribution of phenotype of five possible tetrads at the two marker loci in term of α and *p'*, which was listed as Table II-2.1. In the distribution, a tetrad phenotype was presented as two sequential integers representing two chromosomes. A non-zero integer in the sequence represented the number of A or B alleles and the four integers referred to the four spores.

Table II-2.1. Distribution of tetrad phenotype of two linked SNP markers generated from an autotetraploid parental strain *AB/ab/ab/ab*.

Phenotype	n_i	Underlying Spore Genotype				Probability
		Spore1	Spore2	Spore3	Spore4	
2000/2000	n_1	*AA/BB*	*aa/bb*	*aa/bb*	*aa/bb*	$\alpha(1-p'/2)$
1100/1100	n_2	*Aa/Bb*	*Aa/Bb*	*aa/bb*	*aa/bb*	$(1-\alpha)(1-5p'/12)$
2000/1100	n_3	*AA/Bb*	*aa/Bb*	*aa/bb*	*aa/bb*	$\alpha p'/2$
1100/0200	n_4	*Aa/bb*	*Aa/BB*	*aa/bb*	*aa/bb*	$(1-\alpha)p'/12$
1100/0110	n_5	*Aa/bb*	*Aa/Bb*	*aa/Bb*	*aa/bb*	$(1-\alpha)p'/3$

Scoring on the four spores of the two loci tetrad follows the rules: 2 represents *AA* (*BB*), 1 represents *Aa*(*Bb*) and 0 represent *aa*(*bb*). Here α is the coefficient of double reduction on locus A and p' is the probability of occurring a crossover event within the marker interval.

For a sample of *N* two-locus tetrad phenotypes, let $n_i\,(i=1,\ldots,5)$ be the number of two-locus tetrad with the i^{th} marker phenotype. Then the log-likelihood function of the model parameters α and *p* can be written as

$$L(\alpha, p \mid n_i) \propto \sum_{i=1}^{5} n_i \log(f_i) = n_1 \log\{\alpha(1-p/2)\} + n_2 \log\{(1-\alpha)(1-5p/12)\} + n_3 \log\{\alpha p/2\} + n_4 \log\{(1-\alpha)p/12\} + n_5 \log\{(1-\alpha)p/3\} \quad \text{(II-2.1)}$$

The derivative of likelihood with respect to *p* equals to

$$\log'(L) = n_1 \frac{1}{1-\frac{1}{2}p'}\left(-\frac{1}{2}\right) + n_2 \frac{1}{1-\frac{5}{12}p'}\left(-\frac{5}{12}\right) + (n_3+n_4+n_5)\frac{1}{p'}$$
$$= \frac{n_1}{p'-2} + \frac{5n_2}{5p'-12} + (n_3+n_4+n_5)\frac{1}{p'} \quad \text{(II-2.2)}$$

Let $\log'(L) = 0$, the likelihood function reaches to maximum and we can get the most likely estimates of *p'*. Since $\sum_{i=1}^{5} n_i = N$, the normal equation is given by

$$5Np'^2 - 2(11N - 5n_1 - 6n_2)p' + 24(N - n_1 - n_2) = 0 \quad \text{(II-2.3)}$$

There are two roots for the equation above,

$$\hat{p}'_1 = \frac{11N - 5n_1 - 6n_2 + \sqrt{N^2 + 2N(5n_1 - 6n_2) + (5n_1 + 6n_2)^2}}{5N} \quad \text{(II-2.4)}$$

$$\hat{p}'_2 = \frac{11N - 5n_1 - 6n_2 - \sqrt{N^2 + 2N(5n_1 - 6n_2) + (5n_1 + 6n_2)^2}}{5N} \quad \text{(II-2.5)}$$

Since a meaningful estimate of crossover rate should be fallen in the range of [0, 1], the most likely estimate of *p'* equals to $\hat{p}'_2$.

In autotetraploids, there are two different kinds of two-locus tetrads: from the phenotype of the first one we can directly observe a crossover, which has a number of N_1, and from the phenotype of the second we can only expect a crossover occurring underlying with a probability of $\hat{p}'_2$, which has a number of ($N - N_1$). The expected number of crossover events occurring among the

total N marker intervals should equal to $N_1+\left(N-N_1\right)\hat{p}_2^{'}$. For a particular chromatid in autotetraploids, it can get involved into six different crossovers among the total twenty-four distinct crossovers. So the expected number of crossover events occur on a chromatid in autotetraploids equals to $\left(N_1+\left(N-N_1\right)\hat{p}_2^{'}\right)/4$.

2.3. Real data analysis

The statistical method proposed was used to analyse a set of sequence data for all four meiotic products of diploid and autotetraploid yeast. Firstly, our collaborator used a haploid strain to construct diploid and autotetraploid strains. Diploid or autotetraploid strains were then sporulated in the way detailed as: Freshly created diploid and autotetraploid cells were streaked out to create single colonies on the YPD plate. After 2 days of colony development, 3 large and healthy colonies were patched on a new YPD plate and grew for 13.5 hours, then transferred to the SPM plate (1% KAC) and incubated at 30 ℃. Genomic DNA was then extracted from single-colony cultures of the tetrads and sheared into fragments with an average length of 200 bp using the Covaris S220 (Duty Factor =10%, Intensity Peak Incident Power =140W, Cycles per Burst = 200, Processing Time = 180 seconds, Volume = 130 μl in microtubes). The DNA fragments were then purified by use of the QIAGEN minelute gel extraction kit. Sequencing library was prepared using the NEBNext Ultra DNA Library Prep Kit designed for illumine and whole genome sequencing was performed using an illumina Hiseq-2000 sequencer with a design to generate 2*100 bp paired reads.

We achieved a detailed characterization of recombination outcomes by calling 30,000~60,000 SNP marker in all four spores derived from four meiosis of diploid yeast and seven meiosis of autotetraploid yeast. To make them comparable, we have to consider the common markers among all the individuals, which reached to 8653. Since we assume that there is at most one crossover event occurring within the marker interval, effective marker intervals were selected with length between 2500 to 10000 bp and we got 1950 effective marker interval totally. Parental genotype on the markers can be represented by HS for all the diploid individuals and by HSSS (four samples) or HHHS (three samples) for the autotetraploid individuals.

Crossover counting estimation under the model above was carried out on all the fifteen chromosomes (except chromosome 3 which has large DNA segment missing in some autotetraploid tetrads) of each individual and summarized in Table II-2.2. Applying general

Table II-2.2. Crossover rate estimation for both diploid and autotetraploid samples

Chr	Common marker	Effective marker intervals	HS1		HS2		HS3		HS4		HS5		HSSS1		HSSS2		HSSS3		HSSS4		HHHS1		HHHS2		HHHS3	
			n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
1	128	28	1	0.5	2	1	2	1	5	2.5	3	1.5	1	0.829	2	1.614	0	0	0	0	2	1.565	1	0.829	3	2.357
2	637	154	3	1.5	1	0.5	3	1.5	3	1.5	2	1	5	4.129	2	1.684	3	2.515	4	3.338	3	2.443	3	2.447	6	4.96
4	1137	284	5	2.5	7	3.5	6	5	7	3.5	8	4	6	5.024	5	4.197	3	2.531	5	4.197	7	5.723	7	5.846	3	2.503
5	547	109	1	0.5	3	1.5	3	1.5	2	1	2	1	0	0	5	4.112	2	1.678	1	0.844	3	2.5	2	1.678	3	2.5
6	446	62	2	1	3	1.5	1	0.5	0	0	1	0.5	0	0	2	1.661	2	1.661	2	1.661	0	0	0	0	3	2.463
7	797	186	4	2	6	3	5	2.5	3	1.5	6	3	3	2.521	3	2.521	3	2.521	3	2.521	7	5.792	7	5.734	7	5.622
8	466	103	0	0	1	0.5	1	0.5	2	1	5	2.5	2	1.677	5	4.104	3	2.498	1	0.844	0	0	2	1.677	2	1.677
9	425	84	3	1.5	1	0.5	2	1	1	0.5	2	1	2	1.671	1	0.843	3	2.411	3	2.486	6	4.843	6	4.843	1	0.822
10	344	90	0	0	1	0.5	2	1	3	1.5	3	1.5	2	1.673	4	3.293	3	2.49	3	2.49	1	0.843	4	3.277	2	1.673
11	405	103	2	1	4	2	2	1	1	0.5	2	1	2	1.677	1	0.82	2	1.677	0	0	2	1.661	3	2.498	1	0.843
12	707	175	2	1	5	2.5	6	3	4	2	5	2.5	9	7.341	3	2.493	4	3.345	2	1.587	5	3.999	5	4.164	5	4.164
13	774	171	4	2	5	2.5	6	3	4	2	2	1	1	0.846	2	1.686	2	1.686	3	2.518	4	3.344	2	1.636	3	2.518
14	137	25	1	0.5	0	0	2	1	0	0	0	0	0	0	1	0.826	0	0	1	0.826	0	0	1	0.826	3	2.334
15	911	205	5	2.5	4	2	5	2.5	6	3	5	2.5	6	4.995	9	7.413	5	4.177	2	1.688	3	2.524	7	5.807	4	3.353
16	792	171	3	1.5	3	1.5	5	2.5	8	4	6	3	6	4.974	5	4.162	5	4.162	3	2.518	5	4.162	4	3.221	6	4.974
sum	8653	1950	36	**18**	46	**23**	49	**24.5**	49	**24.5**	52	**26**	45	**37.36**	50	**41.43**	40	**33.35**	33	**27.52**	48	**39.40**	54	**44.48**	52	**42.76**

Here n_1 represents the observed number of crossover events on the chromosome and n_2 represents the estimated average number of crossover events occurring on a chromatid.

linear model analysis to the data indicates significant effects of chromosomes (p-value is 0.000), ploidy (p-value is 0.000) and their interaction (p-value is 0.049) on the outcome crossover events. Here R^2 equals to 66.73%. According to Bonferroni method, there was no difference between two autotetraploid parental genotypes but they are different to diploids. From the real data analysis, we can see that crossover rate is statistically higher in autotetraploids than that in diploids.

2.4. Discussion

This chapter presents a likelihood-based method for estimating the crossover rate in autotetraploids using dense genetic marker data collected from all four products of meiosis, which takes properly account of essential features of tetrasomic inheritance. This method was designed specific to the data with parental genotypes of *AB/ab/ab/ab* in autotetraploids. I demonstrated the method by analysing the datasets of all meiotic products of diploid yeast and autotetraploid yeast. This method provides a way to compare crossover rate between diploids and the corresponding autotetraploids. The analysis result reveals obvious increase of crossover rate in autotetraploid yeast compared with that in the related diploid yeast, suggesting a new hypothesis that the overall crossover rate would increase after polyploidization. Subsequently, it could have effects on genetic diversity and promote adaptive evolutionary change. Although this method presented here was developed to analyse the current dataset, we can also expand the basic idea of the method to analyse various datasets with any parental genotypes of $A_1B_1/A_2B_2/A_3B_3/A_4B_4$ using a computer-based approach. This method also has some limitations

in real data analysis. For example, to compare crossover rate among different individuals, we have to choose common marker intervals among these sample. Thus as the sample size increases, it is more difficult to ensure sufficient common marker intervals distributing uniformly in the whole genome and it would require much higher sequencing depth in the samples. Even more, it is difficult to get perfect data set in practice, such as Chromosome 3 in current dataset analysis. Due to different large segments of DNA sequence missing during polyploidization, we can hardly find common marker intervals among all the individuals and have nothing to do but to discard data analysis for Chromosome 3. Such kind of missing data would probably causes bias in the comparison result and requires further consideration.

2.5. References

Comai, L. (2005) The advantages and disadvantages of being polyploidy. **Nature,** 6: 836-846.

Mancera, M., et al. (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast. **Nature,** 454: 479-485.

Otto, S.P. and Whitton, J. (2000) Polyploid incidence and evolution. **Annu. Rev. Genet.,** 34: 401-437.

Osman, K., et al. (2011) Pathways to meiotic recombination in Arabidopsis thaliana. **New Phytologist,** 190(3): 523-544.

Pecinka, A., et al. (2011) Polyploidization increases meiotic recombination frequency in Arabidopsis. **BMC Biology,** 9: 24.

Soltis, D.E. and Soltis, P.S. (1995) The dynamic nature of polyploidy genomes. **PNAS,** 92: 8089-8091.

Wang, L. and Luo, Z.W. (2012) Polyploidization increases meiotic recombination frequency in Arabidopsis: a close look at statistical modelling and data analysis. **BMC Biology,** 10: 30.

OVERALL DISCUSSION

PROGRESS IN THE THEORETICAL BASIS FOR STATISTICAL GENETIC ANALYSIS IN AUTOTETRAPLOIDS

1.1. Summary of the project

Polyploidy occurs widely in the evolutionary history of nearly all angiosperms (Jiao et al, 2011). Much evidence suggests that polyploidization of a genome could have profound long-term effects on genetic diversity and evolutionary success (Otto and Whitton 2000; Soltis and Soltis 2000; Blanc and Wolfe 2004; Chen 2007; Otto S.P. 2007; Christian 2010). Besides playing an evolutionarily important role in many species, polyploidy is present in several economically important species, in both agriculture and aquaculture, such as cultivated potato, alfalfa, Atlantic salmon and trout. However, in contrast with diploid species, progress in statistical genetic analysis in polyploid species lags far behind due to much more complicated patterns of inheritance in polyploids, which create a significant challenge. Throughout the entire project, I developed several theoretical methods to accelerate progress in statistical genetic analysis in autotetraploid species. To give some insight into these fundamental questions in autotetraploids, I divided the thesis into two parts: Part I aimed to develop theory and methods in QTL analysis in autotetraploid species, which would provide tools for breeding programmes of the world's third most important food crop, cultivated potato; Part II established methods of statistical analysis of crossover events in autotetraploids, giving some insight into the evolutionarily important role played by autotetraploidy.

In Part I of the thesis, I first proposed and developed an orthogonal contrast scales based genetic model in Chapter I-1, for decomposing quantitative genetic effects into independent monogenic, digenic, trigenic, quadrigenic and various epistatic effects independently under tetrasomic inheritance. This quantitative genetics model outperforms its rivals in several aspects: first, this model properly takes account of the key features of tetrasomic inheritance, especially the

phenomenon of double reduction; second, it is general in its use for populations with various genetic structures; third, the bi-allelic model substantially reduces the number of parameters to allow fundamental analysis of genetic effects; finally, this quantitative genetic model has taken the existence of epistasis into consideration in both populations with linkage equilibrium or linkage disequilibrium. The property of orthogonality ensures that the various genetic effects can be estimated independently for any number of loci, which is essential for a model to be consistent and comparable across multiple loci. In addition, to solve the practical problem caused by the finite sample size in QTL mapping experiments, especially for the analysis of two or more loci, I proposed a reduced model to select a subset of statistically significant genetic effects. This progress provides a solid basis for QTL analysis by linking genetic effects of genes at QTL to phenotypes of quantitative traits.

With the development of various high throughput technologies, several recent genome projects have been launched in economically and strategically important autotetraploid species, creating an urgent need for analytical tools to integrate genome sequence information collected from such projects with phenotypic data of quantitative traits. In Chapter I-2 of Part I, I proceeded to develop an interval mapping method for QTL analysis in autotetraploids. This method properly takes account of the key features of tetrasomic inheritance, including the phenomenon of double reduction and multiplex allele segregation. It is worthy of note that this work contributes the first method to successfully taking into account quadrivalent pairing during meiosis in QTL mapping for autotetraploids, which is an essential feature of tetrasomic inheritance. The quadrivalent pairing method was demonstrated to be more robust in the real data analysis than the corresponding bivalent pairing method. This advancement in the theoretical methods provides analytical tools to the recently launched genome projects in autotetraploids, which can be used

to improve breeding efficiency for economically important autotetraploid species, such as cultivated potato. To further investigate the application of this method in QTL mapping in cultivated potato, I also simulated QTL analysis in whole potato genome. Using this single QTL mapping method, we can see that QTLs on different chromosomes can be adequately detected. However, this method still has some limitations when QTLs are closely linked on the same chromosome or if large amount of epistasis existing among QTLs. Thus it would be worthy of future work to develop methods to inferring multiple QTLs simultaneously.

In Part II of the thesis, I developed statistical methods to investigate the process of recombination in autotetraploids, which is the key event in meiosis and enables generation of new combinations of chromosomes segments or alleles at different loci. In Chapter II-1, I established methods for statistical inference of crossover interference in autotetraploids under bivalent pairing and quadrivalent pairing during meiosis, properly taking account of the essential properties of tetrasomic inheritance. Theoretical analysis of crossover interference has been a historically challenging topic and very little work has been done in autotetraploids. I extended the $C_x(C_o)^m$ model developed by Zhao et al (1995) to analyze three-locus gamete/zygote data for autotetraploids in terms of the genetic distances of the two marker intervals, the coefficient of crossover interference and the coefficient of double reduction. This work has therefore filled a longstanding theoretical and methodological gap in the genetic analysis of crossover interference in autotetraploid species. By comparing crossover interference in a real dataset collected from gametes of yeast, we found a decrease in the strength of crossover interference after polyploidization on one of three chromosomes studied, suggesting a new hypothesis worthy of future exploration to explain the observed increase of recombination in autotetraploids compared with diploids (Pecinka 2011, Wang 2011).

To further investigate the crossover events during meiosis in autotetraploids, in Chapter II-2 I proposed a likelihood-based method to predict crossover rate based on whole genome sequencing data collected from tetrads of autotetraploid meiosis. Taking advantage of the next generation sequencing approach, we can implement this method to analyze phenotype data collected from densely distributed SNP markers and estimate crossover rate at the genome level in autotetraploids. By analyzing a real dataset collected from diploid and its related autotetraploid *Saccharomyces cerevisiae*, we found that the crossover rate significantly increased in autotetraploid yeast compared with diploid yeast, providing further evidence that the increase of recombination frequency in autotetraploids may be partly due to an increase in crossover events at the genome level after polyploidization.

All of the above progress in developing the theoretical basis for statistical genetic analysis will bring us closer to understand the genetic architecture of complex traits in autotetraploids. To facilitate widespread application, the methods developed have been implemented as R packages or Fortran programmes freely available upon request.

1.2. Possibilities for future work

In the future, a number of aspects of the methodology for statistical genetic analysis in autotetraploids need to be improved to increase applicability. For example, one prominent feature of autotetrasomic inheritance, the occurrence of mixed bivalent and quadrivalent chromosome pairing during meiosis, has important implications for QTL analysis and breeding schemes of autotetraploid species. The theoretical methods presented here assume complete quadrivalent chromosome pairing or complete bivalent chromosome pairing during meiosis. It

would be desirable to properly incorporate different pairing patterns of homologous chromosomes during meiosis into the statistical genetic analysis.

To solve the problem of unknown parental QTL genotypes in the QTL mapping method, the strategy presented here is to use a computer-intensive search method, which is obviously a time consuming calculation procedure. This is the key step which determines the rum time of the method and could be improved by developing a more efficient strategy. In addition, it is very likely that the genetic variance of a quantitative trait would be contributed by the segregation of multiple QTLs in practice. Therefore when a test reveals a QTL within a marker interval, the effect observed may be due to two or more loci. In this case, methods dealing with single QTL would be biased and simultaneously dealing with multiple QTLs would be required to improve the estimates of mapping positions and genetic effects.

The method of statistical inference of crossover interference assumes that the amount of crossover interference does not vary in different regions within the chromosome. However, this would not necessarily be upheld in practice. A local coefficient of crossover interference would be more desirable than a global coefficient of crossover interference. Moreover, the model could be improved to include both gene conversions and crossovers, helping us understand more about the process of crossing over during meiosis. The method presented here for predicting crossover rate in autotetraploids was designed specifically for data with a particular parental genotype. A more general method would be desirable for analysing tetrad datasets with various parental genotypes in autotetraploids.

1.3. References

Blanc, G., et al. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. **Plant Cell,** 16: 1667-1678.

Chen, Z. (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. **Annu Rev Plant Biol**, 58: 377-406.

Christian, P., et al. (2010) Evolutionary consequences of autopolyploidy. **New Phytologist,** 186 (1): 5-17.

Jiao, Y., et al. (2011) Ancestral polyploidy in seed plants and angiosperms. **Nature**, 473: 97-100.

Otto, S.P. and Whitton, J. (2000) Polyploid incidence and evolution. **Annu. Rev. Genet.,** 34: 401-37.

Otto, S.P. (2007) The evolutionary consequences of polyploidy. **Cell,** 131: 452-462.

Pecinka, A., et al. (2011) Polyploidization increases meiotic recombination frequency in Arabidopsis. **BMC Biology,** 9: 24.

Soltis, P.S. and Soltis, D.E. (2000) The role of genetic and genomic attributes in the success of polyploids. **PNAS,** 97(13): 7051-7067.

Wang, L. and Luo, Z.W. (2012) Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biology, 10:30.

Zhao, H., et al. (1995) Statistical analysis of crossover interference using the chi-square model. **Genetics,** 139(2): 1045-56.