RECOMBINATION STUDIES AND DEVELOPMENT OF QTL-TARGETED TILED INTROGRESSION LIBRARIES BY MARKER-ASSISTED BACKCROSSING ACROSS FOUR MAIZE CHROMOSOMAL SEGMENTS

A Dissertation

by

RUPA SRIDEVI KANCHI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Seth C. Murray
Co-Chair of Committee,	Fred P. Dahm
Committee Members,	David M. Stelly
	Alan R. Dabney
Head of Department,	David D. Baltensperger

May 2015

Major Subject: Plant Breeding

ABSTRACT

Biological characterization of causal genomic variants present in large numbers within a quantitative trait locus (QTL) interval is a challenging problem. QTL-targeted tiling paths of near-isogenic lines (NILs) are useful in fine mapping and gene cloning, and in estimating the effects of closely linked loci. However, their development is costly and involves recurrent cycles of selection-recombination-repetition in a breeding program.

In the first project, a distance-based strategy for fractionating loci into tiled introgression libraries using marker-assisted backcross selection is proposed. Computer simulations were used to investigate the efficiency of the proposed selection strategy in separately producing NIL sets with, on average, 3, 1.5 and 1 cM introgressions across target QTL regions of 15 cM. The proposed distance-based strategy identified NIL sets that led to better control and elimination of linkage drag using fewer backcross generations and smaller progeny sizes. Increasing BC_1 progeny size (>150) had little and increasing the number of backcross individuals per selected individual (N_{BSI}) had significant positive effect on the genomic composition of the NILs and length of backcrossing. More N_{BSI} and backcrossing generations were required to produce NILs with smaller (1.5 and 1 cM) when compared to the larger (3 cM) average introgression lengths.

In the second project, recombination patterns across four previously identified maize photoperiod QTLs (ZmPR1-4) were evaluated based on multiple genetic backgrounds. Marker data from the first backcross generation of crosses between seven tropical and two temperate inbred lines were used. It was found that recombination rates varied between the four photoperiod QTLs. The ZmPR2 region was most recombinogenic while the ZmPR4 region was least recombinogenic among the four QTLs evaluated. The distributions of recombinations along the four ZmPR regions were consistent, and seemed to be uniform within the ZmPR2 and ZmPR4 regions. Within the ZmPR3 region, a 4.221 Mbp region containing at least two orders of magnitude higher recombination than the maize genome average was found. This study suggests that increased marker density will need to be used to gain valid estimates of the genetic diversity for recombination rates.

DEDICATION

To my parents who gave a direction to my life;

I am nothing without your unconditional love, support and encouragement.

To my two beautiful daughters who think I am perfect;

You have changed my life, given it a new meaning, made it more challenging; and yet, my most precious, simple, happy and confident moments have been with you.

To the memories of my uncle Kanchi Seshagiri who did not live to see this day;

You would have been extremely happy today.

ACKNOWLEDGEMENTS

I would like to express my special thanks and gratitude to Dr. Seth Murray, my committee chair, for his invaluable support and encouragement. His knowledge in plant breeding along with wonderful insights in statistics and computing/software greatly influenced my learning process. His wisdom taught me many more life lessons than he can imagine. My heartfelt thanks go to Dr. Murray for his ever-smiling and helpful nature and for standing by me in my efforts to complete my doctoral program.

I wish to thank Dr. Randall Wisser and Dr. Nick Lauter whose vision significantly improved this manuscript. Despite being extremely busy, they took time out for me. I am deeply indebted to them for their time.

My committee co-chair Dr. Fred Dahm, and members Dr. David Stelly and Dr. Alan Dabney supported me throughout my graduate research. Dr. Dahm gave me time to discuss with him my analyses. My appreciation for Dr. Stelly and Dr. Dabney does not cease for their encouraging and kind words. Dr. Stelly has always been willing to help me with understanding the aspects underlying the process of meiotic recombination. I thank them all for their time and for serving on my committee.

I was financially supported by the distance education program in the Department of Statistics at TAMU, and the USDA-NIFA project “Analysis of clinal variation in maize: implementation of an experimental framework for studying crop adaptation” (award no: 2011-67003-30342). I sincerely thank them for allowing me to pursue higher studies.

My special thanks go to Dr. Michael Longnecker and Dr. Wayne Smith, truly the “problem solvers” for students in their respective departments, for their guidance whenever I came across hurdles through these years.

Dr. Michael Longnecker, Dr. Thomas Schlumprecht, Dr. Fred Dahm, Dr. Mohsen Pourahmadi, Dr. Seth Murray and Dr. Bill Rooney made my graduate life a joyful and memorable experience through their academic courses. I thank them for being such awesome and inspiring teachers.

Scott Vajdak, Henrik Schmiediche and Jeff Savage shared their knowledge over my computing issues; my fellow graduate students Robyn Ball, Adam Mahan, Ivan Barrero, Bharath Reddy, John Gill, Schuyler Smith, Jacob Pekar, Justine Christman and Yuanyuan Chen helped me on various occasions. I thank them for being so helpful.

My husband Sagar has been an amazing dad to our two beautiful daughters who were born during my graduate studies. In balancing our conflicting demands for time, we touched many highs and lows. I hope we evolve as better individuals through this roller-coaster ride.

My family helped me immensely and has been my strongest support system. Mom, Dad, Sashi, Sumanth and Abhay have always had time for me, in particular my mom and dad. Their belief in me and encouragement through rough times has made it possible for me to realign myself.

I would also like to sincerely extend my thankfulness to many whom I might have missed here. Thanks to all of you.

NOMENCLATURE

AFRI	Agriculture and Food Research Initiative
ATLAS	Adaptation Through Latitudinal Artificial Selection
CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo [International Maize and Wheat Improvement Center]
DNA	Deoxyribo Nucleic Acid
DP	Donor Parent
DSB	Double Strand Break
ex-PVP	expired Plant Variety Protection
FAOSTAT	Food and Agriculture Organization Statistical Databases
IITA	International Institute of Tropical Agriculture
MAS	Marker Assisted Selection
NAM	Nested Association Mapping
NIFA	National Institute of Food and Agriculture
NIL	Near Isogenic Line
NILAS	Near Isogenic Lines in Allelic Series
QTL	Quantitative Trait Locus
QTLs	Quantitative Trait Loci
RP	Recurrent Parent
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism

SEM	Standard Error of Mean
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
USDA	United States Department of Agriculture
ZmPR	*Zea mays* Photoperiodic Response

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1 INTRODUCTION

1.1 Genetic enhancement of maize

Maize (*Zea mays* ssp. *mays* L.) is one of the three major grass and grain crops in the world along with wheat and rice. Its' total world production in the year 2013 was ~1016 million tons of which about 35% was accounted for by the U.S. (FAOSTAT, verified on 28 Sep 2014). Also known as corn, it is mainly used for animal feed, food and in industry. Maize, supposed to be domesticated from its' wild progenitor teosinte (*Zea mays* ssp. *parviglumis* L.) about 6000 to 10,000 years ago in Mexico (Doebley 2004) has a genome size of about 2.3 gigabases with over 32,000 predicted genes spread across ten pairs of chromosomes (Schnable et al. 2009).

Diversity, both genetic and phenotypic, is extremely high in maize. The abundance of favorable alleles in global maize is scattered across an array of climatic zones and/or populations or landraces (Prasanna 2012). The current germplasm base of temperate maize, however, represents only a small fraction of the immense global diversity. Maize inbred lines used in current U.S. commercial breeding programs trace their ancestry to a few prominent genetic clusters represented by the inbred lines B73, Mo17, PH207, A632, Oh43, and B37 (Nelson et al. 2008). Use of exotic germplasm in U.S. maize programs has been highlighted to broaden the diversity of commercial maize varieties for improved yield, adaptation and disease resistance (Pollak 2003, Goodman 2004, Goodman 2005). Superior tropical maize accessions, in particular, are a useful source of germplasm for commercial U.S. maize breeding programs (Holland and

Goodman 1995). The tropical and subtropical inbred lines possess greater gene diversity than the temperate germplasm (Liu et al. 2003). But a major barrier in adaptation of this tropical maize to temperate environments is photoperiod sensitive delayed flowering, among many agronomic issues such as lodging, increased plant height and susceptibility to common smut (Mungoma and Pollak 1991, Coles et al. 2010).

1.2 Photoperiod sensitivity

Photoperiod sensitivity can be described as the differential response of plants to different photoperiods (day-lengths), i.e., genotype-by-photoperiod interaction. Photoperiod and temperature are the most common seasonal cues that influence the transition from the vegetative to reproductive stage in plants (Huijser and Schmid 2011). Flowering at the right time reflects the adaptation of a plant to its' environment for successful seed and fruit development (Putterill et al. 2004), and in tropical maize it eventually determines crop yields. Classified as short-day or long-day responsive, most temperate plants are day-neutral for which the genotype-by-photoperiod interaction is insignificant. Short-day plants, like tropical maize, require long nights to induce flowering and often display significant sensitivity to longer day-lengths (> 10 to 13.5 hours, Kiniry et al. 1983) whereas long-day plants require short nights to induce flowering. Flowering time in maize is a complex trait governed by numerous small-effect additive QTLs (Chardon et al. 2004, Buckler et al. 2009).

Photoperiod response in maize is quantitatively inherited and governed by a few genes with large, additive effects (Russell and Stuber 1983, Wang et al. 2008, Coles et al. 2010, Hung et al. 2012). The genetic architecture for silking / anthesis under photoperiod response has been described by Coles et al. (2010) in populations between elite US and elite tropical breeding lines across four interrelated maize linkage mapping populations. Four major-effect photoperiodic response quantitative trait loci (QTLs) were identified on chromosomes 1, 8, 9 and 10 and named ZmPR1-4 (for *Zea mays* Photoperiodic Response 1 to 4) respectively. The positions of these ZmPR genomic loci have been further redefined (Hung et al. 2012) based on the high-resolution nested association mapping (NAM, McMullen et al. 2009) map of maize.

1.3 Meiotic recombination

To dissect and integrate favorable genes, plant breeders exploit meiotic recombination which contributes to genetic variability by forming new combinations of chromosomes in resulting gametes and new combinations of chromosome segments within a single chromosome (Fairbanks and Anderson 1999, Chapter 11). Detection of variation at the DNA level through changing marker techniques (RFLP, RAPD, SSR, SNP etc.) have significantly contributed to research in genome-wide recombination. Meiosis produces haploid gametes through a process consisting of chromosomal replication followed by two cell divisions – Meiosis I and II. Meiosis I is the core of the entire process in which the homologous chromosomes pair, recombine and then segregate from each other (Page and Hawley 2003). The precise pairing of duplicated homologous chromosomes during prophase I permit crossing over between the nonsister chromatids. A crossover (CO) results when two nonsister chromatids exchange DNA segments with each other leading to intrachromosomal reshuffling of parental alleles in the gametes and increasing genetic diversity of meiotic products (Jones and Franklin

2006). Multiple exchanges are considered in general to involve any two nonsister chromatids, apart from polyploid plant species, suggesting no chromatid interference (Karlin and Liberman 1994, Zhao et al. 1995). The orientation of one homologous chromosome pair at the spindle equator during metaphase I has no influence on the orientation of any other chromosome pair. That is, the nonhomologous chromosomes assort independently into two daughter cells during meiosis I (Fairbanks and Anderson 1999, Chapter 11). The two homologues of each chromosome pair segregate during anaphase I. The sister chromatids of each chromosome, still attached at the centromere, in these two daughter cells align along the spindle equator and separate in Meiosis II resulting in gametes with haploid chromosome number.

The number and position of COs along the bivalent (four chromatid bundle at metaphase I), the nonsister chromatids involved in COs, independent assortment of nonhomologous chromosomes and sampling of a haploid gamete to produce a diploid cell through fertilization are all believed to be stochastic characters during the process of meiotic cell division.

Crossing over is a controlled process associated with an obligate crossover and a crossover interference process in most organisms tested (Jones and Franklin 2006, Martini et al. 2006). An obligate crossover refers to at least one crossover per chromosome pair while crossover interference refers additional crossovers subject to the observation that a crossover in one chromosomal region reduces the probability of additional crossover(s) in adjacent regions resulting in crossovers being more widely and evenly spaced than would be expected if they occurred independently (Martini et al.

2006, Mézard et al. 2007). It is not fully understood, however, if obligation and interference are governed by two distinct and independent pathways or if they are determined by single process (Jones and Franklin 2006). Because crossover interference is well accepted and chromatid interference has not been demonstrated, interference generally refers to crossover interference only (unless otherwise specified).

2 IN SILICO OPTIMIZATION FOR DEVELOPMENT OF QTL-TARGETED TILED INTROGRESSION LIBRARIES BY MARKER ASSISTED SELECTION

2.1 Introduction

2.1.1 What are NIL libraries?

Near-isogenic lines (NILs) or introgression libraries are genetically similar sets of inbred lines containing one or more small donor chromosomal segment(s) from a donor parent (DP) spread across the genomic region of interest in an elite recurrent parent (RP) background (Eshed and Zamir 1994).

Quantitative trait studies aim to go beyond the mapping of loci (QTL) and to identify and biologically contextualize the causal variant(s). Linkage mapping is a powerful framework for QTL detection but identifying causal variants is often not possible, as the approach typically suffers from poor mapping resolution (~10 cM resolution is typical; Mackay 2001, King et al. 2012) resulting in a challenging biological characterization of hundreds to thousands of genetically inseparable genomic variants. The statistical-genetic framework to detect and estimate the effect of each QTL in the presence of numerous other QTL requires large population samples (Beavis 1994). Therefore, many lower-throughput but informative techniques that can offer more biological insight about the function of a QTL (e.g. physiological, cellular, molecular bases) are not compatible with linkage mapping population studies alone. Sets of NILs or introgression libraries comprised of lines in a common, RP genetic background and differing only by the presence of contrasting recurrent versus donor parent haplotypes at individual loci complement weaknesses of linkage mapping populations (Eshed and Zamir 1994, 1995; Frary et al. 2000).

2.1.2 Why create NIL libraries?

With a NIL library of introgressions tiled across a QTL region it is possible to 1) delineate causal genes(s), 2) address a range of hypotheses about the haplotype structure underlying phenotypic effects such as linkage drag, pleiotropy and epistasis among linked loci, and 3) have a manageable set of lines for more detailed biological investigation. Furthermore, multiple parents can be used as recurrent parents or donor parents to broaden the genetic scope of inference. Using multiple recurrent parents can shed light on the context-dependency and distribution of allele effects at a QTL for which there is currently little information in any species. A practical advantage of such near-isogenic line allelic series (NILAS) libraries is that they can also serve as a stepping-stone toward re-configuring haplotypes (Peleman and van der Voort 2003). Through marker assisted selection, we are creating tiled NILAS libraries for four, previously identified maize photoperiod QTL (*ZmPR1-4* in Coles et al. 2010, redefined by Hung et al. 2012) in two different recurrent parent backgrounds. This effort can identify loci that are linked to the photoperiod responsive genes which is important because if selection of material is done in the Midwestern U.S. “Corn Belt”, the center for commercial maize breeding, genes in linkage with the day-neutral allele at photoperiod genes could be highly selected against and the variation associated with those genes would not be incorporated in the selected material. This inadvertent selection against linked loci could negatively affect any trait that the loci are linked to. A tiling path NIL library can assist in dissecting these effects to better understand what may be missed through exotic introgression.

2.1.3 How to create NIL libraries?

A recurrent selection or select-recombine-repeat breeding procedure is followed to cyclically improve individuals, families or breeding populations (Bernardo 2010). Unlike phenotypic recurrent selection that involves selecting the best individuals based on their phenotypic traits, marker assisted recurrent selection involves identifying the best individuals based on molecular markers linked with the traits of interest.

Marker-assisted backcrossing (MAB) is often used to introgress a defined QTL or gene of interest into a separate genetic background that is more desirable or elite (Hospital 2005). For a single recurrent (RP) and donor parent (DP) pair, the backcrossing process starts by crossing the F_1 individual to the RP. Progeny in BC_1 generation, when taken together, may appear to have chromosomes with staggered segments from the DP in the background of the RP (Figure 2.1), a practical outcome as recombination events rarely occur at the same positions, and desired to isolate the fragment containing the causal gene(s). The most desirable BC_1 individuals are selected and crossed back to the RP. Repeated backcrossing of the selected individuals to the RP increases the proportion of RP genome by half, on average, in each BC generation (Bernardo 2010).

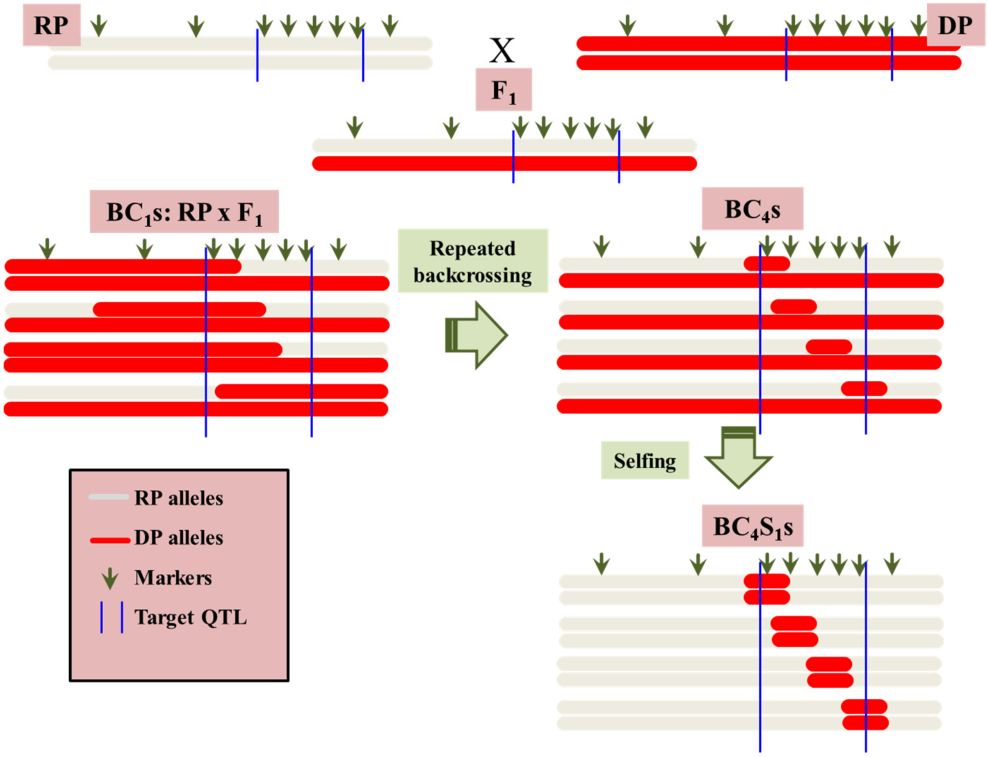


Figure 2.1 Marker assisted backcrossing scheme for four BC and two selfing generations based on a chromosome pair to select progeny individuals forming a tiling path. The region between the blue lines is the targeted QTL region – the statistically most likely position for causal gene(s). The small vertical pointers at the top of each chromosomal set refer to positions of the genotyped markers. The RP and the DP are crossed to obtain F_1 individuals. In each cycle of the repeated backcrossing, individuals are selected based on their marker genotypes only. At the end of backcrossing, individuals are selfed to achieve homozygosity of the chromosomes.

The selection of individuals in each recurrent generation is based on determining the identity by descent (IBD) of alleles in the individuals with the RP and DP. If an allele in an individual can be traced back to an ancestor in the pedigree, it is said to be identical by descent (IBD) with the ancestral allele. Two alleles at a locus may be identical by state (IBS) but not necessarily IBD (Lynch and Walsh 1998). IBS but not IBD can occur when an independent mutation takes place, which is usually extremely rare. In general, single nucleotide polymorphism (SNP) markers are chosen such that each parental allele can be separated without confusing IBD and IBS. In this study, alleles are compared that are one generation apart in the pedigree. Hence, an allele that is IBS is an IBD allele when the parents are polymorphic at that locus. In such cases, IBD can be established with probability 1.

By the end of the backcrossing program, it is desired that a set of individuals be created that are identical by descent with the RP at all loci except in the target region(s) of interest where chromosomal segments identical by descent with the DP are desired in a tiling manner (Figure 2.1). This effort led to an interest in a selection algorithm for constructing QTL-targeted NIL libraries.

2.1.3.1 Which individuals to select?

Selection strategies developed for marker-assisted introgression aim to select individuals with donor parent alleles across a target segment (foreground selection) and recurrent parent alleles across the rest of the genome (background selection). Optimality of selection strategies have been investigated for the introgression of a specific single locus or multiple unlinked loci (Hospital et al. 1992, Frisch et al. 1999, Hospital et al. 2000, Frisch and Melchinger 2001), and for the introgression of tens of loci constituting the entire genome of a donor parent (Sušiƈ 2005, Falke et al. 2009, Herzog et al. 2014). In each case, different criteria have been defined for developing introgression lines in terms of the selection algorithm and breeding scheme.

Previous selection algorithms in constructing introgression libraries have employed selection pressure on the target and non-target chromosomal segments or chromosomes such that in the first step, all the individuals carrying donor alleles in the target segment would be selected and in the second step, individuals with the highest proportion of RP alleles outside the target segment would be selected. This approach optimizes for reduction in linkage drag simultaneously on both sides of the target segment and does not select individuals, if present, with recombinations closest to the target segment under selection. For instance, if the RP proportions for the individuals A to C were the same in the rest of the genome not shown in Figure 2.2, the above approach would select the individual A which has maximum proportion of RP alleles outside the target segment under selection (chromosomal segment between the blue lines in the target QTL region, Figure 2.2) on the target chromosome. A desired selection choice might be individual C with a recombination closer to the target boundaries on at least one side.

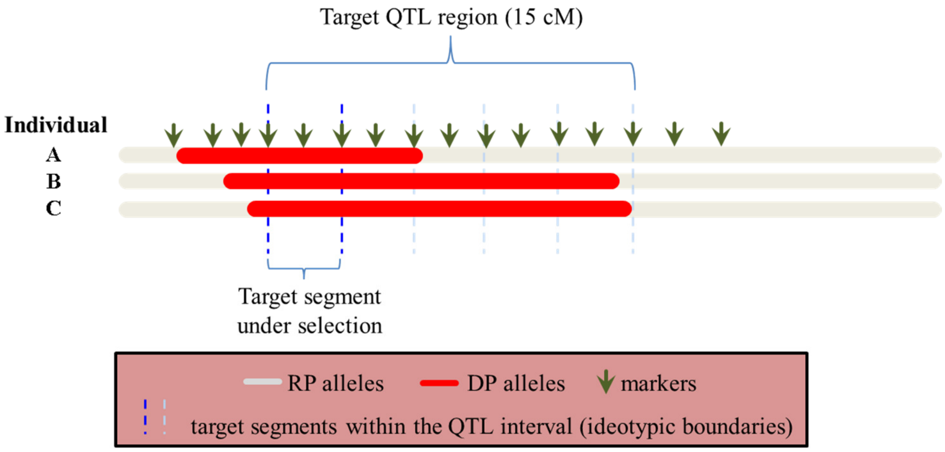

Target QTL region (15 cM)
Individual
A
B
C
Target segment under selection
RP alleles
DP alleles
markers
target segments within the QTL interval (ideotypic boundaries)

Figure 2.2 Visual showing that an individual with the closest recombination on at least one side of the target segment under selection (desired) might be missed (when available) if the selection algorithm maximizes for the RP proportion outside the target segment.

Falke et al. (2009) used a selection index on a 0-2 scale which overcomes this problem of not being able to select individuals with recombinations closer to the target segment. This selection index was used in an additional selection step after preselecting individuals with donor alleles in the target segment. The index was 2 for recombination between the target and both flanking markers, 1 for recombination on at least one side, and 0 for no recombination between the target segment and flanking markers (the *three stage selection* in Falke et al. 2009). Individual(s) with the largest index would then be chosen for the next step of selection for maximum RP proportion. The selection indices for the three individuals A, B and C would be 0, 0 and 1, and individual C would be selected. When there is no individual C which could happen very often when the flanking markers are present at a genetic distance of 0.5 or 1 cM (as in the case of present study), the individual A, by virtue of larger RP proportion in the non-target segments would be selected (since A and B both have the same index 0), which is a downside of the selection index since a preferred individual would be individual B when compared to individual A.

In developing tiled introgressions across a QTL region, the priority is to select a set of individuals with recombinations accumulated (usually over generations) between tightly linked markers within and flanking this targeted region. Reducing the linkage drag is a high priority in order to create individuals with short and evenly spaced donor introgressions across the QTL region. In this dissertation, a distance-based method to develop such resource is proposed.

A major challenge of many plant breeding and genetics research programs is allocating resources appropriately, which are not infinite. Identifying the areas that require low inputs and diverting the related resources to other areas that require higher inputs mark a major shift towards optimized breeding strategies. In this study, using simulations, impact of the number of individuals genotyped, the number of individuals selected for advanced backcrossing and the numbers of backcross progeny produced per selected individual in each backcross generation, on the quality of the set of NILs across a pre-defined locus are evaluated. To my knowledge, there have been no studies concerning the fractionation of a QTL region into tiled introgressions.

2.2 Theory and approaches

A whole genome (10 chromosomes of maize) was simulated to explore scenarios for optimizing the production of tiled NILs at separate 15 cM target QTL regions on chromosome 1 and 10, the largest and smallest chromosomes respectively, based on known ZmPR photoperiod sensitivity genes, and including background control of other chromosomes (non-ZmPR; Table B1). For each target QTL region, 30 marker loci were defined each spaced 0.5 cM apart. Four additional marker loci were defined at evenly spaced positions in pre-defined bins along each chromosome, bound by equally spaced segments along the non-target chromosome space for background selection. Backcross breeding was initiated with a cross between a fully inbred RP and DP, with contrasting SNP alleles at all simulated marker loci.

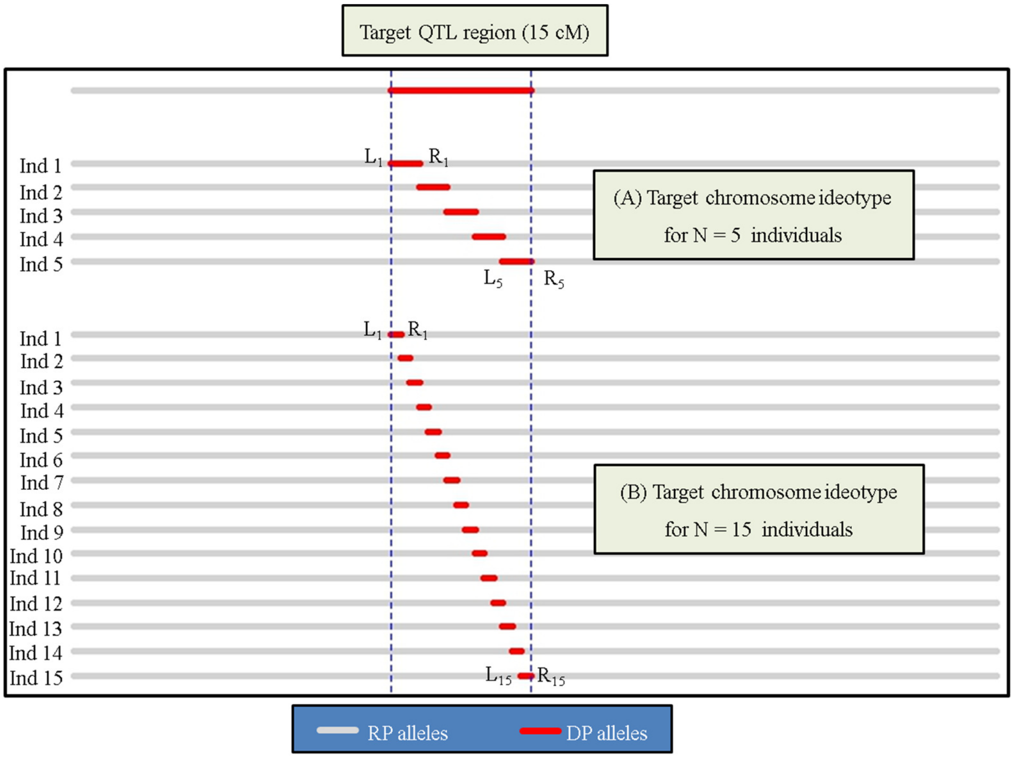
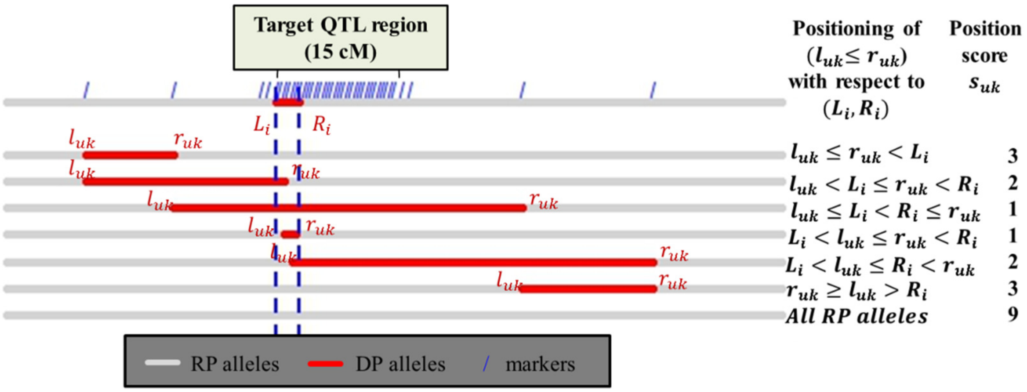


Figure 2.3 Ideotypic set for the target chromosome with donor parent alleles tiled across the target QTL region in the recurrent parent background.

2.2.1 Selection algorithm

A distance-based algorithm was developed to select a subset of progeny that most closely resemble (are the least distant to) a user-defined ideotype. An ideotypic set was defined as a set of N NILs constituting a distance-defined tiling structure of donor-derived segments traversing a target region. For instance, the target regions of length T cM were divided uniformly into N non-overlapping segments $\{(L_i, R_i), i = 1 \dots N\}$ where L_i and R_i are the left and right boundaries respectively of the N segments (Figure 2.3).

Alleles associated with the recurrent and donor parents were denoted by 0 and 1 respectively so that the genotype in diploid state (RP/RP or RP/DP) was the number of donor alleles at each marker locus. Consequently, marker data on the target chromosome in backcross generations were represented as a KxM matrix of 0s and 1s where K and M denote the numbers of individuals and markers respectively, and 0 and 1 represent RP/RP homozygous and RP/DP heterozygous genotypes respectively determined by IBD. Based on the cM positions of these M marker loci, denoted by $M_j, j = 1 \dots M$, the contiguous stretches of 1s for each individual k were indexed as $\{(l_{uk}, r_{uk}), k = 1 \dots K\}$ for the u^{th} contiguous stretch. In each BC generation, independently for each ideotypic target (L_i, R_i), a single 'most desired' individual from the backcross progeny was recovered as described in steps 1-3 below. From step 2 onwards, selection was performed on individuals pre-selected in the previous step if at least two individuals were available.



Positioning of $(l_{uk} \leq r_{uk})$ with respect to (L_i, R_i)	Position score s_{uk}
$l_{uk} \leq r_{uk} < L_i$	3
$l_{uk} < L_i \leq r_{uk} < R_i$	2
$l_{uk} \leq L_i < R_i \leq r_{uk}$	1
$L_i < l_{uk} \leq r_{uk} < R_i$	1
$L_i < l_{uk} \leq R_i < r_{uk}$	2
$r_{uk} \geq l_{uk} > R_i$	3
All RP alleles	9

Figure 2.4 Position score for the contiguous stretches of donor parent alleles with respect to an ideotypic target segment (L_i, R_i).

For an ideotypic target (L_i, R_i):

Step 1: Use the index data $\{(l_{uk}, r_{uk}), k = 1 \ldots K\}$ to assign a position score s_{uk} and a distance measure d_{uk} for each u^{th}contiguous stretch (l_{uk}, r_{uk}). Set the position score s_{uk}as 1, 2, 3, or 9 such that (see Figure 2.4):

$s_{uk} =$ 1 if the contiguous stretch (l_{uk}, r_{uk}) is located completely within (L_i, R_i) or completely spans it (overlaps with both L_i and R_i)

2 if the contiguous stretch (l_{uk}, r_{uk}) overlaps with only L_i or R_i

3 if the contiguous stretch (l_{uk}, r_{uk}) is located outside of (L_i, R_i); or

9 if there is no contiguous stretch along the target chromosome (the gamete descending from the F_1 parent comprises of all RP alleles).

Assign a distance score $d_{uk} = \min\{abs(l_{uk} - L_i), abs(r_{uk} - R_i)\}$ and select among individuals for which $s_{uk} = 1$ and $s_{uk} = 2$ with minimum d_{uk}. If $s_{uk} = 3$ or 9 for all individuals, selection was not performed for that target segment.

Step 2: Select individuals with the highest RP proportion in the chromosome containing the target region.

Step 3: Select individuals with the highest RP proportion in any additional target regions.

Step 4: Select individuals with the highest RP proportion in the remaining fraction of the genome. If there were at least two individuals at this stage, select one of them randomly.

Repeat steps 1-4 for each $(\mathrm{L_i}, \mathrm{R_i})$ to select the most desirable individual from the progeny for each ideotypic target segment.

2.2.2 Simulation parameters

2.2.2.1 Backcrossing strategies

Variations were considered in the number of backcross progeny genotyped, the number selectively advanced for the next cycle of backcrossing, and the number of backcrosses per selected individual. It was hypothesized in the empirical experiment that the number of individuals in the BC_1 (first backcross) generation was critical in finding a good candidate-NIL set. For simulations, five different numbers of BC_1 individuals (N_{BC1}= 50, 150, 250, 350 and 500) were considered for genotyping. A maximum of 10 individuals per QTL per donor and recurrent parent pair could be empirically afforded by the breeders in this study but there was interest to understand the properties of the candidate NILs set if selecting 5 or 15 individuals. Consequently, the selected set of ILs were desired to carry contiguous donor segments, on average, of 3 cM, 1.5 cM and 1 cM respectively in the target QTL region (15 cM) and were represented by N_{SEL} = 5, 10 and 15 respectively for the numbers of individuals expected to be selected at each backcross selection step in each generation in each locus. The numbers of backcrosses per selected individual were N_{BSI}= 5, 10, 20, 50 and 100. All possible combinations of N_{BC1}, and N_{BSI} (5 x 5 = 25) backcrossing strategies (Figure 2.5) were simulated 1000 times each for each N_{SEL}.

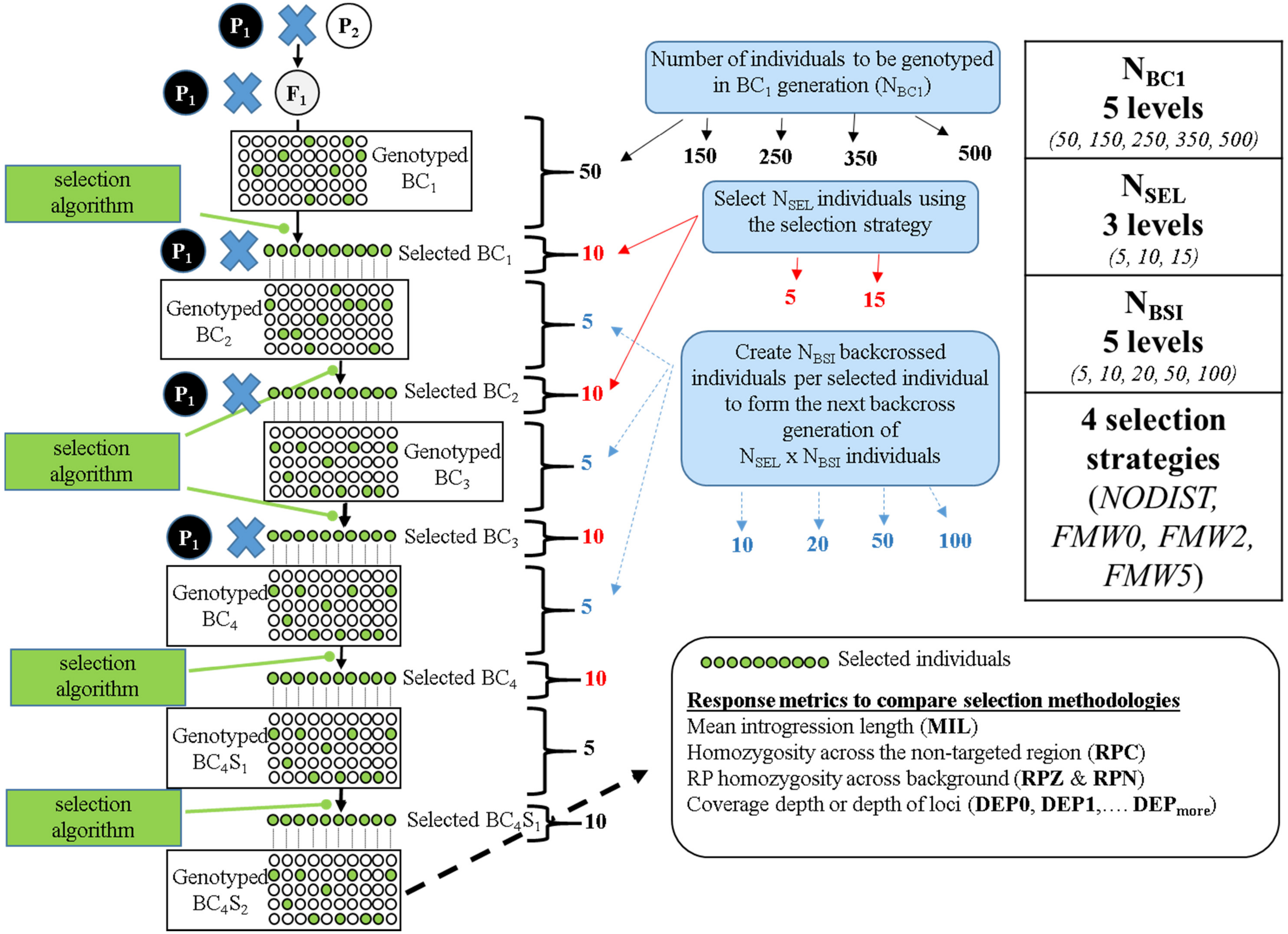

P1
P2
F1
Genotyped BC1
selection algorithm
Selected BC1
Genotyped BC2
Selected BC2
Genotyped BC3
Selected BC3
Genotyped BC4
Selected BC4
Genotyped BC4S1
Selected BC4S1
Genotyped BC4S2
50
10
5
10
5
10
5
10
5
10
Number of individuals to be genotyped in BC1 generation (NBC1)
150
250
350
500
Select NSEL individuals using the selection strategy
5
15
Create NBSI backcrossed individuals per selected individual to form the next backcross generation of NSEL x NBSI individuals
10
20
50
100
NBC1
5 levels
(50, 150, 250, 350, 500)
NSEL
3 levels
(5, 10, 15)
NBSI
5 levels
(5, 10, 20, 50, 100)
4 selection strategies
(NODIST, FMW0, FMW2, FMW5)
Selected individuals
Response metrics to compare selection methodologies
Mean introgression length (MIL)
Homozygosity across the non-targeted region (RPC)
RP homozygosity across background (RPZ & RPN)
Coverage depth or depth of loci (DEP0, DEP1,…. DEPmore)

Figure 2.5 The experimental design for simulation of a breeding scheme with four BC and two selfing generations (BC_4S_2). Simulations were performed with varying numbers of individuals in the BC_1 generation (N_{BC1}), numbers of individuals expected to be selected in each generation (N_{SEL}) as proxies to the desired lengths of donor introgression in each selected individual (3, 1.5 and 1 cM respectively), numbers of backcrosses per selected individual (N_{BSI}) and selection strategies.

2.2.2.2 Selection strategies

Four selection strategies were investigated in which the numbers of foreground (30 markers per target QTL region) and background markers (four markers per chromosome) were constant. The first three strategies selected individuals using the distance method with variations in the number of markers in a flanking marker window (FMW) defined adjacent to the target QTL region on both sides. The fourth method consisted of selection without the distance method.

1) **FMW0 selection:** Flanking marker window was not defined.
2) **FMW2 selection:** A 2 cM window was defined with two markers spaced 1 cM apart on each side of the target QTL region. The four background markers were therefore defined evenly outside of the QTL target + flanking marker region.
3) **FMW5 selection:** A 5 cM window was defined with five markers spaced 1 cM apart on each side of the target QTL region. The four background markers were therefore defined evenly outside of the QTL target + flanking marker region.
4) **NODIST selection:** Instead of the distance measure, proportions of the donor and recurrent alleles were used for selection. For this selection strategy, a flanking marker window was defined similar to the FMW2 strategy. Foreground selection was first carried out such that only individuals carrying DP alleles at all markers in the ideotypic target segment were selected. Background selection was then performed for highest RP proportion on the target chromosome, ZmPR regions and the rest of the genome. This strategy was similar to the three stage selection strategy employed in Falke et al. (2009).

2.2.2.3 Recombination model

Meiotic crossover positions were modeled using two independent pathways, interference and non-interference, based on the distribution of inter-crossover distances of late recombination nodules in maize (Falque et al. 2009), which are thought to mark all crossover positions along the tetrad (Anderson et al. 2003; Falque et al. 2009). Using an interference parameter (υ) and proportion of non-interference crossovers (p) for each chromosome (Table B1), interference crossovers were formed in a sequence according to a stationary renewal process with inter-crossover distances following a gamma distribution with shape and rate parameters υ and 2υ respectively, while non-interference crossovers were formed by an exponential distribution with rate parameter 2 (see Appendix C for details). When both the interference and non-interference statistical models for crossovers generated crossover positions that were greater than the chromosome length resulting in an empty vector of crossover positions, an obligatory crossover was produced which is required for a proper disjunction of homologous chromosomes (Jones and Franklin 2006).

2.2.3 Evaluation criteria

For each generation, data summaries on the selected set of NILs were used to examine the outcomes of different breeding and selection strategies. The following data were recorded about the 15 cM QTL target: mean introgression length in cM (MIL), percentage of target region having a DP allele covered with zero to four and greater than four individuals (%DEP_0, %DEP_1, %DEP_2, %DEP_3, %DEP_4, %$DEP_{>4}$). The following data were recorded about the non-target loci in relation to target (15 cM) region:

percentage (%) of RP homozygosity across the non-target loci of the target chromosome (%RPC), across the loci at the other three (background ZmPR) target regions (%RPZ), and across all loci outside of the target chromosome and the three background ZmPR QTL regions (%RPN). The values obtained from the selected set of individuals in each BC generation based on 1000 simulations were considered as realizations of random variables that describe MIL, %RPC, %RPZ, %RPN, and various DEP measurements.

The MIL was an indicator of how well the linkage drag was addressed and eliminated. In the recurrent process of selecting sets of individuals with 3 cM, 1.5 cM and 1 cM donor introgressions (using an expected N_{SEL}=5, 10 and 15 individuals respectively as proxies) across the 15 cM target region in each generation, a breeding strategy was declared successful if mean donor introgression lengths of up to 3.5 cM, 2 cM and 1.5 cM respectively were achieved because it was only possible to make selections based on the flanking markers and the nearest flanking marker loci from the ideotypic boundaries in this study were defined at 0.5 cM within the target region and at 1 cM outside of the target region (for selection strategies FMW2, FMW5 and NODIST).

The three variables %RPC, %RPZ and %RPN represented three levels used for background selection to recover the RP genome as rapidly as possible. All the other variables were used to quantify and compare the success of the different breeding and selection strategies with respect to the genomic composition of the selected set of NILs.

Simulations were performed using Intel® Core™ i7-2600 CPU at 3.4GHz with 16 GB RAM. The package 'doParallel' of R programming language (R development core team 2012) was used to implement parallel processing of simulation replications.

2.3 Results

2.3.1 Pooling vs maintaining individuals separately

An initial simulation study based on chromosome 10 using 150 BC_1 individuals established how selection was practiced from the BC_2 generation onwards for foreground selection. To create a set of NILs with a mean donor introgression of 1.5 cM in each individual, ten individuals were selected in each BC generation and for each selected individual, 20 backcrosses with the RP were produced. Starting from the BC_2 progeny, two selection routes were examined using the distance based selection strategy FMW2 (section 2.2.2.2) to produce NIL sets with evenly spaced and minimally overlapping donor introgressions: 1) to maintain each selected line separately to select the best plant with the least linkage drag in the next generation from within the backcrosses produced from that line; and 2) to pool the backcrossed plants produced from the selected lines to construct a set of next-generation introgression lines. Pooling produced NIL sets with relatively lower average introgression lengths from BC_3 onwards (Figure 2.6). In fact, by the BC_4 generation, the median and quartiles of the distribution of average introgression lengths for pooling were nearly a full generation ahead of the separation strategy (shown in bold for BC_4 in Table 2.1).

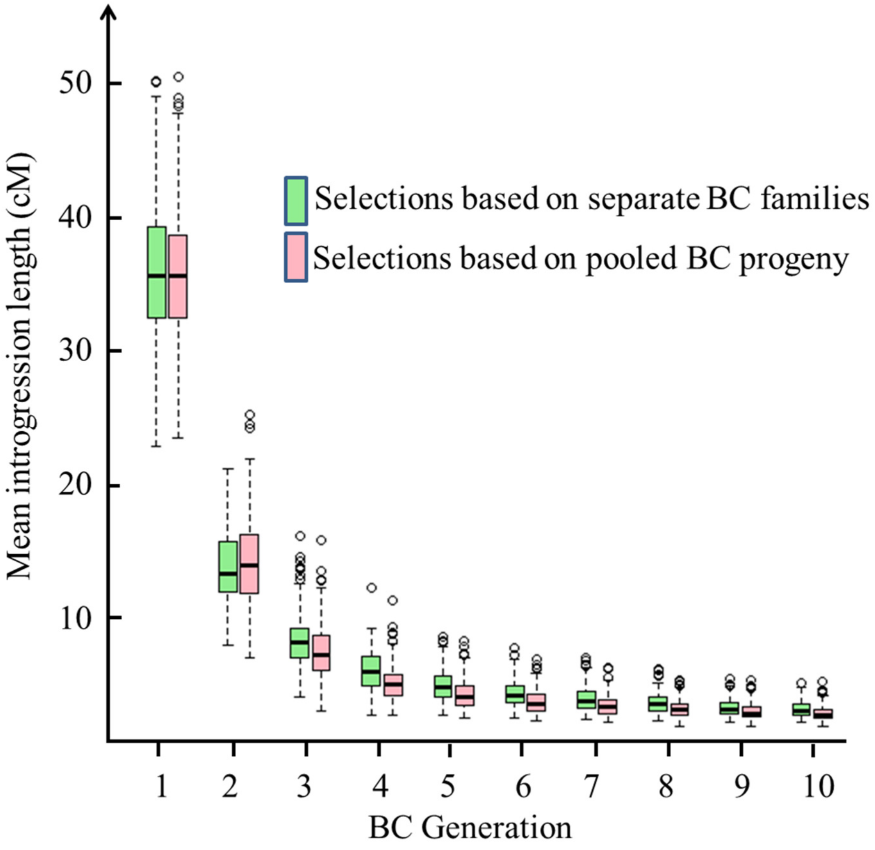


Figure 2.6 Mean introgression length (in cM) for NIL sets selected from pooled BC progeny and from families that were maintained separately so that the best individual from each family could be selected.

Furthermore, the percentage of target region covered by one, two, three, four and more than four individuals (%DEP_{1-4}, and %$DEP_{>4}$, Figure A1 to Figure A5) showed that in the BC_1 generation, almost all of target region was covered by four or more overlapping donor segments in the NIL set (considering that the %DEP_1 and %DEP_2 are zero, and %DEP_3 is close to zero in BC_1). Using the pooling strategy, the percentage of target region covered by a single individual (%DEP_1) seemed to increase throughout and at a faster rate when compared to the separation strategy (Figure A1). It was desirable to maintain more of the %$\text{DEP}_{1\text{-}2}$ where each locus in the target region was covered by only one or two individuals, and less of the undesirable %$\text{DEP}_{3\text{-}4}$ and %$\text{DEP}_{>4}$ where each locus in the target region is covered by more individuals.

Table 2.1 Median and quartiles based on the distribution of mean introgression lengths (cM) of NIL sets selected from pooled BC progeny and from BC families maintained separately.

	Gen	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Separate	**$Q_{0.75}$**	39.3	15.7	9.2	7.1	**5.6**	4.9	4.4	4.0	3.6	3.5
	$Q_{0.5}$	35.6	13.2	8.1	5.8	**4.7**	4.1	3.7	3.4	3.1	3.0
	$Q_{0.25}$	32.4	11.9	6.9	4.9	**4.0**	3.6	3.2	3.0	2.8	2.6
Pooled	**$Q_{0.75}$**	38.6	16.2	8.6	**5.7**	4.9	4.2	3.8	3.5	3.3	3.1
	$Q_{0.5}$	35.6	13.9	7.2	**4.9**	4.0	3.5	3.2	3.0	2.8	2.7
	$Q_{0.25}$	32.4	11.8	6.0	**4.1**	3.4	3.0	2.8	2.6	2.5	2.4

After the first few BC generations (beyond BC_4), %DEP_2 was the highest, followed by %DEP_1, %DEP_3, %DEP_4 and then (close to or) zero %$\text{DEP}_{>4}$. The %$\text{DEP}_{>4}$ decreased with increasing BC generations, and by the BC_5 generation, there were 0% loci in the target region covered with more than four individuals (Figure A5). Since selection in each BC generation reduced the donor introgression lengths (as a result of aiming at producing evenly spaced donor introgressions with minimal linkage drag), the number of loci covered by a multiplicity of the selected individuals decreased. Therefore, the decrease in %$\text{DEP}_{>4}$ corresponded with an increase in %$\text{DEP}_{1\text{-}4}$ in the BC_2 generation. While %DEP_4 decreased from BC_3 generation onwards, %DEP_3 increased until BC_3 generation and decreased from BC_4 generation onwards (Figure A3 and Figure A4). In particular, more loci were covered with one or two individuals (greater %$\text{DEP}_{1\text{-}2}$) and excessive coverage of loci (%$\text{DEP}_{>4}$) was eliminated faster with selection from pooled progeny when compared to that based on separate BC families.

The improvement of the pooled set of backcrossed individuals appeared to be due to multiple plants (primarily in the BC_1 and/or BC_2) with larger donor segments harboring more than one ideotypic target; which then split into two or more separate donor segment lines in further BC generations, in the event of favorable recombinations. Maintaining BC families separately would not allow these favorable recombinations to be captured, particularly in case of fewer backcrosses per selected individual. Therefore, the backcrossed individuals were pooled in each generation to select the individuals for advancement for in the rest of this study.

2.3.2 Impact of selection and backcrossing strategies

Marker assisted selection (MAS) was carried out in simulated populations to evaluate the backcrossing and selection strategies in terms of the achievable level of introgression resolution and the recovery of RP genome given realistic and finite resources.

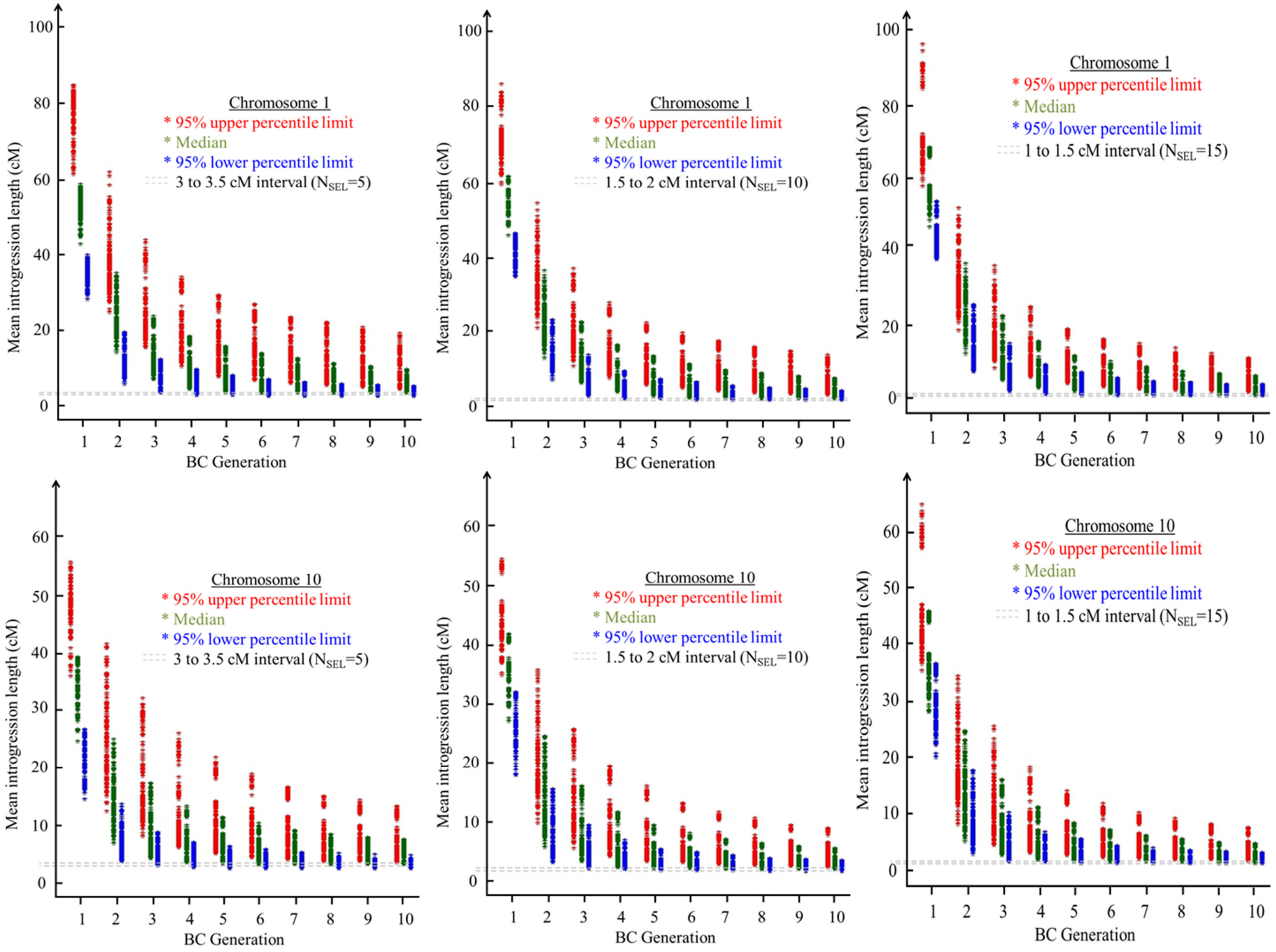


Figure 2.7 Median and the 95% percentile limits for the mean introgression length (MIL) of the selected set of individuals based on 1000 simulations for each combination of selection and breeding strategy (4 selection and 5 N_{BC1} x 5 N_{BSI} = 25 breeding strategies) producing 100 triplets at each BC generation in each plot. The grey bars represent the desired level of introgression lengths in the selected sets of NILs. The top and bottom rows correspond to mean introgressions at QTL on chromosomes 1 and 10 respectively; from left to right, the figures represent strategies for 5, 10 and 15 individuals (N_{SEL}) in the tiling path.

Selection (MAS) had an impact on both the magnitude and variance of the mean donor introgression length (MIL, Figure 2.7), and percentage of RP alleles in the non-target loci of the target chromosome (%RPC, Figure A6), in the ZmPR loci of the background chromosomes (%RPZ, Figure A7) and in the non-ZmPR loci plus non-target chromosomes (%RPN, Figure A8) across all the selection and backcrossing strategies (figures based on Supplementary files 2.1 to 2.4). The MIL decreased and the return to RP genome increased with increasing number of BC generations, but with diminishing returns. As expected (since selection was aimed at creating sets of NILs) and evidenced by the reduction in gap between the 95% lower and upper percentiles, irrespective of the breeding strategy, the variability in the MIL and proportion of RP genome decreased with increasing BC generations (Figure 2.7, Figure A6 to Figure A8).

Across all the selection and breeding strategies in this study, the MIL decreased from a range of 42.7-68.8 cM in BC_1 to 1.5-10 cM in BC_{10} generation for chromosome 1 while the corresponding decrease for the chromosome 10 was from a range of 24.8-46 cM to 1.4-7.7 cM (Table 2.2). For the same selection and breeding strategy, absolute lengths of average donor introgressions were smaller for the chromosome 10 (chromosome length = 101.9 cM, Table B1) when compared to the chromosome 1 (chromosome length = 202.4 cM, Table B1). However, in terms of the percentage of genome homozygous for the RP alleles, %RPC, %RPZ and %RPN, there seemed to not be much difference between the longest and the shortest chromosomes. A few selection and breeding strategies (discussed later) achieved high values of %RPZ (100%) in a minimum of three BC generations for both chromosomes 1 and 10 (Table 2.2).

Table 2.2 Impact of selection on the average values (medians) of genomic characteristics. The ranges (minimum and maximum values on top and bottom respectively within each cell) were across the selection and backcrossing strategies (300 values based on 4 selection strategies x 5 levels of N_{BC1} x 5 levels of N_{BSI} x 3 levels of N_{SEL}) at each BC generation.

BC	**MIL**		**%RPC**		**%RPZ**		**%RPN**	
gen	**Chr 1**	**Chr 10**	**Chr 1**	**Chr 10**	**Chr 1**	**Chr 10**	**Chr 1**	**Chr 10**
1	42.7 68.8	24.8 46	48.4 64.4	45.8 72.5	49.4 55.4	49.8 56.9	49.4 50.7	49.2 51
2	12.9 37.2	5.2 25.1	73 91.4	70.9 95.3	76.3 87.9	76.7 86.8	74.8 77.4	74.8 77.1
3	5.7 23.9	2.7 17.5	85.4 97.1	83.5 98.9	91.2 100	91.2 100	88 92.7	88 92.8
4	3.4 18.6	2 13.6	91.2 98.9	89.1 99.7	97.3 100	97.3 100	94.7 98.1	94.6 98.1
5	2.5 15.9	1.7 11.6	93.5 99.7	92.3 99.8	99.8 100	99.6 100	97.9 99.3	97.8 99.3
6	2 14.2	1.6 10.5	95 99.9	94.3 99.9	100 100	100 100	99.1 99.7	99.1 99.7
7	1.7 12.9	1.5 9.4	95.8 99.9	95.1 99.9	100 100	100 100	99.7 99.9	99.7 99.9
8	1.6 11.6	1.5 8.7	96.5 100	96 100	100 100	100 100	99.9 100	99.9 100
9	1.5 10.7	1.4 8.2	97 100	96.6 100	100 100	100 100	99.9 100	100 100
10	1.5 10	1.4 7.7	97.5 100	96.9 100	100 100	100 100	100 100	100 100

Although the RP genome coverage across the non-ZmPR and non-target chromosomes (%RPN) was lower in the first BC generation (49.4-50.7% for chromosome 1 and 49.2-51% for chromosome 10) when compared to the theoretical allele frequencies of 75% (Bernardo 2010) due to the focus of selection strategies on preselecting individuals with less linkage drag, by the BC_4 generation, some of the selection/breeding strategy combinations had recovered more %RPN (up to 98.1%) than was expected in BC_4 generation (96.8%).

The recovery %RPC of RP genome across the target chromosome was slower when compared to %RPZ and %RPN which were associated with background chromosomes unlinked to the target QTL region. Unlike the smaller ranges of %RPZ and %RPN ($\leq$ 1.5%) across all the selection and breeding strategies from generation BC_5 onwards, the ranges of MIL and %RPC were larger even till the BC_{10} generation. The differences in the values for MIL and hence %RPC were due to the effect of various selection and breeding strategies used in this study.

2.3.3 The BC_1 generation

Across all selection strategies, increasing BC_1 progeny size from 50 to 500 resulted in smaller MIL by at least 7.6 cM (Chr 1, FMW2, N_{SEL}=5, Table 2.3) and up to 20.4 cM (Chr 1, FMW0, N_{SEL}=15, Table 2.3). The introgression lengths decreased, on average, by 2 to 3 cM per 100-individual increase. The largest reduction in MIL, however, seemed to be in increasing the BC_1 progeny size from 50 to 150, especially for N_{SEL}=10 with a reduction of 3.4 to 7.3 cM and N_{SEL}=15 with 7.4 to 12 cM (Table 2.3).

The selection strategy FMW0 appeared to have selected individuals with a relatively smaller MIL in BC_1 generation with the reduction more pronounced for larger chromosomes and smaller N_{SEL}. For instance, using the FMW0 strategy, the MIL level achieved with 100 and 150 individuals for chromosome 1 when N_{SEL}=5 (50.3 and 49.1 with SEM=9.31 and 8.93 respectively) were closer to the level achieved by the other three strategies with 500 individuals (shown in bold and red font in Table 2.3).

Table 2.3 Impact of number of individuals genotyped in BC1 generation on the mean introgression length (MIL) of the selected set of NSEL (5, 10 and 15) individuals. The MIL and its' (standard error) measured after the BC1s were generated are shown for five different BC1 progeny sizes (NBC1=50, 150, 250, 350 and 500) and four selection strategies for chromosomes 1 and 10.

		Chr1				**Chr10**			
	N_{BC1}	**NODIST**	**FMW0**	**FMW2**	**FMW5**	**NODIST**	**FMW0**	**FMW2**	**FMW5**
N_{SEL} = 5	50	58.3 (11.20)	54.5 (11.26)	56.3 (11.31)	58.2 (11.39)	38.4 (6.69)	38.9 (7.63)	39.1 (7.17)	39.3 (6.84)
	150	55.8 (11.19)	**49.1** **(8.93)**	55.0 (10.80)	57.2 (10.99)	33.9 (6.54)	**31.3** **(6.01)**	35.5 (6.63)	36.4 (6.64)
	250	52.9 (10.96)	47.0 (8.68)	53.6 (10.55)	55.0 (10.95)	**31.4** **(6.79)**	28.7 (5.65)	**34.3** **(6.74)**	**34.6** **(6.56)**
	350	50.3 (10.59)	45.4 (8.39)	51.7 (10.33)	52.9 (10.56)	29.4 (6.84)	26.9 (5.46)	32.8 (6.65)	32.9 (6.57)
	500	**47.4** **(10.23)**	43.2 (8.04)	**48.7** **(9.47)**	**49.9** **(9.84)**	27.0 (6.77)	24.8 (5.42)	30.5 (6.59)	30.5 (6.49)
N_{SEL} = 10	50	61.7 (9.67)	61.3 (9.43)	60.9 (9.11)	61.3 (9.15)	40.9 (5.84)	41.8 (5.83)	41.4 (5.63)	41.1 (5.53)
	150	57.5 (7.97)	53.2 (7.00)	56.3 (7.72)	57.2 (7.76)	34.9 (4.65)	33.8 (4.46)	35.6 (4.57)	36.2 (4.61)
	250	54.5 (7.87)	51.0 (6.89)	54.4 (7.51)	54.6 (7.56)	32.2 (4.79)	31.3 (4.27)	33.9 (4.62)	34.2 (4.64)
	350	52.0 (7.90)	48.9 (6.52)	52.2 (7.27)	52.2 (7.32)	30.1 (4.88)	29.5 (4.17)	32.3 (4.65)	32.5 (4.68)
	500	48.9 (7.71)	46.2 (6.23)	49.1 (6.96)	49.1 (6.84)	27.5 (4.93)	27.1 (4.06)	29.8 (4.56)	30.0 (4.69)
N_{SEL} = 15	50	68.7 (9.90)	68.0 (9.29)	67.8 (9.56)	67.8 (9.47)	45.6 (6.27)	46.5 (6.49)	45.7 (6.39)	45.5 (6.28)
	150	58.1 (6.37)	54.6 (5.97)	56.7 (6.33)	57.8 (6.23)	35.7 (3.70)	35.3 (3.75)	36.3 (3.71)	36.6 (3.73)
	250	56.2 (6.33)	53.0 (5.86)	54.8 (6.26)	55.3 (6.13)	33.1 (3.74)	32.8 (3.60)	34.3 (3.72)	34.3 (3.79)
	350	54.1 (6.55)	50.7 (5.66)	52.2 (5.97)	52.5 (5.92)	30.9 (3.96)	30.9 (3.60)	32.3 (3.85)	32.3 (3.85)
	500	50.8 (6.54)	47.6 (5.33)	49.1 (5.92)	49.2 (5.53)	28.3 (4.01)	28.5 (3. 60)	29.7 (3.69)	29.6 (3.82)

Similarly, for chromosome 10, the MIL level based on five selected individuals from 100 and 150 BC$_1$s (33.7 and 31.3 with SEMs 6.73 and 6.01 respectively) using FMW0 strategy seemed to be closer to selection from 250 BC$_1$s using the other three strategies (bold and red font in Table 2.3). These relatively smaller introgressions might not necessarily have been closer to the target segment because the first step in selection of individuals using FMW0 was not based on the recombinations between the target segment and flanking markers (as there was no flanking marker window), but on the target segment and the background markers on the target chromosome. With a relatively higher expected frequency of individuals with recombination between the target segment and the background markers when compared to the target segment and the flanking markers (considering only the proportional genetic distance), the individuals preselected by FMW0 strategy would be more and so would the chances of double recombinants to select from which would lower the average introgression lengths of the selected set of individuals.

Whether large introgressions or smaller, and whether recombinations closer enough to the target boundary or further away, the effect of BC_1 progeny size on the composition of selected NILs was propagated throughout the further generations. The success (or failure) of a backcrossing scheme, however, was not completely dependent on the BC_1 size. The selection strategies and number of backcrosses per selected individual (N_{BSI}) played an important role in optimizing the production of QTL-targeted NILs. Since it was found that increasing the progeny sizes was redundant unless accompanied by a good selection strategy, the selection strategies are discussed first, followed by the number of backcrosses per selected individual.

2.3.4 Selection strategies

Choice of selection strategy seemed to be very important in producing QTL-targeted tiled introgression lines. For instance, the strategy FMW0 was not successful in producing NILs with desired MILs of 3 to 3.5 cM, 1.5 to 2 cM and 1 t o1.5 cM within ten BC generations for chromosome 1 (Table 2.4A-C). For the smaller chromosome 10, FMW0 was successful in creating desired sets of individuals but required large N_{BSI} (=100) to be successful in 8 to 10 BC generations. Use of FMW2 for these same breeding strategies yielded the desired tiled NILs earlier by 7 (N_{SEL}=5, Table 2.4A) and 4 to 5 BC generations (N_{SEL}=10 and 15, Table 2.4B and C) for larger N_{BSI} (=100), and by 3 to 6 BC generations for strategies with smaller progeny sizes (N_{BC1}=50 or 150 and N_{BSI}=20 or 50, Table 2.4A-C). The NODIST strategy, in comparison, required at least 1 to 2 BC generations more than FMW2.

The selection strategies FMW2 and FMW5 appeared to be equivalent in terms of the number of generations required to produce sets of tiled NILs with average introgression lengths of 3, 1.5 and 1 cM within the set margin of 0.5 cM (Table 2.4A-C). When compared to NODIST, the strategy FMW2 (and hence FMW5) seemed to have been successful in creating NILs with the desired MIL earlier and with smaller progeny sizes except for a few breeding strategies with N_{BSI}=100 where NODIST performed equivalent to FMW2 (Table 2.4).

Table 2.4 Impact of N_{BC1}, N_{BSI} and selection strategies on the success and length of the backcrossing scheme. The numbers in each cell represent the BC generation at which desired mean donor introgression length (MIL) was achieved for the selected set of individuals with donor introgressions tiled across a target QTL region of 15 cM. Twenty five breeding strategies that varied in N_{BC1} and N_{BSI}, and four selection strategies were evaluated with an objective to produce NILs with an average per NIL donor introgression length of **(A)** 3 to 3.5 cM by selecting N_{SEL}=5 individuals, **(B)** 1.5 to 2 cM by selecting N_{SEL}=10 individuals, and **(C)** 1 to 1.5 cM by selecting N_{SEL}=15 individuals. The symbol "-" is used to represent the breeding/selection strategy combination for which the desired MIL was not achieved within the 95% percentile limits of the simulated distributions of MIL based on 1000 simulations for each combination of breeding and selection strategy.

(A) N_{SEL}=5									
N_{BC1}	**N_{BSI}**	**Chr 1**				**Chr 10**			
		NODIST	**FMW0**	**FMW2**	**FMW5**	**NODIST**	**FMW0**	**FMW2**	**FMW5**
50	5	-	-	-	-	-	-	-	-
150	5	-	-	-	-	-	-	-	-
250	5	-	-	-	-	-	-	-	-
350	5	-	-	-	-	-	-	-	-
500	5	-	-	-	-	-	-	-	-
50	10	-	-	-	10	-	-	8	8
150	10	-	-	-	10	-	-	9	10
250	10	-	-	-	-	-	-	-	-
350	10	-	-	-	-	-	-	-	-
500	10	-	-	-	-	-	-	-	-
50	20	-	-	7	7	-	-	6	5
150	20	-	-	7	7	-	-	6	6
250	20	-	-	8	8	-	-	6	6
350	20	-	-	9	8	10	-	7	7
500	20	-	-	9	8	9	-	7	6
50	50	8	-	5	4	6	-	4	4
150	50	7	-	5	5	6	-	4	4
250	50	7	-	5	5	5	-	4	4
350	50	6	-	5	5	5	-	4	4
500	50	6	-	5	5	6	-	4	4
50	100	5	-	4	4	4	10	3	3
150	100	5	-	4	4	4	10	3	3
250	100	4	-	4	4	4	10	3	3
350	100	4	-	4	4	3	-	3	3
500	100	4	-	4	4	3	-	3	3

Table 2.4 Continued

(B) N_{SEL}=10									
N_{BC1}	N_{BSI}	**Chr 1**				**Chr 10**			
		NODIST	**FMW0**	**FMW2**	**FMW5**	**NODIST**	**FMW0**	**FMW2**	**FMW5**
50	5	-	-	-	-	-	-	-	-
150	5	-	-	-	-	-	-	-	-
250	5	-	-	-	-	-	-	-	-
350	5	-	-	-	-	-	-	-	-
500	5	-	-	-	-	-	-	-	-
50	10	-	-	-	-	-	-	-	-
150	10	-	-	-	-	-	-	-	-
250	10	-	-	-	-	-	-	-	-
350	10	-	-	-	-	-	-	-	-
500	10	-	-	-	-	-	-	-	-
50	20	-	-	-	-	-	-	10	10
150	20	-	-	-	-	-	-	9	9
250	20	-	-	-	-	-	-	10	10
350	20	-	-	-	-	-	-	10	10
500	20	-	-	-	-	-	-	-	10
50	50	9	-	8	7	7	-	6	5
150	50	10	-	7	7	8	-	5	5
250	50	9	-	7	7	7	-	5	5
350	50	9	-	7	7	7	-	5	5
500	50	9	-	7	7	7	-	5	5
50	100	6	-	5	5	5	9	4	4
150	100	6	-	5	5	5	9	4	4
250	100	6	-	5	5	5	9	4	4
350	100	6	-	5	5	5	9	4	4
500	100	6	-	5	5	5	9	4	4

Table 2.4 Continued

(C) N_{SEL}=15									
N_{BC1}	N_{BSI}	Chr 1				Chr 10			
		NODIST	FMW0	FMW2	FMW5	NODIST	FMW0	FMW2	FMW5
50	5	-	-	-	-	-	-	-	-
150	5	-	-	-	-	-	-	-	-
250	5	-	-	-	-	-	-	-	-
350	5	-	-	-	-	-	-	-	-
500	5	-	-	-	-	-	-	-	-
50	10	-	-	-	-	-	-	-	-
150	10	-	-	-	-	-	-	-	-
250	10	-	-	-	-	-	-	-	-
350	10	-	-	-	-	-	-	-	-
500	10	-	-	-	-	-	-	-	-
50	20	-	-	-	-	-	-	-	-
150	20	-	-	-	-	-	-	-	-
250	20	-	-	-	-	-	-	-	-
350	20	-	-	-	-	-	-	-	-
500	20	-	-	-	-	-	-	-	-
50	50	10	-	9	9	8	-	7	6
150	50	10	-	7	8	8	-	6	6
250	50	10	-	8	8	8	-	6	6
350	50	10	-	8	8	7	-	6	6
500	50	10	-	8	8	7	-	6	6
50	100	7	-	6	6	5	9	5	5
150	100	6	-	5	5	5	8	4	4
250	100	6	-	5	5	5	9	5	5
350	100	6	-	5	6	5	9	4	4
500	100	6	-	6	6	5	9	4	4

For N_{SEL}=5 (desired MIL=3 cM), FMW2 could be used with combinations of (N_{BC1}, N_{BSI}) = (50, 20), (50, 50) and (50, 100) to obtain desired NIL sets in 7, 5 and 4 BC generations for chromosome 1, and in 6, 4 and 3 BC generations for chromosome 10 which meant significantly lesser resources and time when compared to breeding strategies with same and larger progeny sizes using NODIST strategy (shaded in yellow in Table 2.4A). Similarly for N_{SEL} = 10 and 15 (desired MILs 1.5 and 1 cM), FMW2 could be used to better optimize the breeding program by 1 to 2 BC generations when compared to using NODIST strategy for selecting individuals (shaded in yellow in Table 2.4B and C). In a real world situation however, more than 50 individuals are generally used in BC_1 generation to reduce risks for cases where pollination events are unsuccessful.

2.3.5 Number of backcrosses per selected individual (N_{BSI})

N_{BSI} significantly affected the duration of backcrossing to successfully produce, within ten BC generations, tiled introgressions across the targeted QTL for both chromosomes 1 and 10. Increase in N_{BSI} decreased the backcrossing duration (Table 2.4). In general, when the number of BC generations is prefixed, more N_{BSI} were required to achieve the MIL threshold for chromosome 1 when compared to that of chromosome 10 (Table 2.4). The minimum required N_{BSI} was also affected by the choice of the selection strategy (described in section 2.3.4) and increased with decrease in the desired MIL (Table 2.4).

2.3.5.1 Selection for desired MIL of 3 to 3.5 cM (N_{SEL}=5)

For N_{BSI} = 5 and 10 individuals, the breeding strategies were either unsuccessful (selection strategies NODIST, FMW0 and FMW2 for chromosome 1, and NODIST and FMW0 for chromosome 10) or produced desired NIL sets in 10 and 8 to 10 BC generations respectively for chromosome 1 and 10 respectively (Table 2.4A).

Contrary to backcrossing strategies with smaller N_{BSI}, increase in N_{BSI} to 20, 50 and 100 considerably reduced the backcrossing duration to 7, 5 and 4BC generations respectively for chromosome 1, and 6, 4 and 3 BC generations respectively for chromosome 10 with FMW2 strategy (yellow cells in Table 2.4A under FMW2 column). When the NODIST strategy was applied for selection, more N_{BSI} (>20) were required to produce the desired NILs, and this strategy was comparable to FMW2 in the number of BC generations required for success only for large N_{BSI} (=100) and N_{BC1} (>250). The best breeding strategies within a specific level of N_{BSI} were associated with N_{BC1} = either 50 (or 150 in actual practice) for the FMW2 strategy indicating that it probably was redundant to use a large N_{BC1} when using FMW2. Instead, a better option could be to divert those resources towards increased N_{BSI}. NODIST, however, seemed to perform better with increasing N_{BC1} (Table 2.4A).

For the breeding strategies that successfully produced desired NIL sets using FMW2 strategy, 100% RPZ and greater than 96% RPN were recovered when the number of generations of backcrossing was at least 4 (Table 2.5A). All breeding strategies with N_{BSI} = 100 for chromosome 10 produced the desired NILs in as early as 3 BC generations but with %RPZ ranging from 93.3 to 98.2 and %RPN from 90.1 to 92.2 (Table 2.5A).

2.3.5.2 Selection for desired MIL of 1.5 to 2 cM and 1 to 1.5 cM (N_{SEL}=10 and 15)

The minimum N_{BSI} required for the creation of tiled NILs with average introgression lengths of 1.5 cM and 1 cM increased when compared to that required to produce NIL sets with desired introgression lengths of 3 cM (Table 2.4A to C). Irrespective of the selection strategy and the size of N_{BC1} (up to 500 individuals), the breeding strategies did not reach the MIL threshold for chromosome 1 in ten BC generations when N_{BSI} was 5, 10 and 20 for desired NILs of 1.5 and 1 cM introgressions in the target QTL region (Table 2.4B and C). For chromosome 10, however, with N_{BSI}=20 and use of selection strategies FMW2 or FMW5 desired NIL sets were produced in 9 to 10 BC generations for desired MIL of 1.5 cM (Table 2.4B). For all the successful breeding strategies using FMW2 selection, high recovery of %RPZ (100%) and %RPN (97%) were found for both the tiled sets with desired MIL of 1.5 cM and 1 cM (Table 2.5B and C).

Table 2.5 Percentage of RP homozygosity in the background, %RPZ and %RPN, at the BC generation BC_X in which the breeding strategies (combination of N_{BC1} and N_{BSI}) were successful using the selection strategy FMW2 in creating the tiled NILs with the desired mean introgression lengths of (A) 3 to 3.5 cM cM when N_{SEL}=5, (B) 1.5 to 2 cM when N_{SEL}=10, and (C) 1 to 1.5 cM when N_{SEL}=15.

(A) N_{SEL}=5							
		Chromosome 1			**Chromosome 10**		
N_{BC1}	**N_{BSI}**	**BC_X**	**%RPZ**	**%RPN**	**BC_X**	**%RPZ**	**%RPN**
50	10	-	-	-	8	100	100
150	10	-	-	-	9	100	100
50	20	7	100	99.9	6	100	99.6
150	20	7	100	99.9	6	100	99.6
250	20	8	100	100	6	100	99.6
350	20	9	100	100	7	100	99.9
500	20	9	100	100	7	100	99.9
50	50	5	100	99	4	100	96.7
150	50	5	100	99.1	4	100	97.3
250	50	5	100	99.1	4	100	97.4
350	50	5	100	99.2	4	100	97.5
500	50	5	100	99.2	4	100	97.7
50	100	4	100	97.2	3	93.3	90.1
150	100	4	100	97.6	3	94.8	91.1
250	100	4	100	97.7	3	97	91.6
350	100	4	100	97.8	3	97	92
500	100	4	100	98	3	98.2	92.2

Table 2.5 Continued

(B) N_{SEL}=10

		Chromosome 1			Chromosome 10		
N_{BC1}	N_{BSI}	BC_X	%RPZ	%RPN	BC_X	%RPZ	%RPN
50	20	-	-	-	10	100	100
150	20	-	-	-	9	100	100
250	20	-	-	-	10	100	100
350	20	-	-	-	10	100	100
500	20	-	-	-	-	-	-
50	50	8	100	99.9	6	100	99.6
150	50	7	100	99.8	5	100	99.1
250	50	7	100	99.8	5	100	99.1
350	50	7	100	99.8	5	100	99.1
500	50	7	100	99.8	5	100	99.1
50	100	5	100	99	4	100	97.1
150	100	5	100	99.1	4	100	97.5
250	100	5	100	99.2	4	100	97.7
350	100	5	100	99.1	4	100	97.7
500	100	5	100	99.2	4	100	97.8

(C) N_{SEL}=15

		Chromosome 1			Chromosome 10		
N_{BC1}	N_{BSI}	BC_X	%RPZ	%RPN	BC_X	%RPZ	%RPN
50	50	9	100	100	7	100	99.8
150	50	7	100	99.8	6	100	99.6
250	50	8	100	99.9	6	100	99.6
350	50	8	100	99.9	6	100	99.6
500	50	8	100	99.9	6	100	99.6
50	100	6	100	99.6	5	100	99
150	100	5	100	99.1	4	100	97.5
250	100	5	100	99.1	5	100	99.2
350	100	5	100	99.2	4	100	97.8
500	100	6	100	99.6	4	100	97.8

3 RECOMBINATION PATTERNS IN FOUR PHOTOPERIOD QTL REGIONS IN MAIZE USING MULTICROSS LINKAGE DATA

3.1 Introduction

Understanding the recombination rates and patterns in a QTL region is important in the success of a QTLs' fine-mapping to gene or near gene resolution. The efficiency of breeding programs depends on the recombination rates and positions of recombination events between desirable and non-desirable chromosomal regions to ensure that the favorable linkage blocks are retained. For instance, backcrossing is affected, among other factors, by the process of recombination in possibilities to better eliminate linkage drag, and in quickly recovering the recurrent parent.

Recombination rates in plants vary among and within species, across and along chromosomes, both locally and globally (Mézard 2006, Henderson 2012), particularly in maize (Falque et al. 2009, Gore et al. 2009, Bauer et al. 2013). In yeast (*Saccharomyces cerevisiae*), for which recombinations have been well characterized, alternating high and low recombinogenic regions have been observed every 50 kb (on average) along the non-centromeric chromosome space (Gerton et al. 2000). In maize (*Zea mays* ssp. *mays* L.) and other eukaryotes, genic regions have been suggested to be more recombinogenic when compared to the non-genic regions (Thuriaux 1977, Brown and Sundaresan 1990, Civardi et al. 1994). In spite of the requirement of large populations to analyze intra-genic meiotic recombinations (due to low intragenic recombination rate within a specific gene), it has been possible to examine if the recombinations are uniformly distributed within a gene (Dooner 1986, Dooner and Martinez-Ferez 1997) and to identify preferential sites or recombination hotspots within a gene (Eggleston et al. 1995, Patterson et al. 1995, Xu et al. 1995). Hotspots exhibiting much higher recombination in a region of chromosome than the average recombination across the entire genome are thought to contain special sites for recombination initiation (Litchen and Goldman 1995).

Genetic factors underlying variation in recombination either affect the entire genome or specific regions of the genome in increasing or decreasing recombination (Enns and Larter 1962, Moens 1969, Cornu et al. 1988). Recombination rates are also affected by environmental factors such as temperature, and water stress (Maguire 1968, Verde 2003, Parsons 1988, Francis et al. 2007).

3.1.1 NILAS resource

Near isogenic lines in allelic series (NILAS) are tiled introgression paths useful for fine mapping causal polymorphisms in a QTL, as well as for dissecting the linked cryptic phenotypic variation and the loci that contribute to this variation that would be inadvertently selected during the breeding process.

Through marker assisted selection, we are creating tiled NIL libraries for four previously identified maize QTL (*ZmPR1-4*; Coles et al. 2010, Hung et al. 2012) using seven elite tropical inbred lines as donor parents in two different recurrent parent backgrounds and targeting ten tiled lines per family. The large dataset collected through this introgression process is a tangible way to investigate the localized recombination rate at these important QTLs. In this study, thirteen BC_1 populations have been used to

(1) understand and evaluate recombination patterns in the female meioses, and (2) determine if there is an impact of genetic background on the recombination rates across the four ZmPR QTLs.

3.2 Materials and methods

3.2.1 Empirical BC_1 data

Seven tropical inbred lines CML10, CML258, CML277, CML341, CML373, Tzi8 and Tzi9 from the International Maize and Wheat Improvement Center (CIMMYT, the CML lines) and the International Institute of Tropical Agriculture (IITA, the Tzi lines) were used as donor parents (DP) that showed promising yields in North Carolina trials (long day length) when testcrossed with elite U.S. inbreds (Nelson and Goodman 2008, Nelson 2009). Two different ex-PVP temperate inbred lines LH123Ht and 2369 were used as the recurrent parents (RP). LH123Ht is classified as non-stiff stalk and most closely resembles Oh43Ht. 2369 is classified as stiff-stalk and is an improved version of B73 from which the maize genome was sequenced (Schnable et al. 2009).

For each of the 14 tropical x temperate crosses (Table 3.1), 192 (or 156 and 168 in two and one cross respectively) BC_1s were selected in 2011 in Puerto Rico, USA which has a tropical climate. The seven tropical parents were the female lines while the two temperate parents were the male lines.

Table 3.1 The fourteen crosses based on which the BC_1 progeny in this study were produced. Two temperate and seven tropical inbred lines were used as recurrent and donor parents (RPs and DPs).

Cross (RP x DP)	**Cross Code**
2369 x CML10	G11
2369 x CML258	G12
2369 x CML277	G13
2369 x CML341	G14
2369 x CML373	G15
2369 x Tzi8	G16
2369 x Tzi9	G17
LH123Ht x CML10	G18
LH123Ht x CML258	G19
LH123Ht x CML277	G20
LH123Ht x CML341	G21
LH123Ht x CML373	G22
LH123Ht x Tzi8	G23
LH123Ht x Tzi9	G24

For genotyping, an iPLEX assay of 80 bi-allelic single nucleotide polymorphism (SNP) markers, with 14, 11, 20 and 21 SNPs in the ZmPR1-4 regions respectively (Table 3.2) along with 2 to 5 SNPs per background chromosome, was used to identify desirable progeny in the BC_1 generation in the process of creating near-isogenic lines in an allelic series in further BC generations. The marker data in the BC_1 generation across the four ZmPR QTLs for 13 crosses (data for the cross 2369 x CML258, G12, were not available) were used in this study. Since all the four ZmPR loci are segregating in the BC_1 generation, it allowed for a good opportunity to elucidate the trends in segregation and recombination at these four ZmPR QTLs across different genetic backgrounds.

Table 3.2 Single nucleotide polymorphism (SNP) markers used to genotype the maize (*Zea mays* L.) photoperiod response (ZmPR) QTL regions on chromosomes 1, 8, 9 and 10.

	Marker name	**Marker code**	**Physical Distance (bp)**	**Genetic Distance (cM)**
Chromosome 1	Z1-OL_PZE-101077530	M1_1	61667318	72.7
	Z1-OL_PZE-101079397	M1_2	63894760	73.4
	Z1-OL_PZE-101120159	M1_3	66395872	74.3
	Z1-OL_PZE-101082480	M1_4	69699798	75.8
	Z1-PL_PZE-101095352	M1_5	93521773	86.2
	Z1-PK_PZE-101100455	M1_6	96611366	87
	Z1-PR_PZE-101102064	M1_7	99992827	87.5
	Z1-PR_PZE-101102066	M1_8	99992943	87.5
	Z1-BR_PZE-101102630	M1_9	101586103	87.7
	Z1-OR_PZE-101106079	M1_10	109394965	88.5
	Z1-OR_PZE-101106080	M1_11	109395908	88.5
	Z1-BR_PZE-101126954	M1_12	160948965	91.2
	Z1-OR_PZE-101132120	M1_13	170475870	93.4
	Z1-OR_SYN36897	M1_14	180520434	97
Chromosome 8	Z2-OL_PZE-108053921	M8_1	95954176	56.9
	Z2-OL_PZE-108056211	M8_2	101176032	57.8
	Z2-OL_PZE-108063980	M8_3	114014118	62.1
	Z2-OL_PZE-108064003	M8_4	114025247	62.1
	Z2-BL_PZE-108071041	M8_5	124287804	66.7
	Z2-PL_PZE-108072703	M8_6	126080719	67.3
	Z2-PL_PZE-108072730	M8_7	126287026	67.3
	Z2-PR_PZE-108073925	M8_8	129070423	68.2
	Z2-PR_PZE-108074258	M8_9	129496028	68.4
	Z2-BR_PZE-108078317	M8_10	134065087	70.8
	Z2-OR_PZE-108090173	M8_11	147216431	77.5
Chromosome 9	Z3-OL_SYN33106	M9_1	18891784	33.1
	Z3-OL_SYN32163	M9_2	19533785	34.2
	Z3-OL_PZE-109022419	M9_3	22784946	39.6
	Z3-OL_SYN5266	M9_4	23540249	40.9
	Z3-OL_SYN34182	M9_5	23754839	41.3
	Z3-BL_PZE-109031097	M9_6	35902596	43.4
	Z3-PL_PZE-109032519	M9_7	38582303	43.6
	Z3-PL_PZE-109034151	M9_8	42838932	43.8
	Z3-PK_SYN36476	M9_9	45294678	44

Table 3.2 Continued

	Marker name	Marker code	Physical Distance (bp)	Genetic Distance (cM)
	Z3-PR_PZE-109037872	M9_10	55354627	44.6
	Z3-PK_PZE-109038364	M9_11	56918453	44.7
	Z3-PK_PZE-109040941	M9_12	63685732	45.1
	Z3-BR_PZE-109044890	M9_13	76535233	45.8
	Z3-BR_PZE-109045034	M9_14	76758542	45.8
	Z3-OR_PZE-109049849	M9_15	86662249	46.3
	Z3-OR_SYN32485	M9_16	86864564	46.3
	Z3-OR_SYN32492	M9_17	86864855	46.3
	Z3-BR_SYN35232	M9_18	105194953	51
	Z3-OR_PZE-109073536	M9_19	118862807	57
	Z3-OR_PZE-109073618	M9_20	118940679	57.1
Chromosome 10	Z4-OL_PZE-110014880	M10_1	14496146	31.7
	Z4-OL_PUT-163a-21330234-1513	M10_2	24611535	35
	Z4-BL_PZE-110021296	M10_3	28602144	35.6
	Z4-BL_PZE-110022324	M10_4	31311797	36
	Z4-BL_PZE-110022958	M10_5	33615372	36.1
	Z4-BL_PZE-110032266	M10_6	50534422	37
	Z4-BL_PZE-110024035	M10_7	52620884	37.1
	Z4-OL_PZE-110033839	M10_8	63813778	37.8
	Z4-OL_PZE-110041758	M10_9	79940873	39.1
	Z4-BL_PZE-110044652	M10_10	85187030	40.1
	Z4-BL_PZE-110046826	M10_11	87799475	40.7
	Z4-PL_PZE-110048295	M10_12	90479827	41.2
	Z4-PK_SYN15285	M10_13	98704920	42.9
	Z4-PK_PZE-110052325	M10_14	98944143	43
	Z4-PR_PZE-110054162	M10_15	102808150	43.5
	Z4-PR_PZE-110054571	M10_16	103737016	43.6
	Z4-BR_PZE-110055034	M10_17	105629209	43.8
	Z4-OR_PZE-110059287	M10_18	113394727	45.4
	Z4-OR_PZE-110060077	M10_19	114238975	45.7
	Z4-OR_PZE-110071740	M10_20	128025397	53.1
	Z4-OR_PZE-110072863	M10_21	129354223	54.7

3.2.2 Data analyses

For each RP x DP x photoperiod-QTL combination, the associated SNP marker data were converted to 0-1 data and represented as a KxM matrix of 0s and 1s where K and M denote the numbers of individuals and markers respectively, and 0 and 1 represent RP/RP homozygous and RP/DP heterozygous genotypes respectively determined using identity-by-descent (IBD, discussed in Section 2.1.3). The marker data contained known genotyping errors, in the form of homozygous DP/DP genotypes which were not expected in BC progeny, and missing values which were represented by 2 and 9 respectively. Unknown genotyping errors will remain mostly undiscovered with the exception of selected individuals for which there were BC_2 genotyping data.

3.2.2.1 Segregation analysis

In theory, at each locus the BC_1 progeny inherit one allele each from the RP and F_1 hybrid. While the two alleles produced by the inbred RP are the same genotype, the two alleles produced by F_1 hybrid are combinations of the RP and DP genotype equally likely in each gamete. The frequencies of RP/RP (coded as 0) and RP/DP (coded as 1) genotypes at each locus were consequently expected to be in 1:1 ratio. Segregation of the observed genotypic ratios of SNP markers was tested for deviation from the expected 1:1 using the Pearson's Chi-square goodness-of-fit test with 1 df (Agresti 2007). A region within the QTL region was identified as showing distorted segregation from the expected segregation ratio at a predefined type I error rate of 0.05 if at least three consecutive markers showed segregation distortion. This was used to minimize false positives created by incorrect genotyping calls.

3.2.2.2 Recombination frequency

In all crosses, the donor parent was used as a female parent, so it was possible to measure only the recombinations in the context of female meioses. The number of recombinations in each QTL region was counted by the switches in parentage, from 0s to 1s or vice versa, using haplotypes or contiguous stretches of parental sequences in each photoperiod QTL (Supplementary Files 3.2 to 3.5). The agglomerative hierarchical cluster analysis was used to find the crosses that had similar recombination rates. Clustering was performed using the hclust function in R (R development core team 2012).

3.2.2.3 Pairwise recombination

An F_1 x RP cross can produce four different genotypes at every pair of markers – RP/RP double homozygous genotypes at both markers, RP/RP at the first marker and RP/DP at the second marker, RP/DP at the first marker and RP/RP at the second marker or RP/DP heterozygous genotype at both markers (00, 01, 10 or 11 respectively). The frequencies of parental (00 and 11) and recombinant (01 and 10) genotypes with respect to a pair of markers are $0.5(1-r)$ and $0.5r$ where r is the recombination fraction/frequency between the two markers (Xu 2013, Chapter 2).

Recombination fractions between sequential pairs of markers were computed as

$$\hat{r} = \frac{n_{01} + n_{10}}{n_{00} + n_{01} + n_{10} + n_{11}}$$

where n_{00} and n_{11}denoted the counts of parental genotypes 00 and 11 respectively, and n_{01} and n_{10} denoted the counts of recombinant genotypes 01 and 10 respectively.

Observed estimates of pairwise recombination fractions (Supplementary File 3.6) were used to compute map distances between the adjacent markers. The Haldane, Kosambi and Morgan map functions resulted in similar genetic distances with the cM-distance computed using Morgan (complete interference assumption) ≤ Kosambi (partial interference) ≤ Haldane (no interference). The differences in cM-distance estimates, when present, were negligible as the markers within each ZmPR region were tightly linked. The estimates based on Kosambi function (Supplementary File 3.6) were used in computing the average recombination rate of a ZmPR region, expressed as genetic map length of the region or marker interval in cM per megabase pair (cM/Mbp).

3.3 Results

Three, 1, 8 and 6 markers of the total 14, 11, 20 and 21 markers (Table 3.2) genotyped in the ZmPR1-4 QTL regions respectively were removed from any further analysis. These markers were removed from analysis due to one of the following reasons: 1) the marker had more than 20% (up to 79%) missing values and either the RP homozygous or the RP/DP heterozygous genotypes were not called, 2) the marker was non-polymorphic and consisted of only the RP homozygous genotypes whereas heterozygous genotypes were expected in close to 50% of the individuals for which the flanking markers of the discarded marker were heterozygous genotypes, 3) the marker had more than 10% missing values and among the non-missing values, there existed a number of incompatible genotypes indicating double recombinations on either side of the marker within very short inter-marker distances and thus signaled towards genotyping errors in the marker, and 4) the marker was comprised of all homozygous

RP/RP, DP/DP genotypes and missing values (no RP/DP heterozygotes which were expected in the BC_1 progeny).

The markers retained for analysis were located within a genetic distance of 24.3 cM, 20.6 cM, 24 cM and 23 cM on chromosomes 1, 8, 9 and 10 respectively. For these markers, missing values where possible, were imputed by observing the sequence of RP/RP and RP/DP alleles across the QTL region as haplotype blocks (traced back as having been inherited from the same parent). Missing values at a marker with flanking markers carrying the same genotype RP/RP (or RP/DP) were substituted with the genotype RP/RP (or RP/DP) assuming that the probability of odd numbers of recombinations between the marker and its' flanking markers (resulting in even number of recombinations between the adjacent markers), given the genetic distance between the markers, was highly unlikely. After imputation, the percentage of missing values ranged from 0 to 10.2% across all the crosses (%MISS, Table 3.3). Presence of DP/DP double homozygotes which were not expected in a backcross with the RP were found and ranged from 0 to 18.8% (%DP, Table 3.3). Crosses with a large amount of DP homozygosity (> 5%) did not seem to have a high (> 5%) missing percentage and vice versa.

3.3.1 Segregation analysis

The regions of segregation distortion seemed to depend significantly on the parental origins. There was no evidence of significantly distorted regions in the four ZmPR QTLs when the BC_1s were associated with crosses involving 2369 as the RP and

CML10 (G11), CML277 (G13), Tzi8 (G16) or Tzi9 (G17) as the DPs, and LH123Ht as the RP and CML277 (G20) as the DP (Table 3.4A-D, Supplementary File 3.1).

Table 3.3 Percentage of DP/DP homozygotes and missing values in the 13 BC_1 families.

Cross	%DP				%MISS			
	ZmPR1	**ZmPR2**	**ZmPR3**	**ZmPR4**	**ZmPR1**	**ZmPR2**	**ZmPR3**	**ZmPR4**
G11	1.1	0.6	1.1	1.6	0.5	0.6	0.5	0.7
G13	0.8	0.2	0.1	0.0	1.5	1.6	0.9	2.9
G14	1.3	0.5	2.0	1.4	**8.1**	**9.1**	**6.9**	**10.2**
G15	0.1	0.1	0.0	0.0	**8.5**	**8.0**	**5.8**	**10.2**
G16	0.0	0.0	0.0	0.0	0.6	1.2	0.5	0.7
G17	1.8	2.8	4.1	2.7	0.3	0.5	0.4	0.1
G18	**13.5**	**16.7**	**18.8**	**10.7**	0.5	0.8	1.7	0.9
G19	0.0	0.0	0.6	0.0	0.1	0.1	0.2	0.0
G20	0.5	0.7	0.6	0.6	**7.9**	2.8	0.9	0.5
G21	4.4	1.6	4.0	**5.4**	0.4	1.0	0.3	0.3
G22	**11.2**	**10.6**	**9.8**	**14.3**	1.4	1.6	0.3	0.7
G23	0.2	0.0	3.0	0.0	**8.3**	2.3	**5.8**	**9.7**
G24	1.3	0.6	2.2	0.6	0.2	0.8	0.8	0.0

For the remaining BC_1 families (G14, G15, G18, G19, G21, G22, G23, G24), segregation distortion regions were found with preference for either the temperate or the tropical alleles in at least one of the four QTLs. For the ZmPR2 and ZmPR3 regions, in particular, there seemed to be significant segregation distortion when the LH123Ht was RP.

3.3.1.1 ZmPR1 QTL region (Chromosome 1)

The region marked by M1_1 to M1_4 (72.7 cM to 75.8 cM, 61.6 Mbp to 69.7 Mbp, Table 3.2) showed significantly distorted segregation ($p < 0.05$, Table 3.4A) in both the crosses involving CML373 (G15 and G22). The frequency of RP alleles was significantly higher (57.2 to 58.1%) when 2369 was the RP as opposed to the significantly lower RP allele frequency (36.8 to 41.4%) when LH123Ht was the RP. For the remaining markers in the ZmPR1 region (M1_5 to M1_14), the frequency of RP alleles continued to be higher for G15 and lower for G22, although not statistically significant. Another region in which all consecutive markers (M1_5 to M1_14, 87 cM to 97 cM) showed evidence for significant segregation distortion ($p < 0.05$, Table 3.4A) was found in the cross LH123Ht x CML258 with preference for RP alleles (57.3 to 61.6%).

Segregation pattern in the cross 2369 x CML341 (G14), unlike other crosses, showed multiple switches in the frequencies of homozygous RP and heterozygous RP/DP genotypes.

3.3.1.2 ZmPR2 QTL region (Chromosome 8)

The region marked by M8_1 to M8_4 (56.9 to 62.1 cM) showed significant segregation distortion for all consecutive markers ($p < 0.05$, Table 3.4B) in five crosses, one cross involving 2369 as RP (G14) and four crosses (G18, G21, G22 and G23) involving LH123Ht as RP. In three of these crosses (G18, G21 and G23), the distortion extended to the entire QTL region under study (to M8_11 at 77.5 cM). For all these crosses the frequency of homozygous RP alleles seemed to be lesser (26.9 to 44%) when compared to the heterozygous RP/DP alleles, except in the cross LH123Ht x Tzi8 for which the frequency of RP alleles was significantly higher (64.6 to 72.4%) at markers M8_6 to M8_11 (Table 3.4B).

Segregation pattern for the cross 2369 x CML341 (G14) showed multiple switches in the frequencies of homozygous RP and heterozygous RP/DP genotypes across the ZmPR2 QTL, as observed in ZmPR1 QTL region.

3.3.1.3 ZmPR3 QTL region (Chromosome 9)

Deviated segregation was found in five crosses, one cross with 2369 as RP (G14) and four crosses with LH123Ht as RP (G18, G19, G22 and G23), with significantly lower percentage of temperate RP alleles (15.2 to 41.7%, Table 3.4C) across the entire ZmPR3 QTL region (markers M9_2 to M9_17, 34.2 to 46.3 cM). For the BC_1 family associated with the cross LH123Ht x CML373 (G22), however, markers M9_8 to M9_15 were not statistically significant although they had lower percentage of homozygous RP alleles (< 45%).

3.3.1.4 ZmPR4 QTL region (Chromosome 10)

Two regions marked by M10_1 to M10_9 (31.7 cM to 39.1 cM) and M10_13 to M10_21 (42.9 cM to 54.7 cM) showed evidence for distorted segregation ($p < 0.05$, Table 3.4D). The BC_1s associated with the crosses 2369 x CML373 (G15) and LH123Ht x CML10 (G18) showed distortion in both these regions, where as those associated with the cross LH123Ht x CML341 (G21) showed distorted segregation for markers in the region M10_2 to M10_9 (35 cM to 39.1 cM). For all these crosses a significantly lower frequency of homozygous RP alleles was observed ranging from 33.5 to 41.8% (Table 3.4D). Switch in preference from tropical to temperate to tropical alleles was observed for the cross 2369 x CML341 (G14) across the ZmPR4 QTL as seen earlier in the ZmPR1 and ZmPR2 QTL regions.

Table 3.4 Percentage of BC_1 progeny homozygous for RP alleles at markers in the ZmPR QTL regions on (A) chromosomes 1, ZmPR1; (B) chromosome 8, ZmPR2; (C) chromosome 9, ZmPR3; and (D) chromosome 10, ZmPR4. Orange (≤ 45%) and blue (≥ 55%) represent percentages that are not within 5% of the expected 50% of the individuals. The boxed percentages show significantly distorted heterozygosity segregation ($p < 0.05$) at the associated marker. See Supplementary File 3.1 for details on genotypic frequencies.

(A) Chromosome 1, ZmPR1

		cM position	72.7	74.3	75.8		86.2	87	87.5	87.5	87.7	88.5	88.5	97
RP	**DP**	**cross code**	**M1_1**	**M1_3**	**M1_4**		**M1_5**	**M1_6**	**M1_7**	**M1_8**	**M1_9**	**M1_10**	**M1_11**	**M1_14**
2369	**CML10**	**G11**	50.0	50.3	48.7	QTL peak at 84.9 cM (AGP2 89 Mbp)	52.9	52.9	54.0	54.0	54.0	53.7	52.4	54.0
2369	**CML277**	**G13**	43.8	45.5	45.5		47.9	48.4	48.9	48.9	48.7	48.7	48.4	47.6
2369	**CML341**	**G14**	55.1	56.3	55.8		29.3	52.9	52.8	28.2		52.8	52.4	29.2
2369	**CML373**	**G15**	57.2	57.9	58.1		56.0	55.8	55.4		55.1	55.4	55.6	52.4
2369	**Tzi8**	**G16**	44.2	45.8	46.4		45.5		46.1	46.1	46.1	46.1	46.1	47.1
2369	**Tzi9**	**G17**	44.3	46.0	46.8		45.5	46.8	47.1	45.5	45.5	47.9	47.6	48.4
LH123Ht	**CML10**	**G18**	45.1	46.7	46.4		45.5	47.3	46.1	46.3	46.4	46.4	46.4	43.9
LH123Ht	**CML258**	**G19**	51.0	52.1	52.6		56.3	57.3	57.8	57.8	57.8	58.3	58.3	61.6
LH123Ht	**CML277**	**G20**	52.9	53.9	53.6		52.9	51.3	52.4	50.9	51.2	52.1	51.3	50.6
LH123Ht	**CML341**	**G21**	46.6	48.3	49.0		48.0	48.0	48.6		48.6	47.3	47.7	51.4
LH123Ht	**CML373**	**G22**	36.8	41.1	41.4		44.3	46.9	44.6	45.7	44.2	44.9	43.8	42.5
LH123Ht	**Tzi8**	**G23**	47.6	48.4	48.2		48.2		47.9	47.9	47.3	47.6	46.4	44.9
LH123Ht	**Tzi9**	**G24**	43.0		43.5		46.8	45.5	45.5	45.5	45.5	45.5	45.5	48.7

Table 3.4 Continued

(B) Chromosome 8, ZmPR2

		cM position	**56.9**	**57.8**	**62.1**	**62.1**	**67.3**	**67.3**	**68.2**	**68.4**		**70.8**	**77.5**
RP	**DP**	**cross code**	**M8_1**	**M8_2**	**M8_3**	**M8_4**	**M8_6**	**M8_7**	**M8_8**	**M8_9**		**M8_10**	**M8_11**
2369	**CML10**	**G11**	48.4	50.0	49.5	49.2	52.1		52.1	52.1	**QTL peak at 69 cM (AGP2 130.7 Mbp)**		54.0
2369	**CML277**	**G13**	52.7	52.7	51.8		51.6		52.7	52.4		50.5	
2369	**CML341**	**G14**	**28.7**		**27.7**	**26.9**	**55.7**		**28.3**	**29.5**		**56.0**	
2369	**CML373**	**G15**	47.9	49.4	49.2		48.9		47.9	47.6			47.0
2369	**Tzi8**	**G16**	48.7	49.2	49.2		50.0	49.2	50.3	50.5			51.3
2369	**Tzi9**	**G17**	**44.8**	47.1	46.7		47.9	47.6	47.3	47.0			49.7
LH123Ht	**CML10**	**G18**	**33.5**	**36.8**	**37.1**	**36.5**	**38.4**		**38.4**	**38.1**			**39.8**
LH123Ht	**CML258**	**G19**	45.8	45.5	**44.8**		**44.3**		**44.8**	45.3			46.9
LH123Ht	**CML277**	**G20**	**57.0**	53.4	49.1	46.1	45.9			49.1			53.6
LH123Ht	**CML341**	**G21**	**39.1**	**40.3**	**40.7**	**37.6**	**40.4**		**41.3**	**41.3**			**41.7**
LH123Ht	**CML373**	**G22**	**41.6**	**40.2**	**39.7**	**36.8**	**44.0**		45.4	45.4			**44.0**
LH123Ht	**Tzi8**	**G23**	**36.5**	**38.4**	**40.6**		**64.6**		**68.2**				**72.4**
LH123Ht	**Tzi9**	**G24**	48.1	49.3	47.7		47.7	47.7	47.1	46.5			**44.5**

Table 3.4 Continued

(C) Chromosome 9, ZmPR3

		cM position	34.2	39.6	40.9	41.3	43.8	44	44.7	45.1		45.8	46.3	46.3	46.3
RP	**DP**	**cross code**	**M9_2**	**M9_3**	**M9_4**	**M9_5**	**M9_8**	**M9_9**	**M9_11**	**M9_12**		**M9_14**	**M9_15**	**M9_16**	**M9_17**
2369	**CML10**	**G11**	**45.0**	46.6	47.1	46.0	46.6	46.6	46.6	46.6	**QTL peak at 45.2 cM (AGP2 42.8 Mbp)**	46.0	46.0	46.0	47.8
2369	**CML277**	**G13**	51.9	53.7	52.7	53.2	53.9	54.5	54.0	53.7		54.5	54.5	54.2	**56.3**
2369	**CML341**	**G14**	**24.1**	**24.0**	**25.4**	**24.0**	**24.5**	**24.0**	**23.2**	**23.2**		**23.9**	**24.0**	**23.5**	**25.5**
2369	**CML373**	**G15**	47.9	48.5		49.2	48.9	48.9	49.2	48.9			49.5	50.6	51.5
2369	**Tzi8**	**G16**	49.0	48.2	47.6	47.6	48.2	47.6	47.9			47.9	47.9	47.6	48.7
2369	**Tzi9**	**G17**	51.6	53.5	**55.4**	**55.2**	52.5	52.5	53.3	52.7		53.6	53.3	53.3	**55.1**
LH123Ht	**CML10**	**G18**	**37.8**	**36.4**	**37.3**	**37.3**	**32.7**	**32.7**	**26.9**	**27.3**		**27.3**	**27.1**	**26.9**	**29.1**
LH123Ht	**CML258**	**G19**	**38.7**	**39.3**	**39.8**	**39.8**	**39.3**	**39.3**	**39.3**	**39.3**		**39.5**	**39.3**	**39.3**	**40.3**
LH123Ht	**CML277**	**G20**	53.6	52.1	51.8	50.3	48.8	48.5	49.1	49.1			49.1	49.4	50.3
LH123Ht	**CML341**	**G21**	46.4	**44.6**	**44.6**	**44.6**	45.7	45.7	46.0	46.0		46.3	46.0	46.0	46.9
LH123Ht	**CML373**	**G22**	45.4	**40.2**	**39.7**	**40.0**	**43.7**	**44.2**	**44.0**	**44.0**		**44.6**	**44.0**	**41.3**	**44.6**
LH123Ht	**Tzi8**	**G23**	**41.3**	**42.9**	**42.7**	**41.7**	**38.8**	**38.8**	**15.2**	**15.2**			**22.9**	**23.2**	**24.8**
LH123Ht	**Tzi9**	**G24**	46.2	45.5	46.2	45.5	**37.1**	**37.1**	45.5	45.5		45.8	45.5	45.5	48.3

Table 3.4 Continued

(D) Chromosome 10, ZmPR4

		cM position	**31.7**	**35**	**35.6**	**36.1**	**37.1**	**39.1**	**41.2**		**42.9**	**43**	**43.6**	**43.8**	**45.4**	**45.7**	**53.1**	**55**
RP	**DP**	**cross code**	**M10_1**	**M10_2**	**M10_3**	**M10_5**	**M10_7**	**M10_9**	**M10_12**		**M10_13**	**M10_14**	**M10_16**	**M10_17**	**M10_18**	**M10_19**	**M10_20**	**M10_21**
2369	**CML10**	**G11**	52.1		52.7		53.2	52.7	52.7	**QTL peak at 42.9 cM (AGP2 98.7 Mbp)**	51.9	51.6	51.6	52.1	52.2	52.4	52.7	52.9
2369	**CML277**	**G13**	**65.4**		52.9		54.0	53.4	52.9			54.0	53.7	53.9	52.6	52.6	**55.3**	54.1
2369	**CML341**	**G14**	**27.9**		**44.8**		53.9	50.8	**59.7**		**30.4**	**27.3**	**27.3**	**25.5**	53.1	52.1	**25.0**	**24.7**
2369	**CML373**	**G15**	**39.2**		**41.0**		**40.7**	**43.2**	47.0		45.8	**41.3**	**41.3**	**43.2**	**40.4**	**40.7**	**40.7**	**40.7**
2369	**Tzi8**	**G16**	49.7		48.4		49.0	49.5	48.4			47.4	47.9	48.2	47.9	46.8	47.4	47.1
2369	**Tzi9**	**G17**	46.2		**44.9**		47.9	47.6	48.1		46.2	46.2	46.2	46.2	48.9	49.5	47.3	
LH123Ht	**CML10**	**G18**		**37.3**	**38.5**	**38.5**		**37.9**	**38.7**			**38.5**	**39.4**	**39.4**	**38.6**	**38.4**	**35.6**	**33.5**
LH123Ht	**CML258**	**G19**		**55.2**	**55.7**	**55.7**	**55.2**	54.7	53.1			53.1	52.6	52.6	52.1	51.6	53.1	53.1
LH123Ht	**CML277**	**G20**		46.1	46.1	45.5	46.1	47.3	46.1			**43.1**	**42.5**	**42.5**	**42.8**	**43.4**	**44.9**	**43.8**
LH123Ht	**CML341**	**G21**		**40.5**	**39.9**		**40.5**	**41.8**	**43.4**			**44.1**	45.3	45.3	45.9	45.3	**43.9**	**42.8**
LH123Ht	**CML373**	**G22**		47.6	47.6	48.0	**56.2**	**43.4**	45.1			**43.0**	**55.5**	**55.2**	53.4	54.1	46.2	**58.6**
LH123Ht	**Tzi8**	**G23**		48.5	48.5		48.5	47.4	**55.3**			49.1	49.1	47.4	49.1	49.1	50.3	49.4
LH123Ht	**Tzi9**	**G24**		52.9	54.2		54.2	54.2	54.2			54.8	54.2	54.2	53.5	53.5	**55.5**	

3.3.2 Recombination analysis

Observed recombinant haplotypes (unique sequences of recurrent and donor parent genotypes, Supplementary Files 3.2 to 3.5) were used to identify the recombination breakpoints and to evaluate the spatial distribution of recombinations across the four ZmPR QTLs. In nine of the 52 BC_1 families (13 BC_1 populations x 4 ZmPR QTLs; no data for the cross G12) there was no ambiguity in resolving if and how many recombinations occurred in the ZmPR regions (%ambiguous = 0 in Table 3.5, Supplementary files 3.2 to 3.5). Among the remaining 43 BC_1 families, 23 had less than 5% individuals with ambiguous haplotypes. The 20 BC_1 families with more than 5% ambiguous haplotypes (5.1% to 27.1%, shown in bold font in Table 3.5) were associated with cross x QTL combinations having high percentage of DP/DP homozygotes or missing observations (Table 3.3).

The numbers of recombinations ranged from 0 to 6 in the ZmPR1, ZmPR3 and ZmPR4 QTL regions on chromosomes 1, 9 and 10 respectively, and from 0 to 5 in the ZmPR2 QTL region on chromosome 8 (Table 3.5). As expected, for each BC_1 family, the frequencies of individuals with no recombination (x=0) in the ZmPR regions was the highest with a mean of 0.816 (SD=0.097, min=0.538, max=0.94), followed by a mean frequency of 0.141 (SD=0.066, min=0.042, max=0.422) of single recombinations (x=1, Table 3.5). Frequencies of double (x=2) and more (x > 2) simultaneous recombinations within the same ZmPR region were skewed towards zero with means 0.018 (SD 0.03) and 0.007 (SD=0.032) respectively.

Table 3.5 Numbers of recombinations identified in the four ZmPR QTL regions and their frequencies. Multiple recombinations (x ≥ 2) observed with zero frequency are represented using the symbol "-". The %ambiguous individuals refer to the frequency of individuals for which it was not possible to determine if recombinations occurred, typically individuals with more missing values or DP genotypes.

chromosome 1; ZmPR1

cross	total ind	resolved	%ambi-guous	Frequency of the number of recombinations x						
				x = 0	x = 1	x = 2	x = 3	x = 4	x = 5	x = 6
G11	192	187	2.6	0.813	0.160	0.016	0.011	-	-	-
G13	192	188	2.1	0.793	0.197	0.011	-	-	-	-
G14	192	171	**10.9**	0.556	0.205	0.006	0.018	0.012	0.199	0.006
G15	192	188	2.1	0.846	0.144	0.011	-	-	-	-
G16	192	192	0.0	0.792	0.208	-	-	-	-	-
G17	192	187	2.6	0.743	0.225	0.016	-	-	0.016	-
G18	192	158	**17.7**	0.829	0.158	0.006	-	-	-	0.006
G19	192	192	0.0	0.844	0.156	-	-	-	-	-
G20	192	171	**10.9**	0.807	0.170	0.006	0.006	0.012	-	-
G21	156	148	**5.1**	0.865	0.135	-	-	-	-	-
G22	156	128	**17.9**	0.797	0.117	0.008	0.023	0.008	0.008	0.039
G23	192	185	3.6	0.832	0.108	0.059	-	-	-	-
G24	156	154	1.3	0.831	0.162	0.006	-	-	-	-

chromosome 8; ZmPR2

cross	total ind	resolved	%ambi-guous	Frequency of the number of recombinations x					
				x = 0	x = 1	x = 2	x = 3	x = 4	x = 5
G11	192	191	0.5	0.827	0.168	0.005	-	-	-
G13	192	189	1.6	0.794	0.206	-	-	-	-
G14	192	171	**10.9**	0.538	0.199	0.012	0.251	-	-
G15	192	166	**13.5**	0.916	0.084	-	-	-	-
G16	192	192	0.0	0.922	0.073	0.005	-	-	-
G17	192	185	3.6	0.816	0.157	0.011	-	0.011	0.005
G18	192	152	**20.8**	0.842	0.151	0.007	-	-	-
G19	192	192	0.0	0.802	0.198	-	-	-	-
G20	168	165	1.8	0.697	0.291	0.012	-	-	-
G21	156	151	3.2	0.828	0.146	0.026	-	-	-
G22	156	124	**20.5**	0.742	0.161	0.097	-	-	-
G23	192	192	0.0	0.568	0.422	0.010	-	-	-
G24	156	156	0.0	0.808	0.186	0.006	-	-	-

Table 3.5 Continued

chromosome 9; ZmPR3

cross	**total ind**	**resolved**	**% ambiguous**	**Frequency of the number of recombinations x**						
				x = 0	**x = 1**	**x = 2**	**x = 3**	**x = 4**	**x = 5**	**x = 6**
G11	192	189	1.6	0.905	0.090	-	0.005	-	-	-
G13	192	189	1.6	0.910	0.085	0.005	-	-	-	-
G14	192	168	**12.5**	0.940	0.042	-	0.018	-	-	-
G15	192	166	**13.5**	0.911	0.089	-	-	-	-	-
G16	192	192	0.0	0.880	0.104	0.010	0.005	-	-	-
G17	192	184	4.2	0.913	0.082	-	0.005	-	-	-
G18	192	140	**27.1**	0.750	0.221	0.029	-	-	-	-
G19	192	191	0.5	0.911	0.089	-	-	-	-	-
G20	168	167	0.6	0.928	0.060	-	0.012	-	-	-
G21	156	147	**5.8**	0.912	0.088	-	-	-	-	-
G22	156	115	**26.3**	0.722	0.139	0.122	-	0.009	-	0.009
G23	192	174	**9.4**	0.621	0.190	0.103	0.057	0.011	0.017	-
G24	156	154	1.3	0.831	0.065	0.104	-	-	-	-

chromosome 10; ZmPR4

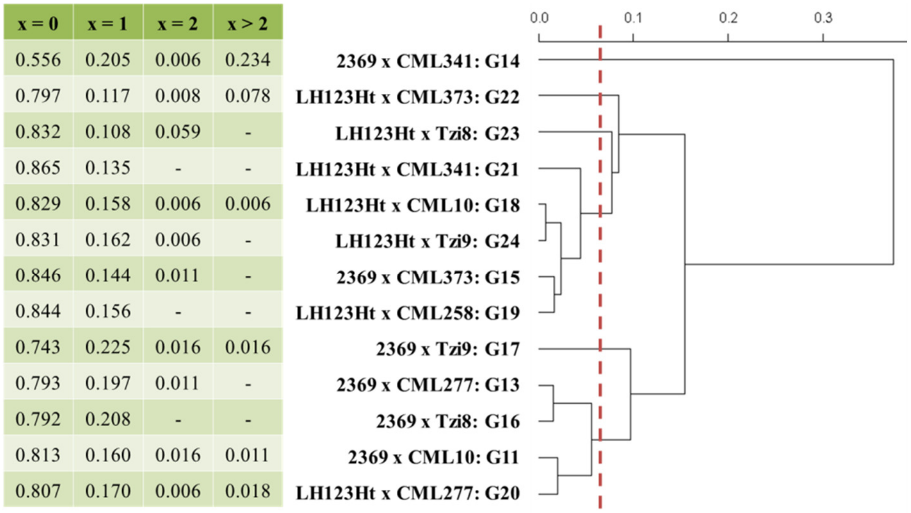
cross	**total ind**	**resolved**	**% ambiguous**	**Frequency of the number of recombinations x**						
				x = 0	**x = 1**	**x = 2**	**x = 3**	**x = 4**	**x = 5**	**x = 6**
G11	192	188	2.1	0.878	0.117	0.005	-	-	-	-
G13	192	190	1.0	0.879	0.105	0.016	-	-	-	-
G14	192	172	**10.4**	0.640	0.058	0.017	-	0.285	-	-
G15	192	176	**8.3**	0.869	0.074	0.057	-	-	-	-
G16	192	192	0.0	0.901	0.089	0.010	-	-	-	-
G17	192	186	3.1	0.876	0.097	-	-	0.027	-	-
G18	192	168	**12.5**	0.857	0.137	-	-	-	0.006	-
G19	192	192	0.0	0.885	0.115	-	-	-	-	-
G20	168	167	0.6	0.844	0.138	0.012	0.006	-	-	-
G21	156	145	**7.1**	0.869	0.117	-	0.007	0.007	-	-
G22	156	122	**21.8**	0.689	0.180	0.025	0.057	0.049	-	-
G23	192	187	2.6	0.770	0.070	0.086	0.005	0.011	-	0.059
G24	156	155	0.6	0.877	0.116	0.006	-	-	-	-

Double recombinations were found in 35 of the 52 BC_1 families with frequencies ranging from .005 to 0.122. Multiple recombinations (x > 2) within the same QTL region were found in 22 of the 52 BC_1 families. In case of multiple (x > 2) simultaneous recombinations, seven of the 22 BC_1 families associated with crosses G14 and G22 for the ZmPR1, G14 for ZmPR2, G23 for ZmPR3, and G14, G22 and G23 for ZmPR4 showed relatively higher and potentially erroneous frequencies (0.075 to 0.285, >5%, Table 3.5) when compared to the remaining 15 families with positive multiple recombination frequencies (0.005 to 0.027, Table 3.5).

3.3.2.1 Overall recombination rates in each QTL region

The numbers of recombinations and the rates at which they occurred in the four ZmPR QTLs showed a pattern with certain parents, but not in all the ZmPR regions. Hierarchical clustering (using the function 'hclust' in R), based on a similarity (or distance) cut-off value of 0.6 for each ZmPR QTL, revealed six, seven, six, and six groups of BC_1 families with varying recombination rates between groups (Figure 3.1 to Figure 3.4). The number of BC_1 families in any group was from a minimum of one (singleton) to a maximum of seven.

Four singleton BC_1 families each were found in groupings for ZmPR1 (G14, G17, G22, G23), ZmPR2 (G14, G20, G22, G23), ZmPR3 (G18, G22, G23, G24) and ZmPR4 (G14, G15, G22, G23) QTL regions. The distribution of recombination events for these singletons was uniquely different and typically consisted of a higher proportion of individuals with one and/or more recombinations when compared to other groups at the same ZmPR QTL (Figure 3.1 to Figure 3.4).



x = 0	x = 1	x = 2	x > 2	
0.556	0.205	0.006	0.234	2369 x CML341: G14
0.797	0.117	0.008	0.078	LH123Ht x CML373: G22
0.832	0.108	0.059	-	LH123Ht x Tzi8: G23
0.865	0.135	-	-	LH123Ht x CML341: G21
0.829	0.158	0.006	0.006	LH123Ht x CML10: G18
0.831	0.162	0.006	-	LH123Ht x Tzi9: G24
0.846	0.144	0.011	-	2369 x CML373: G15
0.844	0.156	-	-	LH123Ht x CML258: G19
0.743	0.225	0.016	0.016	2369 x Tzi9: G17
0.793	0.197	0.011	-	2369 x CML277: G13
0.792	0.208	-	-	2369 x Tzi8: G16
0.813	0.160	0.016	0.011	2369 x CML10: G11
0.807	0.170	0.006	0.018	LH123Ht x CML277: G20

Figure 3.1 Grouping of the 13 BC_1 families based on numbers of recombinations (x) and their frequencies in the ZmPR1 QTL region on the maize chromosome 1. The red bar represents the cut-off point (0.6) at which groups were formed (the bar is placed at a value slightly more than 0.6 for clarity in observing the lines of the dendrogram lines).

In the ZmPR1 QTL region, the two groups with multiple (> 1) BC_1 families were dominated by the presence of each of the two temperate recurrent parent lines 2369 and LH123Ht (Figure 3.1). The group dominated by 2369 as RP appeared to have relatively more recombinations in the ZmPR1 region (mean non-recombinant frequency = 0.199 with SD = 0.01) when compared to the group dominated by the RP LH123Ht (mean non-recombinant frequency = 0.157 with SD = 0.014). In particular, BC_1 families with CML10 (G11 vs G18), CML341 (G14 vs G21) and Tzi9 (G17 vs G24) as donor parents were more recombinogenic when 2369 was used as RP. In contrast, families with CML373 (G15 vs G22) and Tzi8 (G16 vs G23) were more recombinogenic when LH123Ht was used as RP. The families with CML277 (G13 and G20) did not show any specific preference for more or less recombinations with any RP and appeared together in the 2369-dominated group (Figure 3.1).

x = 0	x = 1	x = 2	x > 2
0.916	0.084	-	-
0.922	0.073	0.005	-
0.742	0.161	0.097	-
0.808	0.186	0.006	-
0.794	0.206	-	-
0.802	0.198	-	-
0.827	0.168	0.005	-
0.816	0.157	0.011	0.016
0.842	0.151	0.007	-
0.828	0.146	0.026	-
0.538	0.199	0.012	0.251
0.697	0.291	0.012	-
0.568	0.422	0.010	-

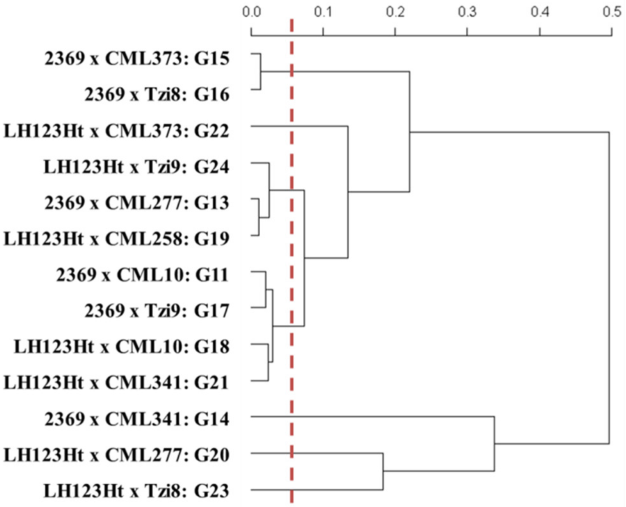


Figure 3.2 Grouping of the 13 BC_1 families based on numbers of recombinations (x) and their frequencies in the ZmPR2 QTL region on the maize chromosome 8. The red bar represents the cut-off point (0.6) at which groups were formed.

In the ZmPR2 QTL region, the group {G15, G16: 2369 x CML373 and 2369 x Tzi8 respectively} consisted of a very low frequency of recombinant BC_1 individuals (mean recombinant frequency = 0.081 with SD = 0.004). The BC_1 families G22 and G23 with recurrent LH123Ht and the same donor parents CML373 and Tzi8 respectively, however, were highly recombinogenic (singletons, Figure 3.2). Among the other two groups in the ZmPR2 region {G24, G13, G19} and {G11, G17, G18, G21}, the former had reatively more single recombination individuals while the latter had more of multiple recombinant individuals. Unlike the pattern observed in ZmPR1 region, families with CML10 as DP had the same level of recombinations irrespective of the RP, and families with CML277 were more recombinogenic when LH123Ht was used as RP (Figure 3.2). CML341, however, showed a pattern as observed in ZmPR1 QTL, that of more recombinations when 2369 was the RP.

x = 0	x = 1	x = 2	x > 2	
0.831	0.065	0.104	-	LH123Ht x Tzi9: G24
0.940	0.042	-	0.018	2369 x CML341: G14
0.928	0.060	-	0.012	LH123Ht x CML277: G20
0.880	0.104	0.010	0.005	2369 x Tzi8: G16
0.905	0.090	-	0.005	2369 x CML10: G11
0.913	0.082	-	0.005	2369 x Tzi9: G17
0.910	0.085	0.005	-	2369 x CML277: G13
0.912	0.088	-	-	LH123Ht x CML341: G21
0.911	0.089	-	-	2369 x CML373: G15
0.911	0.089	-	-	LH123Ht x CML258: G19
0.621	0.190	0.103	0.086	LH123Ht x Tzi8: G23
0.750	0.221	0.029	-	LH123Ht x CML10: G18
0.722	0.139	0.122	0.017	LH123Ht x CML373: G22

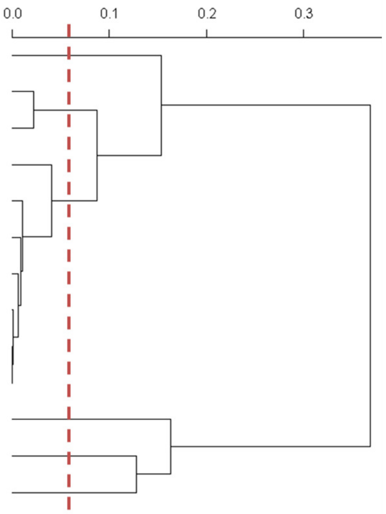


Figure 3.3 Grouping of the 13 BC$_1$ families based on numbers of recombinations (x) and their frequencies in the ZmPR3 QTL region on the maize chromosome 9. The red bar represents the cut-off point (0.6) at which groups were formed.

In the ZmPR3 QTL region, both the groups with multiple BC$_1$ families {G14, G20} and {G16, G11, G17, G13, G21, G15, G19} showed a high proportion of individuals with no recombinations (mean non-recombinant frequency > 0.9). Unlike the pattern of recombination rates in both ZmPR1 and ZmPR2, proportion of recombinant BC$_1$ individuals with CML10 as DP was more when LH123Ht was the RP as opposed to 2369. Families with CML341 (G14 vs G21), CML373 (G15 vs G22), Tzi8 (G16 vs G23) and Tzi9 (G17 vs G24) as DP were more recombinogenic with LH123Ht as RP (Figure 3.3).

x = 0	x = 1	x = 2	x > 2
0.640	0.058	0.017	0.285
0.689	0.180	0.025	0.107
0.770	0.070	0.086	0.075
0.857	0.137	-	0.006
0.844	0.138	0.012	0.006
0.869	0.074	0.057	-
0.901	0.089	0.010	-
0.876	0.097	-	0.027
0.879	0.105	0.016	-
0.869	0.117	-	0.014
0.885	0.115	-	-
0.878	0.117	0.005	-
0.877	0.116	0.006	-

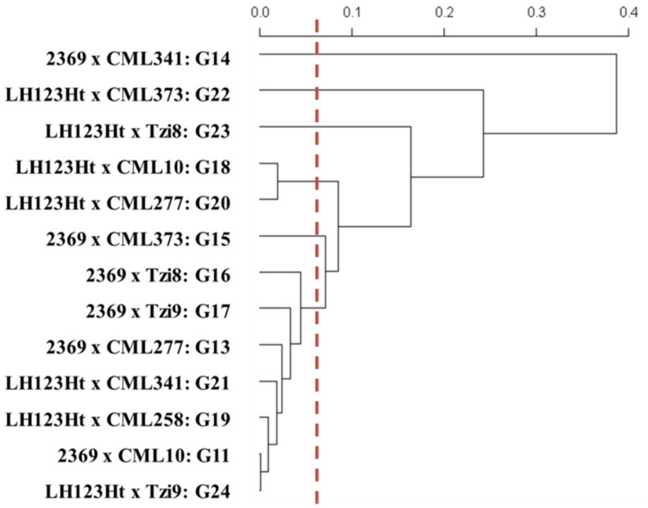


Figure 3.4 Grouping of the 13 BC_1 families based on numbers of recombinations (x) and their frequencies in the ZmPR4 QTL region on the maize chromosome 10. The red bar represents the cut-off point (0.6) at which groups were formed.

In the ZmPR4 QTL region, two groups with multiple BC_1 families were found. The largest group {G16, G17, G13, G21, G19, G11, G24} comprised of families with a relatively lower proportion of recombinant BC_1 individuals (mean recombinant frequency = 0.119 with SD = 0.01). Families with CML10 (G11 vs G18), CML277 (G13 vs G20), CML373 (G15 vs G22) and Tzi8 (G16 vs G23) were more recombinogenic when LH123Ht was used as RP. Families with CML341 (G14 vs G21) appeared to be relatively more recombinogenic when 2369 was used as RP whereas families with Tzi9

(G17 and G24) had similar recombination rates in the ZmPR4 QTL irrespective of the RP (Figure 3.4).

None of the BC_1 families appeared together in the same group across all the four ZmPR QTLs. However, there were families that appeared in the same group at three ZmPR QTLs. The families {G11, G13, G16} and {G19, G21} showed similar recombination patterns within the ZmPR1, ZmPR3 and ZmPR4 QTLs, the families {G11, G17, G21} and {G13, G19} showed similar recombination patterns within the ZmPR2, ZmPR3 and ZmPR4 QTL, and {G19, G24} in ZmPR1, ZmPR2 and ZmPR4 QTLs.

Observed recombination rates were also expressed in terms of cM/Mbp. The unlikely large number of multiple simultaneous recombinations (Table 3.5) resulted in hugely inflated linkage distances in three ZmPR regions each for the BC_1 families G14, G22 and G23 (shown in bold in Table 3.6).

The ZmPR2 region appeared to be more recombinogenic with higher average recombination rates (cM/Mbp) which were also more variable across the BC_1 families (0.38 ± 0.148 cM/Mbp, Table 3.6). The ZmPR4 region, in contrast, seemed to be less recombinogenic (0.127 ± 0.038 cM/Mbp, Table 3.6) among the four ZmPR regions.

3.3.2.2 Spatial distribution of recombinations in the ZmPR 1-4 regions

In nine BC_1 families (G14 in ZmPR1, ZmPR2 and ZmPR4; G22 in ZmPR1, ZmPR3 and ZmPR4; and G23 in ZmPR2, ZmPR3 and ZmPR4), highly unlikely numbers of recombinant BC_1 individuals and/or numbers of multiple simultaneous recombinations within a ZmPR region were found (Table 3.7). Barring these families, on average, there were 37 (SD = 11), 33 (SD = 12), 21 (SD = 10) and 26 (SD = 6) recombination events in the ZmPR1-4 regions respectively in each BC_1 family (based on Table 3.7).

Table 3.6 Average recombination rate (cM/Mbp) in the ZmPR regions. Estimates of the observed genetic lengths (cM) can be found in the Supplementary File 3.6.

Cross	**Observed genetic length (cM)**				**Average recombination rate in the ZmPR region (cM/Mbp)**			
	ZmPR 1	**ZmPR 2**	**ZmPR 3**	**ZmPR 4**	**ZmPR 1**	**ZmPR 2**	**ZmPR 3**	**ZmPR 4**
NAM[1]	24.3	20.6	24	23	0.204	0.402	0.356	0.2
G11	23.4	17.5	11.6	11.6	0.197	0.341	0.172	0.101
G13	22.2	20.3	9.0	14.6	0.187	0.534	0.134	0.127
G14	**154.6**	**114.1**	10.9	**131.8**	1.301	2.993	0.162	1.148
G15	16.9	8.3	8.5	10.5	0.142	0.162	0.126	0.091
G16	21.1	7.9	14.0	7.7	0.178	0.154	0.208	0.067
G17	34.1	23.0	9.6	20.8	0.287	0.449	0.143	0.181
G18	20.0	14.9	26.2	17.2	0.168	0.290	0.390	0.164
G19	15.7	19.4	9.0	11.4	0.132	0.378	0.134	0.109
G20	22.0	27.3	8.8	18.0	0.185	0.532	0.131	0.172
G21	13.6	18.1	8.8	17.2	0.115	0.353	0.131	0.150
G22	**47.2**	32.0	**46.7**	**102.1**	0.397	0.625	0.694	0.974
G23	22.3	**43.6**	**72.8**	**62.0**	0.188	0.850	1.082	0.539
G24	17.6	18.5	30.1	12.8	0.148	0.360	0.447	0.112
mean*					0.175	0.380	0.198	0.127
SD*					0.045	0.148	0.112	0.038

[1] Based on Table B1

* computed after excluding the cM/Mbp values corresponding to the **bold** cM values in G14, G22 and G23

In the ZmPR1 region, two marker intervals, M1_4 to M1_5 and M1_11 to M1_14, individually accounted for more than one-fifth of recombinations in each BC_1 family (Figure 3.6).

Table 3.7 Numbers of recombinant individuals (N.IND) and recombinations (N.REC) found in 13 BC_1 families across four ZmPR QTL regions. The numbers in bold font represent unlikely numbers of N.IND and N.REC.

	ZmPR1		**ZmPR2**		**ZmPR3**		**ZmPR4**	
Cross	**N.IND**	**N.REC**	**N.IND**	**N.REC**	**N.IND**	**N.REC**	**N.IND**	**N.REC**
G11	35	42	33	34	18	22	23	23
G13	39	41	39	39	17	17	23	26
G14	**76**	**232**	**79**	**167**	10	10	**62**	**116**
G15	29	31	14	14	15	15	23	27
G16	40	40	15	16	23	27	19	21
G17	48	63	34	45	16	18	23	39
G18	27	33	24	25	35	35	24	24
G19	30	30	38	38	17	17	22	22
G20	33	42	50	52	12	16	26	30
G21	20	20	26	30	13	13	19	24
G22	26	**64**	32	41	32	**53**	38	**73**
G23	31	40	**83**	**84**	**66**	**118**	43	**110**
G24	26	24	30	30	26	42	19	20

The interval M1_4 to M1_5, in particular, harbored the photoperiod QTL peak. However, these two marker intervals were wide (20.04% for M1_4 to M1_5 and 59.84% for M1_11 to M1_14 in the ZmPR1 region evaluated, 118.8 Mbp from M1_1 to M1_14). Recombination rates ranging, on average, from 0.222 cM/Mbp (SD = 0.215) to 0.385 cM/Mbp (SD = 0.197) were observed along the marker intervals till the marker M1_6 beyond which the recombination rate was very low (0 cM/Mbp to 0.091 cM/Mbp with SD = 0.035, Supplementary File 3.6). Two marker intervals M1_7 to M1_8 and M1_10 to M1_11 were very small, 1.16E-04 Mbp and 0.001 Mbp respectively, for which extremely high (>1000) cM/Mbp values were observed. These high and likely inaccurate estimates of recombination rates (> 1 cM/Mbp and in a few cases >1000) for some marker intervals were also found in ZmPR2-4 regions too for the BC_1 families G14, G22 and G23 (shown in bold in Table 3.7) and a few other crosses in which recombination events occurred in smaller intervals (<1 Mbp in particular, or a few Mbp). These estimates were believed to possibly be artifacts and hence the corresponding marker intervals were not necessarily recombination hotspots.

In the ZmPR2 region, there seemed to be a lower recombination (0 cM/Mbp) for most of the BC_1 families between markers M8_6 to M8_9. All the marker intervals showed a wide range of recombination rates across families, suggesting that the recombination is distributed throughout the ZmPR2 region without any preferential location to resolve.

For the ZmPR3 region, the QTL peak was located in the marker interval M9_12 to M9_14 where the amount of recombinations ranged from 0% to 7.7% (mean = 2.39% with SD = 3.23%, based on counting recombinations in the Supplementary File 3.4). The marker interval M9_1 to M9_5, on the other hand accounted for 44.1 % to 100% of the recombinations across all BC_1 families (mean = 80% with SD = 15.6%, based on counting recombinations in the Supplementary File 3.4). In terms of cM/Mbp, recombination rates were higher (1.238 to 2.798 cM/Mbp) in the region between the markers M9_2 to M9_5 which is at least two orders of magnitude higher than the maize genome average recombination rate of 0.68 cM/Mbp (computed using total genetic map length from Table B1 and maize genome length in Mbp from AGPv2 sequence).

For the ZmPR4 region, all the marker intervals typically showed a recombination rate less than the genome average of 0.68 cM/Mbp. Recombinations in the ZmPR4 region seemed to be spread throughout the region without any preference for a specific intragenic location.

4 SUMMARY AND DISCUSSION

This first study differs from previous simulation studies on developing introgression libraries in the 1) lengths of target region and of desired donor introgressions, and 2) selection strategies investigated. Importantly, a new selection strategy is proposed and its' efficiency is investigated with varying sample sizes in backcross breeding programs to develop QTL-targeted tiled introgressions.

4.1 Lengths of target region and of desired donor introgressions

In contrast to the previous simulation studies for development of genomic libraries with desired donor segment lengths of 20-40 cM (Sušiƈ 2005, Falke et al. 2009, Herzog et al. 2014) across the entire genome, the focus of this study was on target QTLs of 15 cM and in producing sets of individuals with introgressions tiled across this 15 cM target region. The desired average introgression lengths for the tiles evaluated in this study were 3, 1.5 and 1 cM, much less than the lengths evaluated previously. It was revealed by Sušiƈ (2005) that decrease in donor chromosome segment length from 40 to 20 cM substantially increased the total number of individuals required to reach a pre-defined recurrent parent genome threshold for BC strategies with 2 and 3 generations. Subsequently, the optimal strategies for producing QTL targeted NILs would be different.

In producing genomic libraries, the criteria for evaluating different breeding strategies to achieve complete donor coverage in the introgression lines consisted of optimizing for 1) the number of marker data points and BC generations required to reach pre-specified recurrent parent coverage thresholds (95.6% and 92.83% for introgressions of 20 cM and 40 cM respectively; Sušicˇ 2005), 2) optimizing for high donor allele coverage in target segments with minimum overlap outside of the target segments on the target chromosome and a low total donor genome proportion on non-target chromosomes within three BC generations (with an objective of producing 100 introgression lines with 20 cM donor segments on average; Herzog et al. 2014), or 3) attaching a success criterion of expected number of introgression lines within an introgression library carrying donor alleles at markers outside the target segments to be smaller than 1 (Falke et al. 2009) and more. In all these studies, the numbers of BC generations were typically fixed to up to three, and the evaluation/success criteria were appropriate in the context of optimizing for low input resources (like the progeny sizes and marker data points) and high gains in the genomic composition of the resulting genomic introgression library. In producing QTL-targeted introgression libraries, however, more than three BC generations were required. Through repeated backcrossing removal of donor parent segments on the non-target chromosomes that are unlinked to the target segment of interest (through segregation and independent assortment) and removal of donor parent segments linked to the target segment (through recombination) are both simultaneously accomplished. While the former can be simple, the latter definitely would not be, especially in producing individuals with smaller donor introgressions as evidenced in faster rates of increase of %RPZ and %RPN when compared to %RPC (Figure A6 to Figure A8) which clearly emphasize the importance of elimination of linkage drag as the dominant problem in backcrossing programs.

Therefore, achieving a preset average introgression length was considered success for a breeding strategy. Furthermore, since chances of future observations for MIL of up to 3.5 cM, 2 cM and 1.5 cM were of interest rather than the mean level, the 95% percentile intervals were used to determine if a breeding strategy was successful. The percentile intervals represented, along with the sampling error, the variability in the genomic composition (or sizes) of donor introgressions of the selected set of individuals in each BC generation which decreased (and hence the gap between the upper and lower limits reduced, Figure 2.7) in the process of select-backcross cycles.

4.2 Selection strategies

Three variants of the distance based strategy (FMW0, FMW2 and FMW5) were compared in this study, which have not been investigated before. The three stage selection method in Falke et al. (2009) was used as a no distance strategy (NODIST). For all the selection strategies, with the simulated marker data, if multiple individuals appeared to have the same recombination breakpoints after the foreground selection (since marker positions were discrete), background selection was performed at three levels: maximizing proportion of homozygous RP alleles in 1) the non-targeted region of the target chromosome (%RPC), 2) the other three ZmPR regions (%RPZ), and 3) in the remaining non-target genome (%RPN).

The strategies FMW2 and FMW5 were equivalent in magnitude of average introgressions lengths of the selected sets of individuals for the same breeding strategies at each BC generation which suggested that multiple flanking markers might be redundant (Table 2.4). Failure of the strategy FMW0 in producing sets of desired NILs across the target QTL on chromosome 1, and at significantly later BC generations with high progeny sizes for chromosome 10 when compared to other selection strategies (Table 2.4A-C) reiterates the importance of flanking marker windows in MAS and MAB studies, and of the choice of a "good" selection strategy. The average introgressions achieved by FMW0 were always less when compared to FMW2 except in the BC_1 generation.

The NODIST strategy was close to FMW2 and FMW5 for the breeding strategies with high number of backcrosses per selected individual (N_{BSI}=100, Table 2.4) when there was a relatively higher probability of finding recombinations close to the target region (due to higher number of individuals). A new selection strategy that can identify the best candidates early in the breeding process and contributes to its' success with limited resources is a significant step towards an enhanced understanding of how to design efficient backcross introgression schemes to produce sets of individuals to test loci within a QTL region. To benefit by two BC generations, for crops that have a longer growing season, would be significant.

The position score s_{uk} of distance-based strategy (Section 2.2.1, Figure 2.4) is a variable to finely select individuals with tiled NILs. For instance, the NODIST strategy would preselect individuals in which the donor allele stretch (l_{uk}, r_{uk}) completely covered the ideotypic target (L_i, R_i) thus probably missing out on the "chance" individuals with desired double recombinations (accumulated over multiple BC generations) within this ideotypic target segment. In this study, s_{uk} was equally weighted and assigned a score of 1 in both the cases – when a contiguous stretch of donor alleles (l_{uk}, r_{uk}) completely covered the ideotypic target (L_i, R_i) and also when (l_{uk}, r_{uk}) was within the segment (L_i, R_i). However, s_{uk} could be assigned differently to differentiate between these two kinds of donor allele stretches in individuals. With empirical data sets, a useful approach might be to produce multiple sets of individuals in each generation by changing the position score and also the number of individuals to be selected (which affects the length of ideotypic target). Generating multiple sets of individuals might provide an opportunity to better evaluate the individuals to be included for advancing. The proposed selection strategy may be evaluated with variations in marker density which in this study was kept constant for all the breeding schemes. In particular, for constructing NILs with relatively larger desired introgression lengths (for instance 3 cM when compared to 1.5 or 1 cM) the density of markers could be reduced in the target region.

4.3 The BC$_1$ generation

Although increase in BC$_1$ progeny size (N$_{BC1}$) decreased the average introgression lengths of the selected sets of individuals, increasing N$_{BC1}$ beyond ~150 individuals was less helpful in improving the genomic composition of the selected sets of individuals. In creating genome-wide introgression libraries, large N$_{BC1}$ would be useful in achieving complete donor genome coverage (Sušic̆ 2005) but with QTL-targeted introgression libraries, donor genome coverage in the target region is sufficient. The priority, however, would be to find individuals with desired recombinations for which it would be important to have not too large but a sufficiently large progeny size (~150) to account for loss of individuals that inherit the RP target chromosome from the heterozygous parent or have recombinations outside the target region (homozygous RP genotype in the target region).

The increase in the number of individuals in the BC_1 generation would be accompanied by a linear increase in the number, and cost, of total genotyped markers. In this study, the number of markers per individual in the BC_1 generation was a function of the number of target regions for which NILs were being developed, the numbers of markers used to populate the target regions for foreground selection and to track the return to RP genome on the background chromosomes, and the number of markers contributing towards reducing linkage drag in the targets' flanking region. The numbers of markers genotyped per individual in the BC_1 generation were 160, 176 and 200 when the selection strategies applied were FMW0, FMW2 or NODIST, and FMW5 respectively (based on Section 2.2). The time and cost involved in genotyping increased drastically for larger BC_1 sizes (250, 350 and 500) which could rightly be diverted towards genotyping more backcrossed individuals per selected individual (N_{BSI}) which appeared to be an important factor in producing tiled NILs early in the backcross program.

4.4 Number of backcrosses per selected individual (N_{BSI})

The number of backcrosses per selected individual (N_{BSI}) substantially affected the length of the backcrossing scheme (or the least number of BC generations required) to produce sets of NILs with the desired introgression lengths (Table 2.4A-C). Increasing N_{BSI} (to > 20) increased the chances of obtaining individuals with further refinement or reduction in linkage drag. However, this increase if not accompanied with a good selection strategy, did not result in success of the breeding strategy. In general, more cycles of recurrent selection were required for creating NIL sets with smaller donor introgressions and for larger chromosomes. Both NODIST and FMW2 performed better with increasing N_{BSI}.

4.5 Chr 1 vs Chr 10

To construct a set of tiled NILs across a QTL (15 cM) for smaller chromosome 10 in the same number of BC generations, lesser numbers of N_{BSI} were needed, by almost half, when compared to that required for the larger chromosome 1. For instance, QTL-targeted NILs across the target region on chromosome 1 with desired introgressions of up to 3.5 cM could be produced in four BC generations but would require large N_{BSI} (> 50 and up to 100). With chromosome 10, however, the same could be achieved with smaller N_{BSI} (>20 and up to 50, Table 2.4).

Genetic length of chromosome 1 was almost twice that of chromosome 10 (202.4 and 101.9 cM respectively, Table B1). Target segments of 15 cM make up for 7.4 and 14.7% of the linkage distance on chromosomes 1 and 10 respectively. A desired introgression length would be a larger proportion of the chromosome length for smaller chromosomes than for larger chromosomes. Hence, the generation time and sample sizes required to created NILs across smaller chromosomes would be lesser than that required for larger chromosomes. Knowledge of such resource requirements could be valuable in planning an experiment.

4.6 Limitations

In simulating a process, there could be certain features of relevance that cannot be incorporated into the model due to knowledge gap or simply because they are impractical. Simulation methods (also known as Monte Carlo methods) allow comparison of alternative strategies until the end of the research plan and can help assist in the design of experiments that consist of many stochastic factors to identify the best strategy to move forward (Sun et al. 2011, Li et al. 2012) and to learn which combination of input resources performs better. The simulation parameters and the genetic model used in this study were the best available to our knowledge.

4.7 Recombination in ZmPR regions

In the second study, variation in recombination frequencies and patterns for four photoperiod QTL regions located on chromosomes 1, 8, 9 and 10 among 13 BC_1 families was documented. The results suggest, as previously observed (Okagaki and Weil 1997, Yao et al. 2002) that not all genes are recombinational hotspots. Yao et al. (2000) characterized meiotic recombination across the 140 kb multigenic *a1-sh2* interval of maize which contains at least four genes (*a1*, *yz1*, *x1*, and *sh2*) and showed also that not all recombination hotspots are genes. The data used in this study were not generated with an objective to study recombination patterns within the ZmPR1-4 QTLs but instead to create a set of ~10 near-isogneic lines in an allelic series (NILAS) for each of these QTLs via backcrossing. Therefore, the marker resolution in this study was not as high as was found in Yao et al. (2000) and other studies (Brown and Sundaresan 1990, Huihua et al. 2001). However, no other study has been able to look at genetic diversity at this resolution at the ZmPR QTLs and this represents a first step towards understanding recombinational aspects in these four QTL regions.

For the ZmPR1 region, recombination rates were higher in the region between the markers M1_1 to M1_7 (Supplementary File 3.6). In particular, more than 20% (based on counting recombinations in the Supplementary File 3.2) of the recombinations for every BC_1 family investigated were resolved within the marker interval M1_4 to M1_5 (23.8 Mbp) which harbors the photoperiod QTL peak. Although the recombination rate for this interval was an estimated 0.319 cM/Mbp (SD = 0.061), smaller than the maize genome average recombination rate of 0.68 cM/Mbp, it is possible that further fine mapping this interval reveals a recombination hotspot within this region.

ZmPR2 region was the most recombinogenic among the four QTLs studied and widely varied in recombination rates across the BC_1 families (0.38 cM/Mbp with SD = 0.148, and based on Figure 3.1 to Figure 3.4, and Supplementary File 3.6) and seemed to have recombinations across the entire 51.2 Mbp region. The photoperiod QTL peak, located in the marker interval M8_9 and M8_11 (17.7 Mbp) contained, on average, 30% (SD = 13.6%, based on counting recombinations in the Supplementary File 3.3) of the recombinations varying from 4% to 50% among the BC_1 families. One of the reasons for such huge variation in recombination rates in this marker interval could possibly be attributed to genetic crossover interference mechanism (Jones and Franklin 2006, Martini et al. 2006). Since this entire region (M8_1 to M8_11) appeared to be recombinogenic (even if not a recombination hotspot), occurrence of a recombination in a nearby interval might have prevented a recombination from happening in a given interval and vice versa, giving rise to the variation in recombination rates across the region within a BC_1 family, and consequently across the BC_1 families for a given interval.

In the ZmPR3 region, a 4.221 Mbp wide region containing a recombination hotspot was found in the region between the markers M9_2 to M9_5 with recombination rates of 1.238 to 2.798 cM/Mbp which is at least 2 orders of magnitude higher than the maize genome average recombination rate (0.68 cM/Mbp).

ZmPR4 region was the least recombinogenic (0.127 cM/Mbp with SD = 0.038) of the four QTLs evaluated in this study and did not show a preference for any sub-region for more or less recombination.

Identifying chromosomal regions or haplotype blocks in which the linkage disequilibrium is maintained at a high level in populations can avoid collection of redundant information (Jorde 2005). In this study, it was only possible to estimate the marker interval for recombinations. There is a need for further analysis of the ZmPR regions to better understand the finer structure of these regions to understand the mechanisms contributing to high or low recombinations and furthermore, to fill the knowledge gap in understanding adaptation of tropical germplasm in temperate climates.

4.8 Effect of genotypic background

Genotypic background seemed to affect the recombination rate. The ZmPR1-grouping of BC_1 families into two groups dominated with crosses associated with either of the two recurrent parents with a slightly different recombination rate within each group points to an unknown genetic basis underlying recombination. The gap in recombination was much more in the ZmPR2-grouping where the group {G15, G16} was highly non-recombinant when compared to the other groups. The cross G15, in particular, was very non-recombinant across ZmPR1-3 regions. However, independent of the recombination rates in a ZmPR QTL, there seemed to be consistency in spatial patterns of recombination along the ZmPR regions.

REFERENCES

Agresti, A., 2007 An introduction to categorical data analysis, Second edition. John Wiley & Sons Inc., Hoboken, NJ, USA.

Anderson, L. K., G. G. Doyle, B. Brigham, J. Carter, J. Carter, K. D. Hooker et al., 2003 High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. Genetics 165: 849-865.

Bauer, E., M. Falque, H. Walter, C. Bauland, C. Camisan et al., 2013 Intraspecific variation of recombination rate in maize. Genome Biol. 14: R103.

Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies, p 250-266 in Proceedings of the Forty Ninth Annual Corn and Sorghum Industry Research Conference. American Seed Trade Association, Washington DC, USA.

Bernardo, R., 2010 Breeding for quantitative traits in plants, Second edition. Stemma press, Woodbury, MN, USA.

Brown, J., and V. Sundaresan, 1990 A recombination hotspot in the maize *A1* intragenic region. Theor. Appl. Genet. 81: 185-188.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown et al., 2009 The genetic architecture of maize flowering time. Science 325: 714-718.

Chardon, F., B. Virlon, L. Moreau, M. Falque, J. Joets et al., 2004 Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168: 2169-2185.

Civardi, L., Y. Xia, K. J. Edwards, P. S. Schnable, and B. J. Nikolau, 1994 The relationship between genetic and physical distances in the cloned *a1-sh2* interval of the *Zea mays* L. genome. Proc. Natl. Acad. Sci. 91: 8268-8272.

Coles, N. D., M. D. McMullen, P. J. Balint-Kurti, R. C. Pratt, and J. B. Holland, 2010 Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184(3): 799-812.

Cornu, A., E. Farcy, and C. Mousset, 1988 A genetic basis for variations in meiotic recombinations in *Petunia* hybrid. Genome 32: 46-53.

Daley, D.J., and D. Vere-Jones, 2003 An introduction to the theory of point processes: Volume I: Elementary theory and methods, Second edition. Springer-Verlag, New York, NY, USA.

Doebley, J., 2004 The genetics of maize evolution. Annu. Rev. Genet. 38: 37-59.

Dooner, H. K., and I. M. Martínez-Férez, 1997 Recombination occurs uniformly within the *bronze* gene, a meiotic recombination hotspot in the maize genome. The Plant Cell 9: 1633-1646.

Dooner, H. K., 1986 Genetic fine structure of the bronze locus in maize. Genetics 113: 1021-1036.

Eggleston, W. B., M. Alleman, and J. L. Kermicle, 1995 Molecular organization and germinal instability of R-stippled maize. Genetics 141: 347-360.

Enns, H., and E. N. Larter, 1962 Linkage relations of *ds*: a gene governing chromosome behavior in barley and its effect on genetic recombination. Ca. J. Genet. Cytol. 4(3): 263-266.

Eshed, Y., and D. Zamir, 1994 A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. Euphytica 79: 175-179.

Eshed, Y., and D. Zamir, 1995 An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141: 1147-1162.

Fairbanks, D. J. and W. R. Anderson, 1999 Genetics: The continuity of life. Brooks/Cole Publishing Company and Wadsworth Publishing Company, New York, NY, USA.

Falke, K. C., T. Miedaner, and M. Frisch, 2009 Selection strategies for development of rye introgression libraries. Theor. Appl. Genet. 119: 595-603.

Falque, M., L. K. Anderson, S. M. Stack, F. Gauthier, and O. C. Martin, 2009 Two types of meiotic crossovers coexist in maize. The Plant Cell 21: 3915-3925.

Francis, K. E., S. Y. Lam, B. D. Harrison, A. L. Bey, L. E. Berchowitz et al., 2007 Pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. Proc. Natl. Acad. Sci. 104: 3913-3918.

Frary, A., T. C. Nesbitt, A. Frary, S. Grandillo, E. V. Knapp et al., 2000 *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-88.

Frisch, M., and A. E. Melchinger, 2001 Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci. 41: 1716-1725.

Frisch, M., M. Bohn, and A. E. Melchinger, 1999 Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci. 39: 1295-1301.

Gerton, J. L., J. DeRisi, R. Shroff, M. Lichten, P. O. Brown et al., 2000 Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. 97(21): 11383-11390.

Goodman, M. M., 2004 Developing temperate inbreds using tropical maize germplasm: rationale, results, conclusions. Maydica 49(3): 209-219.

Goodman, M. M., 2005 Broadening the U.S. maize germplasm base. Maydica 50(1): 203-214.

Gore, M. A., J-M. Chia, R. J. Elshire, Q. Sun, and E. S. Ersoz et al., 2009 A first-generation haplotype map of maize. Science 326: 1115-1117.

Henderson, I. R., 2012 Control of meiotic recombination frequency in plant genomes. Curr. Opin. in Plant Biol. 15: 556-561.

Herzog, E., K. C. Falke, T. Presterl, D. Scheuermann, M. Ouzunova et al., 2014 Selection strategies for development of maize introgression populations. PLoS ONE 9(3): e92429.

Holland, J. B., and M. M. Goodman, 1995 Combining ability of tropical maize accessions with US germplasm. Crop. Sci. 35(3): 767-773.

Hospital, F., 2005 Selection in backcross programs. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1459): 1503-1511.

Hospital, F., C. Chevalet, P. Mulsant, 1992 Using markers in gene introgression breeding programs. Genetics 132: 1199:1210.

Hospital, F., I. Goldringer, S. Openshaw, 2000 Efficient marker-based recurrent selection for multiple quantitative trait loci. Genet. Res. 75(3): 357-368.

Huihua, F., Z. Zheng, and H. K. Dooner, 2001 Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. Proc. Natl. Acad. Sci. 99(2): 1082-1087.

Huijser, P., and M. Schmid, 2011 The control of development phase transitions in plants. Development 138: 4117-4129.

Hung, H. Y., L. M. Shannon, F. Tian, P. J. Bradbuy, C. Chen et al., 2012 *ZmCCT* and the genetic basis of day-length adaptation underlying the post domestication spread of maize. Proc. Natl. Acad. Sci. 109: E1913-21.

Jones, G. H. and C. H. Franklin, 2006 Meiotic crossing-over: obligation and interference. Cell 126(2): 246-248.

Jorde, L. B., 2005 Where we're hot, they're not. Science 308: 60-62.

Karlin, S. and U. Liberman, 1978 Classification and comparisons of multilocus recombination distributions. Proc. Natl. Acad. Sci. 75: 6332-6336.

Karlin, S. and U. Liberman, 1994 Theoretical recombination processes incorporating interference effects. Theor. Population Biol. 46: 198-231.

King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen et al., 2012 Genetic dissection of a model complex trait using the *Drosophila* synthetic population resource. Genome Res. 22: 1558-1566.

Kiniry, J. R., J. T. Ritchie, and R. L. Musser, 1983 Dynamic nature of the photoperiod response in maize. Agron. J. 75: 700-703.

Li, X., C. Zhu, J. Wang, and J. Yu, 2012 Computer simulation in plant breeding. Advances in Agronomy 116: 219-264.

Litchen, M., and A. S. H. Goldman, 1995 Meiotic recombination hotspots. Annu. Rev. Genet. 29: 423-444.

Liu, K., M. Goodman, S. Muse, J. S. Smith, E. Buckler et al., 2003 Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165(4): 2117-2128.

Lynch, M., and B. Walsh, 1998 Genetics and analysis of Quantitative traits. Sinauer Associates Inc., Sunderland, MA, USA.

Mackay, T. F. C, 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. 35: 303-339.

Maguire, M. P., 1968 Evidence on the stage of heat induced crossover effect in maize. Genetics 60: 353-362.

Martini, E., R. L. Diaz, N. Hunter, and S. Keeney, 2006 Crossover homeostasis in yeast meiosis. Cell 126(2): 285-295.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li et al., 2009 Genetic properties of the maize nested association mapping population. Science 325(5941): 737-740.

Mézard, C., 2006 Meiotic recombination hotspots in plants. Biochemical Society Trans. 34(4): 531- 534.

Mézard, C., J. Vignard, J. Drouaud, and R. Mercier, 2007 The road to crossovers: plants have their say. Trends in Genetics 23(2): 91-99.

Moens, P. B., 1969 Genetic and cytological effects of three desynaptic genes in the tomato. Ca. J. Genet. Cytol. 11: 857-869.

Mungoma, C., and L. Pollak, 1991 Photoperiod sensitivity in tropical maize accessions, early inbreds, and their crosses. Crop Sci. 31: 388-391.

Nelson, P. T., 2009 Evaluation of elite exotic maize inbreds for use in long-term temperate breeding. M. S. Thesis, North Carolina State University, Raleigh, USA.

Nelson, P. T., N. D. Coles, J. B. Holland, D. M. Bubeck, S. Smith et al., 2008 Molecular characterization of maize inbreds with expired U.S. plant variety protection. Crop Sci. 48: 1673-1685.

Okagaki, R. J., and C. F. Weil, 1997 Analysis of recombination sites within the maize waxy locus. Genetics 147: 815-821.

Page, S. L., and R. S. Hawley, 2003 Chromosome choreography: the meiotic ballet. Science 301: 785-789.

Peleman, J. D., and J. R. van der Voort, 2003 Breeding by design. Trends in Plant Sci. 8(7): 330-334.

Parsons, P. A., 1988 Evolutionary rates: effects of stress upon recombination. Biol. Jour. of the Linn. Soc. 35(1): 49-68.

Patterson, G. I., K. M. Kubo, T. Shroyer, and V. L. Chandler, 1995 Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences. Genetics 140: 1389-1406.

Pollak, L. M., 2003 The history and success of the public-private project on germplasm enhancement of maize (GEM). Advances in Agronomy 78: 45-87.

Prasanna, B.M., 2012 Diversity in global maize germplasm: characterization and utilization. J Biosciences 37: 843-855.

Putterill, J., R. Laurie, and R. Macknight, 2004 It’s time to flower: the genetic control of flowering time. Bioessays 26 (4) 363-373.

R Development Core Team, 2012 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Russell, W. K., and C. W. Stuber, 1983 Inheritance of photosensitivity in maize. Crop Sci. 23: 935-939.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei et al., 2009 The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

Sun, X., T. Peng, and R. H. Mumm, 2011 The role and basics of computer simulation in support of critical decisions in plant breeding. Mol Breeding 28: 421-436.

Sušic̆, Z., 2005 Experimental and simulation studies on introgressing genomic segments from exotic into elite germplasm of rye (*Secale cereal* L.) by marker-assisted backcrossing. PhD dissertation, University of Hohenheim, Stuttgart, Germany.

Thuriaux, P., 1977 Is recombination confined to structural genes on the eukaryotic genome? Nature 268: 460-462.

Verde, L. A., 2003 The effect of stress on meiotic recombination in maize (*Zea mays L.*). Retrospective Theses and Dissertations, Paper 1397. Iowa State University, Ames, USA.

Wang, C. L., F. F. Cheng, Z. H. Sun, J. H. Tang, L. C. Wu et al., 2008 Genetic analysis of photoperiod sensitivity in a tropical by temperate maize recombinant inbred population using molecular markers. Theor. Appl. Genet. 117: 1129-1139.

Xu, S., 2013 Principles of statistical genomics. Springer Science + Business Media, New York, NY, USA.

Xu, X., A. P. Hsia, L. Zhang, B. J. Nikolau, P. S. Schnable, 1995 Meiotic recombination break points resolve at high rates at the 5’ end of a maize coding sequence. The Plant Cell 7(12): 2151-2161.

Yao, H., Q. Zhou, J. Li, H. Smith, M. Yandeau et al., 2002 Molecular characterization of meiotic recombination across the 140-kb multigenic *a1-sh2* interval of maize. Proc. Natl. Acad. Sci. 99(9): 6157-6162.

Zhao, H., M. S. McPeek, and T. P. Speed, T. P., 1995 Statistical analysis of chromatid interference. Genetics 139: 1057-1065.

APPENDIX A

Supplementary figures

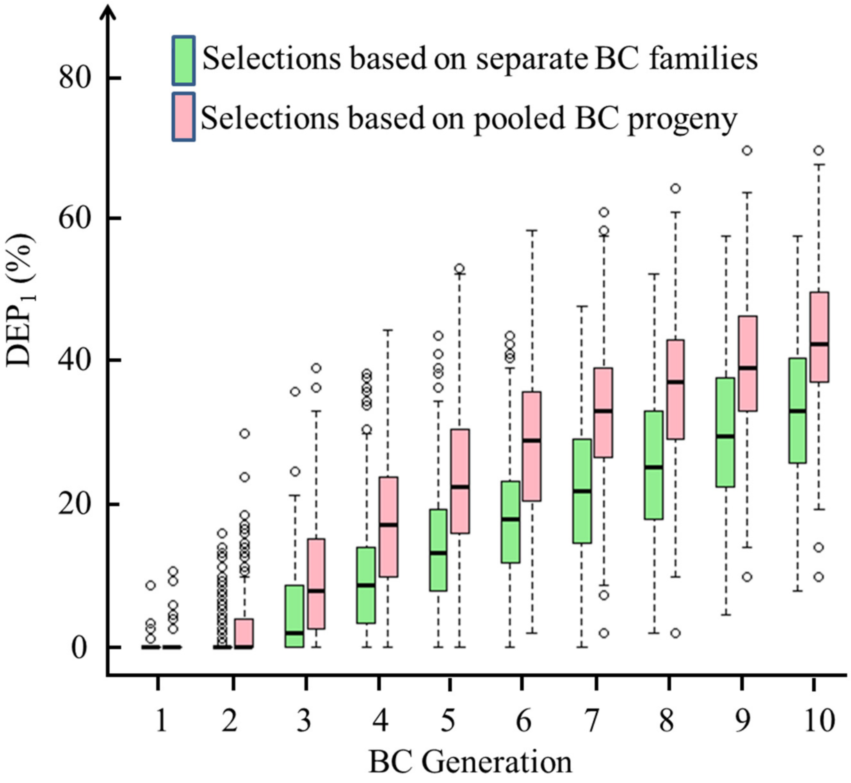


Figure A1 Percentage (%) of target region covered with one individual (DEP$_1$) of the selected NIL set.

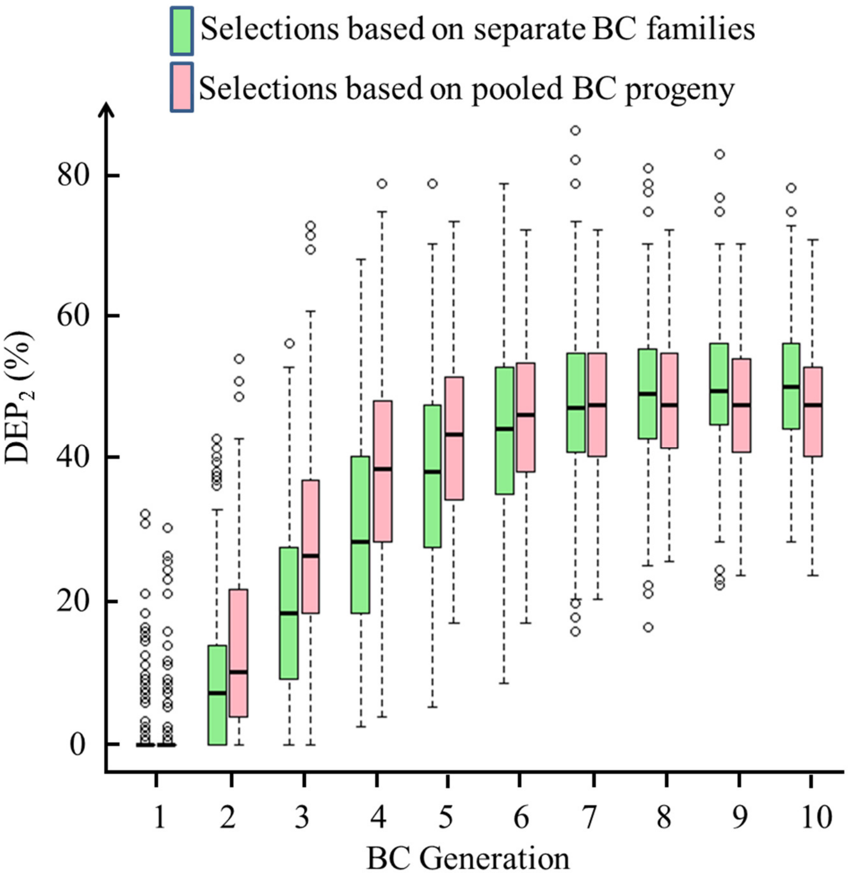


Figure A2 Percentage (%) of target region covered with any two individuals (DEP$_2$) of the selected NIL set.

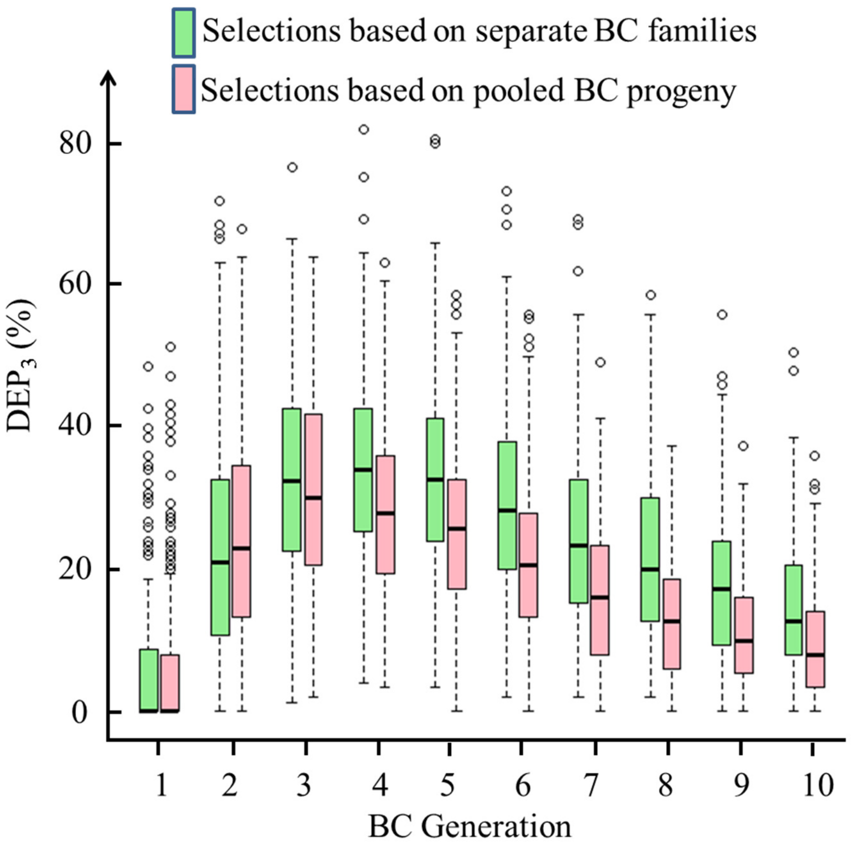


Figure A3 Percentage (%) of target region covered with three individuals (DEP$_3$) of the selected NIL set.

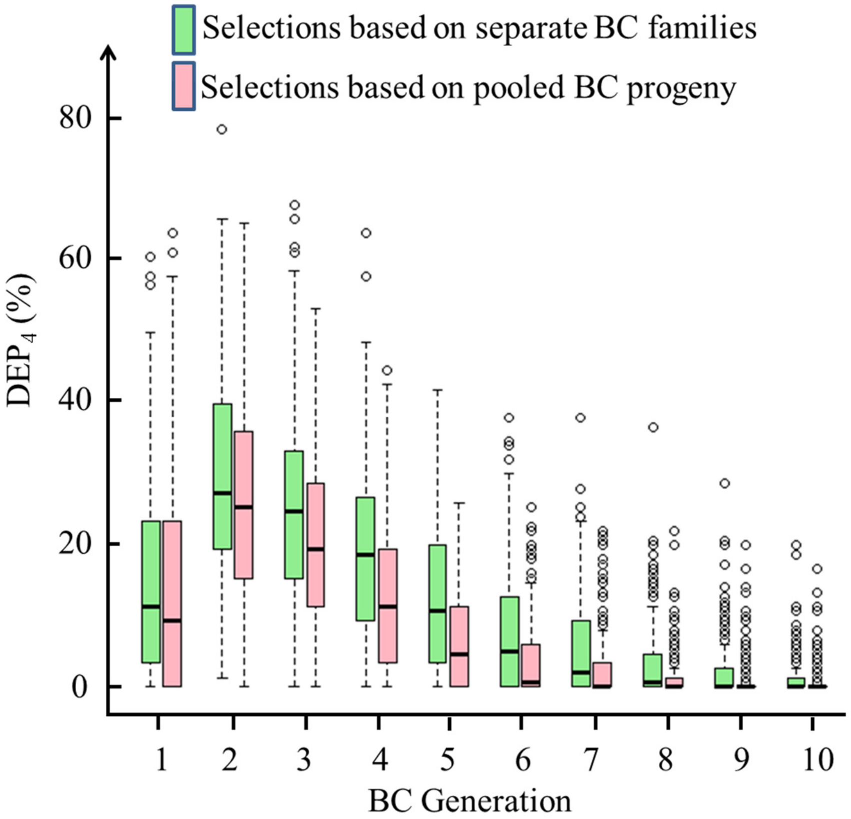


Figure A4 Percentage (%) of target region covered with four individuals (DEP$_4$) of the selected NIL set.

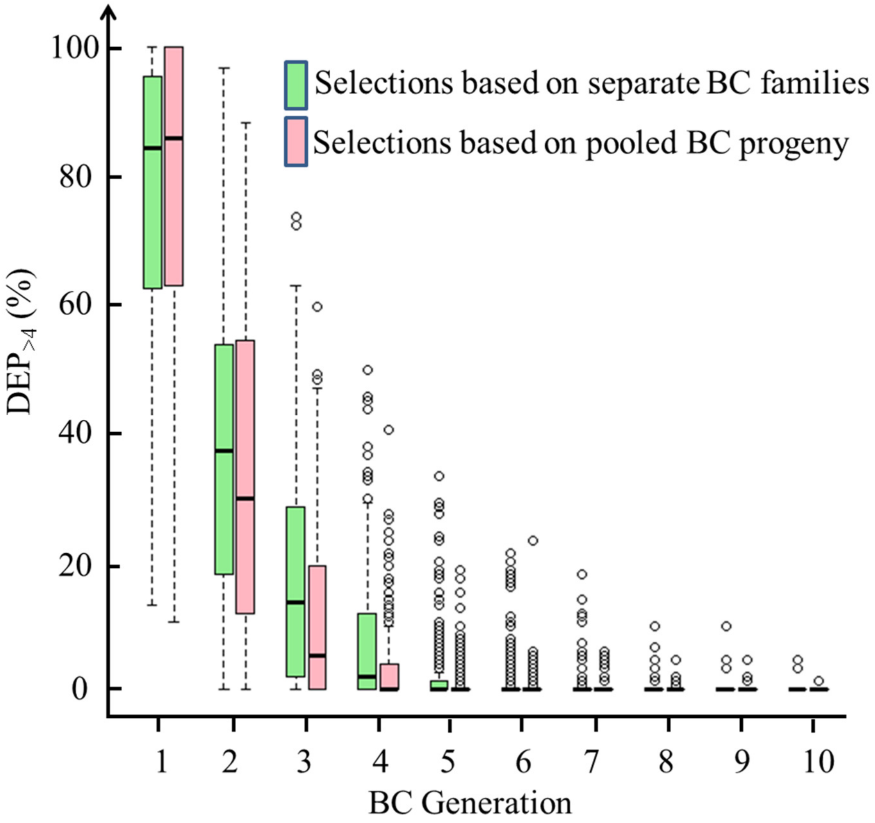


Figure A5 Percentage (%) of target region covered with more than four individuals (DEP$_{>4}$) of the selected NIL set.

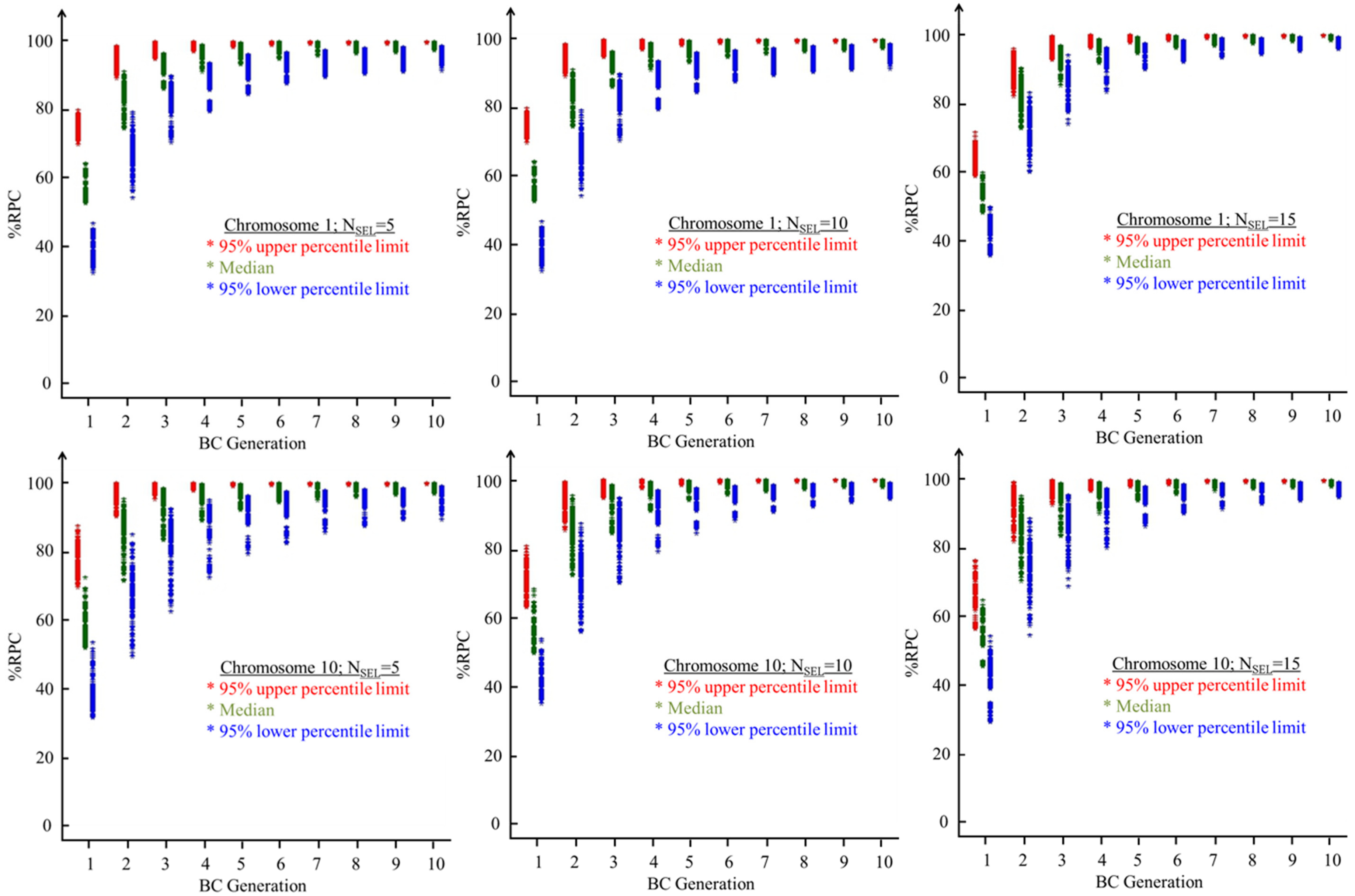


Figure A6 Median and the 95% percentile limits for the percentage of homozygous recurrent genome on the target chromosome (%RPC) of the selected set of individuals based on 1000 simulations for each combination of selection and breeding strategy (4 selection and 5 N_{BC1} x 5 N_{BSI} = 25 breeding strategies) producing 100 triplets at each BC generation in each plot. The top and bottom rows correspond to %RPC on chromosomes 1 and 10 respectively; from left to right, the figures represent strategies for 5, 10 and 15 individuals (N_{SEL}) in the tiling path.

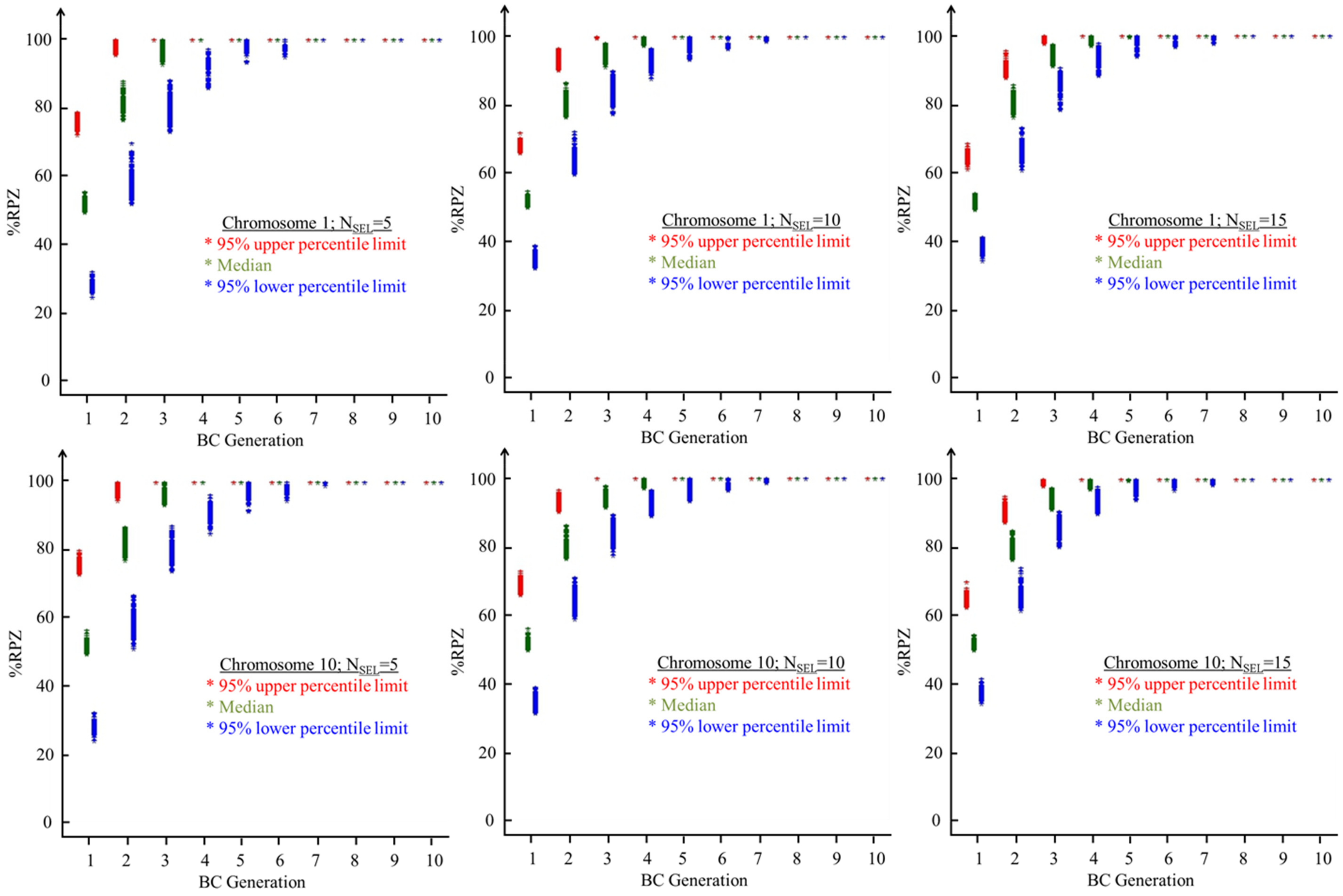


Figure A7 Median and the 95% percentile limits for the percentage of homozygous recurrent genome in the ZmPR loci of background chromosomes (%RPZ) of the selected set of individuals based on 1000 simulations for each combination of selection and breeding strategy (4 selection and N_{BC1} x 5 N_{BSI} = 25 breeding strategies) producing 100 triplets at each BC generation in each plot. The top and bottom rows correspond to %RPZ on chromosomes 1 and 10 respectively; from left to right, the figures represent strategies for 5, 10 and 15 individuals (N_{SEL}) in the tiling path.

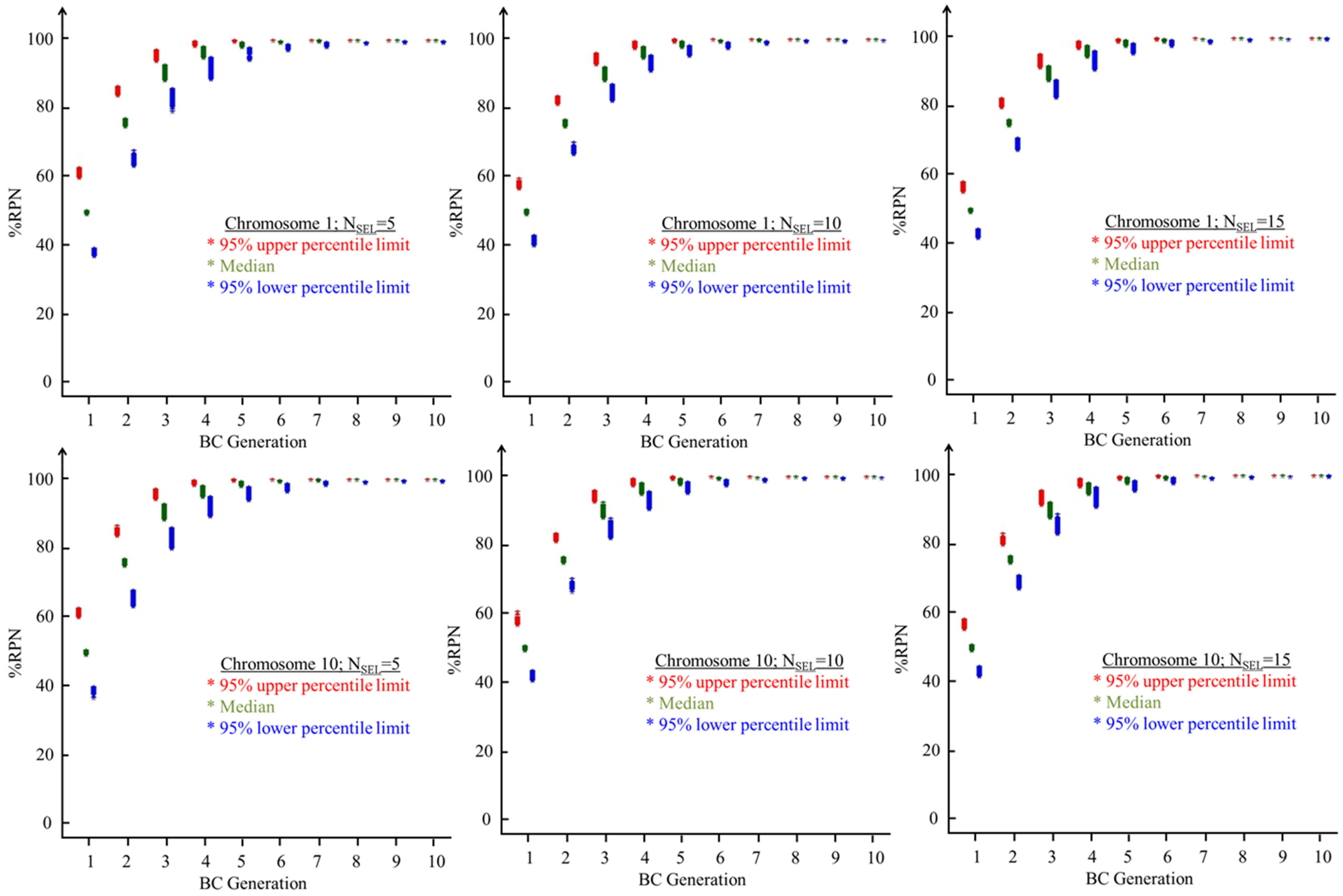


Figure A8 Median and the 95% percentile limits for the percentage of homozygous recurrent genome in the non-ZmPR and non-target loci (%RPN) of the selected set of individuals based on 1000 simulations for each combination of selection and breeding strategy (4 selection and N_{BC1} x 5 N_{BSI} = 25 breeding strategies) producing 100 triplets at each BC generation in each plot. The top and bottom rows correspond to mean introgressions at QTL on chromosomes 1 and 10 respectively; from left to right, the figures represent strategies for 5, 10 and 15 individuals (N_{SEL}) in the tiling path.

APPENDIX B

Supplementary tables

Table B1 Genetic architecture underlying the simulation study.

Chromosome	**Classification**[a]	**Length (cM)**[b]	**QTL peak**[b] ± 7.5 **(cM)**	**Gamma shape parameter** $\boldsymbol{v}$ [c]	**Proportion of non-interference crossovers** $\boldsymbol{p}$ [c]
1	ZmPR	202.4	84.9 ± 7.5	6	0.19
2	Non-ZmPR	155.7		6	0.15
3	Non-ZmPR	155.5		4.5	0.06
4	Non-ZmPR	141.1		6	0.12
5	Non-ZmPR	153.4		6	0.15
6	Non-ZmPR	109.7		4.5	0.12
7	Non-ZmPR	135		6	0.12
8	ZmPR	129.8	69 ± 7.5	9.5	0.18
9	ZmPR	114.8	45.2 ± 7.5	6	0.12
10	ZmPR	101.9	42.9 ± 7.5	4.5	0.06

[a] Coles et al. 2010, [b] Hung et al. 2012, [c] Falque et al. 2009

APPENDIX C

Statistical modelling of a crossover process

Point process models have been used in modelling meiotic crossover events which can be described as phenomena occurring at random points (locations). These models can be specified in three equivalent forms (Daley and Vere-Jones, 2003): positions of crossovers, inter-crossover distances and crossover counts. Let C_i be random variables denoting the positions of crossover occurrences on the four strand chromosome bundle (during prophase of meiosis I) , X_i be random variables denoting the inter-crossover distances where $i > 1, i \in \mathbb{N}$, and $N(s), s > 0$ denote the counting process as the total number of crossovers that have occurred up to and including position s on a chromosome (Figure C1). The inter-crossover distances can be derived from the crossover positions by setting $X_1 = C_1$ and computing $X_i = C_i - C_{i-1}, i > 1, i \in \mathbb{N}$. Conversely, the crossover positions can be computed by taking the cumulative sum of inter-crossover distances, $C_i = \sum_{k=1}^{i} X_k \, , i > 1, i \in \mathbb{N}$. C and X have discrete indices and take continuous values whereas N has a continuous index and takes discrete values. Therefore the probability distributions associated with these random variables have different forms. However, these distinct random variables can express the same events. For instance, the set of positions $\{s: N(s) < j\}$ when the crossover count is less than j and the set of positions $\{s: C_j > s\}$ when the j$^{\text{th}}$ crossover has not yet occurred are equal. Specifying the point process in any one of these three forms specifies the other two.

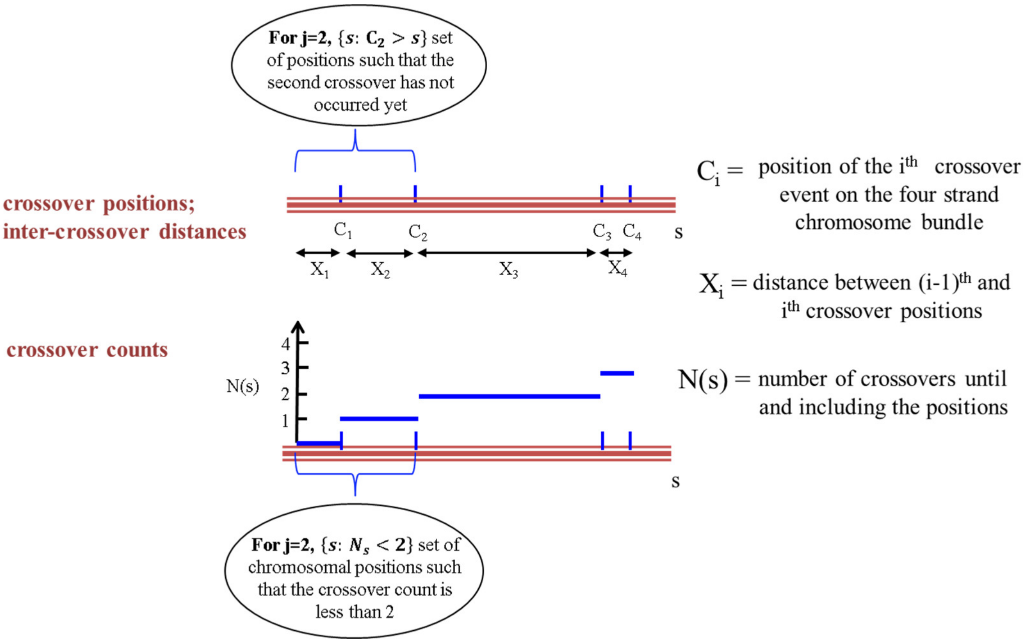

For j=2, $\{s\colon C_2 > s\}$ set of positions such that the second crossover has not occurred yet
crossover positions;
inter-crossover distances
C_1
C_2
C_3
C_4
s
X_1
X_2
X_3
X_4
C_i = position of the i^{th} crossover event on the four strand chromosome bundle
X_i = distance between $(i-1)^{th}$ and i^{th} crossover positions
crossover counts
N(s)
4
3
2
1
s
N(s) = number of crossovers until and including the positions
For j=2, $\{s\colon N_s < 2\}$ set of chromosomal positions such that the crossover count is less than 2

Figure C1 Three equivalent representations of a crossover process in terms of crossover positions, inter-crossover distances and crossover counts.

Stationarity of a crossover process

Stationarity of a crossover process can be described in terms of two important characteristics: 1) the dependence of a crossover in a chromosomal segment on the location of the segment on the chromosome, and 2) the dependence of a crossover at a position on other crossover occurrence(s). While decreased crossover activity in near-centromeric regions and increased activity in distal regions of the chromosome in some species is a dependence behavior of type (1), crossover interference in regions adjacent to other crossovers in many organisms relates to type (2). The nature of this dependence can be used to categorize crossover processes.

A crossover process is stationary in terms of the type (1), or simple stationary, when the distribution of the number of crossovers in a chromosomal segment depends on the length of the segment but not on its' location on the chromosome. That is, $P\{N(s, s+t]\}$ depends on the length of the chromosomal segment t and not on the starting position s of the segment. In this case, non-stationarity would mean that the probability of crossover changes at each position and hence depends on the location of the segment on the chromosome. A point process is stationary in terms of the type (2), or interval stationary, when the crossover at a position does not depend on other crossovers on the chromosome. This would imply that the inter-crossover distributions are independent. Generalized point processes including both simple and interval non-stationarities can be constructed but would require more parameters to define them. The archetypal point processes are the Poisson and renewal processes (Daley and Vere-Jones, 2003).

Simple Poisson process and stationary renewal process

A simple Poisson process is defined by Poisson number of crossovers $N(x)$ in a finite chromosomal segment where x is the length of the segment. For a series of k disjoint chromosomal segments $x_i, i = 1, \dots, k$, the numbers of crossovers in each segment is given by

$$P\{N(x_i) = n_i, i = 1, \dots k\} = \prod_{i=1}^{k} \frac{(\lambda x_i)^{n_i}}{n_i!} e^{-\lambda x_i} \qquad (1)$$

where λ, the rate parameter denotes the number of crossovers in one genetic distance unit (Morgan) on the four-strand chromosome bundle. This definition (equation 1) subsumes both simple and interval stationarities. Therefore to estimate the distribution of inter-crossover distances, it is enough to consider just one chromosomal segment. Distance to the next crossover, using equation 1, is given as

$$P\{N(x) = 0\} = e^{-\lambda x} \qquad (2).$$

Equation 2 represents the probability of no crossovers in a chromosomal segment of length x. It can also be interpreted as the probability that the distance to the next crossover event will have length more than x. That is, in terms of the inter-crossover distances X, $P\{X > x\} = e^{-\lambda x} \Rightarrow P\{X \leq x\} = 1 - e^{-\lambda x}$ which is the exponential distribution function with mean distance to the next crossover $\frac{1}{\lambda}$. Likelihood of a crossover in the vicinity of another crossover is high with exponential inter-crossovers (Figure C2). Poisson processes are a special case of count-location models (Karlin and Liberman 1978) of crossovers which define a discrete count distribution to choose the number of crossovers N and assume that the positions of crossovers given n are independently distributed according to a continuous location distribution.

Contrary to count-location models, renewal processes start by defining a distribution for the inter-crossover distances. A renewal process is a point process in which the probability of a crossover at any position on the chromosome can depend on the occurrence of an adjacent crossover but not on any other crossovers. Stationary renewal processes are an extension of Poisson processes in which simple stationarity still holds. However, likelihood of a crossover in the neighborhood of another crossover can be reduced by choosing an appropriate distribution for inter-crossover distances (Figure C2).

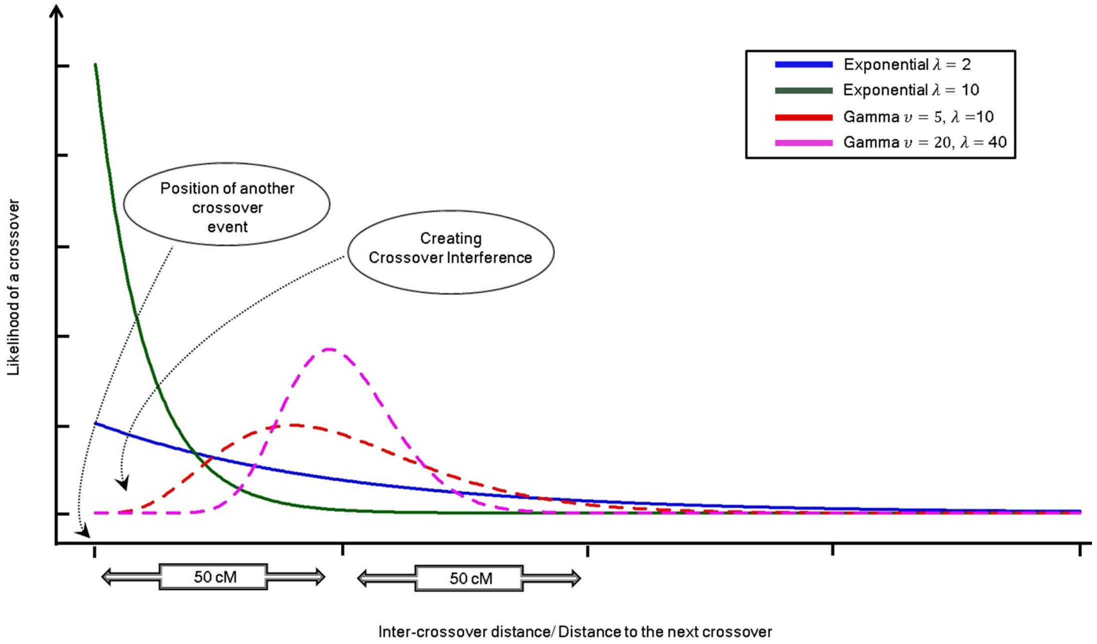


Figure C2 Likelihood of a crossover in the vicinity of another crossover event: visualizing crossover interference using Poisson and Gamma distributions.

Since the distance to the next crossover depends on an adjacent crossover, the process is interval non-stationary. However, it still has independent inter-crossover distances. A common choice for the distribution of inter-crossover distances is the Gamma distribution.

Gamma distribution to model inter-crossover distances

Crossover processes with inter-crossover distances following a gamma distribution are stationary renewal processes, that is, simple-stationary but not interval-stationary. Gamma distribution has a density function given by

$$f(x;\upsilon,\lambda) = \frac{\lambda^{\upsilon}x^{\upsilon-1}e^{-\lambda x}}{\Gamma(\upsilon)}, \upsilon > 0, \lambda > 0, 0 < x < \infty \qquad (3)$$

where the shape parameter υ denotes the intensity of interference and the rate parameter λ denotes the rate at which a crossover occurs after the interference is exerted. If the average number of crossovers per Morgan (100 cM) per bivalent (the four strand chromosome bundle during prophase of meiosis I) is 2, then the mean distance between any two crossovers is 0.5 Morgans (50 cM). Equating this to the gamma mean $\frac{\upsilon}{\lambda}$, we get $\lambda = 2\upsilon$. As υ increases, that is the intensity of interference increases, the likelihood of larger inter-crossover distances increases (Figure C2). A positive interference can be created with $\upsilon > 1$while a no interference model corresponds to $\upsilon = 1$ which reduces the equation 3 to the exponential density.

Gamma distribution can be used to obtain successive crossover positions from left to right on a chromosome of length L_M (in Morgans) by generating a random value x from a gamma distribution with given shape and rate parameters υ and λ, where x stands for the distance from the start (position 0 M) of the chromosome or a previous crossover position. The generation of crossover positions stops whenever $x > L_M$. In this study, interfering crossovers were formed using gamma distributed inter-crossover distances with proportion $(1-p)$, on average, of the total number of crossovers across the bivalent from the interference pathway (Falque et al. 2009). So, for a chromosome of length L_M, the interfering gamma inter-crossovers with shape υ and rate 2υ (Falque et. 2009, Table B1) were generated for a map length of $(1-p)L_M$ because the stopping criterion for crossover generation was defined by the map length here unlike the count-location models with stopping criterion conditional on the number of crossovers. Independently, for a map length of pL_M, non-interference crossovers were generated using gamma distribution with $\upsilon = 1$ and $\lambda = 2\upsilon = 2$. Crossover generation stopped whenever the sum of successive distances exceeded the map length defined for the pathway. Union of the two sets of crossover positions was placed from either left to right or right to left of the chromosome and crossover was simulated for two non-sister chromatids that were randomly selected.

APPENDIX D

List of supplementary data files

Supplementary File 2.1 MIL

Supplementary File 2.2 RPC

Supplementary File 2.3 RPZ

Supplementary File 2.4 RPN

Supplementary File 3.1 Genotypic Frequencies

Supplementary File 3.2 Haplotypes for chr1

Supplementary File 3.3 Haplotypes for chr8

Supplementary File 3.4 Haplotypes for chr9

Supplementary File 3.5 Haplotypes for chr10