

# Combining DNA pooling with selective recombinant genotyping for increased efficiency in fine mapping

Xiao-Fei Chi · Xiang-Yang Lou · Qing-Yao Shu

Received: 9 August 2009 / Accepted: 17 October 2009 / Published online: 8 November 2009  
© Springer-Verlag 2009

**Abstract** One of the key steps in positional cloning and marker-aided selection is to identify marker(s) tightly linked to the target gene (i.e., fine mapping). Selective genotyping such as selective recombinant genotyping (SRG) is commonly used in fine mapping for cost-saving. To further decrease genotyping effort and rapidly screen for tightly linked markers, we propose here a combined DNA pooling and SRG strategy. A two-stage pooled genotyping can be used for identifying recombinants between a pair of flanking markers more efficiently, and a joint use of bulked DNA analysis and two-stage pooling can also save cost for genotyping recombinants. The combined DNA pooling and SRG strategy can further be extended to fine mapping for polygenic traits. The numerical results based on hypothetical scenarios and an illustrative application to fine mapping of a mutant gene, called *xl(t)*, in rice suggest that the proposed strategy can remarkably reduce genotyping amount compared with the conventional SRG.

## Introduction

Positional cloning is a powerful tool for identification of genes underlying human diseases and traits of economic importance in plants and animals, and for forward genetics studies (Glazier et al. 2002; Peters et al. 2003; Tanksley et al. 1995). With advances in genome sequencing and the exploitation of abundant DNA molecular markers scattered throughout the whole genome, positional cloning has been used successfully in numerous instances to isolate genes from human, animals, and plants (Andersson and Georges 2004; Korstanje and Paigen 2002; Remington et al. 2001; Salvi and Tuberosa 2005; Varshney et al. 2006; Yano 2001) and has evolved as a routine technique in several species such as *Arabidopsis thaliana* and rice (*Oryza sativa* L.) (Jander 2006; Jander et al. 2002; Lukowitz et al. 2000; Yano and Sasaki 1997). Meanwhile, marker-assisted selection (MAS) strategy has proven to be a very useful technique in plant and animal breeding as it could improve the efficiency of breeding through precise transfer of genomic regions of interest (foreground selection) and accelerating the recovery of the recurrent parent genome (background selection) (Babu et al. 2004; Collard and Mackill 2008; Dekkers 2004; Ribaut and Hoisington 1998). One of the prerequisites in positional cloning and efficient marker-aided breeding is to identify marker(s) physically close to the target gene. Furthermore, fine mapping offers a way to distinguish between pleiotropy and close linkage in the multiple phenotype context (Christians and Senger 2007).

Usually, several thousand or more meiotic events are needed in fine mapping to narrow a gene down to a small segment size allowed for positional cloning and dissecting multiple linked QTLs, because the segment is defined by the closest flanking crossover events—for example, to have a 95% chance of finding at least one crossover in a 0.1 cM

---

Communicated by M. Xu.

---

X.-F. Chi · Q.-Y. Shu  
Institute of Nuclear Agricultural Sciences, Zhejiang University,  
Hangzhou, People's Republic of China

X.-Y. Lou  
Institute of Bioinformatics, Zhejiang University,  
Hangzhou, People's Republic of China

X.-Y. Lou (✉)  
Department of Biostatistics, University of Alabama  
at Birmingham, RPHB 420B, 1665 University Boulevard,  
Birmingham, AL 35294, USA  
e-mail: xlou@soph.uab.edu; xylou@uab.edu

interval would require approximately 3,000 meiotic events (Durrett et al. 2002; Jander et al. 2002; Tanksley 1993). The standard mapping procedures with experimental populations typically include two steps: preliminary linkage analysis and fine-scale mapping (Jander 2006; Jander et al. 2002; Lukowitz et al. 2000; Tanksley et al. 1995). Initially, tens to hundreds of segregating progeny such as  $F_2$  and backcross (BC) are sparsely genotyped to search for a candidate interval with markers typically spaced roughly several centiMorgans (cM) across the whole genome. Two markers closest to the gene, one on either side, are selected as a pair of flanking markers for fine mapping. Next, a sufficiently large segregating population consisting of several thousand or more individuals is analyzed with another high-density marker set that covers this interval well to pinpoint a candidate region suitable for chromosome walking or chromosome landing. To reduce the genotyping burden in fine mapping, selective recombinant genotyping (SRG) is widely adopted (Jander 2006; Jander et al. 2002; Lukowitz et al. 2000; Ronin et al. 2003; Tanksley et al. 1995). The rationale behind SRG is that the recombinants in the target segment, a minority of a large mapping population, contain condensed recombination information contributing to mapping accuracy and thus only those recombinants are collected for use in fine mapping. The most part, non-recombinants in the interval that are less informative for mapping the gene, and, in relatively rare cases, those carrying an even number of recombination events, can be discarded without further analysis. Although SRG can substantially reduce genotyping effort, it is still labor-intensive and cost-consuming to genotype a large segregating population with the conventional gel electrophoresis-based methods used in many laboratories.

In this report, we propose to apply DNA pooling strategies to SRG for achieving additional cost saving. Assuming that any disparate marker alleles can be detected in a small pool (e.g., 20 individuals) and that the relative frequencies of alleles can be reasonably well estimated in a large bulked sample (e.g., hundreds of individuals), we advocate the use of two pooled genotyping designs for reducing the numbers of PCR reactions and genotyping assays: two-stage pooling (Chi et al. 2009) and bulked-sample analysis (BA) (Michelmore et al. 1991; Sham et al. 2002). The applications include (1) a two-stage pooling for identifying recombinants between a pair of flanking markers, and (2) a combined use of bulked DNA analysis and two-stage pooling for genotyping recombinants.

## Theory and numerical results

In BA, first, approximately equal quantities of DNA from multiple subjects are pooled; then, the pooled DNA is

examined *en masse* with markers of interest to estimate the allele frequencies through comparing the signal intensity of the allelic band on the electrophoretic gel with a reference sample of known allele frequencies (e.g.,  $F_1$  heterozygote). For more detailed information on BA, please refer to the relevant literature (e.g., Sham et al. 2002). In the following we will focus our presentation on two-stage pooling.

### Two-stage pooling genotyping design

In a two-stage DNA pooling study, a pool is subdivided into several sub-pools—from a broader perspective, an individual in a BC, doubled haploid (DH), or recombinant inbred line (RIL) population can be regarded as a sub-pool with size 1 since such an individual contains one independent meiosis, and one in an  $F_2$  population as a sub-pool with size 2 since it has two independent meioses in the context of fine mapping. We first genotype each pool *en masse* (stage one) and then genotype each subgroup or individual (stage two) once a pooled sample is determined to contain the given allele(s).

Assume that  $n$  subjects can be randomly divided into  $k$  pools of size  $r$ , i.e.,  $n = kr$ , and a pool of size  $r$  can be further subdivided into  $l$  sub-pools of size  $s$ , i.e.,  $r = ls$ . As the pooled tests carried out at stage one can inform us whether a pool contains a given allele, we can exclude the negative pools that have no such an allele from genotyping at stage two. Only the sub-pools in those positive pools are required to individually genotype for identifying which contains the allele and which does not. Thus, the expected number of genotypings is (Chi et al. 2009),

$$y = k + k[1 - (1 - \pi)^r] \left[ l - \frac{(1 - \pi)^{s(l-1)} - (1 - \pi)^r}{1 - (1 - \pi)^r} \right], \quad (1)$$

where  $\pi$  is the frequency of the allele of interest. When  $\pi$  is small,  $(1 - \pi)^r \approx e^{-r\pi}$  and then,

$$y \approx k + k(1 - e^{-r\pi}) \left[ l - \frac{e^{-s(l-1)\pi} - e^{-r\pi}}{1 - e^{-r\pi}} \right]. \quad (2)$$

The latter term in the right of Eq. 1 or 2 is the expected number of genotypings in the pools having at least one copy of the allele. For a given  $\pi$ , we can compute optimal pool and subpool allocations that minimize the expected number of genotypings by setting the partial derivative equal to zero and solving the resulting equations. In most cases, the analytical solution is not immediately obvious and a numerical solution may be used (Chi et al. 2009).

From Eq. 1, theoretically, a two-stage pooling approach will reduce genotyping workload if  $\pi < 0.382$  for the case of  $s = 1$  (e.g., BC, DH and RIL populations) or  $\pi < 0.214$  for the case of  $s = 2$  (e.g.,  $F_2$  population). Figure 1 presents the ratio of the expected numbers of two-stage pooled genotyping to those of exhaustive genotyping without

exclusion at stage one in the scenarios potentially occurring in fine mapping. As shown in Fig. 1, DNA pooling requires fewer PCR reaction and genotyping assays than the traditional genotyping does, although the magnitude of reduction depends on the  $\pi$  value and the pooling allocation. For example, the pooling design can reduce the genotyping amount approximately to 40% for  $s = 1$  and to 55% for  $s = 2$  when  $\pi = 0.05$ , while it can reduce the genotyping amount to 6% for  $s = 1$  and to 8% for  $s = 2$  when  $\pi = 0.001$ .

Figure 1 also suggests that the expected number of genotypings varies with pooling scheme and there exists a minimum whose coordinate at horizontal axis represents the optimal pool size. Given the  $\pi$  and  $s$  values, in theory, by differentiating  $y$  with respect to  $r$  or  $k$  in Eq. 1 or 2, and solving the resulting equation, we can obtain the optimal allocation that minimizes the expected number of genotypings. It is possible to get an empirical one from such a plot for avoiding complicated computation.

Around the optimum, there is also a wide range for pool size within which the number of genotypings is close to the minimum. For example, the optimal pool size is approximately 25 with a ratio of 6%, but the ratio is less than 7% for pool size from 20 to 40 when  $\pi = 0.001$  and  $s = 1$ , suggesting a nearly optimal pooling scheme can still considerably reduce genotyping burden. This feature allows us to determine a cost-effective pooling scheme, even though we only roughly know the allele frequency from the primary mapping or a BA.

Without loss of generality, we first use the case of single recessive gene to describe the DNA pooling applications. Then we propose further extension to fine mapping for complex traits (Darvasi and Soller 1994; Ronin et al. 2003). (For convenience of notation, we denote throughout the report a locus in mathematical font and alleles in italic

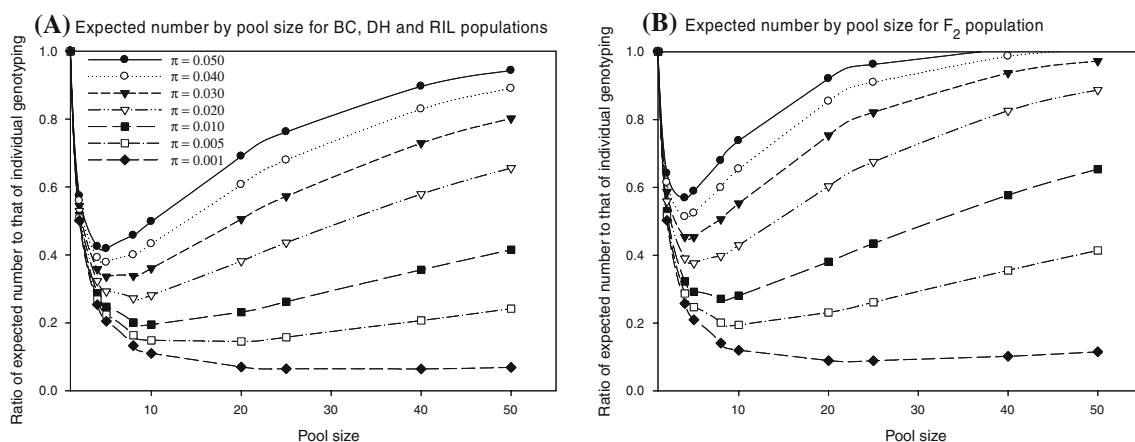
font in which the alleles in uppercase letters are in coupling phase, that is, they come from one parent, and those in lowercase letters from another parent.)

### Two-stage pooling for identifying recombinants between a pair of flanking markers

In the SRG, the recombinants in the vicinity of the gene, the crossover-enriched samples, are first selected from a large mapping population by typing a pair of flanking markers, and subsequently only the recombinants that are more informative portion for fine mapping are further evaluated. Traditionally, in order to identify recombinants, we need to individually genotype all samples (usually several thousand or more individuals) for both markers. A two-stage pooling strategy offers a more efficient tool for the screening of recombinants.

Suppose a target gene  $T$  with alleles  $T$  and  $t$  bracketed by a pair of flanking markers  $A$  with alleles  $A$  and  $a$ , and  $B$  with alleles  $B$  and  $b$ , which are identified by a primary mapping. Consider the subjects homozygous for the target locus in an experimental population derived from a cross between two inbred lines with genotypes  $AATTBB$  and  $aattbb$ , respectively, such as BC,  $F_2$ , DH, or RIL population—for illustrative purpose, in what follows we consider as an example homozygous recessive individuals with  $tt$  that can be viewed as a selected sample from a single tail of the phenotypic distribution. As mentioned previously,  $s = 1$  for an individual in a BC, DH, or RIL population, while each  $F_2$  individual is virtually a pool of size 2. Our objective is to seek for the subjects (i.e., sub-pools) with recombination event(s) either between  $A$  and  $T$  or between  $B$  and  $T$ , i.e., containing gamete either  $Atb$  or  $atB$ .

For a linked marker locus, e.g.,  $A$ , the frequency of allele  $A$  in a  $tt$  pool from BC, DH, and  $F_2$ , the  $\pi$  in Eq. 1 or



**Fig. 1** Expected numbers of genotypings in two-stage pooling strategies. The y-axis is the ratio of the expected number of genotypings to the number of individual genotypings. **a** and **b** are

for BC, DH and RIL populations, and  $F_2$  population, respectively (i.e., sub-pool sizes 1 and 2, respectively)

2, is the recombination rate between A and T, while, in RILs, it is the proportion of recombinant zygote that is higher than the recombination rate per generation because of the accumulation of crossovers occurring at each meiosis. The relationship between the recombination rate and the proportion of recombinant zygote has been reported in the literature (e.g., Broman 2005; Haldane and Waddington 1931; Martin and Hospital 2006),

$$R = \frac{2\theta}{1 + 2\theta},$$

where  $R$  is the proportion of recombinant zygote, and  $\theta$  is the recombination rate.

Given the estimated recombination rates of two flanking markers, we can plan a nearly optimal two-stage pooling scheme. To illustrate the reduction in genotyping, consider an interval of 4 cM, a size suggested by Jander et al. (2002) for primary mapping. A two-stage pooling can reduce at least 70% of genotyping effort in both cases, although the reduction depends on the gene location and varies between two extreme cases: T located in the midpoint of the two markers A and B, and coincident with one of the markers (Fig. 2).

#### Combined use of bulked DNA analysis and two-stage pooling genotyping

In the conventional SRG, we need to individually examine all the recombinants (usually hundreds of individuals) for a set of markers with adequate coverage after collecting the recombinants that contain gamete either *Atb* or *atB* from a large mapping population. Here we suggest to pool the recombinants into two contrasting groups with gametes *Atb* and *atB*, respectively, or into one overall group. The former can supply more information on marker positions and additional accuracy. For example, if a new marker, say X,

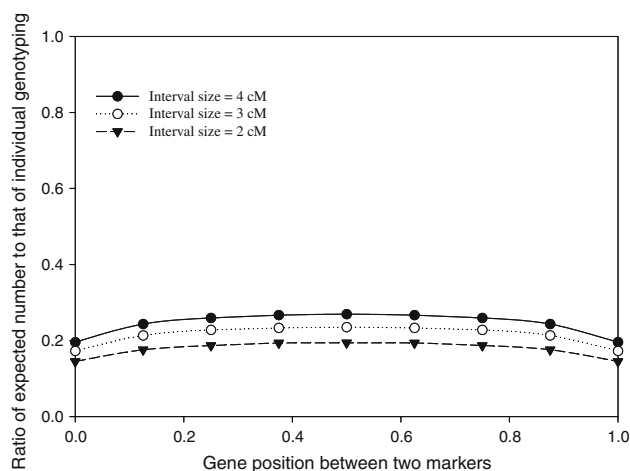
is located between A and T, most, if not all, alleles in the *atB* pool will be  $x$  while the frequency of allele  $x$  varies from 0 to 1 depending on the relative genetic distance of locus T to markers A and B in the *Atb* pool. Likewise, for a marker located between B and T, say Y, allele  $y$  is predominant in the *Atb* recombinants and the frequency of allele  $y$  varies from 0 to 1 in the *atB* recombinants. In most cases, we can determine an interval in which a new marker is located according to the distributions of alleles in the two contrasting pools.

Instead of individual genotyping, we first carry out a BA to estimate allele frequencies at a new marker locus in the recombinant pool(s) and thereby screen for the nearer markers to the target gene. A high frequency (e.g., >10%) of the allele from the same parental source with *T* indicates that this marker is farther away from the gene no matter in one overall pool or in one of the contrasting pools, and thereby the majority of markers in fine mapping can be excluded from further genotyping. Again, applying the two-stage pooling approach can lead to an additional increase in genotyping efficiency to the remained markers according to the estimated minor allele frequency. Considering an exclusion criterion of 10%, the joint BA and two-stage studies can, on average, reduce more than 90% of genotyping effort.

#### Extension to QTL mapping

The theory presented above is equally applicable to fine mapping of genes underlying complex polygenic traits whose variation is quantitative, and under the control of multiple genes and further blurred by differences in environment and developmental noise, known as quantitative trait loci (QTLs). Once a QTL is mapped to a specific chromosomal segment with a certain confidence by a primary mapping analysis, a large segregating population such as  $F_2$  and/or BC derived from the original cross, or a cross between a pair of nearly isogenic lines (NILs) or recombination inbred lines (RILs) are genotyped with a saturated set of markers in the vicinity of the QTL for further genetic dissection (Ashikari and Matsuoka 2006; Darvasi 1998; Flint and Mott 2001; Flint et al. 2005; Nadeau and Frankel 2000; Salvi and Tuberosa 2005; Yano 2001). Since only recombinant individuals within an identified interval will contribute to further mapping accuracy, selective genotyping has been advocated to use in high-resolution mapping (Darvasi 1998; Ronin et al. 2003; Thaller and Hoeschele 2000). As in mapping a single mutant gene aforementioned, BA and two-stage pooling can lead to a potential reduction in genotyping.

Two-stage pooling genotyping can be used for diverse purposes in selective genotyping designs. In SRG (Ronin et al. 2003; Thaller and Hoeschele 2000), two-stage



**Fig. 2** Expected genotyping burden using two-stage DNA pooling to screen for recombinants under three interval sizes (4, 3, and 2 cM)



pooling genotyping can be performed to determine multi-marker haplotypes in any genomic region of interest by generating ordinal pools on the basis of the marker(s) previously assayed, similar to genetic walking along a chromosome (Michelmore et al. 1991). Instead of exhaustive genotyping for all segregants, two-stage pooling genotyping can also be applied to select recombinant individuals at the interval previously defined to contain a QTL, although its efficiency is dependent on the interval size, the magnitude of the QTL effects, and the location of the QTL relative to two flanking markers. For example, if the QTL effects are significant enough and/or one of two flanking markers is sufficiently close to the QTL, the individuals at each of the two phenotypic extremes can be classified into pools for use of a two-stage pooling genotyping strategy. If two flanking markers are close to each other, we can exhaustively genotype one of two flanking markers and then perform a two-stage pooling to genotype the other marker by grouping samples that are identical in the marker genotype.

BA can be used to rapidly screen for the nearer markers to the target QTL from a set of closely spaced markers in SRG. The recombinants, of which the majority have only one recombination event within the target region except for an exceedingly small portion that contain an odd number of multiple crossovers, can be classified into two recombinant groups,  $Ab$  and  $aB$ . Through hitchhiking effects, the change in QTL allele frequencies in pools will cause a parallel change in the allele frequencies at a tightly linked marker (Kearsey 1998; Wang et al. 2007). A nearer marker is more likely to have a similar difference in allele frequency to the target gene. The comparison of phenotypic mean between two groups will reveal the relative location of the targeted QTL to two flanking markers, i.e.,  $\mu_{Ab} > \mu_{aB}$  implies the QTL closer to  $A$  while  $\mu_{Ab} < \mu_{aB}$  implies the QTL closer to  $B$ . Then the comparison of allele frequencies with BA can exclude a portion of markers with a disparate pattern from further individual genotyping. When there are a sufficient number of recombinants available, each group of recombinants can be further divided into several subgroups concordant for high- and low-trait values, respectively. The phenotypically similar individuals are more likely to share the same genotype at a tightly linked marker than phenotypically discordant ones. Comparison of allele frequencies at markers between the phenotypic extremes within the  $Ab$  or  $aB$  group leads to efficient identification of candidate markers.

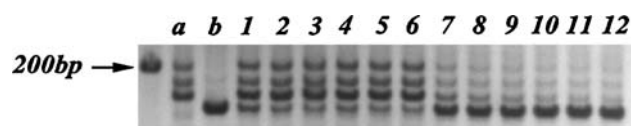
### An illustrative application

To demonstrate the technical feasibility as a rapid and cost-effective tool in fine mapping, we conducted a pooled

genotyping experiment by using an SSR marker, RM3701, polymorphic between two inbred lines in rice (*O. sativa* L.): Nipponbare (*japonica*) and Huangyu B (*indica*). Figure 3 presents a genotyping result of 12 pools with different mixing proportions of Nipponbare to Huangyu B, 4:1, 9:1, 14:1, 19:1, 29:1, 39:1, 1:4, 1:9, 1:14, 1:19, 1:29, and 1:39, respectively. As shown in Fig. 3, pooling with a size of 40 or less could yield an estimated 100% sensitivity by the electrophoresis-based detection technique, suggesting the validity of the proposed pooling procedure with the conventional gel electrophoresis-based detection technique.

Next, we applied the combining of DNA pooling with SRG strategy to fine mapping of a mutant gene, called *xl(t)*, which conditions the xantha leaf phenotype and is recessively inherited (Zhou et al. 2006). BA was first performed to identify linked markers from a total of 344 SSR loci scattered across 12 chromosomes on the basis of two contrasting pools that were composed of 40 mutant yellow and normal green  $F_2$  plants, respectively, from the cross between Nipponbare (a normal *japonica* variety with green leaves) and Huangyu B (a mutant yellow *indica* line carrying the *xl(t)* allele), indicating that the *xl(t)* locus was linked to RM21 on chromosome 11. The nearby polymorphic markers were individually typed for the 40  $F_2$  mutant plants until two tightly linked markers were identified—no single recombinant plant among 40 plants for both RM3701 and RM7226, suggesting that the *xl(t)* locus was potentially located between RM3701 and RM7226.

A total of 1,720 mutant yellow plants of the large  $F_2$  populations derived from two crosses, Nipponbare  $\times$  Huangyu B and OS-lpa-XS110-2 (another normal *japonica* variety)  $\times$  Huangyu B, respectively, were subjected to fine mapping. Using two-stage pooling with a pool size of 5  $F_2$  individuals, 11 and 10 recombinants were identified for RM3701, respectively, from 964 mutant type  $F_2$  plants of Nipponbare  $\times$  Huangyu B and 756  $F_2$  mutant individuals of OS-lpa-XS110-2  $\times$  Huangyu B, suggesting estimated genetic distances of 0.61 and 0.66 cM between RM3701 and the *xl(t)*, respectively. Three and two recombinants were found for RM7226 in Nipponbare  $\times$  Huangyu B and OS-lpa-XS110-2  $\times$  Huangyu B population, respectively, suggesting estimated genetic distances of 0.21 and 0.13 cM, respectively. The number of genotypings was reduced to



**Fig. 3** Electrophoresis-based pooled genotyping for 12 pools with different mixing proportions of Nipponbare to Huangyu B, from left to right (1–12), 4:1, 9:1, 14:1, 19:1, 29:1, 39:1, 1:4, 1:9, 1:14, 1:19, 1:29, and 1:39, respectively. Lanes *a* and *b* represent the bands from parents Nipponbare and Huangyu B, respectively

less than 900 for both the markers, representing a two-thirds saving in screening for recombinants as compared with the individual genotyping. Although the estimated genetic distance between RM3701 and RM7226 is less than 1.0 cM, this interval still extends  $\sim 6$  Mbp as it spans over the centromere region where the crossover frequency is extremely low. Then the DNA samples from the 26 recombinants of either RM3701 or RM7226 were pooled at roughly equimolar ratios and further assayed with another high-density marker set, including 5 SSR and 24 InDel markers (see Fig. 4 for the genomic positions). Eight of the 24 InDel markers are polymorphic between the parents and of good quality in agarose gels (see Table 1 for detailed information). To achieve the maximum genotyping efficiency, BA was performed to screen for tightly linked markers. As shown in Fig. 5, the markers relatively more distant from the gene such as RM536 (Fig. 5b) had a high-allele frequency and could be readily identified, even by visual inspection of silver-stained gels. Eight markers can be eliminated from the subsequent analysis. Two-stage pooling was used to type the remaining markers. Ultimately, the *xl(t)* gene was delimited to a 100-kb region between markers RM7283 and ID3. This worked example shows that DNA pooling can improve genotyping efficiency in the preliminary screen for candidate markers and recombinants.

## Discussion

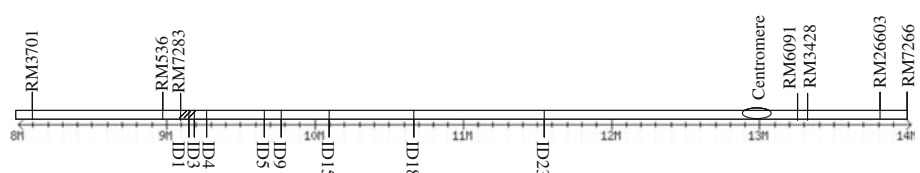
Gene mapping plays an increasingly pivotal role in deciphering the genetic architecture of inherited traits and gene pyramiding for breeding. This course usually involves an intensive genotyping for a number of molecular markers and/or individuals. Although the genotyping cost is decreasing with improved techniques, it remains important to pursue better strategies for a further reduction of genotyping. Selective genotyping (genotypically selected sample, e.g., recombinants), selective sampling (phenotypically selected sample, e.g., that from a single tail of the phenotypic distribution or both the upper and lower tails), and DNA pooling are cost-effective procedures, but in separate use in the literature (Darvasi and Soller 1994; Korol et al. 2007; Ronin et al. 2003, 1998; Sen et al. 2005; Sham et al. 2002; Thaller and Hoeschele 2000; Vision et al. 2000; Xu et al. 2005). We propose in this communication to combine DNA pooling with selective genotyping and sampling to

rapidly and efficiently screen for informative markers and individuals and to achieve additional saving in time and cost.

The combined use of DNA pooling and SRG can potentially reduce the experimental cost when the genotyping cost is not negligible. Although the proposed method requires an additional cost in DNA quantification and pool construction, such a cost is almost one-time for all markers genotyped. We only need to quantify DNA concentration one time for all individual extract, and then put DNA extract together to generate pools according to our needs, e.g., constructing recombinant pools once for BA marker screening. For two-stage pooling in identifying recombinants and in genotyping recombinants, given a wide range of pool size shown in Fig. 1, usually the pools can be used for multiple times. Furthermore, although multi-stage genotyping is more time-consuming, the extra cost for genotyping the pool(s) is relatively smaller as compared with the genotyping burden required by the customary complete genotyping in a regularly equipped laboratory (i.e., high-throughput genotyping not available). Thus, it is possible for the proposed method to achieve an additional saving relative to the traditional SRG.

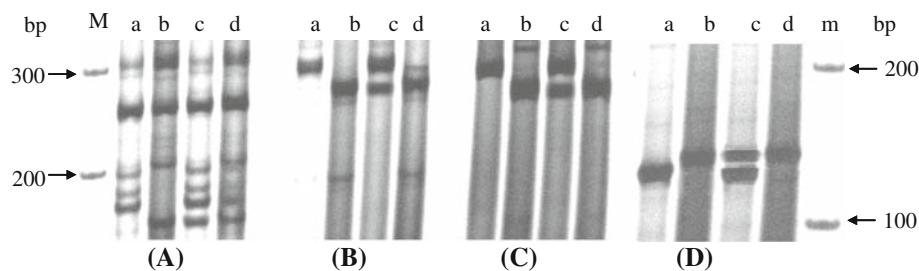
Although less obvious relative to monogenic traits, the application of the proposed procedure to QTL fine mapping is straightforward in principle as each of these genes segregates in a standard Mendelian manner. Fine mapping of a QTL can be performed with the same population in which primary mapping was carried out. Mendelization of QTLs (i.e., making polygenic traits oligogenic) is a more effective approach to zero in on the causative genes because the target QTL becomes the major genetic source of variation in the absence of other segregating QTLs. Several experimental strategies have been proposed for QTL Mendelization through use of specialized segregating populations such as chromosome substitution lines (consomic lines), chromosome segment substitution lines, NILs (congenic strains), recombinant inbred intercrosses, interval-specific congenic lines and others (Darvasi 1998; Flint et al. 2005; Nadeau and Frankel 2000; Salvi and Tuberosa 2005; Shalom and Darvasi 2002; Yano 2001; Zou et al. 2005). Whether with the original population or with a Mendelized population, the required sample size for fine mapping may usually be much larger than that in single gene case not only for a sufficient number of recombination events at the

**Fig. 4** Genomic positions of the markers in fine mapping. The hatched segment between markers RM7283 and ID3 shows a plausible region in which the *xl(t)* gene is located



**Table 1** InDel markers designed for delimiting the *xl(t)* locus

InDel <sup>a</sup>	Forward primer (5'–3')	Reverse primer (5'–3')	Start location (bp)	Amplicon size (bp) <sup>b</sup>
ID1	TGTAGGTCTTGCACAGGC	TGGCAGGAAACACTCATAG	9165591	132
ID3	GGCAAGACTCCCGAAGA	TTTTGAAAGTGCAAGAAGG	9201076	104
ID4	CAAGAACCGTAATGTAAC	CAAATGTGTTGGCACTTTGA	9274678	218
ID5	TTGAAAGTTAAGCACCTTATTG	CGGAGTGTCTATGGGAAAA	9673120	319
ID9	CCAAAGCCAACCAAAAAG	TGATACGGGATGAGGAATA	9772808	178
ID15	ATTACTATCGTCGCCAACCC	GCACTACATACGAATCAAACCTG	10111245	266
ID18	TCTCGCTGTTTGTACCTC	CTGATGCTATGGGCTTCT	10669808	346
ID23	TACTTCTAAAGCTGACGGATCT	GACTATTGTACTCGCTCTAACT	11573808	272

<sup>a</sup> Annealing temperature: 52°C<sup>b</sup> PCR product size for Nipponbare**Fig. 5** Polyacrylamide and agarose gel electrophoresis of two parents, F1 and DNA pool of 26 recombinant individuals for three SSR markers and one InDels marker, respectively. Lanes *M* and *m* are DNA ladders for sizing that have been loaded with fragments of known size in base pair, respectively, in which *M* corresponds to *a*, *b*, and *c*, and *m* corresponds to *d*; lanes *a*, *b*, *c*, and *d* represent Nipponbare (the *japonica* variety OS-lpa-XS110-2 not shown because

it has the identical band pattern), Huangyu B, F1 of Nipponbare × Huangyu B, and DNA pool of 26 recombinant individuals identified in coarse mapping, respectively. *a*, *b*, *c*, and *d* are for SSR markers RM3701, RM536, RM7283, and InDels marker ID3, respectively, in which there are 22, 15, 4, and 1 recombinants, respectively

short chromosome segment harboring the QTL but also for an increased precision through replication and progeny testing. Thus, the saving achieved by the proposed strategy is still substantial in most reasonable QTL mapping scenarios, although the exact efficiency is context-dependent, related to the QTL effect size, interval width, density of markers genotyped, genetic distance away from two flanking markers, and measurement accuracy.

Moreover, complex traits are likely caused by the interplay of multiple genes. Once the interactive QTLs are located in multiple intervals, the proposed method can be extended to narrow down the multiple target intervals based on the fact that nearer markers are more likely to have a similar pattern in allele frequency to the target genes as a result of genetic hitchhiking. Specifically, two-stage pooling genotyping can be performed to determine multi-marker haplotypes and screen for recombinants between the flanking markers defining each target segment. BA can also be used to screen for the combinations of markers more correlated with the segregation mode of the phenotype of interest.

As mentioned earlier, the proposed combined strategy is based on the assumptions that the allelic frequencies can be

reasonably well estimated in BA and that any disparate marker alleles can be detected in two-stage pooling. BA has been extensively used in many linkage and association studies and the relevant technical issues have also been well explored (Barratt et al. 2002; Sham et al. 2002). Further, the role of BA is merely to provide a rough guidance for screening markers and/or planning the two-stage pooling in our approach, and thus the quantification is not necessarily highly accurate. The latter requires a 100% sensitivity of assay (i.e., the proportion of true positives in all true cases). As shown in Fig. 3, pooling with a size of 40 or less could yield an estimated 100% sensitivity for the electrophoresis-based detection technique, suggesting the validity of the proposed two-stage pooling. Another potential limiting factor for two-stage pooling is the specificity of assay, i.e., the probability of true negatives in all false cases (equal to 1 minus false-positive rate). Since the impact of false-positive findings is to increase the number of pools that will have to be assayed, two-stage pooling can allow for a reasonably small false-positive rate. In most scenarios of fine mapping two-stage pooling can reduce genotyping burden for a false-positive rate of <5%.

Further, improved quantification technologies such as real-time PCR densitometry analysis and mismatch amplification mutation assay (MAMA) (Brohede et al. 2005; Chen and Zarbl 1997; Glaab and Skopek 1999; Mattarucchi et al. 2005; Sham et al. 2002) enable highly accurate assessment of allele frequencies. In summary, at the currently technical level, both assumptions, at least approximately, hold true and the proposed strategy is workable.

**Acknowledgments** We thank Dr. Mark C. K. Yang for his helpful comments. This project was funded in part by the National Institutes of Health Grant R01 DA025095 and the National Science Foundation of China 30000097 to XYL, and the National Science Foundation of China 30571131 to QYS.

## References

- Andersson L, Georges M (2004) Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet* 5:202–212
- Ashikari M, Matsuoka M (2006) Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends Plant Sci* 11:344–350
- Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker-assisted selection in crop breeding—prospects and challenges. *Curr Sci* 87:607–619
- Barratt BJ, Payne F, Rance HE, Nutland S, Todd JA et al (2002) Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet* 66:393–405
- Brohede J, Dunne R, McKay JD, Hannan GN (2005) PPC: an algorithm for accurate estimation of SNP allele frequencies in small equimolar pools of DNA using data from high density microarrays. *Nucl Acids Res* 33:e142
- Broman KW (2005) The genomes of recombinant inbred lines. *Genetics* 169:1133–1146
- Chen ZY, Zarbl H (1997) A nonradioactive, allele-specific polymerase chain reaction for reproducible detection of rare mutations in large amounts of genomic DNA: application to human k-ras. *Anal Biochem* 244:191–194
- Chi XF, Lou XY, Yang MC, Shu QY (2009) An optimal DNA pooling strategy for progressive fine mapping. *Genetica* 135:267–281
- Christians JK, Senger LK (2007) Fine mapping dissects pleiotropic growth quantitative trait locus into linked loci. *Mamm Genome* 18:240–245
- Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans Roy Soc B-Biol Sci* 363:557–572
- Darvasi A (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet* 18:19–24
- Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* 138:1365–1373
- Dekkers JC (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J Anim Sci* 82(E-Suppl):E313–E328
- Durrett RT, Chen KY, Tanksley SD (2002) A simple formula useful for positional cloning. *Genetics* 160:353–355
- Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* 2:437–445
- Flint J, Valdar W, Shifman S, Mott R (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* 6:271–286
- Glaab WE, Skopek TR (1999) A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan) and mismatch amplification mutation assay. *Mutat Res* 430:1–12
- Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science* 298:2345–2349
- Haldane JBS, Waddington CH (1931) Inbreeding and linkage. *Genetics* 16:357–374
- Jander G (2006) Gene identification and cloning by molecular marker mapping. *Methods Mol Biol* 323:115–126
- Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM et al (2002) Arabidopsis map-based cloning in the post-genome era. *Plant Physiol* 129:440–450
- Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). *J Exp Bot* 49:1619–1623
- Korol A, Frenkel Z, Cohen L, Lipkin E, Soller M (2007) Fractioned DNA pooling: a new cost-effective strategy for fine mapping of quantitative trait loci. *Genetics* 176:2611–2623
- Korstanje R, Paigen B (2002) From QTL to gene: the harvest begins. *Nat Genet* 31:235–236
- Lukowitz W, Gillmor CS, Scheible WR (2000) Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. *Plant Physiol* 123:795–805
- Martin OC, Hospital F (2006) Two- and three-locus tests for linkage analysis using recombinant inbred lines. *Genetics* 173:451–459
- Mattarucchi E, Marsoni M, Binelli G, Passi A, Lo Curto F et al (2005) Different real time PCR approaches for the fine quantification of SNP's alleles in DNA pools: assays development, characterization and pre-validation. *J Biochem Mol Biol* 38:555–562
- Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA* 88:9828–9832
- Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nat Genet* 25:381–384
- Peters JL, Cnudde F, Gerats T (2003) Forward genetics and map-based cloning approaches. *Trends Plant Sci* 8:484–491
- Remington DL, Ungerer MC, Purugganan MD (2001) Map-based cloning of quantitative trait loci: progress and prospects. *Genet Res* 78:213–218
- Ribaut JM, Hoisington D (1998) Marker-assisted selection: new tools and strategies. *Trends Plant Sci* 3:236–239
- Ronin YI, Korol AB, Weller JI (1998) Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis. *Theor Appl Genet* 97:1169–1178
- Ronin Y, Korol A, Shtenberg M, Nevo E, Soller M (2003) High-resolution mapping of quantitative trait loci by selective recombinant genotyping. *Genetics* 164:1657–1666
- Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci* 10:297–304
- Sen S, Satagopan JM, Churchill GA (2005) Quantitative trait locus study design from an information perspective. *Genetics* 170:447–464
- Shalom A, Darvasi A (2002) Experimental designs for QTL fine mapping in rodents. *Methods Mol Biol* 195:199–223
- Sham P, Bader JS, Craig I, O'Donovan M, Owen M (2002) DNA pooling: a tool for large-scale association studies. *Nat Rev Genet* 3:862–871
- Tanksley SD (1993) Mapping polygenes. *Annu Rev Genet* 27:205–233
- Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends Genet* 11:63–68
- Thaller G, Hoeschele I (2000) Fine-mapping of quantitative trait loci in half-sib families using current recombinations. *Genet Res* 76:87–104



- Varshney RK, Hoisington DA, Tyagi AK (2006) Advances in cereal genomics and applications in crop breeding. *Trends Biotechnol* 24:490–499
- Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD (2000) Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics* 155:407–420
- Wang J, Koehler KJ, Dekkers JCM (2007) Interval mapping of quantitative trait loci with selective DNA pooling data. *Genet Select Evol* 39:685–709
- Xu ZL, Zou F, Vision TJ (2005) Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. *Genetics* 170:401–408
- Yano M (2001) Genetic and molecular dissection of naturally occurring variation. *Curr Opin Plant Biol* 4:130–135
- Yano M, Sasaki T (1997) Genetic and molecular dissection of quantitative traits in rice. *Plant Mol Biol* 35:145–153
- Zhou XS, Shen SQ, Wu DX, Sun JW, Shu QY (2006) Introduction of a xantha mutation for testing and increasing varietal purity in hybrid rice. *Field Crops Research* 96:71–79
- Zou F, Gelfond JA, Airey DC, Lu L, Manly KF et al (2005) Quantitative trait locus analysis using recombinant inbred intercrops: theoretical and empirical considerations. *Genetics* 170:1299–1311