

Positional Cloning in *Brassica napus*: Strategies for Circumventing Genome Complexity in a Polyploid Plant

Gregory G. Brown and Lydiane Gaborieau
*McGill University, Montreal QC,
Canada*

1. Introduction

Positional, or map-based, cloning provides a strategy for isolating genes of interest when there is little or no information available on the molecular characteristics of the gene or its products, i.e. the RNA or protein that it specifies. Instead, the strategy relies on the identification of the chromosomal location of the gene through genetic mapping with molecular markers. Molecular markers, in general, allow for the detection of polymorphisms between the DNA sequences of different individuals of a species. By following the segregation of such polymorphisms through genetic crosses it is possible to construct genetic maps of a species' chromosomes. The first such markers used for this purpose were restriction fragment length polymorphisms (RFLPs, Botstein et al., 1980), in which DNA sequence differences are detected through their capacity to influence the size of fragments generated by restriction endonucleases. Subsequently, a wide range of marker types have been developed, most of which rely on the polymerase chain reaction. These include simple sequence repeat (SSR), or microsatellite, polymorphisms (Hearne et al., 1992) amplified fragment length polymorphisms (AFLPs, Vos et al., 1995) and single nucleotide polymorphisms (SNPs, Brookes, 1999).

Once chromosomal markers positioned on each side of the gene of interest are identified, DNA clones spanning the interval between these are recovered from genomic libraries. The size of the chromosomal interval between flanking markers that must be cloned is dependent on the resolution of the genetic map in the region of interest. This, in turn, depends both on the density of the molecular markers available for the region and the size of, and extent of genetic polymorphism within, the genetic populations employed for mapping.

Until the early 1990s, chromosome walking was usually required to clone specific genomic intervals. This process involved successive rounds of library screening and assembling the recovered clones into progressively larger sets of overlapping fragments (contigs) until the entire interval between the flanking markers was spanned. The development of new methods for cloning larger fragments of DNA, however, especially the development of bacterial artificial chromosomes (BACs, Shizuya et al., 1992) has greatly simplified and to some extent eliminated that need for chromosome walking. For most plant species, a population of several thousand individuals is sufficient to provide the mapping resolution

needed to reduce the size of the interval separating the nearest flanking markers to one that can be contained within one or few BAC clones.

To characterize a cloned chromosomal segment, the region is first sequenced and the sequence analyzed with program such as GENSCAN (Burge and Karlin, 1997) to identify potential protein coding sequences. For plants, introduction of subclones containing the various potential genes into an appropriate recipient plant via genetic transformation (e.g. Moloney et al., 1989; Brown et al., 2003) provides the most effective means of gene identification. Other methods for gene identification can involve higher resolution mapping and/or sequencing corresponding intervals from genetic strains lacking the gene of interest. The overall strategy of map-based cloning was devised on the premise that the target genome was a conventional diploid, with the target individual or line being homozygous for the gene of interest. Many plant genomes, however, including those several economically important crops, are polyploids in which most chromosomal segments are present in several paralogous copies (Adams and Wendel, 2005). This adds a layer of complexity to map-based cloning projects, since it necessitates the development of an additional strategy for identifying which of the several paralogous regions or clones that may be isolated from a genomic library represents the one that contains the gene of interest. Cytoplasmic male sterility (CMS) is a widespread trait in flowering plants that is specified by novel, often chimeric genes in the maternally inherited mitochondrial genome (Chase, 2007). The trait can be suppressed by nuclear restorer of fertility (Rf) genes that act to specifically down-regulate the expression of the corresponding novel, CMS-specifying, mitochondrial gene. The phenomenon of CMS and nuclear fertility restoration is of commercial interest because it can be used for the production of higher yielding hybrid crop varieties (Bonen and Brown, 1993), and of fundamental biological interest because it represents a novel evolutionary process termed an “intragenomic arms race” that has apparently been occurring throughout much of angiosperm evolutionary history (Budar et al., 2003; Fuji et al., 2011).

Out of an interest in characterizing nuclear restorer genes for CMS in the oilseed rape, or canola plant *Brassica napus*, we developed strategies for employing map based cloning approaches for the complex, polyploid genome of this plant. In the sections below we briefly discuss the architecture of genomes for the genus *Brassica* in general, and *Brassica napus* specifically, and our approaches for circumventing problems posed by the genomic complexity this presents.

2. *Brassica* crops

Plants of the genus *Brassica* comprise an exceptionally diverse group of crops and include varieties that are grown as oilseeds, vegetables, condiment mustards and forages. The cytogenetic and evolutionary relationships among the major oilseed and vegetable species are commonly depicted as U’s triangle, named after the Korean scientist who first formulated it (U, 1935). U speculated that *B. carinata*, *B. juncea* and *B. napus* are each allotetraploids formed by interspecific hybridization events between the parental diploid species *B. nigra*, *B. rapa* and *B. oleracea*. Thus, hybridization between *B. nigra* and *B. rapa* resulted in the formation of *B. carinata*, and between *B. nigra* and *B. oleracea* in the formation of *B. juncea*. *B. napus* is a hybrid between *B. oleracea* and *B. rapa*. The relationships among these species first postulated by U have since been confirmed by a large variety of genetic and molecular analyses. The haploid genomes of *B. rapa*, *B. nigra* and *B. oleracea* are

designated A, B and C respectively. Thus diploid *B. rapa* has two copies of the A genome within 20 chromosomes (AA, $n=10$, $2n=20$) and diploid *B. napus* has two copies of both the A and C genomes within 38 chromosomes (AACC, $n=19$, $2n=38$).

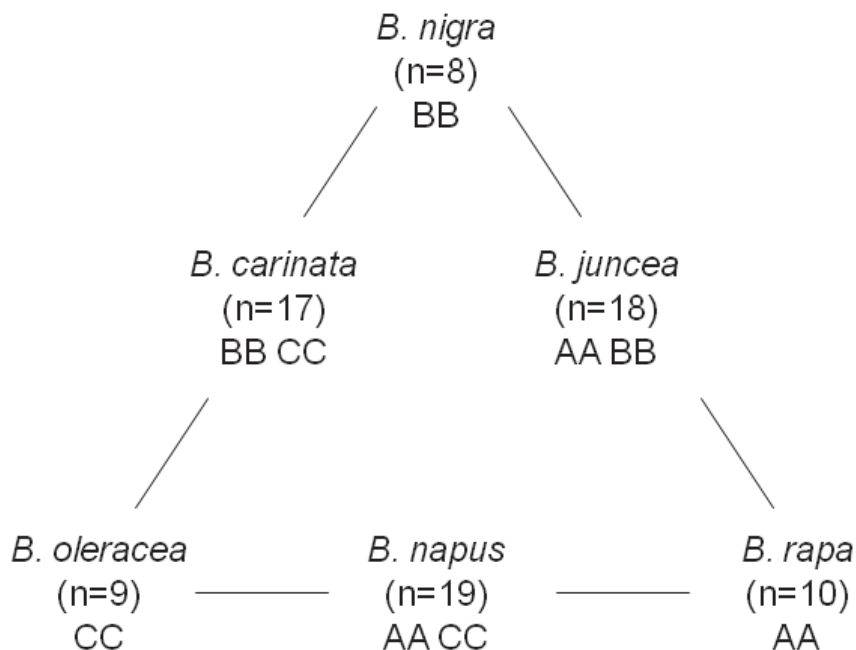


Fig. 1. U's triangle representation of the relationships between the diploid Brassica species *B. nigra*, *B. rapa* and *B. oleracea* and the allotetraploid species *B. carinata*, *B. juncea* and *B. napus*. The haploid genomes of the diploid species of *B. rapa*, *B. nigra* and *B. oleracea*, are referred to as A, B and C respectively. Thus diploid cells of *B. rapa* contain two copies of the A genome, and diploid cells of *B. napus* contain two copies of the A and C genome.

2.1 Brassica genomes

The model plant *Arabidopsis thaliana* and the Brassica species belong to the same plant family, the Brassicaceae. Initial efforts at determining the relationships between the Brassica and Arabidopsis genomes involved using molecular probes for co-linear sets derived from the developing Arabidopsis genomic resources to map RFLP polymorphisms in Brassica species (Kowalski et al., 1994). These studies indicated that there is extensive co-linearity between the Brassica and Arabidopsis genomes, but that most single copy Arabidopsis regions exist as multiple, on average 3 copies in modern Brassica genomes (Lagercrantz et al., 1995; Sadowski et al., 1996; Cavell et al., 1998). This in turn, gave rise to the hypothesis that the modern diploid Brassica species are derived from a hexaploid ancestor whose genome was generated from a diploid, Arabidopsis-like genome through polyploidization events. This view has largely been confirmed through subsequent comparative mapping studies involving much larger numbers of markers whose position was known on the sequenced Arabidopsis genome (Parkin et al., 2005). The latter study allowed for the identification of a

large number of segments of the *B. napus* genome that are co-linear with corresponding regions of *Arabidopsis*. The average length of these co-linear segments was 14.3 cM in *Brassica*, corresponding to 4.3 Mb in *Arabidopsis thaliana*, suggesting that, on average, 1 cM genetic distance in *B. napus* corresponded to 300 kb in physical distance in *Arabidopsis thaliana*. Similar high definition mapping experiments in *B. rapa* indicate that 1 cM in this species corresponds to 341 kb in *Arabidopsis* and thus a similar relationship between physical and genetic distances in the two species.

Recent years have witnessed large scale genome sequencing efforts targeting many crop species, including those of the *Brassica* species *B. oleracea*, *B. napus* and especially *B. rapa* (Town, et al., 2006; Yang et al., 2006; Cheung et al., 2008). On the basis of synonymous base substitution rates in orthologous protein coding sequences, such studies indicate that the *Arabidopsis* and *Brassica* lineages diverged roughly 17 Mya and that the complex structure of the *Brassica* genomes resulted from polyploidization events occurring roughly 14 Mya. In addition, these studies confirmed the extensive co-linearity between the *Brassica* and *Arabidopsis* genomes of species evident from comparative genetic mapping, but further indicated the occurrence of segmental duplications, interspersed deletions and the occasional insertion of noncolinear genes or gene fragments. Comparison of orthologous A and C genome segments of *B. oleracea*, *B. rapa* and *B. napus* has indicated that the A and C genomes diverged about 3.7 Mya. The timing of the hybridization event between the A and C genomes that gave rise to *B. napus* has proven more difficult to determine, since the rates of silent base substitution between the parental and *B. napus* orthologs varies among different regions, suggesting that modern *B. napus* was derived from multiple progenitor varieties bearing varying degrees of similarity to the sequenced *B. rapa* and *B. oleracea* regions used in these investigations (Cheung et al., 2008). Physical distances between corresponding orthologous sequences in *Arabidopsis* and *Brassica* chromosomes appear, in general, to be similar.

Our current view of the structure and evolution of the *Brassica napus* genome is illustrated in Figure 2. Following the divergence of the *Brassica* lineage from that leading to modern *Arabidopsis*, its genome underwent triplication of all or most chromosomal segments. For a given region in such an ancestral *Brassica* plant, each *Arabidopsis* gene (a, b, c) was present in 3 copies (a, a', a'', b, b', b'' etc.). Subsequent chromosomal rearrangements reduced the length of the conserved sequence blocks. In addition, as a result of the genetic redundancy of the triplicated genome, chromosomal deletions led to the loss of specific gene copies, illustrated in the Figure 2 as the loss of gene b in triplicated region 1 of both the A and C genomes. Subsequent to the divergence of the A and C genomes, additional gene loss (and duplications) has taken place in both genomes (illustrated by the loss of gene c_{A'} in region 2 of the A genome).

It is evident from this discussion, that a given "single copy" region of the *Arabidopsis thaliana* genome is present, on average, six times in the *B. napus* genome. Underlying this complexity, it is now widely accepted that more ancient genome duplication events occurred during the evolution of angiosperms (Blanc & Wolfe, 2004; Adams & Wendel, 2005), further increasing the complexity of *Brassica* genomes. Thus, to perform map-based cloning in *B. napus*, it is necessary to be able to distinguish which of the multiple copies of a given genome segment correspond to that linked to the gene of interest. In the sections below, we illustrate the strategy our group has successfully developed to achieve this.

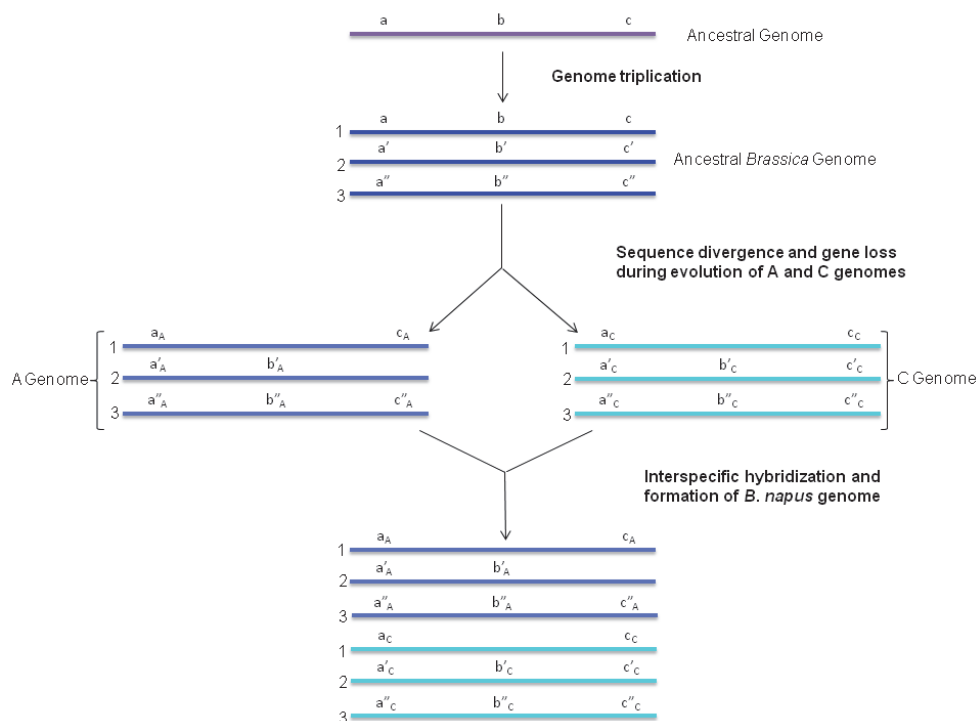


Fig. 2. Structure and evolution of co-linear regions of the *B. napus* genome. Polyploidization events occurring in an Arabidopsis like ancestor ~15 Mya resulted in a genome triplication. These three genomic regions underwent extensive independent rearrangement, sequence divergence (leading to the *a*, *a'* and *a''* etc. gene sequences for regions 1, 2 and 3 respectively) and sequence loss (symbolized by gene *b* from region 1) and duplication. Following the divergence of the A and C lineages, the different sub-genomes underwent additional evolutionary diversification, leading to additional rearrangements, sequence divergence and gene loss (symbolized by the loss of gene *c'* from region 2 of the A genome). Finally, the formation of *B. napus* by hybridization of the A and C genomes resulted in a genome in which many regions are present in six copies.

3. Map-based cloning in *Brassica napus*

The strategy we have employed to identify genes in *Brassica napus*, knowing only the phenotype they specify, is largely that developed by Tanksley and others for map-based cloning in other plant species, but with several specific modifications necessary for dealing with the highly duplicated nature and *Brassica* genomes and for taking advantage of the extensive co-linearity to between the *Arabidopsis* and *Brassica* genomes. The overall strategy is outlined in Figure 3. The gene in question is first mapped to a resolution of approximately 10 cM using molecular markers. Then a fine mapping stage is initiated that involves developing a large population segregating for the gene, and at the same time, developing flanking markers suitable for high-throughput, PCR based screening and developing a set of ordered genetic markers based on the co-linear region of the *Arabidopsis* genome. The gene is then positioned with respect to the ordered markers at a resolution of < 0.5 cM. The most closely positioned markers flanking the gene are then used to select one or more BAC clones that flank the gene. This is the most critical aspect of the process, since it is essential that the region selected is actually that spanning the target gene and not a homeologous chromosomal segment. Finally the BAC is sequenced, annotated and candidate genes are identified. Several approaches can be used to determine which of the candidate genes correspond to the target, the most rigorous of which is genetic transformation.

3.1 Rough genetic mapping

Our initial rough genetic mapping experiments (Jean et al., 1997; Li et al., 1998) were conducted using RFLP markers developed by a research team led by Benoit Landry, initially at Agriculture and Agri-Food Canada then at DNA LandMarks, a plant genetics research company that is now an operating unit of BASF Plant Science. These markers were based on probes selected from a cDNA library (Landry et al., 1991). Now a wide variety of mapped markers, especially SSR and SNP markers are available for this purpose, and necessary information relevant to the development of such markers is available through public resources (<http://Brassica.bbsrc.ac.uk/>; <http://Brassicadb.org/brad/>). In some cases, the *Arabidopsis* genome sequence coordinates corresponding to these markers are also known, which simplifies use of this information for the fine mapping analysis described below. Thus, rough mapping a single gene in *Brassica napus* is now a relatively simple and straightforward process.

Our efforts to map the nuclear restorer gene *Rfp*, for the *B. napus* 'Polima' CMS system, one of two CMS systems native to *B. napus*, were initiated in the early 1990s, when RFLPs were the only available marker system. We used Westar-*Rfp*, a derivative of the standard cultivar "Westar" into which the 'Polima' or *pol* male sterility conferring mitochondrial genome, as well as the corresponding nuclear restorer gene, *Rfp*, had been introgressed through a series of repeated backcrosses, as the male parent for our mapping population. The female parent for the population was Karat-*pol*, a derivative of the cultivar Karat in which the *pol* mitochondrial genome (but not the restorer) had been introgressed. We generated fertile F1 hybrid plants that were then crossed back to the female parental plant to generate a backcross population of approximately 180 individuals, roughly half of which, as expected, were male sterile due to the absence of the dominant *Rfp* allele, and the remaining half were fertile due to the presence of *Rfp*. We identified one RFLP marker, termed cRF1, that showed complete linkage to the *Rfp*

gene (i.e. all fertile progeny and none of the sterile progeny displayed the RFLP specific to the fertile parent; Jean et al., 1997). The cRF1 probe was then used to initiate a short chromosome walk with a cosmid library constructed from a doubled haploid *B. rapa* line into which the *pol* mt genome and *Rfp* gene had been introduced (Formanova et al., 2006). From these clones we were able to identify additional markers mapping close to *Rfp*, all of which were located on the same side of the gene. The next closest flanking marker on the opposite side of the gene was situated 10.8 cM away from *Rfp*.

A critical strategy for dealing with the complexity of the *Brassica* genome was developed at the stage in which we selected our initial clones for this chromosome walk using the cRF1 probe. First, the complexity of the genome we analyzed was reduced by transferring *Rfp* and the *pol* mitochondrial DNA, from the allotetraploid *B. napus* into the diploid *B. rapa* by genetic crosses and microspore culture. Even so, we found that clones selected with the cRF1 probe could fall into anyone of four contigs, Thus, the sequences homologous to the cRF1 probe were present in at least four genomic locations. We were able to determine which of the four locations corresponded to that linked to *Rfp*, however, because only for this contig did the size of the fragments hybridizing to the cRF1 probe match those of linked the *Rfp*-linked polymorphic fragments detected in genomic DNA with the same probe. This allowed us to “anchor” this cosmid, and the cosmids recovered by extending the walk, into the *Rfp* region. Subsequent analysis of the *Rfp* region, as outlined below, confirmed the linkage between this genomic region and the *Rfp* locus.

The *nap* CMS system, like the *pol* system, is also indigenous to *B. napus*. Interestingly, most *B. napus* cultivars possess the CMS-conferring *nap* mt genome, but are male sterile due to the presence of the corresponding *Rfn* nuclear restorer gene. A few natural cultivars, such as ‘Bronowski’ have been found that lack the *nap* mtDNA and *Rfn*; such cultivars possess the *cam* mt genome of *B. rapa* (formerly *B. campestris*), one of the diploid progenitors of *B. napus*. When such cultivars are crossed as males to other *B. napus* varieties and the resulting F1 plants are then backcrossed as females to ‘Bronowski’ plants, male sterility segregates 1:1 in the first backcross generation.

When we performed mapping experiments on populations derived from crosses such as that described immediately above, we discovered that the *Rfn* restorer locus maps to a position indistinguishable from that of *rfp*. From this, and related experiments on the different effects the two genes have on mt transcript profiles, we postulated that they represent different alleles or haplotypes of a single nuclear locus (Li et al. 1998).

Our strategy for fine mapping genes, as outlined below is based, in part, on exploiting the co-linearity between the *Arabidopsis* and *Brassica* genomes. This can be done most effectively if rough mapping allows the genomic region that could potentially encode the gene to be narrowed to a single co-linear *Arabidopsis* region. As mentioned above, the average length of *B. napus* genome segments that are co-linear with the *Arabidopsis* genome is 14.3 cM. Normally, this degree of resolution can be achieved with mapping populations that allow ~200 meioses to be surveyed, but this is highly dependent on the density of available markers, the frequency of crossovers in the target region, and of course, the extent of co-linearity with *Arabidopsis*, which varies significantly from target region to target region. Given the sequence of linked polymorphic markers (e.g the cDNA sequence of an RFLP probe or the sequence of a linked SSR amplicon) the corresponding genomic location in *Arabidopsis* can be determined using readily available software such as BLAST (Altschul et al., 1990; www.blast.ncbi.nlm.nih.gov/).

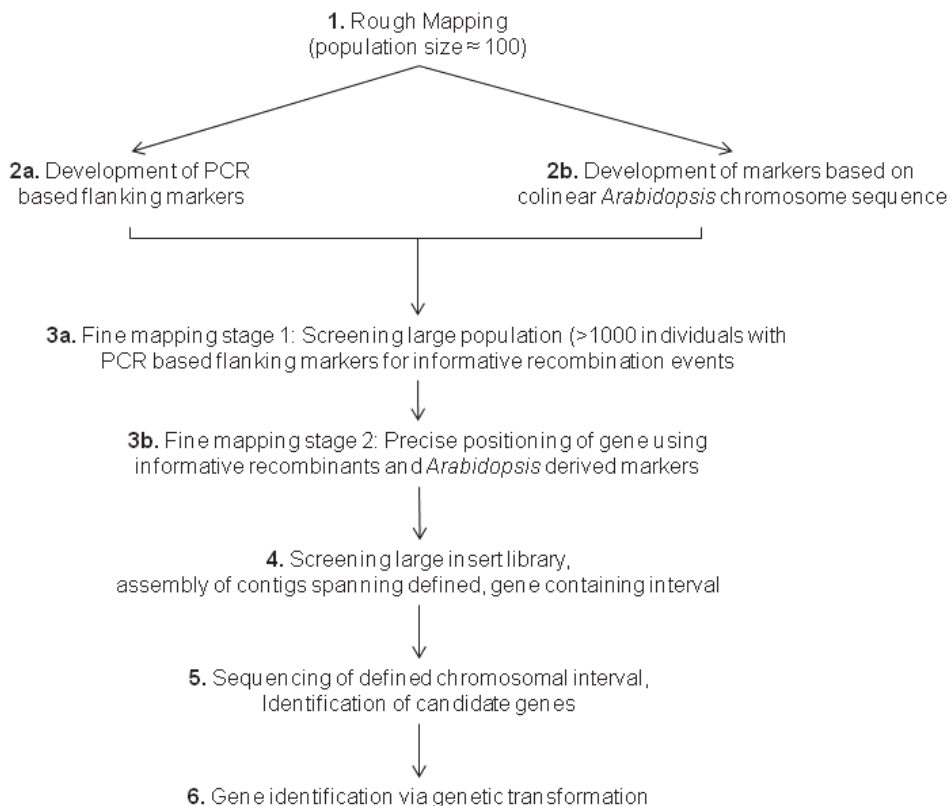


Fig. 3. Strategy for map-based cloning in *Brassica napus*. Following rough mapping with molecular markers (1), markers most closely flanking the target gene are converted, if necessary, into markers that can be used for high-throughput screening of small DNA preparations from a large number of plants (2a) and the sequences of the flanking markers are used to select corresponding allelic sequences in order to “land” in the target and identify the corresponding region or regions of the *Arabidopsis thaliana* genome. Additional markers are then derived from the *Arabidopsis* interval (2b) that will allow more precise positioning of the target gene. A large population of young plants is then screened with the flanking markers (3a) and plants with a different allele configuration between the two markers are raised to maturity, scored for phenotype and scored for genotype using markers derived from the corresponding interval of the *Arabidopsis* genome (3b). Markers mapping closest to the target gene are used to select clones from a large insert DNA library that contain the expected allele (4). The selected clones are then sequenced, the sequence is annotated and candidate genes are identified using criteria described in the text. Finally, the identity of the target gene is determined by genetic transformation and phenotype assessment.

3.2 Fine genetic mapping

Fine genetic mapping is usually accomplished in two stages, the first of which involves raising a large (>1000 individuals) genetic population segregating for the gene in question, then screening that population with PCR-based markers for individuals containing a chromosome that has undergone recombination in the region close to the target gene and which are therefore informative for more precise localization of that gene. If, as is normally now the case, the original mapping was performed using PCR-based markers such as SSRs or SNPs, no additional marker conversion is necessary. If on the other hand, the initial mapping was performed with RFLPs or AFLPs, it is necessary to convert the most closely flanking markers into a form that can be easily interrogated before attempting to screen the population.

3.2.1 Development of PCR based flanking markers

For fine mapping *Rfp*, we already had genomic sequence corresponding to the RFLP probe regions on one side of the gene, and used this to design primers that amplified corresponding regions from plants homozygous for either the *Rfp* or *rfp* alleles. One such pair of primers amplified different sized products from these plants. We used bulked segregant analysis (BSA) to determine if the polymorphism detected between the parent plants of the population was indeed linked to the *Rfp* gene (Michelmore, et al., 1991). In this process DNAs from fertile and sterile plants of the mapping population are pooled prior to analysis. The principle of the process is that linked markers will differentiate between the separately pooled DNA samples, whereas unlinked markers will not. Because the pooled samples from the fertile plants amplified a product that was not observed in the amplification products obtained from sterile plants, we determined the marker was indeed linked to the *Rfp* gene. To identify an PCR-based marker that mapped to the opposite side of *Rfp*, we identified an AFLP that mapped appropriately, then sequenced the differentially amplified product from the fertile and sterile parents to convert the AFLP into a sequenced characterized amplified region (SCAR) polymorphism in which the same primer pair amplified different sized products from pools of DNA from fertile and sterile plants.

In summary, to design PCR based markers to facilitate analysis of the large genetic populations necessary for fine mapping, our overall approach has been to use RFLP probes to select corresponding sequences from a genomic cosmid library, ensure that the selected genomic clones correspond to the region linked to the gene of interest, then sequence these clones and design primers that detect polymorphisms (length polymorphisms, SNPs or cleaved amplified polymorphic sequences [CAPS, Konieczny & Ausubel, 1993]) linked to the gene. Fortunately, the numbers and density of publicly available marker sequences may now obviate the need for the development of such tools for many map based cloning projects in *Brassica napus* and its diploid progenitors. It is critical, however, to ensure that the flanking markers detect a least one recombination event between the marker and the target gene.

3.2.2 Development of markers based on co-linear *Arabidopsis* chromosome sequence

The second important aspect of our fine mapping strategy involves exploiting co-linearity between the *Arabidopsis* and *Brassica* genomes. In general, the strategy for rough mapping outlined above provides mapping resolution such that the closest markers flanking either side of the target gene will match sequences in the same *Arabidopsis* genomic region.

Assuming this is the case, one then simply can amplify coding sequences from the annotated *Arabidopsis* region using either *Arabidopsis* genomic DNA, or preferably the corresponding BAC DNA. Information on the annotation and BACs is available from the TAIR online resource (www.Arabidopsis.org) and the amplicons can be used as RFLP probes to analyze digests of DNA from the ~100 plants whose genomes have experienced informative recombination events. Normally, analysis of digests using three to four restriction endonucleases is sufficient that 10 to 20% of the tested probes will detect a linked polymorphism, although this again, will vary, depending on the extent of genetic difference between the parental lines and the degree of polymorphism within the target region. We have found that in general, the order of sequences in the *Arabidopsis* genome conforms with the genetic localizations of the detected polymorphisms in *Brassica* plants, although we have noted an exception to this rule (Formanova, 2010).

Due to the degree of genetic redundancy in the *B. napus* genome, a given RFLP probe may detect polymorphism between parental lines at multiple genomic sites, only one of which, and possibly none of which, are linked to the target gene. For this reason, it is essential that bulked segregant analysis (BSA) be performed for each probe to confirm that a detected polymorphism is linked to the target region before using it in mapping experiments with the subpopulation of informative recombinant genomes. An example of BSA for an RFLP linked to the *Rfp* gene is shown in Fig. 4, and other examples can be found in Formanova et al. (2010). In addition to using the *Arabidopsis* sequence as a source of ordered RFLP probes, the sequence may also be used to identify other marker types, particularly SNPs, mapping to the target region. For SNP identification we have selected *B. napus* cDNA and EST sequences by interrogating databases with coding sequences derived from the targeted region of the annotated *Arabidopsis* genome, then designed primers from the selected sequences that amplified corresponding regions from the parental line and bulked segregant DNA samples. For linked polymorphisms, sequence differences are evident in the products amplified from the bulked segregant DNA samples. In this manner we were able to identify one SNP that has provided the closest of flanking markers to the *Rfp* gene (Formanova et al., 2010) and that served as the key probe for recovering a BAC spanning the entire target region.

3.2.3 Screening a large mapping population with PCR based markers

A key advantage of using a two step approach to fine mapping the target gene is that the entire population need be grown only to the point where sufficient tissue can be obtained to permit a limited number of PCR based marker assays to be performed while maintaining the viability of each individual plant. For the vast majority of plants in such populations, the marker genotype detected by each of the closely linked markers will be the same. Other than the very small number of plants that may have experienced double crossovers in the region of interest, these plants will not be informative for more precise localization of the gene in question and can be discarded. The remaining plants can be grown to maturity and analyzed for both phenotype and genotype, as described below. Methodology for DNA preparation and analysis by PCR based probes can be found in Formanova et al. (2010).

3.2.4 Precise gene positioning using informative individual and *Arabidopsis* derived markers

Once the markers and informative recombinants have been identified, the actual mapping of the gene is very straightforward. For example, a backcross population segregating for the

dominant nuclear restorer of fertility gene *Rfp*, will consist of sterile plants, lacking *Rfp* and therefore homozygous for the recessive allele *rfp*, and fertile plants that are heterozygous, i.e. contain a copy of each allele *Rfp/rfp*. The further the marker is from the *Rfp* locus, the larger the number of recombination events observed between the marker and the gene. For example, male-sterile plants in this population will be homozygous for the *rfp* allele. If a recombination event has occurred within an individual's chromosome at a site between a marker and the *Rfp/rfp* locus, then that individual will score as heterozygous for the marker alleles derived from the two parents of the initial cross. A more complete example of marker scoring and fine mapping can be found in Formanova et al. (2010).

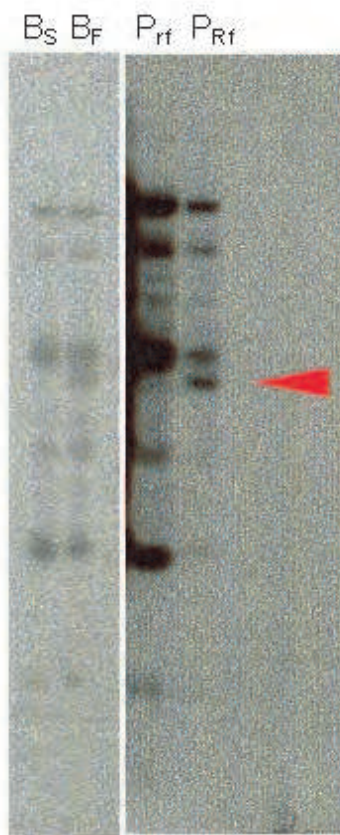


Fig. 4. An example of the use of bulked segregant analysis to demonstrate linkage between a restriction fragment length polymorphism detected between the parents of population designed for mapping population the nuclear restorer gene *Rfp* (P_{rf} and P_{Rf}) and DNA from 12 pooled male-sterile (sterile bulk or B_s) and 12 pooled fertile segregants (fertile bulk or B_f) of the population. Only linked polymorphisms (red arrow) will appear as differences between the samples pooled on the basis of phenotype. Clones selected from a genomic library that carry the linked polymorphic fragment represent the target gene region.

3.3 Screening a large insert library

In the case of the *Rfp* gene, the closest flanking markers we identified using a BC1 population of roughly 10^3 individuals corresponded to sequences on the *Arabidopsis* genome roughly 100 kb apart, which made it likely that the corresponding region of the *B. napus* genome might be captured within one or at most two overlapping BAC clones from a library with a mean insert size of 120kb (Parkin, I.A. unpublished data). It should be noted, however, that the sequence of the marker separated from cRF1 by 12.7 cM in our rough mapping analysis was located only 2 Mb from the position corresponding to cRF1 in the *Arabidopsis* genome, suggesting that the correspondence between genetic distance in *B. napus* and physical distance in *Arabidopsis* for this region of the genome might be considerably less than the 300kb estimated overall. It is suggested therefore, that larger fine mapping populations for future gene identification endeavors be constructed, if possible.

Our initial efforts at cloning the *Rfp* region were directed at assembling a contig of overlapping cosmid clones constructed from the library generated using the doubled haploid *B. rapa* line homozygous for *Rfp* described above (Formanova et al., 2006). The clone insert size in this library was quite small (~25 kb) and given that most probes recovered clones from at least 4 different genomic regions, the task of assembling contigs that could cover the target region proved more difficult than we initially envisioned. Instead, we adopted a different strategy and used the sequence of one cosmid clone, which contained the SNP polymorphism most closely linked to *Rfp*, to design primers to screen an ordered array of a *B. napus* BAC library generated from a *B. napus* homozygous for the *Rfn* allele. We were able to recover several clones from this library that amplified sequences from the library that matched the sequence of the cosmid over 1kb or more and, importantly, contained the expected SNP allele characteristic of the *Rfn* genotype. Thus the clone was anchored in the targeted *Rfn/Rfp* region.

A more generalized version of this strategy would be to generate a fosmid library from the targeted genotype, as we did for the cloning of the radish *Rfo* restorer gene (Brown et al, 2003). This library could then be used to recover clones corresponding to marker sequences by hybridization to colony lifts (Sambrook et al, 1989), which could then be sequenced; sequences of clones carrying linked polymorphisms could then be used to design primers useful for screening available BAC libraries.

3.4 Sequencing the defined chromosomal interval, identification of candidate genes

With the advent of very high throughput sequencing technologies, the sequencing of one or two BAC clones is a relatively straightforward process now best left to service providers. From the standpoint of the cloning effort it is recommended that very highly purified BAC DNA be provided. For *Brassica* genomic regions that have not yet been sequenced, it is recommended that dideoxy sequencing be used to construct a scaffold sequence. After shearing the BAC DNA this to an appropriate size, linkers are attached to individual fragments, and these are amplified and sequenced. Normally 5 to 10 fold coverage is sufficient. Alignment of the sequence contigs with the corresponding *Arabidopsis* sequence can be useful for determining the relationships among the contigs, which can then be used for designing primers that will allow amplification and sequencing of gaps between the contigs. At the time this chapter is being written the *B. rapa* genome sequencing project is well underway, and for many *A* genome regions it is possible that a reference sequence will

be available that will allow higher throughput, shorter read sequencing methodologies such as pyrosequencing to be employed.

Once a complete sequence for the genetically defined interval containing the target gene has been assembled, it is annotated with a bioinformatics tool such as GENSCAN (Burge and Karlin, 1997), which allows for the identification of open reading frames and likely sites of introns, polyadenylation signals and promoters. If some knowledge of the sequence characteristics of the target gene product are known, it is usually possible to quickly identify likely candidates using only the annotation output. For example, most nuclear restorer genes have been found to encode a subclass of P-type PPR proteins (Fuji et al., 2011), and these can be easily identified from an annotated BAC sequence. An additional strategy involve amplification of various gene sequences from genotypes that differ with respect to the allele of the candidate gene using a high fidelity thermostable DNA polymerase. Sequences showing allelic differences would obviously represent strong candidates for further investigation, although, in the case of *B. napus*, distinguishing alleles from closely related paralogs can proved to pose a significant challenge.

3.5 Gene identification via genetic transformation

Certain cultivars of *Brassica napus*, e.g. the spring variety “Westar”, are highly amenable to genetic transformation with *Agrobacterium tumefaciens*. Although explants of different plant tissues can be employed in such experiments, perhaps the most widely used protocol employs petiolar tissue of young cotyledons (Moloney et al., 1989). With this protocol, and using as a vector pRD400, which carries a wild-type kanamycin resistance gene, we have routinely been able to achieve transformation/regeneration efficiencies of >10% of tissue inoculants. Normally, some degree of selection of candidate genes should be attempted. However, this is not essential; in the case of the *Rfo* restorer gene, since we had no information as to the nature of the gene or its product, we successfully identified the gene after testing each predicted gene from a ~100 kb interval by evaluating the floral phenotype of over 100 transformed regenerants generated from 19 different genetic constructs. In our experience, transformation with genomic DNA has proven sufficient for generating the altered phenotype expected of the target gene, although in some cases increased expression using a cDNA and high-level expression vector may be necessary.

4. Conclusion

Positional cloning has been successful at cloning only a few genes thus far from plants of the genus *Brassica* and allied genera. However, the rapid development of high throughput markers and the accumulation of large amounts of genomic sequence, particularly in *Brassica rapa*, should greatly facilitate the application of the approaches described in this chapter. Nevertheless, because of the complexity of these genomes, researchers must be exceedingly careful to ensure that the regions they are analyzing are genetically linked to the target gene before further extending their analyses.

5. Acknowledgments

G.G.B. would like to thank the many collaborators who have helped develop approaches explained in this chapter. Particular thanks are owed to Benoit Landry, Natasa Formanova,

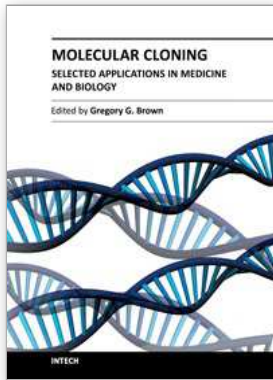
Wing Cheung, Martin Laforest, Prahant Patil, Xiu-Qing Li and Isobel Parkin. Financial support from the Natural Sciences and Engineering Council of Canada and BASF Plant Sciences are gratefully acknowledged.

6. References

- Adams, K.L. & Wendel, J.F. (2005). Novel patterns in gene expression in polyploid plants. *Trends Genet.* 21: 539-543.
- Altschul, S.F.; Gih, W.; Miller, W.; Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410
- Blanc, G. & Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667-1678.
- Bonen, L. & Brown, G.G. (1993). Genetic plasticity and its consequences: perspectives on gene organization and expression in plant mitochondria. *Can. J. Bot.* 71, 645-660.
- Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Brown, G.G.; Formanova, N.; Jin, H.; Wargachuk, R.; Dendy, C.; Patil, P.; Laforest, M.; Zhang, J.; Cheung, W.Y. & Landry, B.S. (2003). The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* 35: 262-72.
- Budar, F.; Touzet, P. & DePaepe R. (2003). The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* 117: 3-16.
- Burge, C. & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 266: 78-94.
- Cavell, A.C.; Lydiat D.J.; Parkin I.A.; Dean, C. & Trick M. (1998). Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* 41: 62-69.
- Chase, C.D. (2007). Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* 23: 81-90.
- Cheung, F.; Trick, M.; Drou, N.; Lim, Y.P.; Kwib, S.-J.; Kim, J.-A.; Scott, R.; Pires, J.C.; Paterson, A.H.; Town, C. & Bancroft, I. (2008). Comparative analysis between homeologous genome segments of *Brassica napus* and its progenitor species reveals extensive, sequence level divergence. *Plant Cell* 21: 1912-1928.
- Formanova, N.; Li, X.-Q.; Ferrie, A.M.R.; DePauw, M.; Keller, W.A.; Landry, B. & Brown, G.G. (2006). Towards positional cloning in *Brassica napus*: generation and analysis of doubled haploid *B. rapa* possessing the *B. napus pol* CMS and *Rfp* nuclear restorer gene. *Plant Mol. Biol.* 61: 269-281.
- Formanova, N.; Stollar, R.; Geddy, R.; Mahe, L.; Laforest, M.; Landry, B.S. & Brown, G.G. (2010). High resolution mapping of the *Brassica napus Rfp* restorer locus using *Arabidopsis*-derived molecular markers. *Theor. Appl. Genet.* 120: 843-851.
- Fuji S.; Bond, C.S. & Small, I.D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl. Acad. Sci. USA* 108:1723-1728.

- Jean, M.; Brown, G.G. & Landry, B.S. (1997). Genetic mapping of fertility restorer genes for the Polima cytoplasmic male sterility in canola (*Brassica napus* L.) using DNA markers. *Theor. Appl. Genet.* 95, 321-328.
- Konieczny, N. & Ausubel, F.M. (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4:403-410.
- Kowalski, S.D.; Lan, T.-H.; Feldman, K.A. & Paterson, A.H. (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order. *Genetics* 138: 499-510.
- Lagercranz, U.; Putterill, J.; Coupland G. & Lydiate, D. (1996). Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant J.* 9:13-20.
- Landry, B.S.; Etoh, T.; Harada, J.J. & Lincoln, S.E. (1991). A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. *Genome* 34:543-552.
- Li, X-Q.; Jean, M.; Landry, B.S. & Brown, G.G. (1998) Restorer genes for different forms of *Brassica* cytoplasmic male sterility map to a single nuclear locus that modifies transcripts of several mitochondrial genes. *Proc. Natl. Acad. Sci. USA* 95, 10032-10037.
- Michelmore R.W.; Paran, I. & Keseli, R.V. (1991). Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect makers in specific genomic regions using segregating populations. *Proc. Natl. Acad. Sci. USA* Vol. 88, pp. 9828-9832.
- Moloney, M.M.; Walker, J.M. & Sharma, K.K. (1989) High efficiency transformation of *Brassica napus* using *Agrobacterium* vectors. *Plant Cell Rep.* Vol. 8, pp. 238-242.
- Parkin, I.A.; Gulden, S.M.; Sharpe, A.G.; Lukens, L.; Trick, M.; Osborn, T.C. & Lydiate, D.J. (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Segmental structure of the *B. napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* Vol. 171, pp. 765-781.
- Sadowski, J.; Gaubier, P.; Delseny, M. & Quiros, C.F. (1996) Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. *Mol. Gen. Genet.* Vol. 12, pp. 298-306
- Sambrook, J.; Frisch, E.F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, N.Y; Cold Spring Harbor Laboratory Press.
- Shizuya, H.; Birren, B.; Kim, U.J.; Mancino, V.; Slepak, T.; Tachiiri, Y. & Simon, N. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* Vol. 89, pp. 8794-8797.
- Town, C.D. et al., (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveals gene loss, fragmentation and dispersal following polyploidy. *Plant Cell* Vol 18, pp. 1348-1359.
- Yang, T.J. et al. (2006). Sequence level analysis of the diploidization process in the triplicated FLC region of *Brassica rapa*. *Plant Cell* Vol. 18, pp. 1339-1347.

- U, N. (1935). "Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization". *Japan. J. Bot* Vol. 7, pp. 389–452.
- Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; Lee, T.V.D.; Hornes M *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* Vol 23, pp. 4407–4414.



Molecular Cloning - Selected Applications in Medicine and Biology

Edited by Prof. Gregory Brown

ISBN 978-953-307-398-9

Hard cover, 324 pages

Publisher InTech

Published online 12, October, 2011

Published in print edition October, 2011

The development of molecular cloning technology in the early 1970s created a revolution in the biological and biomedical sciences that extends to this day. The contributions in this book provide the reader with a perspective on how pervasive the applications of molecular cloning have become. The contributions are organized in sections based on application, and range from cancer biology and immunology to plant and evolutionary biology. The chapters also cover a wide range of technical approaches, such as positional cloning and cutting edge tools for recombinant protein expression. This book should appeal to many researchers, who should find its information useful for advancing their fields.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Gregory G. Brown and Lydiane Gaborieau (2011). Positional Cloning in Brassica napus: Strategies for Circumventing Genome Complexity in a Polyploid Plant, Molecular Cloning - Selected Applications in Medicine and Biology, Prof. Gregory Brown (Ed.), ISBN: 978-953-307-398-9, InTech, Available from: <http://www.intechopen.com/books/molecular-cloning-selected-applications-in-medicine-and-biology/positional-cloning-in-brassica-napus-strategies-for-circumventing-genome-complexity-in-a-polyploid-p>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.