

Microstructure of a *Brassica rapa* genome segment homoeologous to the resistance gene cluster on *Arabidopsis* chromosome 4

Keita Suwabe^{†1,2)}, Go Suzuki^{†3)}, Tsukasa Nunome²⁾, Katsunori Hatakeyama²⁾, Yasuhiko Mukai³⁾, Hiroyuki Fukuoka²⁾ and Satoru Matsumoto^{*2)}

¹⁾ Graduate School of Bioresources, Mie University, 1577 Kurima-machiya, Tsu, Mie 514-8507, Japan

²⁾ NARO Institute of Vegetable and Tea Science, 360 Kusawa, Ano, Tsu, Mie 514-2392, Japan

³⁾ Division of Natural Science, Osaka Kyoiku University, 4-698-1 Asahigaoka, Kashiwara, Osaka 582-8582, Japan

Genome evolution is a continuous process and genomic rearrangement occurs both within and between species. With the sequencing of the *Arabidopsis thaliana* genome, comparative genetics and genomics offer new insights into plant biology. The genus *Brassica* offers excellent opportunities with which to compare genomic synteny so as to reveal genome evolution. During a previous genetic analysis of clubroot resistance in *Brassica rapa*, we identified a genetic region that is highly collinear with *Arabidopsis* chromosome 4. This region corresponds to a disease resistance gene cluster in the *A. thaliana* genome. Relying on synteny with *Arabidopsis*, we fine-mapped the region and found that the location and order of the markers showed good correspondence with those in *Arabidopsis*. Microsynteny on a physical map indicated an almost parallel correspondence, with a few rearrangements such as inversions and insertions. The results show that this genomic region of *Brassica* is conserved extensively with that of *Arabidopsis* and has potential as a disease resistance gene cluster, although the genera diverged 20 million years ago.

Key Words: microsynteny, genome evolution, genome organization, genomic collinearity, BAC library.

Introduction

Arabidopsis thaliana is an excellent model species for dicots, with a small and relatively simple genome (125 Mb) and a diverse range of genetic resources. The completion of its genome sequence has yielded new insights into plant genetics and genomics (Arabidopsis Genome Initiative 2000). Because patterns of chromosomal collinearity have been identified between *Arabidopsis* and other flowering plants, the *Arabidopsis* genome sequences provide an excellent resource for identifying the genes controlling target phenotypes, *i.e.* physically and functionally conserved genes, which are known as orthologues and paralogues (Axelsson *et al.* 2001, Grant *et al.* 1998, Rossberg *et al.* 2001).

The genus *Brassica* is closely related to *Arabidopsis*. The genetic relations within *Brassica* have been well studied and are explained by “U’s triangle” (U 1935). The genomes of three diploid species, *B. rapa*, *B. nigra* and *B. oleracea*, are designated as A, B and C, respectively and those of three amphidiploids, *B. juncea*, *B. napus* and *B. carinata*, as AB, AC and BC, respectively (U 1935). Because natural hybridization both within and between species has created wide genetic and morphological variation, the remarkable diversity

within *Brassica* is an excellent source for understanding how genetic and morphological variation has developed during the evolution of plant genomes. Several studies have established chromosomal collinearity between *Brassica* and *Arabidopsis* (reviewed in Schmidt 2002, Schmidt *et al.* 2001) and the degree of synteny between them provides a good opportunity to study how genetic and morphological variation has developed during the evolution of the genome, including the endurance of certain genetic structures. Such investigation would lead to a better understanding of plant genetics and genomics, including molecular biological and physiological aspects which have emerged during evolution. More generally, it would provide insights into the organization and function of larger and more complex genomes in plants.

It is widely accepted that the diploid *Brassica* species evolved from a common hexaploid ancestor and that large-scale genome duplication (polyploidization) has occurred via evolution (Cavell *et al.* 1998, Parkin *et al.* 2002, Schmidt *et al.* 2001). Studies based on genetic mapping revealed conserved genomic segments with a similar conserved gene content and order between species. These observations were confirmed at the micro-level in selected chromosomal regions (O’Neill and Bancroft 2000, Quiros *et al.* 2001, Sadowski *et al.* 1996) and at the macro-level in the whole genome (Mun *et al.* 2009, Town *et al.* 2006). At the same time, complex genomic relationships have also been identified within *Brassica* and between *Brassica* and *Arabidopsis*.

Communicated by R. Ohsawa

Received August 11, 2011. Accepted March 28, 2012.

*Corresponding author (e-mail: ssmats@affrc.go.jp)

† These authors contributed equally to this work

Microsynteny analysis identified potential inversion, insertion and gene loss in *Brassica* homoeologous genomic segments relative to *Arabidopsis* homologous segments (O'Neill and Bancroft 2000, Rana *et al.* 2004). This supports the idea that fine-scale chromosome rearrangements in *Brassica* occurred both before and after divergence of species. So although comparative analysis should take prominence in plant genetics and genomics, chromosomal evolution between species complicates the picture.

This study focuses on the genome of *B. rapa*, a diploid species ($2n=20$) with the smallest genome (550 Mb) in *Brassica*. During research on clubroot resistance in *B. rapa*, we found three independent QTLs for resistance with distinct functions (Suwabe *et al.* 2006). Two of them, named *Crr1* and *Crr2*, have synteny with a small region of *Arabidopsis* chromosome 4. This region houses a disease resistance gene cluster (major recognition complex: MRC) in the *A. thaliana* genome. This similarity suggests that the genes for clubroot resistance originated from an MRC in a common ancestral genome and were subsequently distributed to the different regions they now inhabit in the process of evolution. We aimed here at characterizing the genomic microstructure for a potential disease resistance gene cluster in the *Brassica* genome through fine-scale genetic mapping and physical mapping, and we discuss the physical and genomic relationships of this region with respect to *Arabidopsis* MRCs. We also discuss characteristic features of this region by comparison with the draft genome sequence of *B. rapa*, which was recently reported (BRGSPC 2011).

Materials and Methods

Plant materials

Chinese cabbage doubled-haploid line G004 and parental line Nou 7 (A9709) were used as the parents for the mapping population and G004 was also used for BAC library construction. G004 is derived from a clubroot-resistant cultivar of European fodder turnip, 'Siloga' (Kuginuki *et al.* 1997). For fine mapping, an F_2 population of 1920 plants was obtained from a cross of A9709 \times G004 and F_3 seeds of each F_2 plant were obtained by bud self-pollination. All F_2 plants were grown in a greenhouse.

Fine mapping of *Crr1* genomic region

Our previous study revealed that *Crr1*, a major QTL for clubroot resistance in *B. rapa* located within a 1.6-cM region between SSR markers BRMS-173 and BRMS-088 in linkage group (LG) 7 (A8), had homology with the genomic region of MRC-H (Holub 1997, Speulman *et al.* 1998) on *A. thaliana* chromosome 4. This region houses three SNP/indel markers, BSA1, 2 and 7, which were designed by synteny with *Arabidopsis* (Suwabe *et al.* 2006). In addition, two CAPS markers, AT27 and BZ2-DraI, were designed by the same strategy. All markers cosegregated completely with BRMS-173/BRMS-088 in the F_2 population of 94 plants an-

alyzed before. For fine mapping of this region, we used these markers for genetic analysis in the F_2 population of 1920 plants. F_3 individuals of the recombinant F_2 plants were used for the clubroot resistance test to identify recombination between BRMS-173 and BRMS-088. The test for clubroot resistance and data evaluation are described in a previous report (Suwabe *et al.* 2003).

Preparation of high-molecular-weight genomic DNA

Young leaves were collected from G004 and immediately frozen in liquid nitrogen before being stored at -80°C . The nuclei were extracted as described by Suzuki *et al.* (2001). The extracted nuclei were embedded in agarose plugs and lysed twice in lysis buffer (100 mM EDTA, 1% Sarkosyl, 0.1 mg/mL proteinase K) for 24 h at 50°C . The plugs were rinsed with TE50 (10 mM Tris-HCl, pH 8.0, 50 mM EDTA) plus 2 mM phenylmethylsulfonylfluoride (PMSF) for 2 h at 50°C and re-rinsed 5 times with TE50 for 2 h at 4°C , and then stored at 4°C until use.

BAC library construction

Plug-embedded high-molecular-weight genomic DNA was partially digested with *Hind*III and size-fractionated by pulsed-field gel electrophoresis (CHEF DR II, Bio-Rad, Hercules, CA, USA), first at 6 V/cm with 90 s pulse time for 4 h and then at 6 V/cm with 6 s pulse time for 12 h. The gel slices containing DNA fragments of ~ 150 kb in size were excised and incubated with 10 U/g β -agarase I (New England Biolabs, Ipswich, MA, USA) for 2 h at 40°C . After incubation, the DNA solution was concentrated at 45°C and dialyzed through 0.025- μm nitrocellulose membrane (Millipore, Billerica, MA, USA) to a suitable concentration. The CopyControl pCC1BAC vector (Epicentre, Madison, WI, USA) was used as a cloning vector for ligation at the molecular ratio of 5 vectors to 1 insert. The ligation reaction followed the manufacturer's instructions (CopyControl BAC cloning kit). Aliquots of ligation solution were electroporated into *E. coli* TransforMax EPI300 cells (Epicentre) in a Gene Pulser II electroporation system (Bio-Rad) at 1.25 kV, 25 mF and 100 Ω . The electroporated cells were spread onto an LB plate containing 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol, X-Gal and IPTG and incubated for 24 h at 37°C . Each recombinant (white colony) was picked up and inoculated into 200 μL LB medium with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol and freezing buffer (Zimmer and Verrinder Gibbins 1997) in a 96-well microtiter plate. After incubation at 37°C overnight, the plates were stored at -80°C .

Screening and construction of BAC contig

BAC libraries were screened by a two-step PCR-based strategy with each anchor and BAC-end marker. First, 96 BAC clones were pooled together as a "bin" and plasmid DNAs were extracted from them by the standard alkali method. The first screening used 2-D scoring of 400 bins and the bins which showed a marker-specific amplification by PCR were selected. The individual positive clone in a bin

was determined by the second screening, in which the 96 individual clones were screened. BAC-end sequencing was conducted with T7 and Reverse universal primers in a model 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA) and the BAC-end specific primers were designed from the region to amplify products of ~200–500 bp. After the specific PCR amplification by the designed BAC-end primers was confirmed with the genomic DNA of G004, BAC contigs for the target region were screened and assembled by chromosome walking on the basis of overlaps between BAC clones by specific PCR amplification.

Comparison of genomic sequence between B. rapa and A. thaliana

The nucleotide sequences of each BAC end of *B. rapa* were aligned with *Arabidopsis* genome sequences by BLASTN search in the *Arabidopsis* database (TAIR: <http://www.arabidopsis.org>). Genes with high homology to *Arabidopsis* genes (threshold value of $E < 10^{-10}$) were regarded as homologous. Genes with lower homology ($10^{-10} < E < 10^{-5}$) were confirmed manually. In addition, to assess the conservation of disease resistance genes (*R*-genes) in the region, we selected 11 candidate genes from the corresponding region of *Arabidopsis* chromosome 4, designed specific primers, and compared nucleotide sequences between *B. rapa* and *A. thaliana*.

Results

Fine mapping around Crr1

For fine mapping of the 1.6-cM genetic region between BRMS-088 and BRMS-173, we genotyped three SNP/indel and two CAPS markers in an F_2 population of 1920 plants. This genetic region had been identified as syntenous with a central region of the long arm of *Arabidopsis* chromosome 4 (Suwabe *et al.* 2006). The region in *Arabidopsis* has been located within a disease resistance gene cluster (MRC-H), in which many *R*-genes with typical motifs such as nucleotide-binding sites (NBSs) and leucine-rich repeats (LRRs) lie close to each other (Holub 1997, Hulbert *et al.* 2001, Speelman *et al.* 1998). All markers were mapped between BRMS-088 and BRMS-173 as expected (Fig. 1A). Their order corresponded well with that in the *Arabidopsis* genome, indicating that this region is homoeologous between *Brassica* and *Arabidopsis*. The genetic distances between markers from BRMS-088 to BRMS-173 totaled 3.9 cM (Fig. 1A). The total value was 2.4 times that of a previous estimate in the F_2 population of 94 plants (Suwabe *et al.* 2006). In that previous population, this region showed a genetic deviation from the Mendelian ratio (1 : 2 : 1) that might not reflect the genetic distance accurately.

We found 147 recombinant F_2 plants in the genomic region between BRMS-173 and BRMS-088. Among them, F_3 seeds (populations) were obtained from 124 F_2 plants while another 23 F_2 plants did not produce enough seeds for the resistance test and the infection test was carried out among 64

F_3 populations. Except of one population, all plants in 24 F_3 populations, harboring resistance genotype in BSA7, showed resistance phenotype (Fig. 1B). In contrast, most of plants in 40 F_3 populations, harboring susceptible genotype in BSA7, showed susceptible phenotype. This indicates that *Crr1* is located around BSA7. In addition, the plants with resistance homozygous in BSA1 and susceptible homozygous in BSA2 segregated their phenotypes both in resistance and susceptible (the 5th group from upper in the graphical genotype in Fig. 1B), another minor effect was also predicted around BSA2. Thus *Crr1* for clubroot resistance is likely to consist of two genetic loci located in the region around BSA2–BSA7.

Assembly of BAC contig around Crr1

For physical mapping, we constructed a large-insert BAC library using *B. rapa* line G004. The library consisted of 38 400 clones. Sixty of these were selected randomly and cut for measurement of insert size by *NotI* digestion (Fig. 2). Insert sizes ranged from 7 to 138 kb, with an average of 67.4 kb. Despite our using DNA fractions of ~150 kb for library construction, some clones had smaller inserts and average size was lower than in earlier reports (O'Neill and Bancroft 2000, Rana *et al.* 2004), likely owing to entrapment of small DNA fragments in the agarose gel, which two-step size fractionation would avoid (Nakamura *et al.* 1997). Overall, our library would provide ca. 4.7-fold redundant representation of the *B. rapa* genome (Arumuganathan and Earl 1991).

Because *Crr1* was delimited to the region of around BSA2–BSA7 by the recombination test (Fig. 1B), we assembled a BAC contig of the 0.6-cM region between BSA2 and BSA7, using BSA2, AT27 and BSA7 as anchors (Fig. 1C). The anchors and BAC-end markers identified 2 to 11 clones, with an average of 5.3 clones per marker (Fig. 1C). However, in PCR analysis to confirm the cross-linking of each clone, some clones did not show any correlation with others that were screened with same marker. Those ambiguous paralogous or chimeric clones were excluded from analysis. Consequently, by screening with three anchor and six BAC-end specific markers, we identified a total of 28 clones in the region covering 0.6 cM. The genomic region between BSA2 and AT27 was covered by only one clone (289F12, 112.4 kb) and the region between AT27 and BSA7 was covered with six overlapping clones and the physical distance was estimated as ~599.9 kb.

Conservation of disease-resistance gene cluster between B. rapa and A. thaliana

To evaluate the microsynteny of *B. rapa* and *A. thaliana*, we used *Brassica* BAC-end sequences to seek homologous genes in the *Arabidopsis* genome. Several homologous genes on chromosome 4 were found in the resulting *Brassica* BAC contig (Fig. 3 and Table 1). The order of genes was well conserved between species, except for one inversion. The inverted region housed genes associated with other

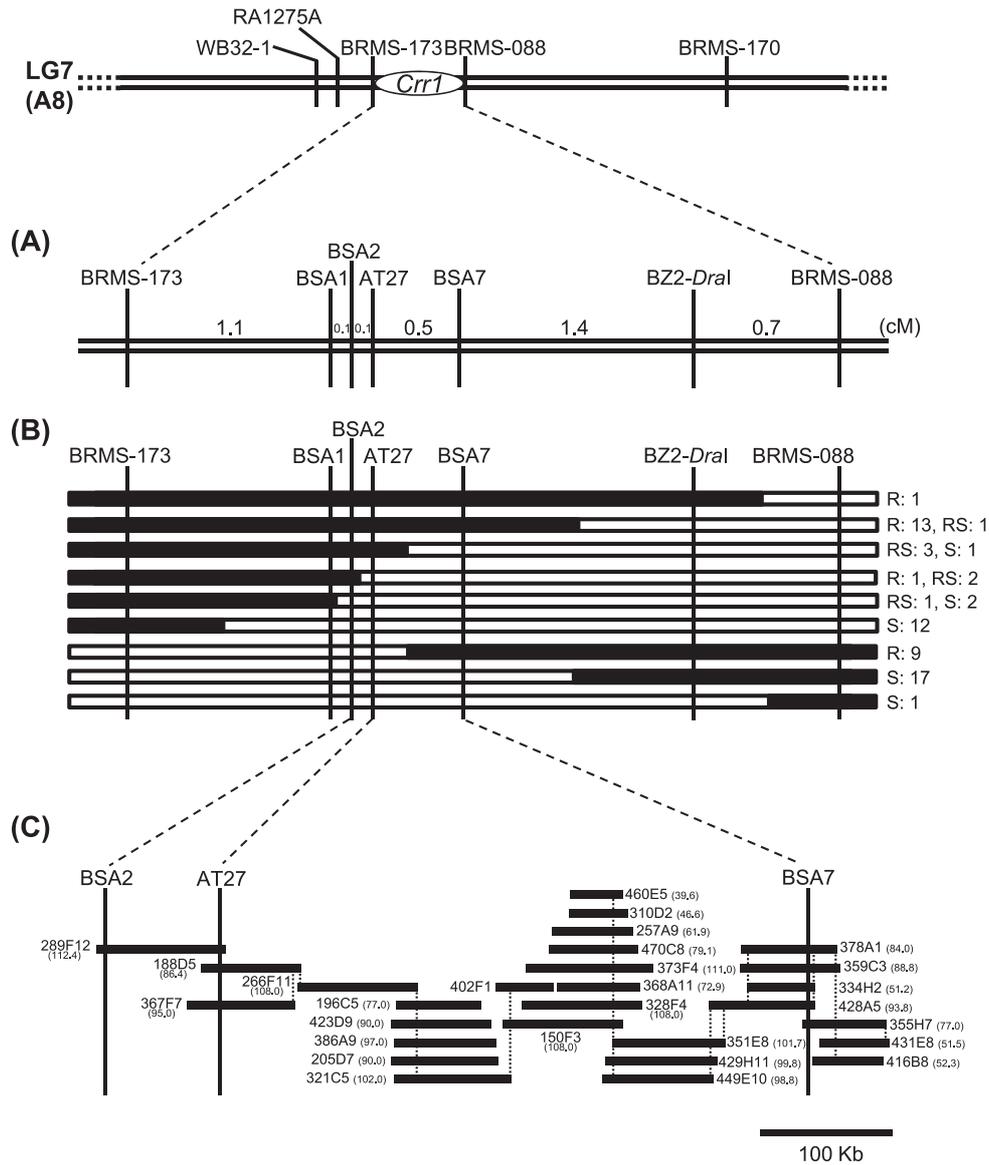


Fig. 1. Genetic and physical maps of the *Crr1* region. (A) Fine genetic map of the 1.6-cM region between BRMS-088 and BRMS-173. Three SNP/indel (BSA1, BSA2 and BSA7) and two CAPS markers (AT27 and BZ2-DraI) which co-segregated with BRMS-088/BRMS-173 in a previous F₂ population of 94 lines (Suwabe *et al.* 2006) were genotyped and mapped in an F₂ population of 1920 lines. (B) Graphical genotypes and phenotypes of 64 recombinant F₃ populations, in the region between BRMS-173 and BRMS-088. Black and white regions represent chromosomal segments homozygous for resistance and susceptible alleles, respectively. The number of F₃ populations in each phenotype is indicated on the right side of the graphical genotype. R: resistant population, RS: population segregating of resistant, susceptible and intermediate, S: susceptible population. (C) Physical map between BSA2 and BSA7. The 0.1-cM region between BSA2 and AT27 was covered by BAC clone 289F12 (112.4 kb). The 0.6-cM region between AT27 and BSA7 was covered by six overlapping clones covering ~599.9 kb. Insert sizes (Kb) of BAC clones, except 402F1, are provided in parentheses.

chromosomes of *Arabidopsis*, most of which encoded multi-copy genes or gene families such as those encoding F-box proteins, protein kinases, homeobox elements and retrotransposons. Because there was no continuous synteny with other chromosome regions, the insertions would have resulted from occasional events after diversification between *Brassica* and *Arabidopsis*.

To assess the conservation of *R*-genes in the region, selected 11 candidate genes were evaluated between *B. rapa*

and *A. thaliana*. The genes have characteristic motifs for *R*-genes, such as NBSs and LRRs, but the functions of some genes have not been characterized yet, so putative genes were predicted. Four of these genes were amplified from BAC clones which cover the BSA2–BSA7 contig: At4g19920 from BAC clone 289F12, At4g20270 and At4g20380 from 289F12 and 188D5 and At4g21450 from 321C5 and 150F3. The position of each gene corresponded with other homologous genes identified by comparison of

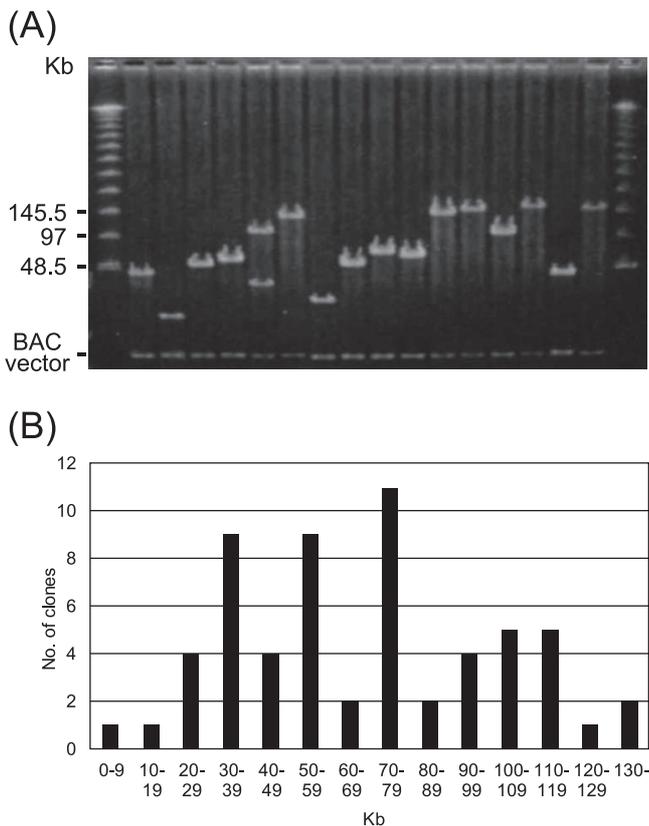


Fig. 2. Features of the *B. rapa* BAC library. (A) Representative electrophoretic gel: 60 randomly selected clones were digested by *NotI*, and insert size was evaluated. (B) Distribution of insert sizes in the BAC library.

the BAC-end sequence with the *Arabidopsis* genome (Fig. 1C, Fig. 3). These results indicated that this syntenous region originated from the region in the common ancestor genome which corresponded to part of *Arabidopsis* chromosome 4, although some rearrangements have occurred since diversification between *Brassica* and *Arabidopsis*. The *Brassica* genomic regions of BSA2 to AT27 (~112.4 kb) and AT27 to BSA7 (~599.9 kb) corresponded to *Arabidopsis* regions of 115 kb and 695.6 kb, respectively, indicating that these regions remain almost parallel between species.

Discussion

Complex structure of the *Brassica* genome

O'Neill and Bancroft (2000) reported that the 222-kb *Arabidopsis* genomic segment containing genes At4g17260–At4g17800, just upstream of our site on chromosome 4, was partially duplicated on chromosome 5 and triplicated in the *B. oleracea* genome, although extensive divergence of gene contents was evident. Those homologous segments of *B. oleracea* were 1.01 to 4.44 times the size of those of *Arabidopsis*, suggesting a different diversification of each after the divergence of *Brassica* and *Arabidopsis*. The genome triplication of diploid *Brassica* species is widely ac-

cepted, as seen in other diploid *Brassica* species (O'Neill and Bancroft 2000) and amphidiploid species (Parkin *et al.* 2005) and can be seen in several genomic regions (Park *et al.* 2005, Rana *et al.* 2004). Recently, the Multinational *Brassica* Genome Project (MBGP) consortium reported the draft genome sequence of *B. rapa*, which is equivalent to half of the whole genome, covering over 98% (283.8 Mb) of the gene space in the *B. rapa* genome (BRGSPC 2011). The MBGP confirmed the almost complete triplication of the *B. rapa* genome relative to *A. thaliana*; 91.13% (259.6 Mb) of the *B. rapa* genome assembly contained blocks that were collinearly syntenous with the *A. thaliana* genome. At the same time, however, there were significant disparities in gene loss across the triplicated blocks: in the least fractionated blocks, 70% of the genes were retained within the syntenous regions; in the moderately fractionated blocks, 46% were retained; while in the most fractionated blocks, 36% were retained. Thus the complexity of the chromosomal rearrangements that occurred both before and after the differentiation of *Brassica* species is clear and evolution has expanded the genome (Quiros *et al.* 2001). However, the region of focus on A8 (LG7) may be different. The physical sizes corresponding to BSA2–AT27 and AT27–BSA7 (Fig. 1) were estimated to be ~112.4 and ~599.9 kb, respectively, in *B. rapa* and 115 and 695.6 kb in *Arabidopsis*. No blocks showing continuous synteny with other chromosomes were found in the region in *B. rapa* (Fig. 3). These results indicate that this region is almost parallel between species, although *B. rapa* has a far larger genome than *Arabidopsis*. The average rate of expansion of a *Brassica* genome segment would be ~1.5 times that of *Arabidopsis* when we account for the genome size of each species and the triplication of the *Brassica* genome. Thus, the evolution of the *Brassica* genome is more complicated than simple triplication. In fact, the region on A8 and the *Crr2* region in A1 (LG6) overlap on chromosome 4 in *A. thaliana* (Suwabe *et al.* 2006), but the fine-scale genetic mapping and physical microstructure of the region reveal no genes in common (Fig. 3). This result is consistent with the observation that each *Brassica* homologous region has a different level of genome rearrangement (BRGSPC 2011, O'Neill and Bancroft 2000) and illustrates a non-uniform genome evolution in *Brassica*, with substantial gene retention and gene loss.

Disease resistance gene cluster in the *Brassica* genome

Brassica rapa and *A. thaliana* originated from a common ancestor and differentiated between about 20.4 and 14.5 million years ago (Koch *et al.* 2000, Yang *et al.* 1999). Recent studies that have gradually identified the genome structure of *Brassica* suggest that functionally conserved genes are maintained as genomic blocks. For example, the self-incompatibility (*S*) locus is inherited as a block in *Brassica* (Sato *et al.* 2002) and in *Arabidopsis* (Kusaba *et al.* 2001). The *S* locus contains at least two genes, *SRK* and *SP11/SCR*, which are the determinants of self-incompatibility in stigma and pollen (Schopfer *et al.* 1999, Takasaki *et al.* 2000,

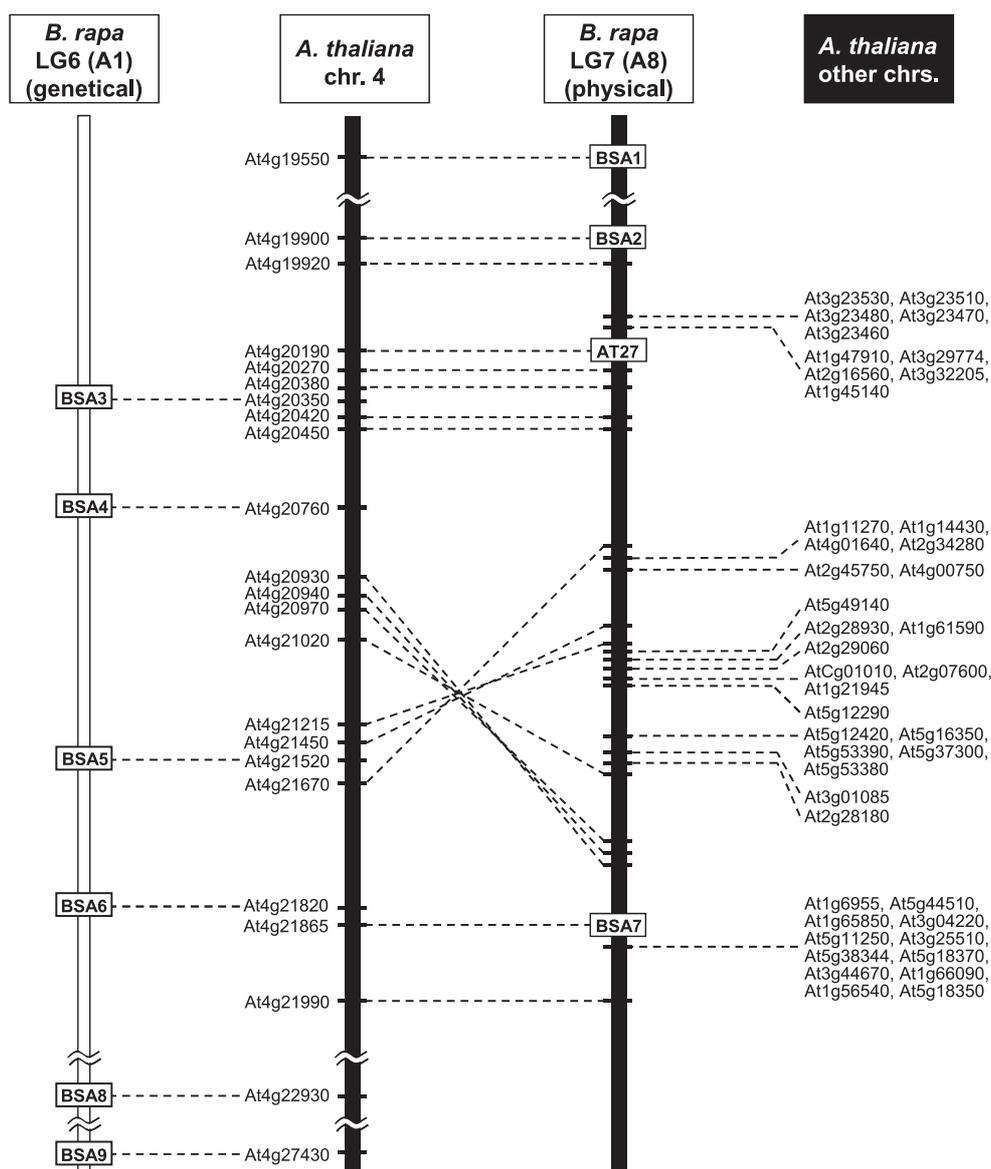


Fig. 3. Microstructure of syntenic *Crr1* region between *B. rapa* and *A. thaliana*. Homologous genes in the *Arabidopsis* genome were sought using *Brassica* BAC-end sequences. Homologous genes are connected by dashed lines. The *Crr2* region, another major QTL region for clubroot resistance (Suwabe *et al.* 2006), is aligned on the left of *Arabidopsis* chromosome 4.

Takayama *et al.* 2000). Although they are highly polymorphic, they are inherited as a unit and have maintained self-incompatibility through co-evolution. The genomic region on A8 is highly conserved between *Brassica* and *Arabidopsis* (Fig. 3 and Table 1). The disease resistance gene cluster (MRC-H) of which it is part includes many genes for resistance to bacterial, fungal, and viral pathogens. Resistance genes often reside as a cluster complex in the genome, because the clustering of multiple genes with different capabilities enables the plant to maintain its defenses against corresponding pathogens (Holub 1997, Wei *et al.* 2002). At the same time, it allows the generation of diversity by recombination, accelerating the evolution of novel specificities for resistance (Hulbert *et al.* 1997). Thus the region on

A8 shows genomic potential as an *R*-gene cluster in the *Brassica* genome. Further functional analyses, such as physiological and resistance analysis of each gene, will shed light on the evolution of *R*-gene clusters in the *Brassica* genome.

The region on A8 has a large inversion, corresponding to about 310 kb in the *Arabidopsis* genome, and insertion of genes homologous to those in other *Arabidopsis* chromosomes (Fig. 3). Gene deletion is also likely, although our strategy would not have shown it. The gene insertions are concentrated in a very small part of the inverted region, suggesting that a hot spot of genomic substitution occurs there. The suppression of genetic modification by the need to maintain function leads to an accumulation of repeat sequences and transposable elements (Fu *et al.* 2002). In

Table 1. BAC clones showing homology with *Arabidopsis* genes in the *Brassica* BSA2-BSA7 region

BAC clone	Side	Homologous gene(s) in <i>A. thaliana</i>	E-value
188D5	– ^a	At4g19920	2e-13
367F7	T7	At3g23530, At3g23510, At3g23480, At3g23470, At3g23460 (acyl-phospholipid synthase)	3e-37
289F12	R	At1g47910, At3g29774, At2g16560, At3g32205, At1g45140 (LINE)	2e-10
289F12	– ^a	At4g20270	6e-69
188D5	– ^a	At4g20380	6e-17
188D5	–	At4g20420	1e-10 ^b
188D5	R	At4g20450	3e-63 ^c
266F11	R	At4g21670	4e-52
423D9	T7	At1g11270, At1g14430, At4g01640, At2g34280 (F-BOX family protein)	1e-27
368A11	T7	At2g45750, At4g00750 (dehydration-responsive family protein)	1e-58
321C5	– ^a	At4g21450	7e-07
328F4	R	At4g21215	8e-16
373F4	T7	At4g21215	5e-02 ^d
321C5	R	At5g49140	8e-04
257A9	T7	At2g28930, At1g61590 (Protein kinase family protein)	5e-05
470C8	R	At2g29060	4e-06
310D2	T7	AtCg01010, At2g07600, At1g21945 (copia-like retrotransposon)	3e-64
460E5	R	At5g12290	9e-16
460E5	T7	At5g12420, At5g16350, At5g53390, At5g37300, At5g53380 (expressed protein)	2e-32
257A9	R	At3g01085	7e-07
470C8	T7	At2g28180	1e-21
373F4	R	At4g21020	3e-34
351E8	R	At4g20970	4e-11
359C3	R	At4g20940	1e-61
378A1	R	At4g20940	9e-75
334H2	R	At4g20930	9e-31
416B8	T7	At1g6955, At5g44510, At1g65850, At3g04220, At5g11250, At3g25510, At5g38344 At5g18370, At3g44670, At1g66090, At1g56540, At5g18350 (TIR-NBS-LRR)	2e-38
431E8	R	At4g21990	4e-12

^a Amplified by PCR using corresponding BAC clone; sequenced; homology calculated.

^b Calculated from internal sequence of homologous *B. rapa* BAC clone KBrB021L13.

^c Calculated from upstream sequence of homoeologous *B. oleracea* BAC clone BOGRL90TF.

^d Confirmed partial high homology.

barley, for example, the *Mla* locus for resistance to powdery mildew shows clustering and conservation in the genome, with a variety of duplications, inversions and transposon insertions (Wei *et al.* 2002). In addition, genes in the Gene Ontology category of “response to environment (including pathogen defense)” are over represented in the *B. rapa* genome, with an apparent bias (BRGSPC 2011). Taken together, these results suggest that the genomic region on A8 is inherited as a unit and the disease resistance cluster has been maintained through genome evolution with rearrangements in a hot spot of genomic substitution.

Our results confirm genomic microsynteny between *B. rapa* and *A. thaliana* and support the view that the origin of the diploid *Brassica* genome is more complicated than simple triplication. A potential for a disease resistance cluster in the *Brassica* genome is also suggested. For an accurate overview of genome evolution and the functional *R*-gene cluster in the *Brassica* genome, the ongoing complete sequencing of *Brassica* species will provide insights into mechanisms underlying the conservation and rearrangement of the genomes of differentiating species.

Acknowledgments

We thank Dr. Ian Bancroft of the John Innes Centre for his valuable comment and advice. We are grateful to Ms. K. Tanaka and Ms. H. Maeda for their technical assistance. This work was supported by the Cooperative System for Supporting Priority Research of the Japan Science and Technology Corporation, by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Rice Genome Project DM-2105) and by a grant-in-aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 14360006).

Literature Cited

- Arabidopsis Genome Initiative (2000) Analysis of the genome of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Arumuganathan, K. and E.D. Earle (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9: 211–215.
- Axelsson, T., O. Shavorskaya and U. Lagercrantz (2001) Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. *Genome* 44: 856–864.
- BRGSPC (*Brassica rapa* Genome Sequencing Project Consortium)

- (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genet.* 43: 1035–1039.
- Cavell, A., D. Lydiate, I. A. P. Parkin, C. Dean and M. Trick (1998) Collinearity between 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* 41: 62–69.
- Fu, H., Z. Zheng and H. K. Dooner (2002) Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proc. Natl. Acad. Sci. USA* 99: 1082–1087.
- Grant, M. R., J. M. McDowell, A. G. Sharpe, M. T. Zabala, D. J. Lydiate and J. L. Dangl (1998) Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 95: 15843–15848.
- Holub, E. B. (1997) Organization of resistance genes in *Arabidopsis*. In: Crute, I. R., E. B. Holub and J. J. Burdon (eds.) *The Gene-for-Gene Relationship in Plant–Parasite Interactions*. CAB International, Wallingford, UK, pp. 5–26.
- Hulbert, S., T. Pryor, G. Hu, T. Richter and J. Drake (1997) Genetic fine structure of resistance loci. In: Crute, I. R., E. B. Holub and J. J. Burdon (eds.) *The Gene-for-Gene Relationship in Plant–Parasite Interactions*. CAB International, Wallingford, UK, pp. 27–43.
- Hulbert, S. H., G. A. Webb, S. M. Smith and Q. Sun (2001) Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* 39: 285–312.
- Koch, M. A., B. Haubold and T. Mitchell-Olds (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabidopsis*, and related genera (*Brassicaceae*). *Mol. Biol. Evol.* 17: 1483–1498.
- Kuginuki, Y., H. Ajisaka, M. Yui, H. Yoshikawa, K. Hida and M. Hirai (1997) RAPD markers linked to a clubroot resistance locus in *Brassica rapa* L. *Euphytica* 98: 149–154.
- Kusaba, M., K. Dwyer, J. Hendershot, J. Vrebalov, J. B. Nasrallah and M. E. Nasrallah (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* 13: 627–643.
- Mun, J. H., S. J. Kwon, T. J. Yang, Y. J. Seol, M. Jin, J. A. Kim, M. H. Lim, J. S. Kim, S. Baek, B. S. Choi *et al.* (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* 10: R111.
- Nakamura, S., S. Asakawa, N. Ohmido, K. Fukui, N. Shimizu and S. Kawasaki (1997) Construction of an 800-kb contig in the bear-centromeric region of the rice blast resistance gene *Pi-ta²* using a highly representative rice BAC library. *Mol. Gen. Genet.* 254: 611–620.
- O'Neill, C. M. and I. Bancroft (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosome 4 and 5 of *Arabidopsis thaliana*. *Plant J.* 23: 233–243.
- Park, J. Y., D. H. Koo, C. P. Hong, S. J. Lee, J. W. Jeon, S. H. Lee, P. Y. Yun, B. S. Park, H. R. Kim, J. W. Bang *et al.* (2005) Physical mapping and microsynteny of *Brassica rapa* ssp. *Pekinensis* genome corresponding to a 222 kbp gene-rich region of *Arabidopsis* chromosome 4 and partially duplicated on chromosome 5. *Mol. Gen. Genomics* 274: 579–588.
- Parkin, I. A. P., D. J. Lydiate and M. Trick (2002) Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. *Genome* 45: 356–366.
- Parkin, I. A. P., S. M. Gulden, A. G. Sharpe, L. Lukens, M. Trick, T. C. Osborn and D. J. Lydiate (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765–781.
- Quiros, C. F., F. Grellet, J. Sadowski, T. Suzuki, G. Li and T. Wroblewski (2001) *Arabidopsis* and *Brassica* comparative genomics: sequence, structure and gene content in the ABI-Rps2-Ck1 chromosomal segment and related regions. *Genetics* 157: 1321–1330.
- Rana, D., T. Boogaart, C. M. O'Neill, L. Hynes, E. Bent, L. Macpherson, J. Y. Park, Y. P. Lim and I. Bancroft (2004) Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J.* 40: 725–733.
- Rossberg, M., K. Theres, A. Acarkan, R. Herrero and T. Schmitt (2001) Comparative sequence analysis reveals extensive microcollinearity in the lateral suppressor regions of the tomato, *Arabidopsis*, and *Capsella* genomes. *Plant Cell* 13: 979–988.
- Sadowski, J., P. Gaubier, M. Delseny and C. F. Quiros (1996) Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 251: 298–306.
- Sato, K., T. Nishio, R. Kimura, M. Kusaba, T. Suzuki, K. Hatakeyama, D. J. Ockendon and Y. Satta (2002) Coevolution of the *S*-locus genes *SRK*, *SLG* and *SP11/SCR* in *Brassica oleracea* and *B. rapa*. *Genetics* 162: 931–940.
- Schmidt, R. (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. *Plant Mol. Biol.* 48: 21–37.
- Schmidt, R., A. Acarkan and K. Boivin (2001) Comparative structural genomics in the Brassicaceae family. *Plant Physiol. Biochem.* 39: 253–262.
- Schopfer, C. R., M. E. Nasrallah and J. B. Nasrallah (1999) The male determinant of self-incompatibility in *Brassica*. *Science* 286: 1697–1700.
- Speelman, E., D. Bouchez, E. B. Holub and J. L. Beynon (1998) Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. *Plant J.* 14: 467–474.
- Suwabe, K., H. Tsukazaki, H. Iketani, K. Hatakeyama, M. Fujimura, T. Nunome, H. Fukuoka, S. Matsumoto and M. Hirai (2003) Identification of two loci for resistance to clubroot (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *Theor. Appl. Genet.* 107: 997–1002.
- Suwabe, K., H. Tsukazaki, H. Iketani, K. Hatakeyama, M. Fujimura, M. Kondo, T. Nunome, H. Fukuoka, M. Hirai and S. Matsumoto (2006) Simple Sequence Repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: the genetic origin of clubroot resistance. *Genetics* 173: 309–319.
- Suzuki, G., A. Ura, N. Saito, G. S. Do, B. B. Seo, M. Yamamoto and Y. Mukai (2001) BAC FISH analysis in *Allium cepa*. *Genes. Genet. Syst.* 76: 251–255.
- Takasaki, T., K. Hatakeyama, G. Suzuki, M. Watanabe, A. Isogai and K. Hinata (2000) The *S* receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature* 403: 913–916.
- Takayama, S., H. Shiba, M. Iwano, H. Shimosato, F. S. Che, N. Kai, M. Watanabe, G. Suzuki, K. Hinata and A. Isogai (2000) The pollen determinant of self-incompatibility in *Brassica campestris*. *Proc. Natl. Acad. Sci. USA* 97: 1920–1925.
- Town, C. D., F. Cheung, R. Maiti, J. Crabtree, B. J. Haas, J. R. Wortman, E. E. Hine, R. Althoff, T. S. Arbogast, L. J. Tallon *et al.* (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* 18: 1348–1359.
- U, N. (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7: 389–452.
- Wei, F., R. A. Wing and R. P. Wise (2002) Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. *Plant Cell* 14: 1903–1917.
- Yang, Y. W., K. N. Lai, P. Y. Tai and W. H. Li (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J. Mol. Evol.* 48: 597–604.
- Zimmer, R. and A. M. Verrinder Gibbins (1997) Construction and characterization of a large-fragment chicken bacterial artificial chromosome library. *Genetics* 42: 217–226.