Theor Appl Genet (2005) 111: 949–955
DOI 10.1007/s00122-005-0029-9

ORIGINAL PAPER

Muqiang Gao · Genyi Li · W. Richard McCombie
Carlos F. Quiros

Comparative analysis of a transposon-rich *Brassica oleracea* BAC clone with its corresponding sequence in *A. thaliana*

Received: 27 July 2004 / Accepted: 15 November 2004 / Published online: 26 July 2005

Abstract We compared the sequence of a 96.7 Kb-long BAC clone (B19N3) from *Brassica oleracea* (broccoli) with its corresponding regions in *Arabidopsis thaliana*. B19N3 contains eight genes and 5 transposable elements (TEs). The first two genes in this clone, Bo1 and Bo2, have its corresponding region at the end of chromosome V of Arabidopsis (24 Mb). The third gene, Bo3, corresponds to an ortholog at the opposite end (2.6 Mb) of the same chromosome. The other five genes, Bo4 to Bo8 also have a corresponding region on the same chromosome but at 7.7 Mb . These five genes are colinear with those found in the corresponding region of Arabidopsis, which contains, however, 15 genes. Therefore, a cluster of 10 genes is missing in *B. oleracea* clone (B19N3). All five genes in common have the same order and orientation in the genomes of both species. Their 36 exons constituting the eight homologous genes have high conservation in size and sequence identity in both species. Among these, there is a major gene involved in aliphatic glucosinolate biosynthesis, *BoGSL-ELONG* (Bo4). Similar to *A. thaliana*, this gene, has a tandem duplicate, Bo5. A contig for this region was constructed by primer walking and BAC-end-sequencing, revealing general gene colinearity between both species. During the 20 million years separating *A. thaliana* from *B. oleracea* from a common ancestor both genomes have diverged by chromosomal rearrangements and differential TE activity. These events, in addition to changes in chromosome number are responsible for the evolution of the genomes of both species. In spite of these changes, both species conserve general colinearity for their corresponding genes.

Communicated by C. Möllers

M. Gao · G. Li · W. R. McCombie · C. F. Quiros (✉)
Department of Vegetable Crops,
University of California, Davis, CA 95616, USA

Present address: G. Li
Department of Plant Science, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada

W. R. McCombie
Cold Spring Harbor Laboratory, Genome Research Center, Woodbury, NY 11797, USA

Introduction

Comparative genomics of *A. thaliana* and Brassica species is of clear importance for understanding the evolution of their genomes. The main events implicated in the formation of these genomes are changes on chromosome number and structure, as well as chromatin duplication and accumulation of transposable elements. Having a completely sequenced *Arabidopsis thaliana* genome (AGI 2000), it is possible to start to ascertain the contribution of most of the events molding the genomes of this species and other species of economic importance in the family, such as *Brassica oleracea*.

A. thaliana has a genome of only 125 Mb (AGI 2000) and $n = 5$ chromosomes whereas the genome of *B. oleracea* is five times larger, with approximately 600 Mb and almost twice as many chromosomes ($n = 9$). It has been estimated that these two species share 85% nucleotide identity and have diverged from a common ancestor 15–20 million years ago (Yang et al. 1999; Zhang and Wessler 2004). Based on the frequency of duplications, both genomes have undergone one or two cycles of polyploidization during their evolution (Quiros and Paterson 2004). Previous studies on comparative genomics for these two species have disclosed general conservation of gene content and colinearity, however, this conservation is incomplete, due to extensive chromosomal rearrangements (Kowalski et al. 1994; Lagercrantz 1998; Cavell et al. 1998; Lan et al. 2000; O'Neill and Bancroft 2000; Ryder et al. 2001; Bancroft 2001; Quiros et al. 2001; Babula et al. 2003; Li et al. 2003; Lukens et al. 2003; Gao et al. 2004a). The present study adds to our previous report comparing the sequences of

B. oleracea BAC clone B21H13 with its corresponding *A. thaliana* sequence (Gao et al. 2004a). This clone contains 23 genes including *BoGSL-ALK*, a major gene involved in the aliphatic glucosinolate pathway. Its corresponding region in Arabidopsis had 37 genes. Blocks of 5 to 7 genes were observed in Brassica. The genes in common conserved order and orientation. We now report the comparative analysis of the sequence of a second *B. oleracea* BAC clone, B19N3, with its corresponding regions in Arabidopsis. B19N3 also contains a major gene on glucosinolate biosynthesis, *BoGSL-ELONG* (Li and Quiros 2002). Sequencing of this clone provides further evidence of chromosomal rearrangements and accumulation of transposable elements (TEs) affecting gene density and size of introns and intergenic spacers.

Materials and methods

Sequencing of BAC clone B19N3

BAC clone B19N3 originates from a *Brassica oleracea* var . *italica* (broccoli) 'Early Big' library (Gao et al. 2004a). It was selected for sequencing because it contains *BoGSL-ELONG*, a major gene in aliphatic glucosinolate biosynthesis (Li and Quiros 2002). The complete and annotated sequence of this clone was submitted to GenBank as accession number AC149635 (Gao et al. 2004b). The BAC clone was subcloned and sequenced as described by Gao et al. (2004a). Five gaps were filled by a combination of primer walking and shotgun sequencing of subclones at both sides of the sequencing gaps. Final error rate was estimated using CONSED.

Sequence analysis and gene-prediction

The *Brassica* B19N3 sequence was analyzed for protein-coding genes with the following gene-prediction *A thaliana* software: GenScan (Burge and Karlin 1997), GlimmerM (Salzberg et al. 1999) and TwinScan, by comparing conserved regions in the DNA of both species (Flicek et al. 2003). The sequence of B19N3 was aligned with its corresponding *A. thaliana* sequences with Blast 2.2.9 (Altschul et al. 1997). The alignment result was viewed using ACT (http://WWW.Sanger.ac.uk/software/ACT), a DNA sequence comparison viewer based on Artemis (Rutherford et al. 2000). The score used for ACT was 30. The BAC sequence also was compared to *Arabidopsis, Brassica*, and *Oryza sativa* ESTs, cDNAs, and CDS using BLAST and FASTA with the NCBI, AGI and TIGR database (http://www.tigr.org/tdb/e2k1/bog1/) to analyze gene conservation. The conserved regions were translated into protein and tBLASTn applied to the GenBank protein database to adjust exon-intron boundaries. The informatic search was done during March 2004 and the last modified dates of all the Arabidopsis gene models were 2003-05-02 or 2003-05-07. last modified dates of all the Arabidopsis gene models were 2003-05-02 or 2003-05-07. We use the program "RepeatMasker"(A.F.A. Smith and P. Green, unpublished) to search and locate TEs in this BAC. Then we used BLASTN and BLASTX searches to the GenBank database to find by comparison all types of transposable elements.

Contig assembly by BAC-end sequencing

We assembled five BAC clones from the right end of B19N3 into a contig along the chromosome. One pair of 20 bp-long primers were constructed based on the B19N3 BAC-right end sequence. Two BAC clones were selected with this sequence from BAC library. These clones were then sequenced at both the ends. Then additional primers were constructed and used to re-screen the BAC library to select the rest of the BAC clones forming the contig.

Results and discussion

Characteristics of BAC clone B19N3

The length of the whole B19N3 sequence is 96,718 bp and contains eight genes, named Bo1 to Bo8, and 5 transposable elements. The G + C content in the whole clone is 35%, and 44% in its protein-coding region. The average gene size (from start codon to stop codon) of the eight genes in this BAC is 2,732 bp. The average gene density in B19N3 is 1 gene/12,089 bp (Table 1), three times less than the average gene density for the *A. thaliana* genome (Gao et al. 2004a).

Identification of protein-coding genes in B19N3

Eight protein-coding genes with 39 exons are predicted from the B19N3 sequence by using three programs, GenScan, GlimmerW and TwinScan, and BLASTP, BLASTN and BLASTX to conserved regions. By comparing conserved regions with Arabidopsis, TwinScan predicts correctly 28 exons of a total of 39 exons in 10 genes. All exons in four genes (Bo1, Bo4, Bo6 and Bo8) were predicted by TwinScan. GenScan predicted correctly only 12 of the 39 exons, and it was unable to predict all the exons of a single gene. The relative low prediction values of these programs for B19N3 are similar to those observed for Brassica BAC clone B21H13 (Gao et al. 2004a).

Comparison of B19N3 with its corresponding region in *Arabidopsis*

The *A. thaliana* sequences corresponding to the Brassica B19N13 sequence are on chromosome V, but there is

Table 1 Features of Brassica BAC clone B19N3 compared to its corresponding *A. thaliana* segment and to Brassica BAC clone B13H21

	Brassica BAC clone B19N3 (AC149635[a])	Brassica BAC clone B13H21[a] (AC122543)	Arabidopsis BAC clones T2007 + MYJ24 (AB026660 + AB006708[b])
Sequence length	96.7 Kp	101.5 Kb	97.2 Kb
G + C Content			
Overall	35.0%	37.1%	34.6%
Protein-coding DNA	44.0%	46.4%	44.0%
Non-coding region	33.8%	32.8%	29.0%
Total number of genes	8	23	15
Average gene size (bp)	2732	2086	3412
Average gene density (bp per gene)	12089	4415	6486
Average number exons per gene	4.5	5.1	7.1
Average exon size (bp)	184	264	320
Average number introns per gene	3.5	4.1	6.1
Average intron size (bp)	467	176	183
Average spacer size (bp)	9,627	2,425	2,695

[a] data extracted from Gao et al. 2004

[b] Arabidopsis data extracted from: Rice Chromosome 10 Sequencing Consortium (2003)

poor gene colinearity. The first two genes in the B19N3 clone, Bo1 and Bo2, have its corresponding region at the end of chromosome V (Arabidopsis clone MAE1, 24.4 Mb). There are three genes in the corresponding region in *A. thaliana*, but the middle gene is absent in *Brassica* clone B19N3 (At5g60820). The *A. thaliana* ortholog (At5g08130) for the third gene in B19N3, Bo3, is further removed from the rest, being at the opposite end of the same chromosome (clone: T22D6, 2.6 Mb). Five genes in B19N3, including Bo4, Bo5, Bo6, Bo7 and Bo8, have orthologous in the Arabidopsis contig formed by

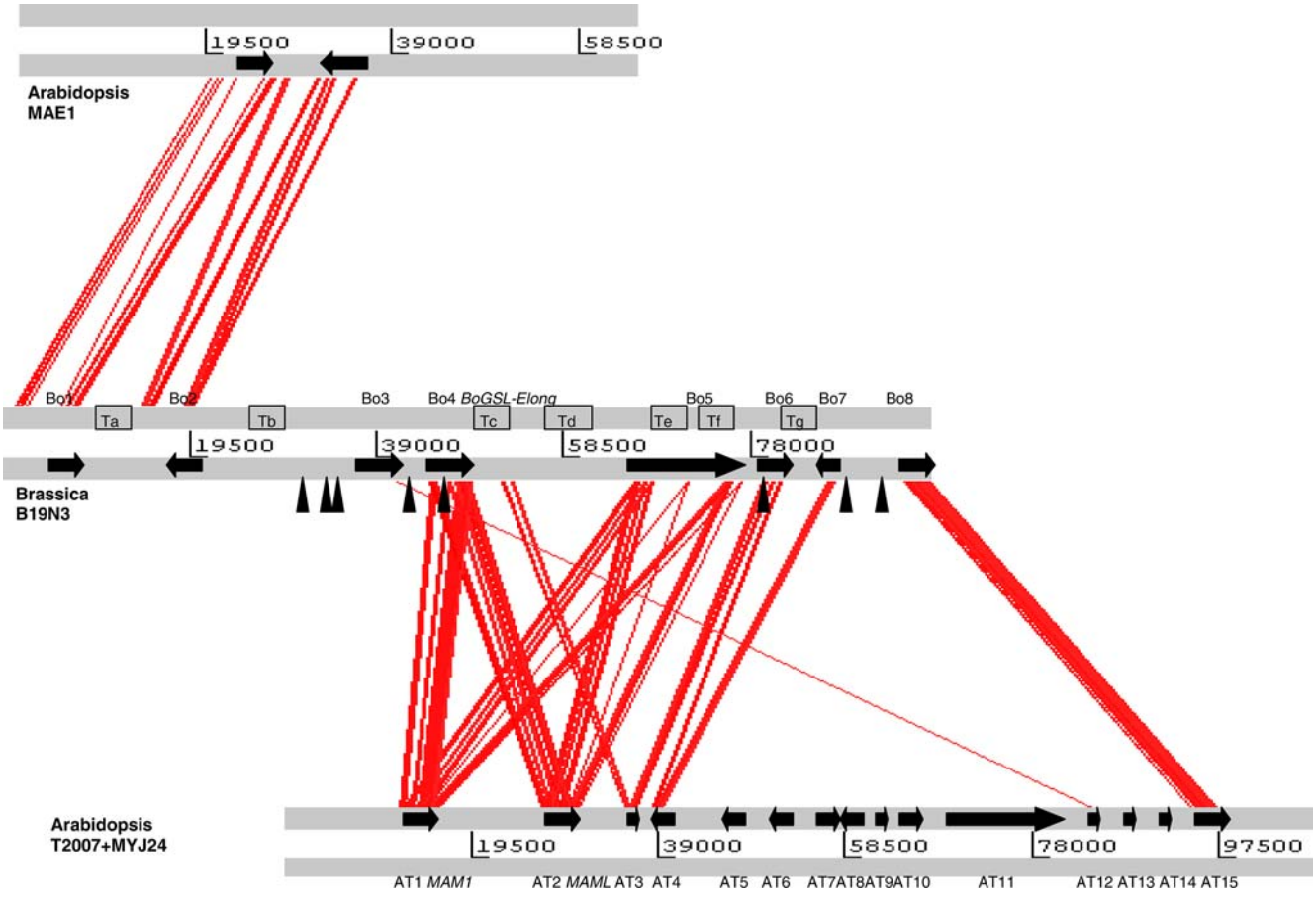

Fig. 1 Predicted genes and transposons in clone B19N3 and sequence alignment of this clone to Arabidopsis BACs MAE1 and T2007 + MYJ24 using the DNA sequence comparison viewer ACT. Main transposable elements are indicated by boxes, "*Filled triangle*" indicates transposable elements shorter than 1,000 bp

Table 2 List and properties of corresponding orthologs in Brassica BAC B19N3 and *A. thaliana*

Brassica				*Arabidopsis*				
Gene No.	Exons[a]	Introns[a]	Spacer[b]	Gene No.	cDNA.EST supported	Exons[a]	Introns[a]	Description
Bo1	531/1	0/0	11816	At5g60830	Partial	621/2	60/1	BZIP transcription factor protein
Bo2	351/3	228/2	18995	At5g60810	Partial	252/2	125/1	Hypothetical
Bo3	654/6	845/5	4573	At5g08130	No found	1230/9	767/8	bHLH family protein
Bo4	1464/10	2792/9		AT1. At5g23010	Full length	1521/10	2076/9	MAM1
BoGSL-ELONG			16967					
Bo5	1272/10	9910/9	2078	AT2. At5g23020	Partial	1512/10	1920/9	*MAM-L*
Bo6	792/2	482/1	5659	AT3. At5g23030	Full length	795/2	96/1	Sequence-associated cc protein
Bo7	504/4	258/3	7304	AT4. At5g23040	Full length	777/5	689/4	Expressed protein
Bo8	1638/3 (partial)	482/2 (partial)		AT15. At5g23150	Partial	1916/3	631/2	PWWP domain-containing protein
Total	7206/39	14504/31	67392/7	Total		8624/43	6364/35	
Average	184	467	9627	Average		200	181	

[a] Exons and introns: total length (bp)/ number
[b] Spacer: spacer length (bp) from the previous gene stop codon to listed gene start codon

clones T2007 and MYJ24 at 7.7 Mb. Colinearity is conserved for these five genes in both species; however, the corresponding region in Arabidopsis has a total of 15 predicted genes. The difference in gene number is due to the fact that a cluster of ten genes in B19N3 (between Bo7 and Bo8) is missing (Fig. 1, Table 2).

Gene structure and DNA sequence conservation

All 39 exons for the eight genes predicted in B19N3 have a high level of DNA sequence conservation with their orthologs in *A. thaliana* (Table 3). Their identity ranged from 70 to 90%, with the majority (25 exons) having 80–89% identity.

The orientations of all eight genes in B19N3 are the same as those of their Arabidopsis orthologs. Also they share the same structural features, with a few exceptions such as Bo1, Bo3 and Bo7, which have fewer exons than their corresponding to *A. thaliana* orthologs (Table 3). Bo8 was truncated, therefore it was not possible to assess whether it has all 21 exons present in its ortholog At5g23150. The genes displaying structural differences were near the breakpoints of the rearrangement events causing loss of colinearity between the chromosomes of both species, with the exception of Bo2. It is unknown whether these rearrangements might be related to the observed changes in exon number or is just coincidental.

Duplication of glucosinolate genes

BoGSL-ELONG (Bo4) is a key gene controlling biosynthesis of 4-carbon side-chain aliphatic glucosinolates in Brassica (Li and Quiros 2002). Its orthologous gene in

Table 3 Comparison of size (in bp) of corresponding exons in the orthologs of *A. thaliana* and *B. oleracea* (identity%)

Gene	Exon1	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7	Exon8	Exon9	Exon10
At5G60830	531	86								
Bo1	531(79%)	0								
At5g60810	91[a]	54	198							
Bo2	91 (83%)	44 (70%)	198 (84%)							
At5G08130	169	103	137	115	70	96	338	112	90	
Bo3	143 (76%)	0	41 (76%)	115 (85%)	70 (88%)	48 (85%)	371 (73%)	121(73%)	0	
At5g23010	441	260	214	81	93	87	101	34	75	132
Bo4	432 (83%)	260 (85%)	214 (78%)	81 (87%)	93 (91%)	87 (89%)	101 (91%)	34 (85)	75 (85%)	141(75%)
At5g23020	444	260	214	81	93	87	101	34	75	123
Bo5	192 (80%)	260 (80%)	214 (82%)	81 (85%)	93 (90%)	87 (88%)	101 (89%)	34 (75%)	75 (85%)	135(78%)
At5g23030	492	303								
Bo6	489 (80%)	303 (84%)								
At4g23040	65	101	383	77	151					
Bo7	0	0	279 (84%)	77 (87%)	148 (80%)					
At5g23150	111	77	1728							
Bo8	111 (89%)	77 (88%)	1638(78%)							

[a]Annotated as 5'UTR and not an exon in Arabidopsis

Arabidopsis, (At5g23010) also has this function and is a member of the isopropyl malate synthase-like gene family (methylthioalkylmalate synthase, *MAM1*) (Kroyman et al. 2001). Another member of this family, (At5g23020) is a tandem duplicate (*MAM2*) in some accessions and both genes together with a third locus, (*MAM-L*), which is approximately 11 kb apart, constitute the *GS-Elong* region in *A. thaliana* reported as an insect resistance quantitative trait locus (Kroyman et al. 2003). The segment harboring the paralogs *MAM1* and *MAM2* is quite variable in *A. thaliana* due to presence of indels and gene conversion events between the two genes, which range in nucleotide similarity between 95 to 99%. Further *MAM2* seems to be subjected to balancing selection. (Kroyman et al. 2003). Similar to *A. thaliana*, Bo4, has a tandem duplicate, Bo5 (*BoGSL-ELONG-L*), which corresponds to *MAM-L*. Bo4 and Bo5 and their corresponding *A. thaliana* homologs have 10 exons and share the same exon size, except those at both ends of the genes whose size is similar but not identical (Table 3). The exons of Bo4 and Bo5 share 90% identity. Bo5 is non-functional in broccoli 'Early Big' due to the insertion of a TE between exon 4 and 5. The tandem duplication of these genes in both species might have occurred before their divergence in separate lineages. However the lost of functionality of Bo5 might be an independent event since in its *A. thaliana* counterpart there is no evidence of TE insertions or presence of premature stop codons.

Transposable elements

We found fifteen TEs, making a total length of 23 kb in BAC B19N3 (Fig. 1). Seven of the TEs, *Ta-Tg*, are longer than 1 kb. The other eight TEs are shorter than 1 kb., a 3.5 kb-element is located between Bo1 and Bo2. The large size of the spacer between Bo2 and Bo3, which is 18,995 bp-long, is due partly to the insertion of *Tb*, a *copia*-type retrotransposon of approximately 5 kb, and insertion of three other shorter TEs. Similarly, the intergenic region between Bo4 and Bo5 is 16,967 bp due in part to the insertion of two TEs, *Tc* and *Td*. The first TE, *Tc*, is a 2 kb-long L1/CIN4 LINEs retroelement whereas *Td* is a 3 kb-long *En-Spm* DNA transposon. *Tg*, which corresponds to a transposase, is inserted in the intergenic region of Bo6 and Bo7. The large intronic sequence of Bo5, which is almost eight times that of its *A. thaliana* counterpart, is due to the insertion of two TEs, *Te* and *Tf* in two of the nine introns of this gene . *Tf* is a 3.5 kb-long *copia*-like-LTR retroelement inserted into intron 5 whereas *Te*, a *Harbinger* DNA Transposon, is inserted into intron 4 of Bo5. Three shorter TEs, a *Mariner* DNA transposon, a DNA tranposon and a *RC/Helitron* tranposon are located between *Tb* and Bo3. A *Pogo* DNA transposon is located between Bo3 and Bo4. A 354 bp-long DNA transposon is located in intron 1 of Bo4. A very short *HAT* DNA transposon is inserted into intron 1 of Bo6. Finally, two short TEs, a *Gypsy* LTR retroelement and a DNA transposon are inserted between Bo7 and Bo8.

TEs are the major component of plant genome size variation (Zhang and Wessler 2004; San Miguel et al. 1996; 1998; Suoniemi et al. 1996; Gribbon et al. 1999; Vicient et al. 1999). Only 6% of the *A. thaliana* is constituted by TEs and *B. oleracea* shares most of the TE lineages present in the former species. However, the number of elements in the genome of the latter species is much greater than in that of the former species, comprising approximately 20% of its genome (Zhang and Wessler 2004). The small size of the *A. thaliana* genome is due perhaps to its inability to amplify many copies of TE, since there is no evidence for their massive elimination. On the other hand the *B. oleracea* genome size, which is approximately five times that of *A. thaliana*, can be explained in part by relative amplification of certain TEs, such as copia-like elements (Zhang and Wessler 2004) and chromosome gain due to polyploidization (Quiros and Paterson 2004). The large proportion of TE elements in clone B19N3, six occupying close to 20% of its total sequence, results in a low gene density (1gene/12,800 bp) by increasing intron size and intergenic spacers, supporting the findings of Zhang and Wessler (2004). This chromosomal region seems to correspond to a TE rich region, considering that another sequenced BAC clone, B21H13 contains only one TE and has a gene density of 1 gene/4,415 bp (Gao et al. 2004a) (Table 1).

Conservation of a Brassica BAC contig with their corresponding *A. thaliana* sequences

After we sequenced the Brassica BAC clone B19N3, we assembled four additional BACs from the right end of B19N3 into a contig by BAC-end sequencing. These four BACs are in the following order: B13D10, B5F12, B63O8 and B42P13. Five of the eight Brassica BAC-end DNA sequences matched the same number of *A. thaliana* genes by BLAST alignment with the Arabidopsis database. We named these partial genes Bo9, Bo10, Bo11, and Bo12 (Fig. 2). The left end of BAC clone B5F12 matches with gene Bo8. Genes, Bo4, Bo5, Bo6, B07, Bo8, B09, B010, Bo11, Bo12 conserve order with their corresponding *A. thaliana* orthologs distributed in a contig of six BAC clones on chromosome V (Fig. 2). However, it is not possible to tell whether there are additional synteny breaks for the intervening genes, as observed in BAC clone B19N3. It is a well known fact that both the *A. thaliana* genome and *B. oleracea* genomes are duplicated (AGI 2000; Lagercrantz 1998; Quiros and Paterson 2004). Thus the question arises whether the regions compared are orthologous. The conservation of the two contigs in both species reinforces further the evidence for orthology of the compared regions. Furthermore, Li et al. (2002) presented evidence of orthology of the two regions based on sequence conservation of the

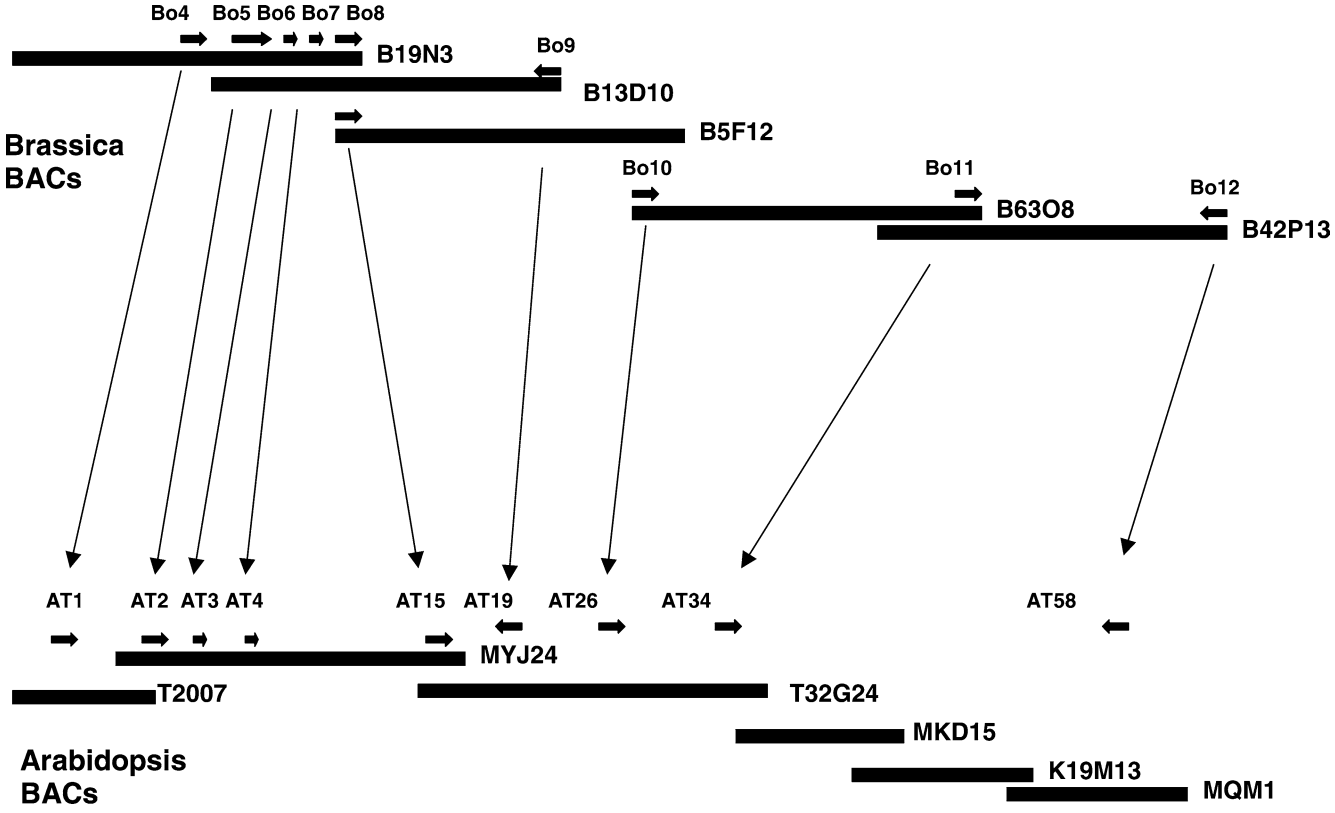

Fig. 2 Sequence conservation between five *B. oleracea* and six *A. thaliana* contiguous BAC clones AT1 = At5G23010, AT2 = At5g23020, AT3 = At5g23030, AT4 = At5g23040, AT15 = At5g23150, AT19 = At5g23190, AT26 = At5g23260, AT34 = At5g23340, AT58 = At5g23580

MAM gene region and overlapping Brassica BAC end sequences with those of *A. thaliana*.

Colinearity between Brassica and Arabidopsis

Complete sequencing of a second BAC clone from *B. oleracea* discloses that although general colinearity is maintained in the genomes of *A. thaliana* and *B. oleracea*, extensive chromosomal reshuffling has taken place during the evolution of these species. In addition to chromosome gain by polyploidization, accumulation of TEs have been responsible for the expansion of the Brassica genome. This accumulation seems to affect gene density in some genomic regions due to the increase in size of introns and intergenic spacers.

Acknowledgements We are indebted to Dr. Lidia Nascimento from Cold Spring Harbor Laboratory for reception and sequencing coordination of BAC clone B19N3, to Ms Bo Yang and Mr Vincent D'Antonio for technical assistance. Research supported by USDA-IFAFS grant# 00-52100-9683. "Development of Genomic Tools and Resources for *Brassica*" . The sequencing was funded under NSF grant NSF DBI 9813578:" A Genetic Approach to Ordered Sequencing of Arabidopsis".

References

AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Babula D, Kaczmarek M, Barakat A, Delseny M, Quiros C, Sadowski J (2003) Chromosomal mapping of Brassica oleracea based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. Mol Genet Genom 268:656–665

Bancroft I (2001) Duplicate and diverge: the evolution of plant genome microstructure. Trends Genet 17:89–93

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M (1998) A 30 centiMorgan segment of Arabidopsis thaliana chromosome 4 has six collinear homologues within the Brassica napus genome. Genome 41:62–69

Flicek P, Keibler E, Hu P, Korf I, Brent MR (2003) Leveraging the mouse genome for gene prediction in human: from whole-genome shotgun reads to a global synteny map. Genome Res 13:46–54

Gao M, Li G, Yang B, McCombie WR, Quiros CF (2004a) Comparative analysis of a Brassica BAC clone containing several major aliphatic glucosinolate genes with its corresponding *Arabidopsis* sequence. Genome 47(4):666–679

Gao M, Quiros CF, McCombie WR, Nascimento L, Balija V, Bell M, de la Bastide M, Spiegel L, Zutavern T, Muller S, Miller B, Katzenberger F, Andrade MV, Dike S, O'Shaughnessy A, Palmer LN (2004b) Genomic sequence for broccoli [*Brassica oleracea* L. (italica group)] BAC clone B19N3, complete sequence. NCBI submission AC149635. Genomic sequence gi: 48762521. http://www.ncbi.nlm.nih.gov/

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ (1999) Phylogeny and transpositional activity of Ty1-copia group retrotransposons in cereal genomes. Mol Gen Genet 261:883–891

Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics 138:499–510

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S (2001) A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127:1077–1088

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait loci. PNAS 100:14587–14592

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228

Lan TH, Del Monte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S, Paterson AH (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res 10:776–788

Li G, Quiros CF (2002) Genetic analysis, expression and molecular characterization of *BoGSL-ELONG*, a major gene involved in the aliphatic glucosinolate pathway of Brassica species. Genetics 162:1937–1943

Li G, Gao M, Yang B, Quiros CF (2003) Gene to gene alignment between the *Brassica* and *Arabidopsis* genomes by transcriptional mapping. Theoret Appl Genet 107:168–180

Lukens L, Zou F, Lydiate D, Parkin I, Osborn T (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics 164:359–372

O'Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. The Plant J 23:233–243

Quiros CF, Paterson AH (2004) Genome mapping and analysis in Brassica. In: Pua EC, Douglas CJ (eds) Biotechnology in Agriculture and Forestry, 54:31–42

Quiros CF, Grellet F, Sadowski J, Suzuki T, Li G, Wroblewski T (2001) Arabidopsis and Brassica comparative genomics: sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. Genetics 157:1321–1330

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944–945

Ryder CD, Smith LB, Teakle GR, King GJ (2001) Contrasting genome organization: two regions of the *Brassica oleracea* genome compared with collinear regions of the *Arabidopsis thaliana* genome. Genome 44:808–817

Salzberg SL, Pertea M, Delcher AL, Gardner MJ, Tettelin H (1999) Interpolated Markov models for eukaryotic gene finding. Genomics 59:24–31

San Miguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 82:37–44 [CrossRef][ISI]

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Suoniemi A, Anamthawat-Jonsson K, Arna T, Schulman AH (1996) Retrotranposon BARE-1 is a major dispersed component of the barley (*Hordeum vulgare* L.) genome. Pl Mol Biol 30:1321–1329

Vicient CM, Kalendar R, Anamthawat-Jonsson K, Suoniemi A, Schulman AH (1999) Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica 107:53–63

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and the other angiosperm lineages. J Mol Evol 48:597–604

Zhang X, Wessler S (2004) Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. PNAS 101:5585–5594