

Abhijit Sanyal · Scott A. Jackson

Comparative genomics reveals expansion of the FLC region in the genus *Arabidopsis*

Received: 22 October 2005 / Accepted: 22 October 2005 / Published online: 9 December 2005
© Springer-Verlag 2005

Abstract Mechanisms of genome evolution are poorly understood although recent genome sequencing is providing the tools to begin to illuminate such mechanisms. Using high-resolution molecular cytogenetic tools, we examined the structural evolution of 790 kb surrounding the evolutionarily important FLC locus of *Arabidopsis thaliana* in three of its relatives, *Arabidopsis halleri*, *Arabidopsis neglecta* and *Arabidopsis arenosa*. Sequenced BACs from *A. thaliana* were used as heterologous probes across these species and genome expansion was found in all three species relative to *A. thaliana*, ranging from 16 to 27%. Expansion was seen along the length of the entire region but molecular analyses revealed no characteristic pattern of either intra- or intergenic expansion among these species. Mapping of BACs on DNA fibers from *A. thaliana* revealed one possible error, ~14 kb missing from the reported sequence, indicating that for comparative studies it is important to confirm the reference sequence to which comparison will be made.

Keywords *Arabidopsis* · Comparative genomics · FISH · Physical mapping · Cytogenetics

Introduction

Plant genomes are highly variable in DNA content due to increases in chromosome number and/or accumulation of repetitive DNA (Leitch et al. 1998). Despite differences in DNA content, many plant genomes maintain remarkable genetic synteny and colinearity of genes along their chromosomes (Ahn and Tanksley

1993). Thus a paradox, how can evolutionarily divergent genomes that differ in DNA content and chromosome number have conserved gene synteny and order? Comparative genome sequencing is the most direct way to address this question but due to prohibitive costs it is unlikely that many plant genomes will be sequenced in the near future. Moreover, those that will be sequenced may not be chosen to address these particular questions. Using a combination of high-resolution, low cost cytogenetic tools, we compared the physical structure and evolution of a 789 kb region surrounding the FLC locus, a repressor of flowering (Michaels and Amasino 1999), amongst four species in the genus *Arabidopsis*.

Arabidopsis thaliana was the first plant to have a sequenced genome (The *Arabidopsis* Genome Initiative 2000). Despite its prominence as a genetic model, relatively little is known about the evolutionary history of genomes within the genus *Arabidopsis*; however, the phylogenetic position of these species to each other and within the Brassicaceae is becoming more defined (Koch et al. 2000; Al-Shebaz and O’Kane 2002). Comparative genetic mapping using RFLP probes from one species as markers in another has shown that there is extensive genetic colinearity between *A. thaliana* and various cultivated *Brassica* species (Kowalski et al. 1994; Lagercrantz et al. 1995; Cavel et al. 1998; Lagercrantz 1998). These studies are complicated due to the duplicated and even triplicated nature of the *Brassica* genomes (Quiros et al. 1994), the evolutionary distance traversed (15–20 MY) and do not result in high-resolution structural comparisons.

Using a set of species from the same genus as *A. thaliana*, we sought to determine, across the 789-kb FLC-containing genomic region in *A. thaliana*, the extent of structural conservation of sequences and genomic reorganization, expansion and/or contraction around this gene. FLC is involved in the vernalization process, whereby flowering occurs only after exposure to cold (winter), an important evolutionary trait/adaptation in many plants. FLC acts as a negative repressor of flowering and FLC activity is repressed during and after cold

Communicated by W.R. McCombie

A. Sanyal · S. A. Jackson (✉)
Purdue University, 915 West State St., West Lafayette,
IN 47907, USA
E-mail: sjackson@purdue.edu
Tel.: +1-765-4963621
Fax: +1-765-4967255

by an epigenetic process involving polycomb group proteins and targeted histone modifications (Bastow et al. 2004). In some accessions of *A. thaliana*, weak alleles of FLC have a transposon inserted in the first intron (Gazzani et al. 2003).

Fiber-FISH, fluorescence in situ hybridization on extended genomic DNA fibers, of BACs (bacterial artificial chromosomes) from *A. thaliana* spanning the FLC locus was used to structurally compare, at the resolution of the Watson–Crick double helix, the orthologous regions from three other *Arabidopsis* species (*Arabidopsis halleri*, *Arabidopsis arenosa* and *Arabidopsis neglecta*). All of these species have more genomic DNA and a greater chromosome number than *A. thaliana* [*A. thaliana* $n=5$, c -value = 0.34 pg) (*A. arenosa* $n=8$ or 14–16, c -value = 0.87 pg) (*A. halleri* $n=8$ or 16, c -value = 0.56 pg) (*A. neglecta* $n=16$, c -value = 0.87 pg) Al-Shehbaz and O’Kane 2002; W.A. Peer and A. Murphy, personal communication] indicating the possibility of structural evolution amongst these species. Unlike observations between *A. thaliana* and *Brassica rapa* where there was little difference in size of an orthologous region (Jackson et al. 2000), expansion was seen in this region among these *Arabidopsis* species ranging from 16% in *A. neglecta* to 26% in *A. arenosa* and 27% in *A. halleri*. Further analysis showed relatively even distribution of expansion along this 789 kb, although molecular analysis did not reveal any characteristic increase in size of either intra- or intergenic regions. The BACs from *A. thaliana* were also mapped back to *A. thaliana* revealing an instance of incongruity between the physical mapping data and the reference genome sequence.

Materials and methods

A 789-kb BAC contig from *A. thaliana* ecotype Col, consisting of eight BAC and two PAC clones (T5E8, F17I14, MYH9, T31P16, F18D22, F12B17, MAJ23, T30N20, T5K6 and F2I11) was chosen from the TAIR website (<http://www.arabidopsis.org>) for comparative physical mapping (Fig. 1). This region, which spans the FLC locus, is located on the short arm of chromosome 5, ~2.9 Mb from the telomere and ~8.1 Mb from the centromere. We were unable to obtain the PAC clones for this study. The BAC clones were obtained from the Arabidopsis Biological Resource Center (ARBC). BAC DNA was extracted using Qiagen large construct kits (Qiagen Corp. Hilden, Germany). Following extraction, 1 µg of DNA was labeled with either digoxigenin-11-dUTP or biotin-16-dUTP (Roche Diagnostics, Basel, Switzerland) using standard nick translation protocols.

Nuclei were extracted from leaf tissue of *A. arenosa*, *A. neglecta*, *A. halleri* spp. *halleri* (provided by Dr. Wendy Ann Peer, Purdue University, West Lafayette, IN) and *A. thaliana* ecotype Columbia according to the method of Liu and Whittier (1994). Genomic DNA fibers were prepared following Fransz et al. (1996) and

Jackson et al. (1998). Probe preparation for fiber-FISH followed the methods described in Jiang et al. (1996), using modifications described by Jackson et al. (2000). Each experiment consisted of two duplicate slides, and each experiment was repeated twice to ensure reliability of the data.

Images were captured digitally using a Hamamatsu Photonics (Japan) charge coupled device (CCD) camera attached to an Olympus BX60 epifluorescence microscope using 60 or 100× objectives. The camera was controlled via Metamorph v 6.0r3 (Universal Imaging Corp., Downingtown, PA, USA) on a PC. Images for red (TXRED), green (FITC) and blue (DAPI) channels were captured as gray scale images and then merged to produce color images. Brightness and contrast of the images were adjusted using Adobe Photoshop v 8.0. Measurements were made on the merged images using Metamorph software and the data was analyzed in Microsoft Excel (2000) using the data analysis package.

Polymerase chain reaction (PCR), using primers described in Table 2, was done using the BD Advantage™ 2 PCR enzyme system (BD Biosciences, San Jose, CA, USA). A 50 µl PCR reaction consisted of 1× BD Advantage 2 PCR buffer, 1 µl of each primer, 0.2 mM dNTP mix, 1× BD Advantage 2 polymerase mix and distilled water. The reaction conditions were 95°C for 3 min, 30 cycles of 94°C for 40 s, 60°C for 1 min 30 s, 72°C for 3–5 min (depending on expected size of PCR product and taking 1 kb/min rate of extension), and final extension at 72°C for 8 min.

Results

Ploidy and copy number of the FLC region in each species

The terminal BACs of this contig (T5K8 and F2I11) were used as FISH probes to pachytene chromosomes and interphase nuclei to determine the ploidy of the species and, where possible, the copy number of this region in the event of segmental duplications. Either of these instances can confound subsequent fiber-FISH analysis if there are structural differences between paralogous regions.

FISH using the two flanking BACs to *A. arenosa* revealed loci on two separate pairs of homologues confirming that this accession is tetraploid with $n=16$ chromosomes (Fig. 2a; W.A. Peer, personal communication). FISH to meiotic mid-prophase chromosomes of *A. halleri* revealed only one set of loci on one pair of homologues indicating that this accession is diploid (Fig. 2b). We were unable to obtain chromosomes from *A. neglecta* but FISH to interphase nuclei showed two distinct signals for each BAC resulting from the unpaired homologues (Fig. 2c), thus, this accession of *A. neglecta* is also diploid.

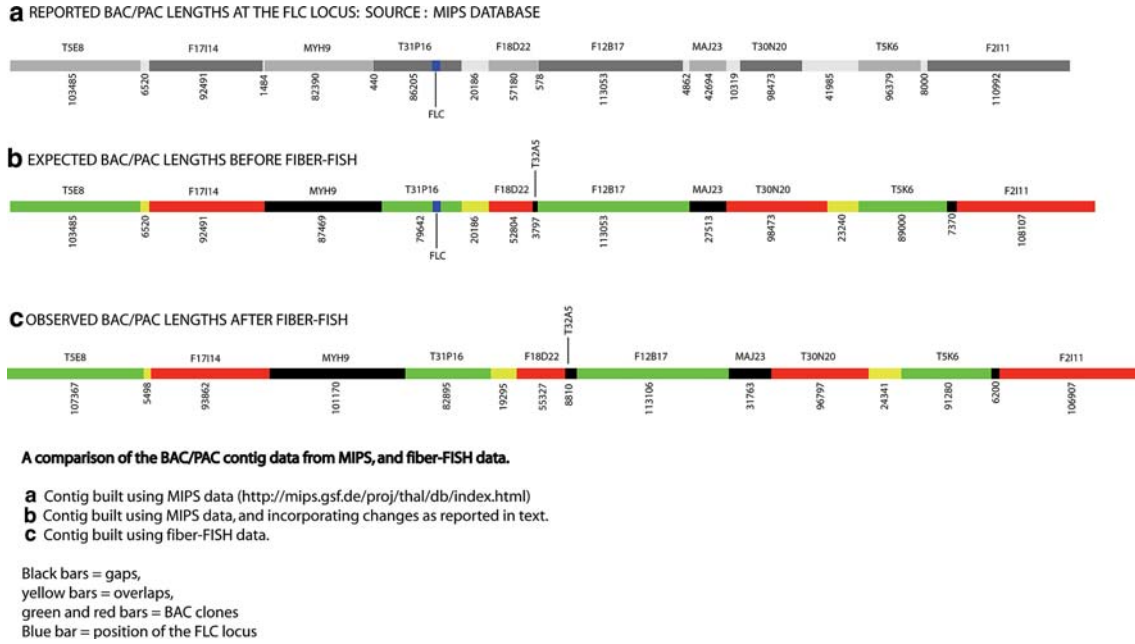


Fig. 1 Schematic of the FLC-containing BAC contig from Chromosome V of *Arabidopsis thaliana* employed in comparative studies. **a** The FLC contig using the data from <http://www.mips.gsf.de/proj/thal/db/index.html>. BACs/PACs are shown as gray or black bars with their names and reported lengths. Overlap between clones is shown as lightly colored bars. **b** Expected BAC/PAC contig after making adjustments because of unavail-

ability of the PAC clones and sequence that was added to several BAC ends to facilitate sequence assembly (see Tables 1, 2). BACs are colored green or red and overlaps yellow to reflect the fiber-FISH data. PACs are shown in black as they will appear as gaps in fiber-FISH data. **c** Schematic of the fiber-FISH data from *A. thaliana* reported in text

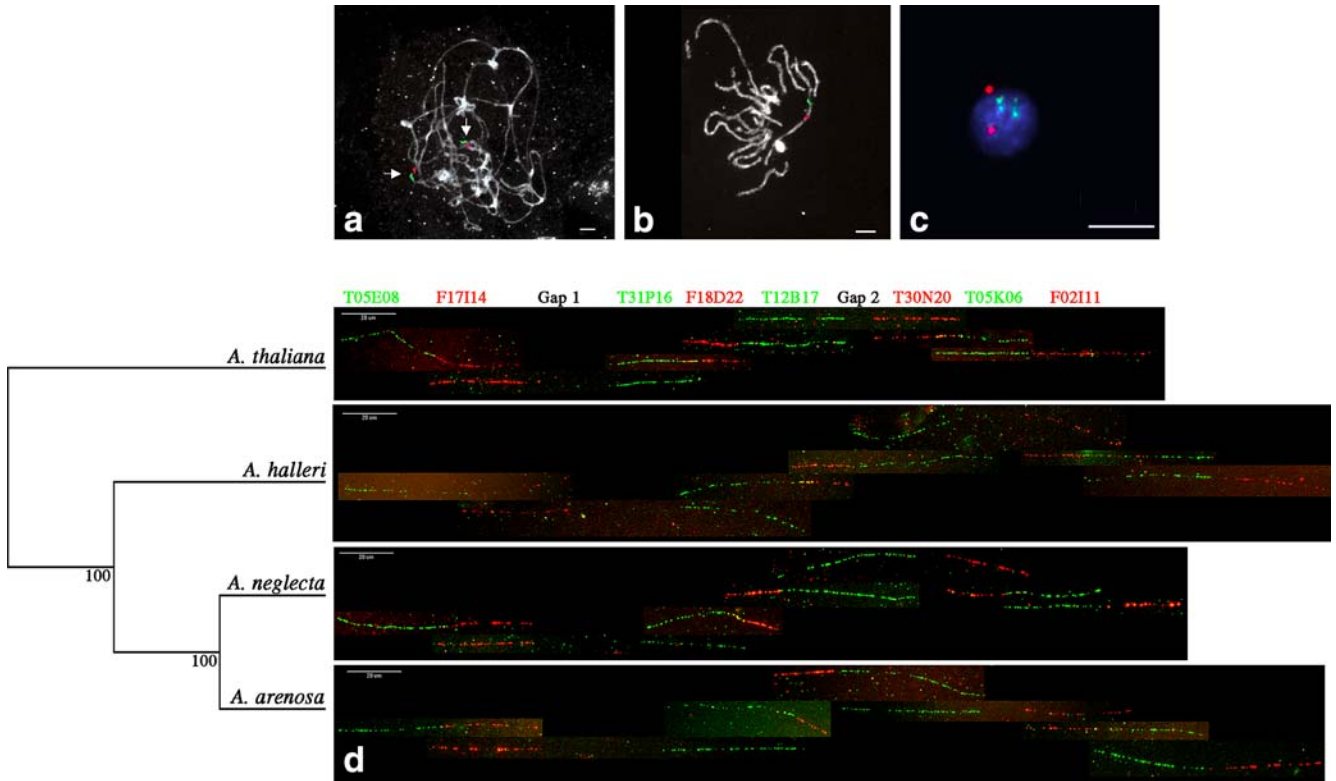


Fig. 2 FISH and fiber-FISH of the FLC region across four species of *Arabidopsis*. **a-c** FISH of terminal BACs, T5E8 (green) and F211 (red), to **(a)** *A. arenosa*, **(b)** *A. halleri* and **(c)** *A. neglecta* showing that *A. arenosa* is tetraploid with two copies (arrows) of this locus and *A. halleri* and *A. neglecta* are both diploid with one copy of this locus. Size bars = 5 μ m. **d** Fiber-FISH analysis of eight *A. thaliana* BACs mapped across the four *Arabidopsis* species as adjacent pairs of BACs. BACs are shown across top of figure with color indicating

the fluorophore used for FISH detection. Size bars = 20 μ m. Species relationships are shown on left. A phylogenetic tree was constructed using PHYLIP and a CLUSTAL alignment of 5.8S rRNA genes, including ITS1 and ITS2, of the four species (GenBank accession numbers: *A. halleri* gi|6686530, *A. arenosa* gi|1244732, *A. neglecta* gi|1245680 and *A. thaliana* gi|16131). Bootstrap values are shown in nodes

Physical size of the FLC region

In our first attempts to integrate the fiber-FISH data with the *A. thaliana* sequence data we uncovered multiple instances of incongruity that could, for the most part, be explained by the way sequences in this region were assembled. On further analysis of the sequence data it was found that four BACs had sequences from other BACs or PCR products added to their ends. These changes are noted in Table 2 (expected column) and shown in Fig. 1 and as follows based on data from MATDB and Dr. Klaus Mayer (personal communication). The results of this analysis led to an expected contig size of 813 kb, to which *A. thaliana*'s fiber-FISH comparisons were made.

Gap 1

The PAC clone MYH9 was unavailable, thus a gap between BACs T31P16 and F17I14 was expected. MYH9 has a reported insert size of 82.39 kb and has overlaps with F17I14 and T31P16 of 1,484 bp and 440 bp, respectively. Therefore the expected gap size is 80.9 but T31P16 has additional sequence added to it (see below) making the expected gap size 87.47 kb.

T31P16

The first 6,563 bp were added from an overlapping clone, T22D10. After subtraction of this value, BAC T31P16 is 79.64 kb.

F18D22

The last 4,376 bp of this clone were derived from overlapping clone T32A5, thus the size of F18D22 is 52.80 kb.

Overlap 3

The overlap of 578 bp between F18D22 and F12B17 occurred because of the addition to F18D22 (above); therefore, a gap of 3,797 bp is expected between F18D22 and F12B17.

GAP 2 (MAJ23)

The PAC clone MAJ23 was also unavailable for our studies. Hence a gap is expected between clones F12B17 and T30N20. When the overlap with F12B17 (4,862 bp) and T30N20 (10,319 bp) are subtracted from the reported size of MAJ23, the expected gap size is 27.51 kb.

Overlap 4

After conferring with Dr. Klaus Mayer at MIPS, we have used the value of 23.24 kb as reported by The

Arabidopsis thaliana Information Resource (TAIR) to be the expected value of this overlap not the MIPS value of 41.98 kb.

T5K6

A Blast of the SP6 BAC end sequence of T5K6 to itself maps to position 88,688–89,006. Since additional sequence was added to this BAC, the expected size is 89 kb.

Overlap 5

The SP6 BAC end sequence of T5K6 also maps to position 309–627 on BAC F2I11. The first 2,885 bp of F2I11 were added from another clone (F22L2), so BACs T5K6 and F2I11 do not physically overlap but have a gap of ~7 kb that may have been bridged by a PCR fragment. This is uncertain as the sequencing lab did not detail how this gap was bridged (Dr. Klaus Mayer, personal communication).

F2I11

The initial 2,885 bp of F2I11 belong to clone F22L2, so the actual size of F2I11 is 108.10 kb.

Comparative mapping using fiber-FISH

Fiber-FISH is essentially mapping at the Watson–Crick resolution on extended DNA fibers released from isolated plant nuclei (Fransz et al. 1996; Jackson et al. 1998) where every micrometer of physical distance is equivalent to ~2.9 kb of DNA (Jackson et al. 1998). A 789-kb BAC contig spanning the FLC locus of *A. thaliana* was chosen to examine the physical structure of this region in the related species *A. halleri*, *A. neglecta* and *A. arenosa*. Two PACs (P1-derived artificial chromosomes, MYH9 and MAJ23), part of this sequencing contig, were unavailable for these experiments; thus, two gaps, gaps 1 and 2, respectively, reflect the missing PACs. Adjacent BACs were mapped in pairs to fibers from all four *Arabidopsis* species and measurement data was quantitatively analyzed to determine the size of the orthologous regions.

Mapping in *Arabidopsis thaliana*

The data derived from measurements of these BACs in *A. thaliana* closely matched the publicly available genome sequence for this region (Fig. 1, Tables 1, 2). Chi-squared tests were made between the fiber-FISH measurements and the expected values in Tables 1, 2. None of the fiber-FISH data deviated significantly from the expected sizes except for gap 1 (PAC MYH9) that was larger by ~14 kb than the expected value (Tables 1, 2).

Table 1 Sequencing data of an *Arabidopsis* BAC contig and fiber-FISH data from *A. thaliana*, *A. halleri* and *A. neglecta*

Sequencing data ^a		Fiber-FISH data																
		<i>A. thaliana</i>				<i>A. neglecta</i>				<i>A. halleri</i>				<i>A. arenosa</i>				
BAC	Reported (kb)	Expected (kb)	Size (kb) ^e	SD ^b	n	Chi-test ^c	Size (kb) ^e	SD ^b	n	t-test ^d	Size (kb) ^e	SD ^b	n	t-test ^d	Size (kb) ^e	SD ^b	n	t-test ^d
T5E8	103.48	103.48	107.36	5.67	28		117.73	6.06	27	**	150.15	7.57	19	**	116.10	6.81	15	**
Overlap 1	6.52	6.52	5.49	1.98	31		4.51	3.83	31		4.06	3.01	21	**	2.91	3.29	18	**
F1714	92.49	92.49	93.86	5.81	43		97.96	11.74	75	*	108.93	7.52	30	**	100.71	13.66	36	*
Overlap	1.48	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Gap 1 (MYH9)	82.39	87.46	101.17	5.39	21	**	117.89	11.90	34	**	119.38	5.66	16	**	131.70	17.88	16	**
Overlap	0.44	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
T31P16	86.20	79.64	82.89	3.89	38		111.21	8.17	60	**	138.06	6.11	36	**	120.89	14.70	34	**
Overlap 2	20.18	20.18	19.29	3.56	20		18.72	8.05	28		18.88	5.76	24		18.56	7.45	27	
F18D22	57.18	52.80	55.32	3.57	33		56.72	4.98	50		56.19	5.42	50		59.62	6.28	40	**
Overlap 3	0.57	3.79 (gap)	8.81 (gap)	6.56	29		2.82 (gap)	4.42	28	**	6.06 (gap)	5.23	30	**	8.68 (gap)	7.47	25	**
F12B17	113.05	113.05	113.10	4.76	37		146.56	7.13	44	**	150.16	21.54	49	**	142.84	12.53	25	**
Overlap	4.86	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Gap 2 (MAJ23)	42.69	27.51	31.76	2.20	20		37.35	8.06	33	**	45.76	4.90	13	**	46.01	13.65	17	**
Overlap	10.31	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
T30N20	98.47	98.47	96.79	4.27	24		88.12	9.59	54	**	108.22	10.50	36	**	135.68	37.15	39	**
Overlap 4	41.98	23.24	24.34	3.70	28		24.31	5.83	36		33.38	8.90	21	**	30.54	8.85	20	**
T5K6	96.37	89.00	91.28	4.06	30		104.86	5.11	40	**	124.06	9.60	33	**	114.37	9.60	38	**
Overlap 5	8.00	7.37 (gap)	6.2 (gap)	3.79	48		10.58 (gap)	5.53	26	**	17.15 (gap)	5.36	16	**	9.91 (gap)	6.55	30	**
F2I11	110.99	108.10	106.90	6.69	23		133.85	10.60	15	**	109.02	11.49	15	**	128.85	9.71	24	**
Contig length	788.98	813.23	846.32				978.11				1076.82				1063.35			

NA not applicable

*Significant at 5% level; ** significant at 1% level

^aData from MIPS (<http://www.mips.gsf.de/proj/thal/db/index.html>)^bStandard deviation^cSize comparison between observed and expected values from *A. thaliana*^dFiber-FISH data comparison to observed values from *A. thaliana*^eThe physical size (kb) was calculated after using a conversion of 2.87 kb/um for the fiber-FISH measurements

Table 2 PCR primers used to amplify around the FLC locus in four species of *Arabidopsis*

	Position		Intra/intergenic	BAC/PAC	Primer	Sequence	Expected PCR size
	From	To					
a	3095320	3097844	Inter/intra	MYH9	MYH9_RP2_F MYH9_RP2_R	ggaagaagcagaaaggcgta agagaagcacggaagaagca	2101
b	3172273	3175274	Intragenic	T31P16	T31P16_FLC_1F T31P16_FLC_1R	gtcgcttctcgtcgtctc ccagttgaacaagagcatcg	3996
c	3189975	3193621	Intergenic	T31P16	T31P16_I3P_F T31P16_I3P_R	tcacgttcggagattcaacc aatggcttggaggacagaga	3403
d	3238082	3240725	Inter/intra	F18D22	F18D22_HYASE_F F18D22_HYASE_R	aagccatagccatcttctgc cgccatcgatctcgtagact	2513
e	3282110	3286262	Inter/intra	F12B17	F12B17_CYCD4_F F12B17_CYCD4_R	aacgatctctcgtctcca ctcgttacaggcgtactcc	3940
f	3297040	3301105	Inter/intra	F12B17	F12B17_PAS2_F F12B17_PAS2_R	gccagtaataaagcggctct cactcacctcccagttccat	3540
g	3360761	3365688	Intragenic	MAJ23	MAJ23_EF1_F MAJ23_EF1_R	tgttctcgtcatctcgtcg aaaccggttccatctctcg	4713
h	3382328	3387540	Intergenic	MAJ23	MAJ23_AHK5_F MAJ23_AHK5_R	agcaatccctgtgcgttac acctgcggaacaatcactc	4812
i	3405169	3408551	Intergenic	T30N20	T30N20_ChIDBP_F T30N20_ChIDBP_R	tcgaccagtaccatacact tgagctccatcgttacaacc	3207
j	3418298	3421452	Inter/intra	T30N20	T30N20_EnRP_F T30N20_EnRP_R	tcggataaccagattccttg ggtgttttccagggtga	2940
k	3500388	3503636	Inter/intra	T5K6	T5K6_MYB_F T5K6_MYB_R	gctgcaagaagcaagacact tcgctagtggagttagctgg	3139
l	3547626	3550865	Intergenic	F2I11	F2I11_SYNAP_F F2I11_SYNAP_R	tggtatcgtatcgtttcc caaccctaaggtctgctgat	3130
m	3591345	3595686	Inter/intra	F2I11	F2I11_BZIP_F F2I11_BZIP_R	aggaccgatggattctctc ggaagcgagtttgaagtg	4000

Comparative mapping in *A. neglecta*, *A. halleri* and *A. arenosa*

Arabidopsis thaliana BACs were hybridized to genomic fibers isolated from the three related *Arabidopsis* species

(Fig. 2d) and compared to the measurements from *A. thaliana* (Table 1). A minimum of 15 measurements were made for any individual segment but ranged up to 75 independent measurements. Comparisons were made to the *A. thaliana* fiber-FISH data using two-tailed

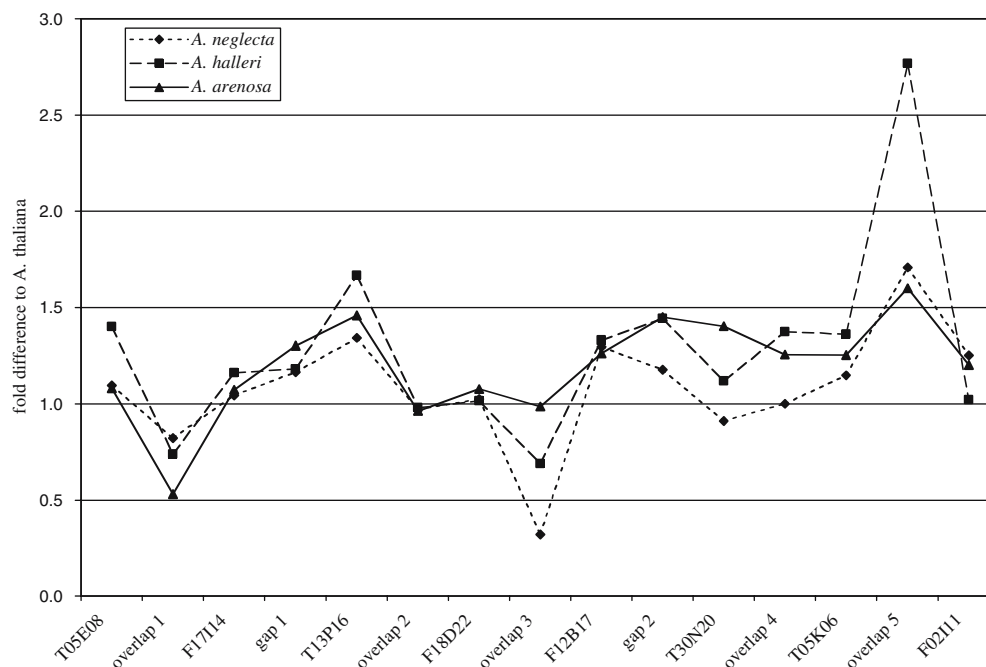


Fig. 3 Change in fiber-FISH-derived size measurements across the FLC-containing region relative to *A. thaliana*. Fiber-FISH measurement data was converted to fold difference relative to the

corresponding *A. thaliana* measurement and graphed linearly. Gap1 and Gap 2 correspond to the PAC clones, MYH9 and MAJ23, respectively, that were unavailable

t-tests and significance at the 5 and 1% levels are reported. A phylogenetic tree of these three species was constructed using ITS 1 and 2 (internal transcribed sequence) of the 5.8S rRNA gene to show the putative relationship of these species (Fig. 2d).

Arabidopsis neglecta is reported to have a diploid chromosome complement of $2n=16$ and a 0.87 pg haploid genome, ~ 2.6 times bigger than *A. thaliana*. Of the 15 comparisons made between the BACs, gaps and overlaps, 11 were significantly different at either the 1 or 5% levels (Table 1). Nine of the 11 differences were larger while BAC T30N20 and the gap at overlap 3 were smaller (Table 1 and Fig. 2d).

Arabidopsis arenosa is reported to have chromosome numbers of either $2n=2x=16$ or $2n=4x=32$ (Al-Shebaz and O’Kane 2002) and a genome size of 0.87 pg per haploid genome. Cytogenetic analysis revealed that this accession is tetraploid. On DNA fibers, 12 out of the 15 measurements from *A. arenosa* (Table 1) were significantly larger than those from *A. thaliana* while overlap 1 was smaller. Similarly for *A. halleri*, previously known as *Cardaminopsis halleri* (L.) Hayek, 10 of the 15 measurements in this region were significantly larger than those from *A. thaliana*. The overall fiber-FISH size increased from 846.3 kb in *A. thaliana* to 978.1 kb in *A. neglecta*, to 1063.4 kb in *A. arenosa* and to 1076.8 kb in *A. halleri*.

The expansion, relative to *A. thaliana*, appears to be similar in distribution along the segment among these species (Fig. 3), although, the overall expansion varies among these three relatives of *A. thaliana*.

Analysis of structural variation in intergenic/intronic regions

To establish the source of structural variation across these species, primer pairs for 15 regions that spanned either introns (intragenic) or genes (intergenic) were used to PCR-amplify across all the species (Table 2). This was done to determine if there was any characteristic increase in size and if it was due primarily to intra- and/or

intergenic expansion. There were a few clear instances of PCR products that increased in size such as product “g” in *A. neglecta*, “h” in *A. halleri*, and products “l” in *A. neglecta* and *A. halleri* but there were also instances of smaller PCR products (Fig. 4). No clear pattern of increasing PCR product size concomitant with increasing genome size was seen nor were there any discernable differences between inter- and intragenic amplicons across these species. Two primer pairs for GNAT on BAC F2I11 and calcineurin on BAC T30N20 were excluded as they did not amplify across all the species. The lack of amplification could be due to loss of primer site(s) or a large insertion between the primer sites that could have contributed to the expansion seen in this region.

Discussion

The genus *Arabidopsis* is uniquely poised for comparative studies as the complete genome sequence of *A. thaliana* is available and the genome sizes of these species are relatively small. Previously, we have shown that *A. thaliana* BACs can be used as heterologous probes to chromosomes and DNA fibers from *B. rapa* (Jackson et al. 2000). This was exploited in the current study to examine the structure of the FLC region in a more closely related group of species. FLC is a key component in vernalization, a trait in plant adaptation, and some accessions of *A. thaliana* with weak alleles of FLC, i.e., Landsberg erecta (Ler), have a 1.2-kb *Mutator*-like transposon inserted in the first intron (Gazzani et al. 2003).

One observation from these data is that the fiber-FISH results from *A. thaliana* closely match the expected values with only one instance of incongruity. Gap 1, a missing PAC clone, was 14 kb larger than the expected size. This could be due to incorrect assembly of sequences within this clone or genomic variation in the Columbia ecotype of *A. thaliana*. Such structural differences have been observed between inbreds of maize (Fu and Dooner 2002). It is unlikely that variation in the

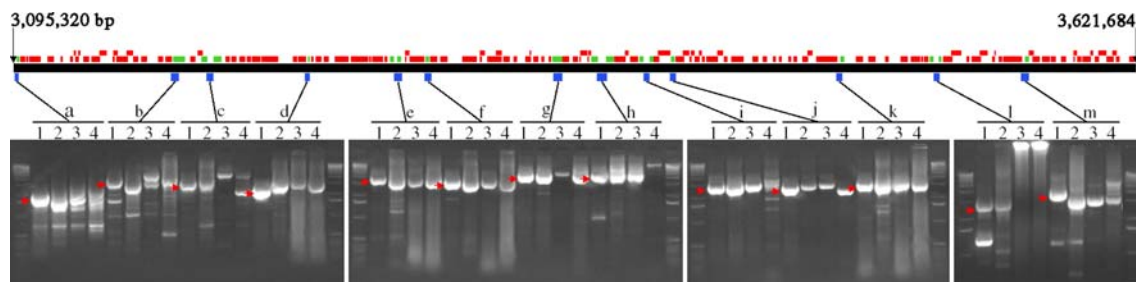


Fig. 4 PCR analysis of intra- and intergenic regions across the four *Arabidopsis* species. A portion of the 789-kb FLC region (3,095,320 bp–3,648,233 bp) is shown across top of the figure. Genes (taken from genome annotation at TAIR) are shown across top in green and red, green indicates that all or part of that gene is included in the PCR product shown across bottom in blue. PCR

products are shown in bottom of figure and Lanes 1, 2, 3 and 4 are *A. thaliana*, *A. arenosa*, *A. neglecta* and *A. halleri*, respectively. Letters across the top refer to PCR primer pairs and positions indicated in Table 2 and red arrows indicate expected PCR product in *A. thaliana*. Molecular weights (1 kb Plus ladder, Invitrogen) are shown in first and last lane of each panel

fiber-FISH technique (Nishio et al. 1996) contributed to this as we did not see other deviations from the expected values in the rest of the contig. However, it would also be fallacious to conclude that the sequence is in error (missing 14 kb) without resequencing of this region in *A. thaliana*.

When the gap 1 discrepancy is subtracted from the fiber-FISH data, the overall length is ~832 kb—only 1.02% larger than the expected value of 813 kb and well below the variation inherent to fiber-FISH. This highlights the importance of verifying the reference genome when doing comparative studies with related genomes as structural rearrangements could be wrongly inferred.

No other major gross rearrangements were observed, i.e., translocations or transpositions, at either the chromosomal or DNA fiber level. This does not preclude the possible rearrangement of elements below the resolution of these mapping techniques. For instance, duplication or deletion of small segments or genes within the region would not be detectable at the chromosomal level, but may be quantitatively detectable at the level of fibers. This would result in increases or decreases in sizes of the BACs across the species. This is complicated by the possibility of invading transposable elements that could lead to increased measurements, e.g., 1.2 kb insertion in *FLC* in some *Arabidopsis* accessions (Gazzani et al. 2003).

Retroelement invasion is a major mechanism for genomic expansion among eukaryotes (reviewed in Kidwell 2002) including the grass family (SanMiguel and Bennetzen 1998). Analysis of fiber-FISH data from our study suggests that invasion of large (5–15 kb) retroelements did not lead to the observed genomic expansion as there were not consistent gaps of that size range. This does not preclude contribution of smaller retroelements undetectable by the fiber-FISH method or the removal of retroelements by either homologous or illegitimate recombination (Devos et al. 2002; Ma et al. 2004). A 1.2-kb Mutator-like element is present in some *Arabidopsis* accessions (Gazzani et al. 2003), an element this small would likely be missed using fiber-FISH as 1.2 kb is less than one micrometer of physical distance.

In order to look at intra- vs intergenic contribution to the expansion of this region we PCR-amplified several regions to see if the PCR fragments from the three other species were uniformly larger than those of *A. thaliana*. The majority of the fragments were similar in size, only a few increased in size in one or all of the species relative to *A. thaliana*. A few primer pairs produced fragments in the relatives smaller than those of *A. thaliana*. A recent report from rice shows that amplification of LTR-retrotransposons has been a primary contributor to increased genome size of *Oryza sativa* cv japonica and *O. sativa* cv indica relative to their progenitor, although this has been counteracted by small deletions (Ma and Bennetzen 2004). Without targeted sequencing of the orthologous regions from *A. arenosa*, *A. halleri* and *A. neglecta*, it is impossible to tell what contribution these mechanisms may have had to the expansion of this region or the entire genomes.

The expansion of this region by 27% in *A. halleri* and 16% in *A. neglecta* is not concordant with the total genome size expansion relative to *A. thaliana* of 64 and 156%, respectively. Only in the case of *A. arenosa* is the local expansion (26%) similar to the whole genome expansion of ~28% if, since this accession is tetraploid, we reduce the *c*-value by one-half. In allopolyploid cotton it was proposed that genome expansion of one genome was regional rather than global (Grover et al. 2004) and in rice there are conflicting reports. Ma and Bennetzen (2004) provide compelling evidence for global expansion counteracted in part by small deletions, but expansion of specific tandem repeats was shown to correlate with increased genome size (Ohmido et al. 2000). It is likely that both expansion of intergenic/non-genic islands (centromeres and other repeat-rich clusters) and the “global” proliferation of transposons contribute to genome “obesity”.

Comparative genetic analyses have primarily proceeded down two paths: genetic mapping of conserved sequences between related species (Devos and Gale 2000) or sequencing of orthologous regions between related species as has been done for barley, rice, sorghum and wheat (Ramakrishna et al. 2002) and among mammalian genomes (Thomas et al. 2003). Both approaches yield valuable information but have limitations. Genetic mapping can be laborious and it may be difficult to establish orthology whereas comparative sequencing is expensive and dependent on the ability to identify contigs spanning orthologous regions between related species. If the species are close enough, an intermediate level of information can be generated from cytogenetic approaches similar to those described here, and for a limited investment of money and time. Care should be taken, however, to verify or confirm the reference sequence for the region of interest.

References

- Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. *Proc Natl Acad Sci USA* 90:7980–7984
- Al-Shebaz IA, O’Kane SL (2002) Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). In: Somerville CR, Meyerowitz EM (eds) *The Arabidopsis book*. American Society of Plant Biologists, Rockville, MD
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C (2004) Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* 427:164–167
- Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M (1998) Colinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* 41:62–69
- Devos KM, Lydiate DJ (2000) Genome relationships: the grass model in current research. *Plant Cell* 12:637–646
- Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 7:1075–1079
- Franz PF, Alonso-Blanco C, Liharska TB, Peeters AJ, Zabel P et al (1996) High-resolution physical mapping in *Arabidopsis*

- thaliana* and tomato by fluorescence in situ hybridization to extended DNA fibers. *Plant J* 9:421–430
- Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implication in maize. *Proc Natl Acad Sci USA* 99:9573–9578
- Gazzani S, Gendall RA, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Phys* 132:1107–1114
- Grover CE, Kim HR, Wing RA, Paterson AH, Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. *Genome Res* 14:1474–1482
- Jackson SA, Wang M, Goodman H, Jiang J (1998) Fiber-FISH analysis of repetitive DNA elements in *Arabidopsis thaliana*. *Genome* 41:566–572
- Jackson SA, Cheng Z, Wang ML, Goodman HM, Jiang J (2000) Comparative FISH mapping of a 431 kb *Arabidopsis thaliana* BAC contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. *Genetics* 156:833–838
- Jiang J, Hulbert SH, Gill BS, Ward DC (1996) Interphase fluorescence in situ hybridization mapping: a physical mapping strategy for plant species with large complex genomes. *Mol Genet* 252:497–502
- Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63
- Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis* and related genera (*Brassicaceae*). *Mol Biol Evol* 17:1483–1498
- Kowalsi 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
- Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusion and frequent rearrangements. *Genetics* 150:1217–1228
- Lagercrantz U, Putterill J, Coupland G, Lydiat D (1995) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant J* 9:13–20
- Leitch IJ, Chase MW, Bennett MD (1998) Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size of flowering plants. *Ann Bot* 82(Suppl A):85–94
- Liu YG, Whittier RF (1994) Rapid preparation of megabase plant DNA from nuclei in agarose plugs and microbeads. *Nucleic Acids Res* 22:2168–2169
- Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101(34):12404–12410
- Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14:860–869
- Michaels SD, Amasino RM (1999) Flowering locus C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11:949–956
- Nishio H, Heiskanen M, Palotie A, Belanger L, Dugaiczky A (1996) Tandem arrangement of the human serum albumin multigene family in the sub-centromeric region of 4q: evolution and chromosomal direction of transcription. *J Mol Biol* 259:113–119
- Ohmido N, Kijima K, Akiyama Y, de Jong JH, Fukui K (2000) Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in indica and japonica rice. *Mol Genet* 263:388–394
- Quiros CF, Hu J, Truco MJ (1994) DNA-based *Brassica* map. In: Vasil I, Phillips RL (eds) *DNA-based markers in plants*. Kluwer Academic, Dordrecht, pp 199–222
- Ramakrishna W, Dubcovsky J, Park YJ, Busso CS, Emberton J et al (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. *Genetics* 162:1389–1400
- SanMiguel 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
- Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM et al (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424:788–793