Plant Molecular Biology **39:** 171–176, 1999.

Short communication

Genetic and physical characterization of a region of *Arabidopsis* chromosome 1 containing the *CLAVATA1* gene

Robert W. Williams**, Steven E. Clark[1,**] and Elliot M. Meyerowitz*
*Division of Biology, 156–29, California Institute of Technology, Pasadena, CA 91125, USA (*author for correspondence); [1] Department of Biology, University of Michigan, Ann Arbor, MI 48109–1048, USA; **These two authors contributed equally to this work*

Received 24 March 1998; accepted in revised form 6 July 1998

Key words: Arabidopsis, *CLAVATA1*, genome sequencing, RFLP, YAC contig

Abstract

With the advance of *Arabidopsis* as a model system for understanding plant genetics, development and biochemistry, a detailed description of the genome is necessary. As such, focused projects are underway to map and sequence the *Arabidopsis* nuclear genome. We have characterized a region of chromosome 1, surrounding the *CLAVATA1* (*CLV1*) locus. Three (RFLP) clones were mapped relative to *clv1-1*, and were used to construct an ca. 700 kb yeast artificial chromosome (YAC) contig. Three cosmids spanning the *CLV1* locus were analyzed and ca. 24 kb of genomic DNA was sequenced, including a continuous stretch of 18 kb. In addition to generating clones in this region of chromosome 1, we have analyzed the size, spacing and organization of several contiguous genes.

Mapping RFLP clones near CLV1

The *CLV1* gene, which encodes a receptor-like kinase, was previously mapped to the bottom of chromosome 1 [1, 2]. This included generating meiotic recombination breakpoints between *clv1-1* and the flanking visible markers *ap1-1* and *ga2* and using these recombinants to map the m532 RFLP clone [1, 3]. Using the same recombinants, we mapped three additional RFLP clones relative to *clv1-1*. N7-24 mapped 19/124 recombinants proximal to *clv1-1*. m237 [3] and g6838 [4] both mapped 9/82 recombinants distal to *CLV1* (Figure 1). This fine-scale mapping revealed that the order of the RFLP clones and their positions relative to *CLV1* are different than previously reported [5].

Construction of the YAC contig

To construct a physical map of the *CLV1* region, three different RFLP clones were hybridized to the yUP, EW and EG YAC libraries [6–8]. The N7-24 clone hybridized to yUP3G5, yUP7C8, yUP17F5, EW4E10 and EW6G11 (Figure 1). The restriction pattern of N7-24 hybridizing fragments was altered in the yUP7C8 clone so it was determined that N7-24 lies at one end of the YAC insert. yUP7C8 did not hybridize to a nearby proximal marker, KG-24, so yUP7C8 probably extends towards the *CLV1* gene (data not shown). Clone m532 hybridized to yUP4B3, yUP7C7, yUP13C10, and EG8D6. Clone g6838 also hybridized to yUP4B3 and EG8D6 as well as to EW1G5. yUP7C7 hybridized to m532 and m237 but not to g6838. This indicates that even though m237 and g6838 mapped the same distance from *CLV1* based on recombinants, g6838 is distal to m237.

The left and/or right ends of some of the YACs were isolated by either inverse PCR or plasmid rescue [9, 10] (Table 1). The left end of yUP13C10 hybridized to both yUP7C8 and yUP3G5, indicating that the region from markers N7-24 to m532 was contiguous (Figure 1). The left end of yUP4B3 was used to isolate a lambda genomic clone called B2. A polymorphism between L-*er* and Col-0 was identified and B2

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number AF049870.

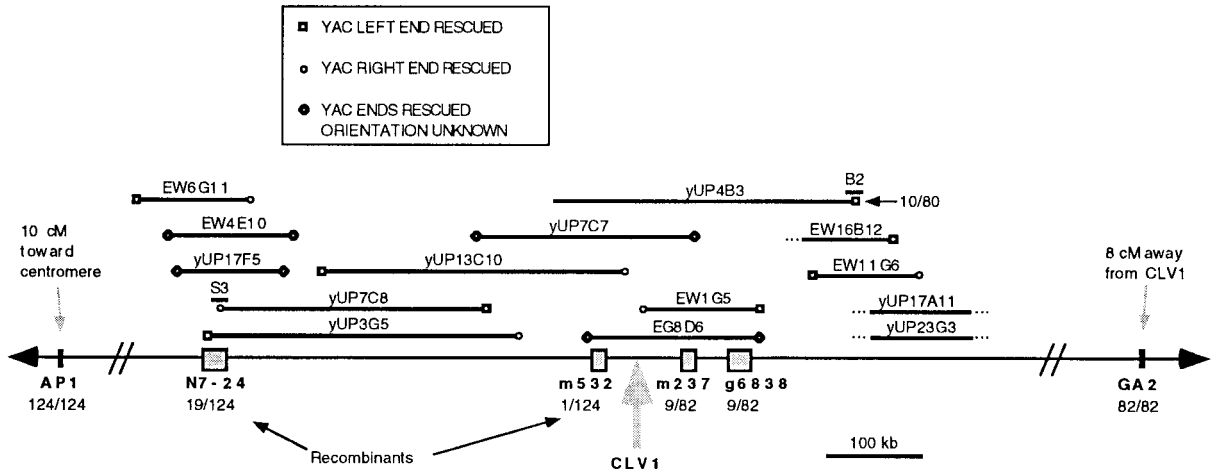

Figure 1. Genetic and physical map of the *CLV1* region. The number of recombinants is listed for the mapped molecular clones. Some data used to generate the contig was from Clark *et al.* [1].

was mapped distal to g6838. B2 was also used to isolate two more YACs, EW11G6, and EW16B12. The right end of EW11G6 was used to isolate two more YACs, yUP17A11 and yUP23G3. Neither yUP17A11 or yUP23G3 hybridized to the left end of EW11G6 so they presumably extend distally.

With the exceptions of EW16B12, yUP17A11 and yUP23G3, the size of each YAC clone was determined by pulsed field gel electrophoresis. Using the YAC sizes and some additional hybridization data summarized in Table 1, a complete physical map of the *CLV1* region was obtained (Figure 1). Based on N7-24 being ca. 2.3 cM from g6838 and the two markers being at least 490 kb apart, this gives an estimate of 170 kb/cM in this region.

Fine-scale mapping and sequencing of the CLV1 *region*

The YAC clone EG8D6 was used as a probe to isolate 31 Col-0 derived cosmid genomic clones [11] in the region spanning *CLV1*. m532 was used to screen this pool of genomic clones, and two cosmids called cosQ and cosY were purified. Both cosmids contained the same *Eco*RI polymorphism as m532 and both mapped 1/124 recombinants proximal to *CLV1*. Insert DNA from cosY and cosQ was used to isolate two more hybridizing clones, cosR and cosT, respectively. cosR mapped 3/124 recombinants proximal to *CLV1* and cosT mapped 1/82 distal to *CLV1* (Figure 2A). Therefore, the overlapping cosQ and cosT clones contained the *CLV1* locus.

Table 1. Characteristics of yeast artificial chromosomes. Each YAC is listed with the restriction enzyme(s) that was used to generate the plasmid based left-end (L.E.) probes and the restriction enzyme and primer, in conjunction with primer C69 [10], used to generate the inverse PCR based right-end (R.E.) specific probes. B2 and S3 are lambda genomic clones that correspond to the yUP4B3 left end and the yUP7C8 right end, respectively. The size of each YAC clone is given as well as the names of the clones that hybridize to it. The clones listed in italics in the 'Hybridizes' column do not hybridize to the specified YAC. HE, *Hin*cII/*Eco*RV double digest.

YAC	L.E.	R.E.	Size (kb)	Hybridizes
yUP3G5	both[1]	all	310	N7-24, *S3, 6G11le*
yUP7C8	*Nde*I	HE70 (S3)	280	N724, *6G11le*
yUP17F5	both	*Alu*71 HE70	100	N724, S3 *6G11le*
EW4E10	both	*Alu*71	120	N724, KG24, S3 *6G11le*
EW6G11	*Nde*I	*Alu*71	120	N724 *KG24, S3*
EW11G6	*Xho*I	all	110	B2, 16B12le
EW16B12	*Xho*I	all	n.d.	B2
yUP17A11	both	*Alu*	n.d.	11G6re
yUP23G3	both	all	n.d.	11G6re
yUP4B3	*Xho*I (B2)	–	320	532, 6838, B2 *16B12le*
EW1G5	*Xho*I	–	120	6838 *B2*
EG8D6	*Xho*I	all	180	532, 6838, *B2*
yUP13C10	*Xho*I	all	320	532 *6838, B2*
yUP7C7	*Xho*I	HE70	230	237, 532 *6838, B2*

[1] 'both' means that *Nde*I and *Xho*I-derived left-end plasmid rescue clones were generated.

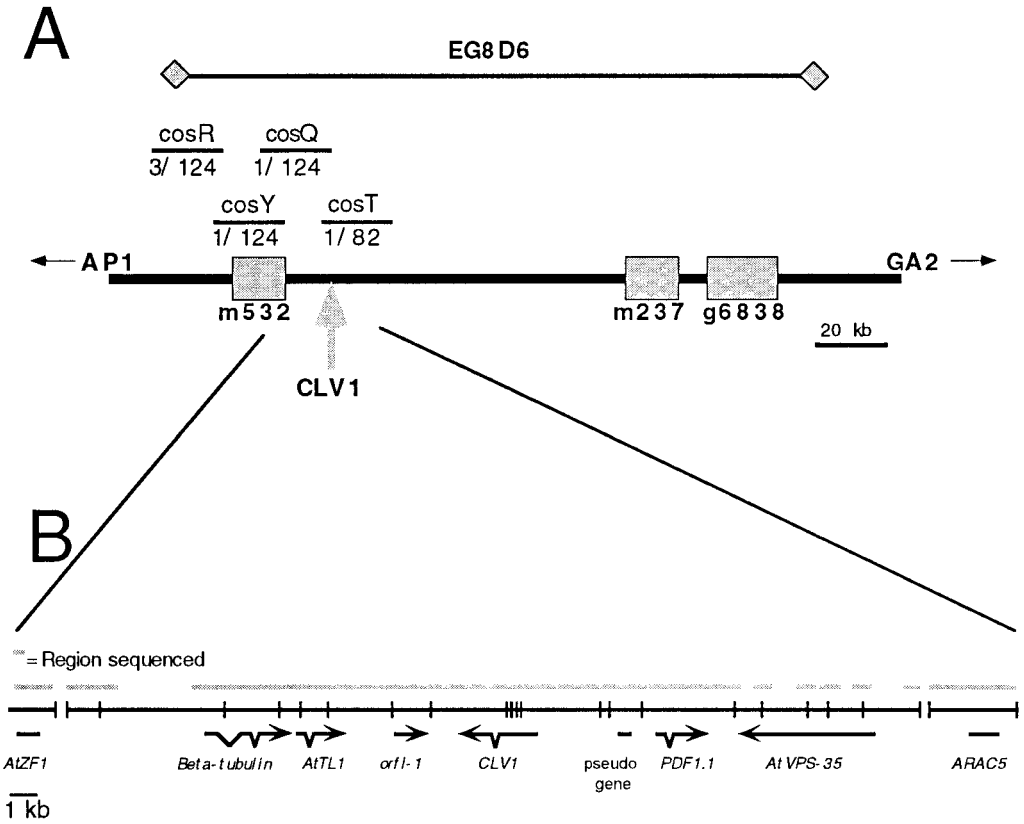

Figure 2. Fine scale mapping and sequencing in the *CLV1* region. A. The four mapped cosmids are listed with the number of recombinants. B. The *Hin*dIII restriction map and locations and direction of transcription of each ORF is given for most of cosQ and cosT and part of cosY. The position and size of the introns are given only for the genes contained in the contiguous 18 kb of genomic sequence.

Both cosQ and cosT were digested with *Hin*dIII and each fragment was cloned into the pGEM 7zf vector (Promega). Most of the fragments were sequenced in their entirety. To generate a complete *Hin*dIII map of the two cosmids, primers were designed 100 to 200 bases from each *Hin*dIII site complementary to the sequence generated from the Universal or Reverse primers. Every primer combination was used in PCR reactions with intact cosQ or cosT as the template. Primer combinations that gave PCR products of the appropriate size (ca. 300 to 400 bp) represented adjacent *Hin*dIII fragments. The identity of each PCR product was verified by sequencing. The largest continuous stretch of genomic sequence is 18 kb (Figure 2B).

Each *Hin*dIII fragment from cosQ and cosT as well as some of the *Hin*dIII fragments from cosY were used to screen a flower cDNA library [12]. Each positive cDNA clone was converted to plasmid and sequenced using a combination of Universal, Reverse and internal primers. A total of eight different genes were identified in the ca. 40 kb region.

Sequence analysis of transcription units

Three of the eight genes have previously been identified: β-tubulin, an antifungal gene called *PDF1.1* [13] and *ARAC5*. There are at least nine β-tubulin genes in the *Arabidopsis* genome [14]; the one identified here is identical to β-tubulin 1 [15]. While this β-tubulin gene is preferentially expressed in roots, it was readily detected in the flower/inflorescence cDNA library. The *ARAC5* gene was identified in a PCR based screen for Ras-like GTPase-encoding genes [16]. Rac proteins are in the Rho family of small GTPases and have been speculated to play a role in transducing signals directly to the actin cytoskeleton. There are at least 10 Rac-like genes in *Arabidopsis* [16].

Of the remaining five genes, four had sequence motifs or sequence similarities that suggest possible functions. A zinc-finger containing gene, called *AtZF1*

(corresponding to EST R30294), lies on cosY (Figure 2). The gene encodes a protein of at least 462 amino acids that contains a Cys-2/His-2 type zinc-finger in the middle of the ORF. We also isolated a cDNA clone, of unknown map location, called *AtZF2*, that is 80% identical and 88% similar to *AtZF1* (corresponding to EST R65207) over a 192 amino acid region.

Immediately distal to β-tubulin there is a thaumatin-like gene, termed *AtTL1*. Thaumatins, and osmotins, are a class of secreted proteins that are part of the plant's response to stress and/or pathogens. The gene with closest sequence similarity to *AtTL1* is an *Arabidopsis* gene called *PR5K* (Blast score of 851), which contains an extracellular thaumatin domain and a cytoplasmic serine/threonine protein kinase domain [17]. Based on the genomic sequence, the AtTL1 protein does not have a kinase domain.

The *CLV1* gene, which is just distal to *AtTL1*, is completely contained on cosT and partially contained on cosQ. It encodes a putative receptor-kinase, with an extracellular domain comprised of 21 leucine-rich repeats. Further analysis of this gene has been presented elsewhere [1].

Distal to *PDF1.1* there is a gene encoding a protein with sequence similarity to a maternally transcribed mouse gene called MEM3 and the yeast vacuolar sorting protein VPS35 [18]. The closest gene in the database is another *Arabidopsis* gene identified while sequencing a BAC on chromosome IV (AF024504). The *VPS35* is required in yeast for proper targeting of a subset of vacuolar proteins. The *Arabidopsis* gene is called *AtVPS-35* [1].

In between *AtTL1* and *CLV1* lies a gene that encodes a small, novel protein. This gene, called *orfI-1*, has an open reading frame of 141 amino acids and contains a putative signal sequence. This protein may be secreted. The remainder of the deduced amino acid sequence has no recognized motifs, but the protein is largely hydrophobic.

Genomic organization of the genes and intergenic regions

The cDNA and genomic sequences were compared for five of the genes: β-tubulin 1, *AtTL1*, *orfI-1*, *CLV1* and *PDF1.1* (table 2). All five genes are contained in the 18 kb of continuous genomic sequence discussed above. The β-tubulin gene contains two introns, one of 794 bp and the other of 88 bp. The *AtTL1* gene contains an 80 bp intron after the putative signal sequence and *PDF1.1* contains a 107 bp intron after the signal sequence. The *CLV1* gene contains a 79 base intron in a conserved position in many other serine/threonine kinase genes [1, 19]. All the intron splice sites conform to the consensus for *Arabidopsis* (A/C)AG|GTAAGT(A/T)-(C/T)AG|G(A/T). Furthermore, the codon usage of these five genes closely follows that of a variety of other *Arabidopsis* genes (http://www.dna.affrc.go.jp/~nakamura/CUTG.html).

Four of the five ORFs are closely spaced, with 585 bp between β-tubulin 1 and *AtTL1* and 2 kb between the *AtTL1* gene and *orfI-1*. There is a larger 5.7 kb gap between *CLV1* and *PDF1.1*, however, located between these two genes is a 104 bp region with sequence similarity to *PDF1.1* that is probably the result of a duplication event. There is a stop codon in the *PDF1.1* pseudo-gene, so it likely does not represent a functional gene.

Analysis of the nucleotide content of coding versus non-coding sequences indicates that coding sequences are noticeably more GC-rich than non-coding regions. On average, intergenic and intron sequences are 60–80% AT whereas coding regions are only 40–60% AT. By plotting %AT content of the 18 kb region with a 50 base window size, the ORFs are easily distinguished as regions where the AT content drops below 60% (data not shown).

Significance

The data reported here confirm several predictions about the *Arabidopsis* genome. First, we estimate that there is 170 kb/cM in the region analyzed. This is similar to previous estimates of 150–200 kb/cM [3]. The density of genes observed in this study, which corresponds to ca. 20 000 genes per haploid genome, is consistent with previous estimates for *Arabidopsis* and other angiosperms [20, 21]. In this region of the genome, the splice site sequences and codon bias closely resemble those for a number of other *Arabidopsis* genes. It is interesting to note that two of the genes in the region, *AtTL1* and *PDF1.1*, contain introns immediately following a potential signal sequence.

In addition to reporting the sequence and exact map positions of eight genes, the data presented have several additional uses. First, the continuous genomic sequence matches up with four BAC end sequences in the TIGR database (T16B16TRD, F05H06Sp6, T19O22TF and T20K17TR/T20M12TR) thus mapping those BAC clones to chromosome 1. In addition,

Table 2. List of identified open reading frames as well as any homology or sequence motifs that indicate a function.

Gene	Homology	Function	Blast Score*	Accession	EST
AtZF1	zinc finger	transcription factor?			R30294
β-tubulin 1	*β-tubulin*	cytoskeleton		M20405	W43629
AtTL1	*PR5K*	stress/pathogen response?	851(PRK5)	AF049870	H37452
orfI-1	novel	secreted?		AF049870	none
CLV1	receptor-kinase	meristem development		U96879	R29811
PDF1.1	cysteine-rich	pathogen defense		Z27258	none
AtVPS-35	*Mem3/VPS35*	vacuolar sorting?	222(AF024504)	U96877/U96878	none
ARAC5	GTP-binding	signal transduction		U52350	N97025

*For the genes presented here that are not identical to known genes but have sequences that suggest a function, BLAST scores are given.

the molecular clones generated here could be used as entry points for positional cloning of other genes that map near *CLV1*. For example, *EMB25*, *EMBRG3*, *PIN*, *IAR1*, *EMB54*, *EMB17*, *EMB120*, *BOT1*, *EMB156*, and *SUS2* all map within 10 cM of *CLV1* (D. Meinke, http://mutant.lse.okstate.edu/).

Finally, as many different plant species have vastly different genome sizes but have regions where the genetic maps are similar [22], these data may be used to identify the cognate homologue of any gene in this region from other flowering plant species. This is particularly useful since many plant genes, including *CLV1* [1] and *AtZF1*, appear to be members of gene families. Identifying the orthologue of a gene in another plant species can be difficult but may be aided by knowing which genes map near the gene of interest.

Acknowledgements

We thank Hajime Sakai and Leonard Medrano for technical advice. We also acknowledge Xuemei Chen, Jennifer Fletcher, Toshiro Ito, Prakash Kumar, Jose-Luis Riechmann, Kevin Roberg, Mark Running and Eva Ziegelhoffer for careful review of this manuscript. *ga2* seeds were provided by the Arabidopsis Biological Resource Center at The Ohio State University. This work was supported by NSF grant MCB-9204839 and a Strategic Research Fund Grant from Zeneca Seeds to E.M.M. S.E.C. was a NSF Plant Biology Postdoctoral Fellow and R.W.W. was supported by NIH Predoctoral Training Grant GM07616 and the Howard Hughes Medical Institute.

References

1. Clark SE, Williams RW, Meyerowitz EM: The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. Cell 89: 575–585 (1997).
2. Koornneef M, Van Eden J, Hanhart CJ, Stam P, Braaksma FJ, Feenstra WJ: Linkage map of *Arabidopsis thaliana*. J Hered 74: 265–272 (1983).
3. Chang C, Bowman JL, De John AW, Lander ES, Meyerowitz EM: Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85: 6856–6860 (1988).
4. Nam HG, Giraudat J, Denboer B, Moonan F, Loos WDB, Hauge BM, Goodman HM: Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. Plant Cell 1: 699–705 (1989).
5. Hauge BM, Hanley SM, Cartinhour S, Cherry JM, Goodman HM *et al.*: An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. Plant J 3: 745–754 (1993).
6. Guzman P, Ecker JR: Development of large DNA methods for plants – molecular-cloning of large segments of *Arabidopsis* and carrot DNA into yeast. Nucl Acids Res 16: 11091–11105 (1988).
7. Ward ER, Jen GC: Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. Plant Mol Biol 14: 561–568 (1990).
8. Grill E, Somerville C: Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. Mol Gen Genet 226: 484–490 (1991).
9. Putterill J, Robson F, Lee K, Coupland G: Chromosome walking with YAC clones in *Arabidopsis*: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. Mol Gen Genet 239: 145–157 (1993).
10. Schmidt R, Cnops G, Bancroft I, Dean C: Construction of an overlapping YAC library of the *Arabidopsis thaliana* genome. Aust J Plant Physiol 19: 341–351 (1992).
11. Olszewski NE, Martin FB, Ausubel FM: Specialized binary vector for plant transformation-expression of the *Arabidopsis thaliana* AHAS gene in *Nicotiana tabacum*. Nucl Acids Res 16: 10765–10782 (1988).
12. Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM: *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell 69: 843–859 (1992).

13. Penninckx IAMA, Eggermont K, Terras FRG, Thomma BPHJ, De Samblanx GW Buchala A, Metraux J-P, Manners JM, Broekaert WF: Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway. Plant Cell 8: 2309–2323 (1996).
14. Snustad DP, Haas NA, Kopczak SD, Silflow CD: The small genome of *Arabidopsis* contains at least 9 expressed beta-tubulin genes. Plant Cell 4: 549–556 (1992).
15. Oppenheimer DG, Haas N, Silflow CD, Snustad DP: The beta-tubulin gene family of *Arabidopsis thaliana*: preferential accumulation of the beta 1 transcript in roots. Gene 63: 87–102 (1988).
16. Winge P, Brembu T, Bones AM: Cloning and characterization of rac-like cDNAs from *Arabidopsis thaliana*. Plant Mol Biol 35: 483–495 (1997).
17. Wang XQ, Zafian P, Choudhary M, Lawton M: The PR5K receptor protein-kinase from *Arabidopsis thaliana* is structurally related to a family of plant defense proteins. Proc Natl Acad Sci USA 93: 2598–2602 (1996).
18. Paravicini G, Horazdovsky BF, Emr SD: Alternative pathways for the sorting of soluble vacuolar proteins in yeast: a *vps35* null mutant missorts and secretes only a subset of vacuolar hydrolases. Mol Biol Cell 3: 415–427 (1992).
19. Chang C, Schaller GE, Patterson SE, Kwok SF, Meyerowitz EM, Bleecker AB: The TMK1-gene from *Arabidopsis* codes for a protein with structural and biochemical characteristics of a receptor protein-kinase. Plant Cell 4: 1263–1271 (1992).
20. Meyerowitz EM: Structure and organization of the *Arabidopsis* nuclear genome. In Meyerowitz EM, Somerville CR (eds), Arabidopsis, pp. 21–36. Cold Spring Harbor Laboratory Press, Plainview, NY (1994).
21. Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Vanstaveren-M *et al.*: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391: 485–488 (1998).
22. Shields R: Pastoral synteny. Nature 365: 297–298 (1993).