to new protein synthesis or recycling from the bud. Also, latrunculin-A (20 min, 500 μM) did not cause a shift in GFP-Ist2p from bud to mother cell in wild-type or *cdc12-6* cells at the permissive temperature. Thus, actin is not required to maintain GFP-Ist2p in the bud.
18. Y. Barral, V. Mermall, M. S. Mooseker, M. Snyder, *Mol. Cell* **5**, 841 (2000).
19. E. Rodriguez-Boulan and W. J. Nelson, *Science* **245**, 718 (1989).
20. M. S. Balda *et al.*, *J. Cell Biol.* **134**, 1031 (1996).
21. B. Winckler, P. Forscher, I. Mellman, *Nature* **397**, 698 (1999).
22. D. Zhou *et al.*, *J. Cell Biol.* **143**, 1295 (1998).
23. C. L. Nehme, M. M. Cesario, D. G. Myles, D. E. Koppel, J. R. Bartles, *J. Cell Biol.* **120**, 687 (1993).
24. Immunoprecipitation of She2p-myc, She3p-myc, Myo4p-myc and untagged control were performed as described (*8*). RNA fractions were prepared from the immunoprecipitates by extraction of the eluate with phenol/chloroform and precipitation with ethanol. The RNA was reverse-transcribed with Superscript II (Life Technologies, Rockville, MD) according to the manufacturer's instructions and using the primer, GTTTCCCAGTCACGATC$(N)_6$. After the reverse transcription, the reactions were incubated for 2 min at 94°C and then held at 8°C. Four units of T7 sequenase (Amersham Pharmacia, Uppsala) was added, and the reaction was ramped to 37°C at 0.1°C per second and then held for 8 min at 37°C. The sequenase step was repeated once. A portion of the reaction was then amplified by PCR using the sequence-specific end of the first primer. The resulting amplified DNA was fluorescently labeled by a subsequent round of PCR (25 cycles) in presence of either Cy3-dUTP or Cy5-dUTP (Amersham). These reactions were purified using the QiaQuick PCR purification kit (Qiagen, Valencia, CA), and then applied to the DNA microarray. DNA microarrays were fabricated and hybridized as described (*19*). Genepix software (Axon Instruments, Forest City, CA) was used for image analysis and quantification. DNA microarray analysis was performed in duplicate.
25. The *IST2* gene was tagged with the method described in (*30*). All strains were derived from w303. Strains containing pGAL-GFP-*IST2* were grown in rich medium (YP) containing 2% raffinose. At absorbance of 0.5 U, galactose was added to 2%, and the cultures were incubated for 1 to 1.5 hours at 30°C. Cells were fixed with 4% formaldehyde and then examined on a Zeiss Axioplan fluorescence microscope. Images were captured with a CCD camera and processed with Adobe Photoshop.
26. To perform FRAP, wild-type, W303 cells containing pGAL-GFP-*IST2* cells were induced (*25*) and then incubated in 2% dextrose for 10 min to turn off production of GFP-Ist2p. Cells were examined on a Leica TCS NT confocal microscope, and a portion of a large bud was photobleached for 1 s. An image was captured immediately after the photobleaching and then every 2 s thereafter. Images were analyzed using NIH Image and Adobe Photoshop.
27. To analyze the effect of a septin mutant on the localization of GFP-Ist2p, *cdc12-6* and its parental strain, A364, containing pGAL-GFP-*IST2* were grown in rich medium (YP) with 2% raffinose at 24°C until absorbance 0.5 U. Galactose was added to 2%, and the cultures incubated for 1 hour at 24°C. Dextrose was added to 2%, and the cultures were incubated for 10 min at 24°C. A sample was removed, and the cultures were shifted to 37°C for 10 min, after which time a second sample was taken. All samples were fixed in 4% formaldehyde and examined as described (*25*). Images were captured and processed as described. Large-budded cells (100 in each sample) were analyzed and counted for asymmetric GFP-Ist2p.
28. K. D. Entian *et al.*, *Mol. Gen. Genet.* **262**, 683 (1999).
29. D. Drubin, personal communication.
30. M. S. Longtine *et al.*, *Yeast* **14**, 953 (1998).
31. We thank A. Sil, I. Herskowitz, and C. Takizawa for critical review of the manuscript. We would also like to thank members of the Herskowitz, Walter, Murray, and O'Shea laboratories for their help in producing the microarrays used in this study. This work was supported in part by National Institutes of Health grant 38496 (R.D.V.), a grant from the Jane Coffin Childs Memorial Fund for Medical Research Fellowship (P.A.T.), and the Sandler Family (J.D.R).

15 June 2000; accepted 24 August 2000

Molecular Analysis of *FRIGIDA*, a Major Determinant of Natural Variation in *Arabidopsis* Flowering Time

Urban Johanson,[1]* Joanne West,[1]† Clare Lister,[1] Scott Michaels,[2] Richard Amasino,[2] Caroline Dean[1]‡

Vernalization, the acceleration of flowering by a long period of cold temperature, ensures that many plants overwinter vegetatively and flower in spring. In *Arabidopsis*, allelic variation at the *FRIGIDA* (*FRI*) locus is a major determinant of natural variation in flowering time. Dominant alleles of *FRI* confer late flowering, which is reversed to earliness by vernalization. We cloned *FRI* and analyzed the molecular basis of the allelic variation. Most of the early-flowering ecotypes analyzed carry *FRI* alleles containing one of two different deletions that disrupt the open reading frame. Loss-of-function mutations at *FRI* have thus provided the basis for the evolution of many early-flowering ecotypes.

A requirement for vernalization—the acceleration of flowering that occurs during a 3- to 8-week period of cold temperature (4°C)—has been bred in many crops to produce winter/spring varieties. Vernalization requirement is also a major factor in determining flowering time in *Arabidopsis thaliana* ecotypes. Despite the large number of genes known to control flowering time (*1*), the vernalization requirement segregates as a single gene trait (*2–6*) mapping to the *FRIGIDA* (*FRI*) locus (*7*). This locus was first described by Napp-Zinn (*8*), who analyzed the progeny of a cross between the late-flowering ecotype Stockholm and the early-flowering ecotype Li5. The action of an active *FRI* allele depends on an active *FLC* allele (*9, 10*).

To analyze the molecular basis of the allelic variation at *FRI*, we cloned the gene using map-based techniques (Fig. 1) (*11*). *FRI* is a single-copy gene in the *Arabidopsis* genome and encodes a predicted open reading frame (ORF) of 609 amino acids (GenBank accession numbers: genomic *FRI*, AF228499; cDNA, AF228500). The predicted protein shows no significant match to any protein or protein domain of known function in available databases. The *FRI* protein is predicted to contain coiled-coil domains in two positions (between amino acids 55 to 100 and 405 to 450, respectively). *FRI* has been shown to increase RNA levels of *FLC*, which encodes a MADS-box protein likely to act as a transcriptional repressor (*12, 13*). Whether the predicted coiled coils in the FRI protein are important for this function remains to be tested.

The rapid-cycling ecotypes Columbia (Col) and Landsberg *erecta* (L*er*) carry recessive *FRI* alleles. To analyze the basis of the recessivity, we compared ~3.6 kb of genomic sequence from the dominant H51 [a derivative of the late-flowering ecotype Stockholm (*8*)] *FRI* allele with the same region from Col and L*er*. Ten polymorphisms were found between H51 and Col. Two result in amino acid differences ($Gly^{146} \rightarrow Glu$ and $Met^{148} \rightarrow Ile$, respectively, the former resulting in loss of a Bsm FI restriction site); another is a 16–base pair (bp) deletion at the end of exon 1, which changes the reading frame and terminates the ORF immediately at the beginning of exon 2 (Fig. 2).

Three differences were detected between H51 and L*er* *FRI* alleles: two single-base changes that do not alter the ORF, and a 376-bp deletion combined with a 31-bp insertion that disrupts the beginning of the ORF, removing the putative translation start codon (Fig. 2). The additional 31 bp appear to

[1]Department of Molecular Genetics, John Innes Centre, Norwich NR4 7UH, UK. [2]Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA.

*Present address: Department of Plant Biochemistry, Lund University, Post Office Box 117, SE-221 00 Lund, Sweden.

†Present address: Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK.

‡To whom correspondence should be addressed. E-mail: caroline.dean@bbsrc.ac.uk

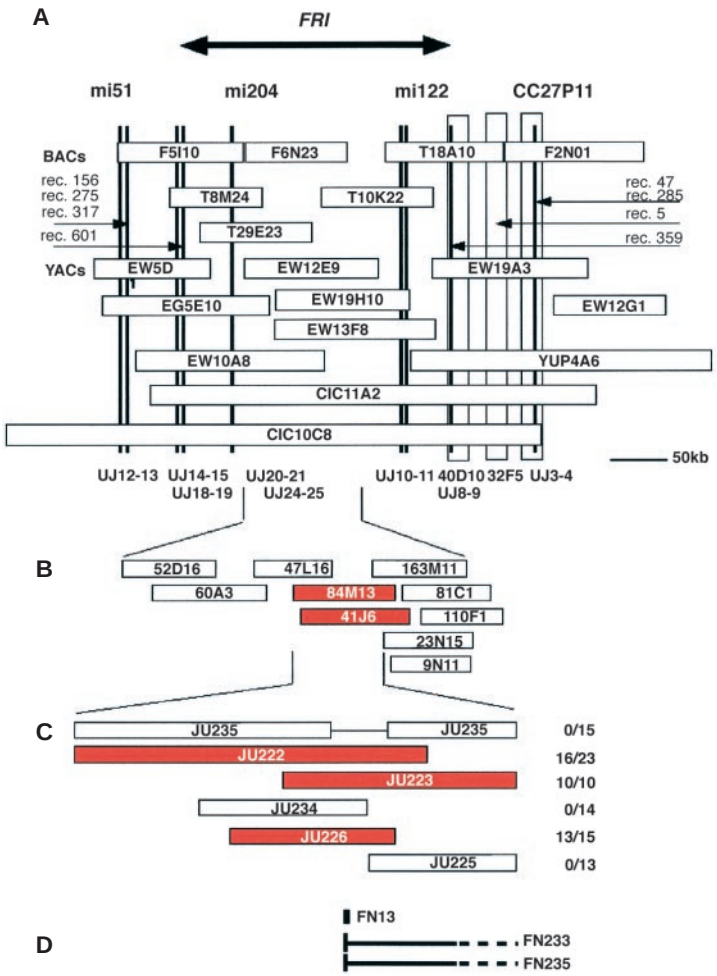

Fig. 1. Cloning of *FRIGIDA*. (**A**) Genetic interval, molecular markers, recombination positions, and yeast and bacterial artificial chromosome (YAC and BAC) clones covering *FRI*. Markers with mi and CC prefixes are described in (*22*), and those with UJ prefixes are new CAPS (cleaved amplified polymorphic sequence) markers distinguishing H51; Col DNA, 40D10, and 32F5 are H51 cosmids. Breakpoints in different recombinants are indicated at ends of horizontal arrows. IGF and TAMU BAC clones carry F and T prefixes, respectively. The different YAC clones are prefixed CIC, EW, EG, and YUP. The vertical lines and rectangles represent the extent of the region covered by the respective marker (*11*). (**B**) H51 cosmid clones covering the BAC clone F6N23. Cosmids complementing the flowering time phenotype (i.e., changing it from early to late) are shaded in red. (**C**) Subclones of cosmid 84M13, with complementing subclones in red. JU235 is an internal deletion of 84M13. The number of late-flowering individuals per number of transformants generated is indicated at the right. (**D**) Extent of deletions and rearrangements in three fast neutron–induced early-flowering mutants, FN13, FN233, and FN235 [from a population of >40 early mutants in the *FRI(Sf-2), FLC-Col* background (*9*)]. Southern analysis showed that FN233 and FN235 carried intact *FRI* promoter regions but lacked most of the coding region of *FRI*. FN13 could be interpreted as having an insertion or a complex rearrangement within a 600-bp Pst I–Sac I fragment of exon 1 of *FRI*, internal to the Sca I fragment missing in subclone JU235. (**E**) Introduction of H51 *FRI* allele into Li5. Wild-type Li5 is shown on the left. A late-flowering primary transformant following *Agrobacterium*-mediated transformation of Li5 with the cosmid 84M13 is shown on the right. Both plants were photographed once they had flowered (after ~3 months for the transformant and 3 weeks for the parent).

be a partial duplication of the subsequent 52 bp, and it carries an ATG codon. If translation occurred, this would very likely act as the start codon and would yield a short out-of-frame 41–amino acid peptide. No *FRI* transcript is detected in L*er* seedlings when transcript can be detected in both H51 and Col seedlings.

To survey a more extensive set of *FRI* alleles, we designed polymerase chain reaction–based tests to score the 16-bp and 376-bp deletions and the Gly → Glu amino acid polymorphism, the latter detected by the presence or absence of the Bsm FI restriction site. We analyzed 38 additional ecotypes that had been classified as early- or late-flowering in the absence of vernalization (*10, 14, 15*). The results of this analysis (Table 1) enable a classification of the ecotypes into five groups. Group 1, represented by Col, contains early-flowering ecotypes having *FRI* alleles with the 16-bp deletion and lacking the Bsm FI site. Group 2, represented by L*er*, contains early-flowering ecotypes having *FRI* alleles with the 376-bp deletion, the 31-bp insertion, and the Bsm FI site. Group 3, represented by H51, and group 4 are all late-flowering ecotypes and have *FRI* alleles with or without the Bsm FI site, respectively, but none have either of the two deletions. Group 5, however, contains six early-flowering ecotypes that (on the basis of the polymorphism tests) carry functional *FRI* alleles.

These results suggest that the *FRI* allele from the late-flowering ecotypes is the ancestral form of the gene, with early flowering evolving independently at least twice from late-flowering ecotypes through deletion events leading to loss of *FRI* function. Group 1 early-flowering ecotypes would seem to be derived from a group 4 ecotype and group 2 from a group 3 ecotype. The late flowering of groups 3 and 4 supports the interpretation that the Gly → Glu and Met → Ile difference does not contribute to the earliness of group 1 ecotypes.

FRI was sequenced from the ecotypes Sf-2 (group 3) and EDI-0 (group 4). The only difference between the EDI-0 and Col alleles was the 16-bp deletion, which suggested that the

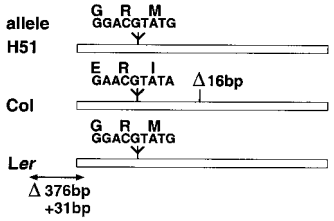

Fig. 2. Molecular structure of different *FRI* alleles. The *FRI* gene contains two introns, 393 and 89 bp long, positioned 954 and 1520 bp downstream of the Met residue of the likely translation start codon. The genomic regions of H51, Col, and L*er* alleles are represented schematically, with the changes in nucleotide and amino acid sequence shown and the positions of the deletions indicated.

deletion occurred relatively recently. Four polymorphisms distinguished Sf-2 and L*er*, one of which was the 376-bp deletion and 31-bp insertion. One other led to a change in the protein (a conservative amino acid substitution, Leu[79] → Ile). The greater number of differences between Sf-2 and L*er* *FRI* alleles as compared to

EDI-0 and Col suggests that the deletion that gave rise to the group 2 ecotypes predated the deletion that gave rise to the group 1 ecotypes. L*er* also carries recessive *FLC* alleles, so it is possible that changes at *FLC* were the primary cause of earliness in L*er* and the *FRI* deletion was secondary. It will be important to determine the dominance of the *FLC* alleles in the other group 2 ecotypes to further address this question.

A *FRI* allele from a group 5 ecotype was also sequenced to establish whether a different mutational event accounted for the earliness of this group. The *FRI* allele from the ecotype Shakhdara, from Tadjikistan, did not contain any deletions or rearrangements relative to the H51 allele. The Shakhdara allele showed six nucleotide differences in comparison to H51, with only one polymorphism resulting in an amino acid substitution ($Phe^{55} \rightarrow Ile$). Northern analysis showed that Shakhdara *FRI* is expressed at the same level as *FRI* in the late-flowering H51 parent. In addition, *FRI* genotype analysis of late- and early-flowering F_2 individuals from a Shakhdara × Columbia cross showed that the Shakhdara *FRI* allele confers late flowering (*16*). These data suggest that the Shakhdara *FRI* allele is fully functional and cannot account for the early flowering of Shakhdara. Crosses of Shakhdara to *FRI flc* and *fri flc* genotypes [an *flc* mutant from the *FRI*/*FLC* mutagenesis (*13*) with or without the active *FRI* allele] resulted in F_1 plants that were early flowering (~12 leaves), whereas crosses to *fri FLC* (wild-type Col) resulted in late-flowering F_1 plants (>35 leaves). Thus, Shakhdara, like L*er*, appears to carry recessive *FLC* alleles, which could account for its early-flowering phenotype. Another ecotype in group 5, Cvi, has dominant *FLC* alleles (*10, 17*), so *FLC* variation cannot account for the earliness of all the group 5 ecotypes. Whether the other group 5 ecotypes carry *FRI* alleles with an as yet unidentified lesion, or whether genes other than *FRI* and *FLC* account for their earliness, remains undetermined. It is interesting that mutations at *FRI* are the basis of earliness in the majority of early-flowering ecotypes, as *fri* mutations accounted for relatively few (3 from >40) of the early-flowering mutants in the *FRI*/*FLC* mutagenesis. This may indicate that loss of function of other genes incurs a fitness penalty in the natural environment, or that the strong phenotypic change caused by *fri* mutations conferred the greatest selective advantage.

The independent appearance of early-flowering ecotypes in the evolution of *Arabidopsis* suggests that there has been strong selection in some environments for ecotypes that do not require vernalization. The longitude and latitude coordinates for the collection points of the ecotypes are shown in Table 1. The ecotypes show a north-south distribution that is significantly different from random ($0.1 > P > 0.05$, Wilcoxon two-sample test). The majority of the late-flowering ecotypes are from northern latitudes, whereas most of the early-flowering ecotypes were collected from central and eastern Europe. This might suggest that a strong vernalization requirement conferred by active *FRI* alleles is a selective advantage in northern regions but a disadvantage in localities where winters are milder. However, there was no correlation between flowering time and mean temperatures (December–February or September–October, 1961 through 1990) or altitude for the ecotypes shown in Table 1. Variation in flowering times of ecotypes collected from the same locality indicates the complexity of the selective forces in different microenvironments (*15, 18*). Flowering early without vernalization may be an advantage where severe winter conditions prevent germination [e.g., the Shakhdara ecotype (*18*)], in climates that would support more than one generation per year, or where there is a selective advantage

Table 1. Classification of *Arabidopsis* ecotypes based on *FRI* alleles and flowering time. Flowering times (FT) of different ecotypes are from (*10, 14, 15*); ecotypes were classified as flowering early (E) or late (L), the latter being scored if the flowering time without vernalization was >75 days or the plants had more than 10 leaves. Where discrepancies occurred in the published data, flowering time was taken from our own unpublished observations. Stock center accession numbers of the ecotypes are included where available. The promoter difference was scored (+, present; –, deletion) using primers 5′-AGTACTCACAAGTCACAAC-3′ and 5′-GAAGATCATCGAATTGGC-3′. The presence of the Gly residue was scored by a CAPS marker (primers 5′-CCATAGACGAATTAGCTGC-3′ and 5′-AGACTCCAGTATAAGAAG-3′) and the presence of a Bsm F1 site. The 16-bp difference was scored (+, present; –, deletion) using primers UJ26 (5′-AGATTTGCTGGATTTGATAAGG-3′) and UJ34 (5′-ATATTTGATGTGCTCTCC-3′). Longitude and latitude coordinates are given for collection points of different ecotypes, except for Columbia whose geographical origin is unclear.

Ecotype	Accession no.	FT	Promoter	Gly	16 bp	Longitude	Latitude
			Group 1				
Col	933	E	+	–	–		
Li-5	3178	E	+	–	–	8°E	50°N
En	1136	E	+	–	–	8°E	50°N
Ws	1638	E	+	–	–	30°E	52°N
Nd	1390	E	+	–	–	9°E	53°N
MT-0	1380	E	+	–	–	38°E	56°N
Köln	6003	E	+	–	–	7°E	51°N
OY-0	1436	E	+	–	–	5°E	61°N
			Group 2				
L*er*	20	E	–	+	+	15°E	52°N
TSU-0	1564	E	–	+	+	136°E	34°N
RSCH-4	1494	E	–	+	+	34°E	56°N
Dijon	910	E	–	+	+	5°E	47°N
Gr		E	–	+	+	16°E	47°N
			Group 3				
H51		L	+	+	+	18°E	59°N
Sf-2	1516	L	+	+	+	3°E	42°N
Te	1550	L	+	+	+	26°E	63°N
Ko	1288	L	+	+	+	12°E	55°N
Can	1064	L	+	+	+	15°W	28°N
Vimmerby		L	+	+	+	16°E	57°N
Lisse	6033	L	+	+	+	5°E	52°N
PU-2-3		L	+	+	+	15°E	50°N
GOT-32		L	+	+	+	10°E	52°N
Lund		L	+	+	+	13°E	56°N
Kent	6009	L	+	+	+	1°E	51°N
Västervik		L	+	+	+	16°E	57°N
TAMM-46		L	+	+	+	23°E	60°N
KZ-13		L	+	+	+	65°E	48°N
PU-2-8		L	+	+	+	15°E	50°N
MS-0	905	L	+	+	+	38°E	56°N
			Group 4				
EDI-0	1122	L	+	–	+	3°W	56°N
NC-6		L	+	–	+	78°W	35°N
DEM-4		L	+	–	+	86°W	40°N
Algutsrum		L	+	–	+	16°E	57°N
DEM-2		L	+	–	+	86°W	40°N
BG-2		L	+	–	+	123°W	42°N
			Group 5				
Cvi	1096	E	+	+	+	24°W	16°N
Wil	1594	E	+	+	+	25°E	55°N
Shakhdara	929	E	+	+	+	70°E	39°N
KZ-9		E	+	+	+	65°E	48°N
Kondara	916	E	+	+	+	70°E	39°N

in escaping agricultural harvesting, succession, or summer drought (*19*). The selective forces driving the molecular evolution of flowering time genes can now be examined in more detail.

There was no clear geographical association of the ecotypes within the groups defined by the *FRI* alleles, or between the early-flowering group 1 ecotypes and the late-flowering group 4 ecotypes from which they may have arisen. Human-induced dispersal has been a major factor in the recent spread of *Arabidopsis* ecotypes and has resulted in little association of geographical and genetic distance (*20, 21*). Human disturbance regularly exposes *Arabidopsis* ecotypes to novel environments, thus maintaining a strong selective pressure for adaptive mutations. This may account for loss-of-function mutations providing the basis for the evolutionary changes in *Arabidopsis* flowering time.

References and Notes

1. G. G. Simpson, A. R. Gendall, C. Dean, *Annu. Rev. Cell Dev. Biol.* **15**, 519 (1999).
2. J. E. Burn, D. R. Smyth, W. J. Peacock, E. S. Dennis, *Genetica* **90**, 145 (1993).
3. J. H. Clarke and C. Dean, *Mol. Gen. Genet.* **242**, 81 (1994).
4. H. Kuittinen, M. J. Sillanpaa, O. Savolainen, *Theor. Appl. Genet.* **95**, 573 (1997).
5. I. Lee, A. Bleecker, R. Amasino, *Mol. Gen. Genet.* **237**, 171 (1993).
6. S. Sanda, M. John, R. Amasino, *J. Hered.* **88**, 69 (1997).
7. K. Napp-Zinn, *Z. Indukt. Abst. Vererb.* **88**, 253 (1957).
8. _____, *Naturwissenschaften* **42**, 650 (1955).
9. I. Lee, S. D. Michaels, A. S. Masshardt, R. M. Amasino, *Plant J.* **6**, 903 (1994).
10. M. Koornneef, H. Blankestijn-de Vries, C. Hanhart, W. Soppe, T. Peeters, *Plant J.* **6**, 911 (1994).
11. For supplementary information on the details of the map-based cloning of *FRIGIDA*, including primer sequences, see *Science* Online (www.sciencemag.org/feature/data/1053859.shl).
12. C. C. Sheldon *et al.*, *Plant Cell* **11**, 445 (1999).
13. S. D. Michaels and R. M. Amasino, *Plant Cell* **11**, 949 (1999).
14. B. H. Karlsson, G. R. Sills, J. Nienhuis, *Am. J. Bot.* **80**, 646 (1993).
15. M. Nordborg and J. Bergelson, *Am. J. Bot.* **86**, 470 (1999).
16. J. Hagenblad, U. Johanson, M. Nordborg, unpublished data.
17. C. Alonso-Blanco, S. El-Din El-Assal, G. Coupland, M. Koornneef, *Genetics* **149**, 749 (1998).
18. K. K. Khurmatov, *Arab. Inf. Serv.* **19**, 63 (1982).
19. T. Mitchell-Olds, *Evolution* **50**, 140 (1996).
20. J. Bergelson, E. Stahl, S. Dudek, M. Kreitman, *Genetics* **148**, 1311 (1998).
21. T. F. Sharbel, B. Haubold, T. Mitchell-Olds, *Mol. Ecol.*, in press.
22. R. Schmidt *et al.*, *Plant J.* **9**, 755 (1996).

23. We thank M. Smith for excellent care of *Arabidopsis*, K. Torney for construction and transformation of the *35S-FRI* cDNA construct, J. Clarke for ecotype DNA, N. Hartley for sequence analysis, J. Hagenblad and M. Nordborg for sharing unpublished results and seeds, and M. Nordborg, E. Coen, G. Simpson, T. Mitchell-Olds, and D. Stern for invaluable comments. We also thank R. Martienssen and M. Marra for unpublished BAC information. Supported by an EMBO fellowship (U.J.) and a UK Biotechnology and Biological Sciences Research Council Core Strategic Grant to the John Innes Centre.

10 July 2000; accepted 17 August 2000

REPORTS

Selfish DNA in Protein-Coding Genes of *Rickettsia*

Hiroyuki Ogata,[1] Stéphane Audic,[1] Valérie Barbe,[2] François Artiguenave,[2] Pierre-Edouard Fournier,[3] Didier Raoult,[3] Jean-Michel Claverie[1]*

***Rickettsia conorii*, the aetiological agent of Mediterranean spotted fever, is an intracellular bacterium transmitted by ticks. Preliminary analyses of the nearly complete genome sequence of *R. conorii* have revealed 44 occurrences of a previously undescribed palindromic repeat (150 base pairs long) throughout the genome. Unexpectedly, this repeat was found inserted in-frame within 19 different *R. conorii* open reading frames likely to encode functional proteins. We found the same repeat in proteins of other *Rickettsia* species. The finding of a mobile element inserted in many unrelated genes suggests the potential role of selfish DNA in the creation of new protein sequences.**

Selfish DNA (*1*, *2*)—repeated elements without obvious cellular function—is thought to be an important factor in genome evolution. Transposons and other extragenic interspersed repeats are responsible for gene (or exon) shuffling and duplication, as well as regulatory changes (*3*–*5*). However, those mechanisms cannot account for de novo creation of protein domains. The finding, in *Rickettsia conorii*, of a palindromic interspersed repeat inserted in several unrelated protein-coding sequences now suggests that selfish DNA could directly participate in the creation of new protein sequences.

Rickettsiae, the closest extant relatives of mitochondria (*6*), are normally found inside the cells of arthropods, such as lice, fleas, and ticks. They occasionally infect humans and cause serious diseases. The causative agent of Mediterranean spotted fever, also known as boutonneuse fever, is *R. conorii*, an intracellular bacterium of ticks (*Rhipicephalus sanguineus*) (*7*). The complete 1.1-Mb genome of *Rickettsia prowazekii*, a close relative of *R. conorii* transmitted by lice (*Pediculus humanus humanus*), was recently reported (*8*). One salient feature of this genome is the abundance of noncoding sequences, suggesting ongoing genome degradation that parallels the evolutionary scenario presumed to have occurred for mitochondria (*8*, *9*). The sequencing of the entire 1.3-Mb *R. conorii* genome has been undertaken to study the evolution of *Rickettsia* and the molecular basis of different life-styles and pathogenicities (*10*). The palindromic repeat described here was found in 19 open reading frames (ORFs) and in 25 distinct extragenic regions of the genome. The lengths of the 44 repeats identified in *R. conorii* varied from 106 to 150 nucleotides [Web fig. 1 (*11*)]. Because of their well-conserved palindromic nature, most of these sequences are able to form stable hairpin-like mRNA secondary structures (*12*). Hairpins, such as those shown in Fig. 1, are predicted to correspond to the minimum free energy secondary structures for 39 of these repeats.

The amino acid sequences encoded by the 19 repeats found in *R. conorii* ORFs are well conserved and correspond to the same reading frame [Web fig. 2 (*11*)] (*13*). The predicted conformation of the putative peptide chains consists of a central, mildly hydrophobic α helix, flanked by two extended or coil regions including two conserved glycine residues (at positions 14 and 34), a conserved proline (position 40), and numerous conserved charged residues. Insertion sites within the ORFs vary from NH_2-terminal to near COOH-terminal ends (Fig. 2), but secondary-structure prediction and hydropathy analyses suggest that they always occur at the surface of the various proteins. In all cases where the three-dimensional (3D) structures of homologs were available, the insertion sites occurred within NH_2-terminals, loops, or short helices exposed to solvent (*14*).

Of the 19 repeat-containing ORFs in *R. conorii*, 16 have homologs in other species (Table 1) and are thus very likely to encode functional proteins. Most of them are predicted to have important roles in *R. conorii* metabolism. Thirteen have homologs in distant phyla, including four ORFs [guanosine 5′-triphosphate (GTP)–binding protein Era, glutamyl–tRNA synthetase, guanilate kinase, and cell cycle protein MesJ] having homologs within the minimal

[1]Information Génétique & Structurale, CNRS-AVENTIS UMR 1889, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France. [2]Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, 91000 Evry, France. [3]Unité des Rickettsies, Université de la Méditerranée, Faculté de Médecine, CNRS UPRES A 6020, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France.

*To whom correspondence should be addressed. E-mail: jean-michel.claverie@igs.cnrs-mrs.fr

Molecular Analysis of *FRIGIDA*, a Major Determinant of Natural Variation in *Arabidopsis* Flowering Time

Urban Johanson, Joanne West, Clare Lister, Scott Michaels, Richard Amasino and Caroline Dean (October 13, 2000)

Science **290** (5490), 344-347. [doi: 10.1126/science.290.5490.344]

Editor's Summary