EMBO reports

scientific report

Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family

Sophie Desloire[1]*, Hassen Gherbi[1]*, Wassila Laloui[1], Sylvie Marhadour[2], Vanessa Clouet[2], Laurence Cattolico[4], Cyril Falentin[2], Sandra Giancola[3], Michel Renard[2], Françoise Budar[3], Ian Small[1], Michel Caboche[1], Régine Delourme[2] & Abdelhafid Bendahmane[1+]

[1]Unité de Recherches en Génomique Végétale INRA CNRS, Evry, France, [2]UMR INRA-ENSAR, Amélioration des Plantes et Biotechnologies Végétales, Le Rheu, France, [3]Station de Génétique et d'Amélioration des Plantes, INRA, Versailles, France, and [4]Centre National de Séquençage, Evry, France

Ogura cytoplasmic male sterility (CMS) in radish (*Raphanus sativus*) is caused by an aberrant mitochondrial gene, *Orf138*, that prevents the production of functional pollen without affecting female fertility. *Rfo*, a nuclear gene that restores male fertility, alters the expression of *Orf138* at the post-transcriptional level. The Ogura CMS/*Rfo* two-component system is a useful model for investigating nuclear–cytoplasmic interactions, as well as the physiological basis of fertility restoration. Using a combination of positional cloning and microsynteny analysis of *Arabidopsis thaliana* and radish, we genetically and physically delimited the *Rfo* locus to a 15-kb DNA segment. Analysis of this segment shows that *Rfo* is a member of the pentatricopeptide repeat (PPR) family. In *Arabidopsis*, this family contains more than 450 members of unknown function, although most of them are predicted to be targeted to mitochondria and chloroplasts and are thought to have roles in organellar gene expression.

EMBO reports **4**, 588–594 (2003)
doi:10.1038/sj.embor.embor848

INTRODUCTION

Cytoplasmic male sterility (CMS) occurs in higher plants and is a maternally inherited trait that prevents the production of functional pollen, but maintains female fertility. Molecular studies have shown that CMS is determined by mitchondrial genes that are cotranscribed with essential genes in bicistronic messenger RNAs (Schnable & Wise, 1998). This is the case for Ogura CMS in radish (Ogura, 1968), which is controlled by the *orf138* mitochondrial locus. This locus comprises two cotranscribed open reading frames, Orf138 and OrfB. *orf138* is similar to several mitochondrial genes, and is the sterility-inducing gene (Bonhomme *et al.*, 1992). *orfB* encodes subunit eight of the ATP-synthase complex (Gray *et al.*, 1998).

Nuclear genes that restore male fertility in plants showing Ogura CMS occur naturally in wild radish populations. The effect of these restorer genes on the transcription and translation of *orf138* has been investigated. It was shown that the restoration of male fertility correlates with a lower accumulation of Orf138 protein, but has no significant effect on the transcription level of *orf138* (Bellaoui *et al.*, 1997).

CMS systems are widely used for the production of commercial F_1 hybrid plants. However, when the harvested crop is the seed, CMS systems are exploitable only if a nuclear restorer gene is introduced to suppress male-sterility in the hybrid plants. Ogura CMS, and the corresponding nuclear restorer locus, *Rfo*, have been introgressed from radish into rapeseed (Pelletier *et al.*, 1983; Heyn, 1976). However, the introgression of the *Rfo* locus introduced linked deleterious genetic characteristics and led to a loss of rapeseed genetic information (Delourme *et al.*, 1998). Classical breeding methods have been used to improve the lines, but with limited success, due to reduced homologous recombination between radish and rapeseed DNA (Delourme *et al.*, 1998).

In this context, it would be useful to clone the male-fertility restorer locus, *Rfo*, and to use it alone to restore fertility. In addition to simplifying the commercial exploitation of the Ogura CMS/*Rfo* two-component system, detailed analysis of Orf138 and the corresponding restorer locus, *Rfo*, will provide insight into the understanding of nuclear–cytoplasmic interactions, as well as into the physiological basis of fertility restoration.

[1]Unité de Recherches en Génomique Végétale INRA CNRS, 2 Rue Gaston Crémieux, CP5708, 91057 Evry Cedex, France
[2]UMR INRA-ENSAR, Amélioration des Plantes et Biotechnologies Végétales, BP 35327, F35653 Le Rheu Cedex, France
[3]Station de Génétique et d'Amélioration des Plantes, INRA, Route de Saint-Cyr, 78026 Versailles Cedex, France
[4]Centre National de Séquençage, 2 Rue Gaston Crémieux, CP5708, 91057 Evry Cedex, France
*These authors contributed equally to this work
[+]Corresponding author. Tel: +33 1608 74502; Fax: +33 1608 74510;
E-mail: bendahm@evry.inra.fr

Received 28 January 2003; revised 19 March 2003; accepted 8 April 2003
Published online 9 May 2003

 ©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

In this study, we describe our strategy to clone the *Rfo* locus using a positional cloning approach and making use of the microsynteny between *Arabidopsis* and radish. From this analysis, we concluded that the protein encoded by *Rfo* is a member of the pentatricopeptide repeat (PPR) family of proteins.

RESULTS

Rfo-linked markers match *Arabidopsis* chromosome I

The fertility restorer gene, *Rfo*, has been mapped previously in radish, and amplified fragment length polymorphism (AFLP) markers linked to *Rfo* have been identified (Giancola *et al.*, unpublished data). The localization of *Rfo*-linked AFLPs spanning a 3-cM genetic region of the *Arabidopsis* genome using BLAST analysis has identified more than one putative *Arabidopsis* syntenic region. This indicates that, near *Rfo,* the sequence colinearity between radish and *Arabidopsis* is limited to a small interval. On the basis of this analysis, we hypothesized that AFLP markers that are tightly linked to *Rfo* in radish should match a single genomic region in *Arabidopsis.* Thus, we combined our AFLP analysis with bulked segregant analysis. The DNA that constitutes the bulk was derived from the progeny of a cross between a male-sterile European radish line, 7ms, and a radish line that is homozygous for *Rfo*. In this analysis, 800 *Pst*I–*Mse*I primer combinations were analysed for polymorphisms, and 3 new AFLP markers that are tightly linked to *Rfo* (R3, R15 and R5) were identified. R3 and R15 cosegregate with *Rfo* in a population of more than 900 segregant plants, and R5 maps at 0.1 cM from *Rfo* (Fig. 1). The markers R3, R5 and R15 that were identified in this AFLP screen, as well as the markers pVR3, R22 and pVR1, which were identified in a previous screen, were cloned, sequenced and mapped in *Arabidopsis* using BLAST analysis. The markers R3, R15 and pVR3 map within a 40-kb genomic region on *Arabidopsis* chromosome I, and R5, R22 and pVR1 match a region 150 kb away from this (Table 1). From these data, we conclude that radish genomic DNA near the *Rfo* locus is likely to be colinear with *Arabidopsis* DNA around the region contained in bacterial artificial chromosome (BAC) F24D7 (Table 1).

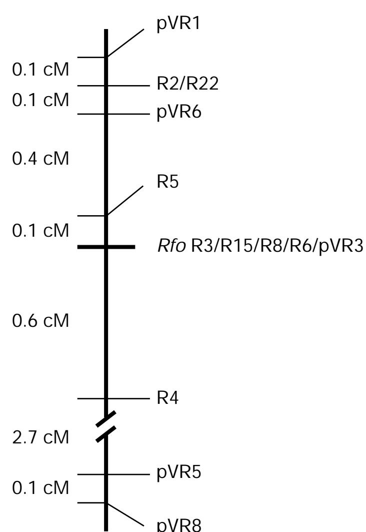

Fig. 1 | Genetic map of the *Rfo* locus in radish. The data are based on the screening of 800 *Pst*I–*Mse*I primer combinations for amplified fragment length polymorphism (AFLP) markers linked to the *Rfo* locus, and 900 segregant plants for recombination events in the pVR1–pVR8 region.

Arabidopsis chromosome I markers map near *Rfo* in radish

To further test the sequence colinearity between *Arabidopsis* chromosome I and the *Rfo* locus in radish, PCR markers were derived from *Arabidopsis* chromosome I in the vicinity of BAC F24D7 (Table 2), and were mapped to *Rfo* in radish. The markers were derived from six *Arabidopsis* BAC clones that spanned a 1100-kb region, and were based on AGI-predicted coding sequences (CDSs). Primers were designed in exons flanking a predicted intron, to increase the chance of identifying polymorphisms between male-fertile and male-sterile radishes. In this analysis, primers were designed for 60 CDSs and PCR was carried out on *Arabidopsis* and on fertile and sterile radishes. All of the primer pairs amplified the CDSs from the corresponding genomic sequence in *Arabidopsis.* However, only 17 primer pairs efficiently amplified the corresponding CDSs in radish. Out of these 17 amplified CDSs (Table 2), 13 were polymorphic between fertile and sterile radishes, and these were mapped relative to *Rfo* in a population of 900 F_2 plants. The markers M-T12P18.15, M-F24D7.4, M-F24D7.7, M-F24D7.9, M-F24D7.13, M-F24D7.14, M-F24D7.16, M-F24D7.17 and

Table 1 | The physical position on the *Arabidopsis* genome of the amplified fragment length polymorphism markers tightly linked to *Rfo*, which were obtained using BLAST analyses

AFLP markers	*A. thaliana* chromosome	BAC name and gene number	Predicted gene function*	*E*-value†
R3	I	F24D7.4	Putative aminopeptidase	8.7×10^{-9}
R15	I	F24D7.7	Disease-resistance protein	4×10^{-7}
R8	I	F10F5.13	Hypothetical protein	6.4×10^{-7}
R22	I	F1N19.28	Unknown protein	1×10^{-4}
R2	I	F25P12.101	Disease-resistance protein	0.003
pVR1	I	F1N19.5	Unknown protein	2.10×10^{-26}
R5	I	F1N19.20	Unknown protein	0.034
pVR3	I	T12P18.15	Unknown protein	2.10×10^{-38}
R4	–	–	No significant match	–
R6	III	21B14	No match with gene prediction	1×10^{-18}
pVR5	I	F20D22.11	ABC transporter family protein	1×10^{-17}
pVR8	II	F24L7.13	26S proteasome regulatory subunit (RPN2; putative)	2×10^{-89}

*Predicted using the *Arabidopsis* Genome Initiative database (http://arabidopsis.org). †Obtained by performing a BLAST search. AFLP, amplified fragment length polymorphism; *A. thaliana*, *Arabidopsis thaliana*; BAC, bacterial artificial chromosome.

©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

Table 2 | Markers derived from *Arabidopsis thaliana* chromosome I map in the vicinity of *Rfo* in radish

BAC	Flagged genes*	Marker
F16M19	Putative scarecrow transcription factor	*M-F16M19.21
F2K11	Kinesin-like protein	*M-F2K11.1
	Putative receptor protein kinase	*M-F2K11.19
F24D7	Aminopeptidase (putative)	*M-F24D7.4
	Disease-resistance gene	*M-F24D7.7
	Hypothetical protein	*M-F24D7.9
	UDP-*N*-acetylmuramoylanalyl-D-glutamate-2-6-diaminoligase	*M-F24D7.13
	Hypothetical protein	*M-F24D7.14
	Putative transcription factor	*M-F24D7.16
	Kinesin-like protein	*M-F24D7.17
T12P18	Putative monodehydroascorbate reductase	M-T12P18.4
	Ring zinc-finger proteins	*M-T12P18.9
	Unknown protein	*M-T12P18.15
F22C12	Hypothetical protein	*M-F22C12.1
	Putative H^+-transporting ATPase protein	M-F22C12.12
F15H21	Putative acyl-CoA synthetase	M-F15H21.7
	Putative endo-β-1,4-glucanase	M-F15H21.9

*Genes that were used as molecular markers. Asterisks indicate markers that are polymorphic between fertile and sterile radishes. BAC, bacterial artificial chromosome.

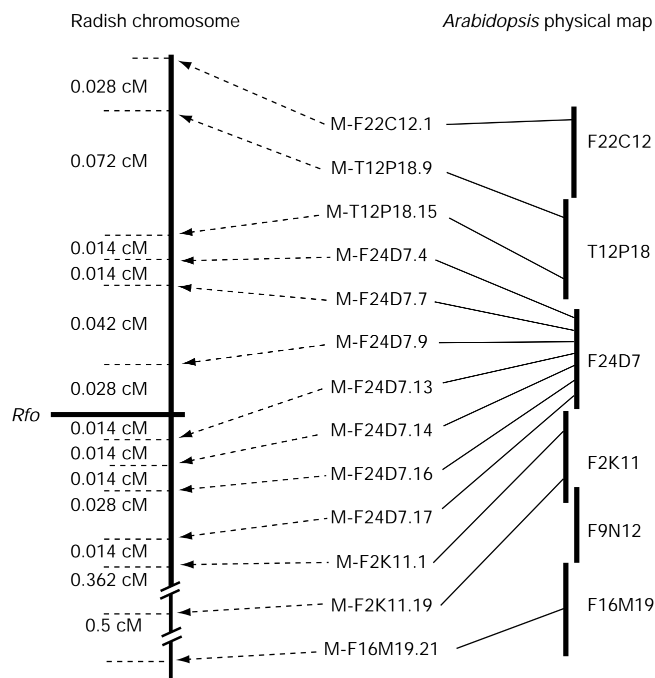

Fig. 2 | Analysis of microsynteny around the *Rfo* locus between radish and *Arabidopsis*. The genetic map of *Rfo* is based on the analysis of 6,907 segregant plants. The *Arabidopsis* physical map was deduced from the AGI database (http://arabidopsis.org). On the right, the physical localization of the markers in the *Arabidopsis* BAC (bacterial artificial chromosome) contig are shown, and the arrows localize them on the radish genetic map, which is shown on the left.

M-F2K11.1 cosegregated genetically with *Rfo*. On one side of *Rfo*, the markers M-F16M19.21 and M-F2K11.19 were mapped at 1 cM and 0.6 cM from *Rfo*, respectively. On the other side, M-F22C12.1 and M-T12P18.9 were mapped at 0.2 cM and 0.1 cM from *Rfo*, respectively. From these analyses, we conclude that *Rfo* is genetically delimited to a region of 0.7 cM between the markers M-F2K11.19 and M-T12P18.9.

A high-resolution genetic map of the radish *Rfo* locus

To more accurately position the nine markers that cosegregate with *Rfo*, it was necessary to identify members of the *Rfo* mapping population that had recombination events close to the *Rfo* locus. This process involved genotyping DNA samples from 6,907 plants in the mapping populations with the *Rfo*- flanking markers M-F2K11.19 and M-T12P18.9. From this screen, 43 plants carrying recombination events in the M-F2K11.19–M-T12P18.9 region were identified. These plants were used to map the nine markers relative to each other and to *Rfo*. In this analysis, *Rfo* was mapped to a genetic region of 0.042 cM, which is delimited by the markers M-F24D7.13 and M-F24D7.9 (Fig. 2; Table 3). In *Arabidopsis*, this region contains three genes that encode an unknown protein (F24D7.12), a putative protein kinase (F24D7.11) and a putative cytochrome P450 (F24D7.10). Thus, we might predict that *Rfo* is orthologous to one of these genes. Alternatively, the *Rfo* locus might be absent from *Arabidopsis*, and the microsynteny between radish and *Arabidopsis* in the M-F24D7.13–M-F24D7.9 region might not be conserved. The construction of a BAC library from radishes homozygous for the *Rfo* locus and the identification of radish DNA fragments carrying both of the *Rfo* flanking markers M-F24D7.13 and M-F24D7.9 was the only reliable way to clone *Rfo*.

A high-resolution physical map of the radish *Rfo* locus

To ascertain if there is micro-colinearity between radish and *Arabidopsis* within the M-F2K11.19–M-T12P18.9 region, a BAC library was constructed from nuclear DNA derived from the line D81, which is homozygous for *Rfo*. The library consists of 120,000 clones and represents the haploid radish genome at least 23 times over. Using a systematic PCR-based procedure, the library was screened with markers tightly linked to *Rfo* (Fig. 2). Positive BAC clones were isolated and aligned in a single predicted contig. The order of the BAC clones in the contig was consistent with the genetic distribution of the markers relative to *Rfo*. The *Rfo* locus was physically delimited to a single BAC clone, BAC64, which is positive for both of the *Rfo* flanking markers, M-F24D7.13 and M-F24D7.9 (Fig. 3).

Sequence analysis of BAC64

BAC64 was sequenced using a shotgun sequencing procedure and the sequence was assembled into a single-sequence contig of 127 kb. The redundancy of the sequence coverage was at least ten times the BAC sequence-length. Two other tests were carried out to check the quality of the sequence consensus. First, the predicted restriction map was compared to the fingerprint of BAC64, and was found to be consistent. Second, the sequences of the markers linked to *Rfo* were aligned with the BAC64 sequence, and the genetic order of the markers was shown to match the physical order in the sequence.

©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

Table 3 | Genotypes of recombinants used for the genetic mapping of the *Rfo* locus

	Recombinants														
	RcR29	RcE1	RcR2	RcE12	RcE14	RcR24	RcR25	RcE9	RcR18	1431	RcR30	RcE4	RcE8	RcR16	RcE3
	Phenotype of plants														
Molecular markers	S	S	S	F	F	F	F	F	F	F	F	F	F	S	S
M-T12P18.9	S	S	S	S	S	S	S	S	S	F	F	F	F	F	F
M-T12P18.15	S	S	S	S	S	S	S	S	F	F	F	F	F	F	F
M-F24D7.4	S	S	S	S	S	S	S	F	F	F	F	F	F	F	F
M-F24D7.7	S	S	S	S	S	S	S	F	F	F	F	F	F	F	S
M-F24D7.9	S	S	S	S	S	F	F	F	F	F	F	F	F	S	S
PPR-C	S	S	S	S	F	F	F	F	F	F	F	F	F	S	S
PPR-B	S	S	S	F	F	F	F	F	F	F	F	F	F	S	S
PPR-A	S	S	S	F	F	F	F	F	F	F	F	F	F	S	S
M-F24D.13	S	S	S	F	F	F	F	F	F	F	F	F	S	S	S
M-F24D7.14	S	S	F	F	F	F	F	F	F	F	F	F	S	S	S
M-F24D.16	S	S	F	F	F	F	F	F	F	F	F	S	S	S	S
M-F24D7.17	S	S	F	F	F	F	F	F	F	S	S	S	S	S	S
M-F2K11.1	S	F	F	F	F	F	F	F	F	S	S	S	S	S	S
M-F2K11.19	F	F	F	F	F	F	F	F	F	S	S	S	S	S	S

Only plants with recombination events tightly linked to *Rfo* are shown. F, fertile; S, sterile.

In the genetic analysis, *Rfo* was delimited to a genetic interval of 0.042 cM between the markers F24D7.9 and F24D7.13. These two markers physically delimit *Rfo* to a 22-kb region. The M-F24D7.9–M-F24D7.13 region encodes three predicted proteins, PPR-A, PPR-B and PPR-C, that belong to the PPR family of proteins (Small & Peeters, 2000) (Fig. 4). We further characterized the three plants, RcE8, RcE14 and RcE12, that carry recombination events in the 22-kb M-F24D7.9–M-F24D7.13 physical region (Table 3). Markers were derived from *Ppr-A*, *Ppr-B* and *Ppr-C*, and were mapped relative to *Rfo* in the RcE8, RcE14 and RcE12 recombinant plants. The RcE8 plant carries a recombination event between the marker M-F24D7.13 and *Ppr-A*. In this plant, *Rfo* cosegregates genetically with *Ppr-A*, *Ppr-B* and *Ppr-C*. Thus, this recombination event eliminates the gene encoding UDP-*N*-acetylmuramoylanalyl-D-glutamate-2,6-diaminoligase (corresponding to the marker M-F24D7.13) as a candidate for *Rfo*. The RcE14 plant carries a recombination event between *Ppr-C* and M-F24D7.9. In this plant, *Rfo* cosegregates genetically with *Ppr-C*, *Ppr-B* and *Ppr-A*. Thus, this recombination event eliminates the gene F24D7.9, which is of unknown function, as a candidate for *Rfo*. The RcE12 plant carries a recombination event between *Ppr-B* and *Ppr-C*. In this plant, *Rfo* cosegregates genetically with *Ppr-B* and *Ppr-A*. Thus, this recombination event eliminates *Ppr-C* as a candidate for *Rfo*. This statement is reinforced by the sequence analysis of *Ppr-C*, which suggests that *Ppr-C* is a pseudogene (Fig. 5). In conclusion, *Rfo* is likely to correspond to *Ppr-A* or *Ppr-B* or both.

DISCUSSION

In the Brassicaceae, genome synteny analysis between *A. thaliana* and other Brassica species has shown that large regions might be colinear, whereas, in some cases, only small sequence islands showed colinearity (reviewed in Schmidt, 2002). One might expect that exploitation of information from the *Arabidopsis* genome will facilitate the cloning and characterization of economically important genes in crop plants, especially in the Brassicaceae family. In this study, we have used the microsynteny between *Arabidopsis* and radish, combined with a positional cloning approach, to clone the *Rfo* locus. It is, to our knowledge, the first study in which exploitation of microcolinearity and map-based cloning were used together for gene cloning. Moreover, it is the first precise synteny study at the physical level between radish and *Arabidopsis* genomic regions. The use of the syntenic region from *Arabidopsis* facilitated the development of closely linked markers for the analysis of the *Rfo* region, and even if the *Rfo* locus itself has no counterpart at the equivalent site in *Arabidopsis*, the comparison accelerated the cloning of *Rfo*.

The genetic and physical maps around the locus formally delimit *Rfo* as one of two highly similar PPR proteins, PPR-A and PPR-B. PPR-C, which was excluded genetically, seems to be a pseudogene, as it contains a 17-bp deletion with respect to the other two genes, that leads to a frameshift and a premature stop codon. PPR-A and PPR-C contain a 12-nucleotide deletion in the third PPR repeat that reduces the similarity of this repeat to the canonical PPR structure, and may prevent these proteins from functioning (Fig. 5). PPR-B is probably the best candidate for *Rfo*.

The PPR gene family is a large family in plants, consisting of more than 450 genes in *Arabidopsis* (Aubourg *et al.*, 2000; Small & Peeters, 2000). The functions of these genes are mainly unknown, although most of them are predicted to be targeted to mitochondria and chloroplasts and may have roles in organellar gene expression (Small & Peeters 2000; Barkan & Goldschmidt-Clermont, 2000). There is some evidence that members of this family are sequence-specific RNA-binding proteins (Lahmy *et al.*, 2000; Mancebo *et al.*, 2001). The few PPR mutants that have been described have defects that are limited to a failure to express specific organellar transcripts (Manthey & McEwen, 1995; Fisk *et al.*, 1999). As most CMS restorer genes seem to function by preventing the expression of mitochondrial CMS-inducing transcripts or proteins (Budar & Pelletier, 2001), PPR

©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

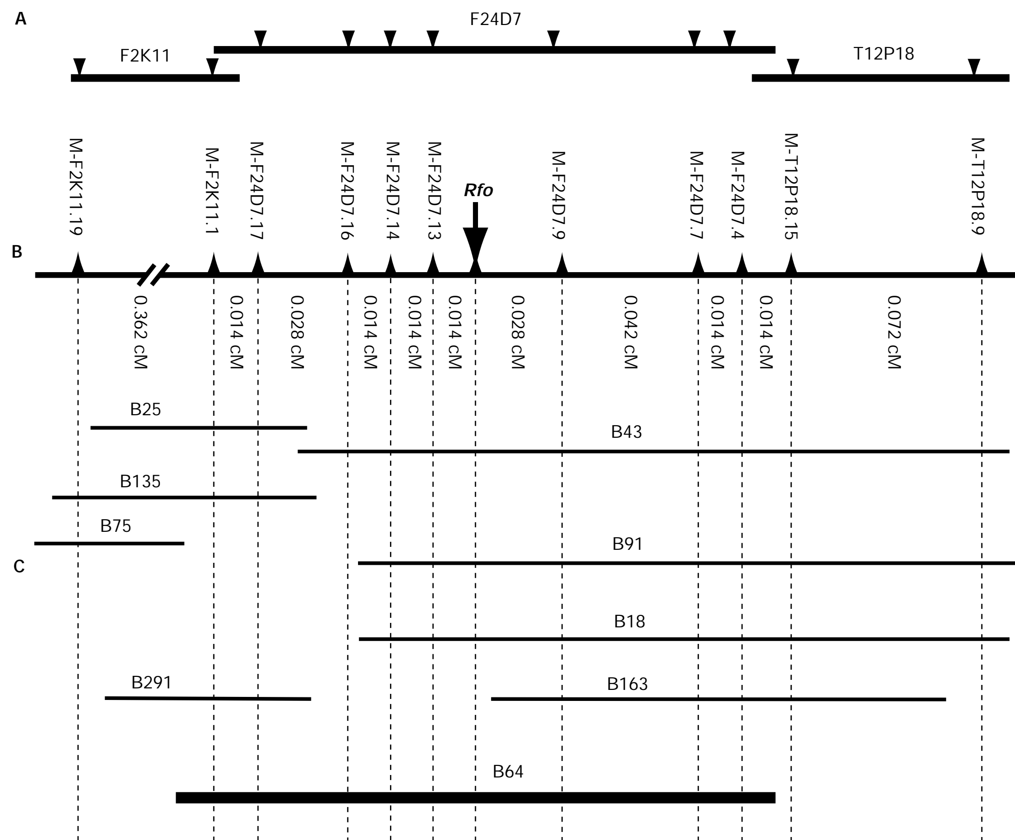

Fig. 3 | High-resolution genetic and physical maps of the *Rfo* locus. (**A**) *Arabidopsis thaliana* BAC (bacterial artificial chromosome) contigs that are syntenic to the *Rfo* locus in radish. The positions of the markers that are shown in (**B**) are indicated by black triangles. (**B**) High-resolution genetic map of the *Rfo* locus in radish. The black arrow indicates the *Rfo* locus. (**C**) Radish BAC contigs that span the *Rfo* locus. Broken lines indicate the positions of the markers from (**B**) in the radish BAC clones. The *Rfo* locus is physically delimited to BAC64.

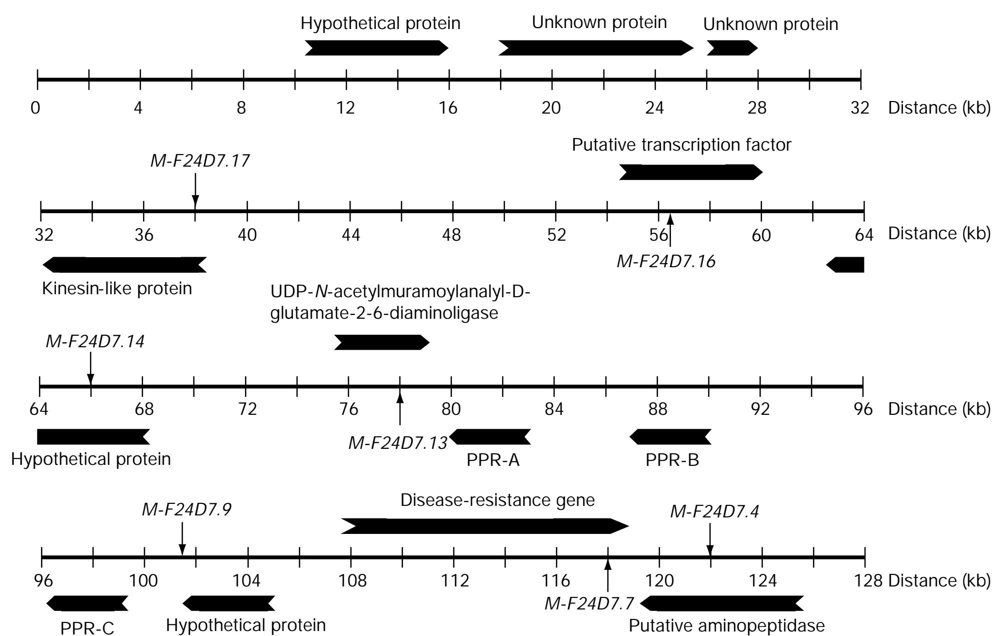

Fig. 4 | Representation of the annotated BAC64 sequence. The prediction of genes and the identification of their functions were carried out using GENSCAN software and BLAST analysis against predicted *Arabidopsis* proteins, respectively. Horizontal arrows indicate the positions and orientations of the predicted genes. Vertical arrows indicate the positions of the markers that are tightly linked to *Rfo*. BAC, bacterial artificial chromosome.

©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

```
PPR-A     1   MLARVCRFESSSSSSVSAARFFCTGSIRHALAEKSRDGESGEAGFRGESL
PPR-B     1   MLARVCGFKCSSSPAESAARLFCTRSIRDTLAKAS--GESCEAGFGGESL
PPR-C     1   MLARVYRSGSSSSPAVSAARLFCTRSIRHALAKKS---RDGESGFGGESL
              *****    .***.: ****:*** ***.:**: *   .. *:** ****

PPR-A     51  KLRSGSYEIKGLEDAIDLFSDMLRSRPLP
PPR-B     49  KLQSGFHEIKGLEDAIDLFSDMLRSRPLP
PPR-C     48  KLRSGFHEIKGLEDAIDLFGDMVRSRPLP
              **:** :************.**:******

PPR-A     80  SVIDFNKLMGAVVRMERPDLVISLYQKMERKQIRC
PPR-B     78  SVVDFCKLMGVVVRMERPDLVISLYQKMERKQIRC
PPR-C     77  SVIDFCKLMGVVVRMGRLDVVISLHRKMEMRRVPC
              **:** ****.**** * *:****::*** ::: *

PPR-A     115 DIYSFTILIKCFCSCSKLPFALSTFGKLTKLGLHP
PPR-B     113 DIYSFNILIKCFCSCSKLPFALSTFGKITKLGLHP
PPR-C     112 NAYSFTILMKCFCSCSKLPFALSTFGKITKLGFHP
              : ***.**:******************:****:**

PPR-A     150 DVVTFTTLLHGLCLDHRVSEALDLFHQI----CRP
PPR-B     148 DVVTFTTLLHGLCVEDRVSEALDFFHQMFETTCRP
PPR-C     147 TVVTFSTLLHGLCVEDRISEALDLFHQM----CKP
               ****:*******::.*:*****:***:      *

PPR-A     181 DVLTFTTLMNGLCREGRVVEAVALLDRMVENGLQP
PPR-B     183 NVVTFTTLMNGLCREGRIVEAVALLDRMMEDGLQP
PPR-C     178 NVVTFTTLMNGLCREGRVVEAVALLDRMLEDGLQP
              :*:**************:**********:*:****

PPR-A     216 DQITYGTFVDGMCKMGDTVSALNLLRKMEEISHIKP
PPR-B     218 TQITYGTIVDGMCKKGDTVSALNLLRKMEEVSHIIP
PPR-C     213 NQITYGTIVDGMCKMGDTVSALNLLRKMEEVSHIKP
              ******:****** ***************:*** *

PPR-A     252 NVVIYSAIIDGLCKDGRHSDSHNLFIEMQDKGIFP
PPR-B     254 NVVIYSAIIDSLCKDGRHSDAQNLFTEMQEKGIFP
PPR-C     249 NVVIWPLERR-------------------------
              ****:.

PPR-A     287 NIVTYNCMIGGFCISGRWSAAQRLLQEMLER-KISP
PPR-B     289 DLFTYNSMIVGFCSSGRWSDAEQLLQEMLER-KISP
PPR-C     259 -----TCMINGFCSSGRWSEAQQLLQEMLERKKISP
                   ..** *** ***** *::******** ****

PPR-A     322 NVVTYNALINAFVKEGKFFEAAELYDEMLPRGIIP
PPR-B     324 DVVTYNALINAFVKEGKFFEAEELYDEMLPRGIIP
PPR-C     290 DVVTYNALINAFVKEGKFFEAEELYDEMLPRGIIP
              :******************** *************

PPR-A     357 NTITYNSMIDGFCKQDRLDAAEDMFYLMATKGCSP
PPR-B     359 NTITYSSMIDGFCKQNRLDAAEHMFYLMATKGCSP
PPR-C     325 STITYSSMIDGFCKQNRLDAAEHMFYLMATKGCSP
              .****.*********:******.************

PPR-A     392 DVFTFTTLIDGYCGAKRIDDGMELLHEMPRRGLVA
PPR-B     394 NLITFNTLIDGYCGAKRIDDGMELLHEMTETGLVA
PPR-C     360 DIITFNTLIAGYCRAKRVDDGIKLLHEMTEAGLVA
              :::**.*** *** ***:***::*****.. ****

PPR-A     427 NTVTYNTLIHGFCLVGDLNAALDLSQQMISSGVCP
PPR-B     429 DTTTYNTLIHGFYLVGDLNAALDLLQEMISSGLCP
PPR-C     395 NTITYTTLIHGFCQVGDLNAAQDLLQEMVSSGVCP
              :* **.******  ******* ** *:*:***:**

PPR-A     462 DIVTCNTLLDGLCDNGKLKDALEMFKAMQKSKMDLDASHPFNGVEP
PPR-B     464 DIVTCDTLLDGLCDNGKLKDALEMFKVMQKSKKDLDASHPFNGVEP
PPR-C     430 NVVTCNTLLDGLCDNGKLKDALEMFKAMQKSKMDIDASHPFNGVEP
              ::***:********************.***** *:***********

PPR-A     508 DVLTYNILICGLINEGKFLEAEELYEEMPHRGIVP
PPR-B     510 DVQTYNILISGLINEGKFLEAEELYEEMPHRGIVP
PPR-C     476 DVQTYNILISGLINEGKFLEAEELYEEMPHRGIVP
              ** ******.*************************

PPR-A     543 DTITYSSMIDGLCKQSRLDEATQMFVSMGSKSFSP
PPR-B     545 DTITYSSMIDGLCKQSRLDEATQMFDSMGSKSFSP
PPR-C     511 DTITYNSVIHGLCKQSRLDEATQMFDSMGSKSFSP
              *****.*:*.*************** *********

PPR-A     578 NVVTFNTLINGYCKAGRVDDGLELFCEMGRRGIVA
PPR-B     580 NVVTFTTLINGYCKAGRVDDGLELFCEMGRRGIVA
PPR-C     546 NVVTFTTLINGYCKAGRVDDGLELFCEMGRRGIVA
              *****.*****************************

PPR-A     613 DAIIYITLIYGFRKVGNINGALDIFQEMISSGVYP
PPR-B     615 NAITYITLICGFRKVGNINGALDIFQEMISSGVYP
PPR-C     581 NAITYITLIHGFRKVGNINGALDIFQEMMASGVYP
              :** ***** ******************::*****

PPR-A     648 DTITIRNMLTGFWSKEELERAVAMLEDLQMSVGMS
PPR-B     650 DTITIRNMLTGLWSKEELKRAVAMLEKLQMSMDLS
PPR-C     616 DTITIRNMLTGLWSKEELKRAVAMLEDLQMSVGYQ
              ***********:******:*******.****:. .

PPR-A     683 FNTFCFQISLLTFIILEKSCSLCCSIRETFLNE
PPR-B     685 FGG------------------------------
PPR-C     651 LEDE-----------------------------
              :

PPR-A     716 WFGVFVLQGISWRMNERMKDTFLFQIYKALLILFCR
PPR-B         ------------------------------------
PPR-C         ------------------------------------
```

Fig. 5 | Alignment of predicted protein sequences of PPR-A, PPR-B and PPR-C. The sequences of the 17 PPR repeats (amino acids 78–684 of PPR-B) are shown. The four amino-acid deletions in the third PPR repeats of both PPR-A and PPR-C are indicated by dashes. The gene encoding PPR-C contains a 17-bp deletion in repeat 6 that leads to a frameshift and a premature stop codon. The PPR-C sequence shown here, with a 30-amino-acid deletion in PPR repeats six and seven, is the hypothetical sequence obtained if this frameshift is spliced out in an intron predicted by GENSCAN. Asterisks indicate amino-acid identity; colons indicate a high level of amino-acid similarity; dots indicate low levels of amino-acid similarity. PPR, pentatricopeptide repeat.

proteins are logical candidates for the products of nuclear restorer genes. An important breakthrough in the understanding of CMS restoration was made last year when the *Rf* restorer locus in *Petunia hybrida* was shown to encode a mitochondrial PPR protein (Bentolila *et al.*, 2002), which was the first demonstration of the involvement of a member of this family in fertility restoration. With the subsequent demonstration that the Ogura restorer locus also encodes PPR proteins, it became even more tempting to speculate that many restorer genes, in a wide range of plant species, are also PPR genes. The only exception described so far relates to Texas maize (*Zea mays*) CMS, in which the nuclear restorer gene *Rf2* encodes an aldehyde dehydrogenase (Cui *et al.*, 1996). *Rf2* has no effect on the RNA or protein levels of the CMS protein URF13. The authors proposed that RF2 protein compensates for a metabolic defect caused by the CMS protein.

A surprising observation is that both petunia *Rf* and radish *Rfo* (whichever of the 2 PPR genes is taken to be *Rfo* in radish) are most similar to the same group of 20 or so *Arabidopsis* PPR proteins, out of the more than 450 possible, when compared using BLASTP searches (data not shown). A distance tree (Fig. 6) shows the relationship of the RFO proteins (and, to a lesser extent, petunia RF) to this group of *Arabidopsis* PPR proteins. However, it is impossible to identify obvious single *Arabidopsis* orthologues of the CMS restorer proteins, whereas putative orthologues of two other characterized PPR proteins (maize CRP1 and radish p67) are easy to identify by sequence similarity. The *Arabidopsis Rfo*-like genes closely resemble each other, and are predominantly arranged in two loose clusters on chromosome I, one of which is close to the zone that is syntenic to the *Rfo* locus. It is intriguing to speculate about the function of these genes in a highly autogamous species where, to our knowledge, CMS has never been described.

METHODS

Mapping of *Rfo*. The *Rfo* segregating population was obtained by selfing F_1 hybrids that carried Ogura CMS, which were derived from the cross between a European male-sterile radish line, 7ms, and the radish line D81.8, which is homozygous for *Rfo*. AFLP analysis was performed on bulked DNA samples of male-fertile and male-sterile plants, as described previously (Bendahmane *et al.*, 1997). To identify plants carrying recombination events linked to *Rfo* in radish, DNA samples were extracted from 6,907 F_2 plants and analysed using the *Rfo*-flanking markers M-T12P18.9 and M-F2K11.19. Map distances are given in centimorgans, and represent the percentage of recombinant plants in the total number of plants analysed.

©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

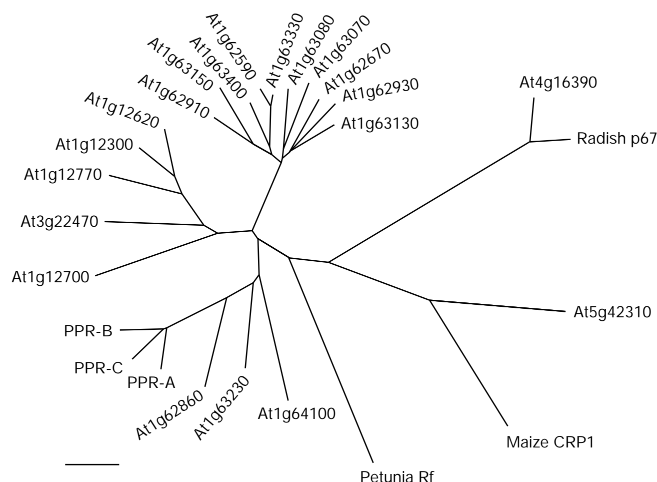

Fig. 6 | Relationships between PPR-A, PPR-B, PPR-C, petunia RF and several PPR homologues from *Arabidopsis*, radish and maize. CRP1 (chloroplast processing 1) is a maize protein that is involved in plastid messenger RNA processing (Fisk *et al.*, 1999). p67 is a radish chloroplast protein of unknown function (Lahmy *et al.*, 2000). At5g42310 and At4g16390 are the closest *Arabidopsis* homologues of these two proteins. The distance tree was produced using ClustalW to align the sequences and using a neighbour-joining algorithm to group them. The *Arabidopsis* homologues were identified by BLASTP searches against predicted *Arabidopsis* proteins. The lengths of the lines connecting the proteins indicate the mean number of estimated substitutions per site (corrected for multiple substitutions). Scale bar, 0.1 substitutions per site. PPR, pentatricopeptide repeat.

Construction and screening of a radish bacterial artificial chromosome library. The BAC library was prepared from nuclei extracted from radish line D81.8, which is homozygous for *Rfo*, based on the method described in Peterson *et al.* (2000). The library consists of 120,000 BAC clones. To assess the insert size of the BAC clones, plasmid DNA was isolated from more than 100 randomly chosen clones, digested with *Not*I and analysed by pulsed-field gel electrophoresis, as described previously (Kanyuka *et al.*, 1999). The insert-size distribution of the library is as follows: 13.2% of the colonies have an insert size in the range of 150–200 kb, 51.5% in the range of 100–150kb, 33.8% in the range of 50–100 kb, and 1.5% have inserts of less than 50 kb. The screening of the BAC library was performed as described previously (Kanyuka *et al.*, 1999).

DNA sequencing and analysis. A shotgun cloning strategy was used for sequencing BAC64. Sequence contigs were assembled using UNIX versions of the Staden programmes package (Staden *et al.*, 1998). Gene prediction was performed using GENSCAN software (Burge & Karlin, 1997). BLAST analysis was used for the prediction of gene function and for mapping radish AFLP sequences in the *Arabidopsis* genome. Multiple sequence alignment was carried out using ClustalW software (Thompson *et al.*, 1994). The genomic DNA sequence containing *Rfo* is deposited in the EMBL nucleotide sequence database under the accession number AJ550021.

ACKNOWLEDGEMENTS
This work was supported by Génoplante, the French consortium for plant genomics. The work was carried out in compliance with the current laws governing research and development programmes in France.

REFERENCES
Aubourg, S., Boudet, N., Kreis, M. & Lecharny, A. (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol. Biol.*, **42**, 603–613.
Barkan, A. & Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in chloroplast gene expression. *Biochimie*, **82**, 559–572.
Bellaoui, M., Pelletier, G. & Budar, F. (1997) The steady-state level of mRNA from the Ogura cytoplasmic male sterility locus in *Brassica* cybrids is determined post-transcriptionally by its 3′ region. *EMBO J.*, **16**, 5057–5068.
Bendahmane, A. Kanyuka, K. & Baulcombe, D.C. (1997) High-resolution genetical and physical mapping of the *Rx* gene for extreme resistance to potato virus X in tetraploid potato. *Theor. Appl. Genet.*, **95**, 153–162.
Bentolila, S., Alfonso, A.A. & Hanson, M.R. (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl Acad. Sci. USA*, **99**, 10887–10892.
Bonhomme, S., Budar, F., Lancelin, D., Small, I., Defrance, M.C. & Pelletier, G. (1992) Sequence and transcript analysis of the *Nco2.5* Ogura specific fragment correlated with cytoplasmic male-sterility in *Brassica* cybrids. *Mol. Gen. Genet.*, **235**, 340–348.
Budar, F. & Pelletier, G. (2001) Male sterility in plants: occurrence, determinism, significance and use. *C.R. Acad. Sci. III.*, **324**, 543–550.
Burge, C. & Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, **268**, 78–94.
Cui, X., Wise, R.P. & Schnable, P.S. (1996). The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science*, **272**, 1334–1336.
Delourme, R. *et al.* (1998) Characterisation of the radish introgression carrying the *Rfo* restorer gene for *Ogu*-INRA cytoplasmic male sterility in rapeseed (*Brassica napus* L.). *Theor. App. Genet.*, **97**, 129–134.
Fisk, D.G., Walker, M.B. & Barkan, A. (1999) Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.*, **18**, 2621–2630.
Gray, M.W. *et al.* (1998) Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.*, **26**, 865–878.
Heyn, F.W. (1976) Transfer of restorer genes from *Raphanus* to cytoplasmic male sterile *Brassica napus*. *Cruciferae News*, **1**, 15–16.
Kanyuka, K., Bendahmane, A., Rouppe van der Voort, J., Van der Vossen, E. & Baulcombe, D.C. (1999) A high resolution molecular map around the *Rx* locus of potato: analysis of a complex locus in a tetraploid background. *Theor. Appl. Genet.*, **98**, 679–689.
Lahmy, S., Barneche, F., Derancourt, J., Filipowicz, W., Delseny, M. & Echeverria, M. (2000) A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett.*, **480**, 255–260.
Mancebo, R., Zhou, X., Shillinglaw, W., Henzel, W. & Macdonald, P.M. (2001) BSF binds specifically to the bicoid mRNA 3′ untranslated region and contributes to stabilization of bicoid mRNA. *Mol. Cell. Biol.*, **21**, 3462–3471.
Manthey, G.M. & McEwen, J.E. (1995) The product of the nuclear gene *PET309* is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of *Saccharomyces cerevisiae*. *EMBO J.*, **14**, 4031–4043.
Ogura, H. (1968) Studies on the new male-sterility in Japanese radish with special reference to the utilisation of this sterility towards the practical raising of hybrid seeds. *Mem. Fac. Agric. Kagoshima Univ.*, **6**, 39–78.
Pelletier, G. *et al.* (1983) Intergeneric cytoplasmic hybridization in *Cruciferae* by protoplast fusion. *Mol. Gen. Genet.*, **191**, 244–250.
Peterson, D., Tomkins, J.P., Frisch, D.A., Wing, R.A. & Paterson, A.P. (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. *J. Agric. Gen.*, **5**, 1–100.
Schmidt, R. (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. *Plant Mol. Biol.*, **48**, 21–37.
Schnable, P.S. & Wise, R.P. (1998) The molecular basis of cytoplasmic male sterility and fertility restoration *Trends Plant Sci.*, **3**, 175–180.
Small, I.D. & Peeters, N. (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.*, **25**, 46–47.
Staden, R., Beal, K.F. & Bonfield, J.K. (1998) The Staden Package. *Methods Mol. Biol.*, **132**, 115–130.
Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION