

Evolutionary dynamics of a locus of fertility restoration in plants

José R. Hernández Mora^{*1}, Eric Rivals^{*2}, Hakim Mireau¹, Françoise Budar¹

¹ Station de Génétique et d'Amélioration des Plantes - IJPB
INRA, route de Saint Cyr, 78026 Versailles, France
budar@versailles.inra.fr

² Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier (LIRMM)
Univ. Montpellier II - CNRS
161 rue Ada, 34392 Montpellier, Cedex 5, France
{rivals}@lirmm.fr

Abstract: *In higher plants, hermaphrodites may genetically lose their male fertility through the cytoplasmic male sterility (CMS) system. In radish, a nuclear locus, denoted Rfo, has evolved that is able to counteract the effect of CMS and restore the fertility. This locus encodes three similar genes in tandem that belong to the pentatricopeptide repeat (PPR) family and each gene encloses a tandem repeat of PPR motif. Among the hundreds of members of this family, some play a role in the post-transcriptional gene regulation in organelles (mitochondria and chloroplasts). In this study, we recently sequenced a European non-restorer allelic locus and compare it to the original Rfo restorer allele to investigate its evolutionary dynamics. We conducted bioinformatic analysis to determine the putative border of tandem duplications both at protein motif level and at gene level. Our results present the picture of complex evolution with multiple gene duplications at a fast evolving locus.*

Keywords: PPR, duplication, tandem repeat, plant genomic, bioinformatic analysis, Rfo locus, radish

1 Introduction

In higher plants, cytoplasmic genes can sometimes prevent the formation of pollen, thus transforming hermaphrodites into females (*i.e.*, sterile males). The presence of nuclear genes named *Restorers of fertility* (Rf) allow the production of pollen in the presence of a male sterility-inducing cytoplasm. This control of the sexual phenotype in plants is known as nuclear-*cytoplasmic male sterility* (CMS). CMS systems are constituted of two actors: a mitochondrial protein that causes male sterility, and a nuclear gene that encodes a mitochondria-targeted protein able to impair the expression of the sterility gene. All Rf genes identified so far (with the only exception of the Rf2 Texas maize restorer) belong to the pentatricopeptide repeat (PPR) family [1,2].

The Rfo locus, first described in Asian radish cultivars and later introduced into Brassica, is involved in the restoration of fertility in the Ogura CMS system. The nuclear restorer gene in the *Rfo-Ogura* system also encodes a protein belonging to the pentatricopeptide repeat (PPR) family [1,2]. The PPR gene family is a very large family in plants, with about 450 members in *A. thaliana*. Their function is largely unknown, although some members proved to be involved in post-transcriptional organelle gene regulation [3].

* J.R.H and E.R. equally contributed to this work

Recent studies analysing chromosomes regions with duplicated PPR genes in rice and in *A. thaliana* suggested high levels of recombination in these regions. In addition, restorer genes, because of their ability to adapt to specificity changing and fast evolving targets, evolve similarly to resistance genes. Nevertheless, the evolutionary dynamics of these loci remains largely unknown.

In the Rfo locus, the restorer gene seems to have been duplicated: three genes encoding highly related PPR proteins are present; however just one of the proteins is able to restorer fertility. We recently sequenced a European maintainer allele of the Rfo locus. In this study, we analysed the sequences of the restorer and maintainer alleles to mine the possible evolutionary events that occurred in each allele, and investigate the evolutionary dynamics of the Rfo locus.

2 Gene prediction and internal structure

We performed gene prediction and similarity search on the DNA sequences of both variants of the Rfo locus. On the restorer variant, three PPR genes are located in tandem (one next to the other), while on the maintainer variant two PPR genes are separated by a gene similar to MOS2. Hence, the orthology relationships are non trivial between the two variants.

The internal structure in term of PPR motifs of each gene have been predicted with specific Hidden Markov Models [5]. Essentially, the structure is similar for all genes, at the exception of the loss of single PPR motif. Alignment at the motif level and phylogenetic analysis show that a group of five adjacent PPR motif has been duplicated in block. The same signal appears in all genes of the restorer locus, showing that the common gene internal structure has evolved before gene duplication occurred.

3 Duplications in the maintainer locus

In our last step, we mine the signals of both sequence similarity between alleles and of duplications inside the maintainer allele to predict the orthology or paralogy relationships between the PPR genes of both alleles. The regions of internal sequence similarity, determined with YASS [4], are multiple and define boundaries of putative deletions and tandem duplications. The agreement of some boundaries lead us to propose a scenario comprising of at least two tandem duplications and one gene loss during the evolution of the two alleles. Overall, these similarities point out the complexity of the evolutionary process at work at the Rfo locus.

References

- [1] S. Aubourg, N. Boudet, M. Kreis, and A. Lechary. In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol. Biol.*, 42(4):603–613, 2000.
- [2] S. Desloire, H. Gherbi, W. Laloui, S. Marhadour, V. Clouet, L. Cattolico, C. Falentin, S. Giancola, M. Renard, F. Budar, I. Small, M. Caboche, R. Delourme, and A. Bendahmane. Identification of the fertility restoration locus, rfo, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Reports*, 2003.
- [3] E. Kotera, M. Tasaka, and T. Shikanai. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, 433(7023):326–30, 2005.
- [4] L. Noe and G. Kucherov. YASS: enhancing the sensitivity of DNA similarity search. *Nucl. Acids Res.*, 33(S2):W540–543, 2005.
- [5] R. Uricaru, L. Bréhélin, and E. Rivals. A new type of hidden markov models to predict complex domain architecture in protein sequences. In C. Brun and G. Didier, editors, *Actes de Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, pages 97–102, July 2007.