

Origins of rice cytoplasmic male sterility genes

Cell Research (2017) 27:3-4. doi:10.1038/cr.2016.158; published online 27 December 2016

Cytoplasmic male sterility (CMS) factors have long been known in some wild plants, and also in some domesticated species, where they are used to produce plants to be used as maternal parents, for example to breed hybrids that display hybrid vigor. Their origins have been mystifying, and now a study recently published in *Cell Research* helps understand how one widely-used rice CMS factor evolved.

Although male sterility mutations are evidently deleterious, maternally inherited cytoplasmic mutations are favored if they confer even a slightly higher female fertility. However, because the females require pollen from non-male-sterile individuals, the mutation will not spread throughout the population, resulting in polymorphism, with both females and non-male-steriles present. Eventually mutation at a nuclear gene may restore male fertility to individuals carrying the maternally inherited sterility factor. One cytoplasmic genotype may take over the population (called an arms-race event) [1], or sometimes both CMS factors and nuclear restorers can remain segregating, potentially for many generations (balancing selection) [2-4].

CMS factors are variants of the mitochondrial genomes, often (perhaps always) involving chimeric open reading frames (ORFs) of mitochondrial genes [5], but their origins, and how their appearance causes male-sterility, are mysterious. Hybrid breeding depends on breeders finding them in their strains, presumably descended from crops' wild ancestors. Understanding the origin of CMS factors might therefore require evolutionary approaches beyond the crop of interest. The new study makes important advances by examining wild rice, *Oryza*

rufipogon [6], from which cultivated rice was domesticated.

The study illuminates the origin of Wild Abortive CMS (CMS-WA), one CMS factor used in domesticated rice, *O. sativa*. CMS-WA has the CMS cytoplasm of a male sterile *O. rufipogon* plant, and the sterility gene responsible had previously been identified [7]. The new paper calls it *WA352c*, as it encodes 352 amino acids; the protein product was shown to accumulate specifically in tapetal cells, where it interacts with the highly conserved mitochondrial protein COX11 (encoded by a nuclear gene), impairing COX11's function, leading to pollen abortion.

The new work searched rice genomes for sequences similar to *WA352c*, or parts of it, first by looking in GenBank, and then using PCR primers targeting possible rearranged segments. More than 200 individuals of both wild and domesticated rice, plus smaller samples of other rice species, were studied, revealing a bewildering multiplicity of chimeric ORF-containing structures in *O. rufipogon*, including one similar to the *O. sativa WA352c* sequence (Figure 1); the true diversity could be even higher, as some plants lacked all these structures. The authors deduced a plausible sequence of evolutionary changes, involving repeated rearrangements creating chimeric structures and copy number differences, probably involving recombination events between different parts of the mitochondrial genome with sequence homology. Intra-genome recombination events are well documented in plant mitochondria.

The finding of different structures within *O. rufipogon* could represent different CMS types maintained within this species by balancing selection (with

shared variants in *O. sativa* reflecting these species' very recent split). *O. rufipogon* populations often have high outcrossing rates [8], therefore male-sterility might be advantageous if it increases female fertility (or does not decrease it greatly), especially if strong inbreeding depression occurs. However, most of the mitochondrial structures related to the WA-CMS factor were not found (and are therefore probably absent or rare) in wild species other than *O. rufipogon*, which probably diverged from *O. rufipogon* within the past 2 million years. Clearly, therefore, the mitochondrial arrangements do not persist for very long, suggesting an alternative – constant generation of new rearrangements. Different structures were sometimes found in single individuals, consistent with very recent rearrangements. However, the study does not exclude the possibility that such heteroplasmy reflects occasional transmission of newly generated variants into new plants through pollen. New variants will initially be rare in the individual's mitochondrial population, but, if their frequencies change, the predominant structures might soon differ in different individuals (perhaps within 2 million years).

In this view, the chimeric structures are a kind of mutation with a high mutation rate due to the propensity of plant mitochondrial genomes to recombine and cause rearrangements. New non-functional structures would at best be neutral, and would probably quickly undergo further damaging sequence changes, or be deleterious and deleted, as occurs for many gene duplicates [9] (if structures with many sequence differences exist in wild rice, the approaches used might not detect them). Some structures, however, might exhibit CMS function (which is

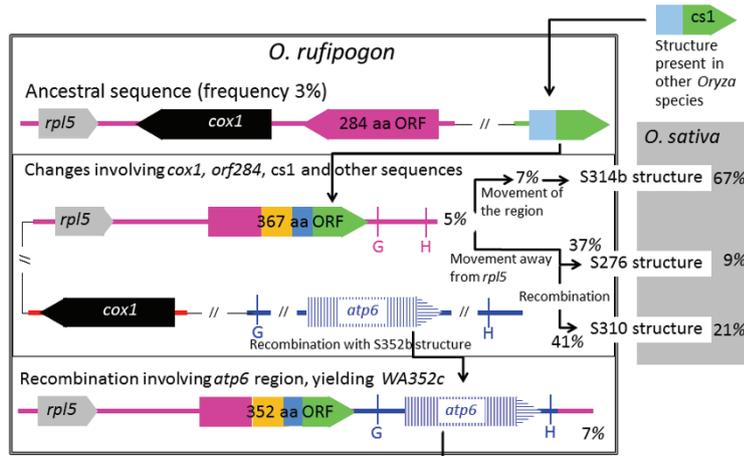


Figure 1 ORF-containing structures in *O. rufipogon* with similarities to portions of the *O. sativa* WA352a, b and c sequences (which differ in the amino acid sequences encoded), including one that could generate the WA352c sequence structure used in rice CMS. Related *O. sativa* structures, and their frequencies, are shown in the grey boxes at the right). The diagram is a greatly simplified version of Figure 4 in the paper, omitting amino acid substitutions, to show just two of the inferred major rearrangements that can generate the many variants found in *O. rufipogon*; their frequencies in the sample of 200 plants studied from this species are also indicated (they sum to < 100% as a few structures rarely found in *O. rufipogon* are omitted). Different mitochondrial genome regions are indicated by horizontal lines in different colors, and thin black lines indicate other parts of the genome, with symbols to indicate gaps where the large-scale organization is not known (the orders of the regions separated by gaps are unknown; for instance, except for the structure at the top, the *cox1* gene was transferred to another location in the genome [7, 12]). The top part of the diagram shows the probable ancestral mitochondrial genome structure (now rare in *O. rufipogon*); the region that carried the *cs1* sequence is also seen in other *Oryza* species and other grasses, but *cs1* is physically separated from the ORF that is involved in the hypothesized rearrangements that yielded a new, chimeric ORF. The middle part of the diagram shows that this new structure gave rise to a diversity of ORF lengths in *O. rufipogon*, one of which predominates in the *O. sativa* sample, although two others are also shared by both species. As shown at the bottom of the diagram, this region later changed again, with insertion of the *atp6* gene from yet another genome region, and loss of the *cox1* gene from the focal region, yielding the WA352c sequence (this is also rarely seen in the *O. rufipogon* sample) that was later inherited by some *O. sativa* cultivars.

simply inhibition of COX11 function, and resembles a known class of deleterious mutations).

Structures with even partial, weak CMS function might be maintained as polymorphic “protogenes” for long enough that natural selection can improve this function by sequence changes that produce more strongly COX11-interacting protein domains, plus high expression in tapetum (and perhaps not expressed in other tissues where accumulation of the protein might be detrimental). Has CMS evolution indeed involved enhancement of protogenes’ male-sterility functions? The paper documents a range of functionality among the structures detected, based on introducing different structures

into cultivated rice. Unlike WA352c and a (already known to have CMS function), some structures exhibited no detectable CMS activity. One of them encodes a protein that could interact with COX11 and was therefore previously thought likely to have activity, but it is not expressed in anther tissues, whereas the proteins expressed by some others interact weakly with COX11. Analyses of sequence changes gave no evidence that natural selection has promoted sequence changes in the COX11-interacting regions of structures that evolved CMS function, or that male-sterility function of new structures has become enhanced, but this would be difficult to detect, especially if different mitochondrial genome

types can occasionally be transmitted by pollen, rather than purely maternally, causing heteroplasmic genotypes whose sequences can recombine [10]. CMS-WA is just one of several CMS factors in wild rice, and, if recombination between different types occurs, this would add to the complexity.

It remains unclear how the rest of the mitochondrial genome is affected by the rearrangements discovered in this study, and before how long the sequence of any other, non-CMS factor regions, remains associated with the causal chimeric ORF. Constant generation of new CMS structures might quickly eliminate associations with CMS types, which might therefore often not be found when examining populations [11]. Rice is probably not the only plant in which mitochondrial genome rearrangements occur frequently, and this study should inspire renewed work in other natural plant populations with CMS polymorphisms, where the maintenance of diversity remains puzzling.

Deborah Charlesworth¹

¹University Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Lab King’s Buildings, W Mains Rd, Edinburgh EH9 3FL, UK

Correspondence: Deborah Charlesworth
E-mail: deborah.charlesworth@ed.ac.uk

References

- 1 Frank SA. *Am Nat* 1989; **133**:345-576.
- 2 Jacobs MS, Wade MJ. *Am Nat* 2003; **161**: 837-851.
- 3 Touzet P. In Maréchal-Drouard L, Ed. Advances Botanical Research 2012; **63**:71-98.
- 4 Couvet D, Ronce O, Gliddon C. *Am Nat* 1998; **152**:59-70.
- 5 Chase C. *Trends Genet* 2007; **23**:81-90.
- 6 Tang H, Zheng X, Li C, et al. *Cell Res* 2016; **27**:131-147.
- 7 Luo D, Xu H, Liu Z, et al. *Nat Genet* 2013; **45**:573-577.
- 8 Barbier P. *Jpn J Genet* 1989; **64**:273-285.
- 9 Walsh JB. *Genetics* 1995; **139**:421-428.
- 10 Pearl SA, Welch ME, McCauley DE. *Mol Bio Evol* 2009; **26**:537-545.
- 11 deHaan AA, Mateman AC, Van Dijk PJ, et al. *Theor Appl Genet* 1997; **94**:539-548.
- 12 Bentolila S, Stefanov S. *Plant Physiol* 2012; **158**:996-1017.