Commentary

Nuclear-mediated mitochondrial gene regulation and male fertility in higher plants: Light at the end of the tunnel?

Roger P. Wise[†‡] and Daryl R. Pring[§]

[†]U.S. Department of Agriculture–Agricultural Research Service, Corn Insects and Crop Genetics Unit, Department of Plant Pathology, Iowa State University, Ames, IA 50011; and [§]U.S. Department of Agriculture–Agricultural Research Service, Crop Genetics and Environment Unit, Department of Plant Pathology, University of Florida, Gainesville, FL 32611

Hybrid vigor, or heterosis, in higher plants is often attained through the use of cytoplasmic male sterility (CMS), a maternally inherited trait characterized by the absence of functional pollen. Hybrids of many plants are produced using CMS, wherein a male-sterile line is grown adjacent to a selected male-fertile line, which functions as a pollen source. Seed formed on the male-sterile line is then grown as a hybrid. When the hybrid plant is harvested for its vegetative parts, sugarbeet for instance, it is immaterial that the F_1 hybrid is male-sterile. However, in plants where the seed is harvested, it is imperative that the F_1 hybrid be male-fertile. Fundamental characteristics of CMS impact on this consideration. Historically, most sources of CMS were discovered as a result of genetic crosses involving normal, male-fertile plants, wherein male-sterile plants were observed among the progeny. Lines that allow male sterility to be expressed were thus identified as potentially useful for plant breeding. It is this fact that enables the utilization of CMS: some lines allow CMS to be expressed and some do not.

CMS systems are found in over 150 plant species and are usually attributed to chimeric ORFs in the mitochondrial genome. These ORFs encode novel proteins, which often interfere with mitochondrial function and pollen development. Evolution has provided the answer to these aberrant mitochondrial proteins through nuclear-encoded, restoration of fertility, or *Rf*, genes, which suppress mitochondrial abnormalities associated with male sterility. This suppression allows normal metabolic processes that lead to successful microsporogenesis. In many instances, this suppression is directly associated with *Rf*-gene-dependent, mitochondrial RNA modification and concurrent reduction of the CMS-associated protein (1). Although many mitochondrial genes associated with CMS have been characterized, the identification of *Rf* genes has proven elusive. Only maize *Rf2a*, a mitochondrial aldehyde dehydrogenase (ALDH), has been identified to date (2, 3). It is in this context that the observations of Bentolila *et al.* (4) represent a critical advance in the fundamental understanding of posttranscriptional, mitochondrial gene regulation and the restoration of male fertility in higher plants. In this issue of PNAS, Bentolila *et al.* report the molecular identification of *Petunia Rf*, the first isolation of a gene that controls the expression of an organellar gene encoding cytoplasmic male sterility.

The CMS phenotype of *Petunia* is due to the expression of a 1.2 kilobase chimeric ORF, designated *pcf* (petunia CMS-associated fused). The *pcf* ORF is comprised of a 5′ segment of the *atp9* gene, parts of the first and second exons of the *coxII* gene, and unidentified sequences (*urfS*) (5). In the presence of *Rf,* the accumulation of *pcf*-derived transcripts is altered, and the accumulation of the 25-kDa PCF protein is reduced. As demonstrated by Bentolila *et al.* (4), *Petunia Rf* encodes a mitochondrially targeted protein almost entirely comprised of 14 repeats of a 35-aa pentatricopeptide repeat (PPR) motif. How then can the molecular identity of *Petunia Rf* be reconciled with its known effect on the *pcf* transcript or its encoded mitochondrial protein? There are over 200 genes harboring the PPR-motif, and its related TPR (tetratricopeptide repeat)-motif in the *Arabidopsis* genome, and two-thirds of these proteins are predicted to be targeted to organelles (6). PPR- and TPR-motifs are found in helical-repeat proteins and would be predicted to have protein-binding properties. A good example of this result is the maize nuclear-encoded protein, CRP1, which, as part of a multisubunit complex, is required not only for the translation of the chloroplast *petA* and *petD* mRNAs but also for the processing of the *petD* mRNA from a polycistronic transcript (7). The CRP1 sequence includes multiple tandem copies of this motif, and appears to activate a site-specific endonuclease independent of the role it plays in translation.

How then can the molecular identity of *Petunia Rf* be reconciled with its known effect on the *pcf* transcript or its encoded mitochondrial protein?

Although TPR-containing proteins, would be predicted to mediate protein—protein interactions, PPR-containing proteins, on average, contain many more repeats that could result in additional ligands (6). Because the width of the central groove of PPR-containing proteins is sufficient to hold an RNA strand and the positively charged surface at bottom of the groove could bind the phosphate backbone, Small and Peeters (6) suggest the possibility that PPR proteins could also be RNA binding. Taken together, an attractive scenario emerges wherein the multifamily of proteins carrying the PPR motif might participate in recognition and facilitation of events such as RNA editing (6) and endonucleolytic processing (7). In yeast, 50% of the cellular proteins are assembled into complexes comprised of one to five proteins; the other 50% are assembled into complexes with even greater numbers (8). If a particular RF protein is part of an RNA processing complex, and the PPR domain is essential, the determination that *Petunia Rf* is a member of one

See companion article on page 10887.

[‡]To whom reprint requests should be addressed. E-mail: rpwise@iastate.edu.

www.pnas.org/cgi/doi/10.1073/pnas.172388899

of the largest gene families in plants (9) presents a plethora of opportunities for the creation of alternate processing specificities. Interaction among various CMS-encoded transcripts and different PPR-containing protein complexes could potentially determine processing specificity. This hypothesis is consistent with the known transcript processing conferred by many *Rf* genes on CMS-associated transcripts (1, 10).

Mitochondrial genomes are models for recombination dynamics. There are hundreds of copies per cell, and, as such, they can undergo a wide range of recombination events through direct and indirect repeats. Hence, mitochondrial DNA is essentially a polyploid genome, sustaining deletions and duplications, with little consequence to the general viability of the organism. CMS-associated genes are often chimeric, derived from portions of known genes fused with previously unknown sequences. In most instances, the sequences of CMS-associated ORFs are unrelated, except for parts of common mitochondrial genes, such as *atp6*, *atp8* (*orfB*), *atp9*, or *cox2* (1, 10). Conserved segments of such mitochondrial genes could be recruited to function as RNA processing targets for nuclear restorer genes, acting on the chimeric template. That these nuclear-specific, differential-processing events need not be associated with fertility restoration genes is readily apparent. The relationship of two altered *atp1* transcripts and fertility restoration in radish was discounted on examination of a number of lines, some of which did not reveal the transcriptional variant, thus indicating that *Rf* genes can also be linked to genes that alter mitochondrial transcripts (11). Nucleus-specific variants for apparent processing of *urf221* transcripts have been demonstrated in N (normal) cytoplasm, male-fertile maize (12, 13). Additionally, pleiotropic effects of *Rf* genes have also been observed. Maize *Rf3*, which alters transcripts of *orf355/77*, associated with S cytoplasm CMS plants (14), also mediates the modification of *cob* and *atp6* transcripts (15).

Nuclear gene duplication in higher plants propagates genetic diversity through multiple cycles of selection and the creation of new gene-family members. Unequal crossover among new gene-family members has the potential to generate ORFs with new or altered functions differing from the progenitor sequence. The products of such recombination, or other mutational events, are often recruited for new or novel functions. Thus, the coevolution of nuclear and organelle genomes has resulted in multiple processes that are coregulated in response to a particular cellular environment; CMS is a typical example of such processes.

The *Petunia Rf* locus contains at least two copies of a PPR gene family. Both copies carry a mitochondrial transit peptide, but, to date, only *Rf-PPR592* has been demonstrated to restore fertility in transgenic plants. A nonrestoring allele, *rf-PPR592,* is expressed in roots, but not in leaves, floral buds, and stem, apparently because of a recombination event between two genes similar to *Rf-PPR591* and *Rf-PPR592* and the loss of part of its promoter. This illegitimate recombination also results in the inability of this allele to restore fertility.

We can probably expect to see additional duplicated loci encoding proteins that function in nuclear-cytoplasmic interactions. For example, the *Rf2a*-encoded mitochondrial ALDH (2, 3) has at least three other family members dispersed throughout the maize genome (16). Two of the *rf2*-encoded proteins are classified as mitochondrial ALDHs, and two are most likely cytosolic. Both mitochondrial and cytosolic proteins exhibit ALDH activity, but, to date, only RF2a activity has been shown to be required for male fertility in T-cytoplasm maize. Interestingly, RF2a is also required for normal anther development in N cytoplasm (3).

The complementary T-cytoplasm restorers, *Rf1, Rf8,* and *Rf*,* show evidence of functional duplication as well. *Rf1* is positioned near the centromere on chromosome *3* (17), whereas *Rf8* and *Rf** are either alleles of one locus or tightly linked genes on chromosome *2L* (D. Pei and R.P.W., unpublished observations). All three of these restorers share a small, conserved target sequence in the T-*urf13* ORF, yet control independently the modification of T-*urf13* CMS transcripts (18). Interestingly, this same target sequence is also highly conserved among sites for *Rf**-mediated T-*urf13* processing and the CMS-associated *orf107* processing regulated by sorghum *Rf3* (18–20). It is unknown whether sorghum *Rf3* is also an orthologue of maize *Rf**; however, with the recent positioning of these two grass genes (ref. 21; and D. Pei and R.P.W., unpublished observations) in reference to molecular markers, this possibility can be tested. The *Rf8/Rf** region on maize *2L* also harbors *Rf3*, another nuclear gene that modifies mitochondrial transcripts, but its action is specific to S-cytoplasm maize (14, 15, 22). Thus, it appears that this RNA processing class of restorers may encode functionally similar gene products that may have been recruited for novel functions via gene duplication and subsequent modification.

Although the specific molecular mechanisms are unknown, nuclear regulation of mitochondrial RNAs is common to many CMS systems, such as sorghum, rice, wheat, sunflower, oilseed rape, and common bean, as well as petunia. A plausible scenario, based on the Bentolila *et al.* (4) and Small and Peeters (6) conclusions, is that the PPR family of genes may indeed be associated with a variety of sequence-specific events involved in the maturation of plant mitochondrial (and chloroplast) mRNAs, such as RNA editing and nuclear-mediated endonucleolytic processing. The multiple descriptions of such events in plant mitochondrial literature provide a fertile field for the investigation of mitochondrial-gene expression and modification by nuclear-encoded factors, specifically how posttranscriptional processing of unique mitochondrial RNAs mediates the restoration of male fertility in CMS systems.

However, because known gene products have not been associated with fertility restoration, *Rf* genes have so far been isolated via genetic methods that rely on a mutant phenotype, such as map-based cloning (23) or transposon tagging (17, 24). The observations of Bentolila *et al.* (4) provide compelling impetus for studies of possible relationships between the PPR gene family and other *Rf* genes. As stated above, PPR-related proteins belong to a large gene family, with possibly over 200 members in *Arabidopsis* (6, 9), and, thus, a significant challenge is posed in using these sequences as candidates to recover *Rf* genes in other higher plants. Nevertheless, a candidate gene approach can be of particular use when searching large complex genomes where little sequence data are available. For example, sequence-specific amplified fragment length polymorphism (AFLP) can be used to quickly map a large number of PPR/TPR-containing genes on existing genetic-mapping populations (25, 26). By using conserved PPR/TPR primers paired with random AFLP primers, whole-genome surveys of PPR-specific polymorphisms can be easily visualized. Use of specialty mapping populations can be used to assay whether a particular clone cosegregates with the *Rf* phenotype. Additionally, computational tools can be used in combination with plant-specific expressed sequence tag (EST) databases to mine PPR/TPR harboring sequences, many of which already have map positions. The favored substrate would be a small ($<$1 Mb) physical contig spanning a candidate *Rf* locus, and, in the near future, draft sequence will be available for investigation in a variety of plants.

However, because CMS-associated organellar genes represent a multitude of diverse targets (1, 10), there likely are additional yet undiscovered, nuclear-encoded *Rf* factors.

The ramifications of the Bentolila *et al.* data (4) in clarifying mitochondrial gene regulation as influenced by nuclear background are evident. The continued use of CMS in the efficient production of higher plant F_1 hybrids, as well as designing novel approaches, underscores the need for elucidation of these phenomena. The nature of *Rf* genes that affect mitochondrial gene transcription, prominently illustrated in the case of CMS, has long been considered a "black box." The possibility that PPR-related proteins may be involved in other, similar events allows the design of strategic approaches in addressing these examples, based on the *Petunia* model.

Note Added in Proof. It has been recently presented that the cloned restorer of fertility, *rfk1* from Kosena radish, also contains a PPR motif.[¶]

[¶]Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T. & Imamura, J., Sixth International Congress on Plant Mitochondria, July 10–14, 2002, Perth, Australia (abstr.).

1. Schnable, P. S. & Wise, R. P. (1998) *Trends Plant Sci.* **3,** 175–180.
2. Cui X., Wise R. P. & Schnable P. S. (1996) *Science* **272,** 1334–1336.
3. Liu, F., Cui, X., Horner, H. T., Weiner, H. & Schnable, P. S. (2001) *Plant Cell* **13,** 1063–1078.
4. Bentolila, S., Alfonso, A. A. & Hanson, M. R. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 10887–10892.
5. Hanson, M. R., Sutton, C. A. & Lu, B. (1996) *Trends Plant Sci.* **1,** 57–64.
6. Small, I. D. & Peeters, N. (2000) *Trends Biochem. Sci.* **25,** 46–47.
7. Fisk, D G., Walker, M. B. & Barkin, A. (1999) *EMBO J.* **18,** 2621–2630.
8. Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M., Cruciat, C.-M., *et al.* (2002) *Nature (London)* **415,** 141–147.
9. Aubourg, A., Boudet, N., Kreis, M. & Lecharny, A. (2000) *Plant Mol. Biol.* **42,** 603–613.
10. Kempken, F. & Pring, D. R. (1999) *Prog. Bot.* **60,** 140–166.
11. Makaroff, C. A. (1995) in *The Molecular Biology of Plant Mitochondria*, eds. Levings, C. S., III, & Vasil, I. K. (Kluwer, Boston), pp. 515–555.
12. Gupta, D. & Abbott, A. G. (1991) *Theor. Appl. Genet.* **82,** 723–728.
13. Rocheford, T. R. & Pring, D. R. (1994) *Theor. Appl. Genet.* **89,** 951–958.
14. Zabala, G., Gabay-Laughnan, S. & Laughnan, J. R. (1997) *Genetics* **147,** 847–860.
15. Wen, L. & Chase, C. D. (1999) *Curr. Genet.* **35,** 521–526.
16. Skibbe, D. S., Liu, F., Wen, T.-J., Yandeau, M. D., Cui, X., Cao, J., Simmons, C. R. & Schnable, P. S. (2002) *Plant Mol. Biol.* **48,** 751–764.
17. Wise, R. P., Dill, C. L. & Schnable, P. S. (1996) *Genetics* **143,** 1383–1394.
18. Dill, C. L., Wise, R. P. & Schnable, P. S. (1997) *Genetics* **147,** 1367–1379.
19. Tang, H. V., Pring, D. R., Shaw, L. C., Salazar, F. A., Muza, F. R., Yan, B. & Schertz, K. F. (1996) *Plant J.* **10,** 123–133.
20. Wise, R. P., Bronson, C. R., Schnable, P. S. & Horner, H. T. (1999) *Adv. Agron.* **65,** 79–130.
21. Wen, Y., Tang, H. V., Chen, W., Chang, R., Pring, D. R., Klein, P. E., Childs, K. L. & Klein, R. R. (2002) *Theor. Appl. Genet.* **104,** 577–585.
22. Kamps, T. L. & Chase, C. D. (1997) *Theor. Appl. Genet.* **95,** 525–531.
23. Bentolila, S. & Hanson, M. R. (2001) *Mol. Genet. Gen.* **266,** 223–230.
24. Schnable, P. S. & Wise, R. P. (1994) *Genetics* **136,** 1171–1185.
25. Waugh, R., McLean, K., Flavell, A. J., Pearce, S. R., Kumar, A., Thomas, B. B. T. & Powell, W. (1997) *Mol. Gen. Genet.* **253,** 687–694.
26. Yu, G. X. & Wise, R. P. (2000) *Genome* **43,** 736–749.