In Silico Identification of Candidate Genes for Fertility Restoration in Cytoplasmic Male Sterile Perennial Ryegrass (*Lolium perenne* L.)

Timothy Sykes\(^1\), Steven Yates\(^1\), Istvan Nagy\(^2\), Torben Asp\(^2\), Ian Small\(^3\), and Bruno Studer\(^1,\ast\)

\(^1\)Institute of Agricultural Sciences, Forage Crop Genetics, ETH Zurich, Zurich, Switzerland
\(^2\)Department of Molecular Biology and Genetics, Research Centre Flakkebjerg, Aarhus University, Slagelse, Denmark
\(^3\)Plant Energy Biology, ARC Centre of Excellence, the University of Western Australia, Crawley, Western Australia, Australia

\(\ast\)Corresponding author: E-mail: bruno.studer@usys.ethz.ch.

Data deposition: The supplementary table provided had all relevant sequence identifiers.

Accepted: February 27, 2016

Abstract

Perennial ryegrass (*Lolium perenne* L.) is widely used for forage production in both permanent and temporary grassland systems. To increase yields in perennial ryegrass, recent breeding efforts have been focused on strategies to more efficiently exploit heterosis by hybrid breeding. Cytoplasmic male sterility (CMS) is a widely applied mechanism to control pollination for commercial hybrid seed production and although CMS systems have been identified in perennial ryegrass, they are yet to be fully characterized. Here, we present a bioinformatics pipeline for efficient identification of candidate restorer of fertility (*Rf*) genes for CMS. From a high-quality draft of the perennial ryegrass genome, 373 pentatricopeptide repeat (PPR) genes were identified and classified, further identifying 25 restorer of fertility-like PPR (*RFL*) genes through a combination of DNA sequence clustering and comparison to known *Rf* genes. This extensive gene family was targeted as the majority of *Rf* genes in higher plants are *RFL* genes. These *RFL* genes were further investigated by phylogenetic analyses, identifying three groups of perennial ryegrass *RFL* genes. These three groups likely represent genomic regions of active *RFL* generation and identify the probable location of perennial ryegrass PPR-*Rf* genes. This pipeline allows for the identification of candidate PPR-*Rf* genes from genomic sequence data and can be used in any plant species. Functional markers for PPR-*Rf* genes will facilitate map-based cloning of *Rf* genes and enable the use of CMS as an efficient tool to control pollination for hybrid crop production.

Key words: cytoplasmic male sterility (CMS), hybrid breeding, pentatricopeptide repeat (PPR) proteins, perennial ryegrass (*Lolium perenne* L.), restoration of fertility, restorer of fertility-like PPR (RFL).

Introduction

The agronomical value of perennial ryegrass (*Lolium perenne* L.) comes from its ability to produce high forage yield of good feed quality in both permanent and temporary grassland systems (Wilkins 1991). Due to the increasing global demand for animal products, improved varieties of forage grasses are becoming an important aspect of global food security. Thus, perennial ryegrass has been the subject of intensive breeding efforts over recent decades. However, these breeding efforts are mainly focused on the improvement of population and synthetic varieties and show limited increases in biomass yield (van der Heijden and Roulund 2010; Pembleton et al. 2015), which is one of the most important traits in forage grasses.

Hybrid breeding, by efficiently exploiting the phenomenon of heterosis, has been successfully used in breeding programs to increase yield in several important crop species including rice (*Oryza sativa* L.), maize (*Zea mays* L.), and rapeseed (*Brassica napus* L.) (Duvick 2001; Melchinger 2010). Due to its significant impact, there are currently considerable efforts to establish hybrid breeding schemes for other crops including wheat (*Triticum aestivum* L.) (Longin et al. 2012). The development and application of hybrid breeding in forage crops has the potential to result in similar yield increases (Pembleton et al. 2015). To employ hybrid breeding in perennial ryegrass, one of the major challenges is the absence of a pollination control strategy that would allow the efficient production of...
hybrid seed on a commercial level. In several plant species including maize, onion (Allium cepa L.), sorghum (Sorghum bicolor L.), sugar beet (Beta vulgaris L.), sunflower (Helianthus annuus L.), rapeseed, common beans (Phaseolus vulgaris L.), and rice, cytoplasmic male sterility (CMS) has been successfully applied to control pollination for hybrid seed production (Ahokas 1983; Yuan and Virmani 1988; Virmani 1994; Schnable and Wise 1998; Havey 2004; Martin et al. 2009; Singh et al. 2010; Kubo et al. 2011). Although CMS systems have been identified in perennial ryegrass (Wit 1974; Connolly and Wright-Turner 1984; Creemersmolenaar et al. 1992), they are yet to be fully characterized (Kiang et al. 1993; Kiang and Kavanagh 1996; McDermott et al. 2008; Islam et al. 2014).

CMS in flowering plants is characterized by a maternally inherited inability to produce functional pollen (Hanson and Bentolila 2004). This functional defect is often attributed to aberrant transcripts originating from the mitochondrial genome, with these CMS causing transcripts usually coding for novel chimeric open reading frames (ORFs) containing part of a functional mitochondrial gene (Chase and Babay-Laughnai 2004; Hanson and Bentolila 2004). The translated products of these chimeric transcripts disrupt normal mitochondrial function such that the energy requirements for pollen formation cannot be met, rendering the pollen uviable (Schnable and Wise 1998).

The CMS phenotype is often restored through the action of nuclear-derived RNA-binding proteins that are generally members of the large family of pentatricopeptide repeat (PPR) proteins (Barkan and Small 2014). Exceptions are the CMS-T restoration in maize (Cui et al. 1996), the restorer of fertility (Rf) gene bvORF20 in sugar beet (Kitazaki et al. 2015) as well as other RNA-binding proteins that have been implicated in fertility restoration (Itabashi et al. 2011; Hu et al. 2012). PPR proteins are particularly numerous in land plants, with 450 PPRs identified in Arabidopsis (Arabidopsis thaliana L.) and 477 in rice (Schmitz-Linneweber and Small 2008; Zehmann et al. 2009; Chateigner-Boutin and Small 2010; Castandet and Araya 2011; Fuji and Small 2011). Although PPR proteins are encoded by the nuclear genome, they most often function within organelles to mediate gene expression, facilitating the processing, and translation of RNAs (Small et al. 2013). PPR proteins contain tandem arrays of a degenerate 35 amino acid motif that bind to RNA in a sequence-specific manner (Schmitz-Linneweber and Small 2008). PPR proteins appear to be functional only in organelles and as such have been described as the chaperones of organelle gene expression (Colcombet et al. 2013). PPR proteins have previously been divided into subclasses based on PPR motif variations and a series of conserved C-terminal domains (Lurin et al. 2004; O'Toole et al. 2008). The two main subclasses of PPR proteins, the P and PLS subclasses, are defined by the organization of the individual PPR motifs within a PPR gene. The P-type PPRs are comprised almost entirely from the canonical 35-amino acid P motif. In contrast, the PLS subclass of PPRs is composed of triplet repeats containing one P motif, one L motif (“long,” usually 36 amino acids) and one S (“short,” usually 31 amino acids). This PLS subclass is also characterized by three distinctive C-terminal motifs; E (extended), E+ (slightly longer version of the E-domain), and DYW (named for terminating with a conserved Asp-Tyr-Trp triplet). All PPRs that have been shown to be involved in RNA editing, in both mitochondria and chloroplasts, are members of these three subgroups (Schallenberg-Ruedinger et al. 2013). The E/E+ domains are believed to provide an essential recognition site for an (as yet unidentified) editing complex. The DYW domain, which usually includes an E domain, shows similarity to deaminases and is possibly directly involved in RNA editing (Hammani et al. 2009; Okuda et al. 2009; Tasaki et al. 2010; Okuda and Shikanai 2012; Toda et al. 2012).

A subgroup of the P-type PPRs is specifically linked to fertility restoration of CMS; the restorer of fertility-like PPR (RFL) proteins. This group is identified by their relative homology from within the CMS family, their identity with other known CMS restorer PPRs from related plant species and their tendency to be present in several homologous copies clustered within the genome. These RFLs comprise around 10–30 members per plant genome from the full set of PPRs (Andres et al. 2007; Fujii et al. 2011). It has been shown previously that RFL genes appear to be under different selection pressures when compared with the rest of the PPR gene family members. Within the RFL subgroup, high ratio of nonsynonymous versus synonymous nucleotide substitutions indicates diversifying selection (Geddy and Brown 2007; Fujii et al. 2011). This suggests, in conjunction with gene duplication, that the generation of new RFL genes and subsequent loss of nonfunctional RFLs is relatively rapid, keeping pace with the generation of novel CMS sources. CMS is also used as a model system for studying nuclear/mitochondrial genome interactions, as its easy detection allows researchers to rapidly identify individuals with a breakdown in nuclear/mitochondrial signaling (Chase 2007).

In order to provide plant breeders with a molecular tool for candidate Rf gene identification and thus facilitate the implementation of hybrid breeding schemes in perennial ryegrass, this study aimed to locate, in silico, regions of active RFL generation in the perennial ryegrass genome by 1) the development and validation of a bioinformatics pipeline for the identification of PPRs and RFLs from genomic sequence, 2) utilizing this pipeline for identification of PPR genes within the perennial ryegrass genome, 3) classifying these PPR genes in order to isolate the RFLs as potential candidate Rf genes, 4) phylogenetically analyzing the RFL genes from several grass species to identify groups of rapidly diverging RFL genes within the perennial ryegrass genome, and 5) using this analysis to locate genomic regions of novel RFL generation.
Materials and Methods

Identification of PPR Proteins

To identify, in silico, members of the PPR protein family in the genome assembly of perennial ryegrass (http://185.45.23.197:5080/ryegrassgenome), all available PPR domain sequences from the Pfam database (http://pfam.xfam.org) were collected and used for the development of a Hidden Markov Model (HMM) profile matrix using the hmmsbuild program of the HMMER package (v3.1b1, http://hmmer.org). This HMM profile matrix was used to identify members of the PPR family in a total of 71,009 translated DNA transcript sequences obtained from ab initio and evidence-based predictions from a high-quality genomic draft of the perennial ryegrass genome sequence (Byrne et al. 2015).

Classification of PPR Proteins

PPR-containing transcript sequences were analyzed on a standalone PfamScan pipeline to ascertain the exact co-ordinates of each PPR domain within a scaffold sequence as well as information on the frequencies and distribution of the PPR domains. Predictive information on protein functions and conserved sequence elements was obtained by sending all PPR containing sequences through a standalone InterProScan (version 5; Jones et al. 2014) pipeline by scanning the PANTHER, PROSITE profiles, Pfam, and SUPERFAMILY databases. Sequences were identified as belonging to the P or PLS subfamilies through analysis of PPR motif lengths, with the PLS subfamily having longer (L) and shorter (S) subdomains (Lurin et al. 2004). The identified members of the PLS family were processed using the online domain elicitation tool MEME (Bailey and Gribskov 1998) and conserved blocks representing the E, E +, and DYW C-terminal domains identified. To ensure all possible C-terminal domains have been identified, the PPR domains were masked out using the maskfeat program of the EMBOSs package (Rice et al. 2000). The masked sequences were aligned and clustered to identify any conserved regions outside of the PPR domains. All sequences were also searched using HMM profiles for the E, E +, and DYW domains.

Identification of RFL Proteins

All identified PPR genomic sequences were clustered using CD-hit (Li and Godzik 2006) at 90%, 80%, 70%, 60%, and 50% identity. Clustering at 90–70% revealed no clusters of more than three members. Clustering at 60% revealed three clusters containing 9, 6, and 4 PPRs, respectively. All PPR sequences were then aligned, using the NCBI BLAST platform (http://blast.ncbi.nlm.nih.gov), to known or predicted restorer genes from; brachypodium (gi|357139997), rice (gi|33859441), and maize (gi|662249846). Hits with at least 50% identity and 50% query cover were collected. PPRs that were present on at least three of these four lists were considered candidate RFLs.

Databases

The coding sequences (CDS) of the following species were downloaded from Ensembl Plants (http://plants.ensembl.org/index.html, on April 10, 2014) (Flicek et al. 2012) using the Perl API tool (McLaren et al. 2010); A. thaliana (TAIR10), Brachypodium distachyon (V1.0), Hordeum vulgare (European Nucleotide assembly (ENA): GCA_000326085.1), Musa acuminata (ENA: GCA_000313855.1), O. sativa Japonica (ENA: GCA_000005425.2), Setaria italica (ENA: GCA_000263155.1), S. bicolor (ENA: GCA_000003195.1), Triticum urartu (ENA: GCA_000347455.1), Z. mays (ENA: GCA_000005050.5). The following CDS of Phyllostachys heterocla da (v1.0) was downloaded from http://www.bamboogdb.org/ (Peng et al. 2013). The CDS of L. perenne was received from Ruttink et al. (2013) and the respective CDS of Lolium multiflorum and Festuca pratense were kindly provided by Stoces et al. (in preparation). The Eragrostis tef cDNA was downloaded from http://www.tef-research.org/genome.html (Extended,gte200.cDNA.fa, Cannarozzi et al. 2014) and its CDS determined using ORFpredictor (Min et al. 2005). The cDNA was then searched against a protein BLAST database comprising of A. thaliana, Glycine max, O. sativa Japonica, Populus trichocarpa, and Manihot esculenta, using BLASTP (Altschul et al. 1990) with minimum e-value $1 \times 10^{-5}$. The BLASTP results were used to infer coding frame, all other parameters and methods used were as described by Min et al. (2005).

Orthologous Clustering of Species

To cluster the protein sequences into orthologous clusters, the offline version of OrthoMCL (Li et al. 2003) was used. Briefly, the protein names within a fasta file (per species) were first changed for consistency (also for simplicity) and to ameliorate any problems arising later from special characters and similarities between names. This was done using an in house Perl script. The resulting fasta file was then formatted to make it compliant with the OrthoMCL algorithm (a short species-specific prefix was added to each name for subsequent species identification). The sequences were then filtered for low quality, based on sequence length (>30 aa, retained) and percentage of stop codons (>10%, discarded). From these high quality proteins, an all-versus-all BLASTP was run where all proteins were searched against all proteins (minimum E-value $1 \times 10^{-5}$); the database was not split into subgroups when doing this so no corrections for E-score where necessary. The results of the BLASTP were collated and then parsed before loading into a local MySQL orthoMCL database. In the next stage, pairs of proteins that are potentially orthologs, paralogs or co-orthologs were identified using the OrthoMCL algorithm (Li et al. 2003), where protein pairwise connections were normalized for ortholog pairs between and within species. The resulting potential pairs were then organized in clusters using the MCL algorithm (Enright et al. 2002).
were output and the names were changed back to their original for subsequent work.

Phylogenetic Reconstruction and Analysis
The phylogenetic relationships between the protein sequences from the OrthoMCL generated cluster containing RFLs, including the nine putative ryegrass RFLs not present, were reconstructed and analyzed using web tools made available by The Montpellier Laboratory of Informatics, Robotics and Microelectronics LIRMM (http://www.phylogeny.fr; Dereeper et al. 2008). Sequence alignments were completed using MUSCLE (Edgar 2004), phylogenetic analysis using PhyML (Guindon and Gascuel 2003; Anisimova and Gascuel 2006) and the resulting tree viewed using TreeDyn (Chevenet et al. 2006).

Results

PPR and RFL Gene Identification and Classification
A draft of the perennial ryegrass genome sequence (Byrne et al. 2015) was scanned to identify PPR genes using a HMM profile matrix (Finn et al. 2011). From a total of 71,009 genes, obtained from ab initio and evidence-based gene predictions in the perennial ryegrass genome, 373 PPR genes were identified. These 373 PPR genes were classified into two subfamilies, P and PLS, based on the arrangements of the repeated PPR motifs. Each of these subfamilies contained roughly half of the identified PPR genes with the P subfamily being slightly larger with 207 members, representing 55% of the total PPRs. The PLS subfamily was further grouped based on the presence or absence of the C-terminal domains implicated in RNA editing. From a total of 166 PLS subfamily genes, 40 were missing RNA editing-specific C-terminal motifs (PLS subclass), while the remaining 126 were organized into the E class (72), the E+ class (23), and the DYW class (31) (fig. 1). Analysis of the 25 RFLs, identified by homology to known proteins or proteins of unknown function, revealed that all identified RFLs have an average of 16 PPR domains as compared with 9.7 PPR domains for the remainder of the PPR genes.

RFL Gene Comparison in Multiple Species
Orthologous clustering of protein sequences from 14 species was performed to ascertain whether the identified perennial ryegrass RFL genes are similar to RFL genes from other plant species. For this clustering, the canonical CDS (McLaren et al. 2010) of 14 species were used, comprising a total of 561,090 protein sequences. Of these, 554,468 passed the quality checking by OrthoMCL (Li et al. 2003), of which 403,713 proteins were grouped into 44,672 clusters (fig. 2A). A subset of 5,054 clusters contained proteins from all species, representing 30.6% of the 403,713 clustered proteins. In contrast, 17.3% of the sequences were species-specific and contained in 39.7% of clusters (fig. 2A).

Further analysis identified 287 clusters that contain at least one of the 373 perennial ryegrass PPRs found previously. Plotting the number of species represented in these 287 clusters against the number of proteins present revealed a linear relationship with one clear outlier. This outlying cluster contained 154 proteins originating from 13 species and is more than three times bigger than the second largest cluster. This cluster was entirely composed of PPR proteins and contained 16 of the previously identified 25 RFLs from perennial ryegrass. The nine RFLs not present in this cluster were found to be either pseudogenes or poorly annotated genes leading to them not clustering with the remainder of the RFLs. The following species were dropped: Italian ryegrass (Lolium multiflorum L.) and meadow fescue (Festuca pratensis L.) as their sequences from the OrthoMCL generated cluster (red dot in fig. 2A) were orthologous with other PPR genes. The remaining species were dropped: banana (M. acuminate L.) and teff (E. tef L.) as, although their genomes have been sequenced into scaffolds, these were not organized into contiguous sequences and thus did not provide precise information about genome positions. No RFLs were identified from bamboo (P. heteroclava L.) and teff (E. tef L.) as, although their genomes have been sequenced into scaffolds, these were not organized into contiguous sequences and thus did not provide precise information about genome positions. No RFLs were identified from banana (M. acuminate L.) This approach not only showed that RFL genes form a distinct orthologous group, but also validated the approach used for RFL identification within the perennial ryegrass genome.

Phylogenetic Analysis of the RFL Cluster
Having identified a set of RFL genes from multiple species (supplementary table S1, Supplementary Material online), a phylogenetic analysis was performed in order to understand the evolutionary ancestry underpinning the RFL genes. Protein sequences from the OrthoMCL generated RFL cluster (red dot in fig. 2B) were phylogenetically analyzed, revealing four major clades of RFLs (fig. 3 and table 1). The only dicot
included, Arabidopsis, was represented within an entire clade of its own (clade 3). The other three clades encompassed all the monocot sequences with perennial ryegrass and Brachypodium (B. distachyon L.) being the only species represented in only one clade and wild einkorn wheat (T. urartu L.) being the only species represented in all three monocot clades. All species, with the exceptions of wild einkorn wheat and foxtail millet (Se. italica L.), had a majority of sequences present in only one clade.

To identify the RFL genes from each species that most recently evolved, detailed phylogenetic trees of each clade were coupled with genome location data available from Ensemble Plants (http://plants.ensembl.org/index.html). This revealed that within each clade, RFLs from the same species tend to cluster together with the tightest clusters containing RFLs from the same genomic region of a single species (fig. 4 and table 2). Clades 1–4 had 65%, 23%, 75%, and 52% of the RFLs represented in these species-specific clusters, respectively.

Considering only those species with whole-genome sequence information available, 68% of their RFL genes were present in 13 clusters comprising 0.13% of their combined genomes. For example, in rice, 50% of the identified RFLs were found within 320 kb of chromosome 10 (table 2).

Given the abundance of RFLs within these relatively small genome regions, these sites can be considered hotspots for RFL recombination that exhibit elevated rates of recombination relative to a neutral expectation. RFL genes within these clusters, at the same genome region, will be the youngest as they are still present within this RFL recombination hotspot. This implies that any list of candidate PPR-Rf genes can be further narrowed to RFLs present within these zones. These regions of active RFL generation contained known Rf genes, with the rice Rf1 (Wang et al. 2006) and Rf4 (Luo et al. 2013) genes being present in the RFL-rich region of rice chromosome 10. This allowed us to further refine the list of possible Rf genes in perennial ryegrass by looking for groups of tightly clustering sequences that show a similar pattern to other species. From Clade 2, three groups of perennial ryegrass RFLs meeting these criteria were identified, comprising of seven, eight, and five sequences, respectively (given in blue, fig. 4B). The first of these groups contained only sequences present in the OrthoMCL RFL cluster, the second group four sequences from this cluster and four from the original RFL genome scan and the third cluster four from the RFL cluster and one from the genome scan.

**Synteny Analysis**

In order to identify the genome position of RFL generation in perennial ryegrass, a comparative genomics approach based on the Genome Zipper (Pfeifer et al. 2013) was applied. The RFL-rich zones from species with available genomic
information were searched for conserved synteny with the genomes of other related plant species. The comparative genomics tool available at Ensemble Plants (http://plants.ensembl.org/index.html) was used to discern if any synteny exists between the RFL-rich genomic regions, from different species, within a single clade. This revealed no synteny between any RFL-rich regions both within each clade and between clades.

**Discussion**

A bioinformatics pipeline targeting candidate Rf genes for CMS was successfully established and identified three clusters of possible RFL generation in perennial ryegrass. This pipeline, consisting of three complementing steps (supplementary fig. S1, Supplementary Material online), is based on genomic sequence data and thus can be used in any plant species for which such data are available. Validation of the pipeline in fully sequenced grass species such as rice, Brachypodium, and sorghum revealed that 50–90% of candidate RFL genes are found within no more than three genomic regions consisting of 0.1–0.01% of the genome. A similar approach could now be applied to cereals, where efficient access to Rf genes is an integral part of CMS-based hybrid production (Whitford et al. 2013).

**FIG. 3.**—Phylogenetic tree showing the four identified clades within the RFL cluster containing a total of 154 proteins originating from 14 species. Colors are species-specific with abbreviated species names (see table 1) shown along with the number of RFLs present.
RFLs not only identifies candidate PPR genes but also enables efficient identification of dynamic orthologous clustering of multiple species, to identify RFL genes from translated CDS. The second step involves orthologous clustering and sequence comparisons to identify PPR and RFL proteins from different species, but in contrast to RFLs, most of these clusters contained only one orthologous PPR gene per species.

Although the orthologous clustering and phylogenetic approach is an effective method to identify regions of active RFL generation, it was unable to identify all PPRs and was also less successful at identifying RFL genes, from perennial ryegrass, than the genome scanning approach. The effectiveness of the orthologous clustering and phylogenetic approach is dependent upon the type and quality of the input data. The type of data used is important as genomic sequence information may contain a more complete set of RFL genes than transcriptome data because of the tissue- and time-specific expression of RFL genes (Prasad et al. 2003). This is highlighted by the Italian ryegrass and meadow rescue transcriptomes, comprising relatively few RFLs. On the other hand, due to this tissue- and time-specific expression, transcriptome data could also be used to enrich for Rf genes by sampling from tissues known to be expressing Rf proteins, such as anthers (Kazama and Toriyama 2014). Although the use of genomic sequence data is preferable, individual RFLs can still be overlooked by orthologous clustering if they are poorly annotated or pseudogenes. Moreover, using incomplete genome assemblies as input data may not reveal all RFL clusters as they can be difficult to assemble, due to the repetitive features of RFL-rich genomic regions (Tsai et al. 2010). This was observed in barley, where an RFL was identified on an unordered contig from the same chromosome 6HS containing a recently mapped Rf locus that could not be associated with an RFL cluster (Ui et al. 2015). In cereals, further functional restorer loci have been described in wheat (Ma et al. 1995) and rye (Secale)

Table 1
The Number of RFL Genes in Each Clade as well as the Total Number of RFLs Identified Are Given for Each Species

| Species                  | Number of Sequences | Totals |
|--------------------------|---------------------|--------|
|                          | Clade 1  | Clade 2 | Clade 3 | Clade 4 |        |
| Perennial ryegrass        | —       | 25      | —       | —       | 25     |
| Wild einkorn Wheat        | 4       | 9       | —       | 7       | 20     |
| Barley                   | —       | 6       | —       | 1       | 7      |
| Foxtail millet            | 8       | —       | 5       | 13      |
| Rice                     | —       | 2       | 10      |
| Sorghum                  | 1       | —       | 17      |
| Maize                    | 7       | —       | 2       |
| Brachypodium             | —       | 9       |
| Arabidopsis              | —       | 28      |
| Totals                   | 20      | 51      | 32      |

In Silico Identification of Candidate Genes

PPR and RFL Genes in Perennial Ryegrass

In the draft genome sequence of perennial ryegrass (Byrne et al. 2015), 373 PPR genes were identified and classified, revealing 25 RFLs. The number of RFLs identified here is consistent with other studies that have reported 10–30 RFLs per genome (Fujii et al. 2011), for example in Arabidopsis (Lurin et al. 2004). These RFLs have, on average, six more PPR domains than non-RFL PPR proteins. This possibly indicates that in perennial ryegrass, RFLs have a higher RNA sequence specificity than other PPR proteins. This was expected as known PPR-Rf proteins bind to a specific mRNA sequence whereas other PPRs have been shown to bind to multiple mRNAs (Zehrmann et al. 2009). Further evidence for multiple binding specificities comes from the number of transcript editing sites being present in mitochondrial genomes compared with PPRs with editing domains. The Arabidopsis mitochondrial genome encompasses 441 cytosine to uracil editing sites, although only 193 PPR genes, containing the E domain required for transcript editing, can be found in the nuclear genome (Giege and Brennicke 1999; Lurin et al. 2004). It appears that RFL proteins, unlike some other PPR proteins, are highly specialized, targeting a single transcript within the mitochondria (Barkan and Small 2014).
The use of sequence data from restoring individuals, in conjunction with the pipeline described here could help to identify candidate Rf-PPR genes within the identified regions.

Genome Regions of Active RFL Gene Generation

To identify the likely location of any PPR-Rf genes, a phylogenetic approach was applied to find clusters of highly similar RFL genes within single species, allowing the genomic regions of RFL generation to be distinguished. By comparing species with genome location information to perennial ryegrass, three regions of possible RFL generation were identified. Through further phylogenetic analysis of RFLs from several species, the fine structure of RFL organization in grasses was resolved and regions of novel RFL generation in species with positional genome information identified. This understanding of the architecture of RFL genes within other grass species led to the identification of similar groups of RFL genes in perennial ryegrass. Given the phylogenetic similarities between these groups, we can confidently assume that each of these groups of RFL genes in perennial ryegrass will be represented.

Fig. 4.—Phylogenetic trees generated using protein sequences from the OrthoMCL generated RFL cluster. (A) Clade 1. (B) Clade 2. (C) Clade 3. (D) Clade 4. Colors represent clusters of sequences originating from the same genomic region of that species or showing a similar arrangement when genome information is not available.

cereal L.) (Hackauf et al. 2012). The use of sequence data from restoring individuals, in conjunction with the pipeline described here could help to identify candidate Rf-PPR genes within the identified regions.
at single loci within the genome. These loci could be elucidated with more detailed genomic information or the use of a mapping population for genetic linkage mapping. Wild einkorn wheat, another species without genome location information, also showed a similar pattern with three tight clusters indicating the likelihood of three RFL generation loci.

The rate of recombination within the mitochondrial genome, which is the source of novel CMS mechanisms, is high (Kubo et al. 2011; Sloan et al. 2012), requiring a relatively rapid generation of new functional genes through recombination driven diversifying selection (Fuji et al. 2011). The likelihood of functional PPR-Rf genes being present in these zones of active RFL generation is a function of how long it takes for fertility restoration to become fixed within a population (the time it takes for an Rf gene to restore CMS in an entire population) and the rate at which RFL genes are shuffled throughout the genome (how long a newly functional Rf gene is likely to stay within the genome region of active RFL generation). This suggests that if the rate of fixation is faster than the rate of shuffling, Rf genes will always be found within these RFL clusters. This is further borne out by the genome synteny results, showing a breakdown of synteny in the region of RFL generation zones, indicating that novel RFL generation occurs faster than speciation, unlike other PPR genes that are highly conserved between species. Similar findings were reported for barley and rye where Rf containing regions showed synteny to regions from rice, Brachypodium, and sorghum that contained no RFLs (Hacka et al. 2012; Ui et al. 2015). These results indicate not only that RFLs are being shuffled around the genome at a rate faster than that of speciation but also that they are being rapidly lost when nonfunctional (Dahan and Mireau 2013).

In the four clades identified within the RFL cluster, all the dicot RFL genes fell within a single clade, representing the split between monocots and dicots. Although the dicot sequences were in a separate clade, the fact that RFLs from both monocot and dicot species were identified within a single cluster based on orthologous clustering is consistent with the hypothesis that monocot and dicot RFL genes share a common ancestor. This also suggests that this common ancestor is distinct from all other PPR genes and predates the monocot/dicot split, meaning that RFL genes evolved before this split (O’Toole et al. 2008).

**Table 2**
Details of the Species-Specific Genomic Regions Having a High Density of RFL Genes

| Region of High RFL Density | No. of Genes Present in Cluster |
|----------------------------|---------------------------------|
| **Clade** | **Species** | **Genome Location (bp)** | **Size (kb)** |
| 1 | *Se. italica* | Ch8:29882484—31204264 | 1,322 | 5 |
| 2 | *Z. mays* | Ch2:22771686—22863247 | 917 | 6 |
| 3 | *Se. italica* | Ch7:15683154—15692828 | 9 | 2 |
| 4 | *O. sativa* | Ch4:16884906—16757223 | 9 | 2 |
| 5 | *H. vulgare* | Ch1:47176692—50263441 | 3,087 | 3 |
| 6 | *B. distachyon* | Ch2:38479458—39012768 | 533 | 7 |
| 7 | *A. thalians* | Ch1:4183066—4335929 | 172 | 4 |
| 8 | *A. thalians* | Ch1:23176930—23988740 | 812 | 17 |
| 9 | *S. bicolor* | Ch2:5169697—5744703 | 575 | 3 |
| 10 | *Z. mays* | Ch8:76606724—76690742 | 84 | 2 |
| 11 | *S. bicolor* | Ch5:2222303—2776884 | 554 | 9 |
| 12 | *O. sativa* | Ch10:18823675—19143586 | 320 | 6 |
| 13 | *O. sativa* | Ch3:2874091—383986 | 10 | 2 |

**Accuracy and Usefulness of This Approach**

The approach presented here allows efficient targeting of RFL containing genomic region(s) in multiple species. These regions have previously been shown to contain RF-PPR genes (Bentolila et al. 2002; Kazama et al. 2008; Uyttevwaal et al. 2008; Barr and Fishman 2010; Jo et al. 2010; Jordan et al. 2011; Kazama et al. 2014; Bish et al. 2015). In grasses, examples can be found in maize with the Rf8 locus mapping to an RFL cluster on chromosome 2 (Meyer et al. 2011), and in rice with the Rf1 (Wang et al. 2006) and Rf4 (Luo et al. 2013) genes being present within the RFL cluster of rice chromosome 10. The most recent example is the Rf6 restorer in rice (Huang et al. 2015). Rf6 was mapped to a 200 kb region on rice chromosome 8 which contains three RFLs identified in this study with one of these genes (Os08g01870) being located within 15 kb of the marker shown to be cosegregating with the restorer gene (Huang et al. 2012). The only identified PPR-Rf gene that is located outside of the RFL-rich regions is Rf1 from sorghum. The Rf1 locus, most likely encoded by PPR13, is located as a single PPR-Rf gene on chromosome 8 although PPR13 was not cloned from a restoring genotype (Klein et al. 2006). PPR13 is different in its structure from all other identified RFL-Rf genes as it is of the PLS subtype and contains domains linked with RNA editing, indicating that the mechanism for restoration of the CMS phenotype may also be unique (Schmitz-Linneveber and Small 2008; Dahan and Mireau 2013). PPR13 also exemplifies the complementarity of protein domain profile matrix scans and orthologous clustering, the latter of which would have been unable to detect a gene like PPR13.

The clustering approach assumes that newly functional PPR-Rf genes are the result of recombination events within an RFL genomic cluster and not an existing RFL that has gained a restoring function through the serendipitous recognition of a novel CMS causing transcript within the mitochondria. This balance will most likely differ between species and between populations of the same species under differing environmental conditions. It is important to note here that this approach will be most successful in identifying PPR-Rf genes in naturally occurring CMS systems (where the rapid evolution of RFLs has had time to overcome the damage in the
mitochondria), but will also find traction in induced CMS systems where the CMS phenotype still has a mitochondrial ORF as its source and as such a possible PPR-Rf gene as a restorer.

The Value of Rf Genes for CMS-Based Pollination Control in Forage Grasses

This pipeline provides an efficient first approach for Rf gene identification as it permits researchers to target the most likely genomic regions to contain Rf genes. Rapid identification of Rf candidate RFL genes will facilitate the development of functional markers for restoration of fertility, enabling efficient exploitation of CMS as a tool to control pollination for hybrid breeding in forage grasses. However, fertile hybrid seed is not necessarily needed for temporary forage production as biomass and not seed is the primary yield target (Islam et al. 2014). Indeed, it is often unwelcome as any partial or full restoration of male fertility during hybrid seed production would decrease the purity and value of that seed. Nevertheless, Rf gene identification is important to ensure that markers can be designed and populations screened to prevent unwanted fertility restoration. This will help to overcome the main challenge in outbreeding forage grasses with highly heterozygous genomes which is the maintenance of the CMS trait. The ability to rapidly identify individuals carrying an Rf gene within a breeding population would assist breeders in maintaining the commercially important CMS phenotype as well as ensuring hybrid seed purity. For breeding purposes, the exact position of the Rf gene does not need to be identified as genetic markers tightly linked to the functional Rf gene might be sufficient to identify restoring phenotypes. The approach used in this study can provide this by identifying RFL clusters within the genome allowing the relatively rapid identification of useful markers. Further dissection of RFL clusters, possible through BAC library screen and subsequent BAC clone sequencing, would allow the identification and cloning of the responsible Rf gene.

Conclusion

Here, we have designed and implemented an in silico pipeline to identify candidate Rf-PPR genes and demonstrated its effectiveness by pinpointing known Rf genes. This study focused on perennial ryegrass and identified three regions of active RFL generation, providing excellent targets for marker development and future mapping approaches. Information is also provided for other species such as wild einkorn wheat, showing the wider applications of this method. As demonstrated, this pipeline can also be used to characterize RFLs in both monocots and dicots, to provide new insights into their evolution. The predictive power of this approach will improve as more genome sequence data becomes available. Knowledge of RFL-rich genomic regions within a genome might also be used for targeted sequencing of such regions in restorer plants and facilitate the expedient determination of Rf genes, the knowledge of which would not only be useful for breeding programs but also for fundamental research into nuclear/mitochondrial interactions.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

This project has received funding from the European Union’s Seventh Framework Program for research, technological development and demonstration under grant agreement no PITN-GA-2013-608422—IDP BRIDGES. The project was also supported by the Swiss National Science Foundation (SNSF Professorship grant no: PPO02P_138988). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors sincerely thank Prof Dr Achim Walter for hosting the Forage Crop Genetics group at ETH Zurich.

Literature Cited

Ahokas H. 1983. Cytoplasmic male sterility in barley. XV. PI 296897 as a restorer of fertility in msml and msml2 cytoplasms. Hereditas 99(1):157–159.

Altschul SF, et al. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Andrés C, Lurin C, Small I. 2007. The multifarious roles of PPR proteins in plant mitochondrial gene expression. Physiol Plant. 129(1):14–22.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 55(4):539–552.

Bailey TL, Gribskov M. 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14(1):48–54.

Barkan A, Small I. 2014. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol. 65:1–23.

Barr CM, Fishman L. 2010. The nuclear component of a cytonuclear hybrid incompatibility in Mimulus maps to a cluster of pentatricopeptide repeat genes. Genetics 184(2):455–466.

Bentolila S, Alfonso AA, Hanson NR. 2002. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A. 99:10887–10892.

Bisht DS, Chamola R, Nath M, Bhat SR. 2015. Molecular mapping of fertility restorer gene of an alloplasmic CMS system in Brassica juncea containing Moricandia arvensis cytoplasm. Mol Breeding. 35(1):1–11.

Byrne S, et al. 2015. A synteny-based draft genome sequence of the forage grass Lolium perenne. The Plant J. 84(4):816–826.

Cannarozzi G, et al. 2014. Genome and transcriptome sequencing identifies breeding targets in the orphan crop teff (Eragrostis tef). BMC Genomics 15:S81.

Castandet B, Araya A. 2011. RNA editing in plant organelles. Why make it easy? Biochemistry (Mosc) 76(8):924–931.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23(2):81–90.

Chase CD, Babay-Laughnan S. 2004. Cytoplasmic male sterility and fertility restoration by nuclear genes. Molecular biology and biotechnology of plant organelles. The Netherlands: Springer. p. 593–622.
Chateigner-Boutin AL, Small I. 2010. Plant RNA editing. RNA Biol. 7(2):213–219.
Chevenet F, et al. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7(1):439.
Colcombet J, et al. 2013. Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA Biol. 10(9):1557–1575.
Connoly V, Wright-Turner R. 1984. Induction of cytoplasmic male-sterility into ryegrass (Lolium perenne). Theor Appl Genet. 68(5):449–453.
Cremersmolenaaar J, Hall RD, Krens FA. 1992. Asymmetric protoplast fusion aimed at intraspecific transfer of cytoplasmic male sterility (CMS) in Loliun perenne L. Theor Appl Genet. 84(5-6):763–770.
Cui XQ, Wise RP, Schnable PS. 1996. The rice nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272(5266):1334–1336.
Dahan J, Mireau H. 2013. The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA Biol. 10(9):1469–1476.
Dereeper A, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:465–469.
Desloire S, et al. 2003. Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep. 4(6):588–594.
Duvick DN. 2001. Biotechnology in the 1930s: the development of hybrid maize. Nat Rev Genet. 2(1):69–74.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.
Enright AJ, van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30(7):1575–1584.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:29–37.
Flicek P, et al. 2012. Ensembl 2012. Nucleic Acids Res. 40(1):84–90.
Fuji S, Bond CS, Small I. 2011. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci U S A. 108(4):1723–1728.
Fuji S, Small I. 2011. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol. 191(1):37–47.
Geddy R, Brown GG. 2007. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics 8:13.
Giege P, Brennicke A. 1999. RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. Proc Natl Acad Sci U S A. 96(26):15324–15329.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52(3):696–704.
Hackauf B, Korzun V, Wortmann H, Wilde P, Whehling P. 2012. Development of conserved ortholog set markers linked to the restorer gene Rf1 in rye. Mol Breed. 30:1507–1518.
Hammani K, et al. 2009. A study of new arabidopsis chloroplast rna editing sites. Plant J. 55:619–628.
Kazama T, Nakamura T, Watanabe M, Sugita M, Toriyama K. 2008. Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J. 55:351–362.
Kazama T, Toriyama K. 2014. A fertility restorer gene, Rf4, widely used for hybrid rice breeding encodes a pentatricopeptide repeat protein. Rice 7(1):1–5.
Kiang AS, et al. 1993. Cytoplasmic male sterility (CMS) in Lolium perenne L.: 1. Development of a diagnostic probe for the male-sterile cytoplasm. Theor Appl Genet. 86(6):781–787.
Kiang AS, Kavanagh TA. 1996. Cytoplasmic male sterility (CMS) in Lolium perenne L. The mitochondrial genome of a CMS line is rearranged and contains a chimaeric atp9 gene. Theor Appl Genet. 92(3-4):308–315.
Kitazaki K, et al. 2015. Post-translational mechanisms are associated with fertility restoration of cytoplasmic male sterility in sugar beet (Beta vulgaris). Plant J. 83(2):290–299.
Klein RR, et al. 2006. Fertility restorer locus Rf1 of sorghum (Sorghum bicolor L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor Appl Genet. 112(2):388–388.
Kubo T, et al. 2011. Male sterility-inducing mitochondrial genomes: how do they differ? Crit Rev Plant Sci. 30(4):378–400.
Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178–2189.
Li SB, et al. 2012. Phylogenetic genomewide comparisons of the pentatricopeptide repeat gene family in indica and japonica rice. Biochim Genet. 50(11-12):978–989.
Li WZ, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658–1659.
Longin CFH, et al. 2012. Hybrid breeding in autogamous cereals. Theor Appl Genet. 125(6):1087–1096.
Luo D, et al. 2013. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat Genet. 45(5):573–U157.
Lurin C, et al. 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16(8):2089–2103.
Ma ZQ, Zhao YH, Sorrells ME. 1995. Inheritance and chromosomal locations of male fertility restoring gene transferred from Aegilops umbellulata Zhuk. to Triticum aestivum L. Mol Genet. 247:351–357.
Martin AC, et al. 2009. Chromosome engineering in wheat to restore male fertility in the msh1 CMS system. Mol Breed. 24(4):397–408.
McDermott P, Connolly V, Kavanagh T. 2008. The mitochondrial genome of a cytoplasmic male sterile line of perennial ryegrass (Lolium perenne L.) contains an integrated linear plasmid-like element. Theor Appl Genet. 117(3):459–470.
McLaren W, et al. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26(16):2069–2070.
Melchinger AE. 2010. The International Conference on “Heterosis in Plants”. Theor Appl Genet. 120(2):201–203.

Meyer J, Pei D, Wise RP. 2011. RfB-Mediated T-Transcript accumulation coincides with a pentatricopeptide repeat cluster on maize chromosome 2L. Plant Genome 4(3):283–299.

Min XJ, et al. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33:677–680.

O’Toole N, et al. 2008. On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol. 25(6):1120–1128.

Okuda K, et al. 2009. Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in arabidopsis chloroplasts. Plant Cell 21(1):146–156.

Okuda K, Shikanai T. 2012. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. Nucleic Acids Res. 40(11):5052–5064.

Pembleton LW, et al. 2015. Design of an F1 hybrid breeding strategy for ryegrasses based on selection of self-incompatibility locus-specific alleles. Front Plant Sci. 6:764–764.

Peng Z, et al. 2013. The draft genome of the fast-growing non-timber forest species moso bamboo (Phyllostachys heterocycla). Nat Genet. 45(4):456–461.

Pfeifer M, et al. 2013. The perennial ryegrass GenomeZipper: targeted use of genome resources for comparative grass genomics. Plant Physiol. 161(2):571–582.

Prasad K, Kushalappa K, Vijayraghavan U. 2003. Mechanism underlying regulated expression of RFL, a conserved transcription factor, in the tomato. Front Plant Sci. 6:764–764.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology open software suite. Trends in Genet. 16(6):276–277.

Ruttink T, et al. 2010. Orthology Guided Assembly in highly heterozygous crops: creating a reference transcriptome to uncover genetic diversity in Lolium perenne. Plant Biotechnol J. 11(5):605–617.

Schallenberg-Ruedinger M, et al. 2013. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. RNA Biol. 10(9):1549–1556.

Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 13(12):663–670.

Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115.

Schnable PS, Wise RP. 1998. The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci. 3(5):175–180.

Singh SK, Chatrath R, Mishra B. 2010. Perspective of hybrid wheat research: a review. Indian J Agric Sci. 80(12):1013–1027.

Sloan DB, et al. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 10(1):53.

Small I, Rackham O, Filipovska A. 2013. Organelle transcriptomes: products of a deconstructed genome. Curr Opin Microbiol. 16(5):652–658.

Tasaki E, Hattori M, Sugita M. 2010. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmFc transcript. Plant J. 62(4):560–570.

Toda T, et al. 2012. Rice MPR25 encodes a pentatricopeptide repeat protein and is essential for RNA editing of nad5 transcripts in mitochondria. Plant J. 72(3):460–460.

Tsi II, Otto TD, Herriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 11(4):841.

Ui H, et al. 2015. High-resolution genetic mapping and physical map construction for the fertility restorer Rfm1 locus in barley. Theor Appl Genet. 128:283–290.

Uyttewaal M, et al. 2008. Characterization of Raphanus sativus pentatricopeptide repeat sequences encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. Plant Cell 20:3331–3345.

van der Heijden SAG, and Roulund N, editors. 2010. Genetic gain in agromonic value of forage crops and turf: a review. Sustainable use of genetic diversity in forage and turf breeding. p. 247–260.

Virmanni SS, editor. 1994. Hybrid rice technology: new developments and future prospects. Hybrid rice technology: new developments and future prospects. p. 296.

Wang Z, Zon Y, Li X. 2006. Cytoplasmic male sterility of rice with BoroII cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18:676–687.

Whitford R, et al. 2013. Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J Exp Bot. 64(18):5411–5428.

Wilkins PW. 1991. Breeding perennial ryegrass for agriculture. Euphytica 52(3):201–214.

Wit F. 1974. Cytoplasmic male sterility in ryegrass (Lolium spp.) detected after intergenic hybridization. Euphytica 23:31–38.

Yuan LP, Virmanni SS. 1988. Status of hybrid rice research and development. Hybrid rice. Proceedings of an international symposium; 1986 October 6-10; Changsha, China. p. 7–24.

Zehrmann A, et al. 2009. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. Plant Cell 21(2):558–567.