A New and Unified Nomenclature for Male Fertility Restorer (RF) Proteins in Higher Plants

Simeon O. Kotchoni1*, Jose C. Jimenez-Lopez2*, Emma W. Gachomo3, Manfredo J. Seufferheld4

1 Department of Agronomy, Purdue University, Lilly Hall, West Lafayette, Indiana, United States of America, 2 Department of Biochemistry, Cell and Molecular Biology of Plants, Estacion Experimental del Zaidin (EEZ), Consejo Superior de Investigaciones Cientificas (CSIC), Granada, Spain, 3 Department of Botany and Plant Pathology, Purdue University, Lilly Hall, West Lafayette, Indiana, United States of America, 4 Department of Crop Science, University of Illinois U-C, Urbana-Champaign, Illinois, United States of America

Abstract

The male fertility restorer (RF) proteins belong to extended protein families associated with the cytoplasmic male sterility in higher plants. Up till now, there is no devised nomenclature for naming the RF proteins. The systematic sequencing of new plant species in recent years has uncovered the existence of several novel RF genes and their encoded proteins. Their naming has been simply arbitrary and could not be adequately handled in the context of comparative functional genomics. We propose in this study a unified nomenclature for the RF extended protein families across all plant species. This new and unified nomenclature relies upon previously developed nomenclature for the first ever characterized RF gene, RF2A/ALDH2B2, a member of ALDH gene superfamily, and adheres to the guidelines issued by the ALDH Genome Nomenclature Committees. The proposed nomenclature reveals that RF gene superfamily encodes currently members of 51 families. This unified nomenclature accommodates functional RF genes and pseudogenes, and offers the flexibility needed to incorporate additional RFs as they become available in future. In addition, we provide a phylogenetic relationship between the RF extended families and use computational protein modeling to demonstrate the high divergence of RF functional specializations through specific structural features of selected members of RF superfamily.

Introduction

Cytoplasmic male sterility (CMS) is a maternally inherited trait observed in numerous plant species, resulting in the formation of non-functional microspores or pollen grains [1,2]. The most pronounced cytological events accompanying CMS concern the tapetum tissue surrounding the differentiating pollen mother cells (PMC), which involve its abnormal vacuolization, fusion of cells into multinucleate syncytia, and disturbances in the time of the programmed tapetum death [3,4]. Development of PMC, is arrested either during meiosis or in postmeiotic phase, and is usually related to the failure in the deposition of the microspore (pollen) wall [1]. Mitochondrial function depends on the coordinate action of nuclear and mitochondrial genomes. CMS is generally determined by mitochondrial genomes. The regions whose expression is associated with CMS contain unusual ORFs that are often chimeric in structure and frequently co-transcribed with conventional mitochondrial genes [2].

In cells, nuclear genes called restorers of fertility (RF) have the ability to suppress the male-sterile phenotype and, hence, restore the production of pollen to plants carrying the deleterious mitochondrial genome. CMS RF systems greatly facilitate hybrid seed production by eliminating the need for tedious hand emasculation and ensuring that each seed is a result of cross-pollination [5]. The RF allele from the pollen parent therefore restores fertility and seed production in the heterotic hybrid progeny. Apart from its commercial exploitation, CMS offers one of the few opportunities to examine the regulation of mitochondrial gene expression by a nuclear gene in multicellular organisms. Up to date, the mechanism by which CMS causes male sterility in higher plants is not fully known, and the functional features of male sterility restorer proteins, RFs, is completely unknown. In this study, we exclusively focused our attention on the RF extended gene families in higher plants. The study of nature and origin of genes that determine CMS, have provided new insights into plant mitochondrial-nuclear communication. This study has revealed the implication of mitochondrial signaling pathways, including those involved in regulating cell death and nuclear gene expression [6]. Generally, the nuclear RF genes encode pentatricopeptide-repeat (PPR) proteins as key regulators of plant mitochondrial gene expression [5]. However, in maize, the sterility restorer gene, RF2, which acts in conjunction with the RF1 gene to restore fertility to T-cytoplasm maize, is an unusual restorer gene, and is the only one that has been well characterized [7]. Rather than affecting the expression of the CMS protein (URF13), the RF2 is an aldehyde dehydrogenase [8] that acts by compensating for a metabolic defect caused by the low levels of URF13 protein. However, it is the presence of RF1 that is responsible for reduction of the toxic protein (T-URF13) [9] and the alteration of the T-URF13 transcript profile [10,11]. In other words, the RF proteins

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: skotchon@purdue.edu

Copyright: © 2010 Kotchoni et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published: December 28, 2010

Citation: Kotchoni SO, Jimenez-Lopez JC, Gachomo EW, Seufferheld MJ (2010) A New and Unified Nomenclature for Male Fertility Restorer (RF) Proteins in Higher Plants. PLoS ONE 5(12): e15906. doi:10.1371/journal.pone.0015906

Received September 14, 2010; Accepted November 28, 2010; Published December 28, 2010

Copyright: © 2010 Kotchoni et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Supporting Information

The data presented in this study are available as part of the manuscript and do not require further support information.
are able to suppress mitochondrial abnormalities associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation. In many instances, the suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.

In many instances, suppression is directly associated with male sterility. This suppression allows for normal metabolic processes leading to normal male reproductive organogenesis, successful microsporogenesis, pollen development and maturation.
A Unified Naming System for RF Proteins

Table 1. Molecular consensus defining fertility the restorer protein structural patterns.

| Molecular consensus | Name of sequence consensus | Code of consensus |
|---------------------|-----------------------------|------------------|
| [LIVMFGA] - E - [LIMSTAC] - [GS] - G - [KNLNM] - [SADN] - [TAPVV] | ALDH_GLU_active site | PS00687 |
| [FYLV] - x - (GVEP) - (DILV) - G - [OE] - (LPGY) - C - [LIVMGSTANC] - (AGCN) - (HE) - [GSTADEKR] | ALDH_CYS_active site | PS00070 |
| Arg repeat regions | Pentatricopeptide (PPR) repeat profile | PS51375 |
| Pro repeat regions | Arginine-rich region profile | PS50323 |
| (AG) - x(4) - G - K - [ST] | Proline-rich region profile | PS50099 |
| L - x(6) - L - x(6) - L - x(6) - L | TIR domain | PS50104 |
| Folyopolyglutamate synthase signature 1 | UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (Mur ligase) | PS51011 |
| [LIVMFY] - x - [LIVM] - [STAG] - G - T - [NK] - G - K - x - [STG] - x(4) - (A) - x - (EAD) - [LIVM](2) - x(3,4) - [GSKQT] | UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (Mur ligase) | PS51012 |
| Folyopolyglutamate synthase signature 2 | Kinesin motor domain | PS50067 |
| [GSAT] - [KRHPSTQGME] - [LIVMF] - x - [LIVM] - [IVC] - [DN] - [LS] - [AH] - G - [SAN] - E | Neutral zinc metallopeptidases, zinc-binding region signature | PS50142 |
| [GSTA][LI][VNW] - (PCR) - [KND] - H - E - [LIVMFYW] - [DEHRS]P - H - (EKP) - [LIVMFYWGSPO] | Myc-type, “helix-loop-helix” domain profile | PS50888 |
| Amphipathic helix 1 | Ginkbilobin-2 (Gnk2)-homologous domain profile | PS51473 |
| C-X-(I)-(C)-X2(-C) | Cysteine residues form three intramolecular disulfide bridges: C1-C5, C2-C3, and C4-C6 | PS51473 |
| [STAG] - [T] - [STAG] - [LIVMF] - R - L - (LP) - [SAGV] - N - [LIVMT] | ATP synthase a subunit | PS00449 |
| [GSTA] - R - [NQ] - P - x(5) - [A] - x - [F] - x(2) - [LIVMFYW](2) - x(3) - [LIVMFW] - x - [DE] | ATP synthase c subunit | PS00605 |
| P - [SAF] - [LIV] - [DNH] - [LKGK] - (F) - (S) - S - (DCF) - S | ATP synthase alpha/beta subunits | PS00152 |
| doi:10.1371/journal.pone.0015906.t001 |

mays (Table 2, Table S2). At this time, more than half of the catalogued RF families encode members of single gene most of which represent the PPR repeat RF proteins and other less characterized functional domains (Table 1, Table 2, Table S2).

The total number of genes in the RF superfamily is expected to increase steadily with time, mainly due to the genomic sequencing of additional species. Regardless of the plethora of RF genes yet to be identified/characterized, their classification and relationship to the entire extended RF gene superfamily will be easy owing to this nomenclature building block that catalogues newly identified/characterized RF gene products only on the basis of sequence similarity to previously characterized RF gene products.

Phylogenetic analysis of the extended RF protein families

The retrieved full-length RF-related sequences were aligned to determine phylogenetic relationships within the male sterility restorer (RF) extended family. A phylogenetic tree of the RF extended sequences is depicted in Figure 1. The phylogenetic tree shows that the 51 RF extended families, although highly divergent, are split into three clades, with clades 1 and 2 representing mainly members of the PPR repeat RF proteins with the exception of family 16 RFs. Clade 3 represents uniquely members of ALDH proteins and the highly variable RF proteins distantly related to the PPR repeat proteins, but clustering together with the ALDH-RF proteins (Figure 1). The evolutionary relationships reveal some interesting observations.

Family 1 exclusively represented by the ALDHs are male sterility restorers of monocots such as maize and rice, while the other RFs including the PPR repeat RF proteins are generally sterility restorers of other higher plant species (Figure 1). However, some members of family 2, 8–15, and family 32 PPR repeat RF proteins have also been identified in maize, rice (Table 2, Table S2).

RF protein superfamilies: Structural and conformational variability

The crystallographic structural coordinates of relatively few RFs have been deposited in the Protein Database (PDB) so far. To our knowledge, detailed comparative studies of structural and conformational features of members of the RF extended protein families have not been performed in higher plants. Using computational modeling analysis, we here determined for the first time the structural features and uniqueness of the 3D structures of selected members of the RF extended families. We wanted to appreciate in detail the structural divergence of the RFs mediating various functional specificities. Each sequence was modeled based on the ten best structural templates using the structural parameters summarized in Table 3.

A general structural comparison (Figure 2) and phylogenetic analysis (Figure 1) provided a clearer and unexpected insight into the structural divergence of the RF extended protein families. Our protein modeling data demonstrates the divergence of RF functional specializations highlighted here by very striking structural features of...
the selected members of RF extended protein families (Figure 2). The divergence in the molecular function is reflected by the differences in the structural subunit of the active RFs (Figure 2), i.e. each subunit of the dimeric or tetrameric enzyme ALDH of family 1 for instance. Each subunit of the active ALDHs is characterized by the "Rossmann fold", and contains an NAD-binding domain, a catalytic domain and an oligomerization domain (Figure 2). At the interface of these domains there is a funnel-shaped opening leading to a putative catalytic pocket. In order to fully understand the structural characteristics of "Rossmann-type fold" of the RFs/ALDHs we depicted in Figure 3 the structural features of the RF/ALDH active site and the NAD-binding domain containing the Rassmann-type fold feature. The "Rossmann fold" represents the structural motif found in proteins that bind nucleotides, especially the NAD cofactor.

Table 2. The fertility restorer protein superfamily: new and unified nomenclature.

| RF Family | Revised annotation | Previous annotation | GeneBank acc. number | Protein Acc. number | Molecular pattern(s) | Putative functional characterization | Source |
|-----------|--------------------|---------------------|----------------------|---------------------|----------------------|------------------------------------|--------|
| Family 1  | RF1A1              | RF2A                | AF215823             | Q43274              | PS00687 PS00070      | ALDH (NAD⁺)                        | Zea mays |
| Family 2  | RF2A1              | RF1B                | DQ311054             | Q2PPE6              | PS1375               | PPR repeat                         | Oryza sativa |
| Family 3  | RF3A1              | RF1                 | AP008216             | Q76C20              | PS1375               | PPR repeat                         | Oryza sativa |
| Family 4  | RF4A1              | RF                  | DQ445625             | Q84KB7              | PS1375               | PPR repeat                         | Raphanus sativus |

doi:10.1371/journal.pone.0015906.t002
The Rossmann fold structural feature is composed of three or more parallel beta strands linked by two alpha helices (Figure 3 C–F). Many members of the ALDH protein family possess different NAD-binding modes and catalytic sites, with a mechanism for enzymatic specificity and activity. Members of the pentatricopeptide repeat extended protein families are characterized by tandem repeats of a degenerate 35 amino acid motif that have a structure predicted to fold into a helix-turn-helix, similar to those found in previously characterized PPR proteins [20], and a degenerate 34 amino acid sequence in tandem arrays of 3–16 motifs, which form scaffolds to mediate protein-protein interactions (Figure 2).

**Discussion**

Although many nuclear and mitochondrial genes associated with CMS have been characterized, the identification and characterization of RF genes has proven elusive, and only the maize RF2A, which encodes a class 2 mitochondrial ALDH is well characterized [2,21]. However, orthologs of maize FR2A have been subsequently characterized in rice, and other plant species [19,22].

RF is often associated with genes encoding pentatricopeptide repeat (PPR) proteins [23,24]. PPR proteins constitute a large
family, with more than 400 members in Arabidopsis, rice, maize, petunia, and Raphanus that are thought to be RNA binding proteins involved in posttranscriptional processes (RNA processing and translation) in mitochondria and chloroplasts [24]. Up to now, the RF gene extended families are deposited into the databases with arbitrary naming system by authors. This arbitrary

| Family | Accession number | Specie | Protein previous name | Template | Identity (%) |
|--------|------------------|--------|-----------------------|----------|--------------|
| RF1A2  | Q94G64           | Z. mays| RF-2                  | 1o01     | 63           |
| RF12A1 | B6USK0           | Z. mays| Fertility restorer    | 1w3dbA   | 15           |
| RF5A1  | C4WRH2           | R. sativus| AGD                  | 1e8cb    | 28           |
| RF49A1 | D0R6K2           | R. sativus| MOS-2                | 2ckkA    | 23           |
| RF1A4  | Q9LLR2           | O. sativa| Aldehyde dehydrogenase| 1agB    | 61           |
| RF1A4  | A7J144           | O. sativa| Fertility restorer    | 2q7fA    | 17           |

AGD: UDP-N-acetylglucosamine:lipopolysaccharide-2-6-diaminoligase.
doi:10.1371/journal.pone.0015906.t003

Figure 2. Three-dimensional structure analysis of selected members of cytoplasmic male sterility restorer proteins. The model proteins are depicted as cartoon diagrams. The secondary elements of the crystallographic structures are rainbow colored, with N-terminus in blue, and C-terminus in red.
doi:10.1371/journal.pone.0015906.g002
nomenclature is not sustainable for adequate comparative mega-functional genomics studies, especially as the numbers of RF genes have increased steadily with the completion of more plant genome sequences. With the increase in genome sequencing of novel plant species, there are currently more than 800 genes encoding for RF or RF like proteins in plants. 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 [24]. PPR- and TPR-motifs are found in helical-repeat proteins and would be predicted to have protein-binding properties. However, our data only revealed ~100 RFs because we focused on RF genes that are indeed characterized as restorers of male sterility. These RF genes encode members of 51 gene families with Family 1 representing exclusively the aldehyde dehydrogenase cluster. The highly expanded PPR repeat RF genes encode for more than half (28

Figure 3. Rossmann-type fold of RF/ALDH proteins. (A, B) General structure (cartoon diagram) showing the secondary structural elements of RF1A2 (A) and RF1A4 (B). (C–F) Overview of the “Rossmann-type fold” structural features at coenzyme-binding domain and catalytic domain are depicted for RF1A2 (C, D) and RF1A4 (E, F). (G–J) Protein surface representation of RF1A2 (G) and RF1A4 (I). (H–J) Magnification overview of the coenzyme binding cavity of RF1A2 (H) and RF1A4 (J). (K–N) Electrostatic surface potential view of RF1A2 (K) and RF1A4 (M). Magnification overview of the electrostatic surface potential of the coenzyme binding cavity of RF1A2 (L) and RF1A4 (N). The surface colours are clamped at red (−10) or blue (+10).

doi:10.1371/journal.pone.0015906.g003
families) of the entire 51 family (Figure 1). With the exception of family 1, we know very little about the functions of members of other RF protein families. The PPR proteins have been hypothesized to function as sequence-specific adaptors for a variety of other RNA-associated proteins [23]. This was supported by the fact that the maize PPR protein CRP1 influences expression of chloroplast genes through association with specific mRNAs [36], and that the fact that PPR proteins are involved in mRNA editing in chloroplasts [27]. In addition, the first RF gene identified in petunia, encoding the PPR protein RF-PPR592, was suggested to have an mRNA processing function [5]. Our data reveals that about half of RF superfamily does not belong to the PPR repeat protein group (Table 2, Figure 2, Table S2). This new unified nomenclature provides essential inventories for comparative genomic analyses of the RF superfamily in flowering plants and the grass species. From the data, it appears that plant RFs have undergone functional specialization over time. Family 1 RF proteins are major fertility restorers of type T-cytoplasmic male sterility (T-CMS) in monocots, especially in maize and rice [28], while the cluster of PPR repeat protein families restore fertility to type S- and BT-CMS in various plant species [29,30]. Several CMS/restorer systems are defined by the different origins of CMS with distinct genetic features. For instance, the BT-CMS (Boro II-type of CMS), WA-CMS (wild abortive-type of CMS) and HL-CMS (Honglian-type of CMS) are CMS types of rice, while the S-CMS (severely affected-type of CMS) and T-CMS (Texas-type of CMS) are of maize genotype, which arose spontaneously in a breeding line, and PET1-CMS of sunflower arose from an inter-specific cross between Helianthus petiolaris and H. annuus.

Our developed unified nomenclature system is helpful in a quick functional prediction of any newly cloned RF gene(s), because from the nomenclature point of view, the newly cloned gene(s) will always be characterized/named with sequence similarity with previously characterized RF genes/proteins. This modified and unifying nomenclature preserves also the widely arbitrary naming system used so far and referenced this old naming designation since the new name is linked to the gene accession number that will automatically pulled out the old naming system from the databases. The changes that have been introduced reflect into which extended family or subfamily a certain RF protein belongs. Accordingly, the new nomenclature will have no significant impact on already published data with old/arbitrary naming system. However, we urge scientists working on RFs to adopt this new and easy nomenclature system. In this regard, we have made an effort to preserve the user friendly linkage between the old and the new designations, which we hope will help researchers to adapt the new names. As the revised nomenclature should facilitate communication and understanding within the community interested in RFs, we advocate that this new naming system be used in all future studies.

Materials and Methods

Database search for RF genes

Restorer of fertility (RF) and RF-like gene sequences were retrieved from the US National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) genome, the rice (TIGR Rice Annotation Release 4, http://tigrblast.tigr.org/ckblast/index.cgi?project=osal), the maize (http://www.maizesequence.org) genome databases, and from the non-redundant expressed sequence tag (EST) databases using BLASTX, BLASTN and BLAST (low complexity filter, Blossum62 substitution matrix) [31]. The searches were conducted using previously characterized maize RF2A (GenBank Accession number AF215823), rice RF1A (GenBank Accession number DQ811052), rice RF1B (GenBank Accession number DQ311054), Brassica PPR-B-L1 (GenBank Accession number FJ455059) and Raphanus Murligase (putative UDP-N-acetylglucosaminidase-1-2-6-diaminoligase) (GenBank Accession number AJ500021). Full-length amino acid sequences for fertility restorer proteins were compiled and aligned using ClustalW [32]. Genetic distances between pairs of amino acid sequences were calculated with Bioedit V7.0.5.3 [33]. Consensus protein sequences were derived from these original alignment, and further analyzed for the presence of putative functional motifs using the PROSITE database [34,35], of biologically meaningful motif descriptions derived from multiple alignments and the ScanProsite program [36], from the Expert Protein Analysis System (ExPaSy) procomics server of the Swiss Institute of Bioinformatics [37]. Finally, the consensus protein sequences (Table 1) were submitted to BLASTP analysis to identify homologous proteins from other plant species. A comparative search for restorer protein homologous was performed using Uniprot database to confirm the identity of the retrieved RF proteins [38].

Restorer of Fertility (RF) proteins: Revised/unified nomenclature

In order to provide a revised and unified nomenclature for RF gene superfamily, we developed a sequence based similarity approach to classify all the retrieved sequences using previously developed gene nomenclature model [15–19]. The criteria for cataloguing the RF protein superfamily was based on the established nomenclature criteria for cataloguing aldehyde dehydrogenase (ALDH) gene superfamily [15–19]; because ALDH (ALDH2B2/RF2A) was identified as the first ever characterized plant RF gene, which was cloned from maize [7]. For this new nomenclature, RF protein sequences that are more than 40% identical to previously identified RF sequences compose a family, and sequences more than 60% identical within a family, compose a gene subfamily. Protein sequences that are less than 40% identical would describe a new RF gene family (Table S1). Taking maize RF1A (previous name P22A) as an example for the revised nomenclature, RF indicates the root; the first digit (1) indicates a family and the first letter (A) a subfamily, while the final number (1) identifies an individual gene within a subfamily. The revised nomenclature is therefore composed of an assigned gene symbol (RF) (abbreviated gene name) for the whole gene superfamily. The gene symbol must be (i) unique and representative of the gene superfamily; (ii) contain only Latin letters and/or Arabic numerals, (iii) not contain punctuation, and (iv) without any reference to species. These newly developed criteria have been applied to database curators to generate the unified RF gene families/classes regardless of the source of the cloned gene(s).

Sequence alignments and phylogenetic analyses.

The retrieved fertility restorer protein families were used to generate a phylogenetic tree using ClustalW [32]. The alignment was created using the Gonnet protein weight matrix, multiple alignment gap opening/extension penalties of 10/0.5 and pairwise gap opening/extension penalties of 10/0.1. These alignments were adjusted using Bioedit V7.0.5.3 [33]. Portions of sequences that could not be reliably aligned were eliminated. Phylogenetic tree was generated by the neighborjoining method [34], and the branches were tested with 1,000 bootstrap replicates. The three was visualized using TreeDyn program [35].

RF superfamily: protein modeling and structural characterization.

In order to study the structural and conformational variability between the RF protein families, selected members of the RF
superfamily were modeled using the best closed PDB templates structures using SWISS-MODEL, a protein structure homology-modeling server, via the ExPASy web server [40–42]. The initial modeled RF structures were subjected to energy minimization with GROMOS96 force field energy [43] implemented in DeepView/Swiss-PDBViewer v3.7 [44] to improve the van der Waals contacts and to correct the stereochemistry of the improved models. The quality of the models was assessed by checking the protein sterology with PROCHECK [45] and the protein energy with ANOLEA [46]. Ramachandran plot statistics for the models were calculated to show the number of protein residues in the favored regions.

Supporting Information

Table S1 Cross diagram representation of the sequence identity of the RF proteins. The RF percentage identity to each other is used to the catalogue the RF families and subfamilies as detailed in materials and methods section.

(XLS)

Table S2 The fertility restorer protein superfamily: new and unified nomenclature (continued).

(DOC)

Author Contributions

Conceived and designed the experiments: SOK. Performed the experiments: JCJ-L, EWG. Analyzed the data: JCJ-L EWG SOK. Contributed reagents/materials/analysis tools: MJS. Wrote the paper: SOK.

References

1. Leroi KD, Leuten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. Bot Rev 38: 425–454.
2. Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility. Trends Plant Sci 3: 175–180.
3. Shi S, Ding D, Mei S, Wang J (2010) A comparative light and electron microscopic analysis of microspore and tapetum development in fertile and cytoplasmic male sterile radish. Prototaenia 241(1-4): 37–49.
4. González-Melendi P, Uytteremaal M, Morcillo CN, Hernández Moro JR, Fajardo S, Buda F, et al. (2008) A light and electron microscopy analysis of the events leading to male sterility in Oryza/INA CMS of rareded (Bosassan nup). J Exp Bot 59(4): 827–838.
5. Bentolila S, Alfonso AA, Maureen R, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad USA 99: 10897–10899.
6. Chase CD (2006) Cytoplasmic male sterility: a window to the world of plant pathology, and molecular biology of T-cytoplasm male sterility in maize. Adv Agron 65: 79–130.
7. Dill CL, Wue RP, Schnable PS (1997) Mutator-induced mutations of the rf1 nuclear fertility restorer of T-cytoplasm maize alters the accumulation of T-urf13 associated with cytoplasmic male sterility in the T cytoplasm of maize. Proc Natl Acad USA 94: 431–436.
8. Wise RP, Dill CL, Schnable PS (1996) Mutator-induced mutations of the rf1 nuclear fertility restorer of T-cytoplasm maize alters the accumulation of T-urf13 mitochondrial transcripts. Genetics 143: 1383–1394.
9. Kennell JC, Wise RP, Pring DR (1987) Influence of nuclear background on transcription of a maize mitochondrial region associated with Texas male sterile cytoplasm. Mol Gen Genet 210: 399–406.
10. Ullstupa J (1972) The impacts of the southern corn leaf blight epidemics of 1970-1971. Annu Rev Phytopath 10: 37–50.
11. Pring DR, Lonsdale M (1989) Cytoplasmic male sterility and maternal inheritance of disease susceptibility in maize. Annu Rev Phytopathol 47: 483–502.
12. Dill CL, Wise RP, Schnable PS (1997) Rf-P and Rf*-mediate unique Turf13 transcript accumulation, revealing a conserved motif associated with RNA processing and restoration of pollen fertility in T-cytoplasm maize. Genetics 147: 1367–1379.
13. Yoshida A, Rabetsky A, Hsu LC, Chang C (1998) Human aldehyde dehydrogenase gene family. Eur J Biochem 251: 549–557.
14. Pirozich J, Nicholas H, Wang BC, Lindahl R, Hempel J (1999) Relationships within the aldehyde dehydrogenase extended family. Protein Sci 8: 137–146.
15. Vasiliou V, Bairoch A, Tipton FK, Nebert DW (1999) Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. Pharmacogenetics 9: 421–434.
16. Sophos NA, Vasiliou V (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. Chem Biol Interact 143-144: 5–22.
17. Kochtchi SO, Jimenez-Lopez JC, Gao D, Edwards V, Gachomo EW, et al. (2010) Modeling-dependent protein characterization of the rice aldehyde dehydrogenase (ALDH) superfamily reveals distinct functional and structural features. PLoS ONE 5(7): e11516.
18. Small ID, Peeters N (2000) The PPR motif, a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25(2): 46–47.
19. Wise RP, Broussard CN, Schnable PS, Horner HT (1999) The genetics, pathology, and molecular biology of T-cytoplasm male sterility in maize. Adv Agron 63: 79–130.
42. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37: D387–D392.

43. Christen M, Hünenberger PH, Bakowies D, Baron R, Birgi R, et al. (2005) The GROMOS software for biomolecular simulation: GROMOS96. J Comput Chem 26(16): 1719–51.

44. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18(15): 2714–2723.

45. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereo-chemical quality of protein structures. J Appl Cryst 26: 283–291.

46. Melo F, Frymans E (1998) Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 277(5): 1141–1152.