D170–D174 *Nucleic Acids Research, 2019, Vol. 47, Database issue* *Published online 26 October 2018*
doi: 10.1093/nar/gky1026

Plant editosome database: a curated database of RNA editosome in plants

Man Li[1,2,3,†], Lin Xia[1,2,3,†], Yuansheng Zhang[1,2,3], Guangyi Niu[1,2,3], Mengwei Li[1,2,3], Pei Wang[1,2,3], Yang Zhang[1,2,3], Jian Sang[1,2,3], Dong Zou[1,2], Songnian Hu[2,3], Lili Hao[1,2,*,†] and Zhang Zhang[1,2,3,4,*]

[1]BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [2]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [3]University of Chinese Academy of Sciences, Beijing 100049, China and [4]School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, China

Received August 12, 2018; Revised October 15, 2018; Editorial Decision October 16, 2018; Accepted October 17, 2018

ABSTRACT

RNA editing plays an important role in plant development and growth, enlisting a number of editing factors in the editing process and accordingly revealing the diversity of plant editosomes for RNA editing. However, there is no resource available thus far that integrates editosome data for a variety of plants. Here, we present Plant Editosome Database (PED; http://bigd.big.ac.cn/ped), a curated database of RNA editosome in plants that is dedicated to the curation, integration and standardization of plant editosome data. Unlike extant relevant databases, PED incorporates high-quality editosome data manually curated from related publications and organelle genome annotations. In the current version, PED integrates a complete collection of 98 RNA editing factors and 20 836 RNA editing events, covering 203 organelle genes and 1621 associated species. In addition, it contains functional effects of editing factors in regulating plant phenotypes and includes detailed experimental evidence. Together, PED serves as an important resource to help researchers investigate the RNA editing process across a wide range of plants and thus would be of broad utility for the global plant research community.

INTRODUCTION

RNA editing is a post-transcriptional process that changes specific nucleotides of a transcript by nucleotide conversion/insertion/deletion. Studies have shown that a large number of plants undergo RNA editing, albeit to date confined to plant organelles (plastid and mitochondria), which mainly occurs in coding regions and may cause recoding effects (1). Meanwhile, there are also some editing sites locating in non-coding regions, which are essential for posttranscriptional regulations such as splicing (2). Collectively, RNA editing plays important roles in plant developmental processes (3), plant flowering (4), organelle biogenesis (5), response to particular environmental conditions (6,7) and signal transduction (8). It is believed that RNA editing acts as an indirect repair mechanism to correct DNA mutations on the RNA level, enlisting a number of editing factors in the editing process and consequently revealing the diversity of plant editosomes for RNA editing (see a review in (9)).

Over the past several years, a variety of editing factors have been identified as essential components of plant RNA editosome, including pentatricopeptide repeat (PPR) proteins (10), RNA editing factor interacting proteins (RIPs)/multiple organellar RNA editing factors (MORFs) (11,12), organelle RNA recognition motif (ORRM) proteins (13), organelle zinc-finger (OZ) proteins (14), and protoporphyrinogen oxidase 1 (PPO1) (15). Undoubtedly, a complete picture of RNA editing factors is of fundamental significance to better understand the diversity of plant editosomes for RNA editing machinery and potential applications. To date, however, there is still no comprehensive collection of RNA editing factors in plants, although several databases have already been developed to integrate various RNA editing data. For instance, REDIdb (16) is specialized for collecting RNA editing events in plant organelles, PREPACT 3.0 (17) aims to facilitate RNA editing data analysis by integrating information on characterized editing factors for editing site recognition specificities, dbRES (18) integrates RNA editing sites for plant, metazoan, protozoa, fungi and virus, RESOPS (19) focuses on the correspondence of RNA editing sites of land plant organelles to

*To whom correspondence should be addressed. Zhang Zhang: Tel: +86 10 84097261; Fax: +86 10 84097720; Email: zhangzhang@big.ac.cn
Correspondence may also be addressed to Lili Hao. Email: haolili@big.ac.cn
†The authors wish it to be known that, in their opinion, first two authors should be regarded as Joint First Authors.

protein three-dimensional structures, and REDIportal (20) and RADAR (21) are two popular databases containing A-to-I RNA editing data in human. Therefore, a specialized database of plant RNA editosomes that houses diverse plant RNA editing factors as well as their associated data is highly desirable.

Here we present Plant Editosome Database (PED; http://bigd.big.ac.cn/ped), a curated database of plant RNA editosomes that is dedicated to the curation, integration and standardization of plant editosome data. Unlike extant relevant databases, PED features high-quality editosome data manually curated from related publications and organelle genome annotations. Accordingly, PED incorporates a complete collection of RNA editing factors and associated data across a wide range of plants, which would be of broad utility for the global plant research community.

METHODS AND MATERIALS

All editing events integrated in PED were derived from a total of 2651 organelle flat files downloaded from NCBI RefSeq Organelle Genome Resources (https://www.ncbi.nlm.nih.gov/genome/organelle/). Specifically, an abundance of editing information, including edited gene, editing position, editing region, editing type, codons before and after editing, amino acids before and after editing, etc., were extracted and parsed by a python script. To obtain high-confidence RNA editing information, genome sequences and related publications were then retrieved by manual curation when some RNA editing information were missing or questionable. As a result, 20 836 RNA editing events in 203 organelle genes were obtained, involving a total of 1621 plant species and 1673 organelle genomes (http://bigd.big.ac.cn/ped/dataset).

In addition, RNA editing factors as well as their associated editing events were manually curated from published literature. Accordingly, a collection of 98 editing factors were obtained, which were classified into several groups (PPR protein, MORFs/RIPs, ORRM protein, OZ protein, etc.) based on their conserved motifs as detailed in (9). To facilitate RNA editing data curation in a standardized manner, a curation model containing controlled vocabularies and conclusion terms was developed, involving a variety of data items such as editing factor, factor family, associated editing events, edited gene(s), editing type, experimental information, mutant effect relative to wild type, etc. More details about this curation model are publicly available at http://bigd.big.ac.cn/ped/help.

PED was constructed with MySQL (http://www.mysql.org; a free and popular relational database management system) as its database engine and web interfaces were developed by using JSP (JavaServer Pages; a technology facilitating rapid development of dynamic web pages based on the Java programming language) and AJAX (Asynchronous JavaScript and XML; a set of web development techniques to create asynchronous applications without interfering with the display and behavior of the existing page). In addition, Bootstrap (https://getbootstrap.com; an open source toolkit for developing with HTML, CSS and JavaScript) was adopted as a front-end framework, which provides a series of templates for designing web pages with consistent interface components.

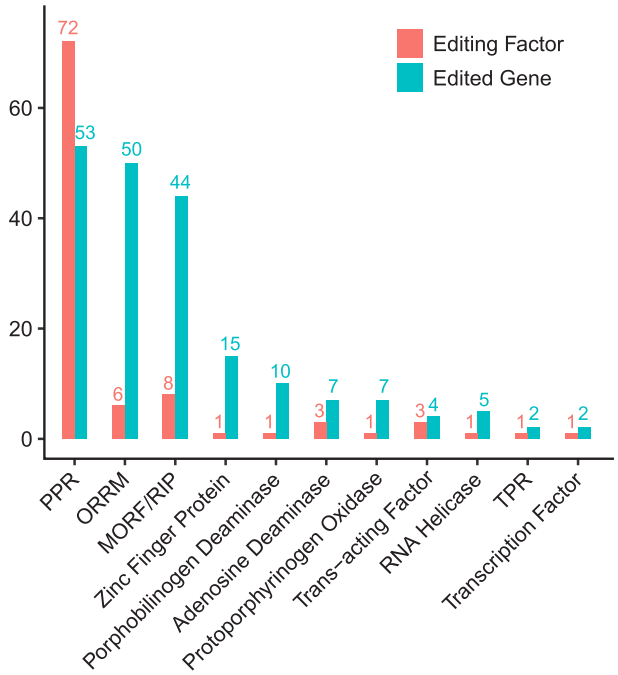

Figure 1. Statistics of experimentally validated RNA editing factors and associated edited genes in PED (as of August 2018).

DATABASE CONTENT AND USAGE

PED features comprehensive integration of RNA editing factors and their associated data manually curated from related publications and organelle annotation files. Collectively, the current version of PED houses a total of 98 RNA editing factors and 20 836 editing events, involving 203 organelle genes and covering 1621 plant species and 1673 plant organelles. In addition, PED includes interactions between editing factors and editing events in 8 model species, functional effects of editing factors in regulating plant phenotypes as well as detailed experimental evidence.

PED contains 98 experimentally validated RNA editing factors and their associated edited genes manually curated from 94 publications (Figure 1), corresponding to a complete collection to date of known plant editing factor families as well as their associated experimental treatments and mutant phenotypes. To have a complete picture of plant RNA editing factors, PED is equipped with a specific page for editing factors (Figure 2A), demonstrating that, according to the current collection, the majority of plant editing factors are PPR proteins (72 in count) and the following are MORF/RIP and ORRM. Accordingly, it facilitates access to all collected factors and their associated information, including species, protein family, subclass, associated phenotype(s), edited gene(s), organelle and publication(s). For each specific editing factor, PED summarizes all relevant editing events and provides an abundance of detailed information, including editing position, editing region, editing type, changes of codons & amino acids (if available), editing effect (recoding or synonymous) and detailed infor-

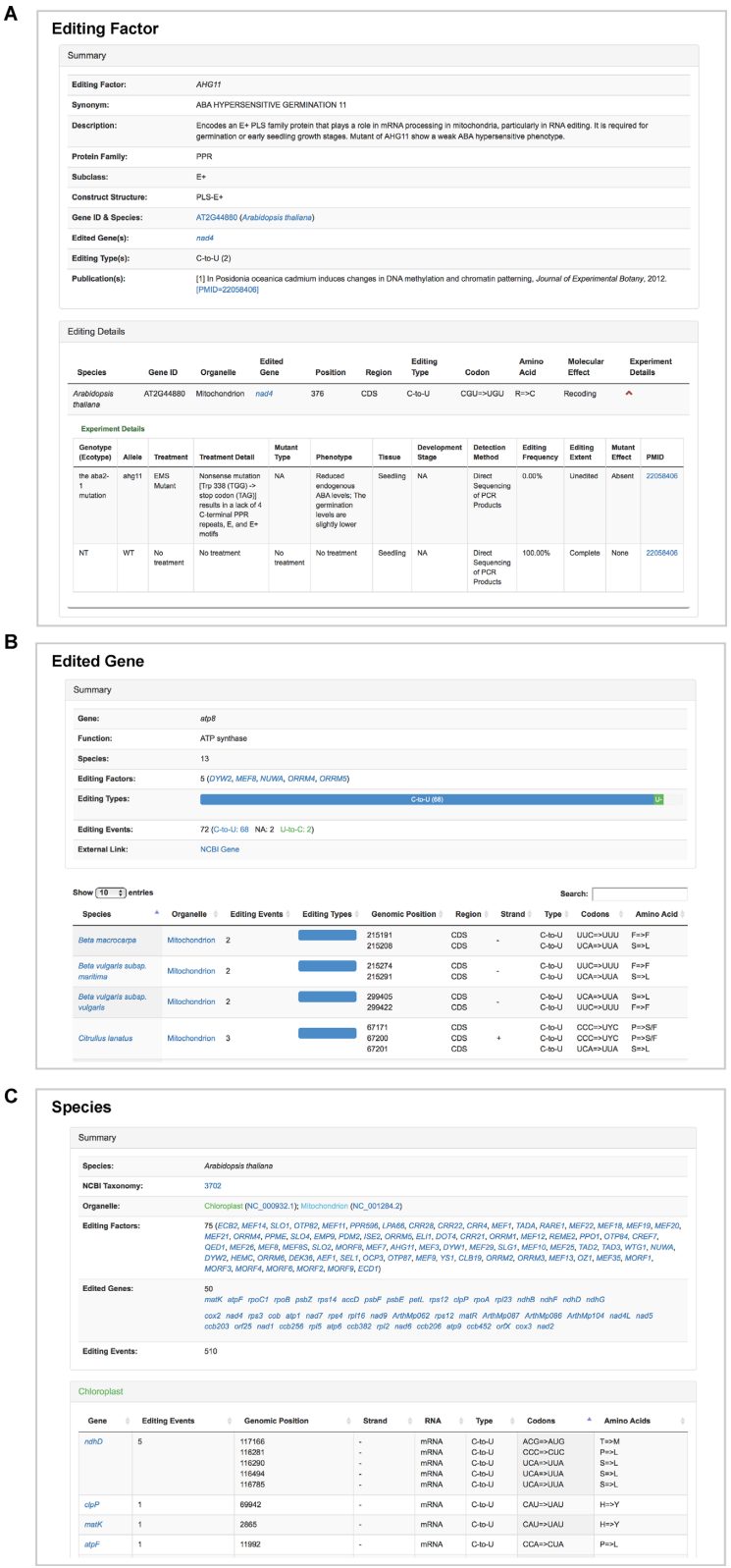

Figure 2. Screenshots of PED web pages, including (**A**) RNA editing factors, (**B**) edited genes and (**C**) species.

mation of the experiment evidence. Interestingly, *DYW2*, an editing factor in *Arabidopsis thaliana* encoding an atypical PPR-DYW protein, is found to have a larger number of edited genes (41 in count) according to the current implementation of PED. Together, all editing factors collected in PED are publicly accessible at http://bigd.big.ac.cn/ped/browse/factors.

PED collects a total of 203 edited genes associated with 20 836 RNA editing events (http://bigd.big.ac.cn/ped/browse/genes). It is revealed from the current version of PED that *rpl2*, a ribosomal protein L2 locating in mitochondria, chloroplast and plastid, contains a large number of editing events in more than 600 plant species, and that *ndhB*, an edited gene of NADH dehydrogenase subunit 2 locating in chloroplast and plastid, is related with 20 different editing factors. According to the current collection of PED, editing events are classified into five types: C-to-U (11, 903), U-to-C (2, 613), A-to-G (15), G-to-A (9) and G-to-U (6), agreeing well with previous studies that C-to-U and U-to-C conversions are most common in plants (22). For a given edited gene, PED provides both basic descriptive information and associated detailed information of editing events (Figure 2B). Basic information includes gene symbol, gene function, organelle type, functional category, number of associated species, number and list of experimentally validated editing factors, number of editing events occurring on edited gene, editing type, and external links to NCBI. Detailed information of editing events contains editing position, editing type, strand, change of codons & amino acids. For instance, *atp6*, a mitochondrial gene (ATP synthase F0 subunit 6), contains >300 RNA editing events among 27 species, associates with three editing factors (*DYW2*, *EMP5*, *NUWA*) and results in multiple different amino acid changes (e.g. proline to leucine, glutamine to stop codon), which, collectively, are helpful for users to investigate RNA editing patterns and explore the editing machinery across a wide range of species.

PED also organizes editing information in terms of species (http://bigd.big.ac.cn/ped/browse/species), where each species has a summary of editing information (Figure 2C). Basic descriptive information includes organelle type, organelle accession number, number of edited genes, number of editing events, list of edited genes. Further detailed features of edited genes contain genomic position, strand, RNA type, editing type, change of codons & amino acids. Currently, PED incorporates 1673 plant organelle genomes from 1621 species, corresponding to a wide diversity of plants and algae, including monocots, eudicots, ferns, club-mosses, bryophytes, and other *Viridaeplantae* and accounting for a total of 20 836 RNA editing events. Of these RNA editing events currently stored in PED, 16 578 occur in mRNAs (16 351 in protein-coding sequences), 252 in tRNAs and 4006 in unannotated transcripts. Also, according to the current collection in PED, organelle genomes from *Anthoceros angustus* (chloroplast; 78 edited genes), *Nelumbo nucifera* (mitochondrion; 38 edited genes) and *Nothoceros aenigmaticus* (plastid; 88 edited genes), contain more editing events (1018, 771 and 226, respectively). In addition, the currently stored RNA editing events are mainly from chloroplast, accounting for a total of 10 916 events from 1255 chloroplasts genomes. In spite of the higher total number of editing sites in chloroplast, the frequency of editing is much lower than that in mitochondria, with 67 editing sites/per genome for mitochondria, but only about nine editing sites/per genome for chloroplast.

DISCUSSION AND FUTURE DEVELOPMENTS

PED is a curated database of RNA editosome in plants, featuring high-quality editosome data manually curated from published literature and organelle genome annotations. The current implementation of PED houses a complete collection of 98 RNA editing factors and integrates 203 edited genes and 20 836 editing events across a wide range of plant organelle genomes. As a featured resource of BIG Data Center (http://bigd.big.ac.cn) (23,24), PED is committed to the curation, integration and standardization of plant editosome data and thus has the great potential to help researchers conduct systematic investigations on RNA editing machinery in a variety of plant species. Considering the increasing number of published studies on plant RNA editing, a specialized team is responsible for PED to make it up-to-date and comprehensive. Future directions of PED include integration of more editing factors, and importantly, their potential interactions (25) through literature curation. Since some editing factors are evolutionarily related and belong to one common protein family, we will build phylogenetic trees for closely related editing factors, with the aim to facilitate researchers to study RNA editing across species. In addition, we will develop online tools to help users to explore the RNA editing process for a specific editing factor, compare RNA editing pattern between different factors/genes, and investigate the conservation of RNA editing sites across a variety of species. For any given editing factor, we will also develop new functionalities to visualize its associated edited gene(s) and editing sites across a wide range of species.

ACKNOWLEDGEMENTS

We thank a number of users for reporting bugs and providing suggestions as well as the anonymous reviewers for their valuable comments on this work.

FUNDINGS

Strategic Priority Research Program of the Chinese Academy of Sciences [XDA19050302, XDA08020102]; The Youth Innovation Promotion Association of Chinese Academy of Science [2018134]; National Key Research & Development Program of China [2017YFC0907502]; National Programs for High Technology Research and Development [2015AA020108, 2012AA020409]; International Partnership Program of the Chinese Academy of Sciences [153F11KYSB20160008]; The 13th Five-year Informatization Plan of Chinese Academy of Sciences [XXH13505-05]; National Natural Science Foundation of China [31100915]. Funding for open access charge: Strategic Priority Research Program of the Chinese Academy of Sciences.

Conflict of interest statement. None declared.

REFERENCES

1. Yan,J., Zhang,Q. and Yin,P. (2018) RNA editing machinery in plant organelles. *Sci. China Life Sci.*, **61**, 162–169.
2. Castandet,B., Choury,D., Begu,D., Jordana,X. and Araya,A. (2010) Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res.*, **38**, 7112–7121.
3. Takenaka,M., Zehrmann,A., Verbitskiy,D., Hartel,B. and Brennicke,A. (2013) RNA editing in plants and its evolution. *Annu. Rev. Genet.*, **47**, 335–352.
4. Shi,X., Germain,A., Hanson,M.R. and Bentolila,S. (2016) RNA recognition motif-containing protein ORRM4 broadly affects mitochondrial RNA editing and impacts plant development and flowering. *Plant Physiol.*, **170**, 294–309.
5. Lurin,C., Andres,C., Aubourg,S., Bellaoui,M., Bitton,F., Bruyere,C., Caboche,M., Debast,C., Gualberto,J., Hoffmann,B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089–2103.
6. Yuan,H. and Liu,D. (2012) Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. *Plant J.*, **70**, 432–444.
7. Garcia-Andrade,J., Ramirez,V., Lopez,A. and Vera,P. (2013) Mediated plastid RNA editing in plant immunity. *PLoS Pathog.*, **9**, e1003713.
8. Tang,J., Kobayashi,K., Suzuki,M., Matsumoto,S. and Muranaka,T. (2010) The mitochondrial PPR protein LOVASTATIN INSENSITIVE 1 plays regulatory roles in cytosolic and plastidial isoprenoid biosynthesis through RNA editing. *Plant J.*, **61**, 456–466.
9. Sun,T., Bentolila,S. and Hanson,M.R. (2016) The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.*, **21**, 962–973.
10. Kotera,E., Tasaka,M. and Shikanai,T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, **433**, 326–330.
11. Bentolila,S., Heller,W.P., Sun,T., Babina,A.M., Friso,G., van Wijk,K.J. and Hanson,M.R. (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E1453–E1461.
12. Takenaka,M., Zehrmann,A., Verbitskiy,D., Kugelmann,M., Hartel,B. and Brennicke,A. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 5104–5109.
13. Sun,T., Germain,A., Giloteaux,L., Hammani,K., Barkan,A., Hanson,M.R. and Bentolila,S. (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E1169–E1178.
14. Sun,T., Shi,X., Friso,G., Van Wijk,K., Bentolila,S. and Hanson,M.R. (2015) A zinc finger motif-containing protein is essential for chloroplast RNA editing. *PLos Genet.*, **11**, e1005028.
15. Zhang,F., Tang,W., Hedtke,B., Zhong,L., Liu,L., Peng,L., Lu,C., Grimm,B. and Lin,R. (2014) Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2023–2028.
16. Lo Giudice,C., Pesole,G. and Picardi,E. (2018) REDIdb 3.0: A comprehensive collection of RNA editing events in plant organellar genomes. *Front. Plant Sci.*, **9**, 482.
17. Lenz,H., Hein,A. and Knoop,V. (2018) Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinformatics*, **19**, 255.
18. He,T., Du,P. and Li,Y. (2007) dbRES: a web-oriented database for annotated RNA editing sites. *Nucleic Acids Res.*, **35**, D141–D144.
19. Yura,K., Sulaiman,S., Hatta,Y., Shionyu,M. and Go,M. (2009) RESOPS: a database for analyzing the correspondence of RNA editing sites to protein three-dimensional structures. *Plant Cell Physiol.*, **50**, 1865–1873.
20. Picardi,E., D'Erchia,A.M., Lo Giudice,C. and Pesole,G. (2017) REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.*, **45**, D750–D757.
21. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.
22. Chateigner-Boutin,A.L. and Small,I. (2010) Plant RNA editing. *RNA Biol*, **7**, 213–219.
23. BIG Data Center Members. (2017) The BIG Data Center: from deposition to integration to translation. *Nucleic Acids Res.*, **45**, D18–D24.
24. BIG Data Center Members. (2018) Database resources of the BIG Data Center in 2018. *Nucleic Acids Res.*, **46**, D14–D20.
25. Yan,J., Zhang,Q., Guan,Z., Wang,Q., Li,L., Ruan,F., Lin,R., Zou,T. and Yin,P. (2017) MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat. Plants*, **3**, 17037.