



Original article

FairBase: a comprehensive database of fungal A-to-I RNA editing

Jinding Liu^{1,2,3}, Dongbo Wang^{1,2}, Yinna Su^{2,3}, Kun Lang^{1,2}, Rongjing Duan^{2,3}, YuFeng Wu³, Fei Ma^{4,5} and Shuiqing Huang^{1,2,*}

¹College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095, China, ²Research Center for Correlation of Domain Knowledge, Nanjing Agricultural University, Nanjing 210095, China, ³Bioinformatics center, Nanjing Agricultural University, Nanjing 210095, China, ⁴College of Life Science, Nanjing Normal University, Nanjing 210023, China and ⁵Laboratory for Comparative Genomics and Bioinformatics, Nanjing Normal University, Nanjing 210023, China

*Corresponding author: Tel: +86 25 8439 5988; Fax: +86 25 8439 5988; Email: sqhuang@njau.edu.cn

Citation details: Liu, J., Wang, D., Su, Y. *et al.* FairBase: a comprehensive database of fungal A-to-I RNA editing. *Database* (2019) Vol. 2019: article ID baz018; doi:10.1093/database/baz018

Received 19 November 2018; Revised 4 January 2019; Accepted 22 January 2019

Abstract

Frequent A-to-I RNA editing has recently been identified in fungi despite the absence of recognizable homologues of metazoan ADARs (“Adenosine Deaminases Acting on RNA”). In particular, there is emerging evidence showing that A-to-I editing is involved in sexual reproduction of filamentous fungi. Here, we report on the creation of FairBase — a fungal A-to-I RNA editing database that provides a platform for deep exploration of fungal RNA editing to relevant academic communities. This database includes a comprehensive collection of A-to-I editing sites in six filamentous fungal species, together with extensive annotations for each editing site. In FairBase, users can conveniently search editing sites and obtain editing levels for each editing site in various RNA-seq samples. In addition, the pathways involving RNA editing are built in FairBase to help users understand the functions of RNA editing. Furthermore, each fungal species has a genome browser (JBrowse) that allows users to explore A-to-I editing in a genomic context. FairBase is the first fungal RNA editing database.

Database URL: bioinfo.njau.edu.cn/fairbase

Introduction

RNA editing is a class of co-/post-transcriptional modification that can alter hereditary information in the genome by changing the nucleotide sequence of RNA molecules (1–3). A-to-I RNA editing catalyzed by members of the adenosine

deaminase acting on RNA (ADAR) family of enzymes is the main type of RNA editing in animals; this editing can expand the transcriptome and proteome (4). Studies have revealed that A-to-I editing can be involved in the regulation protein recoding, alternative splicing, RNA interference

(RNAi) and microRNA expression (5–8). Moreover, A-to-I editing can also be involved in the progressions of many diseases including cancer, neurological deterioration, and metabolic disorders (9–13). In recent years, the advent of high-throughput RNA sequencing (RNA-seq) has enabled transcriptome-wide identification of RNA editing sites and has greatly accelerated the discovery of A-to-I editing sites. The boom in identification of A-to-I editing site has necessitated the development of RNA editing databases to help elucidate biological functions of RNA editing. Some RNA editing databases have been created and are used for large-scale collection and annotation of RNA editing for animals. For example, DARNED is the earliest A-to-I RNA editing database, including editing sites for human, mouse, and fruit fly (14). RADAR not only expanded the number of editing sites but also incorporated editing levels in RNA-seq samples and conservation of editing to other species (15). Recently, REDportal was designed to embed RADAR database and represents the first editing resource designed to answer functional questions, enabling the inspection and browsing of editing levels in a variety of human samples, tissues and body sites (16).

In fungi, A-to-I RNA editing is usually thought to be scarce due to the lack of orthologs of animal ADARs. Until now, only a small number of large-scale RNA editing discovery studies and very few functional experiments for A-to-I RNA editing have been reported in fungi (17–21). Based on RNA-seq data, thousands of RNA editing sites have been identified in *Ganoderma lucidum* (17) and *Fomitopsis pinicola* (21), but no preference for A-to-I over other forms of editing was found, and the functional significance of A-to-I RNA editing was not studied. In *Fusarium graminearum*, more than 26 000 stage-specific A-to-I RNA editing sites were identified during sexual development (18). Furthermore, the editing of RNA encoding a protein kinase PUK1 (perithecium unique kinase) was experimentally shown to play an important role in ascospore formation and release (18). Similarly, more than 40 000 stage-specific A-to-I RNA editing sites were identified at different sexual stages in *Neurospora crassa* (19). The comprehensive analyses of A-to-I editing in *N. crassa* showed A-to-I editing was generally adaptive, and might to be functionally related to repeat induced point mutation and meiotic silencing by unpaired DNA (19, 22). A-to-I RNA editing during sexual development also occurred in *Sordaria macrospora* and *Pyronema confluens*, two distantly related filamentous ascomycetes, suggesting that stage-specific A-to-I editing might be an evolutionary conserved feature during sexual development in filamentous ascomycetes (20).

Despite recent advances in fungal RNA editing, the exact biological role of A-to-I RNA editing, and the adeno-

sine deamination mechanism have not been elucidated in fungi. Systematically gathering and sorting the resources of RNA editing could provide the reference information when further probing the functions and molecular mechanisms of RNA editing, even when performing cross-species functional verification in different species. In this work, using almost all available RNA-seq data of filamentous fungi, we developed FairBase, the first fungal A-to-I editing database. FairBase includes a comprehensive collection of A-to-I editing sites, together with extensive annotations and editing levels for each site. In addition, FairBase provides various query functions and graphical visualization pages to facilitate access to RNA editing data.

Material and methods

RNA-seq data collection and preparation

Only RNA-seq samples with available reference genome assemblies in the Ensembl Fungi database (23) and available RNA-seq data in the NCBI Sequence Read Archive (SRA) database (24) were collected for identification of A-to-I RNA editing. A total of 6253 raw RNA-seq data in SRA format were downloaded and converted into FASTQ format. Then, Trimmomatic (25) was used to cut sequencing adapters and low quality sequences from the 3' and 5' end until the base quality score was at least 10. To obtain high-quality clean data, reads with an overall mean Phred-scaled value less than 20 were discarded. Clean reads of each RNA-seq sample were aligned to a reference genome using Hisat2 (26) and read alignments were saved in mapped bam files. Only RNA-seq samples with more than 70% mapped reads were used for the identification of RNA editing events.

A-to-I RNA editing calling

Duplicated reads were removed from mapped bam files using the MarkDuplicates program in the Picard package (<https://broadinstitute.github.io/picard>). REDIttool was used to call RNA variants present in at least five reads with a minimum frequency of 3% and minimum coverage of 10 reads (27). For non-strand-specific RNA-seq data, mismatch type was inferred by gene annotation and RNA variants occurring in the intergenic region were discarded. Sine noncanonical RNA variants tend to be false-positives in animals (28–30), to improve the reliability of A-to-I RNA editing event, RNA-seq samples without significant preference for A-to-I RNA variants also were discarded. Finally, a total of 32 RNA-seq samples were retained, including 7 samples designed for fungal RNA editing studies (18, 19) and 25 newfound RNA-seq samples.

In the newfound RNA-seq samples, there were 13 samples with RNA-seq strains were inconsistent with reference genome strains (Table S1). The single nucleotide polymorphisms (SNPs) reported in previous studies (31, 32) were used to exclude SNPs at the transcript level in six *F. graminearum* RNA-seq samples with inconsistent reference genome strains. For the remaining seven RNA-seq samples, the whole genome shotgun sequencing (WGS) data from RNA-seq strains were aligned to reference genomes to add genomic support for the exclusion of SNPs at the transcript level (Table S1). Furthermore, A-to-I RNA variants with extreme degree of variance (>90%), which were most likely to be genomic variants, were excluded from the 13 RNA-seq samples.

Annotation of A-to-I RNA editing sites

The editing events reported in the seven RNA-seq samples from previous studies (18, 19) and the editing events identified in the remaining 25 RNA-seq samples were included in FairBase. The positions of RNA editing events of each fungal species were merged to yield a comprehensive and non-redundant catalogue of editing sites. For each editing site, we curated annotations including genome context, codon change, amino acid change, and edited gene. To support the data retrieval with functions of edited gene, Gene ontology (GO) and Protein family (Pfam) function annotations were predicted using InterProScan (33). Furthermore, the pathways involving RNA editing were built using Kyoto Encyclopedia of Genes and Genomes (KEGG) database (34).

Implementation of FairBase

We implemented the FairBase database using MySQL as the back-end database, PHP and Perl for the server-side scripting, as well as JavaScript and JQuery plugins as front-end interface. Moreover, the JBrowse genome browser

(35) was also embedded in FairBase for users to explore editing sites in genomic context.

Results

Database content

Currently, a total of 32 RNA-seq samples with significant preference for A-to-I editing are deposited in FairBase. These RNA-seq samples across six fungi, including *Fusarium graminearum*, *Fusarium verticillioides*, *Neurospora crassa*, *Neurospora tetrasperma*, *Pyronema omphalodes* and *Sordaria macrospora* and are all related to fungal sexual reproduction (Table S1). In the 32 RNA-seq samples, a total of 512 719 A-to-I editing events are detected on 137 213 editing sites (Table 1). Of these editing sites, about 70% occur in coding regions (CDSs) and 56% editing sites can result in nonsynonymous recoding, which can result in amino acid change.

Web Interface

User-friendly web interfaces are designed for users to access the FairBase database. Data retrieval can be achieved in the search, blast and pathway page and the retrieved editing sites are listed in a sortable and downloadable table below (Figure 1). The complete description of an editing site is shown in the editing detail page (Figure 2). All editing sites and editing events can be explored in genome context in the JBrowse genome browser (Figure 3).

In the Search page, editing sites can be searched by genomic position including species, chromosome and coordinate (Figure 1A). Users can define extra filters to restrict retrieval results to the editing sites of interest: (i) Gene, selecting editing sites in edited genes with specified symbol as well as Pfam and GO terms; (ii) Context, selecting editing sites in specified genomic context such as 5'UTR (Untranslated region), CDS (Coding region), 3'UTR, Intronic and Intergenic region; and (iii) AA change,

Table 1. Statistic of A-to-I editing events and sites collected in FairBase

Species	RNA-seq samples	Editing events	Editing sites (Nonsyn%*)
<i>F. graminearum</i>	12	303 279	48 508 (60%)
<i>N. crassa</i>	11	161 197	47 346 (50%)
<i>N. tetrasperma</i>	5	32 355	28 492 (53%)
<i>F. verticillioides</i>	1	5227	5227 (72%)
<i>P. omphalodes</i>	1	8238	5217 (73%)
<i>S. macrospora</i>	2	2423	2423 (65%)
Total	32	512 719	137 213 (56%)

*Nonsyn%, Percentage of nonsynonymous editing sites occurred in coding regions.

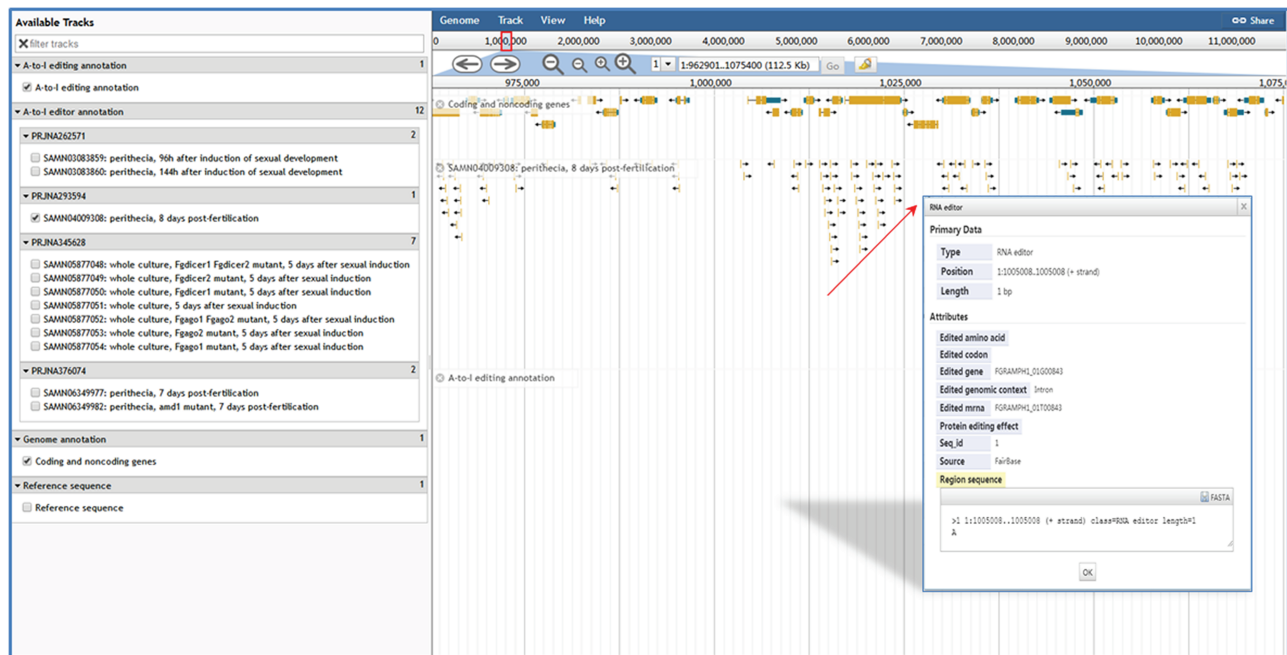


Figure 3. JBrowse in FairBase. Editing sites can be explored in a genomic context using JBrowse. Users can obtain the editing details by clicking on an editing event.

FairBase provides the BLAST search function, which not only enables the retrieval of editing sites through sequence similarity, but also contributes to the cross-species functional verification between edited genes in FairBase and out of FairBase. Furthermore, KEGG pathways and JBrowse genome browser are built in FairBase for users to explore editing sites in a pathway and genome context. Compared with several animal RNA editing databases (14–16), FairBase collects more species and provides more retrieval functions (Table S2).

It is challenging work to identify RNA editing sites from public available data, because many factors, such as sequencing errors, SNPs among different strains, and other problems can affect the reliability of results. Thus the strict strategies were used in this study. For example, only RNA-seq samples with at least 70% mapped reads can be used for identification of RNA editing events. In addition, in view of the fact that noncanonical RNA edits are found to be false-positives (28–30, 36) and A-to-I RNA editing are enriched in animals RNA-seq samples (37, 38), RNA-seq samples without significant enrichment for A-to-I RNA editing also were discarded in this study. Although thousands of RNA-seq samples were gathered for identification of A-to-I editing sites, ultimately only 32 RNA-seq samples related to sexual reproduction were retained. The results seem to show that A-to-I RNA editing occurs specifically during sexual reproduction. However, whether A-to-I RNA editing is related to other developmental or infection stages still is an open issue (22). It is anticipated that high-throughput

sequencing will be continually applied in various fungal studies, including the function and mechanism of RNA editing. With more high-throughput omics data available, we will regularly harvest fungal A-to-I editing data and keep FairBase up-to-date. In addition, the retrieval functions will be further enriched, such as support for the retrieval of editing sites according to codon-change and editing efficiencies and so on. We believe this database and its future updates will be a valuable resource to boost fungal RNA editing researches.

Supplementary data

Supplementary data are available at *Database* Online.

Acknowledgements

We acknowledge the works of all the fungal genome and RNA-seq data producers.

Funding

This work has been supported by the National Nature Science Foundation of China (31301691) and Fundamental Research Funds for the Central Universities (KYZ201667, KJQN201430).

Conflict of interest. None declared.

References

- Farajollahi,S. and Maas,S. (2010) Molecular diversity through RNA editing: a balancing act. *Trends Genet*, **26**, 221–230.

2. Gray,M.W. (2012) Evolutionary Origin of RNA Editing. *Biochemistry-Us*, 51, 5235–5242.
3. Knoop,V. (2011) When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci*, 68, 567–586.
4. Mallela,A. and Nishikura,K. (2012) A-to-I editing of protein coding and noncoding RNAs. *Crit Rev Biochem Mol*, 47, 493–501.
5. Maas,S. (2010) Gene Regulation Through RNA Editing. *Discov Med*, 10, 379–386.
6. Nishikura,K. (2006) Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Bio*, 7, 919–931
7. Laurencikieni,J., Kallman,A.M., Fong,N. *et al.* (2006) RNA editing and alternative splicing: the importance of co-transcriptional coordination. *Embo Rep*, 7, 303–307.
8. Yang,W.D., Chendrimada,T.P., Wang,Q.D. *et al.* (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol*, 13, 13–21.
9. Fritzell,K., Xu,L.D., Lagergren,J. and Ohman,M. (2018) ADARs and editing: The role of A-to-I RNA modification in cancer progression. *Semin Cell Dev Biol*, 79, 123–130.
10. Gan,Z.J., Zhao,L.Y., Yang,L. *et al.* (2006) RNA editing by ADAR2 is metabolically regulated in pancreatic islets and beta-cells. *J Biol Chem*, 281, 33386–33394.
11. Khermesh,K., D'Erchia,A.M., Barak,M. *et al.* (2016) Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *Rna*, 22, 290–302.
12. Kubota-Sakashita,M., Iwamoto,K., Bundo,M. *et al.* (2014) A role of ADAR2 and RNA editing of glutamate receptors in mood disorders and schizophrenia. *Mol Brain*, 7.
13. Wang,C., Zou,J., Ma,X.Y. *et al.* (2017) Mechanisms and implications of ADAR-mediated RNA editing in cancer. *Cancer Lett*, 411, 27–34.
14. Kiran,A.M., O'Mahony,J.J., Sanjeev,K. *et al.* (2013) Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res*, 41, D258–D261.
15. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res*, 42, D109–D113.
16. Picardi,E., D'Erchia,A.M., Lo Giudice,C. *et al.* (2017) REDIPortal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res*, 45, D750–D757.
17. Zhu,Y.J., Luo,H.M., Zhang,X. *et al.* (2014) Abundant and Selective RNA-Editing Events in the Medicinal Mushroom *Ganoderma lucidum*. *Genetics*, 196, 1047.
18. Liu,H.Q., Wang,Q.H., He,Y. *et al.* (2016) Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. *Genome Res*, 26, 499–509.
19. Liu,H.Q., Li,Y., Chen,D.P. *et al.* (2017) A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in *Neurospora crassa*. *P Natl Acad Sci USA*, 114, E7756–E7765.
20. Teichert,J., Dahlmann,T.A., Kuck,U. *et al.* (2017) RNA Editing During Sexual Development Occurs in Distantly Related Filamentous Ascomycetes. *Genome Biol Evol*, 9, 855–868.
21. Wu,B.J., Gaskell,J., Held,B.W. *et al.* (2018) Substrate-Specific Differential Gene Expression and RNA Editing in the Brown Rot Fungus *Fomitopsis pinicola*. *Appl Environ Microb*, 84.
22. Wang,C.F., Xu,J.R. and Liu,H.Q. (2016) A-to-I RNA editing independent of ADARs in filamentous fungi. *Rna Biol*, 13, 940–945.
23. Kersey,P.J., Allen,J.E., Armean,I. *et al.* (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res*, 44, D574–D580.
24. Leinonen,R., Sugawara,H., Shumway,M. *et al.* (2011) The Sequence Read Archive. *Nucleic Acids Res*, 39, D19–D21.
25. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
26. Kim,D., Landmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 12, 357–U121.
27. Picardi,E. and Pesole,G. (2013) REDIttools: high-throughput RNA editing detection made easy. *Bioinformatics*, 29, 1813–1814.
28. Kleinman,C.L. and Majewski,J. (2012) Comment on "Widespread RNA and DNA Sequence Differences in the Human Transcriptome". *Science*, 335.
29. Piskol,R., Peng,Z.Y., Wang,J. *et al.* (2013) Lack of evidence for existence of noncanonical RNA editing. *Nat Biotechnol*, 31, 19–20.
30. Ramaswami,G., Zhang,R., Piskol,R. *et al.* (2013) Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods*, 10, 128–132.
31. Laurent,B., Moinard,M., Spataro,C. *et al.* (2017) Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. *Bmc Genomics*, 18.
32. Cuomo,C.A., Gueldener,U., Xu,J.R. *et al.* (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*, 317, 1400–1402.
33. Jones,P., Binns,D., Chang,H.Y. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240.
34. Kanehisa,M., Furumichi,M., Tanabe,M. *et al.* (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*, 45, D353–D361.
35. Skinner,M.E., Uzilov,A.V., Stein,L.D. *et al.* (2009) JBrowse: A next-generation genome browser. *Genome Research*, 19, 1630–1638.
36. Lin,W., Piskol,R., Tan,M.H. *et al.* (2012) Comment on "Widespread RNA and DNA Sequence Differences in the Human Transcriptome". *Science*, 335.
37. Peng,Z.Y., Cheng,Y.B., Tan,B.C.M. *et al.* (2012) Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol*, 30, 253.
38. Bahn,J.H., Lee,J.H., Li,G. *et al.* (2012) Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Research*, 22, 142–150.