Abundant and Selective RNA-Editing Events in the Medicinal Mushroom *Ganoderma lucidum*

Yingjie Zhu,^{*,1} Hongmei Luo,^{*,1,2} Xin Zhang,^{*} Jingyuan Song,^{*} Chao Sun,^{*} Aijia Ji,^{*} Jiang Xu,[†] and Shilin Chen^{*,1,2}

^{*}The National Engineering Laboratory for Breeding of Endangered Medicinal Materials, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China, and [†]Institute of Chinese Materia Medica, China Academy of Chinese Medicinal Sciences, Beijing 100700, China

ABSTRACT RNA editing is a widespread, post-transcriptional molecular phenomenon that diversifies hereditary information across various organisms. However, little is known about genome-scale RNA editing in fungi. In this study, we screened for fungal RNA editing sites at the genomic level in *Ganoderma lucidum*, a valuable medicinal fungus. On the basis of our pipeline that predicted the editing sites from genomic and transcriptomic data, a total of 8906 possible RNA-editing sites were identified within the *G. lucidum* genome, including the exon and intron sequences and the 5′-/3′-untranslated regions of 2991 genes and the intergenic regions. The major editing types included C-to-U, A-to-G, G-to-A, and U-to-C conversions. Four putative RNA-editing enzymes were identified, including three adenosine deaminases acting on transfer RNA and a deoxycytidylate deaminase. The genes containing RNA-editing sites were functionally classified by the Kyoto Encyclopedia of Genes and Genomes enrichment and gene ontology analysis. The key functional groupings enriched for RNA-editing sites included laccase genes involved in lignin degradation, key enzymes involved in triterpenoid biosynthesis, and transcription factors. A total of 97 putative editing sites were randomly selected and validated by using PCR and Sanger sequencing. We presented an accurate and large-scale identification of RNA-editing events in *G. lucidum*, providing global and quantitative cataloging of RNA editing in the fungal genome. This study will shed light on the role of transcriptional plasticity in the growth and development of *G. lucidum*, as well as its adaptation to the environment and the regulation of valuable secondary metabolite pathways.

RNA editing is a post-transcriptional process that can enhance the diversity of gene products by altering RNA sequences. This can involve site-specific nucleotide modifications, nucleotide insertions/deletions, and nucleotide substitutions (Gott and Emeson 2000; Bass 2002; Farajollahi and Maas 2010). RNA editing can affect messenger RNAs (mRNAs), transfer RNAs (tRNAs), ribosomal RNAs, and small regulatory RNAs present in all cellular compartments and has been described in a wide range of organisms from viruses to fungi, mammals, and plants (Gott and Emeson

2000; Gott 2003; Kawahara *et al.* 2007). The prevalent RNA-editing pathways in higher eukaryotes involves the deamination of cytidine (C) to create uridine (U) and deamination of adenosine (A) to create inosine (I) (Bass 2002; Gott 2003). The editing process generally involves a specificity factor (RNA or protein) that recognizes the editing site, as well as an enzyme, occasionally with other accessory factors, that catalyzes the reaction (Bass 2002). The functions of RNA editing include regulation of gene expression, increasing protein diversity and reversion of the effect of mutations in the genome (Gott and Emeson 2000; Gott 2003). In many cases, RNA editing is essential for correcting protein production (Young 1990) or for modulating the functional properties of proteins encoded by a single RNA (Seeburg *et al.* 1998). RNA editing can be regulated in a developmental or tissue-specific manner, and although editing mechanisms can be similar in different organisms, the editing patterns vary considerably between species (Gott 2003). The fact that editing may vary with tissue, developmental,

Copyright © 2014 by the Genetics Society of America

doi: 10.1534/genetics.114.161414

Manuscript received November 5, 2013; accepted for publication January 22, 2014; published Early Online February 4, 2014.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161414/-/DC1>.

¹These authors contributed equally to this work.

²Corresponding authors: No. 151, Malianwa North Rd., HaiDian District, Beijing 100193, China. E-mail: hmluo@implad.ac.cn; and No. 16, Dongzhimenneinanxiaojie, Dongcheng District, Beijing, 100700, China. E-mail: chenshilin@cmm.ac.cn.

and environmental conditions makes it extremely difficult to effectively screen for systematic editing events at the genomic level. Previously, high-throughput systematic screening techniques, such as complementary DNA (cDNA) sequencing or primer extension (Iwamoto *et al.* 2005), were prohibitively expensive, labor intensive, or lacked sensitivity. Now, the next generation of high-throughput methods, including RNA-Seq technology, allows quantification of editing and large-scale scanning of transcripts for new editing sites without any prior knowledge of their nature or location (Sasaki *et al.* 2006; Park *et al.* 2012).

RNA-Seq is an effective and accurate high-throughput method for transcriptome profiling and offers increased sensitivity to detect rare transcripts and abundant transcripts with superior discrimination (Graveley 2008; Marioni *et al.* 2008; Shendure 2008). The highly accurate measurement of transcript abundance by RNA-Seq can also be used to discover novel transcripts and to identify alternative splicing isoforms and RNA-editing events (Graveley 2008; Shendure 2008; Bahn *et al.* 2012). In particular, the depth of sequence read coverage per reference nucleotide enhances the qualities of base calling and can improve the large-scale detection of RNA-editing sites (Bahn *et al.* 2012; Peng *et al.* 2012; Lagarrigue *et al.* 2013; Ramaswami *et al.* 2013). Although the alignment of RNA-Seq reads to a reference genome can infer RNA-editing events, distinguishing these from genome-encoded single nucleotide polymorphisms (SNPs) and technical artifacts caused by sequencing or read-mapping errors is still the major challenge in identifying RNA-editing sites using RNA-Seq data (Ramaswami *et al.* 2013). Recently, unbiased and in-depth strategies for revealing editing sites in transcripts corresponding to coding, noncoding, and small RNA genes have been developed and applied in both humans and *Drosophila* (Peng *et al.* 2012; Ramaswami *et al.* 2013). These strategies are beginning to unravel RNA-editing events and its implications for the transcriptome.

Ganoderma lucidum, belonging to the Basidiomycota, is a widely used medicinal fungus that possesses multiple therapeutic activities, including antitumor, antiviral, and immunomodulatory activities (Boh *et al.* 2007). More than 400 different compounds have been identified in *G. lucidum* (Shiao 2003), and the secondary metabolites of triterpenoids and polysaccharides comprise the major pharmacologically active ingredients. Therefore, *G. lucidum* is a model organism for understanding the complexity of pathways controlling a diverse set of bioactive secondary metabolites in fungi (Chen *et al.* 2012). In addition, *G. lucidum*, like other white rot Basidiomycetes, produces enzymes that can effectively degrade both lignin and cellulose (Ko *et al.* 2001). The transcriptome of *G. lucidum* fruiting bodies, which are the major medicinal fungal structures used in traditional Chinese medicine, is well annotated and functionally characterized (Luo *et al.* 2010; Chen *et al.* 2012; Yu *et al.* 2012). Therefore, characterization of RNA editing in fruiting bodies will greatly augment our understanding of genetic and metabolic diversity and regulation in *G. lucidum*.

In this study, we combined data from different sequencing platforms and used software to distinguish RNA-editing sites from genomic variant sites to improve the accuracy of RNA-editing prediction (Peng *et al.* 2012). To our knowledge, this represents the most comprehensive identification and analysis of RNA editing in Basidiomycetes. Our findings reveal genome-wide occurrence of RNA editing in *G. lucidum* and highlight the importance of RNA editing in environment adaptation and secondary metabolite biosynthesis and regulation in this valuable medicinal fungus.

Materials and Methods

Transcriptome sequencing of the fruiting bodies of G. lucidum

G. lucidum dikaryotic strain CGMCC5.00226 was used in our study, and the transcriptome data were obtained from GenBank (Chen *et al.* 2012) (see [Supporting Information, Table S1](#)). For Roche 454 RNA-Seq library preparation, ~2 µg of total RNA from fruiting bodies was isolated by using the miRNeasy kit (Qiagen) according to the manufacturer's protocol. RNA was converted into cDNA by using a modified SMART cDNA synthesis protocol (Clontech) similar to that used in our previous study (Sun *et al.* 2010). The cDNA samples were prepared and sequenced by using the Roche 454 GS FLX platform according to the manufacturer's specifications (Roche).

For Illumina RNA-Seq library preparation, the total RNA and poly(A)⁺ RNA were isolated from the fruiting bodies of *G. lucidum* by using the RNeasy Plus Mini Kit (Qiagen) and the Poly(A) Purist Kit (Life Technologies), respectively. RNA-Seq was performed as recommended by the manufacturer (Illumina).

Diploid genome resequencing and mapping

Genomic DNA from the dikaryotic strain CGMCC5.0026 of *G. lucidum*, the source of the sequenced monokaryotic strain G.260125-1 created by protoplasting, was sequenced by using the Roche 454 GS FLX (Roche) sequencing platform (Chen *et al.* 2012). Diploid whole genome resequencing (DWGR) data were obtained. Genome mapping and detection of heterozygous loci were performed by using SSAHASNP version 2.5.3 (Ning *et al.* 2001).

Read mapping and single nucleotide variant detection

To predict the RNA-editing sites at high sequencing depth, we first used Illumina GA IIx sequencing data (RNASeq-Illumina), followed by the Roche 454 FLX+ sequencing data (RNASeq-454). Before the reads were mapped, we executed a strict filter for RNASeq-Illumina sequences. Duplicated reads were filtered with 100% similarity and were considered to be generated from PCR duplication. Afterward, the short paired-end reads were aligned to the *G. lucidum* haploid genome by using the SOAP2 program (Li *et al.* 2009b). Single nucleotide variants (SNVs) were identified by using

SOAPsnp (Li *et al.* 2009a). In addition, long RNASeq-454 reads were also mapped to the *G. lucidum* haploid genome with SSAHASNP and default parameters (Ning *et al.* 2001). To avoid indel sequencing error from 454 reads, only a single base variant with two base substitutions was used for subsequent analysis.

Identification of RNA-editing sites

We developed a pipeline to execute a “regular filter,” a “heterozygous loci filter,” and a “double-insurance filter.” Initially, identified SNVs were filtered by the following steps: (1) For the regular filter, a minimum of five supported reads was used to determine the editing sites and to eliminate sequencing errors. Reads mapping was unique to avoid repeat sequences and paralogous genes. The distance of a potential SNV to the end of the supporting read was determined (≥ 15 bp). The BLAT mapping was executed to further address the potential paralogous sequences. (2) For the heterozygous loci filter, DWGR data were used to exclude heterozygous loci called by using SSAHASNP following basic filtering. (3) For the double-insurance filter, RNASeq-454 data were used to confirm single nucleotide variants. SNVs were identified by using SSAHASNP via genome mapping. Only sites predicted to be common between the RNASeq-Illumina and RNASeq-454 were used for the analysis of RNA editing.

Validation of RNA-editing sites with Sanger sequencing

RNA and genomic DNA (gDNA) were extracted from the fruiting bodies of *G. lucidum*. Nucleic acid quality and quantity were assessed on 1% ethidium bromide-stained agarose gels by using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). The RNA was DNase-treated with TURBO DNase according to the manufacturer's recommendations (Ambion Inc.) at a concentration of 1.5 units/ μ g of total RNA and converted into cDNA by using a Prime-Script 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. Primers were designed around the editing sites located in the randomly selected genes. The PCR amplifications for these sequences were performed by using 30 ng of cDNA and gDNA as templates with the following program: 95° for 3 min, followed by 30 cycles (95° for 30 sec, 58° for 30 sec, and 72° for 2 min), and 72° for 10 min. PCR products were detected on an agarose gel (1.5%) stained in ethidium bromide and then extracted and purified by using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. The purified PCR amplicons were sequenced by Sanger sequencing.

Phylogenetic relationships of adenosine deaminases acting on RNA and adenosine deaminases acting on tRNA

The phylogenetic relationships of adenosine deaminases acting on RNA (ADARs) and adenosine deaminases acting on tRNA (ADATs) were reconstructed by using their amino

acid sequences. Alignments of the protein sequences were executed by using ClustalW 1.83. A maximum-likelihood tree was constructed with the Jones–Taylor–Thornton substitution model by using MEGA5 (Tamura *et al.* 2011). The search strategy of the nearest-neighbor interchange was used to improve the likelihood of the tree. Bootstrap value was set as 1000.

Protein target prediction

Target prediction was executed by using two web-based programs: Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>) (Small *et al.* 2004) and TargetP (<http://www.cbs.dtu.dk/services/TargetP>) (Emanuelsson *et al.* 2007). Both programs predict subcellular localization based on N-terminal amino acids by using default parameters.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis

The InterPro database was used to assign Gene Ontology (GO). The resulting GO terms were classified based on the Gene Ontology Database (<http://www.geneontology.org/>). We executed the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis by using the KO-Based Annotation System (Xie *et al.* 2011). A hypergeometric test was used for statistical testing; the Benjamini and Hochberg method was used to determine multiple testing correction by using the false discovery rate.

Prediction of RNA and protein structures

RNA secondary structures were predicted by using the Mfold web server (Zuker 2003) (<http://mfold.rna.albany.edu/?q=mfold>) and default parameters. Protein structures were predicted by Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2>) (Kelly and Sternberg 2009).

Results

Deep sequencing the transcriptome

To investigate RNA editing in *G. lucidum*, we used the RNA-Seq data sets of *G. lucidum* fruiting bodies obtained from Illumina GA IIx (SRX105469) and Roche 454 FLX+ (SRX113866) sequencing platforms (see Table S1) (Chen *et al.* 2012). The filtered RNASeq-Illumina data were equivalent to 56 $\times$ of the haploid *G. lucidum* genome (43.3 Mb) and 98 $\times$ of the predicted area of protein-coding genes (Chen *et al.* 2012). The RNASeq-454 data were used to confirm the prediction results and to improve the filtering efficiency. After filtering low-quality and duplicated reads, 605,884 high-quality 454-reads, averaging 418 bp and representing 253 Mb expressed sequences, were generated.

Identification of RNA-editing sites

To identify accurately the RNA-editing sites in *G. lucidum*, multiple filtering steps were performed by using the combined

Figure 1 Pipeline flowchart for calling RNA-editing sites. Raw inputs include RNA-Seq reads sequenced by 454 and Illumina and genomic 454-reads produced by diploid genome resequencing. These sequences were filtered and analyzed further in this pipeline.

sequence data to exclude genomic variants (Figure 1). A total of 27,586 SNVs were detected after initial filtering by using only the RNASeq-Illumina data set. Among these sites, a total of 6888 heterozygous loci were excluded based on resequencing sequence mapping. After filtering with the RNASeq-454 data set, we identified 8906 possible RNA-editing sites (see Table S2). Around 73% (6526) of these editing sites were mapped to gene models. Among these, the vast majority (6014; 68%) were predicted within protein-coding regions: 103 sites within introns, 96 sites within 5'-untranslated regions (UTRs), and 313 sites within 3'-UTRs (Figure 2). However, these sites were predicted in only 2991 genes (~19%) among the 16,113 gene models. The remaining editing sites (2380; 27%) were generated mainly from the UTRs of the upstream or downstream of coding genes.

Characterization of RNA-editing sites

The characteristics of *G. lucidum* RNA-editing sites were analyzed. Four primary RNA-editing types were predicted in *G. lucidum*, comprising the conversion from C to U, A to G, G to A, and U to C (Figure 3, A–D). The distribution of edits was fairly consistent across exons, introns, and 3'-UTRs (Figure 3, A, C, and D), whereas the 5'-UTR editing sites primarily consisted of C-to-U changes (Figure 3B). The editing degree was calculated as the ratio of uniquely mapped RNA-Seq reads showing the edited nucleotide, relative to those showing the genome-encoded nucleotide, at a single editing position. In our analysis, the average editing degree was 42%. As shown in Figure 3, E and F, the editing degree most often varied between 40 and 50%. In the protein-coding regions, the codon preference of the predicted edited nucleotide was examined and was found to exist for the third codon position (Figure 4A). The ratio of nonsynonymous and

Figure 2 Distribution of RNA-editing loci in fruiting bodies.

synonymous amino-acid substitutions showed a significant difference between C to U, U to C, A to C, and A to U (Figure 4B).

Contextual analysis of editing sites

The sequence context of the primary RNA-editing sites in the *G. lucidum* genome was analyzed to discover potential motifs that may determine the interaction with the RNA-editing enzyme complexes. The neighboring bases (–20 to +20) of each editing site were analyzed by using the MEME software. The motifs enriched for editing sites were identified irrespective of genomic location. Four motifs, all 15 bp in length, were identified for each of the four primary editing types, with significantly low *E*-values (<1e–150) (see Figure S1, A–D). The two motifs associated with purine transitions (A to G and G to A) had similar structures (see Figure S1, B and C). Likewise, the two motifs overrepresented for the pyrimidine transitions (C to U and U to C) were similar (see Figure S1, A and D). In addition, the general nucleotide composition of the flanking regions of the editing sites showed overrepresentation for the G nucleotides flanking the pyrimidine transition edits (see Figure S2, A and B) and the C nucleotides flanking the purine transitions (see Figure S2, C and D).

Putative RNA-editing enzymes in *G. lucidum*

Three ADATs were identified and named *GLADAT1*, *GLADAT2*, and *GLADAT3* on the basis of homology with known ADATs. Phylogenetic analysis showed the nearest relationship of GLADATs with ADATs from *Saccharomyces cerevisiae* (Figure 5). Structural analysis of the ADAT proteins in *G. lucidum* revealed that each protein contained at least one deaminase domain, but was missing a double-stranded RNA-binding domain. GLADAT1 (GL18117) alone contained an adenosine deaminase domain, whereas GLADAT2 (GL26770) contained a cytidine deaminase domain. In contrast, GLADAT3 (GL19633) contained a deoxycytidylate deaminase zinc-binding region, suggesting different functions of GLADATs. In addition, the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like

Figure 3 RNA-editing types and RNA-editing degree in different genomic regions. (A–D) Distribution of RNA-editing types on the gene models, including CDS (A), 5′-UTR (B), 3′-UTR (C), and introns (D). (E–F) Distribution of RNA-editing levels on the entire genome (E) and protein-coding regions (F).

(APOBEC) cytidine deaminases, were also identified in *G. lucidum*. Interestingly, the APOBEC domain was found in tRNA-specific adenosine deaminase (GLADAT2) and deoxycytidylate deaminase (GL30624).

In plant organelles, RNA editing commonly involves the substitution of C to U, or less commonly, U to C (Gray 1996; Shikanai 2006; Castandet and Araya 2011). Pentatricopeptide repeat (PPR) proteins, which form a large protein family, have been identified as site-specific factors for recognizing RNA-editing sites in plant organelles (Zehrmann *et al.* 2011). In this study, PPR proteins were also identified in the *G. lucidum* genome by searching the tandem PPR motifs against the Pfam and InterPro databases by using the hidden Markov model method. We identified seven PPR genes, in which each gene included at least one PPR motif (see Figure S3). These predicted genes were not structurally similar; however, two (GL29438 and GL23468) were predicted to target mitochondria using both the Predator and the TargetP software (see Table S3 and Table S4).

Validation of editing sites

To validate the RNA-editing predictions in *G. lucidum*, a selection of transcripts containing the predicted editing sites were amplified and sequenced by using the Sanger method (Figure 6). Overall, 94 of 97 sequenced sites showed clear signals with double peaks exhibiting the predicted base variation (see Table S5 and Table S6). Among these sites, four editing types were validated for low, medium, and high editing degrees (see Figure S4). This low false rate of 3.1% (3/97) validates the effectiveness and accuracy of our RNA-editing prediction pipeline. Moreover, editing sites with a low editing degree were also validated, suggesting that our pipeline has high sensitivity. Meanwhile, 33 new sites that were not predicted in the RNA-Seq data were found by Sanger sequencing. This finding may result from the high stringency of filtering and uncertain mapping. The four main editing types were validated in our analysis, as well as the less common editing sites, such as A to C, C to A, C to G, G to C, G to U, and U to G.

Figure 4 Influence of RNA editing on amino acid substitutions in the coding regions. (A) Frequency of codon change at three codon positions. (B) Proportion of nonsynonymous and synonymous substitution of amino acids for different editing types. RNA-editing types above the horizontal dashed line indicate that the proportion of nonsynonymous substitution and synonymous substitution is >1 .

Functional classification of genes subjected to RNA editing

An understanding of the genes and possible gene networks affected by RNA editing in *G. lucidum* may provide insight into the functional importance of this process in fungi. The 2991 gene models predicted to be targeted for RNA editing were subjected to GO and KEGG analysis. A total of 2700 edited genes were assigned to the KEGG pathways. KEGG enrichment analysis revealed the enrichment for 129 pathways, containing 617 genes, with a P -value of <1.0 compared with the related Basidiomycete, *Postia placenta* (brown rot fungus; see Table S7). However, only 33 pathways involving 351 genes were significantly enriched ($P < 0.05$). These pathways include multiple primary and secondary metabolism processes, such as pyruvate metabolism, fatty acid metabolism, terpenoid backbone biosynthesis, and porphyrin and chlorophyll metabolism (see Table S7). GO assignment was also analyzed, with genes being classified into one of the three GO categories: biological process, cellular component, and molecular function (see Figure S5). Within these categories, the represented genes were further assigned to secondary classifications, and metabolic process, binding, and catalytic activity functions were the most abundant.

RNA editing involved in wood degradation and transcriptional regulation

G. lucidum is a white rot Basidiomycete and thus can effectively decompose the complex carbohydrates cellulose and lignin in its plant host (Ko *et al.* 2001). Previously, 417 genes were annotated as carbohydrate-active enzymes in the *G. lucidum* genome (Chen *et al.* 2012). Among these genes, 98 contained a total of 269 RNA-editing sites (see Table S8). The genes involved in lignin digestion included laccases, ligninolytic peroxidases, and peroxide-generating oxidases (Ruiz-Duenas and Martinez 2009). A total of 46 genes encoding enzymes involved in lignin digestion exist in *G. lucidum*, and only two, namely, a multicopper oxidase (GL15267) and an alcohol oxidase (GL18144), contained RNA-editing sites (see Table S9).

In *G. lucidum*, 541 genes were annotated as transcription factors (TFs) (Chen *et al.* 2012). Among which, 97 TF genes contained a total of 164 editing sites, with 29 C2H2-

containing genes containing the most RNA-editing sites, followed by 15 helix-turn-helix-containing genes and 12 fungal family TFs (see Table S10).

RNA editing involved in bioactive secondary metabolite biosynthesis

Triterpenoids and polysaccharides are the major secondary metabolites in *G. lucidum*. A total of 13 genes in *G. lucidum* putatively encode 11 enzymes involved in the upstream mevalonate (MVA) pathway of triterpenoid biosynthesis, whereas 16 putative cytochrome P450 genes (CYPs) perform downstream oxidation and reduction reactions (Chen *et al.* 2012). Among which, 8 genes encoding seven upstream enzymes (see Table S11 and Figure S6) and 53 genes encoding CYPs contained RNA-editing sites (see Table S12). The 3-hydroxy-3-methylglutaryl-CoA synthase gene (*GHMGS*) contained the highest number of editing sites in the upstream MVA pathway, in which six sites are in the CDS and one is in the 3'-UTR. The 53 CYPs contained a total of 135 RNA-editing sites, and 10 of these genes were co-expressed with *Lanosterol synthase*, which is required for the production of the triterpenoid precursor, lanosterol. These 10 genes were also phylogenetically related to the genes involved in the hydroxylation of testosterone in *Phanerochaete chrysosporium* and *P. placenta* (Chen *et al.* 2012) (see Table S12).

A total of 14 genes encoding 10 enzymes that were predicted to be involved in polysaccharide biosynthesis were found to contain 51 RNA-editing sites (see Table S13). Two genes (*GL20535* and *GL24465*) encode 1,3- β -glucan synthase, which has a key role in the biosynthesis of 1,6- β -glucans in *S. cerevisiae* (Shahinian and Bussey 2000) (see Table S13).

Potential impact of RNA editing

RNA editing led to the variation of a single nucleotide in this fungus and further affected the RNA and protein structures. Phosphoglucomutase is important in polysaccharide biosynthesis. Two adenines were changed into guanines because of RNA editing. This effect increased the matching bases in the stem region of the RNA secondary structure and decreased the free energy (Figure 7A). In addition, RNA

Figure 5 Phylogenetic and domain analysis of ADATs and ADARs. (A) Unrooted tree of ADATs and ADARs. The number of bootstrap replications was set as 1000 for testing the confidence level of the phylogenetic tree. Only bootstrap values >50 are shown. (B) A schematic of ADAR and ADAT enzymes. The domain structures were highlighted by using different colors. dsRBD: double-stranded RNA-binding domain (PF00035); A_deamin: adenosine deaminase (PF02137); z-alpha: adenosine deaminase z-alpha domain (PF02295); dCMP_cyt_deam_1: cytidine and deoxycytidylate deaminase zinc-binding region (PF00383).

editing results in nonsynonymous protein-coding substitution. However, the protein secondary structure has not been significantly changed in this phosphoglucomutase (Figure 7B). Mevalonate kinase was a key enzyme involved in the MVA pathway. RNA editing (C to U) opened the matched bases in the RNA secondary structure. When the RNA secondary structure was changed in the local region, two small loops formed a large loop (Figure 7C). In *G. lucidum*, a total of 48 start codons were changed, and 18 stop codons were identified in the edited genes, suggesting new protein products (see Table S14).

Discussion

This study reports the first high-throughput analysis of RNA editing in a fungal species to date. A pipeline was developed for high-throughput, next generation (NG) analysis of RNA editing, incorporating multiple filtering steps and integration of data generated from different NG sequencing platforms. Basing on the analysis of the *G. lucidum* genome, we demonstrated that the RNA-editing events are abundant and selective in the fungal genome.

Several previous studies have used RNA-Seq data for transcriptome-wide identification of editing sites in several organisms (Gott 2003; Peng *et al.* 2012); however, global profiles of RNA-editing events in fungi are limited. In this study, we identified 8906 RNA-editing sites in the fruiting bodies of *G. lucidum* by using a combination of RNA-Seq data and genome data. Unlike in other organisms, most of the RNA-editing sites in humans are located in intronic regions and untranslated regions of the coding genes; however, in *G. lucidum* they were located mainly in the coding regions (Peng *et al.* 2012). The differentiation between the

RNA-editing sites and genomic variants obtained by using RNA-Seq data is one of the major challenges of accurate RNA-editing prediction. Another challenge facing this type of analysis is the discrimination between technical artifacts caused by sequencing or read-mapping errors in the identification of RNA-editing sites. In this study, the diploid *G. lucidum* genome was resequenced by using the Roche 454 platform and used to filter genomic variants effectively. Several studies have suggested that the deep sequencing of both the transcriptome and the genome can minimize erroneous variant calls caused by sequencing or read-mapping errors (Bahn *et al.* 2012; Peng *et al.* 2012). Therefore, we also performed Roche 454 transcriptome sequencing, which can generate very long reads of high accuracy relative to other platforms (Balzer *et al.* 2010). The integration of these data with Illumina RNA-Seq data, followed by multiple filtering steps, provided high accuracy of RNA-editing site identification. Validation of a number of randomly selected sites in the genome revealed 96.9% accuracy of RNA-editing prediction by using this pipeline (see Table S5).

Four single-nucleotide changes composed the majority of the RNA-editing types in *G. lucidum*, which include the transitions (1) C to U (cytosine to uracil), (2) A to G (adenine to guanine), (3) G to A, and (4) U to C (Figure 3). Among these transitions, although U-to-C and G-to-A changes were known, the preferences for these changes were first reported. These reports are not entirely consistent with previous findings in animals or plants, which show a preference for A-to-I (adenine to inosine) changes (Bahn *et al.* 2012; Peng *et al.* 2012) and C-to-U changes (Shikanai 2006; Picardi *et al.* 2010), respectively. Four 15-bp motifs for purine transitions and pyrimidine transitions were identified, which were inconsistent with previous reports, indicating the diversity in the

Figure 6 Validation of predicted RNA-editing sites by Sanger sequencing. (A-D) The chromatogram traces of both strands, which were generated from gDNA and cDNA, are shown. The blue vertical line indicates editing sites on both strands.

recognition of different substrates and the specificity in the recognition of conserved motif for RNA-editing enzyme complexes (Bahn *et al.* 2012). On the basis of the four major editing transitions in *G. lucidum*, the putative editing enzymes involved were identified in the genome. The editing of A to I has been extensively studied in animals and is mediated by ADAR enzymes (Nishikura 2010). To date, no ADAR candidates have been found in fungi. Nonetheless, ADAR has been thought to have evolved from ADATs, which have been identified from almost all eukaryotes (Gerber and Keller 2001; Keegan *et al.* 2004). In *S. cerevisiae*, three ADATs were identified involving A-to-I and C-to-U deamination (Gerber *et al.* 1998; Gerber and Keller 1999). A total of three putative *GLADATs* were discovered in the *G. lucidum* genome on the basis of homology analysis and domain structure analysis (Figure 5B). Compared with ADATs in *S. cerevisiae*, *GLADAT1* and *Tad1p* have a common domain, that is, the adenosine deaminase domain (Figure 5B), which has a key role in A-to-I changes. Both *GLADAT2* and *Tad2p* have one cytidine deaminase domain (Figure 5B), which is indicative of cytidine deaminase activity and a function in the RNA editing of C to U. Similarly, *GLADAT3* and *Tad3p* contain a deoxycytidylate

deaminase zinc-binding region (Figure 5B), which is functionally annotated as being able to hydrolyze dCMP into dUMP.

Cytosine-to-uracil editing has been studied in relation to the APOBEC cytidine deaminases in mammals (Harris *et al.* 2002) and PPR proteins in plant organelles (Uchida *et al.* 2011). Editing of the apoB mRNA is regarded as a model system for C-to-U editing (Wedekind *et al.* 2003) and translates as a glutamine (CAA) to translational stop codon (UAA) change in the apoB protein (Smith and Sowden 1996; Wedekind *et al.* 2003). C-to-U editing is mediated by a minimal complex composed of apoB-editing catalytic subunit 1 (APOBEC1) and a complementing specificity factor (Chester *et al.* 2003). The APOBEC1 subunit contains a zinc-dependent deaminase domain, which is conserved among cytidine deaminases (Mian *et al.* 1998) and is essential for cytidine deamination. A gene, *GL30624*, encodes a deoxycytidylate deaminase containing an APOBEC domain, suggesting that this gene may function in RNA editing in *G. lucidum*. In addition, seven *PPR* genes were identified in the *G. lucidum* genome, and two (*GL29438* and *GL23468*) were putatively targeted to mitochondria.

Figure 7 Comparison of RNA and protein secondary structures for RNA-editing transcripts. (A) RNA secondary structure of phosphoglucomutase involved in polysaccharide biosynthesis. RNA-editing sites were located at 306,124 and 306,136 of GaLu96scf_34. Minimum free energy was shown as the value of "dG." Blue letters and arrows represent original bases or amino acids. Red letters and arrows represent edited bases or amino acids. (B) Protein secondary structure of phosphoglucomutase. For each protein, the α -helix (Alpha helix) represents protein secondary structures; the question mark represents disordered/unstructured amino acids; "confidence" refers to the confidence of secondary structure and disordered/unstructured amino acids. (C) The change of RNA secondary structure of mevalonate kinase in the mevalonate pathway. RNA-editing sites were located at 1,010,080 of GaLu96scf_8.

This finding suggests that these two PPR proteins may function similarly to plant PPRs, which are involved in RNA editing in plant mitochondria.

In general, RNA-editing events that change mRNA sequences result in the production of altered proteins (Gott and Emeson 2000). The creation of new start and stop codons by RNA editing (e.g., C to U) has been observed in many organisms (Stuart *et al.* 1997; Steinhäuser *et al.* 1999). A total of 66 RNA-editing sites predicted to alter start or stop codons were identified in the aforementioned gene classes, with potential to produce new protein products in the *G. lucidum* fruiting bodies. Alterations within 5'- and 3'-UTRs by RNA editing are rare and may affect mRNA stability, translatability, and processing (Gott and Emeson 2000). Consistent with this finding, only 409 editing sites (5%) were located within 5'- and 3'-UTRs in *G. lucidum*. KEGG-enrichment analysis and GO assignment demonstrated that RNA editing was involved in the diverse metabolic processes. Furthermore, genes involved in transcriptional regulation, lignin degradation, and bioactive compound biosynthesis were discovered to contain RNA-editing sites

in our study (see Table S8, Table S9, Table S10, Table S11, Table S12, and Table S13). This result may translate to a diverse molecular base for the regulation of growth and development in *G. lucidum* and its adaptation to the environment.

In conclusion, we revealed the abundance, nucleotide preference, and specificity of the genome-wide RNA-editing events and the putative RNA-editing enzymes involved in an important medicinal fungus, *G. lucidum*. The pipeline developed for this analysis was highly accurate and provides improved detection of reliable editing sites in *G. lucidum*. The GO analysis and KEGG assignment of genes containing RNA-editing sites provide a global profile of RNA-editing events in *G. lucidum* fruiting bodies with functional implications. The identification of RNA-editing sites in genes involved in transcriptional regulation, bioactive compound biosynthesis, and adaptation to the environment may have significant value in enriching the translated genomic repertoire without a concomitant increase in genome size. This finding may assist in uncovering new dimensions of molecular genetic plasticity in the important medicinal fungus, *G. lucidum*.

Acknowledgments

This work was supported by the Key National Natural Science Foundation of China (grant no. 81130069 and 81102761), the National Key Technology R&D Program (grant no. 2012BAI29B01), and the Program for Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (grant no. IRT1150).

Literature Cited

- Bahn, J. H., J. H. Lee, G. Li, C. Greer, G. Peng *et al.*, 2012 Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* 22: 142–150.
- Balzer, S., K. Malde, A. Lanzen, A. Sharma, and I. Jonassen, 2010 Characteristics of 454 pyrosequencing data: enabling realistic simulation with flowsim. *Bioinformatics* 26: i420–i425.
- Bass, B. L., 2002 RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71: 817–846.
- Boh, B., M. Berovic, J. Zhang, and L. Zhi-Bin, 2007 *Ganoderma lucidum* and its pharmaceutically active compounds. *Biotechnol. Annu. Rev.* 13: 265–301.
- Castandet, B., and A. Araya, 2011 RNA editing in plant organelles. Why make it easy? *Biochemistry (Mosc)* 76: 924–931.
- Chen, S., J. Xu, C. Liu, Y. Zhu, D. R. Nelson *et al.*, 2012 Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nat. Commun.* 3: 913.
- Chester, A., A. Somasekaram, M. Tzimina, A. Jarmuz, J. Gisbourne *et al.*, 2003 The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J.* 22: 3971–3982.
- Emanuelsson, O., S. Brunak, G. Von Heijne, and H. Nielsen, 2007 Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2: 953–971.
- Farajollahi, S., and S. Maas, 2010 Molecular diversity through RNA editing: a balancing act. *Trends Genet.* 26: 221–230.
- Gerber, A. P., and W. Keller, 1999 An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* 286: 1146–1149.
- Gerber, A. P., and W. Keller, 2001 RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* 26: 376–384.
- Gerber, A. P., H. Grosjean, T. Melcher, and W. Keller, 1998 Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J.* 17: 4780–4789.
- Gott, J. M., 2003 Expanding genome capacity via RNA editing. *C. R. Biol.* 326: 901–908.
- Gott, J. M., and R. B. Emeson, 2000 Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* 34: 499–531.
- Graveley, B. R., 2008 Molecular biology: power sequencing. *Nature* 453: 1197–1198.
- Gray, M. W., 1996 RNA editing in plant organelles: a fertile field. *Proc. Natl. Acad. Sci. USA* 93: 8157–8159.
- Harris, R. S., S. K. Petersen-Mahrt, and M. S. Neuberger, 2002 RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* 10: 1247–1253.
- Iwamoto, K., M. Bundo, and T. Kato, 2005 Estimating RNA editing efficiency of five editing sites in the serotonin 2C receptor by pyrosequencing. *RNA* 11: 1596–1603.
- Kawahara, Y., B. Zinshteyn, P. Sethupathy, H. Iizasa, A. G. Hatzigeorgiou *et al.*, 2007 Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315: 1137–1140.
- Keegan, L. P., A. Leroy, D. Sproul, and M. A. O’Connell, 2004 Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. *Genome Biol.* 5: 209.
- Kelly, L. A., and M. J. E. Sternberg, 2009 Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4: 363–371.
- Ko, E. M., Y. E. Leem, and H. T. Choi, 2001 Purification and characterization of laccase isozymes from the white-rot basidiomycete *Ganoderma lucidum*. *Appl. Microbiol. Biotechnol.* 57: 98–102.
- Lagarigue, S., F. Hormozdiari, L. J. Martin, F. Lecerf, Y. Hasin *et al.*, 2013 Limited RNA editing in exons of mouse liver and adipose. *Genetics* 193: 1107–1115.
- Li, R., Y. Li, X. Fang, H. Yang, J. Wang *et al.*, 2009a SNP detection for massively parallel whole-genome resequencing. *Genome Res.* 19: 1124–1132.
- Li, R., C. Yu, Y. Li, T. W. Lam, S. M. Yiu *et al.*, 2009b SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966–1967.
- Luo, H., C. Sun, J. Song, J. Lan, Y. Li *et al.*, 2010 Generation and analysis of expressed sequence tags from a cDNA library of the fruiting body of *Ganoderma lucidum*. *Chin. Med.* 5: 9.
- Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad, 2008 RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18: 1509–1517.
- Mian, I. S., M. J. Moser, W. R. Holley, and A. Chatterjee, 1998 Statistical modelling and phylogenetic analysis of a deaminase domain. *J. Comput. Biol.* 5: 57–72.
- Ning, Z., A. J. Cox, and J. C. Mullikin, 2001 SSAHA: a fast search method for large DNA databases. *Genome Res.* 11: 1725–1729.
- Nishikura, K., 2010 Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79: 321–349.
- Park, E., B. Williams, B. J. Wold, and A. Mortazavi, 2012 RNA editing in the human ENCODE RNA-seq data. *Genome Res.* 22: 1626–1633.
- Peng, Z., Y. Cheng, B. C. Tan, L. Kang, Z. Tian *et al.*, 2012 Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* 30: 253–260.
- Picardi, E., D. S. Horner, M. Chiara, R. Schiavon, G. Valle *et al.*, 2010 Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Res.* 38: 4755–4767.
- Ramaswami, G., R. Zhang, R. Piskol, L. P. Keegan, P. Deng *et al.*, 2013 Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* 10: 128–132.
- Ruiz-Duenas, F. J., and A. T. Martinez, 2009 Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* 2: 164–177.
- Sasaki, T., Y. Yukawa, T. Wakasugi, K. Yamada, and M. Sugiura, 2006 A simple in vitro RNA editing assay for chloroplast transcripts using fluorescent dideoxynucleotides: distinct types of sequence elements required for editing of ndh transcripts. *Plant J.* 47: 802–810.
- Seeburg, P. H., M. Higuchi, and R. Sprengel, 1998 RNA editing of brain glutamate receptor channels: mechanism and physiology. *Brain Res. Brain Res. Rev.* 26: 217–229.
- Shahinian, S., and H. Bussey, 2000 beta-1,6-Glucan synthesis in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 35: 477–489.
- Shendure, J., 2008 The beginning of the end for microarrays? *Nat. Methods* 5: 585–587.
- Shiao, M. S., 2003 Natural products of the medicinal fungus *Ganoderma lucidum*: occurrence, biological activities, and pharmacological functions. *Chem. Rec.* 3: 172–180.

- Shikanai, T., 2006 RNA editing in plant organelles: machinery, physiological function and evolution. *Cell. Mol. Life Sci.* 63: 698–708.
- Small, I., N. Peeters, F. Legeai, and C. Lurin, 2004 Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4: 1581–1590.
- Smith, H. C., and M. P. Sowden, 1996 Base-modification mRNA editing through deamination: the good, the bad and the unregulated. *Trends Genet.* 12: 418–424.
- Steinhauser, S., S. Beckert, I. Capesius, O. Malek, and V. Knoop, 1999 Plant mitochondrial RNA editing. *J. Mol. Evol.* 48: 303–312.
- Stuart, K., T. E. Allen, S. Heidmann, and S. D. Seiwert, 1997 RNA editing in kinetoplastid protozoa. *Microbiol. Mol. Biol. Rev.* 61: 105–120.
- Sun, C., Y. Li, Q. Wu, H. Luo, Y. Sun *et al.*, 2010 De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics* 11: 262.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei *et al.*, 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.
- Uchida, M., S. Ohtani, M. Ichinose, C. Sugita, and M. Sugita, 2011 The PPR-DYW proteins are required for RNA editing of *rps14*, *cox1* and *nad5* transcripts in *Physcomitrella patens* mitochondria. *FEBS Lett.* 585: 2367–2371.
- Wedekind, J. E., G. S. Dance, M. P. Sowden, and H. C. Smith, 2003 Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.* 19: 207–216.
- Xie, C., X. Mao, J. Huang, Y. Ding, J. Wu *et al.*, 2011 KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39: W316–W322.
- Young, S. G., 1990 Recent progress in understanding apolipoprotein B. *Circulation* 82: 1574–1594.
- Yu, G. J., M. Wang, J. Huang, Y. L. Yin, Y. J. Chen *et al.*, 2012 Deep insight into the *Ganoderma lucidum* by comprehensive analysis of its transcriptome. *PLoS ONE* 7: e44031.
- Zehrmann, A., D. Verbitskiy, B. Hartel, A. Brennicke, and M. Takenaka, 2011 PPR proteins network as site-specific RNA editing factors in plant organelles. *RNA Biol.* 8: 67–70.
- Zuker, M., 2003 Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31: 3406–3415.

Communicating editor: E. U. Selker

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161414/-/DC1>

Abundant and Selective RNA-Editing Events in the Medicinal Mushroom *Ganoderma lucidum*

Yingjie Zhu, Hongmei Luo, Xin Zhang, Jingyuan Song, Chao Sun, Aijia Ji,
Jiang Xu, and Shilin Chen

Figure S1 Consensus motifs identified by MEME software in 41 neighboring bases centered around the four main editing types: C to U (A), A to G (B), G to A (C), and U to C (D). The position of each base in this motif is shown on the horizontal ordinate. Vertical ordinate shows the probability of each base occurring at the specific position

Figure S2 Nucleotide composition of flanking regions (20 bp both upstream and downstream) of the four main RNA editing types: C to U (A), U to C (B), A to G (C), and G to A (D).

Figure S3 Domain organization of seven PPR proteins. PPR domains in red were identified from the Pfam database by using Pfamscan.

Figure S4 Validation of RNA editing sites for different editing degrees. Four main types of RNA editing were listed at different editing degrees, including A) low (0% to 5%), B) medium (30% to 60%), and C) high (80% to 100%) editing degrees.

Figure S5 Gene ontology (GO) classification of genes affected by RNA editing. GO terms were obtained according to InterPro ID. These genes were classified into cellular component, molecular function, and biological process, as well as their subclasses by using GO databases.

Figure S6 RNA editing of genes involved in the mevalonate pathway. Each asterisk represents one editing locus on this gene. A total of 135 loci were subjected to RNA editing in CYPs.

Tables S1-S14

Available for download as Excel files at

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161414/-/DC1>

Table S1 List of data stored in GenBank and used in this paper.

Table S2 List of RNA editing sites from fruiting bodies identified in this study. Sequencing depth was generated from RNASeq-Illumina datasets.

Table S3 Target prediction of seven PPR proteins using TargetP v1.0.

Table S4 Target prediction of seven PPR proteins using Predotar v1.03.

Table S5 PCR validation of predicted RNA editing sites based on Sanger sequencing. Validated results were shown as three types in this table: 'TRUE' indicates a true discovery, 'FALSE' indicates a false prediction, 'Not predicted' indicates that a true locus was not predicted.

Table S6 PCR primers used for validation of inferred RNA editing sites.

Table S7 KEGG enrichment results of genes containing RNA editing sites.

Table S8 List of RNA editing genes identified as CAZy.

Table S9 List of editing sites identified in genes involved in lignin degradation.

Table S10 List of editing sites identified in genes involved in transcriptional regulation.

Table S11 List of editing sites identified in genes involved in MVA pathway.

Table S12 List of editing sites identified in CYP genes.

Table S13 List of editing sites identified in genes involved in polysaccharide biosynthesis.

Table S14 Changes for start codons and generation for stop codons.