Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 341598, 7 pages
http://dx.doi.org/10.1155/2015/341598

Research Article

Genome-Wide Identification, Evolutionary, and Expression Analyses of Histone H3 Variants in Plants

Jinteng Cui,[1] Zhanlu Zhang,[2] Yang Shao,[1] Kezhong Zhang,[1] Pingsheng Leng,[1] and Zhe Liang[3]

[1]*College of Landscape, Beijing University of Agricultural, Beijing 102206, China*
[2]*Shum Yip Group Limited, Shenzhen 518040, China*
[3]*Department of Biological Sciences, National University of Singapore, Singapore 117543*

Correspondence should be addressed to Zhe Liang; zhe.liang@nus.edu.sg

Received 9 January 2015; Accepted 5 February 2015

Academic Editor: Andrey Cherstvy

Histone variants alter the nucleosome structure and play important roles in chromosome segregation, transcription, DNA repair, and sperm compaction. Histone H3 is encoded by many genes in most eukaryotic species and is the histone that contains the largest variety of posttranslational modifications. Compared with the metazoan H3 variants, little is known about the complex evolutionary history of H3 variants proteins in plants. Here, we study the identification, evolutionary, and expression analyses of histone H3 variants from genomes in major branches in the plant tree of life. Firstly we identified all the histone three related (HTR) genes from the examined genomes, then we classified the four groups variants: centromeric H3, H3.1, H3.3 and H3-like, by phylogenetic analysis, intron information, and alignment. We further demonstrated that the H3 variants have evolved under strong purifying selection, indicating the conservation of HTR proteins. Expression analysis revealed that the HTR has a wide expression profile in maize and rice development and plays important roles in development.

1. Introduction

Histones wrap DNA to form nucleosome particles that compact eukaryotic genomes [1]. Histone variants have evolved crucial roles in chromosome segregation, transcriptional regulation, DNA repair, sperm packaging, and other processes [2]. Histone H3 is one of the five main histone proteins involved in the structure of chromatin in eukaryotes. Histone H3 variants genes can be divided into two major groups: replication dependent and replication independent [3]. The replication dependent variants are highly expressed just before S-phase and then repressed at the completion of DNA replication. The replication-independent histone variants are constitutively expressed throughout the cell cycle [4]. In human the replication dependent variants are represented by H3.1 and the replication-independent variants best characterized are H3.3 and centromeric H3 variants (CenH3) [4]. H3.1 is similar in length and amino acid sequence to H3.3 except at few positions. CenH3 variants differ from the other H3 variants by a long extension of the N-terminal tail, which are not conserved among eukaryotes [5, 6]. Histone chaperones are escort factors associated with soluble histones involved in their transfer without being part of the final product [4, 7].

In animals, four amino acid substitutions distinguish H3.1 featuring A_{31}-S_{87}-V_{89}-M_{90} from H3.3 featuring S_{31}-A_{87}-I_{89}-G_{90} [8]. In plants, H3.1 and H3.3 are distinguished at positions 31, 87, and 90 and involve a different set of amino acids, implying that H3 variants evolved independently in plants and animals [5, 9], In addition, position 41 is a plant-specific substitution that discriminates H3.3 from H3.1 variants [6, 9]. Phylogenetic analyses also showed that H3.1 and H3.3 variants evolved independently, suggesting that H3 variants in plants and animals are analogous and result from convergent evolution [10].

There are 15 HTR genes coding for H3.1, H3.3, and CenH3 proteins in *Arabidopsis thaliana* [9]; H3.1 and H3.3 proteins play strikingly different roles. H3.3 enrichment positions are positively correlated with gene expression and to be biased

towards the transcription termination site [11], which are common feature found in yeast, *Drosophila* and mammals [12–14]. The modification marks associated with transcriptional activity such as H3K27me and H3K36me are enriched in H3.3 in comparison to H3.1 [11].

Until now, genome-wide analyses of H3 variants have been conducted in *Arabidopsis* and several other plant species [6, 9, 11, 15]. However, comprehensive analyses of H3 variants in major plants are still lacking. Here, we studied the identification, evolutionary, and expression analyses of histone H3 variants. We searched for H3 variants in genomes representing a broad taxonomic sampling from distantly related plant evolutionary lineages, including eudicots, monocots, gymnosperm, lycophyte, bryophyte, and chlorophyte species. Subsequently, we classified the HTRs by phylogenetic analysis and alignment. We found the HTR genes in *Capsella rubella* might be a new class of H3 variants because of the unique sequence feature. Moreover, the selection and expression analysis suggested the functional conservation of H3 variants.

2. Methods

2.1. Sequence Retrieval. We performed BLASTP and TBLASTN searches among sequenced genomes of planta in Phytozome (http://www.Phytozome.net/) using *Arabidopsis* H3.1, H3.3, and CenH3 proteins as queries to identify HTRs (e-value $< e^{-10}$). The species represented a broad range of the plant lineages from unicellular green algae to multicellular plants (http://www.jgi.doe.gov/). We used a relatively strict criterion to collect HTRs with high-quality sequences. To verify the reliability of our results, each sequence was then searched against the protein conserved domain database (CDD) [16], SMART database [17], and the PFAM database [18], respectively. The detailed information (i.e., accession number, intron information) of all HTR sequences presented here is listed in Supplementary Table (see Supplementary Material available online at http://dx.doi.org/10.1155/2015/341598).

2.2. Construction of the Alignment. The HTR sequences were then aligned by MAFFT version 7 using the G-INS-i algorithm [19], followed by manual editing in MEGA 6.0 [20]. Only positions that unambiguously aligned were included in the further analyses. ProtTest 3.0 was used for amino acid substitution model selection using the Akaike Information Criterion (AIC) to choose the best-fitting tested model (LG + G) for phylogenetic analyses [21].

2.3. Phylogenetic Analyses. The neighbor joining (NJ) phylogeny was performed by MEGA version 6.0 with 1000 replicas; since LG + G is not available in MEGA, we use next available best model JJT + G model, and pairwise deletion. The maximum likelihood (ML) phylogeny was performed by PhyML software [22], with 100 replicas and the LG + G model.

2.4. Selection Analyses. Sequences used for selection analysis are listed in Supplementary Table. H3.1, H3.3, and CenH3 codon alignment was performed using MAFFT version 7 with the G-INS-i algorithm and then loaded into Hyphy [23] (along with a corresponding NJ phylogenetic tree). The HyPhy batch file NucModelCompare.bf with model rejection level of 0.0002 was used to establish the best fit of 203 general time-reversible (GTR) models of nucleotide substitution. The Hyphy batch file QuickSelectionDetection.bf was used to estimate site-by-site variation in rates.

2.5. Gene Expression Analysis. Maize and rice microarray-based datasets, with accession numbers GSE27004 [24] and GSE19024 [25], were downloaded from the NCBI Gene Expression Omnibus (GEO) [26]. A hierarchical cluster was created using the Cluster 3.0 [27] and viewed using the Java TreeView [28].

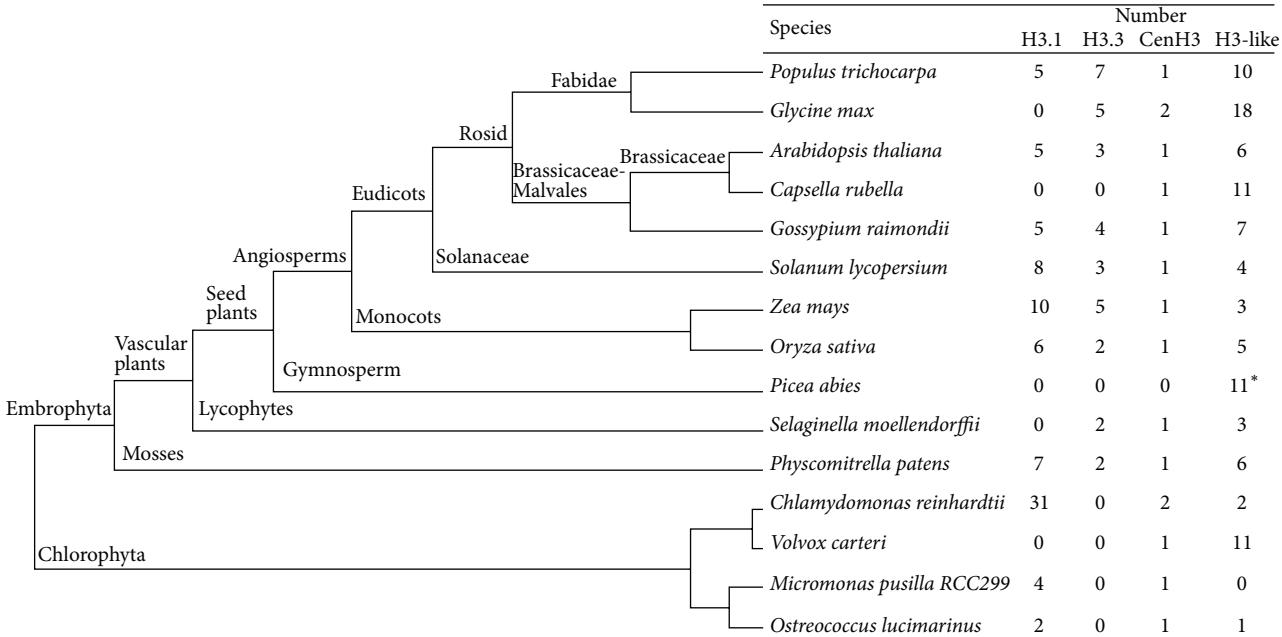
3. Result and Discussion

3.1. Identification of H3 Variants in Planta. To identify H3 variants in planta, we performed BLASTP and TBLASTN searches of the complete genomes of eudicots (*Arabidopsis thaliana*, *Capsella rubella*, *Populus trichocarpa*, *Glycine max*, and *Solanum lycopersicum*), monocots (*Zea mays* and *Oryza sativa*), gymnosperm (*Picea abies*), lycophyte (*Selaginella moellendorffii*), bryophyte moss (*Physcomitrella patens*), and chlorophytes (*Volvox carteri*, *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, and *Micromonas pusilla*) (Figure 1). Each matching sequence was then used to search the respective genome databases until no new sequences were found.

After removing incomplete or redundant sequences (nucleic acid sequence redundant) and predicted alternative splice variants, we identified 230 H3 variants (Supplementary Table). The numbers are variable between species (Figure 1). Interestingly, our analysis identified 33 HTR genes in *Chlamydomonas reinhardtii*, most of them were not previously annotated as HTR genes [6]; the variable number of HTR suggested multiple gene loss and gain events during planta evolution.

We first used phylogenetic analysis to define the different H3 groups present in the plant genomes; the phylogenetic tree contains all the H3 variants sequences from (Supplementary Table); the result showed that most of the branches have very low statistics supports (data not shown). However, the clade that contains *Arabidopsis* CenH3 AT1G01370 has good statistics supports, enabling us to identify all the CenH3 variants (Figure 3). All the species investigated have one or two CenH3 gene(s) except the gymnosperm *Picea abies* [29]. One possible explanation is that the recently sequenced Norway spruce genome still has some gaps. In order to discriminate H3.3 from H3.1 in planta, we used previously published criteria [6, 9]: the presence of introns in the H3.3 genes and absence of introns in the H3.1 genes and the four amino acid substitutions commonly found at positions 31, 41, 87, and 90 ($T_{31}Y_{41}H_{87}L_{90}$ for H3.3; $A_{31}F_{41}S_{87}A_{90}$ for H3.1).

In our analysis, all members of the H3.3 (Figure 2, Supplementary Table) contain introns and corresponding proteins carry the $T_{31}Y_{41}H_{87}L_{90}$ signature. Members of the H3.1 class are intronless genes and corresponding proteins carry the $A_{31}F_{41}S_{87}A_{90}$ signature (Figure 2, Supplementary Table). For



Species	Number H3.1	H3.3	CenH3	H3-like
Populus trichocarpa	5	7	1	10
Glycine max	0	5	2	18
Arabidopsis thaliana	5	3	1	6
Capsella rubella	0	0	1	11
Gossypium raimondii	5	4	1	7
Solanum lycopersium	8	3	1	4
Zea mays	10	5	1	3
Oryza sativa	6	2	1	5
Picea abies	0	0	0	11*
Selaginella moellendorffii	0	2	1	3
Physcomitrella patens	7	2	1	6
Chlamydomonas reinhardtii	31	0	2	2
Volvox carteri	0	0	1	11
Micromonas pusilla RCC299	4	0	1	0
Ostreococcus lucimarinus	2	0	1	1

Figure 1: Phylogenetic relationships between all species investigated in this study. The total number of histone three related (HTR) genes found in each genome is indicated on the right. Data was obtained from the phytozome (http://www.phytozome.net/) and Norway spruce genome project (http://congenie.org/). *The intron information is not available.

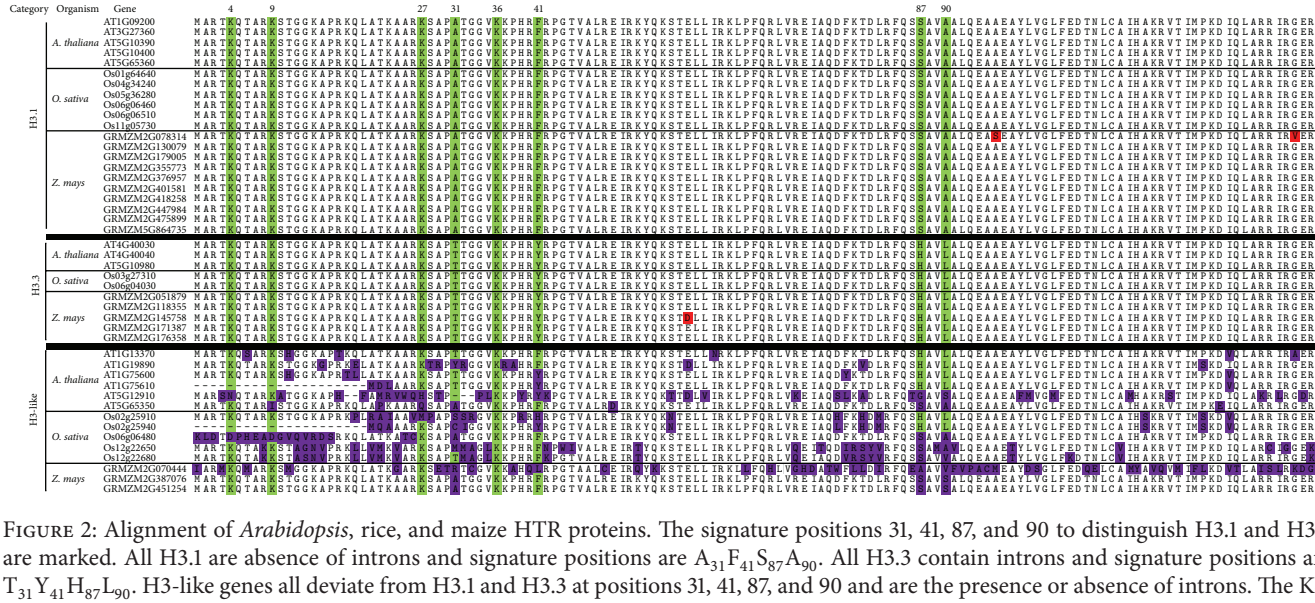


Figure 2: Alignment of *Arabidopsis*, rice, and maize HTR proteins. The signature positions 31, 41, 87, and 90 to distinguish H3.1 and H3.3 are marked. All H3.1 are absence of introns and signature positions are $A_{31}F_{41}S_{87}A_{90}$. All H3.3 contain introns and signature positions are $T_{31}Y_{41}H_{87}L_{90}$. H3-like genes all deviate from H3.1 and H3.3 at positions 31, 41, 87, and 90 and are the presence or absence of introns. The K4, K9, K27, and K36 commonly involved in histone methylations are highly conserved in H3.1 and H3.3. Amino acid substitutions are colored in red (H3.1 and H3.3) and purple (H3-like).

the nonflowering plants, the intronless genes with corresponding proteins carry the $A_{31}Y_{41}(S/Q)_{87}L_{90}$ and we defined them as H3.1A [6]. All the H3-like genes have heterogeneous features [6]: absence or degeneration of N-terminal part of HTR proteins, deviation from H3.1 and H3.3 consensuses at positions 31, 41, 87, and 90, and presence or absence of introns. Our analysis suggested that H3.1, H3-like, and CenH3 already present before the split between chlorophyte and charophyte algae, which is consistent with previous finding [6]. We did not find H3.3 in all the four green algae species, which suggested that H3.3 sequences are more divergent; the signature might be lost in chlorophyta but retained in land plant evolution.

3.2. Conserved Characteristics in the H3 Variants. We are particularly interested in grass H3 variants and we performed alignments using HTR protein sequences from *Arabidopsis*, maize, and rice (Figure 2). Analysis revealed nearly identical sequence within H3.1 group and H3.3 group. H3.1 and H3.3

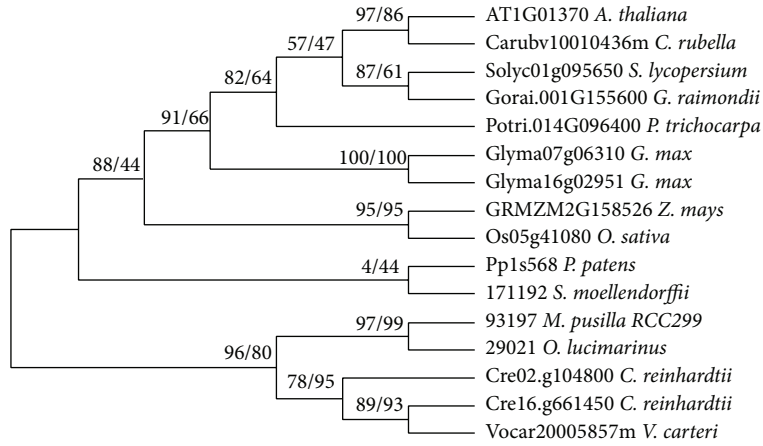


Figure 3: Phylogeny of plant centromeric H3 variants CenH3. For each node, statistical support values are marked (numbers from left to right: neighbor joining inferred under JTT + G models and maximum-likelihood bootstraps inferred using LG + G model).

are highly similar except the four signature sites. In *Arabidopsis*, the histone H3 lysine-4 trimethylation (H3K4me3) and H3 lysine-36 di- and trimethylation (H3K36me2/me3) are linked with active gene expression; H3 lysine-9 methylation and H3 lysine-27 trimethylation (H3K27me3) are associated with gene repression [30, 31]. The K4, K9, K27, and K36 are highly conserved in H3.1 and H3.3. K36 is conserved in all the H3 types (Figure 2), indicating that the major lysine posttranslational modifications are conserved. In addition, we observed many substitutions of K4, K9, and K27 in H3-like genes. One interesting substitution is the K27M found in rice Os02g25910. In animal overexpression of histone H3.3K27M results in loss of H3K27 methylation and derepression of polycomb target genes; lysine-to-methionine mutants could inhibit methylation pathways that also function as biochemical reagents for capturing site-specific histone-modifying enzymes [32, 33]. Whether the K27M substitution in rice is involved in inhibiting K27 methylation will be a subject for further investigation.

We found that the HTR genes in *Capsella rubella* are significantly different from all the other species' HTRs. We performed an alignment compared with *Arabidopsis thaliana* (Supplementary Figure 1). The *Capsella rubella* HTRs were highly conserved within the species. However, even though *Capsella rubella* is the close relative to *Arabidopsis*, we did not find the H3.1 or H3.1 signature. Furthermore, we observed multiple deletions, the deletions were found in all the H3-like genes except CenH3, and the corresponding sequences were highly conserved in *Arabidopsis*. This suggested that CenH3 evolved independently to other H3 variants; the deletion event happened before the duplication of H3-like variants in *Capsella rubella*.

3.3. Molecular Evolution of Plant HTR Genes. Comparing rates of dN and dS is a common way to examine selection pressures on coding regions. Commonly, a dN/dS value of 1 is used to indicate neutral selection and values lesser or greater than 1 to, respectively, indicate purifying and positive selection [34]. To analyze the selective pressures acting during the expansion of plant HTR genes, we investigated the influences of selective constraints on the three group H3 variants coding region. By globally fitting an evolutionary model, we first calculated the dN/dS ratios for each group. The dN/dS values were substantially <1 in all groups, providing a crude indication that the strong purifying selection has been maintained across plants, implying the H3 variants functional conservation. At the individual codon level, most of the residues were under significant negative selection ($P < 0.05$) (Supplementary Figure 2). We also observed several sites under relaxed constrain (Supplementary Figure 2), which may contribute to the functional divergent [9].

3.4. Expression Analysis of HTR Genes at Different Developmental Stages. To understand the temporal and spatial expression patterns of grass HTR genes, we compared their expression patterns during maize and rice development.

Microarray data of 60 different tissues and developmental conditions of maize were used [24]. Several genes were not detected in this dataset, suggesting that they might be pseudogenes. The expressed genes were detected in all samples examined (Figure 4). The two H3.3 genes showed different expression patterns: H3.3 GRMZM2G051879 was constitutively expressed in all organs and developmental stages; in contrast, H3.3 gene GRMZM2G176358 was highly expressed in all stage roots and leaves, but low expressed in seeds, embryo, and endosperm. All the H3.1 genes (GRMZM2G475899, GRMZM2G078314, GRMZM2G447984) showed higher expression in roots, stem, seed, and embryo, but lower expression in all stages of leaves and endosperm, which suggested they play important roles in root, meristem, and embryo development. The H3-like gene GRMZM2G387076 has similar expression to H3.1. GRMZM2G387076 is intronless also suggest that this gene is evolutionary more close to H3.1. The other H3-like gene GRMZM2G070444

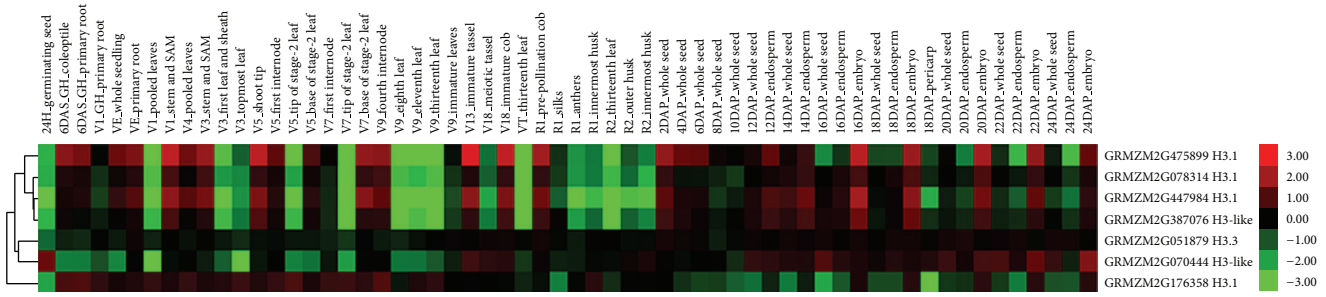


Figure 4: Expression profiles of HTR genes in maize across different developmental stages and organs. The genes IDs are on the right. The tissues used for expression analysis are indicated at the top of each column. The color bar represents log2 expression values.

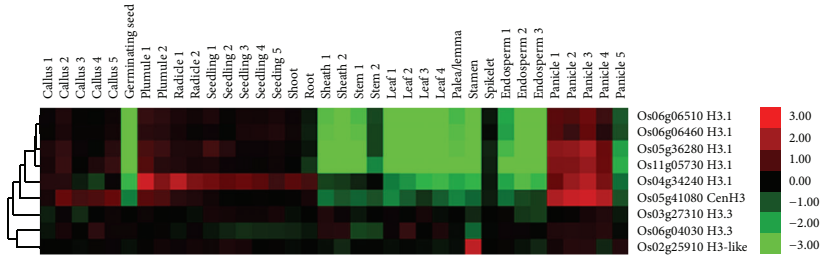


Figure 5: Expression profiles of HTR genes in rice across different developmental stages and organs. The genes IDs are on the right. The tissues used for expression analysis are indicated at the top of each column. The color bar represents log2 expression values.

contains intron; its expression has heterogeneous feature: low expressed in roots and leaves and highly expressed in seed, embryo, and endosperm.

We next analysed the expression profiles of rice HTR genes (Figure 5). The genome arrays from 39 tissues collected throughout the life cycle of the rice were used [25]. The H3.3 genes were highly expressed in majority samples; the intron-containing H3-like gene LOC_Os02g25910 has similar expression patterns with H3.3. All the H3.1 genes have the same expression features: they expressed at higher level in callus, seedlings, shoot, and root and lower level in leaves, endosperms, and so forth, which is highly similar with maize, indicating functional conservation of HTRs in plants.

3.5. Functional Conservation and Diversity of HTR Genes. The comparative phylogenetic analysis of plant HTR proteins enables us only to identify the CenH3 variants. Therefore, we performed alignment to classify H3.1, H3.3, and H3-like variants using the established signatures [6]. In *Arabidopsis*, the H3.1 genes showed high level expression in tissues containing rapidly dividing cells; the H3.3 genes exhibited high level of expression in most of the tissues examined. *Arabidopsis* H3.1 and H3.3 were proposed as replication dependent and replication independent, respectively [9]. Our expression profiles showed the H3.1 and H3.3 expression was similar among maize, rice, and *Arabidopsis* [9], indicating the function conservation in angiosperm. The purifying selection also supports the functional constraint during plant evolution. Collectively, these findings indicate that in flowering plant H3.1 variants are replication dependent and the H3.3 variants are replication independent.

Our analysis identified a big number of H3-like genes. The numbers of H3-like genes are different between species. The H3-like genes' expression patterns are similar with H3.1 or H3.3 or have heterogeneous feature [9] (Figures 4 and 5). In *Arabidopsis*, the H3-like gene At1g19890 is male gamete specific; disruption of this gene might be compensated by other H3 genes [9]. Therefore, the H3-like genes may be full/partial redundancy to other HTRs or have a dosage dependent manner when forming nucleosome. Recent paper showed that the histone concentrations affect gene expression through nucleosome repeat length (NRL) [35]. Considering the number of H3-like genes in each species may affect the histone concentration; it is tempting to hypothesize that H3-like variants contribute to control gene expression by regulating NRL. Null mutants of H3-like genes are required to test the hypothesis.

4. Conclusion

Histone H3 variants in animals and *Arabidopsis* are known to be crucial for a multitude of physiological and intracellular processes. Here, we report an identification, evolution, and expression of the H3 variants in plants; the comparative genomic analyses of H3 variants establish a framework for understanding the evolutionary mechanisms involved in the origin and expansion of plant HTR genes, and it provides a basis for investigating cellular functions of HTR genes.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

This work was supported by research grants from Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges under Beijing Municipality (IDHT20150503) and Beijing Municipal Natural Science Foundation (6122004).

References

[1] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[2] P. B. Talbert and S. Henikoff, "Histone variants—ancient wrap artists of the epigenome," *Nature Reviews Molecular Cell Biology*, vol. 11, no. 4, pp. 264–275, 2010.

[3] D. Schumperli, "Cell-cycle regulation of histone gene expression," *Cell*, vol. 45, no. 4, pp. 471–472, 1986.

[4] D. Filipescu, S. Müller, and G. Almouzni, "Histone H3 variants and their chaperones during development and disease: contributing to epigenetic control," *Annual Review of Cell and Developmental Biology*, vol. 30, no. 1, pp. 615–646, 2014.

[5] H. S. Malik and S. Henikoff, "Phylogenomics of the nucleosome," *Nature Structural Biology*, vol. 10, no. 11, pp. 882–891, 2003.

[6] M. Ingouff and F. Berger, "Histone3 variants in plants," *Chromosoma*, vol. 119, no. 1, pp. 27–33, 2010.

[7] L. de Koning, A. Corpet, J. E. Haber, and G. Almouzni, "Histone chaperones: an escort network regulating histone traffic," *Nature Structural and Molecular Biology*, vol. 14, no. 11, pp. 997–1007, 2007.

[8] K. Ahmad and S. Henikoff, "Histone H3 variants specify modes of chromatin assembly," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, supplement 4, pp. 16477–16484, 2002.

[9] T. Okada, M. Endo, M. B. Singh, and P. L. Bhalla, "Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3," *Plant Journal*, vol. 44, no. 4, pp. 557–568, 2005.

[10] J. H. Waterborg, "Evolution of histone H_3: emergence of variants and conservation of post-translational modification sites," *Biochemistry and Cell Biology*, vol. 90, no. 1, pp. 79–95, 2012.

[11] H. Wollmann, S. Holec, K. Alden, N. D. Clarke, P. É. Jacques, and F. Berger, "Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome," *PLoS Genetics*, vol. 8, no. 5, Article ID e1002658, 2012.

[12] C. Wirbelauer, O. Bell, and D. Schübeler, "Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias," *Genes and Development*, vol. 19, no. 15, pp. 1761–1766, 2005.

[13] A. D. Goldberg, L. A. Banaszynski, K.-M. Noh et al., "Distinct factors control histone variant H3.3 localization at specific genomic regions," *Cell*, vol. 140, no. 5, pp. 678–691, 2010.

[14] E. S. Choi, J. A. Shin, H. S. Kim, and Y. K. Jang, "Dynamic regulation of replication independent deposition of histone H3 in fission yeast," *Nucleic Acids Research*, vol. 33, no. 22, pp. 7102–7110, 2005.

[15] M. Ingouff, S. Rademacher, S. Holec et al., "Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*," *Current Biology*, vol. 20, no. 23, pp. 2137–2143, 2010.

[16] A. Marchler-Bauer, C. Zheng, F. Chitsaz et al., "CDD: conserved domains and protein three-dimensional structure," *Nucleic Acids Research*, vol. 41, no. 1, pp. D348–D352, 2013.

[17] I. Letunic, T. Doerks, and P. Bork, "SMART: recent updates, new developments and status in 2015," *Nucleic Acids Research*, 2014.

[18] R. D. Finn, A. Bateman, J. Clements et al., "Pfam: the protein families database," *Nucleic Acids Research*, vol. 42, pp. D222–D230, 2014.

[19] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772–780, 2013.

[20] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.

[21] D. Darriba, G. L. Taboada, R. Doallo, and D. Posada, "ProtTest 3: fast selection of best-fit models of protein evolution," *Bioinformatics*, vol. 27, no. 8, Article ID btr088, pp. 1164–1165, 2011.

[22] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0," *Systematic Biology*, vol. 59, no. 3, pp. 307–321, 2010.

[23] S. L. Kosakovsky Pond, S. D. W. Frost, and S. V. Muse, "HyPhy: hypothesis testing using phylogenies," *Bioinformatics*, vol. 21, no. 5, pp. 676–679, 2005.

[24] R. S. Sekhon, H. Lin, K. L. Childs et al., "Genome-wide atlas of transcription during maize development," *Plant Journal*, vol. 66, no. 4, pp. 553–563, 2011.

[25] L. Wang, W. Xie, Y. Chen et al., "A dynamic gene expression atlas covering the entire life cycle of rice," *Plant Journal*, vol. 61, no. 5, pp. 752–766, 2010.

[26] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, pp. D991–D995, 2013.

[27] M. J. L. de Hoon, S. Imoto, J. Nolan, and S. Miyano, "Open source clustering software," *Bioinformatics*, vol. 20, no. 9, pp. 1453–1454, 2004.

[28] A. J. Saldanha, "Java Treeview—extensible visualization of microarray data," *Bioinformatics*, vol. 20, no. 17, pp. 3246–3248, 2004.

[29] B. Nystedt, N. R. Street, A. Wetterbom et al., "The Norway spruce genome sequence and conifer genome evolution," *Nature*, vol. 497, no. 7451, pp. 579–584, 2013.

[30] T. Kouzarides, "Chromatin modifications and their function," *Cell*, vol. 128, no. 4, pp. 693–705, 2007.

[31] S. Henikoff and A. Shilatifard, "Histone modification: cause or cog?" *Trends in Genetics*, vol. 27, no. 10, pp. 389–396, 2011.

[32] P. W. Lewis, M. M. Müller, M. S. Koletsky et al., "Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma," *Science*, vol. 340, no. 6134, pp. 857–861, 2013.

[33] H.-M. Herz, M. Morgan, X. Gao et al., "Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling," *Science*, vol. 345, no. 6200, pp. 1065–1070, 2014.

[34] J. C. Fay and C.-I. Wu, "Sequence divergence, functional constraint, and selection in protein evolution," *Annual Review of Genomics and Human Genetics*, vol. 4, pp. 213–235, 2003.

[35] D. A. Beshnova, A. G. Cherstvy, Y. Vainshtein, and V. B. Teif, "Regulation of the nucleosome repeat length in vivo by the DNA sequence, protein concentrations and long-range interactions," *PLoS Computational Biology*, vol. 10, no. 7, Article ID e1003698, 2014.