Genome-wide Identification and Functional Analysis of the H3 Gene Family in Cotton

ZHANG Youping  
Chinese Academy of Agricultural Sciences

ZHANG Jia  
Chinese Academy of Agricultural Sciences

WANG Qiaolian  
Chinese Academy of Agricultural Sciences

Li Simin  
Chinese Academy of Agricultural Sciences

ZUO Dongyun  
Chinese Academy of Agricultural Sciences

CHENG Hailiang  
Chinese Academy of Agricultural Sciences

ASHRAF Javaria  
Chinese Academy of Agricultural Sciences

LV Limin  
Chinese Academy of Agricultural Sciences

KONG Linglei  
Chinese Academy of Agricultural Sciences

FENG Xiaoxu  
Chinese Academy of Agricultural Sciences

JZ Yu  
USDA ARS: USDA Agricultural Research Service

guoli song  
Chinese Academy of Agricultural Sciences Cotton Research Institute  https://orcid.org/0000-0003-3236-9286

Research

Keywords: H3 gene family, VIGS, stomata, chloroplasts

DOI: https://doi.org/10.21203/rs.3.rs-86043/v1

License: ☑️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background:** Histones are major components of chromatin, which is a nucleosome structure associated with chromosome segregation, DNA packaging and transcriptional regulation. Histone H3 is encoded by many genes in most eukaryotic species, but little information is known about the Histone H3 gene family in cotton.

**Results:** In this study, we identified and analyzed the evolution and expression of histone H3 gene family in cotton. First, 34 *G. hirsutum* genes were identified belonging to the H3 gene family which were divided into four subclasses: CENH3, H3.1, H3.3 and H3-like. Among these H3.1 subclass contained the highest number of genes (22 members) followed by H3.3 subclass (9 members). In addition, there were 18 and 16 H3 genes identified in *G. arboreum* and *G. raimondii*, respectively. Furthermore, we conducted conserved sequence analysis of H3 proteins, and found that the four amino acids signature including A31F41S87A90 for H3.1 and T31Y41H87L90 for H3.3 could be used to discriminate H3.1 from H3.3. The expression of H3 gene family varied in different tissues and developmental stages of *G. hirsutum*, where H3.1 subclass genes play a critical role in pistil development. By virus-induced gene silencing of GhCENH3 (Gh_D07G1382) gene, the size of leaf got smaller with pYL156-CENH3 than that with pYL156 in TM-1. Whereas, the number of the stomata in the leaf epidermis and number of chloroplasts in the leaf stomatal guard cells by pYL156-CENH3 was more than that by pYL156 and pYL156-PDS.

**Conclusions:** Four sub-classes (CENH3, H3.1, H3.3 and H3-like) of H3 gene family were highly conserved in cotton during the rapid phase of evolution among which CENH3 is necessary for leaf growth. These findings are useful for providing further insights into cotton biology and breeding.

**Background**

Haploid breeding helps to shorten the breeding time significantly and boost the QTL expression (Zhang Zili 2008; Seymour DK et al. 2012; Gihwan Yi, Kyung-Min Kim and ae-Keun Sohn 2013). However, one of the most important challenges in haploid breeding is to produce a stable haploid. Since the first haploid cotton was discovered by Harland (1920; 1936), haploid plants have been produced in 70 genera and 206 crops in the 1980s (Pan Jiaju 1994). Initially, doubled-haploid (DH) technology was used in tomato (Wu Penghao 2014). Subsequently, DH technology was applied in 12 species of pepper (Toole MGS and Bamford R 1947), maize (Chase SS 1952), cotton (Turchotte EL and Feaster CV 1963), rape (Thompson KF), barley (Ho KM and Jones GE 1980), and wheat (Thomas WTB, Gertson B and Forster BP 2003). Cotton haploid was generated through Vsg, a semigamy with virecent characteristic marker that was most efficient to produce haploid. However, the reproducibility of the semigamy is unstable which hinders its utility for the development of cotton haploid (Sugita M, Ichinose M, Ide M and Sugita C 2013), and making it difficult to apply DH technology and conduct further genetic research.

**CENH3** (Centromeric Histone H3) gene is a centromere-specific histone H3 mutant, which plays an important role in the development of embryos and zygotes during the process of parental hybridization. There are up to 45% haploid plants with male wild type and a few double haploid by gene CENH3 in Arabidopsis (Maruthachalam Ravi and Simon WL 2002). In barley, the hybridization of Hordeum vulgare × Hordeum bulbosum showed that the CENH3 gene is inactive or not fused into the paternal line, resulting a haploid maternal with a failure of centromere assembly (Sanei M and Pickering R 2011). The amino terminal tail of CENH3 gene is essential for survival of Drosophila, which comprised of amino (amino terminal tail) and carboxyl (histone folding domain, HFD) terminals with a highly variable sequence that has three conservative domains in 22 different Drosophila species (Malik HS, Vermaak D and Henikoff S 2002). Recent studies also showed the abnormal development of seeds because of the difference in amino terminal of CENH3 (Maheshwari S and Tan EH 2015). In contrast, the histone fold domains (HFD) are conserved having 4 alpha helices (a1, a2, a3) and 2 rings (Loop1 and Loop2). The a1, a2 helices and Loop1 rings are diverse and related to the DNA binding of CENH3 (Malik HS and Vermaak D 2002; Paul B and Talbert Ricardo 2002; Masonbrink RE et al. 2014). In Drosophila and Arabidopsis, it was also found that adaptive evolution occurred in proximity region of the CENH3 Loop1 and core sequence (Henikoff S and Dalal Y 2005). Histone 3 is very conservative in evolution, while CENH3 is rapidly evolving. This difference is mainly due to the fact that histone 3 interact with DNA of the whole genome, while CENH3 only interacts with the rapidly evolving centromeric satellite DNA (Malik HS and Henikoff S 2001). However, the rapid evolution of centromeric sequences is not the main driving force to lead the diversification of CENH3 sequences (Masonbrink RE et al. 2014). CENH3 is widely spread in plants and animals, and its main function has been conserved in monocotyledonous and dicotyledonous plants (Malik HS and Vermaak D 2002). Recent studies have shown that the E2F transcription factor regulates the expression of CENH3 in Arabidopsis (Heckmann S et al. 2011).

In this study, based on a comprehensive genome-wide analysis of three Gossypium species (*G. hirsutum* and its probable ancestral diploids *G. raimondii* and *G. arboreum*), we analyzed the conserved proteins sequence of the H3 gene family.. Nullification of the GhCENH3 (Gh_D07G1382) gene in cotton by virus-induced gene silencing (VIGS) significantly reduce the leaf size in TM-1. Our current study provides important insights about CENH3 gene in *Gossypium spp.* and will be useful for cotton biology and breeding.
Methods

Cotton genome and RNA-seq resources

The genome sequence and annotation information of three *Gossypium* species (*G. raimondii*, *G. arboreum* and *G. hirsutum*) were downloaded from the CottonGen (https://www.cottongen.org/). RNA-seq data for gene expression analysis in *G. hirsutum* was downloaded from ccNET database (http://structuralbiology.cau.edu.cn/gossypium/), which mainly includes the gene expression data in root, stem, leaf, petal, stamen, pistil and fibers at 5, 10, 20 and 25 days post anthesis (DPA).

Identification of *H3* genes and proteins

To identify the *H3* gene family, *H3* gene of *Arabidopsis* (At1g09200) was used to retrieve the whole genome database of three cotton species by BLASTX.

The presence of the *H3* domains in the protein structure was further validated using SMART software (http://smart.embl-heidelberg.de/) and PfamScan website (https://www.ebi.ac.uk/Tools/pfa/pfamscan/). The redundant sequences without *H3* domain were manually checked and then removed. Molecular weight (MW), theoretical isoelectric point (pI), and size of the *H3* genes were investigated with the online tool ExPASy (http://expasy.org/tools/). Subcellular locations were predicted by software WoLF PSORT (http://wolfpsort.org/).

Phylogenetic analysis

The multiple sequence alignment of *H3* domain sequence from three cotton species was accomplished by ClustalX2 software (Thompson JD, Gibson TJ and Higgins DG 2002) with default parameters. The unrooted phylogenetic tree was constructed by the neighbour joining tree (NJ) in MEGA 6 software (http://www.megasoftware.net/history.php). The bootstrap analysis was set for 1000 iterations.

Chromosomal mapping and gene duplication

The physical location data of *H3* genes were retrieved from genome sequence data of three cotton species. Mapping of these *H3* genes was performed using Mapchart software (Voorrips RE 2002). Synonymous and non-synonymous rates of evolution was computed by the maximum likelihood method from Ka/Ks calculator 2.0 software (Wang D, Zhang Y and Zhang Z et al. 2010).

Gene structure and domain analysis

The exon and intron organizations of *H3* genes were inferred through comparison of genomic sequences and CDS sequences by the gene structure display server (http://gsds.cbi.pku.edu.cn/). The conserved motifs in *H3* genes were identified by MEME (http://meme-suite.org/tools/meme) with the following parameters: the 20 motifs and optimum width from 6 to 250.

Genome-wide syntenic analysis of *H3* genes

A BLASTP comparison provided the pairwise gene information between *G. hirsutum* and two diploid cotton species (*G. raimondii* and *G. arboreum*). According to the BLASTP output, the syntenic analysis was constructed using circos-0.69-3 software package (http://circos.ca/software/) with default parameters.

Quantitative reverse transcription-PCR (qRT-PCR)

Total RNA was isolated from leaves PYL156-CENH3 and PYL156 using EASY spin Plant RNA Kit (Aidlab). Afterwards, first-strand cDNA was synthesized by PrimeScript™ II 1st strand cDNA Synthesis Kit (TaKaRa) according to the manufacturer’s instructions. qRT-PCR was carried out in 20 μL volume containing 2 μL cDNA, 0.8 μL 10 μM forward and reverse primer, 10 μL SYBR® Premix Ex Taq II (2×), 0.4 μL ROX Reference Dye II (50×) and 6 μL d₃H₂O. PCR amplification was performed under the following conditions: denaturation at 95 °C for 30 sec; 40 cycles at 95 °C for 5 sec and 60 °C for 30 sec; followed by 95 °C for 15 sec, 60 °C for 1 min, (ABI PRISM@7500-Fast Real-Time PCR system) (Cheng HL et al. 2016). qRT-PCR was carried out by the gene-specific primers.
Identification and syntenic analysis of Gossypium H3 gene family

A total of 34 G. hirsutum genes were identified belonging to the H3 gene family (Table 1), which were distributed on 16 chromosomes of A and D subgenomes, respectively (Fig. 1). Among them, multiple genes were clustered on one chromosome such as three genes were clustered together on At_chr10 and Dt_chr10 respectively, showing a higher similarity. The H3 gene distributed on chromosomes 5, 6, 8, 10 and 13 of the A subgenome showed a higher collinearity with those in the D subgenome (Fig. 1).

In addition, 18 H3 genes were distributed on 8 chromosomes and 2 scaffolds (scaffold3134 and scaffold 3624) of G. arboreum, while 16 H3 genes of G. raimondii where distributed in 7 chromosomes and 4 scaffolds (scaffold 371, scaffold266, scaffold266 and scaffold23), respectively (Supplementary Table S1). The syntenic analysis of H3 gene in different cotton species revealed that three genes of G. arboreum (Cotton_A_23780, Cotton_A_28035 and Cotton_A_34310) and one gene of G. raimondii (Cotton_D_gene_10004657) were lost in evolution, while five genes of G. hirsutum (Gh_A05G1915, Gh_A07G1271, Gh_D07G0357, Gh_D07G1382 and Gh_D10G0981) were appeared during evolution (Fig. 2).

Evolution and structure analysis of Histone H3 genes in G. hirsutum

These identified 34 H3 genes of G. hirsutum were divided into four sub-classes including CENH3, H3.1, H3.3 and H3-like (Fig. 4) by comparing with the H3 gene of Arabidopsis: Among these 34 H3 genes, 22 and 9 genes belonged to the subclasses H3.1 and H3.3, while subclasses CENH3 and H3-like contained only two and one genes (Fig. 3). Structural analysis of the H3 genes of G. hirsutum showed that the H3.1 sub-class contained no introns except Gh_D10G0981 with an intron, while the sub-classes H3.3, CENH3 and H3-like contained multiple introns. These results were similar to previous studies (Okada T and Endo M 2005; Cui J and Zhang Z 2015).

Conserved sequences analysis in the H3 gene family

The conserved sequence of the H3 protein family were analyzed in G. hirsutum (Fig. 5). Out of the 34 G. hirsutum H3 proteins, 31 had a highly conserved sequence, including 22 H3.1 proteins and 9 H3.3 proteins, in which only four conserved amino acid substitutions found at sites 31, 41, 87 and 90 between the H3.1 and H3.3 subclass, where the four conserved amino acids are A31F41S87A90 for H3.1 and T31Y41H87L90 for H3.3. These four conserved amino acid signatures could be used to discriminate H3.1 from H3.3. Additionally, 2 CENH3 proteins (Gh_A07G1271 and Gh_D07G1382) and H3-like protein (Gh_D07G0357) had a highly diverse sequence and InDel. The CENH3 and H3-like subclasses carried the R31(P/R)41S87(H/L)90 and N31Q41P87Y90 signatures, respectively.
Previous studies showed that the histone H3 lysine 4 trimethylation (H3K4me3), and H3 lysine 36 di- and trimethylation (H3K36me2/me3) were associated with active gene expression while H3 lysine 9 methylation and H3 lysine 27 trimethylation (H3K27me3) were involved in gene repression (Cui J, Zhang Z and Shao Y 2015). In our study, the K4, K9, K27 and K36 were highly conserved in H3.1 and H3.3, while the members of CENH3 carry K4, K9, S27 and K/N36, and H3-like carried H4, L9, A27 and E36, indicating that the lysine modifications were conserved in the two sites of K4 and K9.

Based on the Ka/Ks ratio, it can be assumed that Darwinian positive selection was linked with the H3 gene divergence after duplication (Prince VE and Pickett FB 2002; Vandepoele K 2003). In our study, we found that 16 gene pairs had low Ka/Ks ratios (smaller than 0.5) and one gene pair had the Ka/Ks ratios between 0.5 and 1.0. Only one gene pair (Gh_A07G1271-Gh_D07G1382) had Ka/Ks larger than 1, might be due to relatively rapid evolution following duplication (Table 2). As most of the Ka/Ks ratios were smaller than 1.0, we presumed that the cotton H3 gene family had undergone strong purifying selection pressure with limited functional divergence that occurred after segmental duplications and whole genome duplication (WGD).

**Expression profile of H3 genes across different tissues and developmental stages**

To understand the temporal and spatial expression levels of different H3 genes, a publicly deposited RNA-seq data was used to assess the expression profile across different tissues and developmental stages. Among the 34 H3 genes, four groups of genes have been identified with FPKM > 0.5 in at least one of the selected tissues and developmental stages (Fig. 6). We found that H3 genes were widely expressed in the vegetative (root, stem and leaf) and reproductive (petal, stamen and pistil) tissues as well as in the fiber (5, 10, 20 and 25 DPA), indicating these have multiple biological functions in different tissues. Interestingly, we found that genes belong to H3.1 subgroup were up-regulated in pistil and all vegetative (root, stem, and leaf) tissues and down-regulated in other tissues, indicating their critical role in pistil development. Some genes were up-regulated in one tissue while down-regulated in the other tissues. For example, the high expression of Gh_A02G0886 in stamen suggests that it may play a critical role in stamen development. Similarly, Gh_D07G1382 was up-regulated in pistil and stamen while down-regulated in other tissues. In contrast, some genes were highly expressed in different fiber developmental stages such as Gh_A11G1633 and Gh_D02G0973 were up-regulated in the early stage (5 and 10 DPA) of fibre development, while down-regulated in the 20 and 25 DPA of fiber development.

**Silencing of the GhCENH3 gene affecting the leaf size and the number of the stomata and stroma chloroplast**

Functional analysis of the gene GhCENH3 was performed in TM-1 using VIGS technology to validate its role in leaf size and the number of the stomata and stroma chloroplast in cotton. After the Agrobacterium-mediated infection, the phenotype of both silenced and non-silenced plants was observed regularly to check the efficiency of gene silencing. The mutant phenotypes of the VIGS-treated plants started to emerge after one week of Agrobacterium-mediated infection. The plants injected with pYL156-PDS showed the loss of normal green coloration with albino phenotype in leaves (Fig. 7), showing the efficiency of gene silencing. The significantly smaller size leaves of the plants infiltrated with pYL156-CENH3 were observed (Fig. 7), while, the plants infiltrated with pYL156 had no effect on leaves (Fig. 7). To check the silencing efficiency, qRT-PCR was performed which showed a significant lower expression of candidate gene in the plants infected by pYL156-CENH3 (Fig. 8). In additional, the number of the stomata was 57 ± 1.414 with the coefficient variation 0.055 by pYL156-CENH3, 21 ± 0.632 with the coefficient variation 0.067 by pYL156 and 15 ± 0.748 with the coefficient variation 0.109 by pYL156-PDS (Fig. 9). The number of the stomatal chloroplast was 46.8 ± 5.805 with the coefficient variation 0.1240 by pYL156-CENH3, 21.2 ± 2.864 with the coefficient variation 0.1351 by pYL156 and 15 ± 0.748 with the coefficient variation 0.109 by pYL156-PDS (Fig. 9). The number of the stomatal chloroplast showed size variations in plants injected by pYL156-CENH3, which would need further research to identify the reason of this size variation.

**Discussion**

In this study, 34 G. hirsutum genes were identified belonging to the H3 gene family and they were divided into four subclasses: CENH3, H3.1, H3.3 and H3-like, among which the H3.1 subclass contained the highest number of genes (22 members) followed by H3.3 subclass (9 members). The genes of H3.1 subclass had the high expression level in pistil indicated that the members of the H3.1 subclass may play a critical role in pistil development. In addition, 18 and 16 H3 genes were identified in G. arboreum and G. raimondii, respectively. Syntenic analysis of the H3 gene family in three cotton species revealed that three and one genes of G. arboreum and G. raimondii were lost during evolution, while 5 genes were appeared in G. hirsutum, showing that these genes play a critical role in cotton evolution.
Structural analysis of the H3 gene family showed that H3.1 subclass contains almost no introns, while rest of the H3 genes contain multiple introns, which is consistent with the previous results (Okada T1 and Endo M 2005; Cui J and Zhang Z 2015). Additionally, the signature of four conserved amino acids A31F41S87A90 for H3.1 and T31Y41H87L90 for H3.3 was discovered through sequence alignment. The signature characteristics of four conserved amino acids and absence of intron could be used to distinguish H3.1 from H3.3 subclass.

Previous study indicated that the H3K4me3 and H3K36me2/me3 are associated with gene expression, while H3 lysine 9 methylation and H3K27me3 are involved in gene repression (Cui J and Zhang Z 2015). In this study, the K4, K9, K27 and K36 are highly conserved in H3.1 and H3.3 subclasses, while rest of the H3 proteins has more sequence variations. For example, CENH3 carry K4, K9, S27 and K/N36, and H3-like carry H4, L9, A27 and E36, indicating that lysine modifications are conserved in two sites (K4 and K9), and 27 and 36 sites may associated with diverse functions, however more work need to validate it.

Haploid breeding is a challenging task in cotton due to the difficulty of the production of stable haploid plants and chromosome doubling of the haploid plant. The haploid plants were thin and stunted with small leaves, small buds and no-cracked anthers in the early flowering stage (Turcotte EL and Feaster CV 1963; Owings AD, P. Sarvella and Meyer JR 1964). A rapid identification of the haploid plants is the key to map and clone the gene (Prigge V and Melchinger AE 2012; Dong X and Xu X 2013; Kelliher et. al. 2017). The number of stomata on the leaf epidermis per unit area in haploid plant is two or three times higher than that in diploid plants, even in segregating population (Zhang Jun and Yi Chengxin 2002). In addition, stoma chloroplast can also be used to identify the haploid plants because the ratio of the chloroplast in the leaf stomatal guard cells between the haploid (n = 2x = 26) and the diploid (2n = 4x = 52) plants is 1: 2 even in the chimera leaf, which is similar to the ratio of the chromosome number (Chaudhari HK and Barrow JR 1975; Howman EV and Fowler KJ 2000). In current study, the smaller leaf size has been observed in plants (TM-1) infected by pYL156-CENH3 than the plants (TM-1) infected by pYL156. The number of the stomata on the leaf epidermis per unit area of the infected plants by pYL156-CENH3 was more than that by pYL156 and pYL156-PDS, which is consistent with previous research (Zhang Jun and Yi Chengxin 2002). Interestingly, the number of chloroplasts in the leaf stomatal guard cells by pYL156-CENH3 is more with dispersed states than that by pYL156 which is not consistent with previous studies (Chaudhari HK and Barrow JR 1975). It may be due to the homozygosity after VIGS injection. In mice, the knockout of the CENPa gene (CENP-A) resulted the development of healthy and fertile heterozygous, while the development of the homozygous mutant embryos took place no more than 6.5 days, and the damaged embryo has an abnormal mitotic division (Howman EV and Fowler KJ 2000). In Arabidopsis, a homozygote that completely lose CENH3 gene could only develop into the spherical stage, whereas heterozygous mutants developed normally (Yang Yuanyuan 2013). However, further studies are needed to explore the role of CENH3 for the production of haploid and double haploid plants of cotton.

Conclusions

This study reveals that GhCENH3 gene which belongs to the H3 gene family plays an important role in leaf size development and also significantly affects the number of the stomata and chloroplasts in leaf epidermis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests

The authors declare that they have no competing interests.
Funding

None

Authors' contributions

Song GL and Yu JZ managed the project and designed the research. Zhang YP, Zhang J, Wang QL, Li SM, Zuo DY, Cheng HL, Ashraf J, Lv LL, KONG LL and Feng XX performed the experiments and prepared figures and tables. Zhang YP, Song GL, Yu JZ and Wang QL wrote and revised the paper. All authors reviewed the manuscript.

Acknowledgments

We thank State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences in China.

References

1. Chase SS. Production of homozygous diploids of maize from monoploids. Agron J. 1952; 44: 263-26.
2. Chaudhari, HK and Barrow JR. Identification of Cotton Haploids by Stomatal Chloroplast-count Technique, Crop Sci. 1975; 15: 760-763.
3. Cheng HL, et al. Fine mapping and candidate gene analysis of the dominant glandless gene Gl2e in cotton (Gossypium spp.). Theor Appl Genet. 2016; 129: 1347-55.
4. Cui J, Zhang Z, Shao Y, Zhang K, Leng P, Liang Z. Genome-wide identification, evolutionary, and expression analyses of histone H3 variants in plants. Biomed Res Int. 2015; 341598.
5. Dong X, Xu X, Miao J, Li L, Zhang D, Mi X, Liu C, Tian X, Melchinger AE, Chen S. Fine mapping of qhir1 influencing in vivo haploid induction in maize. Theor Appl Genet. 2013; 126: 1713-20.
6. Gihwan Yi, Kyung-Min Kim, Jae-Keun Sohn. Use of Androgenesis in Haploid Breeding. Current Research on Agriculture and Life Sciences. 2013; 31: 75-82.
7. Harland SC. A note on a peculiar type of 'rogue' in Sea Island cotton. Agr. News, Barbados. 1920; 19: 29.
8. Harland SC. Haploids in polyembryonic seeds of Sea Island cotton. Journal of Heredity. 1936; 27(6): 229-232.
9. Heckmann S, et al. The E2F transcription factor family regulates CENH3 expression in Arabidopsis thaliana. The Plant Journal. 2011; 68: 646-656.
10. Henikoff S, Dalal Y. Centromeric chromatin: what makes it unique Curr Opin Genet Dev. 2005; 15: 177-84.
11. Ho KM, Jones GE. Mingo barley. Can J Plant Sci. 1980; 60: 279-280.
12. Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, Choo KH. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci. USA. 2000; 97: 1148-1153.
13. Kellieher, et al. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature. 2017; 542: 105-109.
14. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods. 2001; 25: 402-408.
15. Maheshwari S, Tan EH, West A, Franklin FC, Comai L, Chan SW. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. 2015; 11 .
16. Malik HS, Henikoff S. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics. 2001; 157: 1293-1298.
17. Malik HS, Vermaak D, Henikoff S. Recurrent evolution of DNA-binding motifs in the rosophila centromeric histone. Proc Natl Acad Sci USA. 2002; 99: 1449-1454.
18. Maruthachalam Ravi and Simon WL. Chan. Haploid plants produced by centromere-mediated genome elimination. Nature. 2002; 464: 615-618.
19. Masonbrink RE, et al. CenH3 evolution in diploids and polyploids of three angiosperm genera. BMC Plant Biology. 2014; 14: 383.
20. Okada T1, Endo M, Singh MB, Bhalla PL. Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J. 2005; 44: 557-68.
21. Owings AD, P Sarvella, Meyer JR. Twinning and haploidy in a strain of Gossypium barbadense L. Crop Sci. 1964; 4: 652-653.
22. Pan Jiaju. Crop breeding subjects. China Agricultural Publishing, BeiJing. 1994.
23. Pang JH, et al. Development of Agrobacterium-mediated virus-induced gene silencing and performance evaluation of four marker genes in *Gossypium barbadense*. PLoS One. 2013; 8: e73211.

24. Paul B, Talbert, Ricardo, Masuelli A, Tyagi P, Comai, Luca and Steven Henikoff. Centromeric Localization and Adaptive Evolution of an Arabidopsis Histone H3 Variant. The Plant Cell. 2002; 14: 1053–1066.

25. Prigge V, Melchinger AE. Production of haploids and doubled haploids in maize. Methods Mol. Biol. 2012; 877: 161-172.

26. Prince VE, Pickett FB. Splitting pairs: the diverging fates of duplicated genes. Nature reviews Genetics. 2002; 3: 827–37.

27. Sanei M, Pickering R, Kumke K, Nasuda S, Houben A. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc Natl Acad Sci USA. 2011; 108: E498–E505.

28. Seymour DK, et al. Rapid creation of Arabidopsis doubled haploid lines for quantitative trait locus mapping. Proc Natl Acad Sci USA. 2012; 109: 4227-32.

29. Suguira M, Ichinose M, Ide M, Sugita C. Architecture of the PPR gene family in the moss *Physcomitrella patens*. RNA Biol. 2013; 10: 1439-45.

30. Thomas WTB, Gertson B, Forster BP. Doubled haploids in breeding. p. 2003; 337–350 . In: Maluszynski M, Kasha KJ, Forster BP, Szarejko I. (eds.), Doubled haploid production in crop plants: A manual. Kluwer Academic Publ, Dordrecht, Boston, London.

31. Thompson JD, Gibson TJ, Higgins, DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics; Chapter 2, Unit 2.3. 2002.

32. Thompson KF. Oil-seed rape. In Reports of the Plant Breeding. Cambridge University Press, Cambridge. pp 94-96.

33. Toole MGS, Bamford R. The formation of diploid plants from haploid peppers. J Hered. 1947; 36:67-70.

34. Turcotte EL, Feaster CV. Haploids: high-frequency production from single-embryo seeds in a line of Pima cotton. Science. 1963; 140: 1407-1408.

35. Turcotte EL, Feaster, CV. Haploids: high-frequency production from single-embryo seeds in a line of Pima cotton. Science. 1963; 140: 1407-1408.

36. Tuttle JR, et al. Geminivirus-mediated gene silencing from Cotton leaf crumple virus is enhanced by low temperature in cotton. Plant Physiol. 2008; 148: 41-50.

37. Vandepeele K. Evidence that Rice and other Cereals Are Ancient Aneuploids. The Plant Cell Online. 2003; 15: 2192–202.

38. Voorrips RE. Map Chart: Software for the Graphical Presentation of Linkage Maps and QTLs. Journal of Heredity. 2002; 93: 77–78.

39. Wang D, Zhang Y, Zhang Z, et al. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics, Proteomics & Bioinformatics. 2010; 8: 77-80.

40. Wu Penghao. The study for haploid male fertility of in vivo induction in maize (Zea may. L). China Agricultural University. 2014.

41. Yang Yuanyu. Construction of EMS Mutant Library and TILLING Analysis of Gene Opaque and CENH3 in Maize. Sichuan Agricultural University. 2013.

42. Zhang Jun, Yi Chengxin, Guo Wangzhen, Zhang Tianzhen. Identification of Cotton Haploids Induced by Viressently Marked Semigamy Line in Early Developmental Stake. Cotton science. 2002; 14: 138-142.

43. Zhang Zili. Study on the mechanism of in vivo haploid induction by Stock6 derived line. Huazhong Agricultural University. 2008.

Tables

Due to technical limitations, table 1 and table 2 are only available as a download in the Supplemental Files section.