Distribution and Characteristics of Transposable Elements in the Mulberry Genome

Bi Ma, Youchao Xin, Lulu Kuang, and Ningjia He*

State Key Lab. of Silkworm Genome Biology, Southwest Univ., Beibei, Chongqing, China.

ABSTRACT Mulberry (Morus notabilis C. K. Schneid) leaves have been used as the food for the domesticated silkworm, Bombyx mori, for more than 5000 yr, and the mulberry–silkworm relationship is one of the best-known and oldest models of plant defense–insect adaptation. The availability of a genome assembly of mulberry provides us with an opportunity to mine the characteristics and distribution of transposable elements (TEs) in this species and to examine their relationship to genes and gene expression. In this study, a significantly correlated inverse relationship between the percentage coverage of genes and TEs was observed. The TE-rich regions appeared to have a lower percentage of putatively expressed genes. Distribution patterns between different TE superfamilies were detected in the mulberry genome. The Copia elements (the TE making up the greatest proportion of the mulberry genome) were significantly overrepresented within genes in the mulberry genome, and they may have a dominant influence on the evolution of the mulberry genome. Approximately 96.93% (330/344) of the TE-containing genes assigned to pathways were assigned to metabolism-related pathways. The results will be valuable in improving our understanding of the important roles of TEs in mulberry genome evolution.

CORE IDEAS

- Percentage coverage of TEs is significantly negatively correlated with that of genes.
- Different TE superfamilies exhibit distinct distribution patterns in mulberry genome.
- Copia elements may have a dominant influence on the regulation of mulberry genes.
- TE-containing genes assigned to pathways were mainly in metabolism-related pathways.
- TEs are a driving force in the formation of alternatively spliced genes.

Transposable elements have been identified in all eukaryotic genomes and influence the evolution, structural rearrangement, and transcriptional regulation of genes and genomes (Bennetzen, 2000, 2005; Bucher et al., 2012; Chuong et al., 2017; Feschotte, 2008; Lisch, 2013). Transposable elements can be classified into two classes based on their transposition mechanism: (i) retrotransposons (class I), transposition by a copy-and-paste mechanism; and (ii) DNA transposons (class II), transposition by a cut-and-paste mechanism (Wicker et al., 2007). Genome size and TE content in angiosperm species are positively correlated (reviewed in Tenaillon et al. [2010]). Transposable elements occupy large proportions of plant genomes, ranging from 14% in Arabidopsis thaliana (L.) Heynh. (genome size, 125 Mb) (The Arabidopsis Genome Initiative, 2000), 35% in rice (Oryza sativa L.) (genome size, 389 Mb) (International Rice Genome Sequencing Project, 2005), to 85% in maize.

Abbreviations: Ac/Ds, activator–dissociator; AS, alternatively spliced; KEGG, Kyoto Encyclopedia of Genes and Genomes; LTR, long-terminal repeat; PCR, polymerase chain reaction; TE, transposable element; WGS, whole-genome shotgun.

Citation: Ma, B., Y. Xin, L. Kuang, and N. He. 2019. Distribution and characteristics of transposable elements in the mulberry genome. Plant Genome 12:180094. doi: 10.3835/plantgenome2018.12.0094

*Corresponding author (hejia@swu.edu.cn).

© 2019 The Author(s). This is an open access article distributed under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
exon of b1 kernels in maize when a retroelement inserted into the first expression in the vegetative tissues to expression in the expression patterns of the activator gene, which increased expression of the transcriptional activator–dissociator (Ac/Ds) can induce genome structural rearrangements in maize (Yu et al., 2011). Research in Brassica species suggested that some LTR–retrotransposons also preferentially insert within other TEs, namely nested LTRs, playing a critical role in centromere formation (Wei et al., 2013). The distribution patterns of TEs in plant genomes suggested that various TEs have undergone distinct selection pressure or they exhibited distinct insertion preference. For example, LTR–retrotransposons including Copia, Gypsy, and other types found in centromere regions of the plant genome have been shown to play critical roles in the function and formation of centromeres (Neumann et al., 2011; Wei et al., 2013), while another nonautonomous LTR–retrotransposon, Dasheng, is a main component of pericentromeric regions in rice (Jiang et al., 2002). When a TE is located inside the gene regions, the gene can be regarded as a TE-containing gene, and the TE may influence gene function (Jiao and Deng, 2007). For instance, an initial insertion and rearrangement of an LTR element (Gretl) located close to the VvmbyA1 gene lead to color variants in grapevine (Vitis vinifera L.) fruit (Kobayashi et al., 2004). In addition to simply disrupting normal gene expression, TEs can introduce new regulatory information, resulting in alternative expression patterns of genes and genomes (Bennetzen, 2000, 2005; Bucher et al., 2012; Chuong et al., 2017; Feschotte, 2008; Lisch, 2013), it would be valuable to analyze the characteristics and the distribution of TEs in mulberry.

MATERIALS AND METHODS

Data Sources

The unmasked WGS assembly of the mulberry genome, genes, and the transcriptome data associated with five tissues (root, flower, leaf, bud, and bark) of wild-type mulberry were obtained from the Morus Genome Database (MorusDB, v 1.0; http://morus.swu.edu.cn/morusdb/) (Li et al., 2014b). All mulberry TEs were retrieved from the MnTEdb (v 1.0, http://morus.swu.edu.cn/mntedb/) (Ma et al., 2015). All other plant TEs were retrieved from Repbase (http://www.girinst.org/repbase/) (Bao et al., 2015).

Transposable Element Annotation in Mulberry

The coverage and distribution of the mulberry TEs were analyzed by RepeatMasker (v 4.0.3; http://www.repeatmasker.org/) against a custom database with default parameters except for the Smith–Waterman cut-off (a parameter of the search engine, RMBlast), which was set at 255. The custom database contained all plant TEs retrieved from Repbase (Bao et al., 2015) and all mulberry TEs deposited in MnTEdb (Ma et al., 2015). Based on the output file of RepeatMasker (.out), a Perl program, which was encoded and kindly provided by Robert Hubley to work with RepeatMasker data (http://www.repeatmasker.org/; Institute for Systems Biology, Seattle, WA), was used to parse the matched regions of the respective TEs in the mulberry genome.
Relationship of Genes to Transposable Elements

First, correlations between the proportional coverage of TEs and genes were determined in each of the scaffolds whose length was 390,115 bp or more (González and Deyholos, 2012). Then, 30 representative scaffolds (10 gene-rich scaffolds, 10 TE-rich scaffolds, and 10 with a similar percentage of TEs and genes) were selected to construct the proportional coverage graphs to present the distribution of TEs and genes visually. To make the graphs more readable, all these selected scaffolds were divided into 50-kb bins and the Multi-Experiment Viewer (Saeed et al., 2006) was used to construct the graphs (González and Deyholos, 2012). Hypergeometric tests (R, version 3.4.0) (R Development Core Team, 2017) were used to determine whether any of the TE superfamilies had a specific distribution pattern in these representative scaffolds. Putative expression of the genes in scaffolds equal to or longer than scaffold N50 was determined based on the transcriptome data associated with the five tissues (root, flower, leaf, bud, and bark) of wild-type mulberry, which were downloaded from the MorusDB (Li et al., 2014b) (http://morus.swu.edu.cn/morusdb/). If the reads per kilobase per million mapped reads values of one gene from any of five tissues was more than one, the gene was considered to be putatively expressed (Mortazavi et al., 2008). Then, correlation between the proportions of TE coverage (total TE length in one scaffold/length of the scaffold) and putatively expressed genes (the number of putatively expressed genes in one scaffold/the number of genes in the scaffold) was determined (González and Deyholos, 2012).

To find out whether any of the TE superfamilies had a specific distribution pattern within genes or not, all gene-related regions were categorized into three regions: within genes, adjacent to the genes (within 2 kb up- or downstream of the gene), and adjacent 2 to 5 kb to the genes (up- or downstream from the gene). According to the position (start and end position) of one TE sequence in the scaffold, the middle point position of this TE was used to identify whether it belonged to the corresponding region or not (Supplemental Fig. S1) (González and Deyholos, 2012). After this, the number of TE hits within genes was compared with that in all scaffolds whose length was ≥390,115 bp (N50) in length using the hypergeometric tests (Johnson et al., 1992) and R (version 3.4.0) (R Development Core Team, 2017). The same analyses were performed to compare the TE distributions in the other two regions.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Analysis of Transposable Element–Containing Genes

If there is one or more TEs present within a gene region, the gene is regarded as a TE-containing gene (Supplemental Fig. S1) (Jiao and Deng, 2007). Our own Perl script was designed to identify all TE-containing genes according to the positions of the TEs and the genes in the mulberry genome. After this, the TE-containing genes were assigned to three functional classes by Blast2GO (v 2.6.4) (Conesa et al., 2005): biological processes, cellular components, and molecular functions. In addition, these TE-containing genes were annotated and analyzed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to further understand the gene functions (Kanehisa et al., 2008).

Transposable Element–Related Alternatively Spliced Genes

If one gene has two or more alternative transcript isoforms and the corresponding splice sites of this gene were within one TE sequencer, the gene was regarded as a TE-derived alternatively spliced gene (Li et al., 2014a). We designed our own Perl programs to identify the TE-derived genes according to the positions of TEs and alternatively spliced genes. The graph was constructed using the heatmap package (version 1.0.8; https://github.com/raivokolde/heatmap, designed by Raivo Kolde) within R (version 3.4.0) (R Development Core Team, 2017). Our own Perl script was used to construct the gene structure of TE-derived alternatively spliced genes. Tissues (root, bark, flower, and leaf) of wild-type mulberry, which were collected from Ya’an, Sichuan Province, China, were used to perform reference transcript–polymerase chain reaction analysis of the selected gene. Primer 3 (http://bioinfo.ut.ee/primer3/) was used to design gene-specific primers with default parameters (Untergasser et al., 2012). All polymerase chain reaction (PCR) products of alternatively spliced (AS) events should be amplified by the gene-specific primers. Total RNA of tissues was extracted with RNAiso plus kit (Takara) according to standard protocols. Then, RNA was treated with RQ1 RNase-Free DNase (Promega) to eliminate DNA according to standard protocols. First-strand cDNA was synthesized from 3 μg total RNA from each sample by Moloney–Murine Leukemia Virus reverse transcriptase (Promega). The PCR amplification process was performed using the StepOnePlus system (Applied Biosystems). Each reaction was mixed with 20 ng cDNA, 10 pmol of each primer (BGI, 0.25 mM of each dNTP (Takara), 10× PCR buffer (including Mg2+; Takara), and 1 U of rTaq polymerase (Takara) in a final volume of 20 μL. The PCR protocol was used as follows: 95°C for 30 s; 35 cycles of 95°C for 10 s, 55°C for 30 s, and 72°C for 10 s, with a final elongation step at 72°C for 1 min. The PCR products were purified from 1.5% agarose gels by Agarose Gel DNA Extraction Kit (Takara). Then, PCR products were cloned into the pMD19-T vector (Takara) according to the manufacturer’s instruction. Positive clones were verified by PCR again and sequenced using M13 universal primers by Sangon Biotech.

RESULTS

Correlation Between the Proportional Coverage of Transposable Elements and Genes

There were 110,759 scaffolds in the assembly genome and the scaffold N50 length (weighted median contig size) was 390,115 bp (He et al., 2013). We limited our analysis to the 245 scaffolds that were 390,115 bp or longer, since this allowed correlation analyses of more contiguous TEs and
genes (Fig. 1A). These 245 scaffolds occupied >50% of the length of the assembly genome and contained 55% of the genes (Supplemental Table S1, S2). A significant inverse correlation was observed between the proportional coverage of genes and TEs \( (r = -0.759, p < 0.01) \), which indicated that there are specific distribution patterns of different TE superfamilies in the mulberry genome (Fig. 1A). Some scaffolds had a high proportion of genes (with few TEs), while other scaffolds had dense TE coverage (with few genes), while still, other scaffolds had similar proportion of genes and TEs (Fig. 1A). To visualize the distribution patterns of TEs and genes in those scaffolds, 30 representative scaffolds (red dots in Fig. 1A, with detailed information on these 30 scaffolds being shown in Supplemental Table S3) were chosen to build heat maps and coverage graphs to present the distribution trends of TEs and genes. Each dot represents the two \((x, y)\) values from one scaffold.

### Correlation Between the Percentage of Putatively Expressed Genes and Transposable Element Proportional Coverage

Out of a total of 14,909 genes in these 245 scaffolds, 11,162 (74.9%) were considered to be putatively expressed (Supplemental Table S1, S4). It was found that the TE proportional coverage was significantly negatively correlated \( (r = -0.556, p < 0.01) \) with the percentage of putatively expressed genes in each scaffold (Fig. 1C).

### Distribution Patterns of Transposable Elements in the Mulberry Genome

In light of the observation that some bins in TE-rich scaffolds had higher frequencies of genes than TEs, and some bins in gene-rich scaffolds had higher frequencies of TEs than genes (Fig. 1B), hypergeometric tests were performed to determine whether any of the TE super-families had a significantly greater tendency to be distributed close to or within genes.
There was a significant difference between the actual number of TEs distributed within genes and the number expected by chance for each of the Copia, Lard, Gypsy, hAT, Helitron, RTE, MULE, and L1 superfamilies ($p < 0.01$; Fig. 2A). Among these TEs, Lard, Copia, RTE, and L1 showed a greater tendency to be distributed within genes, while Gypsy, hAT, Helitron, and MULE were found outside genes more often than expected.

The same analysis was also performed in the other two regions, namely adjacent (within 2 kb), and adjacent (2–5 kb) (upstream or downstream of the genes in each case). The results were not significantly different for the L1 superfamily (Fig. 2B), but there were significant
 differences for the Trim and RTE superfamilies \( (p < 0.05) \)
and for the Copia, Lard, Gypsy, hAT, Helitron, and MULE superfamilies \( (p < 0.01) \) (Fig. 2B). Lard, hAT, PIF-Harbin-
ger, CACTA, RTE, and TcMar showed a greater tendency to be located in the adjacent 2-kb regions surrounding genes, while Copia, Gypsy, Helitron, Trim, and MULE were significantly underrepresented in this region.

Seven TE superfamilies showed significant differences in frequency in the adjacent 2- to 5-kb regions surrounding genes (Fig. 2C), whereas hAT, PIF-Harbin-
ger, Trim, TcMar, and CACTA were significantly overrepre-
sented in the adjacent 2- to 5-kb region around genes, with two superfamilies (Copia and Gypsy) being signifi-
cantly underrepresented in this region (Fig. 2C).

Kyoto Encyclopedia of Genes and Genomes Pathway and Gene Ontology analysis of Transposable Element–Containing Genes

There were 4596 (30.83%, 4596/14,909) TE-containing genes across all 245 scaffolds (Supplemental Table S5). A total of 2691 (58.55%, 2691/4,596) TE-containing genes could be classified into one or more GO terms (Supplemental Fig. S4).

The KEGG pathway-based annotations and analyses were performed to further understand the biological function of these TE-containing genes (Fig. 3; Supple-
mental Table S6). A total of 344 genes were mapped to 93 pathways. These genes were classified into four main categories. About 95.93% (330/344) of the genes were assigned to metabolism-related pathways followed by genetic information processing (3.78%, 13/344), environ-
mental information processing (1.16%, 4/344), and organismal systems (0.29%, 1/344).

Transposable Element–Related Alternatively Spliced Genes

In the 245 scaffolds, there were 1414 TE-containing genes having a total of 5302 alternative transcript isoforms (Supplemental Table S7). Only when the splicing sites of one gene were within TE sequences was the gene regarded as being a TE-derived AS gene. Our results indicated that a total of 402 TE-derived AS events, which accounted for 7.58% (402/5,302) of all AS events, were identified in 117 genes (Fig. 4; Supplemental Table S7; Supplemen-
tal Fig. S5). Specifically, for the genes that exhibit TE-
derived AS, alternative 3′ splicing accounted for 12.69% (51/402), alternative 5′ splicing for 16.17% (65/402), exon skipping for 65.17% (262/402), and intron retention for 5.97% (24/402) of the AS events. Some computationally predicted TE-derived AS events were selected for valida-
tion by experiment, and our experiments results were consistent with the AS events predicted by computational methods (Fig. 4; Supplemental Fig. S6).

DISCUSSION

Relationship Between Transposable Elements and Genes

A significant inverse correlation between the propor-
tional coverage of genes and TEs suggested that the distribution of TEs in the mulberry genome was not completely random (Fig. 1). The results were consistent with those from previous studies, and the explana-
tion proposed for the earlier research was that purifying selection had operated against TEs in gene coding regions to avoid disadvantageous effects on gene function (Paterson et al., 2009; Pereira, 2004; Wright et al., 2003).

Distribution pattern analysis was performed to test whether any of the TEs had a greater tendency to be located close to or within genes (Fig. 2). Lard elements were always overrepresented in the regions within genes and up to 2 kb from genes. These results were in contrast to those from previous studies, which had suggested that Lard elements were mainly located in pericen-
tromic regions of the rice and barley \((Hordeum vulgare\) L.) genomes (Jiang et al., 2002; Kalendar et al., 2004) and heterochromatin regions of the pepper \((Capsicum annuum\) L.) genome (Park et al., 2012). Meanwhile, Copia retrotransposons (representing the largest proportion of the mulberry genome occupied by any one element, 10.44%) were significantly overrepresented within genes, an observation that is consistent with that reported in \(A. thaliana\) (Lockton and Gaut, 2009). Copia elements were...
Fig. 4. Analysis of transposable element (TE)–derived alternative splicing types in the mulberry genome. (A) Summary of TE-derived alternatively spliced genes in the five tissues: root, bark, bud, flower, and leaf. A5SS, alternative 5' splicing; A3SS, alternative 3' splicing; ES, exon skipping; IR, intron retention. The number displayed in each cell indicates the number of TE-derived alternatively spliced genes in the corresponding tissues. The graph was constructed using the pheatmap package (version 1.0.8, designed by Raivo Kolde) within R (version 3.4.0). (B) An example of alternative-splicing (AS) events in the mulberry genome. The top black line indicates the gene position and direction. A black box, dashed line, and red box represent exon, intron, and TE, respectively. The y-axis indicates reads number. The bent lines above each figure denote a regular junction. The bent lines under each figure represent an alternative junction. Only AS events with more than three reads coverage were used to carry out the analysis. The graph was constructed using our own Perl script. (C) An example of experimental test of AS events. Primer 3 (http://bioinfo.ut.ee/primer3/) was used to design gene-specific primers with default parameters. The letters F and R with arrows indicate the position and direction of the primer pair. Primer sequences can be found in Supplemental Table S8. Red box represents a TE position. A black box denotes a corresponding polymerase chain reaction (PCR) product. All PCR products were sequenced by Sangon Biotech (Shanghai, China). As a result, two AS events, namely IR and A3SS, were validated by experiment.
overrepresented within gene-adjacent region between 1 kb up- and down-stream of flax (*Linum usitatissimum* L.) genes (González and Deyholos, 2012). Similar phenomena had also been found in soybean [*Glycine max* (L.) Merr.], in which *Copia*-type retrotransposon (SOR-E-1) insertions are more recently and frequent in chromosome arms than in pericentromeric regions (Nakashima et al., 2018). A typical example is that it inserted in the first intron of the *GmFT2a* gene of soybean where it attenuated the expression level of this gene, causing delays in flowering (Zhao et al., 2016). Random patterns of *Copia* insertions reported in *A. thaliana* (Pereira, 2004) may also be the reason that they insert within or close to gene regions of those species. As for *Gypsy* elements, they were always underrepresented in and around genes, a finding that is in agreement with the situation in *A. thaliana* (Lockton and Gaut, 2009), poplar (*Populus trichocarpa* Torr. & A. Gray) (Cossu et al., 2012), flax (González and Deyholos, 2012), sunflower (*Helianthus annuus* L.) (Cavallini et al., 2010), and several grass species (Miller et al., 1998). Previous studies had suggested that these elements seemed to be clustered around the centromeric regions of chromosomes (Neumann et al., 2011; Pereira, 2004; Staton et al., 2009), and a specific domain in the integrase protein of *Gypsy* elements may be the cause of this distribution pattern (Neumann et al., 2011; Pereira, 2004).

Several DNA TE superfamilies fell within or close to genes more often than would be expected by chance alone. For example, *CACTA*, *hAT*, *PIF-Harbinger*, and *TcMar* superfamilies were overrepresented in the 2-kb region adjacent to genes in the mulberry genome, which is consistent with the situation in grass species such as false brome (*Brachypodium distachyon* (L.) P. Beauv.), rice, sorghum (*Sorghum bicolor* (L.) Moench), and maize (Han et al., 2013). We also compared the distribution pattern of DNA TEs based on their size (Supplemental Fig. S7). Taken together, these data indicate that the size of DNA TEs should have an impact on their distribution pattern. Small DNA TEs tend to be located within or close to genes more often than would be expected by chance. Those small non-coding elements were enriched in the 1-kb region flanking the 5' and 3' ends of the coding regions and displayed conserved distribution patterns with the highest peak being in the promoter region of these grass species where it had an impact on the expression of adjacent genes (Han et al., 2013). Some typical examples were shown in previous studies. For example, experiments in rice suggested that one of a few active DNA transposons (belonging to the *hAT* superfamily), nDart1 and its relatives, tended to be inserted within or very close to genes (Takagi et al., 2010). The transposition of a reversed *Ac* (belonging to the *hAT* superfamily) element generated novel genes in maize (Zhang et al., 2006), while a *CACTA* element trapped a promoterless *ALP* gene leading to the duplication of the *ALP* family in wheat (Akhunov et al., 2007).

The correlation between the percentage of putatively expressed genes and TE coverage was analyzed to determine whether there were general patterns of gene expression regulation by TEs (Fig. 1C). Such a distribution pattern might have important implications. Transposable element can be a positive or negative regulator of gene expression depending on whether they have cis-acting elements or are targets for epigenetic silencing (Hollister et al., 2011). Our results suggested that genes situated in close proximity to TEs are affected negatively because these regions are often targeted for heterochromatinization and silencing, suggesting that genes may be disrupted by these TEs (Hollister et al., 2011). According to the KEGG pathway analysis results, a total of 344 TE-containing genes were mapped to 93 KEGG pathways (Fig. 3; Supplemental Table S6) and 95.93% (330/344) genes were assigned to metabolism-related pathways.

### Transposable Elements are a Driving Force Behind the Formation of Alternatively Spliced Genes

Alternative splicing creates multiple alternative transcript isoforms from a single gene and has greatly contributed to gene regulation and proteome diversity. Alternative splicing is involved in most plant processes such as cell fate, the circadian clock, plant defense, and response to external cues (Huang et al., 2015). A total of 117 TE-derived alternatively spliced genes were shown to have one or more alternative transcript isoforms in the mulberry genome (Fig. 4; Supplemental Table S7; Supplemental Fig. S5). Previous studies had suggested that TE insertion has greatly contributed to AS regulation (Huang et al., 2015; Li et al., 2014a). Transcriptomics analysis suggested that a total of 98 TE-derived intron retention events had been found in maize (Huang et al., 2015). As reported by Li et al. (2014a), 43% (2246) of all retained introns were directly related to TEs in *Gossypium raimondii* Ulbr. Similar percentages of TE-related retained introns have also been found in soybean (45%, 2025), sorghum (51%, 2323) and rice (49%, 2094) (Li et al., 2014a). The efficiency of intron splicing in genes would be reduced when TEs were distributed upstream of the 3'-splice sites of genes. As a result, the intron would be retained (Li et al., 2014a). Taking these results in conjunction with those from previous studies, TEs are believed to play important roles in the regulation of adjacent genes, using several mechanisms (Feschotte, 2008; Hollister et al., 2011).

In summary, we showed that the distribution of TEs was not completely random. A significant inverse correlation relationship was observed between the proportional coverages of genes and TEs. The TE-rich regions appeared to have a lower percentage of putatively expressed genes. Distribution patterns of different TE superfamilies were detected. *Copia* elements, which occupied the largest proportion of the mulberry genome of all TE superfamilies, were found within genes more often than expected, while *Gypsy* elements (the second largest TE component of the mulberry genome) were less commonly found within genes than expected. Considering the content and distribution patterns of *Copia* elements, they are regarded as primary shapers of the mulberry genome. The result that four classical AS events...
accounted for 7.58% (402/5302) of all AS events in the mulberry genome suggested that TEs have greatly contributed to AS regulation. Transposable elements are one of the driving forces behind the formation of AS genes. Further studies should focus on the important roles TEs play in the regulatory effects on adjacent genes and in the architecture of the mulberry genome.

Conflict of Interest
The authors declare that there is no conflict of interest.

Supplemental Material Available
Supplemental material is available online for this article.

Author Contributions
BM conceived and designed the experiments; performed the experiments; analyzed the data; contributed reagents, materials, and analysis tools; wrote the paper; prepared figures and tables; and reviewed drafts of the paper. YX and LK performed the experiments, analyzed the data, and reviewed drafts of the paper. NH conceived and designed the experiments and reviewed drafts of the paper. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS
The authors thank all the laboratory members who provided advice during this work. The authors would also like to thank everyone who made the data and tools freely available for the present study. This project was funded by Chongqing Research Program of Basic Research and Frontier Technology (cstc2018jcyjAX0407), Fundamental Research Funds for the Central Universities (SWU118040, XDJK2015C116), and China Postdoctoral Science Foundation (2016M592622).

REFERENCES


