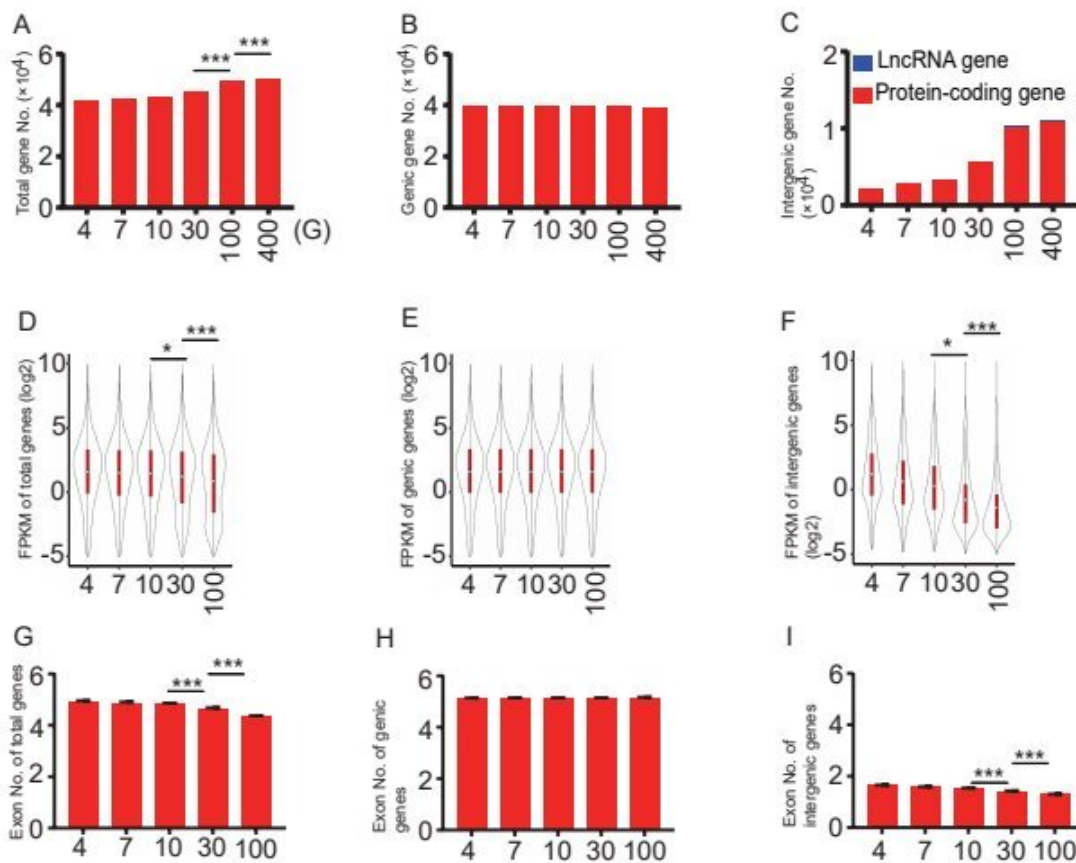


Large-scale long terminal repeat insertions found to produce a significant set of novel transcripts in cotton

May 24 2023

Figure 1



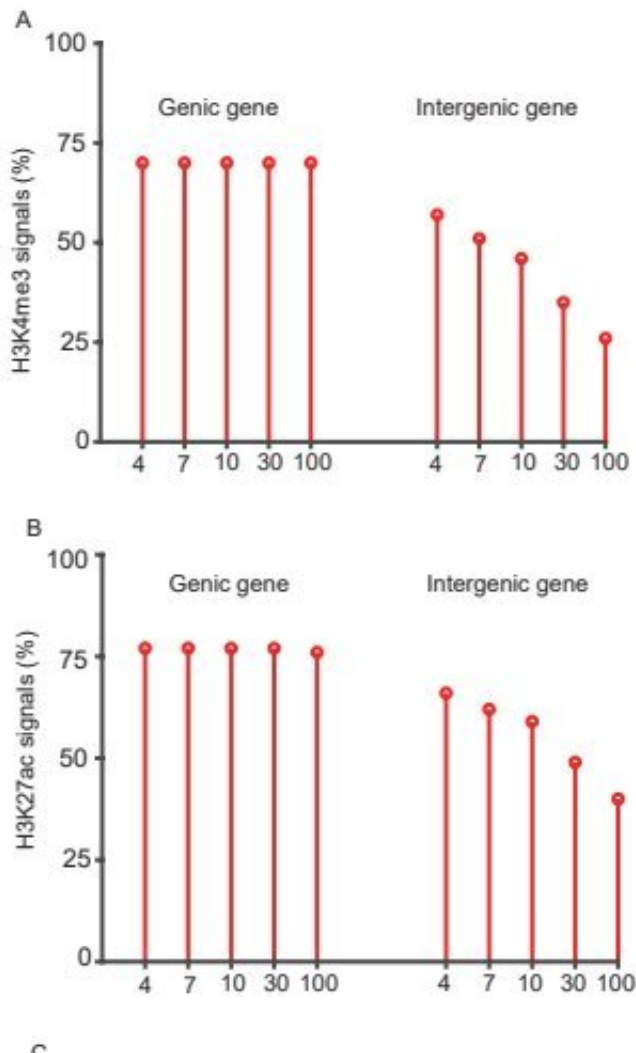
Comparisons of the gene numbers obtained from diploid cotton *G. arboreum* transcriptomes at different sequencing depths. Credit: Science China Press

TEs (transposable elements), especially LTRs, are known to play an important role in determining the basic genome structure and influencing the expression of functional genes. Insertion of TE or LTR fragments may also create novel transcription start sites (TSSs) to initiate transcription in the host genome. New intergenic transcripts were thought to be created by terminal repeat retrotransposon insertions using a combination of de novo and homology-based strategy in maize.

Although these studies have predicted the possibility of new transcript production by transposon insertion, they do not reveal the evolutionary, regulatory and functional mechanisms of these new transcripts. Furthermore, there is not even one systematic study on the extensiveness of intergenic transcript production at the genomic level so far.

In a study published in the journal *Science China Life Sciences*, Yuxian Zhu and their colleagues applied extremely deep-sequencing techniques (from 10 G to over 100 G) in each cotton sample to discover more than 10,000 novel genes that were largely not identified in previous genome assembly and annotations. Most of these transcripts were protein-coding in nature and were created by LTR insertions in various ways.

Figure 4



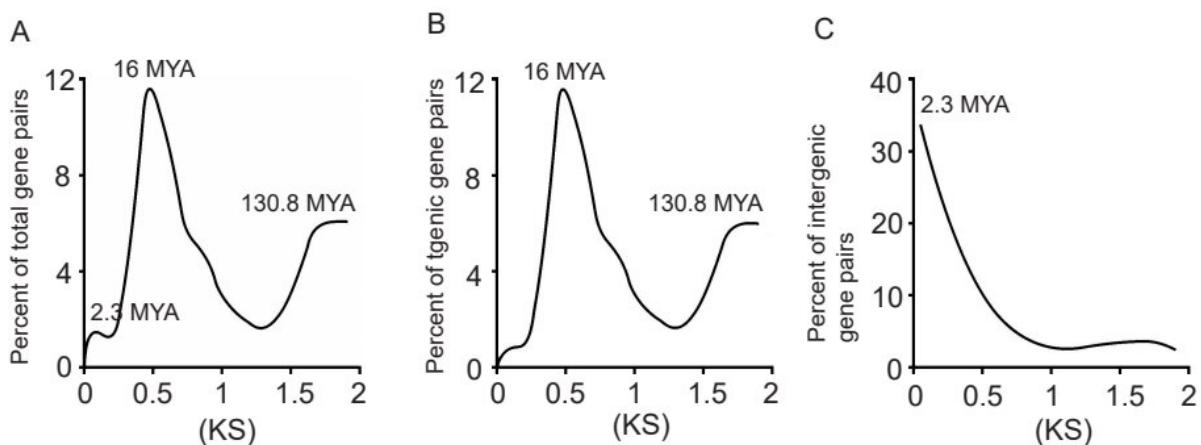
ChIP-seq analysis of H3K4me3, H3K27ac, and H3K9me2 markers in genic and intergenic regions at different sequencing depths. Credit: Science China Press

The team found that more transcripts appeared mainly in intergenic regions as identified in the previously published genome. In the 100 G data set, a total of 10,284 new intergenic genes were discovered. In total, 10,032 are [protein-coding genes](#) and 252 were lncRNA genes. There was no significant increase in genic gene numbers between these two groups. Generally, these new intergenic transcripts were expressed at very low

levels, and most of them were single exon transcripts.

These new intergenic transcripts appeared only when the sequencing depth reached to 30 G to 100 G due to their low expression level. ChIP-seq analysis with antibodies against H3K4me3, H3K27ac and H3K9me2 revealed that most of these new transcripts might not be transcribed by RNA polymerase II. Only 30% of these intergenic transcripts possessed one or two transcription activation markers while greater than 70% of the genic genes contained these markers.

MNase-seq analysis revealed that genes without transcription activation markers formed their +1 and -1 nucleosomes significantly more closely (only 117 ± 1.4 bp apart), while twice as big the spaces (about 403.5 ± 46.0 bp apart) were found for genes with the activation markers. Genes without one of these two markers intended to form -1 nucleosomes at the close vicinity of their +1 nucleosomes. This may impede the binding of the RNA polymerase.



Evolutionary analysis for the origin of genic genes and intergenic transcripts in the *G. arboreum* genome. Credit: Science China Press

Evolutionary analysis showed that genic genes were originated during one of the whole genome duplication events around 130.8 or 16 MYA, while ITG transcripts were evolved around 2.3 MYA, resultant of the last retrotransposon insertion.

Characterization of these low-transcribed ITG transcripts will help us understand the biological roles of retrotransposons during speciation and diversifications. This study may help elucidate the mechanisms related to intergenic [transcript](#) expression and cotton fiber development.

More information: Yan Yang et al, Large-scale long terminal repeat insertions produced a significant set of novel transcripts in cotton, *Science China Life Sciences* (2023). [DOI: 10.1007/s11427-022-2341-8](https://doi.org/10.1007/s11427-022-2341-8)

Provided by Science China Press

Citation: Large-scale long terminal repeat insertions found to produce a significant set of novel transcripts in cotton (2023, May 24) retrieved 7 May 2024 from <https://phys.org/news/2023-05-large-scale-terminal-insertions-significant-transcripts.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.