Persistence of sub-genomes in paleopolyploid cotton after 60 million years of evolution

Simon Renny-Byfield, Lei Gong, Joseph P. Gallagher, Jonathan F. Wendel*

Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50014, USA.

*Corresponding author
email: jfw@iastate.edu

keywords: whole genome duplication, gene fractionation, biased fractionation, transposable element, gene expression, GC content

Abstract

The role of gene duplication as an evolutionary force has long been recognized. Indeed, ancient whole genome multiplication (WGM) is cyclical and ubiquitous in flowering plants, with many lineages having experienced multiple rounds of duplication, each followed by substantial duplicate gene loss (fractionation). The process of fractionation may be biased with respect to homoeologs, often with over-expression of genes in the less fractionated (LF) relative to more fractionated (MF) regions. The correlation between fractionation and expression has been hypothesized to arise via positional-effect down-regulation of genes via silencing of local transposable elements (TEs). Using comparative genomics and analysis of synteny, we report that the genomic signatures of biased fractionation of homoeologous segments remain evident in the paleopolyploid genome of modern diploid cotton (*Gossypium*), which underwent a 5- to 6-fold ploidy increase around 60 million years ago. Our data stand in contrast to other reports by extending the evolutionary time over which these signatures can be detected, suggesting that, hitherto, the long term impact of bias fractionation has not been fully appreciated. As in other species, biased gene loss is associated with over-expression of genes in LF regions. Furthermore we report that LF and MF fractions can be clearly differentiated via several genomic signatures despite the antiquity of the WGM event. For example, MF regions have elevated GC content, higher transposon (TE) density and elicit preferential mapping of siRNAs. Contrary to previous observations, we show that global gene expression levels are not influenced by TE proximity, nor local siRNA targeting, both of which are therefore unlikely to be the primary drivers of biased fractionation. We propose an alternative scenario in which repeat content *per se* and differential recombination rates may be responsible for biased fractionation of sub-genomes following WGM.