Iowa State University



Ecology, Evolution, and Organismal Biology

of science and technology

July 21, 2018

Dear Editorial Board of *Genome Biology and Evolution*,

We are pleased to submit a manuscript entitled “Insights into the evolution of the New World diploid cottons (*Gossypium*, subgenus *Houzingenia*) based on genome sequencing” for consideration as a *Research Article* in *Genome Biology and Evolution*. In this manuscript, we leverage newly generated genome sequences for all species of the cotton subgenus *Houzingenia* to understand molecular and genome evolution among closely related species on a recent time scale.

Here, we employ phylogenomic methods to study molecular evolutionary processes and phylogeny in the geographically widely dispersed New World diploid cottons (*Gossypium*, subg. *Houzingenia*) using whole genome resequencing data (average of 33X genomic coverage) generated here. We find that the subgenus likely originated following trans-oceanic dispersal from Africa about 6.6 mya, but that nearly all of the biodiversity evolved following rapid diversification in the mid-Pleistocene (0.5-2.0 mya), with multiple long-distance dispersals required to account for range expansion to Arizona, the Galapagos Islands and Peru. We compare the sequences of the chloroplast with nuclear genes, which indicate that this history was accompanied by several clear cases of interspecific introgression.

We also assess various aspects of molecular evolution in the subgenus. While repetitive DNAs comprise roughly half of the genome, most transposable element families are relatively old and stable among species. Over 1.1 million indels were detected and phylogenetically polarized, revealing a two-fold bias toward deletions over insertions. We suggest that this genome down-sizing bias counteracts genome size growth by TE amplification and insertions, and helps explain the relatively small genomes that are restricted to this subgenus. Compared to the rate of nucleotide substitution (estimated here at approximately 1% per million years), the rate of indel occurrence is much lower, averaging about 17 nucleotide substitutions per indel event.

We believe that this manuscript will be of broad interest to readers and those interested in genome evolution. All authors have read and approved the manuscript, and we look forward to receiving reviews on this manuscript.

Thanks for your consideration,

Corrinne Grover, PhD

Associate Scientist

Iowa State University

Ames, IA 50011

corrinne@iastate.edu