RESEARCH

Thirteen genomes

Corrinne E Grover[1*†], Mark A Arick II[2], Justin C Conover[1], Jodi A Scheffler[3], William S Sanders[2], Daniel G Peterson[2], Brian E Scheffler[3] and Jonathan F Wendel[1]

*Correspondence: corrinne@iastate.edu
[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Background: Text for this section.

Results: Text for this section.

Conclusions: Text for this section.

Keywords: genome sequence; cotton; *Gossypium*; molecular evolution

Background

The American diploid "D-genome" cottons (subgenus *Houzingenia*) comprise a monophyletic clade of cytogenetically and morphologically distinct species largely distributed from Southwest Mexico to Arizona, with additional disjunct species distributions in Peru and the Galapagos Islands Corrinne. Among the 13 species currently included in the D-genome Corrinne are*G. harknessii* (D2-2), an important species for cytoplasmic male sterility in cotton, and *G. raimondii* (D5), the model diploid progenitor to wild and domesticated allopolyploid cotton Corrinne. The close relationship of *Houzingenia* species to the agronomically important polyploids, combined with the relative ease of sampling this subgenus for early cotton taxonomists, facilitated much of the current understanding of the relationships among D-genome species.

These early taxonomists divided subgenus *Houzingenia* into two sections and six subsections, whose species alliances have largely been retained by subsequent phylogenetic studies Corrinne. Several molecular datasets have been used to evaluate these relationships, including chloroplast restriction sites Corrinne; simple sequence repeat (SSR) and expressed sequence tag (EST)-SSR markers Corrinne; random amplified polymorphic DNA (RAPD) markers Corrinne; internal transcribed sequences (ITS) Corrinne; and few single-copy nuclear genes Corrinne. Relationships among the six subsections, however, remain unclear despite numerous, and often conflicting, studies Corrinne. Determining the closest living relative of the D-genome ancestor to the polyploid, however, has been met with greater success. Early morphological and cytogenetic comparisons using intergenomic hybrids quickly identified *G. raimondii* as the closest living relative to the D-genome ancestor of polyploid cotton species Corrinne. Subsequent analyses have largely supported this observation (Abdalla et al., 2001; cronn 1999, liu 2001, Cronn et al., 1996 Seelanan et al., 1997 Small et al., 1998; Small and Wendel, 2000a,b), with few conflicts Corrinne.

A secondary outcome of this research has been the elucidation of multiple instances of hybridization among D-genome (i.e, *Houzingenia*) species Corrinne, and, in one remarkable case (i.e., *G. gossypioides*), between a *Houzingenia* species and

Corrinne: which c
tions here, Endri
et al 1985

Corrinne: althoug
see Ulloa 2013

Corrinne: Endriz
1985, others?

Corrinne: Cronn
al., 1996; Seelana
et al., 1997; Smal
and Wendel, 200
Wendel and Albe
1992; Wendel et a
1995b; alvarez 20

Corrinne: citation

Corrinne: citation

Corrinne: citation

Corrinne: citation

Corrinne: citation

Corrinne: Cronn
al., 1996; Liu et a
2001b; Small and
Wendel, 2000b m
citations

Corrinne: Stephe
1944b Hutchinso
et al., 1945, 1947
Gerstel and Phill
1958; Phillips, 19
reviewed in Wend
and Cronn 2003

another, geographically isolated subgenus from Africa (either A-, B-, or, F-genome Corrinne). Notably, *G. gossypioides* is multiply introgressant, with subsequent hybridization to a member of the *G. raimondii* lineage resulting in chloroplast, if not further (and cryptic), nuclear introgression (Cronn 2003, cryptic trysts). Cytoplasmic introgression, and possibly cryptic nuclear, is also present in some populations of *G. aridum*, i.e., the Mexican Colima populations; *G. aridum* accessions derived from this location possess a *G. davidsonii*- or *G. klotzschianum*-like cytoplasm.

Corrinne: cite cryptic Trysts again

Modest attempts at understanding the evolution of the repetitive fraction of this genus support the inference of African introgression in *G. gossypioides* Corrinne; however, little else is understood with respect to the evolution of the non-genic fraction of *Houzingenia*. The D-genome cottons possess the smallest genome sizes in the genus, ranging only 1.11 fold, from 841 Mb – 934 Mb. Notably, the distribution of genome sizes among the subsections suggests that this subgenus has experienced differential growth and/or reduction in genome size among species Corrinne; however, the patterns of sequence gain and loss have not been characterized for the subgenus. While the differences in genome size are not dramatic, there is evidence that the transposable element types which have accumulated in *G. raimondii* are different than those that have achieved higher copy numbers than the remainder of the genus Corrinne. Furthermore, research comparing the two sister genera to cotton (i.e., *Kokia* and *Gossypioides*; Corrinne) reveals that their apparently static genome sizes belies both gain and loss of repetitive sequence, a result similar to that of the extant members of the A-genome (subgenus *Gossypium*), whose small change in genome size (1.05X) masks differences in element accumulation Corrinne.

Corrinne: Zhao 1998

Corrinne: need a figure

Corrinne: Jennifer, A domestication, kokia

Corrinne: cite Kokia paper

Corrinne: cite Agenome domestication

Modern sequencing techniques make it easy to produce a substantial amount of genomic sequencing suitable for addressing these basic questions in a more genomically comprehensive manner. Here we use modest coverage Illumina sequencing to present an in-depth view of the evolution subgenus Houzingenia, the cotton D-genome clade. We leverage newly generated genome and plastome sequences, representing the first for many species, to address questions surrounding genome evolution in a monophyletic group of closely related species. We characterize the patterns of molecular evolution of both genes and repetitive sequences to provide insight into the pace and pattern of evolution in this subgenus. For the first time, intergenic regions are evaluated to characterize the amount of divergence outside of genes, and due to indels or single-nucleotide polymorphisms (SNPs). Finally, we revisit the phylogeny of the D-genome, both adding additional insight into the relationships among species using hundreds of nuclear genes, as well as addressing questions regarding sequence gain and loss among closely related species. The genome characterized here not only provides insight into molecular evolution on a relatively recent timeframe, but it also provides resources for comparative research and the cotton community at large.

Results

Text for this section ...

Results subsection

Text for this sub-heading ...

Discussion
Conclusions
Methods
Sequence generation and initial processing
DNA was extracted from (LEAVES) using (WHAT KIT), and sent to (WHERE) for library construction and sequencing. Sequencing was completed on the Illumina (WHAT MACHINE) using (WHICH SEQUENCING). The data were trimmed and filtered with Trimmomatic v0.32 Corrinne with the following options : (1) sequence adapter removal, (2) removal of leading and/or trailing bases when the quality score (Q) <28, (3) removal of bases after average Q <28 (8 nt window) or single base quality <10, and (4) removal of reads < 85 nt. Detailed parameters can be found at https://github.com/williamssanders/D_Cottons_USDA. Corrinne

Corrinne: citation

Corrinne: Let's port this repo to a lab site after and give the new url

Genome assembly and annotation
The trimmed data was independently assembled for each species via ABySS v2.0.1 Corrinne, using every 5th kmer value from 40 through 100. A single assembly with the highest E-size (an alternative statistic to N50; Corrinne) was selected for each species and subsequently annotated with MAKER v2.31.6 Corrinne using evidence from: (1) the NCBI *G. raimondii* EST database Corrinne, (2) *G. raimondii* reference genome predicted proteins, as hosted by CottonGen.org Corrinne, and (3) three *ab initio* gene prediction programs, i.e. Genemark v4.30 Corrinne, SNAP v2013-11-29 Corrinne, and Augustus v3.0.3 Corrinne. Both the SNAP and Augustus models were trained using BUSCO v2.0 Corrinne.

Corrinne: citation

Corrinne: citation Salzberg 2011

Corrinne: citation

Corrinne: citation

Corrinne: citation

Corrinne: citation

Corrinne: citation

Gene stuff
Gene orthology and family designations were determined via OrthoFinder Corrinne...

Corrinne: citation

Repetitive characterization
Phylogenetics and ancestral state reconstruction

Competing interests
The authors declare that they have no competing interests.

Author's contributions
Text for this section . . .

Acknowledgements
We would like to thank....

Author details
[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA. [2]Institute for Genomics, Biocomputing, and Biotechnology, Mississippi State University, Mississippie State, MS 39762, USA. [3]Jamie Whitten Delta States Research Center, USDA-ARS, Stoneville, MS 38776, USA.

References
Figures

Figure 1 Sample figure title. A short description of the figure content should go here.

Figure 2 Sample figure title. Figure legend text.

Table 1 Sample table title. This is where the description of the table should go.

	B1	B2	B3
A1	0.1	0.2	0.3
A2	...	..	.
A3	..	.	.

Tables

Additional Files

Additional file 1 — Sample additional file title

Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title

Additional file descriptions text.