



The Genome Sequence of *Gossypioides kirkii* Illustrates a Descending Dysploidy in Plants

Joshua A. Udall^{1*}, Evan Long², Thiruvarangan Ramaraj^{3,4}, Justin L. Conover⁵, Daojun Yuan^{5,6}, Corrinne E. Grover⁵, Lei Gong⁷, Mark A. Arick II⁸, Rick E. Masonbrink⁹, Daniel G. Peterson⁸ and Jonathan F. Wendel^{5*}

¹ Crop Germplasm Research, USDA, College Station, TX, United States, ² Plant Breeding and Genetics, Cornell University, Ithaca, NY, United States, ³ National Center of Genome Resources, Santa Fe, NM, United States, ⁴ School of Computing, DePaul University, Chicago, IL, United States, ⁵ EEOB Department, Iowa State University, Ames, IA, United States, ⁶ College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, ⁷ Key Laboratory of Molecular Epigenetics of the Ministry of Education, Northeast Normal University, Changchun, China, ⁸ Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, MS, United States, ⁹ Genome Informatics Facility, Iowa State University, Ames, IA, United States

OPEN ACCESS

Edited by:

Martin A. Lysak,
Masaryk University, Czechia

Reviewed by:

Robin Van Velzen,
Wageningen University & Research,
Netherlands
Tae-Soo Jang,
Chungnam National University,
South Korea
Adam Lukaszewski,
University of California, Riverside,
United States

*Correspondence:

Joshua A. Udall
joshua.udall@usda.gov
Jonathan F. Wendel
jfw@iastate.edu

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 16 May 2019

Accepted: 05 November 2019

Published: 27 November 2019

Citation:

Udall JA, Long E, Ramaraj T, Conover JL, Yuan D, Grover CE, Gong L, Arick MA II, Masonbrink RE, Peterson DG and Wendel JF (2019) The Genome Sequence of *Gossypioides kirkii* Illustrates a Descending Dysploidy in Plants. *Front. Plant Sci.* 10:1541. doi: 10.3389/fpls.2019.01541

One of the extraordinary aspects of plant genome evolution is variation in chromosome number, particularly that among closely related species. This is exemplified by the cotton genus (*Gossypium*) and its relatives, where most species and genera have a base chromosome number of 13. The two exceptions are sister genera that have $n = 12$ (the Hawaiian *Kokia* and the East African and Madagascan *Gossypioides*). We generated a high-quality genome sequence of *Gossypioides kirkii* ($n = 12$) using PacBio, Bionano, and Hi-C technologies, and compared this assembly to genome sequences of *Kokia* ($n = 12$) and *Gossypium* diploids ($n = 13$). Previous analysis demonstrated that the directionality of their reduced chromosome number was through large structural rearrangements. A series of structural rearrangements were identified comparing the *de novo* *G. kirkii* genome sequence to genome sequences of *Gossypium*, including chromosome fusions and inversions. Genome comparison between *G. kirkii* and *Gossypium* suggests that multiple steps are required to generate the extant structural differences.

Keywords: speciation, chromosome evolution, cotton, structural rearrangements, *Gossypioideae*

INTRODUCTION

One of the extraordinary aspects of plant genomes is how variable they are in terms of chromosome number. Haploid chromosome counts among angiosperms span more than two orders of magnitude, from a low of $n = 2$ in six different species spread among four angiosperm families (Vanzela et al., 1996; Roberto, 2005), to 320 in the genus *Sedum* (Crassulaceae) (Uhl, 1978). Driving this diversity are mechanisms that both expand and shrink chromosome numbers, either saltationally *via* polyploidy, or in a more stepwise fashion *via* ascending or descending dysploidy. These processes have long been recognized as important in speciation (Stebbins, 1971; Grant, 1981) because of the impact of chromosome number divergence on reproductive isolation. Reflective of this, it is not uncommon for congeneric species to display either ascending or descending chromosome counts. From a mechanistic perspective, ascending or descending dysploidy can arise from several chromosome rearrangement processes (Jones, 1998; Guerra, 2008; Heslop-Harrison and Schwarzacher, 2011;

Lysák and Schubert, 2013; Weiss-Schneeweiss and Schneeweiss, 2013; Hoang and Schubert, 2017), including ascending dysploidy *via* chromosome fission along with the evolution of neocentromeres (Giannuzzi et al., 2013; Lysák and Schubert, 2013), and descending dysploidy through various chromosome fusion processes, including the difficult to distinguish telomere-to-telomere fusions and Robertsonian translocations (Schubert, 1992; Lysák and Schubert, 2013; Chiatante et al., 2017; Jarvis et al., 2017), and the acquisition of chromosome segments into other chromosomes (Luo et al., 2009; Murat et al., 2010; Vogel et al., 2010; Wang and Bennetzen, 2012; Fonsêca et al., 2016).

A prerequisite for understanding the directionality of chromosome number change in any taxonomic group is the availability of a well-established phylogenetic framework, so that hypotheses regarding ancestral and derived conditions are phylogenetically justified. Illustrative of this is the small monophyletic tribe *Gossypieae*, which contains the economically important cotton genus (*Gossypium*) as well as eight other lesser known genera (including *Thepparattia*) (Fryxell, 1979; Seelanan et al., 1997; Phuphathanaphong, 2006). More than 20 years ago, the Hawaiian *Kokia* and the East African/Madagascan *Gossypioides* were shown to belong to a single clade (Figure 1). Because these two genera have one fewer chromosomes ($n = 12$) than their sister genus *Gossypium* ($n = 13$), and because this assemblage is nested within other genera (e.g., *Hampea*, *Thespesia*) with a chromosome number of 13, they proposed an explanation involving aneuploid reduction in the lineage leading to *Kokia* and *Gossypioides* after divergence of this branch from *Gossypium*. Temporal perspectives to this reduction are the recent divergence time estimates of 5 million years (MY) for *Kokia* and *Gossypioides* and about 10 MY for the divergence of this clade from *Gossypium* (Wendel and Cronn, 2003).

Here we describe the genomic consequences of descending dysploidy in the *Kokia/Gossypioides* clade. We present a high quality *de novo* genome assembly for *Gossypioides kirkii* and compare this assembly to *Gossypium*, for which multiple assemblies have been generated. Comparison of our high quality genome assembly to other *Gossypium* genomes suggests that aneuploid reduction was accompanied by chromosome fusion

and other structural rearrangements. Assuming the *Gossypium* genome was representative of the ancestral genome, we developed a model of aneuploid reduction that included several structural rearrangements reducing three chromosomes to two chromosomes during the evolution of the ancestor to the *Kokia* and *Gossypioides* genera.

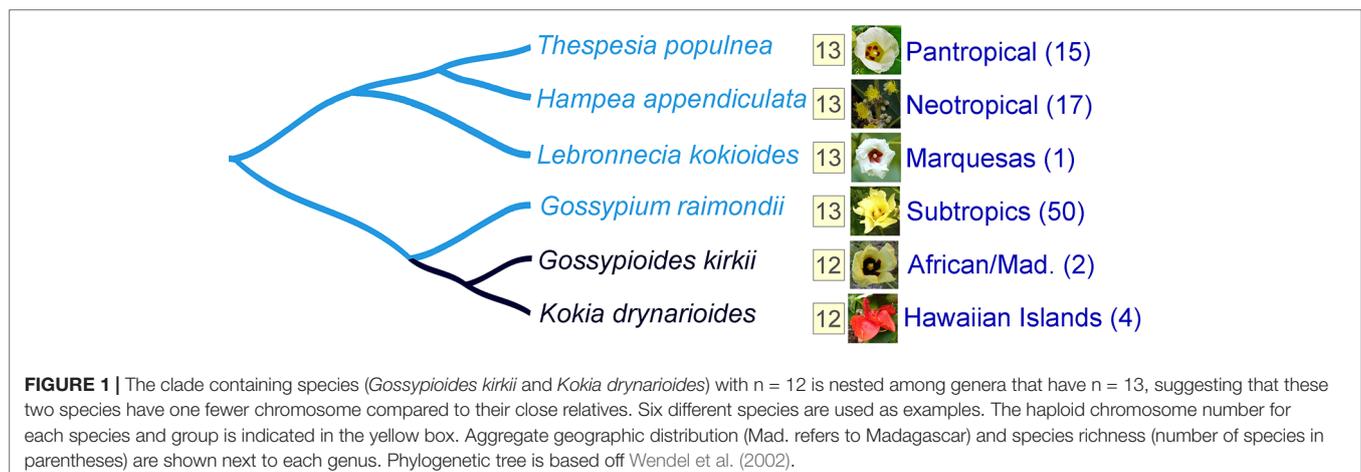
MATERIALS AND METHODS

Plant Material, Sequencing, and Assembly

G. kirkii leaves were collected from the Pohl Conservatory at Iowa State University and shipped to Brigham Young University for DNA extraction. Seven PacBio cells were sequenced at BYU from two libraries created from the same DNA source. Sequence reads were assembled (Table S1) using Canu V1.6 (Koren et al., 2017).

Leaf tissue of *G. kirkii* was also shipped to Phase Genomics (Seattle, WA) for DNA extraction and construction of HiC sequencing libraries. The sequenced HiC libraries generated 47× coverage of 125 bp paired-end Illumina reads; these were used to establish connections between contigs (Table S2). Illumina reads were mapped to the reference genome and a proximity guided assembly (PGA) was performed by Phase Genomics. High-molecular weight (HMW) DNA was extracted and labeled following the Bionano Plant protocol for the Irys system.

The optical map was aligned to the PGA assembly using an *in silico* labeled reference sequence. Conflicts between the Bionano map and the PGA assembly were manually identified in the Bionano Access software by comparing the mapped Bionano contigs and reference sequence to a bed file containing sequence contigs. These inter-species alignments, along with the Bionano alignments, guided manual rearrangements of scaffolds. Corrections to the PGA assembly removed conflicts between datasets by repositioning and reorienting sequence contigs in PGA ordering files. Corrections to the HiC scaffolding were made if more than one other genome agreed with the rearrangement and if the rearrangements coincided with contig breakpoints (i.e. scaffolding rearrangements). The contig order was arranged



to maximize the frequency of close linkages throughout the genome. The resulting fasta file of the scaffolded assembly was produced by concatenating PacBio contigs with 100 N bases between them. Several iterations of correction and realignment resolved nearly all of the conflicts between the sequence and Bionano assemblies. Similar iterations of HiC interaction maps were created using Juicer v1.5 and Juicebox v1.8.8, respectively, for the final manual adjustments to the genome sequence. Specifically, HiC reads were re-mapped to the modified sequence and the association frequency between each paired-end was used to adjust the genome sequence using JuiceBox (Durand et al., 2016). A custom python script from Phase Genomics was used to adjust the initially assembled pseudomolecules with the changes made to the genome *via* JuiceBox. Based on the HiC data, the *G. kirkii* pseudomolecule corrections consisted of two inversions and three translocations (involving seven of 12 chromosomes). These corrections established near-complete congruence between mapped paired-ends along the entire genome. The *G. kirkii* genome sequence is available from GenBank (Accession numbers: CP032244–CP032255).

Genome Alignments

The *G. kirkii* assembly was separately aligned by Minimap2 (Li, 2018) to other genomes in *Gossypium*, including *G. arboreum* (Du et al., 2018), *G. raimondii* (Paterson et al., 2012), and *G. hirsutum* (Zhang et al., 2015), and visualized using dotPlotly (Poorten, 2018). The alignments identified assembly errors in chromosomes Chr09 and Chr12 of *G. raimondii*. Telomere sequences were also used to confirm assembly completeness and structure by searching for the canonical telomere repeat (McKnight et al., 1997; Fajkus et al., 2005; Watson and Riha, 2010) in the *G. kirkii* genome using Geneious (Biomatters, New Zealand). The telomere repeat were also visualized and manually annotated in Geneious to verify telomere location on each chromosome.

Phase Genomics also constructed and sequenced a Hi-C library made from leaf tissue of *Kokia drynarioides*, a member of the genus sister to *Gossypioideis* that also shares $n = 12$, to further verify the structure. The Hi-C reads were mapped to the final, corrected version of the *G. kirkii* genome assembly using BWA. Approximately 10.7 M contacts (11% of the total paired reads) passed mapping filters and were used as Hi-C interaction evidence. Contact maps were visualized using Juicer v1.5 and Juicebox v1.8.8.

Transcriptomic Sequencing and Gene Annotation

Total RNA was extracted from 3-cm seedling leaves. Illumina TruSeq RNA-sequencing libraries were prepared for each replicate and were sequenced (Paired-end 150 bp) Berry Genomics Co. Ltd. (Beijing, China). Gene annotations were created using GenSAS 5.0 (Lee et al., 2011), an online integrated genome sequence annotation pipeline. BUSCO analysis was conducted to test for annotation completeness. Repetitive elements were detected by RepeatModeler (Smit and Hubley, 2013) and RepeatMasker (Smit et al., 2019). AgriGO tested for enrichment of Gene Ontology categories of gene functions in the rearranged segments (Tian et al., 2017). RNA-seq of *G. kirkii* is

available from GenBank under SRX5894875. Gene and repetitive annotations are available from CottonGen under <https://www.cottongen.org/analysis/213>.

Analysis of Paleo-Genome Duplications

Protein sequences of *G. kirkii* and the D_T genome of *G. hirsutum* were clustered using OrthoFinder v2.1 (Emms and Kelly, 2015) with the Diamond alignment tool. Single copy orthologs from OrthoFinder were used as input to MCScanX_h (.homology file), with default settings (Wang et al., 2012). The collinearity plots between chromosomes Chr02 (Chr15, if the tetraploid chromosomes were numbered sequentially), Chr04 (Chr17) and Chr06 (Chr19) in *G. hirsutum* and chromosomes KI_2_4 and KI_06 in *G. kirkii* were created using the circle_plotter downstream tool of the MCScanX package. From the OrthoFinder output, all of the intraspecific paralogs were extracted for *G. kirkii*. Within each group of putatively orthologous genes, Ks values for every possible pairwise combination of paralogs were calculated using the codeml package of PAML, using custom python scripts.

ChIP-seq

Leaves and leaf buds were also collected from *G. kirkii* (specimen voucher ISC 418555, Ada Hayden Herbarium, Iowa State University). Rabbit polyclonal CenH3 antibody was made to the CenH3 amino acids 9–20 and conjugated to KLH (Covance, Inc.), a conserved peptide in *Gossypium* species of CenH3. Immunostain on *G. raimondii* root tips ensured centromere specificity of the CenH3 antibody. Chromatin immunoprecipitation was performed using the Epigentek EpiQuik Plant ChIP Kit (P-2014) with modifications. DIECA (2%) and PVP-40 (4%) were added to the fixative and to final solutions of CP3C, CP3D, and CP3E. DNA samples were sonicated at 60% amplitude for three total minutes of sonication/rest (15 s/15 s). Divided samples were incubated with either rabbit pre-immune sera, anti-CenH3, or polyclonal H3K9ac (ABCam, ab10812, LOT GR171780). Four replications of each reaction was pooled for whole genome amplification using the SeqPlex Enhanced DNA Amplification Kit (SeqXE, Sigma) then sequenced (Illumina PE150 bp) at the Beijing Genomics Institute (BGI). ChIP-seq reads were mapped to the genome using BWA (Li and Durbin, 2009). ChIP-seq data are available from NCBI under SRX5894872–SRX5894874.

FISH

Preparation of chromosomes and staining were performed as previously described for maize (Masonbrink and Birchler, 2010). FISH was performed as specifically described for cotton (Wang et al., 2006).

RESULTS

Sequencing and *De Novo* Assembly of the *G. kirkii* Genome

Two different genome technologies were used to assemble the *G. kirkii* genome sequence (**Figure 1**). First, approximately

68× coverage of raw SMRT data (40 Gb) was generated using the PacBio Sequel System (Table S1). The contig-level assembly was 544 Mb composed of 389 contigs with a contig N50 of 9.92 Mb and a maximum contig size of 31.1 Mb (Table 1). After scaffolding with HiC (Burton et al., 2013), the 12 pseudomolecules assembly was 92.5% of the expected genome size of 588 Mb (Wendel et al., 2002) with only 277

gaps (Table 1, Table S2 and Figure S1). Chromosomes were manually adjusted (Figures S2 and S3) and named according to the convention used in *Gossypium hirsutum* (Zhang et al., 2015). These pseudomolecules represented the 12 chromosomes of the *G. kirkii* genome (Figure 2, Table S3). Chromosome KI_2_4 contained the largest number of sequence contigs (65 contigs, 41.5 Mb) and Chromosome KI_08 contained the fewest (seven contigs, 39.6 Mb), even though these two chromosomes contained approximately the same total sequence length. Chromosome KI_06 was the largest chromosome (see below).

An optical map (Bionano Genomics, Inc.) was used to validate the assembly of individual contigs and the HiC connections between contigs (Table S4). Optical map data typically serves as an independent validation of the assembled sequence because the image data of Bionano labeled DNA molecules is assembled independently and aligned to DNA sequences using restriction patterns matching the labels in the Bionano contigs (Udall and Dawe, 2017). While the percentage of alignments between optical maps and contigs was relatively low, we note that over half of the genome sequence was validated by optical map alignment. The Bionano alignments also spanned 62% of the 71 eligible sequence gaps (i.e. gaps flanked by contigs >100 kb on each side, since

TABLE 1 | Assembly metrics of the *Gossypioides kirkii* genome.

Genome Statistics	Stats for Contigs	Stats for Pseudomolecules (Hi-C)	Stats for Chromosome Assembly (BioNano)
Assembly size (in Mb)	544	538	354
Number of sequences (number of gaps)	389 (0)	12 (277)	1,079 (3,721)
Longest Scaffold Length (Mb)	31.1	60.3	16.7
N50 in Mb (mean number of contigs per chromosome)	9.92 (17)	42.97 (17)	0.296

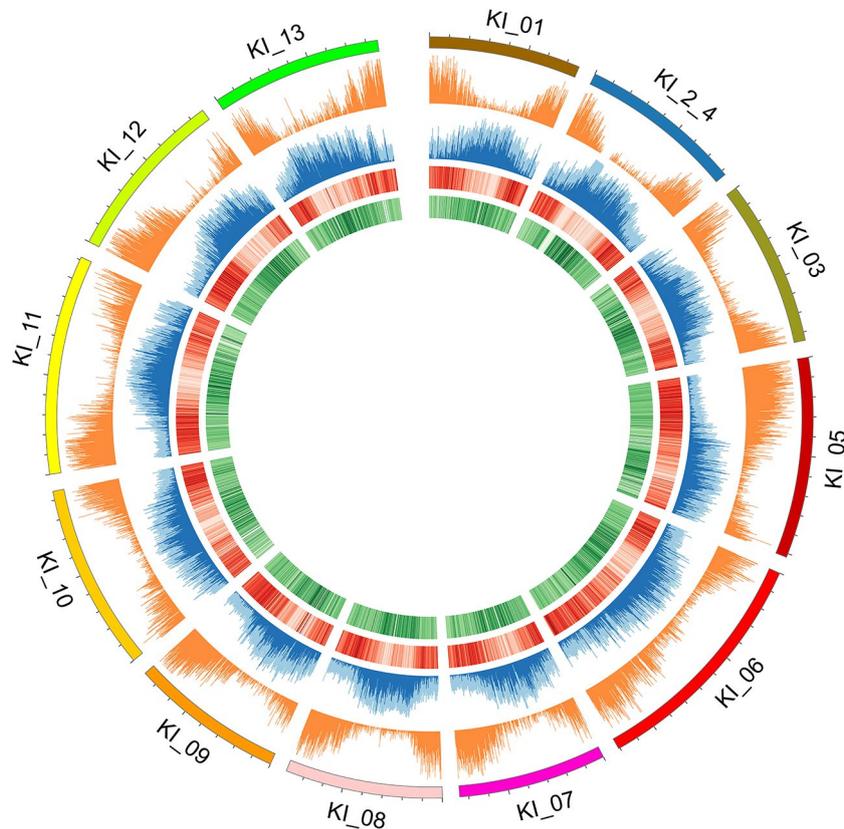


FIGURE 2 | Individual chromosomes (KI_ labels) of the *G. kirkii* genome are illustrated by 5 tracks in a Circos plot. Darker shades of colors represent a higher value or frequency of genomic features with the 100 kb window. From outside to inside: chromosome graph and scale; plot of gene density; TE content (light blue is total TE content, dark blue is Gypsy content), ChIP-seq of H3K9ac (Darker red lines of H3K9ac indicate a higher frequency of H3K9 acetylation); and ChIP-seq of CENH3 (Darker green lines of the CENH3 track indicate a higher frequency of CENH3 binding). Centromeres are inferred in the regions where H3K9ac is low (light red) and CENH3 is high (dark green).

Bionano contigs do not generally match smaller contigs due to limitations in nick-pattern matching) (**Dataset S1**).

A common measure of genome quality is the percentage of expected genes recovered in an annotated assembly. Here, the percentage of genes identified in the *G. kirkii* genome sequence provided confidence that nearly the entire genome was represented, with 95% (1,364/1,440) of conserved genes from BUSCO identified (Simão et al., 2015). The remaining genes were either fragmented ($n = 18$, 1%) or missing ($n = 58$, 4%). That a few of the BUSCO genes were missing is not surprising due to the previously reported genome downsizing and gene loss in this species (Grover et al., 2017), and therefore may reflect a combination of genome completeness as well as historical evolution. Of the 36,669 gene annotations, 64% had RNA-seq reads (> 20 reads) mapping to them, suggesting that we assembled much of the leaf transcriptome.

Collinearity With Other *Gossypieae* Genomes

The integrity of the *G. kirkii* genome assembly was also assessed by comparing it to genome sequences recently published for *Gossypium* (Paterson et al., 2012; Du et al., 2018; Wang et al., 2019) (**Figures S4–S7**). Occasionally, we used these comparisons to correct scaffolding errors in the *G. kirkii* genome if the *G. kirkii* contigs and optical map contigs supported such corrections. These manual rearrangements (**Dataset S2**) utilized evidence from both contig ends of each initial non-collinear placement of *G. kirkii* sequence.

We further assessed the completeness of the assembly and the orientation of terminal scaffolds by searching for telomere sequences in the *G. kirkii* pseudomolecules. We identified 20 loci with characteristic sequence of telomere repeats at the ends of our pseudomolecules (**Table S5**); eight pseudomolecules had telomere repeats on both chromosome arms, four had telomere repeats on a single arm, and two pseudomolecules had telomere repeats that were confidently embedded within a single scaffold. The longest telomeric repeat (> 24 kb) was identified on KI_04. Since this length was longer than most of the length of our trimmed reads used for assembly ($N50 = 16,192$), it is likely that many reads containing a majority of telomeric sequence collapsed during sequence assembly. Indeed, these regions had a higher read coverage compared to the adjacent chromosome sequence (data not shown). Different telomere sequences were identified in different combinations on each of the chromosome ends, suggesting the existence of multiple telomerases or at a minimum multiple guide RNAs in *Gossypioides*.

Because typical centromeres do not have conserved sequences (Ma et al., 2007; Lysak, 2014; Birchler and Han, 2018), we leveraged additional data to identify centromeric regions. That is, we evaluated the density of both ChIP-seq reads and gene density to infer putative centromeric regions. Euchromatic and histone modifications of H3K9ac and CENH3, respectively were used to estimate centromeric regions (Masonbrink et al., 2014). Typical distributions of epigenetic marks were identified (e.g. increasing frequency of CENH3 marks near the centromeric regions, **Figure 2**). In some cases, chromosomes had a single contig

assembled across the centromeric region (e.g. chromosomes KI_06, KI_10, KI_11) suggesting proper assembly and density of CENH3 marks in centromeric regions. The centromeric regions of other chromosomes contained multiple contigs. While their assembly depended on both correct sequence assembly and correct scaffolding, their density of CENH3 marks was similar to those regions composed of a single contig.

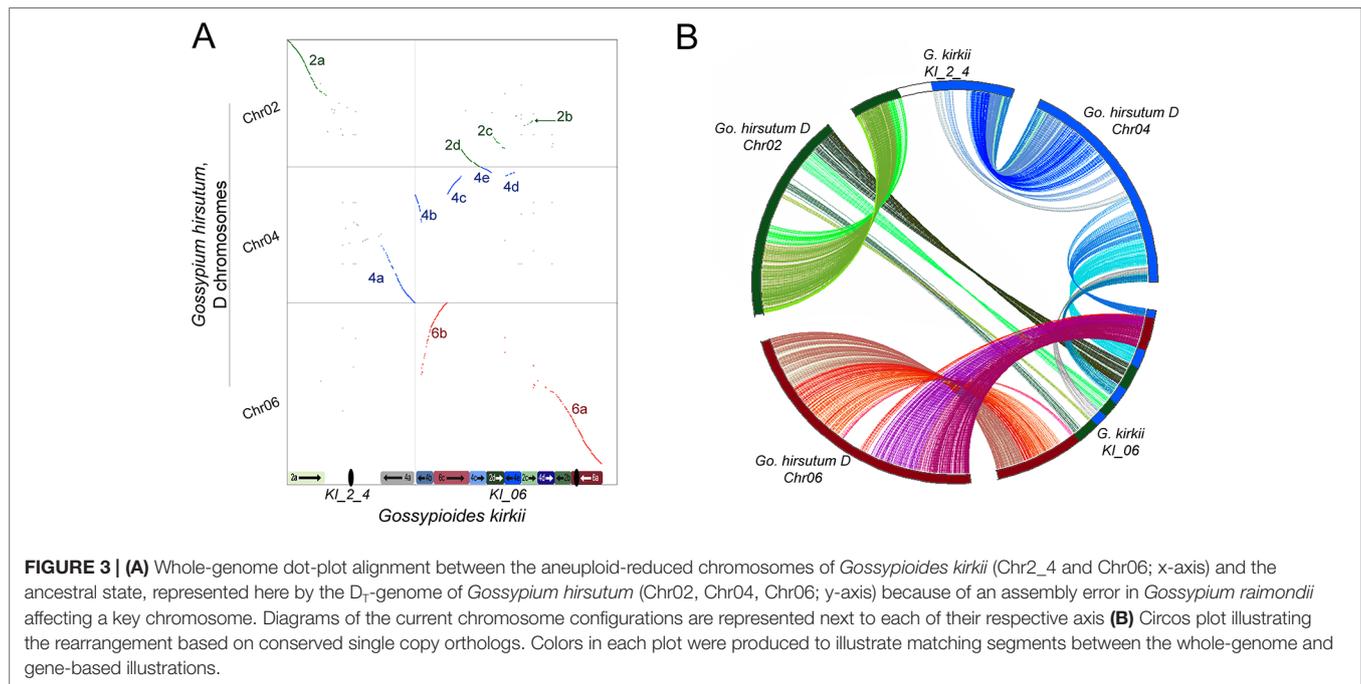
Annotation of Genes and Repetitive Elements

Gene annotation recognized 36,669 genes, somewhat higher than previously reported (Grover et al., 2017); these differences are likely due to both genome quality and annotation method. All *G. kirkii* genes were aligned to their closest intragenomic paralog to calculate synonymous substitutions (Ks); the plot of these pairwise Ks values exhibits a peak congruent with previous findings (Conover et al., 2019) of an ancient polyploidization event shared with *K. drynarioides* and all members of *Gossypium* (**Figure S8**). Because genes comprise useful genomic anchors, gene annotations were used to inform analyses of the chromosome rearrangements in *G. kirkii* (below).

Repetitive elements were detected by RepeatModeler (Smit and Hubley, 2018) and RepeatMasker (Smit et al., 2019). As a whole, the genome contained $\sim 30\%$ interspersed repeats and 1.7% simple repeats. The interspersed repetitive elements corresponded to transposable elements, namely *Gypsy* and *Copia* retrotransposons (**Table S3**). We detected the TEs on each chromosome to assess the class distribution of TE elements throughout the genome (Baillly-Bechet et al., 2014). While TEs can be associated with chromosome rearrangements, we found no bias in terms of TE number, total length, or class between chromosomes. In general, the number and total length of *Gypsy* elements greatly outweighs *Copia* elements, as is common for many plant genomes (**Table S3**).

Comparative Genomics Between *G. Kirkii* and Related Genomes

The base chromosome number (x) of *G. kirkii* and *K. drynarioides* is $x = 12$, but the remainder of the cotton tribe (*Gossypieae*) in which this lineage is nested has a base chromosome number of $x = 13$ (**Figure 1**). To explore which chromosomes may be involved in this derived state, we identified chromosome rearrangements that occurred after divergence between *G. kirkii* and *Gossypium* (represented by the ancestral “G” chromosomes). In this analysis, the genome of *Gossypium* was assumed to represent the ancestral genome to the *Gossypium–Gossypioides–Kokia* clade, and the *G. kirkii* genome was considered derived due to the presence of necessary changes during chromosome reduction, although we cannot discount the possibility of some structural changes in *Gossypium*. Whole genome comparisons suggested that an entire arm of chromosome G2 and an entire arm of chromosome G4 (intact within modern-day *G. raimondii* and *G. arboreum*) fused to form a single chromosome, while the other chromosome arms were fragmented and inserted into KI_06 (**Figure 3A**). A comparison of annotated genes to *G. hirsutum* in these regions



also supports our inferred genome alignments (**Figure 3B**). The insertion of these G2 and G4 fragments into KI_06 explain the absence of a single chromosome that is twice the size of other metacentric chromosomes, as might be expected if a simple chromosome fusion had occurred (Hutchinson, 1943). We confirmed the absence of an unusually long chromosome in *G. kirkii* by chromosome staining (**Figure S9**). The inserted portion on KI_06 consists of alternating segments of ancient chromosomes G2 and G4 with six segments accounting for approximately 30 MB of the chromosome. Details of the rearranged segments are found in **Table S6**. These segments each contained between 117 and 458 genes. A GO enrichment test of each segment found no enriched GO categories.

These findings may be summarized as three salient facts regarding the genomic history of *G. kirkii*. First, one chromosome arm each of G2 and G4 were inserted into what became part of *Gossypioideis kirkii* chromosome KI_06. Second, before or after the insertion, segments of these two chromosome arms were interleaved through unknown evolutionary processes. It is worth noting that all of the G2/G4 'junctions' in KI_06 have strong support of PacBio and Bionano coverage. Third, the remaining, entire chromosome arms of G2 and G4 fused to create KI_2_4. We further support these inferences by mapping a *K. drynarioides* HiC library to the *G. kirkii* assembly. The resulting HiC contact heatmap (**Figure S10**) also shows a linear contact pattern along KI_06 and KI_2_4 suggesting that the chromosome rearrangements we describe are shared between these sister genera *Kokia* and *Gossypioideis*.

DISCUSSION

Among the many opportunities afforded by genome sequencing is the possibility of gaining insight into long-standing cytogenetic

phenomena that remain unexplained at the sequence level. A promising example is dysploid evolution, which is a well-known and common pattern of cytogenetic variation in both plants and animals (Grant, 1971; Stebbins, 1971; White, 1973). From a mechanistic standpoint, it has long been thought that dysploidy arises primarily from chromosome translocations. This view was promulgated in George Ledyard Stebbins' 1971 classic *Chromosomal Evolution in Higher Plants*, in which he stated that "aneuploid alterations of the basic chromosome number are usually the outcome of successive translocations" (5, pg. 86). Similarly, in Verne Grant's widely used 1971 textbook "Plant Speciation", he stated "the mechanism of aneuploid reduction at the diploid level involves unequal reciprocal translocations" (4, pg. 359). More recently, telomeric (end-to-end) fusion and Robertsonian translocation have been recognized as processes leading to aneuploid reduction (Schubert, 1992; Lysák and Schubert, 2013; Chiatante et al., 2017; Jarvis et al., 2017), as has the insertion of one chromosome into another (Luo et al., 2009; Murat et al., 2010; Vogel et al., 2010; Wang and Bennetzen, 2012; Fonsêca et al., 2016). Remarkably, while we were completing the present work, Birchler and Han (Birchler and Han, 2018) published a thought-provoking explication of how the Breakage-Fusion-Bridge cycle, as illuminated by McClintock 80 years ago for understanding various chromosome anomalies in maize (McClintock, 1939; McClintock, 1941), likely has causal connections to common mechanisms of karyotypic evolution in plants, and by extension possibly all eukaryotes.

Here we provide sequence-based evidence for chromosome number reduction where related members of the cotton tribe establish the polarity of the descending dysploidy (from $x = 13$ to $x = 12$). The foundation for our conclusions is the high-quality assembly of the *G. kirkii* genome sequence presented here. The accuracy of this assembly was determined by multiple congruent datasets (PacBio, HiC, and Bionano) and

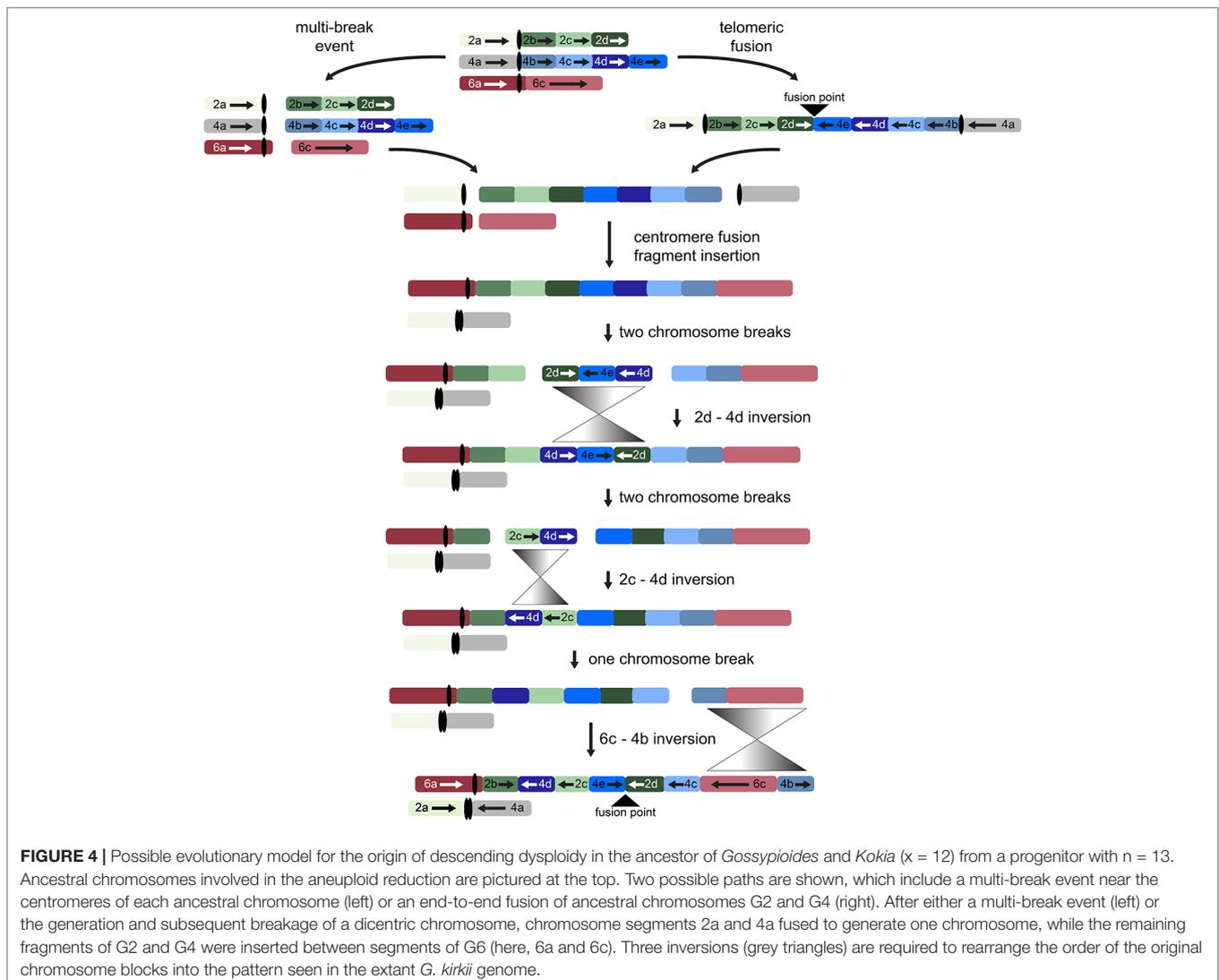
by comparative analyses that demonstrate consistency with previously published cotton genomes (Paterson et al., 2012; Zhang et al., 2015; Du et al., 2018; Wang et al., 2019). Analyses of colinearity revealed a complex pattern of inter-digitating chromosome segments. The identified rearrangements also are congruent with previous cytogenetic observations of *Gossypioides brevilanatum*, in the fact that *G. kirkii* does not display an ‘extra-large’ chromosome (Hutchinson, 1943) as might be expected from a simpler scenario of a 2 to 1 chromosome fusion event.

Explanations for the inferences depicted in **Figure 3** for the derivation of the reduced chromosome number in *G. kirkii* (“KI” chromosomes) relative to the G chromosomes of ancestral *Gossypieae* need to account for the following observations: (1) identification of end-to-end G2 and G4 (2d and 4e) segments (including internal telomeric sequences) in KI_06 that implicate a historical end-to-end fusion of ancestral chromosome arms G2 and G4 (2) chromosome KI_2_4 contains entire chromosome arms of G2 and G4, suggestive

of chromosome fusion at, or close to, the centromeres for the intact G2 and G4 arms; and (3) because the terminal inversion on KI_06 included 2.0Mb of the G4 chromosome (in addition to 8.4 Mb of the original G6 chromosome arm), it must have occurred after the insertion event above.

We recognize that by assuming the *Gossypium* genome represented the ancestral *Gossypieae* genome, unique *Gossypium* changes were confounded with the differences between *G. kirkii* and *Gossypieae*. We are comfortable with this assumption based on previous cytological work that prompted the previous generation of botanists to coin the phrase ‘cryptic structural differentiation’ when working with the cotton tribe (Fryxell, 1979). They understood the chromosomes were different based on pairing data, but observable structural differentiation was not sufficient to differentiate members of the cotton tribe, other than the descending dysploidy of *Gossypioides* and *Kokia*.

While other interpretations could be made when additional genomes from tribe *Gossypieae* (e.g., *Thespesia*, *Hampea*, or *Lebronnecia*) are sequenced, we use the above three key



observations (and one modest assumption) to create a hypothesis for the order of events following initial dicentric chromosome formation (**Figure 4**). The end-to-end fusion of G2 and G4 strongly supports evolutionary models that begin with a dicentric chromosome, although we note that a multi-break-fusion event could bypass the need for a dicentric chromosome (as noted in **Figure 4**). Myriad alternative models are also possible where end-to-end fusion are coincidental instead of contributory; however, they are not considered further because of the key evidence of the telomeric sequence and directionality of the end-to-end fusion fragments.

Chromosome comparisons between *Gossypioides* and *Gossypium* suggest that the origin of *G. kirkii* KI_06 involved both fusion and a series of inversions to generate the observed interleaved pattern (**Figure 4**). While this fusion could have been the result of breaks occurring on each of the involved chromosomes, the presence of internal telomeres supports an end-to-end fusion, generating a dicentric chromosome. Perhaps, the nascent dicentric chromosome somehow contributed to subsequent breaks near each centromere (G2- and G4-derived). If an additional break was concurrent in G6, then translocations followed by subsequent paracentric inversions could create the extant chromosomes of *Gossypioides*. As depicted in **Figure 4**, two-breaks of a dicentric chromosome created an acentric fragment containing most of the arms of G2 and G4, which inserted into G6, and centromeric fusion between the G2- and G4-chromosome arms containing centromeric sequence. Three inversions are then required to transform the initial fusion of *G. kirkii* Chr06 into the extant chromosome morphology. The two unshared inversions would only involve portions of the inserted segment. Notably, the two chromosome inversions of *G. kirkii* were approximately 9.6 Mb and 12.7 Mb, respectively (**Table S6**), which is similar to the average inversion size for plants and animals (i.e., 8.4 Mb, (Wellenreuther and Bernatchez, 2018)). While inversions often are associated with TEs (Kidwell and Lisch, 2000), we do not find an increased density of TEs in KI_06 (**Figure 2**) to support this for *G. kirkii*. Although the responsible inversion mechanism is not known, it is possible that recombination between a hemizygous insertion KI_06 and a normal KI_06 could have played a role.

The foregoing hypothesis explains a novel “3” to “2” route for chromosome number reduction, as opposed to the more conventional “2” to “1”. It certainly invokes a series of seemingly unlikely events, including formation of multiple inversions requiring two simultaneous double-strand breaks and repair (Kirkpatrick, 2010), either through a known mechanism such as breakage-fusion-bridge or unknown accidents of aberrant recombination (as described here). Because the likelihood of each rare event multiplies when each is considered as independent of the others, perhaps it is more parsimonious to postulate that chromosome number reduction occurred within a single generation, in a series of germ-line cell divisions with subsequent ‘healing’ in the sporophyte. In contrast, it remains possible that this entire process unfolded in a stepwise fashion during long evolutionary timescales.

Unfortunately, we lack surviving intermediates that might testify to this temporal possibility, and we are unaware of other methods that might be used to distinguish between the “fast” and “slow” scenarios. Several other studies have detected aneuploid reduction between related species in the Brassicaceae (Lysak et al., 2006), or in the genomes of grasses (Zhang et al., 2008; Murat et al., 2010; Wang et al., 2015; Luo et al., 2017), based on patterns of FISH or using sequence comparisons, and others have noted instances of “chromosome shattering” with possible mechanisms (Zhang et al., 2008; Tan et al., 2015; Mandáková et al., 2019). As more plant and animal genomes are sequenced and assembled by robust methods, the spectrum of causative mechanisms and their frequency in explaining patterns of karyotypic evolution are likely to become much clearer.

DATA AVAILABILITY STATEMENT

The *G. kirkii* genome sequence is available on GenBank (Accession numbers: CP032244-CP032255).

AUTHOR CONTRIBUTIONS

JU and JW developed the idea. JU, JW, JC, and CG designed the project. EL, DY, LG, MA, RM, and DP generated the data. JU, TR, JC, DY, CG, and MA analysed the data. JU, JC, CG, and JW wrote the manuscript. All authors read, edited, and approved the final manuscript.

FUNDING

Primary funding was provided by the National Science Foundation Plant Genome Research Program (#1339412). Additional support was provided from Cotton Incorporated and USDA-ARS (58-6402-1-644 and 58-6066-6-59).

ACKNOWLEDGMENTS

We thank the BYU Fulton SuperComputer lab for use of their computation resources. We thank ResearchIT (<https://researchit.las.iastate.edu/>) for computational support at Iowa State University. We thank Rise Services for office accommodations in Orem, UT.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01541/full#supplementary-material>

Supplementary material for this article can be found in (Udall_Supplemental.docx).

REFERENCES

- Bailey-Bechet, M., Haudry, A., and Lerat, E. (2014). "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mob. DNA* 5, 13. doi: 10.1186/1759-8753-5-13
- Birchler, J. A., and Han, F. (2018). Barbara McClintock's unsolved chromosomal mysteries: parallels to common rearrangements and karyotype evolution. *Plant Cell* 30, 771–779. doi: 10.1105/tpc.17.00989
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727
- Chiatante, G., Giannuzzi, G., Calabrese, F. M., Eichler, E. E., and Ventura, M. (2017). Centromere destiny in dicentric chromosomes: new insights from the evolution of human chromosome 2 ancestral centromeric region. *Mol. Biol. Evol.* 34, 1669–1681. doi: 10.1093/molbev/msx108
- Conover, J. L., Grover, C. E., Wendel, J. F., Karimi, N., Stenz, N., Ané, C., et al. (2019). A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *J. Integr. Plant Biol.* 61 (1), 12–31. doi: 10.1111/jipb.12746
- Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* 50 (6), 796–802. doi: 10.1038/s41588-018-0116-x
- Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3 (1), 99–101. doi: 10.1016/j.cels.2015.07.012
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2
- Fajkus, J., Sýkorová, E., and Leitch, A. R. (2005). Telomeres in evolution and evolution of telomeres. *Chromosom. Res.* 13, 469–479. doi: 10.1007/s10577-005-0997-2
- Fonseca, A., Ferraz, M. E., and Pedrosa-Harand, A. (2016). Speeding up chromosome evolution in Phaseolus: multiple rearrangements associated with a one-step descending dysploidy. *Chromosoma* 125, 413–421. doi: 10.1007/s00412-015-0548-3
- Fryxell, P. A. (1979). *The natural history of the cotton tribe (Malvaceae, tribe Gossypieae)*. 1st ed. College Station: Texas A&M University Press.
- Giannuzzi, G., Pазienza, M., Huddleston, J., Antonacci, F., Malig, M., Vives, L., et al. (2013). Hominoid fission of chromosome 14/15 and the role of segmental duplications. *Genome Res.* 23, 1763–1773. doi: 10.1101/gr.156240.113
- Grant, V. (1971). *Plant speciation*. Columbia University Press.
- Grant, V. (1981). *Plant speciation*. New York: Columbia University Press.
- Grover, C. E., Arick, M. A., Conover, J. L., Thrash, A., Hu, G., Sanders, W. S., et al. (2017). Comparative genomics of an unusual biogeographic disjunction in the cotton tribe (Gossypieae) yields insights into genome downsizing. *Genome Biol. Evol.* 9, 3328–3344. doi: 10.1093/gbe/evx248
- Guerra, M. (2008). Chromosome numbers in plant cytotoxicology: concepts and implications. *Cytogenet. Genome Res.* 120, 339–350. doi: 10.1159/000121083
- Heslop-Harrison, J. S. P., and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. *Plant J.* 66, 18–33. doi: 10.1111/j.1365-3113X.2011.04544.x
- Hoang, P. T. N., and Schubert, I. (2017). Reconstruction of chromosome rearrangements between the two most ancestral duckweed species *Spirodela polyrhiza* and *S. intermedia*. *Chromosoma* 126, 729–739. doi: 10.1007/s00412-017-0636-7
- Hutchinson, J. B. (1943). A note on *Gossypium brevilanatum* Hochr. *Trop. Agric. Agric.* 20, 4.
- Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi: 10.1038/nature21370
- Jones, K. (1998). Robertsonian fusion and centric fission in karyotype evolution of higher plants. *Bot. Rev.* 64, 273–289. doi: 10.1007/BF02856567
- Kidwell, M. G., and Lisch, D. R. (2000). Transposable elements and host genome evolution. *Trends Ecol. Evol.* 15 (3), 95–99. doi: 10.1016/S0169-5347(99)01817-0
- Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biol.* 8 (9), e1000501. doi: 10.1371/journal.pbio.1000501 pii.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive κ -mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Lee, T., Peace, C., Jung, S., Zheng, P., Main, D., and Cho, I. (2011). "GenSAS - An online integrated genome sequence annotation pipeline," in *Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011*. doi: 10.1109/BMEI.2011.6098712
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi: 10.1093/bioinformatics/bty191
- Luo, M. C., Deal, K. R., Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., et al. (2009). Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc. Natl. Acad. Sci. U. S. A.* 106, 15780–15785. doi: 10.1073/pnas.0908195106
- Luo, M.-C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551, 498. doi: 10.1038/nature24486
- Lysák, M. A., and Schubert, I. (2013). "Mechanisms of Chromosome Rearrangements," in *Plant Genome Diversity Volume 2* (Vienna: Springer Vienna), 137–147. doi: 10.1007/978-3-7091-1160-4_9
- Lysak, M., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5224–5229. doi: 10.1073/pnas.0510791103
- Lysak, M. A. (2014). Live and let die: centromere loss during evolution of plant chromosomes. *New Phytol.* 203, 1082–1089. doi: 10.1111/nph.12885
- Ma, J., Wing, R., Bennetzen, J. L., and Jackson, S. A. (2007). Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet.* 23, 134–139. doi: 10.1016/j.tig.2007.01.004
- Mandáková, T., Pouch, M., Brock, J. R., Al-Shehbaz, I. A., and Lysak, M. A. (2019). Origin and evolution of Diploid and Allopolyploid *Camelina* genomes was accompanied by chromosome shattering. *Plant Cell* 31 (11), 2596–2612. doi: 10.1105/tpc.19.00366
- Masonbrink, R. E., and Birchler, J. A. (2010). Sporophytic nondisjunction of the maize B chromosome at high copy numbers. *J. Genet. Genomics* 37, 79–84. doi: 10.1016/S1673-8527(09)60027-8
- Masonbrink, R. E., Gallagher, J. P., Jareczek, J. J., Renny-Byfield, S., Grover, C. E., Gong, L., et al. (2014). CenH3 evolution in diploids and polyploids of three angiosperm genera. *BMC Plant Biol.* 14, 383. doi: 10.1186/s12870-014-0383-3
- McClintock, B. (1939). The behavior in successive nuclear divisions of a chromosome broken at Meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 25 (8), 405–416. doi: 10.1073/pnas.25.8.405
- McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26 (2), 234–282.
- McKnight, T. D., Fitzgerald, M. S., and Shippen, D. E. (1997). Plant telomeres and telomerases. A review. *Biochem. (Mosc.)* 62, 1224–1231. <http://www.ncbi.nlm.nih.gov/pubmed/9467846>.
- Murat, F., Xu, J. H., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., et al. (2010). Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.* 20, 1545–1557. doi: 10.1101/gr.109744.110
- Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798
- Phuphathanaphong, L. (2006). *Theparatia* (Malvaceae), a new genus from Thailand. *Thai. For. Bull.* 34, 195–200.
- Poorten, T. (2018). <https://github.com/tpoorten/dotPlotly.dotPlotly>. Available at: .
- Roberto, C. (2005). Low chromosome number angiosperms. *Caryologia* 58 (4), 403–409. doi: 10.1080/00087114.2005.10589480
- Schubert, I. (1992). Telomeric polymorphism in *Vicia faba*. *Biol. Zentralblatt* 111, 164–168.
- Seelanan, T., Schnabel, A., and Wendel, J. F. (1997). Congruence and consensus in the cotton tribe (Malvaceae). *Syst. Bot.* 22, 259. doi: 10.2307/2419457

- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smit, A., and Hubley, R. (2013). doi: 10.1007/s00572-016-0720-5 RepeatModeler. RepeatModeler.
- Smit, A., and Hubley, R. (2018). RepeatModeler Open-1.0. GitHub.
- Smit, A. F. A., Hubley, R., and Green, P. (2019). <http://www.repeatmasker.org>. RepeatMasker Open-4.0. Available at: .
- Stebbins, G. L. (1971). *Chromosomal evolution in higher plants*. London: Edward Arnold LTD. 87–89.
- Tan, E. H., Henry, I. M., Ravi, M., Bradnam, K. R., Mandakova, T., Marimuthu, M. P., et al. (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* 4, e06516. doi: 10.7554/eLife.06516
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45 (W1), W122–W129. doi: 10.1093/nar/gkx382
- Udall, J. A., and Dawe, R. K. (2017). Is it ordered correctly? Validating genome assemblies by optical mapping. *Plant Cell* 30, 7–14. doi: 10.1105/tpc.17.00514
- Uhl, C. H. (1978). Chromosomes of Mexican Sedum II. Section Pachysedum. *Rhodora* 80, 491–512. <http://www.jstor.org/stable/2331126>. Available at: .
- Vanzela, A. L. L., Guerra, M., and Luceno, M. (1996). *Rhynchospora Tenuis* Link (Cyperaceae), a species with the lowest number of holocentric chromosomes. *Cytobios* 88, 219–228. Available at: isi:A1996XT51300004.
- Vogel, J. P., Garvin, D. F., Mockler, T. C., Schmutz, J., Rokhsar, D., Bevan, M. W., et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463 (7282), 763–768. doi: 10.1038/nature08747
- Wang, H., and Bennetzen, J. L. (2012). Centromere retention and loss during the descent of maize from a tetraploid ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 109 (51), 21004–21009. doi: 10.1073/pnas.1218668109
- Wang, K., Song, X., Han, Z., Guo, W., Yu, J. Z., Sun, J., et al. (2006). Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybridization mapping. *Theor. Appl. Genet.* 113, 73–80. doi: 10.1007/s00122-006-0273-7
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40 (7), e49. doi: 10.1093/nar/gkr1293
- Wang, X., Jin, D., Wang, Z., Guo, H., Zhang, L., Wang, L., et al. (2015). Telomere-centric genome repatterning determines recurring chromosome number reductions during the evolution of eukaryotes. *New Phytol.* 205 (1), 378–389. doi: 10.1111/nph.12985
- Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., et al. (2019). Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* 51, 224–229. doi: 10.1038/s41588-018-0282-x
- Watson, J. M., and Riha, K. (2010). Comparative biology of telomeres: where plants stand. *FEBS Lett.* 584, 3752–3759. doi: 10.1016/j.febslet.2010.06.017
- Weiss-Schneeweiss, H., and Schneeweiss, G. M. (2013). “Karyotype diversity and evolutionary trends in angiosperms,” in *Plant Genome Diversity*, Springer-Verlag Wien. 209–230. doi: 10.1007/978-3-7091-1160-4_13
- Wellenreuther, M., and Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33 (6), 427–440. doi: 10.1016/j.tree.2018.04.002
- Wendel, J. F., and Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 139–186. doi: 10.1016/S0065-2113(02)78004-8
- Wendel, J. F., Cronn, R. C., Spencer Johnston, J., and James Price, H. (2002). Feast and famine in plant genomes. *Genetica* 115, 37–47. doi: 10.1023/a:1016020030189
- White, M. (1973). *Animal cytology and evolution*. London: Cambridge University Press.
- Zhang, P., Li, W., Friebe, B., and Gill, B. S. (2008). The origin of a “Zebra” chromosome in wheat suggests nonhomologous recombination as a novel mechanism for new chromosome evolution and step changes in chromosome number. *Genetics* 179 (3), 1169–1177. doi: 10.1534/genetics.108.089599
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Udall, Long, Ramaraj, Conover, Yuan, Grover, Gong, Arick, Masonbrink, Peterson and Wendel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.