Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa

Levi Yant,1,3 Jesse D. Hollister,1,3,4 Kevin M. Wright,1 Brian J. Arnold,1 James D. Higgins,2 F. Chris H. Franklin,2 and Kirsten Bomblies1,*

1Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
2School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Summary

Whole genome duplication (WGD) is a major factor in the evolution of multicellular eukaryotes, yet by doubling the number of homologs, WGD severely challenges reliable chromosome segregation [1–3, 8–11], a process conserved across kingdoms [4]. Despite this, numerous genome-duplicated (polyploid) species persist in nature, indicating early problems can be overcome [1, 2]. Little is known about which genes are involved—only one has been molecularly characterized [5]. To gain new insights into the molecular basis of adaptation to polyploidy, we investigated genome-wide patterns of differentiation between natural diploids and tetraploids of Arabidopsis arenosa, an outcrossing relative of A. thaliana [6, 7]. We first show that diploids are not preadapted to polyploid meiosis. We then use a genome scanning approach to show that although polymorphism is extensively shared across ploidy levels, there is strong ploidy-specific differentiation in 39 regions spanning 44 genes. These are discrete, mostly single-gene peaks of sharply elevated differentiation. Among these peaks are eight meiosis genes whose encoded proteins coordinate a specific subset of early meiotic functions, suggesting these genes comprise a polygenic solution to WGD-associated chromosome segregation challenges. Our findings indicate that even conserved meiotic processes can be capable of nimble evolutionary shifts when required.

Results and Discussion

Meiotic Chromosome Behavior in Tetraploid A. arenosa

At least initially, WGD is commonly associated with deleterious chromosome missegregation arising from multivalent associations among available homologs (e.g., [1–3, 8–11]). This is especially challenging for autopolyploids, which arise from within-species duplication and have multiple approximately equally homologous chromosomes. We asked if for autotetraploid A. arenosa (1) the tetraploid material we are working with has diploid-like chromosome behavior, and (2) the diploid genome we are comparing to is not preadapted for polyploid meiosis, as has been seen in some species [e.g., 12]. Though bivalent formation among homologs appears to be random and inheritance tetrasomic in natural autotetraploid A. arenosa [13], metaphase I chromosomes associate predominantly as bivalents like in diploids ([14] and Figure 1A). There are structural differences, however: tetraploids have significantly more rod bivalents (and fewer ring bivalents) than diploids do, which indicates natural tetraploid A. arenosa averages fewer chiasmata per bivalent than diploid A. arenosa (Table S1 available online). A reduction in chiasma number to one per bivalent has previously been suggested as a possible mechanism for meiotic diploidization in autopolyploids because limiting crossovers to one per chromosome prevents multivalent associations (e.g., [10, 11]).

We induced WGD in two diploid A. arenosa genotypes using colchicine and examined chromosome behavior of confirmed neotetraploids in diakinesis and metaphase I, when multivalents are readily discernable. Unlike natural autotetraploids, synthetic neotetraploids exhibit extensive multivalent formation and ectopic connections between the chromosomes (Figure 1A; Table S1). The cytological abnormalities in the neotetraploid lines correlate with sharply reduced pollen viability: the two colchicine-doubled lines had only 3% and 5% pollen viability, in contrast to two natural autotetraploid lines that had 91% and 92% pollen viability. Thus, diploid A. arenosa provides an “unevolved” comparison for the natural tetraploid. Bivalent associations and reduced estimated chiasma frequency in natural autotetraploids, and the aberrant meiosis of neotetraploids, are consistent with data from many other autopolyploids (e.g., [8–11]), suggesting A. arenosa is a representative model for studying the molecular basis of adaptation to autopolyploid meiosis.

Evidence of Polygenic Selection in Autotetraploid A. arenosa

Because of its connection to fertility [1, 2], selection for meiotic stability immediately following WGD should be intense. Thus, we reasoned that alleles contributing to stable chromosome segregation in the autopolyploid should show reduced allelic diversity and excess differentiation between autotetraploids and diploids. High genetic diversity suggests A. arenosa autotetraploids did not undergo a severe recent bottleneck associated with WGD [7, 13] and/or have ongoing gene flow with diploids [15]. We have previously shown evidence that autotetraploid A. arenosa has undergone selective sweeps [13], but because diploids were not included, it remained unknown whether top outliers reflect adaptation to polyploidy or species-wide patterns shared with diploids.

We used a genome scanning approach to compare the genomes of diploid and tetraploid A. arenosa. We short-read sequenced whole genomes from 16 natural autotetraploids and 8 diploid individuals from six natural populations (Figure 1B; Table S2). We aligned reads to the closely related A. lyrata genome [16]. More than 46 million sites had coverage in all 24 individuals, of which about 5.6 million are polymorphic relative to the A. lyrata reference (Table 1). There is extensive shared variation between diploids and autotetraploids (>1.7 million sites) and remarkably few fixed differences (26 genome wide; Table 1).

We scanned for signatures suggestive of selective sweeps by analyzing consecutive windows of 100 polymorphic sites
(55,769 windows total) for 0.5% outliers in the distributions of three metrics: \( F_{ST} \) [17] (Figure 2A), the two-dimensional site frequency spectrum (2dSFS) [18] (Figure 2B), and the 0.5% most negative values of linear regression residuals from the relationship between diversity and differentiation. Outlier values for this “residuals” metric indicate excess differentiation for a given level of diversity (Figures 2C and 2D). All 0.5% outlier windows for all three tests are given in Table S3.

We generated an overlap list of windows found both among 0.5% outliers for 2dSFS and the residuals (Table S4). Though both the residuals and \( F_{ST} \) quantify genetic differentiation, we favored the former because it accounts for the positive relationship between differentiation and diversity (e.g., see Figure 2D).

The overlap list contains 39 distinct differentiated regions spanning 44 genes; most contain only a single gene (Figures 3A and 3B; Table S4), with rapid decay to background (e.g., Figure 3B). Using paired end information and de novo assemblies aligned to \( A. \) \textit{lyrata}, we verified gene order in these regions (see Supplemental Experimental Procedures). This analysis showed that neighboring loci in these regions are syntenous meiosis genes (the GO designation contains 219 genes, many of which have no known role in meiosis); thus, we generated a new list by searching \( A. \) \textit{thaliana} gene descriptions (TAIR10; http://www.arabidopsis.org) to identify 71 (out of 25,550) genes that are clearly annotated as having a role in meiosis. Of these 71 genes, 62 have good read alignment in \( A. \) \textit{arenosa} (Table S5). A random list of 44 genes would not be expected to contain any meiosis genes on average (the probability of one is 0.1).

We next asked whether meiosis genes as an overall class have consistently high differentiation, which could indicate they are under selection as a group even if most do not meet stringent 0.5% cutoffs. We compared differentiation of 100 SNP windows mapping within meiosis genes (Table S5) to windows in the rest of the genome using the residuals metric. Aside from the eight outliers, the distribution of values for windows falling in the remaining 54 well-aligned meiosis genes were not distinguishable from the genome-wide distribution (Figures 2D and 2E; \( t \) test \( p = 0.60 \)). Thus, meiosis-related genes show no gene set enrichment for differentiation apart from the eight outliers. This result suggests

between \( A. \) \textit{arenosa} and \( A. \) \textit{lyrata}, confirming that the rapid decay of differentiation reflects low linkage disequilibrium, not an alignment artifact. Low levels of linkage disequilibrium are likely to be the result of the large effective population size of \( A. \) \textit{arenosa} [6]. Six of the 44 genes overlap with our previous scan, even though the analyses used different methods and sample sets [11].

\textbf{Meiosis Genes Are Overrepresented among Genome Scan Outliers}

Eight meiosis-related genes were on our overlap list of 39 regions and 44 genes (Figures 2 and 3; Table S4). In GO category analysis, meiosis was the only significantly overrepresented functional category. However, there is some ambiguity in the GO category designation for meiosis genes (the GO designation contains 219 genes, many of which have no known role in meiosis); thus, we generated a new list by searching \( A. \) thaliana gene descriptions (TAIR10; http://www.arabidopsis.org) to identify 71 (out of 25,550) genes that are clearly annotated as having a role in meiosis. Of these 71 genes, 62 have good read alignment in \( A. \) \textit{arenosa} (Table S5). A random list of 44 genes would not be expected to contain any meiosis genes on average (the probability of one is 0.1).

We next asked whether meiosis genes as an overall class have consistently high differentiation, which could indicate they are under selection as a group even if most do not meet stringent 0.5% cutoffs. We compared differentiation of 100 SNP windows mapping within meiosis genes (Table S5) to windows in the rest of the genome using the residuals metric. Aside from the eight outliers, the distribution of values for windows falling in the remaining 54 well-aligned meiosis genes were not distinguishable from the genome-wide distribution (Figures 2D and 2E; \( t \) test \( p = 0.60 \)). Thus, meiosis-related genes show no gene set enrichment for differentiation apart from the eight outliers. This result suggests...
the eight meiosis-associated genes with strong differentiation in \textit{A. arenosa} represent a polygenic, naturally evolved solution to WGD-associated challenges. Among these eight, three were represented in a previous scan (\textit{ASY1}, \textit{SMC3}, and \textit{PDS5}) \cite{13}, whereas three others did not align in our previous study and were thus not included (\textit{ZYP1a}, \textit{ZYP1b}, and \textit{ASY3}).

### Functional Implications of Identified Meiosis Genes

In seven of the eight meiosis genes, sites with excess derived allele frequency encode predicted amino acid substitutions, and these are more common in tetraploids (Table S6). For example, \textit{ZYP1a} harbors 16 high-frequency-derived (relative to the \textit{A. lyrata} reference) substitutions in the tetraploid that encode predicted coding changes, but none in the diploid. \textit{ASY3}, however, shows highly divergent polymorphism in both ploids.

The eight meiosis genes in our outlier set are not a random sample: selection appears to have acted on multiple unlinked loci to shift the allelic landscape of coordinated events in early prophase I. All eight genes encode proteins crucial for the organization of chromosome structure, alignment, and synopsis of homologous chromosomes, and the controlled formation of crossovers \cite{19–21}. First, PRD3 participates in the early initiation of homologous recombination \cite{22, 23}. Coordination of subsequent events in recombination is dependent on the interplay between the recombination machinery and the chromosome axes. In yeast, this involves Red1, Hop1, and Rec8 \cite{24} whose functional homologs in \textit{A. thaliana} are \textit{ASY3}, \textit{ASY1}, and \textit{SYN1} \cite{20, 25–29}. Their roles appear to be largely conserved \cite{20, 25–29}, and all are differentiated between \textit{A. arenosa} ploidy levels. At zygote-ligulated homologous chromosomes are brough into close apposition by the formation of the synaptonemal complex (SC) \cite{30} crosslinked by a transverse filament protein, Zip1 \cite{31}, which also affects crossover fate \cite{32}. In \textit{A. thaliana} the SC transverse filament is encoded by tandem duplicates, \textit{ZYP1a} and \textit{ZYP1b} \cite{33}, both of which lie under a strong peak of ploidy differentiation in \textit{A. arenosa}. Two other differentiated genes are \textit{SMC3} and scaffold 202722.1 (At1g77600 in \textit{A. thaliana}). The function of \textit{At1g77600} is unknown, but the encoded protein has high homology to PDS5/SP076, which is required in fungi and animals for sister chromatid cohesion and regulation of SC formation in cooperation with cohesins, including SMC3 \cite{34}. All of the meiosis proteins we identified are involved in coordinated processes that contribute to chromosome juxtaposition and chiasma formation. \textit{ASY1} and \textit{ASY3} proteins directly interact and their localization to the chromosome axis requires the presence of \textit{SYN1} \cite{25}. \textit{SYN1} in turn has been shown by mass spectrometry to coprecipitate with \textit{ASY1}, \textit{ASY3}, and \textit{ZYP1} (K. Osman and F.C.H.F., unpublished data). The finding of differentiation in interacting proteins suggests that adaptation to WGD-associated meiotic chromosome segregation challenges might have been multigenic; whether this reflects coevolution or additive contributions to phenotype remains to be tested.

### Conclusions

Understanding the genetic basis of naturally evolved solutions to chromosome segregation with extra homologous copies is relevant to a range of WGD contexts, including crop improvement, polyploid human cancers, and our basic understanding of an evolutionarily important phenomenon. The genes that are sharply differentiated between diploid and tetraploid \textit{A. arenosa} encode proteins that affect the initial juxtaposition and alignment of homologous chromosomes, formation of the SC, and the controlled maturation of recombination intermediates into crossovers or noncrossovers \cite{19–34}. Altering these processes can ultimately affect the number and distribution of crossover events \cite{19, 24–29, 32, 33}. Some cytological studies have found evidence that established polyploids can have reduced crossover frequencies relative to neotetraploids or diploid relatives, and this has been hypothesized as a mechanism of suppressing multivalent formation and thereby stabilizing polyploid meiosis \cite{10, 11}. Our cytological results are consistent with this, and our genome scan results provide a candidate set of genes that could mediate this outcome. It merits mention that an alternative possibility is that some of these alleles may promote unreduced gamete formation in diploids and thus directly contribute to polyploid formation.

There is evidence of parallels with other systems. For example, we observed strong differentiation in \textit{ASY1}, whose homolog has been implicated in meiotic stability in allopolyploid wheat. The wheat gene \textit{Ph1}, the only “diploidization factor” molecularly characterized to date \cite{5, 35}, promotes bivalent formation by solidifying similarity-based pairing fidelity. In the absence of \textit{Ph1}, transcription of the wheat homolog of \textit{ASY1} is increased and its localization is affected, while decreased \textit{ASY1} activity in transgenic lines promoted homologous pairing \cite{36}. Though the genes themselves are not homologs, there are functional similarities among the genes we identified and those critical to tetraploid, but not diploid, yeast cells, which include genes involved in homologous recombination and sister chromatid cohesion \cite{37}. Finally, in humans cancer cells are often polyploid \cite{3}. Though they divide mitotically, a suite of meiosis genes, including a vertebrate homolog of \textit{ASY1} (\textit{HORMAD1}), as well as homologs of \textit{ZYP1} and \textit{SYN1/RE8}, are overexpressed in at least some cancers, where they may contribute to genomic instability and show promise as therapeutic targets \cite{3, 38, 39}. With ours, these studies indicate parallels between kingdoms in processes that affect chromosome segregation after WGD, while our work shows that this conserved process can make evolutionary shifts when necessary.

### Experimental Procedures

#### Plant Material

Plant growth and DNA preparation were previously described \cite{13}. To generate neotetraploids, diploid SN seeds were treated with 0.1% colchicine for 24 hr. We confirmed tetraploidy with chromosome spreads. We assayed pollen viability (n = 90–120 grains/line) using Alexander’s stain \cite{40}.
Cytological Procedures
We fixed inflorescences in 3:1 ethanol:acetic acid. Anthers were isolated and prepared as previously described [41]. Chromosomes were stained with DAPI, mounted in Vectashield (Vector Lab) and visualized using a Nikon 90i Eclipse fluorescent microscope with NIS elements software.

Genome Sequencing
Sequencing libraries were prepared using the Illumina Genomic Sample Preparation Kit and sequenced on an Illumina HiSeq2000. Reads were mapped to the repeatmasked Lyrata1.0 genome [16] using bowtie2 [42], and bam files were processed with Samtools [43] and Picard (http://picard.sourceforge.net). We used GATK [44, 45] for indel realignment, SNP discovery, and genotyping using standard parameters for diploids and the ’’2ploidy 4’’ option for tetraploids. See the Supplemental Experimental Procedures for diploid de novo assembly.

Genomic Analysis
For details, see the Supplemental Experimental Procedures. Sites with coverage in all 24 individuals were binned into 55,570 100-SNP consecutive windows. We calculated $F_{ST}$ between diploids and tetraploids following [17, 46]. We also used a composite likelihood ratio test of the diploid-tetraploid two-dimensional Site Frequency Spectrum (2dSFS) [18] and tested for regions with excess allelic differentiation between diploids and tetraploids for a given diversity within tetraploids. Our final set of differentiated regions was defined as the overlap between these latter two tests.

Accession Numbers
All genomic sequencing reads are available from the NCBI SRA database (bioproject number SRP021057) under accession numbers SRX340942, SRX340943, SRX340944, SRX340945, SRX340946, SRX340947, SRX340948, SRX340949, SRX340950, SRX340951, SRX340952, and SRX341006.

Supplemental Information
Supplemental Information includes Supplemental Experimental Procedures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.08.059.

Acknowledgments
We thank Nancy Kleckner and Dan Hartl for critical reading of the manuscript and thoughtful discussions, Ben Hunter for technical assistance, and members of the Bomblies and Kleckner labs for helpful discussions. This work was supported through a Harvard University Milton Fund Award to K.B. and Ruth L. Kirschstein National Research Service Awards from the National Institutes of Health to L.Y. (1 F32 GM096699) and K.M.W. (1 F32 GM105293). Work in the F.C.H.F. lab is funded by the BBSRC, UK.
Figure 3. Most Differentiated Regions and Examples of Differentiation in Two Sweep Candidates (A) Differentiated regions (vertical lines), with meiosis genes labeled. ZYP1a consists of tandem duplicates, ZYP1b and ZYP1c. (B) Two examples of differentiated regions in meiosis genes. Dots represent polymorphic SNPs. The x axis shows the chromosome location. The y axis shows the degree of differentiation calculated by subtracting diploid from tetraploid allele frequency. Short gaps are regions in which reads did not align as a result of repeat masking, high intergenic polymorphism, or deletions in *A. arenosa* relative to *A. lyrata*. These were verified with alignment of an *A. arenosa* de novo assembly and paired end read information. See also Table S5.

References


