The Molecular and Genetic Characterization of Second Chromosome Balancers in *Drosophila melanogaster*

Danny E. Miller,* † Kevin R. Cook, ‡ Elizabeth A. Hemenway,* Vivienne Fang,§ Angela L. Miller,* Karen G. Hales,§ and R. Scott Hawley*,**,1

*Stowers Institute for Medical Research, Kansas City, Missouri 64110, †MD-PhD Physician–Scientist Training Program and **Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, Kansas 66160, ‡Department of Biology, Indiana University, Bloomington, Indiana 47405, and §Department of Biology, Davidson College, North Carolina 28035

ORCID IDs: 0000-0001-6096-8601 (D.E.M.); 0000-0001-9260-364X (K.R.C.); 0000-0002-9200-1873 (A.L.M.); 0000-0002-6478-0494 (R.S.H.)

ABSTRACT

Balancer chromosomes are multiply inverted and rearranged chromosomes used in *Drosophila melanogaster* for many tasks, such as maintaining mutant alleles in stock and complex stock construction. Balancers were created before molecular characterization of their breakpoints was possible, so the precise locations of many of these breakpoints are unknown. Here, we report or confirm the positions of the 14 euchromatic breakpoints on the 2nd chromosome balancers SM1, SM5, CyO, and SM6a. This total includes three breakpoints involved in a complex rearrangement on SM5 that is associated with the duplication of two genomic regions. Unbiased sequencing of several balancers allowed us to identify stocks with incorrectly identified balancers as well as single and double crossover events that had occurred between 2nd chromosome balancers and their homologs. The confirmed crossover events that we recovered were at least 2 Mb from the closest inversion breakpoint, consistent with observations from other balancer chromosomes. Balancer chromosomes differ from one another both by large tracts of sequence diversity generated by recombination and by small differences, such as single nucleotide polymorphisms (SNPs). Therefore, we also report loss-of-function mutations carried by these chromosomes and unique SNP and InDel polymorphisms present on only single balancers. These findings provide valuable information about the structure of commonly used 2nd chromosome balancers and extend recent work examining the structure of X and 3rd chromosome balancers. Finally, these observations provide new insights into how the sequences of individual balancers have diverged over time.

KEYWORDS

balancer chromosomes inversion breakpoints meiosis whole-genome sequencing crossing over

Balancer chromosomes are multiply inverted and rearranged chromosomes that both suppress recombination during meiosis and prevent the recovery of recombinant chromosomes. Balancers in *Drosophila melanogaster* are used for a number of tasks, such as maintaining deleterious alleles in stock, preserving linkage relationships among alleles, and allowing complex stock construction. Most balancers carry recessive lethal or sterile mutations that prevent them from becoming homozygous in stock, as well as dominant visible markers to make their inheritance easy to follow in crosses. Balancers are available for all chromosomes in *D. melanogaster* except the small 4th chromosome, which does not normally undergo exchange, and the Y chromosome.

Because balancers were created before DNA sequencing was available, the precise locations of most inversion breakpoints and the nature of most of the marker alleles carried by balancers have remained unknown. Recently, the positions of most of the inversion breakpoints on the X chromosome balancer *FM7* and the 3rd chromosome balancers *TM3*, *TM6*, and *TM6B* were reported or confirmed (Miller et al. 2016a; b).
These analyses revealed that several genes are disrupted by breakpoints including the highly conserved tumor suppressor gene p53, which was bisected by an inversion breakpoint on TM3 (Miller et al. 2016a).

Furthermore, whole-genome sequencing (WGS) of multiple X and 3rd chromosome balancers revealed that large chromosomal segments had been exchanged with sequences from non-balancer homologs by meiotic recombination. For example, double crossovers within an inverted segment of the X chromosome balancer FM7c led to loss of the female sterile allele snr52 on multiple occasions (Miller et al. 2016b). Likewise, double crossovers were reported within inverted segments of the 3rd chromosome balancers TM3 and TM6B, and single-exchange events were found to be common in the unbalanced distalmost 7 Mb of chromosome arm 3L in TM3 (Miller et al. 2016a). There was also evidence that balancers had diverged at the nucleotide level since their common origins. Here, we present a similar analysis of sequence diversity for the commonly used 2nd chromosome balancers SM1, SM5, SM6a, and CyO.

These four 2nd chromosome balancers have their origins in two naturally occurring paracentric inversions of 2L and 2R that were on the same chromosome. This chromosome was first described by Ward (1923), who was studying the Curly (Cy) mutation (recently renamed Duxo5; Hurd et al. 2015). She reported that this chromosome (later known as In(2L)Cy + In(2R)Cy) showed reductions in crossing over in each chromosome arm when heterozygous with normal chromosomes. Sturtevant (1926) proposed that these reductions in exchange were due to paracentric inversions and later (Sturtevant 1931) demonstrated that the gene order had indeed changed for 2L. A similar study by Graubard (1932) verified the presence of the paracentric inversion on 2R. By 1936, Calvin Bridges and Ju-Chi Li had confirmed the presence of the two inversions by polytene chromosome analysis and had mapped their breakpoints to 22D:33F for In(2L)Cy and 42A:58A for In(2R)Cy (cited in Morgan et al. 1936; Bridges 1937; and Bridges and Warren 1944). Ward also identified and described a mutation in cinnabar (cnr) on the In(2L)Cy + In(2R)Cy chromosome.

Although In(2L)Cy + In(2R)Cy can be used as a 2nd chromosome balancer, it is not very effective because it still allows exchange with a normal sequence homolog, albeit at a reduced frequency (Ward 1923). In an attempt to create a better whole-chromosome balancer, SM1 was made by irradiating an In(2L)Cy + In(2R)Cy chromosome marked with aP1, Duxo5, cnr2, and sp2 (Lewis and Mislove 1953). Polytene chromosome analysis revealed that irradiation had induced a large pericentric inversion with breakpoints at 22A and 60B. While the addition of an inversion reduced recombination substantially, SM1 still allows low levels of recombination, especially in the large region of 2R not disrupted by inversion breakpoints (Lindsley and Zimm 1992). To isolate a still better balancer, Mislove and Lewis (1955) repeatedly X-rayed SM1, introducing two additional inversions and a complex rearrangement. This balancer, called SM5, is marked with aP1, Duxo5, cnr2, lrP, and sp2, and is associated with good fertility and viability for a chromosome so extensively rearranged (12 euchromatic breakpoints) and, as we will discuss below, at least 4 heterochromatic breakpoints. Unlike the other second chromosome balancers that are euploid, the complex rearrangement in SM5 resulted in the duplication of segments 42A to 42E and 58A to 58F (Mislove and Lewis 1955).

In 1956, Oster reported a new balancer obtained after irradiating males carrying In(2L)Cy + In(2R)Cy marked with Duxo5, dpy30, pr4, and cnr2 (Oster 1956). This balancer is referred to as CyO (Lindsley and Zimm 1992). EMS treatment of CyO several years after it was created changed the weak allele cnr2 to cnr2p, a null allele now present on some, but not all, CyO chromosomes (Craymer 1980). Finally, SM6 was created through a series of single exchanges between CyO and SM1 (Craymer 1984). Two versions of SM6 were made: SM6a marked with aP1, Duxo5, dpy30, crnP, and sp2; and SM6b, which carries the additional marker amosRoi-1.

In this study, we report or confirm the genomic positions of a majority of the breakpoints present on SM1, SM5, CyO, and SM6a. We also identify two previously unannotated marker mutations carried by these chromosomes and precisely define the two large duplicated segments carried by SM5. Furthermore, we find many novel loss-of-function
Although significant sequence diversity due to recombination has been observed among different versions of the X chromosome balancer FM7 and the 3rd chromosome balancers TM3 and TM6B (Miller et al. 2016a; b), we find that, except for two single exchange events in the distal unbalanced region of 2R on CyO and SM5, few large tracts of sequence diversity exist on the SM5, CyO, or SM6a balancers. This suggests that all three of these balancers allow very little recombination with normal-sequence homologs and thus are near-complete balancers. Because SM5 arose from SM1 through two inversions (In(2L)SM5-1 followed by In(2L)SM5-2) and a complex rearrangement, we have given symbols to these component aberrations to replace the single aberration In(2L)SM5. The symbol Dp(2;2)SM5 is chosen for the complex rearrangement to emphasize the duplicated segment from the progenitor over the inverted segment.

This breakpoint could not be localized molecularly, but recessive lethality presumably associated with this breakpoint was mapped to the interval 2L:8,529,124–8,700,124 by complementation tests with molecularly defined chromosomal deletions.

The proximal side of the breakpoint is present in two presumably identical copies juxtaposed to low-complexity sequence and maps to the 2L:6,916,809–8,700,124 interval defined by chromosomal deletions. The sequence on the distal side of the breakpoint suggests this gene is disrupted.

A second, intact copy of this gene is present elsewhere on SM5.

**MATERIALS & METHODS**

**Stocks used for breakpoint identification**

All balancers sequenced in this study were obtained from stocks at the Bloomington Drosophila Stock Center (Table S1). Sequence data for four balancers from a previous study (Miller et al. 2016a) were also included. Before sequencing, multiple balancer-carrying males were crossed to multiple ISO-1 virgin females. ISO-1 is the Drosophila reference genome stock and was obtained in 2014 from the Berkeley Drosophila Genome Project (Hoskins et al. 2015).

**DNA preparation and genome alignment**

DNA for sequencing was prepared from balancer/ISO-1 males using the Qiagen DNeasy Blood and Tissue Kit. Flies were starved for 1 hr before freezing at –80°C. Libraries were prepared and quantified as described in Miller et al. (2016a). All libraries were pooled, requantified, and sequenced in 150-bp paired-end mode on the Illumina NextSeq 500 instrument. Illumina Real Time Analysis version 2.4.6 was run to demultiplex reads and generate FASTQ files following sequencing. Alignment to the D. melanogaster reference genome (dm6) was performed using bwa version 0.7.15-r1140 (Li and Durbin 2009). SNPs were called using SAMtools version 1.3 and BCFtools version 1.4.1 (Li et al. 2009). DNA preparation, sequencing, and alignment for balancers from stocks 504, 22239, 24759 and the Hawley lab SM1 stock are described in Miller et al. (2016a).

**Identification of balancer breakpoints**

Breakpoints were identified as in Miller et al. (2016a). Briefly, split and discordant read pairs were isolated using SAMBLASTER (Faust and Hall 2014) from regions where rearrangements were previously reported to be present (Lindsley and Zimm 1992). Split and discordant pairs were then de novo assembled and BLAST was used to identify assembled fragments that aligned to two distinct regions of the genome. Breakdancer was used to validate our custom analysis and to search for novel rearrangements (Chen et al. 2009).

**PCR and Sanger sequencing**

Primers for PCR validation were designed using Primer3 (Rozen and Skaletsky 2000). Six of 10 inversion breakpoints in which the molecular position was confirmed or identified in this study were validated using
PCR and Sanger sequencing (Table S2). Briefly, ExTaq polymerase was used according to the manufacturer’s instructions. Extension times and annealing temperature for each breakpoint are given in Table S2.

Identification of shared and unique SNPs

Shared and unique SNPs were identified using VCFtools version 0.1.15 (Danecek et al. 2011). Only SNPs with VCF quality scores >220 were considered for analysis. VCF files were merged using vcf-merge (part of the VCFtools package), which allowed the counting of shared and unique SNPs. SnpEff version 4.3p (Cingolani et al. 2012) was used to annotate VCF files. Filtering of annotated VCF files was done using custom scripts.

Complementation testing

Fly crosses were made on standard medium and reared under routine conditions (details provided on request). Crosses listed in Table S3 evaluated recessive phenotypes associated with balancer breakpoints. Stocks used to evaluate breakpoints were tested with the control crosses listed in Table S4. Genotypes and the sources of stocks are given in Table S1.

Testis dissection and microscopy

Flies for microscopy were grown on Ward’s Instant Drosophila Medium and maintained at 25°. dpy^{wy} wg^{1}/SM5 males were crossed to Df(2L) Exc6005/CyO females and male offspring with genotype Df(2L)Exc6005/SM5 were selected and used for testis dissections. dpy^{wy} wg^{1}/Df(2L)Exc6005 males were used as a control. Testes were dissected in TB1 buffer (15 mM 1:1 K2HPO4:KH2PO4, pH 6.7, 80 mM KCl, 16 mM NaCl, 5 mM MgCl2, 1% PEG 6000) on a microscope slide. A coverslip was placed on the dissection and excess buffer was drained slowly using a Kimwipe. Samples were visualized and documented using the phase-contrast setting of an Olympus BX60 Upright Compound Microscope equipped with an Olympus DP73 Color Camera.

Data availability

All stocks are available from the Bloomington Drosophila Stock Center, with the exception of the Hawley lab SM1, TM3 stock, which is available upon request. Raw sequencing data for all samples have been uploaded to the National Center for Biotechnology Information (NCBI) at http://ncbi.nlm.nih.gov and can be found under BioProject PRJNA413446. Sequencing data for balancers from stocks 504, 22239, 24759 and the Hawley lab SM1 stock were submitted to NCBI previously and can be found under BioProject PRJNA315473. Scripts used to align data, call SNPs, create heatmaps, and identify shared and unique mutations are available on Github: https://github.com/danrdanny/2ndChromosomePaper. Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at http://www.stowers.org/research/publications/libpb-1257.

RESULTS

Sequencing and identification of inversion breakpoints

To identify the breakpoints carried by the 2nd chromosome balancers SM1, SM5, CyO, and SM6a, we crossed 18 stocks carrying one of these balancers to the Drosophila reference genome stock, ISO-1, and

![Figure 1 Second chromosome balancer inversion breakpoints and rearrangements. Chromosome bands here and in the text are those predicted from the genomic coordinates of breakpoints in Table 2. (A) Breakpoints whose genomic positions are known are shown as solid lines, those approximated (22D1 and 33F4) are shown as dashed lines, and those with unknown genomic coordinates (29D5–E4 and 40F) are shown as dotted lines. (B) Multiple rearrangements and inversions have resulted in novel configurations for each second chromosome balancer.](image)
selected balancer/ISO-J males for WGS. We also analyzed data from four 2nd chromosome balancers sequenced previously (Miller et al. 2016a) (Table 1). Among this panel of 22 balancer chromosomes, we identified four cases where balancers had been incorrectly identified in the original stocks (Table 1). Throughout the text we refer to the mislabeled stocks by their stock numbers and group them with their correct balancer chromosome in figures.

Previous studies based on polytene squashes reported the presence of 14 euchromatic breakpoints and one heterochromatic breakpoint on the SM1, SM5, CyO, and SM6a balancer chromosomes (Table 2; Figure 1) (summarized in Lindsley and Zimm 1992). Eleven of the euchromatic breakpoints as well as the heterochromatic breakpoint are associated with simple inversions. The remaining three euchromatic breakpoints are associated with the complex rearrangement on SM5 that duplicated two genomic regions: 42A7–42E1 and 58A4–59A2 (Table 2).

Using a combination of standard and large-insert libraries, we confirmed or identified the precise molecular positions of 8 of the 11 euchromatic inversion breakpoints (Table 2) and validated 6 of those using PCR and Sanger sequencing (Table S2). Although we were unable to identify the precise positions of three of these breakpoints (Table 2), we were able to map the two In(2L)Cy breakpoints in polytene bands 22D1 and 33F4 to within 10 kb using large-insert libraries. The exact position of the euchromatic 29D5–E4 breakpoint on SM5 was difficult to determine because the 40F breakpoint of the 29D5–E4 to 40F inversion lies in centric heterochromatin, making it challenging to analyze using either short-read or large-insert libraries. Nevertheless, we will present evidence below that it can be localized to the 16-gene interval 2L:8529124–8700124 corresponding to 29D5E4.

The complex rearrangement on SM5 (Dp(2;2)SM5) that duplicated 42A7–42E1 and 58A4–59A2 is shown in Figure 2. From the order and orientations of chromosomal segments in SM5, it is apparent this rearrangement was induced on an SM1 chromosome that already had two additional inversions (In(2L)SM5-1 and In(2L)SM5-2) (Table 2, Figure 2C). Euchromatic breaks in 42E1, 53D1 and 59A2 were involved to invert the 42E1–53D1 segment and insert a second copy of a segment from the progenitor chromosome that spanned an existing inversion breakpoint and comprised two subsegments (42E1–42A7 and 52A4–59A2) (Figure 2B) (Mislove and Lewis 1955; Lindsley and Zimm 1992). Our sequencing showed the presence of low-complexity sequences juxtaposed to the distal ends of both 42A7–42E1 segments, and they provide a clue to how the unusual mirror-image arrangement of segments may have arisen. They suggest the distal end of the 42A7–42E1 segment in the progenitor was first joined to a heterochromatic region and then, after replication, breaks within the heterochromatic region on sister chromatids joined to produce the mirrored configuration. The exact overall configuration of the intermediate rearrangement is not clear, but it likely involved inversion of the 42E1–53D1 segment as well. The low-complexity sequences made it difficult to determine the exact distal extent of the 42A7 to 42E1 segments (Table 2), and to know if there are sequences deleted or duplicated relative to the distal side of the 42E1 breakpoint. We determined the minimal extent of the duplicated 42A7 to 42E1 segment to be 2R:6,012,459–6,916,809 (Table 3; Figure 2A), an interval containing 83 protein-coding genes, 14 non-protein-coding (CR) genes, and 19 tRNA genes (6.17 annotations) (Table S5). In contrast, it was relatively straightforward to characterize the 42E1;53D1 and 53D1;59A2 junctions, which flank the inverted segment and join it to the duplicated segment. Depth of coverage analysis allowed us to determine the precise extent of the duplicated 58A4 to 59A2 segment to be 2R:21,972,072–22,689,962 (Table 3, Figure 2A), an interval containing 117 protein-coding genes, 18 CR genes, and
Table 3 Details of the duplicated segments on SM5

| Balancer | Start | End | Size | Predicted bands |
|----------|-------|-----|------|-----------------|
| SM5      | 2R:6,012,459 | 2R:6,916,809 | 904,350 | 42A7–42E1 |
|          | 2R:21,972,072 | 2R:22,689,962 | 712,929 | 58A4–59A2 |

1Breakpoint position predicted from genomic coordinates using FlyBase correlation table.

2This coordinate represents the minimal distal extent of this interval. The maximal extent is 2R:6,917,405.

6 snoRNA genes (Table S5). Knowing exactly which genes are duplicated is valuable for the construction of stocks. For example, SM5 has been used to maintain deficiencies of the haploinsufficient locus M(2)88F, which corresponds to RpS24 and/or RpS16, two genes contained within the 58A4 to 59A2 duplicated segment (Marygold et al. 2007).

Inversion breakpoints affect protein-coding genes

Several of the breakpoints we identified or confirmed lie within the transcribed regions of protein-coding genes (Table 2). Breakpoints that lie in intergenic regions and those that we mapped to small intervals may also affect the activities of genes. To test the breakpoints for strong phenotypic effects, we performed complementation tests with deficiencies spanning these breakpoints and scored for lethality, female sterility, or grossly abnormal morphology (Table S6). Only two of the breakpoints, 21E2 and 29D5–E4 on SM5, gave phenotypes in these tests. The 21E2 breakpoint in the gene dachisous (ds) was lethal, with the few escapers having short appendages typical of ds mutants, consistent with previous reports (Craymer 1980; Clark et al. 1995). The 29D5–E4 breakpoint on SM5 also appears to be lethal. Four deletions chosen to span the region of the breakpoint defined by polytene chromosome analysis (29C–E; (Lindsley and Zimm 1992)) were all lethal in combination with SM5. Assuming that the inversion breakpoint is the only lethal mutation present, it maps to the 16e region common to all the deletions (2L:12,726,221–12,736,221), which corresponds to 29D5–E4. While the 30D1 breakpoint disrupting nAChRa6 on SM6a and CyO is not associated with lethality or female sterility, previous studies showed that it and other loss-of-function nAChRa6 mutations confer insecticide resistance (Perry et al. 2007; Watson et al. 2010).

We were unable to evaluate the phenotypic effects of two breakpoints. The 59A2 breakpoint of the complex rearrangement on SM5 disrupts CR44763, but an intact copy is present elsewhere on SM5. The 42E1 breakpoint of the complex rearrangement appears to disrupt CG30158, because the sequence of the 42E1–53D1 junction shows that the break lies at 2R:6,917,406 within an intron. Nevertheless, the difficulties we encountered in characterizing the 42E1 ends of the mirror-image 42A7–42E1 segments leave us unable to say with certainty that they terminate at the same site. If, as we expect, the break is not unusually complicated and CG30158 is disrupted, our complementation tests (Table S6) indicate that knocking out CG30158 has no severe consequences.

Although it did not confer lethality or female sterility in complementation tests, the 22A3 breakpoint on SM1, SM5, and SM6a bisects the gene CG10869, which is expressed only in adult testes (Chintapalli et al. 2007). Complementation testing of SM5 with Df(2L)Exel6005, which encompasses CG10869, showed that males were sterile and that sperm failed to individualize during spermatogenesis (Figure 3). Consequently, we have renamed this gene no individualized sperm (nis). Although nis lies within an intron of hattifattener (haf), it is unlikely that disruption of haf contributes to male sterility, because haf shows negligible expression in the testis (Chintapalli et al. 2007). We did not examine SM5/Df(2L)Exel6005 flies for the incompletely penetrant muscle innervation defects seen when haf expression is reduced by RNAi or mutations (Kurusu et al. 2008).

That we were able to discover a previously unknown mutation and determine the role of an uncharacterized gene demonstrates the value of molecularly mapping the inversion breakpoints on these commonly used balancer chromosomes. Furthermore, knowledge of the genes disrupted by inversion breakpoints is useful for researchers studying those genes in terms of their choice of balancer.

Sequence diversity among 2nd chromosome balancers

As seen in studies of X and 3rd chromosome balancers (Miller et al. 2016a; b), stretches of unique SNPs indicate the occurrence of recombination events between balancers and nonbalancer homologs. To investigate how much sequence diversity, if any, exists among each of the four 2nd chromosome balancers, we plotted the positions of SNPs that were unique among balancers of the same type and found very few large tracts on the SM5, CyO, and SM6a balancers (Figure 4 B–D). Two exchange events were observed at the distal tip of 2R: one on CyO from stock 504 (Figure 4C, Figure 5A) and one on SM5 from stock 240 (Figure 5B). The single crossover (SCO) on CyO occurred at approximately 2R:24,509,500, ~2.5 Mb away from the distalmost 58A4 breakpoint. The SCO on SM5 occurred at approximately 2R:25,241,500, ~1.1 Mb away from the distalmost 60B11 inversion breakpoint. What appears to be a double crossover (DCO) event on CyO from stock 31 (Figure 4C) is in reality a duplication event discussed below.

Unlike the other three balancers, we did find extensive sequence diversity within one region of 2R among the three SM1 chromosome sequences (Figure 4A). This was not surprising, as it was previously noted that the large 16-Mb interval between the 42A7 and 58A4 inversion breakpoints is vulnerable to DCO events (Lindsley and Zimm 1992). It is possible that one of the three SM1 sequences is identical to the sequence of the original SM1 balancer and the other two sequences represent DCO events, or that all three sequences represent DCO events. To determine which scenario is correct, we compared the SNP distribution of each SM1 chromosome against SM5, which was created from SM1, and which we assume is an ancestral snapshot of SM1. This revealed that only the SM1 chromosomes from stock 24759 and the Hawley lab stock had experienced DCO events (Figure 4E, Figure 5C–D). That only two of the three SM1 chromosomes differed from SM5 suggests that the SM1 chromosome used to make SM5 carried the SNP profile of the original SM1 chromosome.

Interestingly, the crossovers giving rise to these DCOs on SM1 occurred at least 2.3 Mb away from the flanking inversion breakpoints (Figure 5C–D). This distance and the ~2.5 Mb distance observed for the SCO on CyO from stock 504 are consistent with previous observations that found SCO events between the 3rd chromosome balancer TM3 and a normal-sequence chromosome occurred no closer than ~2 Mb to an inversion breakpoint (Miller et al. 2016a; b). The SCO that occurred ~1.1 Mb from the distalmost breakpoint on 2R in SM5 from stock 240 is challenging to interpret in this regard. It lies approximately 30 kb from telomeric repeats, and a region with increased recombination was reported previously for a similar subtelomeric region of the X chromosome (Anderson et al. 2008). This SCO may reflect a similar region of increased

1166 | Miller et al. | G3 Genes Genomes Genetics
Neither of these duplications contain any of the haploinsufficient observations show that there are unique structural variations within cataloged by Cook to prove Stock Center shortly after the sequence analysis, so follow-up mapping fortunately, the duplicated segment was 50% higher than background, the duplicated segment did not seen on any other corresponding de 2R carried a novel 402-kb duplication with approximate coordinates 64 spermatids elongate together in smooth bundles (E) and later individualize (F) in the control, while Df(2L)Exel6005/SM5 spermatid bundles (K) appear vacuolated (arrow), suggesting tissue degeneration, and sperm fail to individualize (L). The syncytial appearance of cells in many panels is a well-known artifact of live testis squash preparations in which ring canals between cells in a cyst are often broken open. Scale bar, 10 μm.

Figure 3 Spermatogenesis phenotypes from disrupting CG10869. The noncomplementation of Df(2L)Exel6005 and the SM5 inversion breakpoint in CG10869 results in recessive male sterility and degradation of late elongation-stage spermatid bundles. Phase-contrast micrographs of live squashed testis preparations from (A–F) Df(2L)Exel6005/dpylin pert wgl−1 (indistinguishable from wild type) and (G–L) Df(2L)Exel6005/SM5 males. (A, G) Mitochondria are phase dark, small, and diffuse throughout the cytoplasm in primary spermatocytes (arrows denote mitochondria). (B, H) After meiosis, mitochondria aggregate and fuse beside each phase light nucleus, forming the nebenkern (arrows denote nebenkern), which appears normal in testes from Df(2L)Exel6005/SM5 males. (C, D, I, J) The nebenkern unfurls, and mitochondrial derivatives appear to elongate normally beside the growing flagellar axoneme during early elongation stages (arrows denote elongating mitochondrial derivatives). Cysts of 64 spermatids elongate together in smooth bundles (E) and later individualize (F) in the control, while Df(2L)Exel6005/SM5 spermatid bundles (K) appear vacuolated (arrow), suggesting tissue degeneration, and sperm fail to individualize (L). The syncytial appearance of cells in many panels is a well-known artifact of live testis squash preparations in which ring canals between cells in a cyst are often broken open. Scale bar, 10 μm.

recombination on 2R. Whether subtelomeric exchanges reflect normal meiotic recombination or events associated with the maintenance of chromosome ends remains unclear (Kern and Begun 2008).

Structural differences exist among balancer stocks

Unbiased sequencing of a large number of 2nd chromosome balancers also allowed us to identify novel and unexpected structural events within each stock. For example, we sequenced six balancers from stocks with deficiencies (Table 1) and identified two novel duplications (Figure 6). The CyO in stock 1602 balances Df(2L)TW65, a deletion of 37F–39E (approximate coordinates 2L:19,675,000–21,700,000), and carries a novel 1.2-Mb tandem duplication with coordinates 2L:20,422,204–21,606,670. Stock 31 contained CyO balancing Df(2R)H3D3, a deletion of 44D–44F (approximate coordinates 2R:8,430,000–8,850,000), and it carried a novel 402-kb duplication with approximate coordinates 2R:8,376,000–8,777,700. This duplicated region carried polymorphisms not seen on any other CyO balancer, suggesting that it originated from a chromosome other than CyO (Figure 4C). Both its proximal and distal ends were bounded by low-complexity or repetitive sequence, making it difficult to determine using paired sequencing reads where it was positioned in the genome. Because the sequencing depth of the duplicated segment was 50% higher than background, the duplicated segment appears to have co-segregated with CyO in the outcross to ISO-1. Unfortunately, the Df(2R)H3D3 stock was rebalanced at the Bloomington Stock Center shortly after the sequence analysis, so follow-up mapping to prove CyO carried the duplication was impossible. Regardless, these observations show that there are unique structural variations within balancer stocks and that balancers themselves can carry duplications. Neither of these duplications contain any of the haploinsufficient genes cataloged by Cook et al. (2012), yet they both partially overlap corresponding deficiencies and restore many genes to their normal two copies, which likely explains why they were retained and why, in general, a balancer carrying a duplication can have a competitive advantage when it arises in a stock.

Marker alleles carried by 2nd chromosome balancers

Whole-genome sequencing provides an opportunity to identify and define the molecular nature of marker alleles carried by balancer chromosomes. Among these four balancers, there are both shared and unique visible mutations, which may be used to differentiate the balancers. For example, SM1 is marked with all Durox/ crv/sp, while SM5 carries those markers as well as ds55 and ltv, and SM6a is marked with all Durox dpy/pr/ct sp. Most CyO balancers, meanwhile, are marked with Durox dpy/pr/ct, although some have cv instead of ct (Craymer 1980). Of these nine marker alleles, five (ds55, Durox, pr, ct, and sp) have been sequenced previously, and we were able to confirm the molecular nature of all five (Table 4) as well as the nature of the nACHRα6 mutation associated with insecticide resistance (Perry et al. 2007; Watson et al. 2010). We also identified molecular abnormalities in two (al2 and ltv) of the four previously unsequenced markers (al2, Durox, pr, ct, and cv). Note, however, that ltv is a variegating allele of lt, and it is possible that the 40F breakpoint in SM5 induces position-effect variegation rather than the variegation being attributable to the transposon insertion we identified. We were unable to determine the nature of the lesions in dpypr and crsp.

Mutations that are shared by and unique among balancer chromosomes

Many balancer chromosomes were created using X-ray mutagenesis, which generated not only useful inversions but also unrecognized mutations. Furthermore, as balancers have been kept in stock, they have diverged from one another over time through the accumulation of de novo SNP and InDel polymorphisms. Finally, crsp was induced by EMS treatment, which undoubtedly increased the number of mutations present on balancers carrying this marker (Craymer 1980).
From the five SM5, nine CyO, and five SM6a balancers sequenced, we identified 234,623 high-quality SNPs shared by at least two of these three balancer types. (We did not include SM1 in this analysis because we felt that three stocks, two of which had experienced large DCO events, were not sufficient to determine accurately which mutations were shared or novel.) We used SnpEff (Cingolani et al. 2012) to determine how many of these SNPs affect genes (both protein-coding and noncoding) and found 35 nonsense mutations, 62 splice site mutations, 9 start-loss mutations, and 8,898 missense mutations (summary statistics in Table 5, list of all shared mutations in Table S7). We also found 1,558 high-quality SNPs shared only among all SM5 balancers, 13,888 shared only among all CyO balancers, and 20,567 shared only among all SM6a balancers. These polymorphisms also introduce nonsense, splice acceptor, and missense mutations that are unique to these chromosomes (Table 5, Table S7). For example, all CyO balancers we sequenced have nonsense mutations in 2 genes, CG33310 and CG33750, not present on any of the SM5 or SM6a balancers sequenced. Finally, we identified an average of 2,627 unique SNPs per

Figure 4 Heatmaps of unique SNPs reveals sequence diversity among 2nd chromosome balancers. (A) Extensive sequence diversity appears to exist on the right arm of the 2nd chromosome among SM1 balancers from three stocks, indicating that these regions are susceptible to double crossover events. (B–D) Little sequence diversity exists among the SM5, CyO, and SM6a chromosomes sequenced, except for a single exchange event on the distal 2R tip of CyO from stock 504. One apparent double crossover event on CyO from stock 31 is actually a duplication. (E) Comparing SNPs present on all SM1 stocks to SM5, which SM5 carries the ancestral SM1 polymorphism profile, reveals that SM1 chromosomes from two stocks experienced double crossover events while SM1 from stock 325 did not.
individual balancer chromosome. These unique SNPs resulted in a total of 683 unique missense mutations, 10 unique splice site mutations, and 11 unique nonsense mutations among all the balancers we studied (Table 5, list of all unique mutations in Table S8). As an example, CyO from stock 1602 has a nonsense mutation in Nplp4 that is not observed on any other balancer. That there are many mutations affecting genes both shared and unique among balancer chromosomes demonstrates that stock-to-stock variability exists among balancers and this variability may impact the fitness of stocks or the interpretation of experimental results.

DISCUSSION

Whole-genome sequencing of balancer chromosomes provides several important types of information. For example, it can tell us whether a balancer carries any mutations aside from its marker alleles, whether it has acquired any additional structural variations such as duplications or deletions, and which of its regions have undergone crossing over. It can also help us determine the exact molecular locations of breakpoints, which are useful for understanding how breakpoints may directly or indirectly affect genes.

Knowledge of the exact positions of inversion breakpoints has important consequences for choosing the appropriate balancer for maintaining an allele in stock. Because these breakpoints often bisect or lie very close to genes, they can disrupt gene function or affect it by position-effect suppression. For example, the 15D3 inversion breakpoint on the FM7 balancer bisects the predicted peptidase gene CG45002 (Miller et al. 2016b), and the 84B1 inversion breakpoint on TM6B affects the regulatory region of Antp (Miller et al. 2016a). In this study, we report or confirm the exact or approximate positions of eleven euchromatic inversion breakpoints present on the 2nd chromosome balancers SM1, SM5, CyO, and SM6a (Table 2) along with the positions of the three euchromatic breakpoints of the duplication-associated complex rearrangement on SM5 (Table 3). Knowing the exact genes carried by the duplicated segments is important for researchers studying genes in these intervals who may not realize that their gene of interest is present in three copies.

Moreover, we find that balancers are not immune to de novo structural variation that may provide selective advantages in stock. In our study, two out of six balancers maintained in stock over a deficiency were associated with a duplication covering a large portion of the deficiency (Figure 6). At least one, and possibly both, of these duplications arose after the chromosomes were placed in stock with the deficiency, confirming that copy-number variation arises frequently enough for such duplications to be a concern. This serves as a reminder that balancers kept in stock for long periods of time may carry unexpected structural variation that could affect experimental results.

By comparing WGS data of related balancer chromosomes, we can identify exchange events between a balancer and its normal sequence homolog. Balancers have previously been shown to experience SCO events within terminal, noninverted segments and DCO events within...
inverted segments (Miller et al. 2016a; b), and both types of exchange events were observed in this study as well (Figure 5). It is notable that, like exchange events that occurred between an internal region of the 3rd chromosome balancer TM3 and normal-sequence homologs (Miller et al. 2016a), the non-subtelomeric exchange events observed in this study occurred at least 2 Mb from the nearest inversion breakpoint. This strengthens the conclusion that inversion breakpoints suppress exchange over distances of 2–3 Mb.

Similar to observations of single exchange events in the distal unbalanced region of the 3rd chromosome balancer TM3, the 2nd chromosome balancer CyO allows exchange near the tip of 2R (and may also allow single exchange events near the tip of 2L). We therefore suggest that mutations in the intervals distal to 22D1 on 2L and 58A4 on 2R be balanced with SM5—not with CyO. (However, SM6a or SM1 should work well to maintain mutations in the distal tip of 2R.) Even though an SCO event was observed in the distal portion of 2R on SM5, this extremely distal event occurred in the last 30–40 kb of the chromosome and affected only 5 genes. Thus, we still encourage the use of SM5 for balancing distal genes. We also recommend maintaining more than one independent culture of stocks where there is risk of losing mutations from exchange with balancers.

Finally, WGS also provides an opportunity to investigate both the shared and unique SNPs affecting gene function that are carried by balancer chromosomes. Here, we were able to confirm or molecularly characterize 7 of the 9 previously reported marker alleles carried by these four 2nd chromosome balancers (Table 4). In addition to known visible markers, all chromosomes carry a number of other mutations, most of which remain uncharacterized phenotypically. A study by Araye and Sawamura (2013) reported the presence of novel recessive lethal mutations carried by two balancers maintained in their lab. They suggested that WGS should reveal additional novel mutations affecting gene function. We identified 97 nonsense and splice site mutations (Table 5) shared among all four balancers studied (Table S7) as well as 26 nonsense and splice site mutations (Table 5) that are either unique among a family of balancers (Table S7) or unique to a specific balancer (Table S8). Knowing the genes that are mutated on balancers is important for researchers working with a specific gene who may not realize the balancer chromosome they are using is creating a heteroallelic loss-of-function genotype.

Whole-genome sequencing using short-read technology has now been completed for the Drosophila X chromosome balancers FM7a and FM7c (Miller et al. 2016b); the 2nd chromosome balancers SM1, SM5, CyO, and SM6a; and the 3rd chromosome balancers TM3, TM6, and TM6B (Miller et al. 2016a). These studies have revealed surprising findings about the structures of these chromosomes, the mutations carried by them, and the sequence diversity that exists among them. Sequencing a panel of presumably identical balancer chromosomes has also added to our understanding of the mechanisms by which inversion

### Table 4 Marker alleles present on 2nd chromosome balancers

| Balancers | Allele | Mutation |
|-----------|--------|----------|
| SM1, SM5, SM6a, CyO | Duox<sup>Cy</sup> | Nonsynonymous C-to-A mutation changing Gly 1505 to Cys, consistent with Hurd et al. (2015).<br>cn<sup>2</sup> has two roo insertions: one with a 70-nt target site duplication (TSD) of 2R:7,784,487–7,784,556 that includes the 1st intron and 2nd exon of cn, and the other in the 1st intron of cn with a 6-nt TSD of 2R:7,784,682–7,784,687. Previously reported as an 8-kb insertion in cn (Warren et al. 1996). |
| SM1, SM5, CyO | cn<sup>2</sup> | Previously unknown. 10-nt deletion in 4th exon of aI resulting in a frameshift. Deletion is from 2L:386,781–386,793. The deletion may be a single deletion or two deletions, as 3 nt of sequence aligns in multiple positions within the 13-nt interval. |
| SM6a, CyO | cn<sup>2P</sup> | Previously reported 412 insertion at 5’ UTR of Dat. Previously known (Eric Spana, personal communication). |
| SM1, SM5, SM6a | aI<sup>2</sup> | 21D2–36C inversion breakpoint is in 1st intron of ds. Previously reported (Clark et al. 1995). |
| SM5 | ds<sup>55</sup> | Previously unknown. Doc insertion at near 2L:22,927,436, which may disrupt the splice acceptor site of the 6th exon. Alternatively, the variegating phenotype may be due to the 40F breakpoint, which may be near it. |
| CyO | nAChRs<sup>6a</sup> | 30F–50D inversion breakpoint is within an intron of nAChRs<sup>6a</sup>. Previously reported (Perry et al. 2007). |
| CyO | dpy<sup>1</sup> | Unknown. |

### Table 5 Shared or unique mutations carried by 2nd chromosome balancers

| Balancer | Missense | Splice site | Nonsense | Start-loss |
|-----------|----------|-------------|----------|------------|
| All SM5 only | 49 | 0 | 1 | 0 |
| All CyO only | 548 | 2 | 2 | 0 |
| All SM6a only | 520 | 0 | 0 | 0 |
| Shared among all balancers | 8898 | 62 | 35 | 9 |
| Present on only one balancer | 683 | 10 | 11 | 0 |
breakpoints suppress exchange, demonstrating that studies of these commonly used genetic tools can provide insights into challenging biological questions.

ACKNOWLEDGMENTS
We thank members of the Hawley lab for critically reading the manuscript and offering feedback, our colleagues at the Bloomington Drosophila Stock Center for their support, Bob Levis for helpful discussions, and Eric Spana for sharing his unpublished result that sp2 is an allele of Dat. We also thank members of the molecular biology core at the Stowers Institute for their expert help with library preparation and sequencing. KRC is supported by grant P40 OD018537. VF’s work in KGH’s lab was supported by funds from the Arnold and Mabel Beckman Scholar Program at Davidson College. RSH is an American Cancer Society Research Professor.

LITERATURE CITED
Anderson, J. A., Y. S. Song, and C. H. Langley, 2008 Molecular population genetics of Drosophila subtelomeric DNA. Genetics 178(1): 477–487. https://doi.org/10.1534/genetics.107.083196
Araya, Q., and K. Sawamura, 2013 Genetic decay of balancer chromosomes in Drosophila melanogaster. Fly (Austin) 7(3): 184–186. https://doi.org/10.4161/fly.24466
Bridges, C. B., 1937 Correspondences between linkage maps and salivary chromosome structure, as illustrated in the tip of chromosome 2R of Drosophila melanogaster. Cytologia (Tokyo) Fuji jubil(2): 745–755. https://doi.org/10.1508/cytologia.FujiiJubilaei.745
Bridges, C. B., and K. B. Warren, 1944 The Mutants of Drosophila melanogaster, Carnegie Institution of Washington, Washington, DC.
Chen, K., J. W. Wallis, M. D. McLellan, D. E. Larson, J. M. Kalicki et al., 2009 Breakdancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods 6(9): 677–681. https://doi.org/10.1038/nmeth.1363
Chintapalli, V. R., J. Wang, and J. A. T. Dow, 2007 Using FlyAtlas to screen of cell-surface molecules identified through chemically induced target site resistance, resistance gene identification, and heterologous expression. Insect Biochem. Mol. Biol. 40(5): 376–384. https://doi.org/10.1016/j.ibmb.2009.11.004
Clark, H. F., D. Brentrup, K. Schneitz, A. Bieber, C. Goodman et al., 1995 Dachshous encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in Drosophila. Genes Dev. 9(12): 1530–1542. https://doi.org/10.1101/gad.9.12.1530
Cook, R. K., S. J. Christensen, J. A. Deal, R. A. Coburn, M. E. Deal et al., 2012 The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. Genome Biol. 13(3): R21. https://doi.org/10.1186/gb-2012-13-3-r21
Craymer, L., 1980 New mutants report. DIS 55: 197–200.
Craymer, L., 1984 New mutants report. DIS 60: 715–720. https://doi.org/10.1534/genetics.116.029330
Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks et al., 2011 The variant call format and VCFTools. Bioinformatics 27(15): 2156–2158. https://doi.org/10.1093/bioinformatics/btr330
Faust, G. G., and J. M. Hall, 2014 SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics 30(17): 2503–2505. https://doi.org/10.1093/bioinformatics/btu314
Graubard, M. A., 1932 Inversion in Drosophila melanogaster. Genetics 17: 81–105.
Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez et al., 2015 The Release 6 reference genome sequence of the Drosophila melanogaster genome. Genome Res. 25(3): 445–458. https://doi.org/10.1101/gr.185579.114
Hurd, T. R., F.-X. Liang, and R. Lehmann, 2015 Curly encodes dual oxidase, which acts with heme peroxidase Curly Su to shape the adult Drosophila wing. PLoS Genet. 11(11): e1005625 (erratum: PLoS Genet. 12(8): e1006285). https://doi.org/10.1371/journal.pgen.1005625
Kern, A. D., and D. J. Begun, 2008 Recurrent deletion and gene presence/absence polymorphism: telomere dynamics dominate evolution at the tip of chromosome 3L in Drosophila melanogaster and D. simulans. Genetics 179(2): 1021–1027. https://doi.org/10.1534/genetics.107.078345
Kim, N., J. Kim, D. Park, C. Rosen, D. Dorsett et al., 1996 Structure and expression of wild-type and suppressible alleles of the Drosophila purple gene. Genetics 142: 1157–1168.
Kuru, M., A. Cording, M. Tanimichi, K. Menon, E. Suzuki et al., 2008 A screen of cell-surface molecules identifies leucine-rich repeat proteins as key mediators of synaptic target selection. Neuron 59(6): 972–985. https://doi.org/10.1016/j.neuron.2008.07.037
Lewis, E. B., and R. F. Mislove, 1953 New mutants report. DIS 27: 57–58.
Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16): 2078–2079.
Lindsley, D. L., and G. G. Zimm, 1992 The Genome of Drosophila melanogaster, Academic Press, San Diego, CA.
Marygold, S. J., J. Roote, G. Reuter, A. Lambertsson, M. Ashburner et al., 2007 The ribosomal protein genes and Minute loci of Drosophila melanogaster. Genome Biol. 8(10): R216. https://doi.org/10.1186/gb-2007-8-10-r216
Miller, D. E., K. R. Cook, A. V. Arvanitakis, and R. S. Hawley, 2016a Third chromosome balancer inversions disrupt protein-coding genes and influence distal recombination events in Drosophila melanogaster. G3 (Bethesda) 6(7): 1959–1967. https://doi.org/10.1534/g3.116.029330
Miller, D. E., K. R. Cook, N. Yeganeh Kazemi, C. B. Smith, A. J. Cockrell et al., 2016b Rare recombination events generate sequence diversity among balancer chromosomes in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 113(10): E1352–E1361. https://doi.org/10.1073/pnas.1601232113
Mislove, R. F., and E. B. Lewis, 1955 SM5: Second Multiple 5. DIS 29: 75.
Morgan, T. H., C. B. Bridges, J. Schultz, and J. Schultz, 1936 Constitution of the germinal material in relation to heredity. Year B. Carnegie Inst. Wash. 55: 289–297.
Oster, I. L., 1956 A new crossing-over suppressor in chromosome 2 effective in the presence of heterologous inversions. DIS 30: 145.
Perry, T., J. A. McKenzie, and P. Batterham, 2007 A Drosophila melanogaster confers a high level of resistance to spinosad. Insect Biochem. Mol. Biol. 37(2): 184–188. https://doi.org/10.1016/j.ibmb.2006.11.009
Rozen, S., and H. Skalskay, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132: 365–386.
Sturtevant, A. H., 1920 A crossover reducer in Drosophila melanogaster due to inversion of a section of the third chromosome. Biol. Zent. Bl. 46: 697–702.
Sturtevant, A. H., 1931 Known and probable inverted sections of the autosomes of Drosophila melanogaster. In: Contributions to the genetics of certain chromosome anomalies in Drosophila melanogaster, Carnegie Institution of Washington, Washington, DC. 421: 1–27.
Ward, L., 1923 The genetics of Curly wing in Drosophila. Another case of balanced lethal factors. Genetics 8: 276–300.
Warren, W. D., S. Palmer, and A. J. Howells, 1996 Molecular characterization of the cinnabar region of Drosophila melanogaster: Identification of the cinnabar transcription unit. Genetics 98(3): 249–262. https://doi.org/10.1007/BF00057589
Watson, G. B., S. W. Chouinard, K. R. Cook, C. Geng, J. M. Gifford et al., 2010 A spinosyn-sensitive Drosophila melanogaster nicotinic acetylcholine receptor identified through chemically induced target site resistance, resistance gene identification, and heterologous expression. Insect Biochem. Mol. Biol. 40(5): 376–384. https://doi.org/10.1016/j.ibmb.2009.11.004
Communicating editor: B. Reed