TITLE : High levels of intra-strain structural variation in *Drosophila simulans* X pericentric heterochromatin

AUTHORS : Cécile Courret[1] and Amanda M. Larracuente[1]

AFFILIATION :

1. Department of Biology, University of Rochester, Rochester, 14627, NY, USA

ORCID:

0000-0001-5849-8014 (CC)

0000-0001-5944-5686 (AML)

SHORT RUNNING HEAD: Pericentric rearrangements in Drosophila

Keywords : structural variation – pericentromeric heterochromatin – satellite repeats - Drosophila

ABSTRACT

Large genome structural variations can impact genome regulation and integrity. Repeat-rich regions like pericentric heterochromatin are vulnerable to structural rearrangements although we know little about how often these rearrangements occur over evolutionary time. Repetitive genome regions are particularly difficult to study with genomic approaches, as they are missing from most genome assemblies. However cytogenetic approaches offer a direct way to detect large rearrangements involving pericentric heterochromatin. Here we use a cytogenetic approach to reveal large structural rearrangements associated with the X pericentromeric region of *Drosophila simulans.* These rearrangements involve large blocks of satellite DNA—the *500-bp* and *Rsp-like* satellites—which colocalize in the X pericentromeric heterochromatin. We find that this region is polymorphic not only among different strains, but between isolates of the same strain from different labs, and even within individual isolates. On one hand, our observation raises questions regarding the potential impact of such variation at the phenotypic level and our ability to control for such genetic variability. On the other hand, this highlights the very rapid turnover of the pericentric heterochromatin most likely associated with genomic instability of the X pericentromere. It represents a unique opportunity to study the dynamics of pericentric heterochromatin, the evolution of associated satellites at a very short time scale, and to better understand how structural variation arises.

INTRODUCTION

Structural variants are duplicated, deleted, transposed, or inverted sequences, that can contribute to complex traits (Sudmant *et al.* 2015; Chakraborty *et al.* 2019), diseases (Stankiewicz and Lupski 2010), and genome evolution (Chakraborty *et al.* 2021). Variants involving rearrangements of large genome regions, such as chromosomal translocations and inversions, are associated with diseases involving intellectual disabilities and cancers (Weischenfeldt *et al.* 2013). Pericentric heterochromatin is rich in repetitive sequences like transposable elements and satellite DNAs (Charlesworth *et al.* 1986) and may be particularly prone to structural rearrangements from replication stress, non-homologous recombination, transposable element activity, and a decreased efficiency of some DNA repair pathways (reviewed (Janssen *et al.* 2018)). Structural rearrangements in pericentric heterochromatin may have consequences: although the density of conventional protein-coding genes is low, these regions have roles in genome defense (Andersen *et al.* 2017), coordinating chromosome segregation and nuclear organization (Folco *et al.* 2008; Peng and Karpen 2009) and genomic stability (Janssen *et al.* 2018).

Repeats in the pericentric heterochromatin are highly dynamic over long evolutionary time periods (Lohe and Roberts 1988), as species tend to have their own unique profiles of pericentric repeats. This divergence in the pericentric heterochromatin can lead to genetic incompatibilities between closely related species (Ferree and Barbash 2009; Cattani *et al.* 2012; Jagannathan *et al.* 2017; Jagannathan and Yamashita 2021). We know less about the dynamics of pericentric heterochromatin and its functional consequences over short evolutionary timescales, although satellite DNA copy number varies within species (*e.g.*, (Wei *et al.* 2014)) and can be associated with chromosome rearrangements (Flynn *et al.* 2023). However, some functional variation within

species maps to highly heterochromatic regions of the genome. For example, variation in Y-linked heterochromatin can impact gene expression across the genome and affect male fertility (Dimitri and Pisano 1989; Chippindale and Rice 2001; Lemos *et al.* 2008; Sackton *et al.* 2011; Brown *et al.* 2020).

The repetitive nature of pericentric heterochromatin makes it difficult to study at the genomic level (Treangen and Salzberg 2012), although the relatively compact genome of Drosophila species make them mighty models for repeat biology. Drosophila species have a large genetic toolkit and many Drosophila species can be isogenized and inbred, making the genome homozygous and amenable to experiments (Hoskins *et al.* 2015; Hales *et al.* 2015). High quality genome assemblies exist for species of the *melanogaster* clade: *D. melanogaster* (Chang and Larracuente 2019), *D. simulans*, *D. mauritiana,* and *D. sechellia* (Chakraborty *et al.* 2021; Chang *et al.* 2022). Comparing these assemblies revealed structural divergence between species that may contribute to important phenotypes. Structural rearrangements involving pericentric heterochromatin are difficult to ascertain with genomic approaches—the most densely repetitive regions of the genome including large blocks of tandem satellite repeats are not yet fully assembled (Chakraborty *et al.* 2021; Chang *et al.* 2022). However, cytogenetic approaches indicate that the distribution and type of heterochromatic satellite repeats differs even between these closely related species (Larracuente 2014; Jagannathan *et al.* 2017; Sproul *et al.* 2020), implying that large structural variations contribute to species divergence. Large structural rearrangements in pericentromeric satellite repeats within species are less well documented.

Here we describe striking structural variation in the pericentric heterochromatin of the *X* chromosome in *Drosophila simulans.* We use a cytogenetic approach to document high levels of

structural polymorphism in satellite DNAs in the X pericentromere: *Rsp-like* and *500-bp* satellite. *Rsp-like* is a complex satellite specific to the X pericentromere in *D. simulans* (Sproul *et al.* 2020) and the *500-bp* satellite is associated with the centromere and pericentromere of the *X* chromosome and the autosomes in *D. simulans* (Talbert *et al.* 2018; Courret *et al.* 2023b). The structural polymorphisms we detect involve large blocks of satellite repeats and occur between different strains, within a strain, and even within individual isolates of strains kept in a single lab. This extreme structural polymorphism may not be conspicuous at the DNA sequencing level, but affects large regions of the pericentromere, and could conceivably have functional impacts.

MATERIALS AND METHODS

Fly strains

We use 'strain' to refer to a genotype and give a unique name (*i.e.,* appending lab initials) to 'isolates', which are lineages of a strain from a particular lab. We have three isolates of the w^{501} strains that originated from three different laboratories: Larracuente (*w501-i1*), Presgraves (*w501-i2*), and Andolfatto (*w501-i3*). *w501-i1* and *w501-i2* have a common origin,but have been maintained separately for 7 years. We have two isolates of the w^{XD1} strain that originated from two different labs: Presgraves (*wXD1-i1*) and Meiklejohn (*wXD1-i2*). The *wXD1-i2* isolate originated from the *wXD1-i1* isolate ~10 years ago. We also used other non-white isofemale *D. simulans* strains: *SR (*collected from *Seychelles* in 1981*)*, *ST8 (*collected from *Tunisia* in 1983*)*, *C167.4 (*collected from *Kenya* in 1973*)*, *sim006 (*collected from *California* in 1961*)* (described in (Courret *et al.* 2023a).

Fluorescence in situ hybridization

The FISH was performed using primary oligopaint probes for *Rsp-like* and *500-bp* (Courret *et al.* 2023b) coupled with sec6 and sec5 adaptors (Beliveau *et al.* 2014). Sec5 is coupled with Cy5 while sec6 is coupled with Cy3. We dissected brains from third instar larvae in PBS, incubated 8min in 0.5% Sodium citrate. We fixed for 6 min in 4% formaldehyde, 45% acetic acid before squashing. We squashed the brains between the slide and coverslip and before immersing in liquid nitrogen. After 10 min in 100% ethanol, we air dried slides for at least one hour before proceeding to the hybridization. For the hybridization, we used 20 pmol of primary probes and 80 pmol of the secondary probes in 50 ul of hybridization buffer (50% formamide, 10% dextran sulfate, 2xSSC). We heated slides for 5 min at 95°C to denature and incubated them overnight at 37°C in a humid

chamber. We then washed the slides 3 times for 5 min with 4XSSCT and 3 times for 5min with 0.1SSC before mounting in slowfade DAPI. We imaged using a LEICA DM5500 microscope and cropped and pseudocolored the images using Fiji.

We analyzed 4-10 mitotic spreads for each individual brain, to determine without ambiguity the number of foci carried by the *X* chromosomes. We confirmed that all spreads within an individual brain had the same number of foci. To estimate the allele frequency in each isolate, around 20 individual brains were tested, both male and female (full genotype details in SupTable1). The frequency reported in Table 1 corresponds to the frequency of each type of *X* chromosome among all individual brains tested.

Genotyping and genome analysis

We designed primers around SNPs located on the *X* chromosome. The primer position - alleles on Segkk236 from the reference genome in (Chang *et al.* 2022) and sequences are: 9814904 - T/G (forward primer - GCAAAGTCTTTTAAGCGCGC and reverse primer-CCGGGGGAAAATCTGCTTCT); 17904265 - A/G (forward primer - GTTGTCGCTCTCCTTGACCA and reverse primer-GCTGGCCATCTTCACCATCT); and 18025547 - C/T (forward primer - CTGCTCCGCGTGTATATGGT and reverse primer-ACAGTTCGCGATGAGCTTCT). For each primer pair, we performed a PCR with NEB Taq polymerase (NEB #M0495) following the manufacturer's instructions (hybridization temperature: 53°). We sequenced each PCR product using the Sanger method (ACGT company) and visualized sequence profiles using Geneious.

We downloaded reads for w^{XD1} (SRR8247551; (Meiklejohn *et al.* 2018), *ST8*, *SR*, and *C167.4* (PRJNA905841; (Courret *et al.* 2023a)) and w^{501} (SRR520334 ; (Hu *et al.* 2013)), trimmed and

processed reads with trimgalore (v0.6.2) (Krueger *et al.* 2021) (--paired --nextera --length 75 --phred33 --fastqc). We mapped reads with *BWA-MEM* (v0.17 default parameters) to the *D. simulans* genome assembly (Chang *et al.* 2022) and estimated coverage (in reads per million) with bamCoverage (-bs 1000) in deeptools (v3.5.1) (Ramírez *et al.* 2016) across the *X* chromosome. We plotted in R to look for large-scale differences in coverage that would suggest structural polymorphisms.

To estimate the per-site heterozygosity, we called SNPs using bcftools (v1.6) (Li 2011) *mpileup* and *call* commands, keeping all sites. We filtered the vcf file using vcftools (v0.1.15/b1) (Danecek *et al.* 2011) (--remove-indels --minQ 30 --minDP 10 --maxDP 200) and then extracted the number of homozygous and heterozygous sites using the bcftools *stats* command.

RESULTS

We focus our study on two commonly used *D. simulans* lab strains: w^{501} and w^{XD1}. Both carry a *white* mutation on the *X* chromosome, conferring the white-eyed phenotype. These inbred strains are frequently used for genetic manipulation (Stern *et al.* 2017) or genetic mapping (Matute and Ayroles 2014; Meiklejohn *et al.* 2018) and have abundant genomic resources (Garrigan *et al.* 2012; Hu *et al.* 2013; Chakraborty *et al.* 2021; Chang *et al.* 2022).

We collected isolates of the w^{501} strain from three different laboratories, denoted with initials (*w501-i1; w501-i2*, and *w501-i3*). *w501-i1* and *w501-i2* have a common origin but have been maintained separately for 7 years (91-119 generations). The *w501-i3* was maintained independently. We also collected isolates of the w^{XD1} strains from two different labs: *wXD1-i1* and *wXD1-i2*. The *wXD1-i2* originated from the *wXD1-i1* strains 10 years ago (130-170 generations).

The two satellites that we use as markers for pericentric structural variation, *500-bp* and *Rsp-like,* are adjacent on the *X* chromosome and their localization pattern is always similar (*i.e.*, in adjacent blocks). We did not observe any genotypes where *500-bp* and *Rsp-like* did not co-vary in the number of foci. We show that these blocks are highly variable both within and between strains. We observe three general colocalization patterns for *500-bp* and *Rsp-like* at 1, 2 or 3 foci in the *X* pericentric heterochromatin.

Structural variation within and between isolates of a single strain

The three isolates of the w^{501} strain appear to be polymorphic both between and within isolates. The *w501-i1* and *w501-i3* isolate are polymorphic for two and three-focus *X* chromosomes (Figure 1A and C). Within the *w501-i1* isolate we estimated the frequency of the three-locus and two-locus

178 *X* chromosomes at 66% and 34%, respectively (Table1). While the *w501-i3* has estimated
179 frequencies of 93% and 7%, respectively (Table 1). *w501-i2* shows both two and one-focus *X*
180 chromosomes (Figure 1B), at estimated frequencies of 79% and 21%, respectively (Table 1).
181
182 This degree of polymorphism and divergence within a single strain is surprising as the *w501-i1*
183 isolate originated from the *w501-i2* isolate only 7 years ago (91-119 generations). This suggests
184 that duplication events in the pericentromeric region happened recently and may happen
185 recurrently.
186
187 We observe similarly striking structural variation in the pericentromeric region of the w^{XD1} *X*
188 chromosomes. Consistent with previous observations (Sproul *et al.* 2020), we find that the *wXD1-*
189 *i1 X* chromosome pericentromere has a three-focus pattern (Figure 2A). However, the *wXD1-i2 X*
190 chromosome pericentromeric region appears to be polymorphic for the one-focus and three-focus
191 patterns (Figure 2B), with estimated frequencies of 14% and 86%, respectively (Table1).
192
193 Structural polymorphisms involving large blocks of the *Rsp-like* and *500-bp* satellite repeats may
194 generally be detectable through differences in read depth (Larracuente 2014). However, when these
195 polymorphisms exist within a single isolate, they are not obvious in genomic data (Supplemental
196 Figure 1). In our analysis of sequencing libraries created from pooled individuals, detecting
197 alternative alleles based on read depth is extremely challenging, as it will depend on the frequency
198 of alternative alleles in the pools. Biases in library preparation, tissue, and DNA extraction can all
199 contribute to variation in read mapping in repetitive sequences between biological replicates
200 (Shinde 2003; Aird *et al.* 2011; Ross *et al.* 2013; Wei *et al.* 2018). We suggest that true structural
201 polymorphisms, either between individuals of a single isolate or between tissue and cells within an

individual, can also contribute to variable read coverage. We would need multiple biological replicates from the same isolates and, ideally, a contiguous assembly of pericentric heterochromatin to assess the potential for recovering information about these structural rearrangements in genomic data. Currently, a cytogenetic approach is necessary to characterize such structural polymorphisms, especially within isolates.

These white-eyed lab strains have independent origins and therefore these structural mutations should also be independent. To be sure that the structural variation is not due to strain contamination and/or recombination between the two white-eyed lab strains, we genotyped the *X* chromosomes. We designed primers to genotype three SNPs that allow us to differentiate w^{XD1} and w^{501} *X* chromosomes by PCR re-sequencing. As expected, if pericentromeric variation is due to structural polymorphisms within an *X* chromosome, the different w^{501} isolates carry the same alleles and the w^{XD1} isolates carry the same alternative alleles at all three sites. This suggests that the structural variants arose on their respective *X* chromosome backgrounds and that the *X* pericentric heterochromatin is likely unstable in these white-eyed lab strains.

Within-isolate structural variation seems limited to lab strains

To understand if the chromosomal instability is strain or species specific, we studied satellite organization in four different *D. simulans* strains that do not carry *white* mutations: *SR, ST8, sim006,* and *C167.4*. Each of these strains has a single focus of *Rsp-like* and *500-bp* in their X pericentric heterochromatin (Figure 3). While more strains should be tested in the future, this pattern suggests that the large structural variations may be limited to the w^{501} and w^{XD1} strains.

Isogenization should purge any segregating sequence variants (including structural ones) within strains, although sequence variation may exist due to: *1)* mutations that accumulate over time while strains are maintained in labs (Lack *et al.* 2016); 2.) residual heterozygosity from incomplete inbreeding or linkage to balanced deleterious mutations that cannot be made homozygous. To determine if the structural polymorphism correlates with the extent of inbreeding of each strain, we estimate the per-site heterozygosity (*H*) of the *X* chromosome in available genomic data (Hu *et al.* 2013; Meiklejohn *et al.* 2018; Courret *et al.* 2023a). Despite being polymorphic in the *X* pericentromere, we estimate very low levels of per-site heterozygosity across the *X* chromosome arm in *wXD1-i2* (H=1.254x10^{-5}) and *w501-i3* (H=5.93x10^{-5}). The non-white strains appear less inbred–*ST8* (H=0.000468), *SR* (H=0.000733) and *C167.4* (H=0.000459), which are similar to a previous estimate for the *sim006* strain (H=0.00039) (Kim *et al.* 2021).

Therefore, the structural polymorphism is in the strains with the lowest heterozygosity across the *X* chromosome arm, further supporting our hypothesis that the structural variants arose recently and may be associated with genomic instability in the *X* pericentromere.

DISCUSSION

In summary, we find large X-linked structural polymorphisms segregating within single isolates of two commonly used lab strains of *D. simulans.* These types of polymorphisms are not obvious in genomic data, although they may contribute to variation in read depth between biological replicates in repetitive regions. Because we observe different variants even within single isolates of the same strain (*i.e.,* within single vials of flies), we hypothesize that this region of the *X* pericentromere is unstable and associated with recurrent structural rearrangements. We cannot completely rule out the possibility that these variants were already segregating in the original strains and then sorted differently between lab isolates. Labs may differ in their maintenance conditions, which may impose different selection pressures. Different isolates of the same strain maintained in different labs can accumulate isolate-specific TE landscapes (Rahman *et al.* 2015). Further experiments are necessary to determine the mutation rate in the *X* pericentromere. A recent origin for these structural variants appears more likely based on multiple observations. First, if there was a pre-existing variation we would expect more similarity between the *w501-i1* and *w501-i2* isolates, based on their recent history, than between *w501-i1* and *w501-i3*. Second, two independent strains (w^{501} and w^{XD1}) exhibit structural polymorphism in the same region, suggesting that this X pericentric heterochromatin may experience genomic instability. Finally, the two white strains where we see the variation are highly inbred compared to the four non-white strains which do not have detectable structural polymorphisms.

The structural variation we observe may have functional implications, as pericentric heterochromatin has effects on chromosome dynamics (Dernburg *et al.* 1996; Karpen *et al.* 1996), genome stability (Peng and Karpen 2009), genome structure (Falk *et al.* 2019; Lee *et al.* 2020), and nuclear organization . These regions also contain, or flank, essential genetic elements, including

the centromeres. For example, variation in pericentromeres may affect adjacent centromeres (Kumon *et al.* 2021; Jagannathan and Yamashita 2021), chromosome structures that are essential for coordinating chromosome segregation during cell divisions (Allshire and Karpen 2008). In most species, the *rDNA* are also embedded in heterochromatin (McStay 2016) and in Drosophila species, the *rDNA* locus is generally located in the X pericentromere (Stage and Eickbush 2007). Variation in *rDNA* copy number is associated with reduced translation capacity in *D. melanogaster* (Mohan and Ritossa 1970; Terracol and Prud'homme 1986). Pericentric heterochromatin may also contain piRNA clusters—discrete loci rich in fragments of transposable elements and other repeats that generate precursors for the small RNAs that are important for the silencing of transposable element activity all over the genome (Brennecke *et al.* 2007; Aravin *et al.* 2008). Complex satellite DNAs like those involved in these rearrangements also generate piRNAs that may play a role in establishing heterochromatin in the early embryo (Wei *et al.* 2021). Finally, while gene density in heterochromatin is generally low, species like *D. melanogaster* do contain hundreds of protein coding genes (Marsano *et al.* 2019) some of which are essential (Devlin *et al.* 1990; Gatti and Pimpinelli 1992). For some of these genes, a heterochromatic environment is essential for their proper expression and structural rearrangements can disrupt their function (Wakimoto and Hearn 1990; Eberl *et al.* 1993) and the function of nearby euchromatic genes (Elgin and Reuter 2013).

Structural variation in pericentric heterochromatin can also have global effects on genome stability and regulation. Large blocks of heterochromatin can act as a sink for heterochromatin proteins, titrating them away from other genomic locations (Tartof *et al.* 1984; Dimitri and Pisano 1989; Eissenberg *et al.* 1990; Wallrath and Elgin 1995; Francisco and Lemos 2014; Brown *et al.* 2020). One potential consequence of this sink effect is through its impact on the transcription of

euchromatin genes and transposable elements, both which may ultimately impact individual fitness (Francisco and Lemos 2014; Abramov *et al.* 2016; Nguyen and Bachtrog 2021; Huang *et al.* 2022).

On one hand, our observation is concerning. Having different variants of the pericentric heterochromatin segregating in a single isolate might introduce both genetic and phenotypic variation to experiments. It also raises the question of the reproducibility of the results between laboratories. It is important to keep track of, and report, the origin of each isolate. Because the variation we described here is not easy to assay and thus difficult to control for, we recommend limiting potential variation within isolates by periodically re-isogenizing strains. We caution researchers to consider the impact this structural variation may have on their experiments.

On the other hand, this is an intriguing observation. While we expect structural rearrangements in heterochromatic sequences within and between species, these X pericentromeres we study here are highly dynamic even within a single isolates of inbred *D. simulans* strains. Our observations raise several questions. Why is this region particularly unstable? Is this instability specific to the X pericentromere? Is it specific to *D. simulans*? Further investigation will be necessary to better understand the dynamics of structural variation in pericentric heterochromatin and its consequences. The structural rearrangements we describe here are likely associated with genome instability and may represent a unique opportunity to better understand factors promoting the disruption of heterochromatin structure in general. The mechanisms involved in generating these structural rearrangements may be similar to those associated with structural variations involved in human diseases.

DATA AVAILABILITY

Strains are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

ACKNOWLEDGEMENTS

We are grateful to Colin Meiklejohn, Daven Presgraves, Peter Andolfatto, and Catherine Montchamp-Moreau for generously sharing fly stocks and to Grace Lee and members of the Larracuente lab for discussion.

LITERATURE CITED

Abramov Y. A., A. S. Shatskikh, O. G. Maksimenko, S. Bonaccorsi, V. A. Gvozdev, *et al.*, 2016 The differences between *Cis* - and *Trans* -gene inactivation caused by heterochromatin in *Drosophila*. Genetics 202: 93–106. https://doi.org/10.1534/genetics.115.181693

Aird D., M. G. Ross, W.-S. Chen, M. Danielsson, T. Fennell, *et al.*, 2011 Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12: R18. https://doi.org/10.1186/gb-2011-12-2-r18

Allshire R. C., and G. H. Karpen, 2008 Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat. Rev. Genet. 9: 923–937. https://doi.org/10.1038/nrg2466

Andersen P. R., L. Tirian, M. Vunjak, and J. Brennecke, 2017 A heterochromatin-dependent transcription machinery drives piRNA expression. Nature 549: 54–59. https://doi.org/10.1038/nature23482

Aravin A. A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, *et al.*, 2008 A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell 31: 785–799. https://doi.org/10.1016/j.molcel.2008.09.003

Beliveau B. J., N. Apostolopoulos, and C. Wu, 2014 Visualizing genomes with oligopaint FISH probes. Curr. Protoc. Mol. Biol. 105. https://doi.org/10.1002/0471142727.mb1423s105

Brennecke J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128: 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043

Brown E. J., A. H. Nguyen, and D. Bachtrog, 2020 The Drosophila Y chromosome affects heterochromatin integrity genome-wide, (J. Parsch, Ed.). Mol. Biol. Evol. 37: 2808–2824. https://doi.org/10.1093/molbev/msaa082

Cattani M. V., S. B. Kingan, and D. C. Presgraves, 2012 *CIS* - and *TRANS* -acting genetic factors contribute to heterogeneity in the rate of crossing over between the *Drosophila simulans* clade species. J. Evol. Biol. 25: 2014–2022. https://doi.org/10.1111/j.1420-9101.2012.02578.x

Chakraborty M., J. J. Emerson, S. J. Macdonald, and A. D. Long, 2019 Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. Nat. Commun. 10: 4872. https://doi.org/10.1038/s41467-019-12884-1

Chakraborty M., C.-H. Chang, D. E. Khost, J. Vedanayagam, J. R. Adrion, *et al.*, 2021 Evolution of genome structure in the *Drosophila simulans* species complex. Genome Res. 31: 380–396. https://doi.org/10.1101/gr.263442.120

Chang C.-H., and A. M. Larracuente, 2019 Heterochromatin-enriched assemblies reveal the sequence and organization of the *Drosophila melanogaster* Y chromosome. Genetics 211: 333–348. https://doi.org/10.1534/genetics.118.301765

Chang C.-H., L. E. Gregory, K. E. Gordon, C. D. Meiklejohn, and A. M. Larracuente, 2022 Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes. eLife 11: e75795. https://doi.org/10.7554/eLife.75795

Charlesworth B., C. H. Langley, and W. Stephan, 1986 The evolution of restricted recombination and the accumulation of repeated dna sequences. Genetics 112: 947–962. https://doi.org/10.1093/genetics/112.4.947

Chippindale A. K., and W. R. Rice, 2001 Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. 98: 5677–5682. https://doi.org/10.1073/pnas.101456898

Courret C., D. Ogereau, C. Gilbert, A. M. Larracuente, and C. Montchamp-Moreau, 2023a The evolutionary history of *Drosophila simulans* Y chromosomes reveals molecular signatures of resistance to sex ratio meiotic drive, (H. Malik, Ed.). Mol. Biol. Evol. 40: msad152. https://doi.org/10.1093/molbev/msad152

Courret C., L. Hemmer, X. Wei, P. D. Patel, B. Santinello, *et al.*, 2023b Rapid turnover of centromeric DNA reveals signatures of genetic conflict in Drosophila. Evolutionary Biology.

Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dernburg A. F., J. W. Sedat, and R. S. Hawley, 1996 Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell 86: 135–146. https://doi.org/10.1016/S0092-8674(00)80084-7

Devlin R. H., B. Bingham, and B. T. Wakimoto, 1990 The organization and expression of the light gene, a heterochromatic gene of *Drosophila melanogaster*. Genetics 125: 129–140. https://doi.org/10.1093/genetics/125.1.129

Dimitri P., and C. Pisano, 1989 Position effect variegation in *Drosophila melanogaster*: relationship between suppression effect and the amount of Y chromosome. Genetics 122: 793–800. https://doi.org/10.1093/genetics/122.4.793

Eberl D. F., B. J. Duyf, and A. J. Hilliker, 1993 The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of *Drosophila melanogaster.* Genetics 134: 277–292. https://doi.org/10.1093/genetics/134.1.277

Eissenberg J. C., T. C. James, D. M. Foster-Hartnett, T. Hartnett, V. Ngan, *et al.*, 1990 Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. 87: 9923–9927. https://doi.org/10.1073/pnas.87.24.9923

Elgin S. C. R., and G. Reuter, 2013 Position-Effect Variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb. Perspect. Biol. 5: a017780–a017780. https://doi.org/10.1101/cshperspect.a017780

Falk M., Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, *et al.*, 2019 Heterochromatin

drives compartmentalization of inverted and conventional nuclei. Nature 570: 395–399. https://doi.org/10.1038/s41586-019-1275-3

Ferree P. M., and D. A. Barbash, 2009 Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila, (M. A. F. Noor, Ed.). PLoS Biol. 7: e1000234. https://doi.org/10.1371/journal.pbio.1000234

Flynn J. M., K. B. Hu, and A. G. Clark, 2023 Three recent sex chromosome-to-autosome fusions in a *Drosophila virilis* strain with high satellite DNA content, (J. Sekelsky, Ed.). GENETICS 224: iyad062. https://doi.org/10.1093/genetics/iyad062

Folco H. D., A. L. Pidoux, T. Urano, and R. C. Allshire, 2008 Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science 319: 94–97. https://doi.org/10.1126/science.1150944

Francisco F. O., and B. Lemos, 2014 How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. J. Genomics 2: 94–103. https://doi.org/10.7150/jgen.8043

Garrigan D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark, *et al.*, 2012 Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res. 22: 1499–1511. https://doi.org/10.1101/gr.130922.111

Gatti M., and S. Pimpinelli, 1992 Functional elements in *Drosophila melanogaster* heterochromatin. Annu. Rev. Genet. 26: 239–276. https://doi.org/10.1146/annurev.ge.26.120192.001323

Hales K. G., C. A. Korey, A. M. Larracuente, and D. M. Roberts, 2015 Genetics on the Fly: A Primer on the *Drosophila* Model System. Genetics 201: 815–842. https://doi.org/10.1534/genetics.115.183392

Hoskins R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez, *et al.*, 2015 The Release 6 reference sequence of the *Drosophila melanogaster* genome. Genome Res. 25: 445–458. https://doi.org/10.1101/gr.185579.114

Hu T. T., M. B. Eisen, K. R. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. Genome Res. 23: 89–98. https://doi.org/10.1101/gr.141689.112

Huang Y., H. Shukla, and Y. C. G. Lee, 2022 Species-specific chromatin landscape determines how transposable elements shape genome evolution. eLife 11: e81567. https://doi.org/10.7554/eLife.81567

Jagannathan M., N. Warsinger-Pepe, G. J. Watase, and Y. M. Yamashita, 2017 Comparative Analysis of Satellite DNA in the *Drosophila melanogaster* Species Complex. G3 GenesGenomesGenetics 7: 693–704. https://doi.org/10.1534/g3.116.035352

Jagannathan M., and Y. M. Yamashita, 2021 Defective satellite DNA clustering into chromocenters underlies hybrid incompatibility in *Drosophila*, (A. Larracuente, Ed.). Mol. Biol.

434 Evol. 38: 4977–4986. https://doi.org/10.1093/molbev/msab221

435 Janssen A., S. U. Colmenares, and G. H. Karpen, 2018 Heterochromatin: guardian of the genome.
436 Annu. Rev. Cell Dev. Biol. 34: 265–288. https://doi.org/10.1146/annurev-cellbio-100617-062653

437 Karpen G. H., M.-H. Le, and H. Le, 1996 Centric heterochromatin and the efficiency of
438 achiasmate disjunction in *Drosophila* female meiosis. Science 273: 118–122.
439 https://doi.org/10.1126/science.273.5271.118

440 Kim B. Y., J. R. Wang, D. E. Miller, O. Barmina, E. Delaney, *et al.*, 2021 Highly contiguous
441 assemblies of 101 drosophilid genomes. eLife 10: e66405. https://doi.org/10.7554/eLife.66405

442 Krueger F., F. James, P. Ewels, E. Afyounian, and B. Schuster-Boeckler, 2021
443 FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo

444 Kumon T., J. Ma, R. B. Akins, D. Stefanik, C. E. Nordgren, *et al.*, 2021 Parallel pathways for
445 recruiting effector proteins determine centromere drive and suppression. Cell 184: 4904-
446 4918.e11. https://doi.org/10.1016/j.cell.2021.07.037

447 Lack J. B., J. D. Lange, A. D. Tang, R. B. Corbett-Detig, and J. E. Pool, 2016 A thousand fly
448 genomes: An expanded *Drosophila* genome nexus. Mol. Biol. Evol. 33: 3308–3313.
449 https://doi.org/10.1093/molbev/msw195

450 Larracuente A. M., 2014 The organization and evolution of the responder satellite in species of
451 the *Drosophila melanogaster* group: dynamic evolution of a target of meiotic drive. BMC Evol.
452 Biol. 14: 233. https://doi.org/10.1186/s12862-014-0233-9

453 Lee Y. C. G., Y. Ogiyama, N. M. C. Martins, B. J. Beliveau, D. Acevedo, *et al.*, 2020
454 Pericentromeric heterochromatin is hierarchically organized and spatially contacts H3K9me2
455 islands in euchromatin, (G. Bosco, Ed.). PLOS Genet. 16: e1008673.
456 https://doi.org/10.1371/journal.pgen.1008673

457 Lemos B., L. O. Araripe, and D. L. Hartl, 2008 Polymorphic Y chromosomes harbor cryptic
458 variation with manifold functional consequences. Science 319: 91–93.
459 https://doi.org/10.1126/science.1148861

460 Li H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping
461 and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–
462 2993. https://doi.org/10.1093/bioinformatics/btr509

463 Lohe A., and P. A. Roberts, 1988 Evolution of satellite DNA sequences in Drosophila., pp. 148–
464 186 in *Heterochromatin, Molecular and Structural Aspects*, Cambridge University Press,
465 Cambridge.

466 Marsano R. M., E. Giordano, G. Messina, and P. Dimitri, 2019 A new portrait of constitutive
467 heterochromatin: lessons from *Drosophila melanogaster*. Trends Genet. 35: 615–631.
468 https://doi.org/10.1016/j.tig.2019.06.002

469 Matute D. R., and J. F. Ayroles, 2014 Hybridization occurs between *Drosophila simulans* and *D.*

sechellia in the Seychelles archipelago. J. Evol. Biol. 27: 1057–1068. https://doi.org/10.1111/jeb.12391

McStay B., 2016 Nucleolar organizer regions: genomic 'dark matter' requiring illumination. Genes Dev. 30: 1598–1610. https://doi.org/10.1101/gad.283838.116

Meiklejohn C. D., E. L. Landeen, K. E. Gordon, T. Rzatkiewicz, S. B. Kingan, *et al.*, 2018 Gene flow mediates the role of sex chromosome meiotic drive during complex speciation. eLife 7: e35468. https://doi.org/10.7554/eLife.35468

Mohan J., and F. M. Ritossa, 1970 Regulation of ribosomal RNA synthesis and its bearing on the bobbed phenotype in *Drosophila melanogaster*. Dev. Biol. 22: 495–512. https://doi.org/10.1016/0012-1606(70)90165-X

Nguyen A. H., and D. Bachtrog, 2021 Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome, (R. S. Hawley, Ed.). PLOS Genet. 17: e1009438. https://doi.org/10.1371/journal.pgen.1009438

Peng J. C., and G. H. Karpen, 2009 Heterochromatic genome stability requires regulators of Histone H3 K9 Methylation, (R. S. Hawley, Ed.). PLoS Genet. 5: e1000435. https://doi.org/10.1371/journal.pgen.1000435

Rahman R., G. Chirn, A. Kanodia, Y. A. Sytnikova, B. Brembs, *et al.*, 2015 Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. Nucleic Acids Res. 43: 10655–10672. https://doi.org/10.1093/nar/gkv1193

Ramírez F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, *et al.*, 2016 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44: W160–W165. https://doi.org/10.1093/nar/gkw257

Ross M. G., C. Russ, M. Costello, A. Hollinger, N. J. Lennon, *et al.*, 2013 Characterizing and measuring bias in sequence data. Genome Biol. 14: R51. https://doi.org/10.1186/gb-2013-14-5-r51

Sackton T. B., H. Montenegro, D. L. Hartl, and B. Lemos, 2011 Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in *Drosophila*. Proc. Natl. Acad. Sci. 108: 17046–17051. https://doi.org/10.1073/pnas.1114690108

Shinde D., 2003 Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Res. 31: 974–980. https://doi.org/10.1093/nar/gkg178

Sproul J. S., D. E. Khost, D. G. Eickbush, S. Negm, X. Wei, *et al.*, 2020 Dynamic evolution of euchromatic satellites on the X chromosome in *Drosophila melanogaster* and the *simulans clade*, (J. Parsch, Ed.). Mol. Biol. Evol. 37: 2241–2256. https://doi.org/10.1093/molbev/msaa078

Stage D. E., and T. H. Eickbush, 2007 Sequence variation within the rRNA gene loci of 12 *Drosophila* species. Genome Res. 17: 1888–1897. https://doi.org/10.1101/gr.6376807

Stankiewicz P., and J. R. Lupski, 2010 Structural variation in the human genome and its role in disease. Annu. Rev. Med. 61: 437–455. https://doi.org/10.1146/annurev-med-100708-204735

Stern D. L., J. Crocker, Y. Ding, N. Frankel, G. Kappes, *et al.*, 2017 Genetic and transgenic reagents for *Drosophila simulans*, *D. mauritiana* , *D. yakuba* , *D. santomea* , and *D. virilis*. G3 GenesGenomesGenetics 7: 1339–1347. https://doi.org/10.1534/g3.116.038885

Sudmant P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, *et al.*, 2015 An integrated map of structural variation in 2,504 human genomes. Nature 526: 75–81. https://doi.org/10.1038/nature15394

Talbert P. B., S. Kasinathan, and S. Henikoff, 2018 Simple and complex centromeric satellites in Drosophila sibling species. Genetics 208: 977–990. https://doi.org/10.1534/genetics.117.300620

Tartof K. D., C. Hobbs, and M. Jones, 1984 A structural basis for variegating position effects. Cell 37: 869–878. https://doi.org/10.1016/0092-8674(84)90422-7

Terracol R., and N. Prud'homme, 1986 Differential elimination of rDNA genes in bobbed mutants of *Drosophila melanogaster*. Mol. Cell. Biol. 6: 1023–1031. https://doi.org/10.1128/mcb.6.4.1023-1031.1986

Treangen T. J., and S. L. Salzberg, 2012 Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. 13: 36–46. https://doi.org/10.1038/nrg3117

Wakimoto B. T., and M. G. Hearn, 1990 The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. Genetics 125: 141–154. https://doi.org/10.1093/genetics/125.1.141

Wallrath L. L., and S. C. Elgin, 1995 Position effect variegation in Drosophila is associated with an altered chromatin structure. Genes Dev. 9: 1263–1277. https://doi.org/10.1101/gad.9.10.1263

Wei K. H.-C., J. K. Grenier, D. A. Barbash, and A. G. Clark, 2014 Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. 111: 18793–18798. https://doi.org/10.1073/pnas.1421951112

Wei K. H.-C., S. E. Lower, I. V. Caldas, T. J. S. Sless, D. A. Barbash, *et al.*, 2018 Variable rates of simple satellite gains across the Drosophila phylogeny. Mol. Biol. Evol. 35: 925–941. https://doi.org/10.1093/molbev/msy005

Wei X., D. G. Eickbush, I. Speece, and A. M. Larracuente, 2021 Heterochromatin-dependent transcription of satellite DNAs in the *Drosophila melanogaster* female germline. eLife 10: e62375. https://doi.org/10.7554/eLife.62375

Weischenfeldt J., O. Symmons, F. Spitz, and J. O. Korbel, 2013 Phenotypic impact of genomic structural variation: insights from and for human disease. Nat. Rev. Genet. 14: 125–138. https://doi.org/10.1038/nrg3373

FIGURES LEGEND

Figure 1: FISH on mitotic chromosomes from larval brain in (A) *w501-i1*, (B) *w501-i2*, (C) *w501-i3* strains. We used oligopaints probes targeting the *Rsp-like* (red) and *500-bp* (blue) satellites. The scale bar represents 5µm.The inset zooms in on the *X* chromosome revealing a heterozygote for 2-focus (1) and a 3-focus (2) *X* chromosome in *w501-i1* (A), a heterozygote for a 2-focus (1) and a 1-focus *X* chromosome (2) in *w501-i2* (B) and a heterozygote for a 3-focus (1) and a 2-focus *X* chromosome (2) in *w501-i3*. The arrows within the inset point to each foci associated with the *X* chromosome.

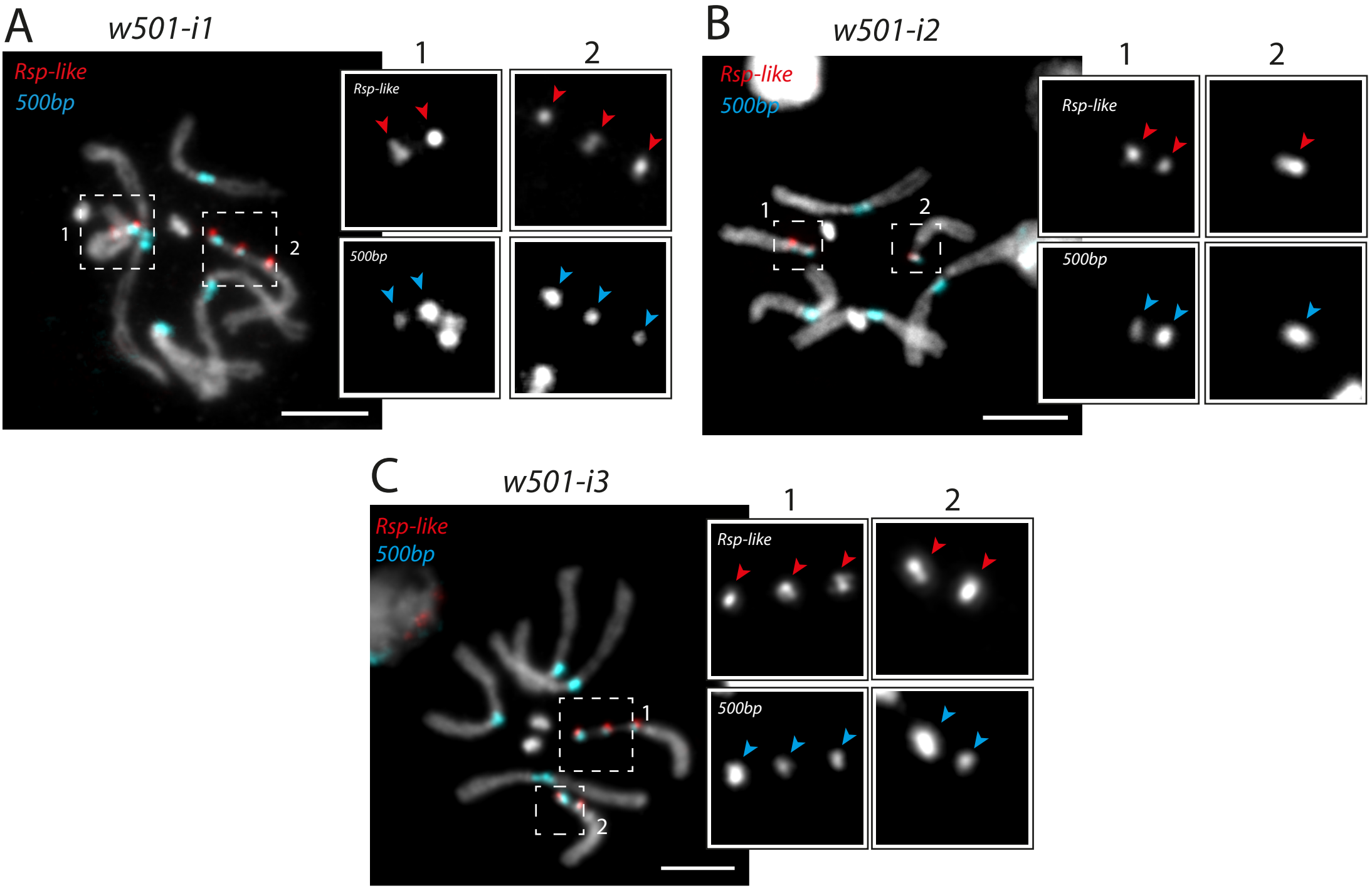
Figure 2: FISH on mitotic chromosomes from larval brains of (A) *wXD1-i1* and (B) *wXD1-i2* strains. We used oligopaint probes targeting the *Rsp-like* (red) and *500-bp* (blue) satellites. The scale bar represents 5µm. The inset zooms in on the *X* chromosome revealing a 3-focus *X* chromosome in *wXD1-i1* (A) and a heterozygote for a 1-focus (1) and a 3-focus *X* chromosome (2) in *wXD1-i2* (B). The arrows within the inset point to each focus associated with the *X* chromosome.

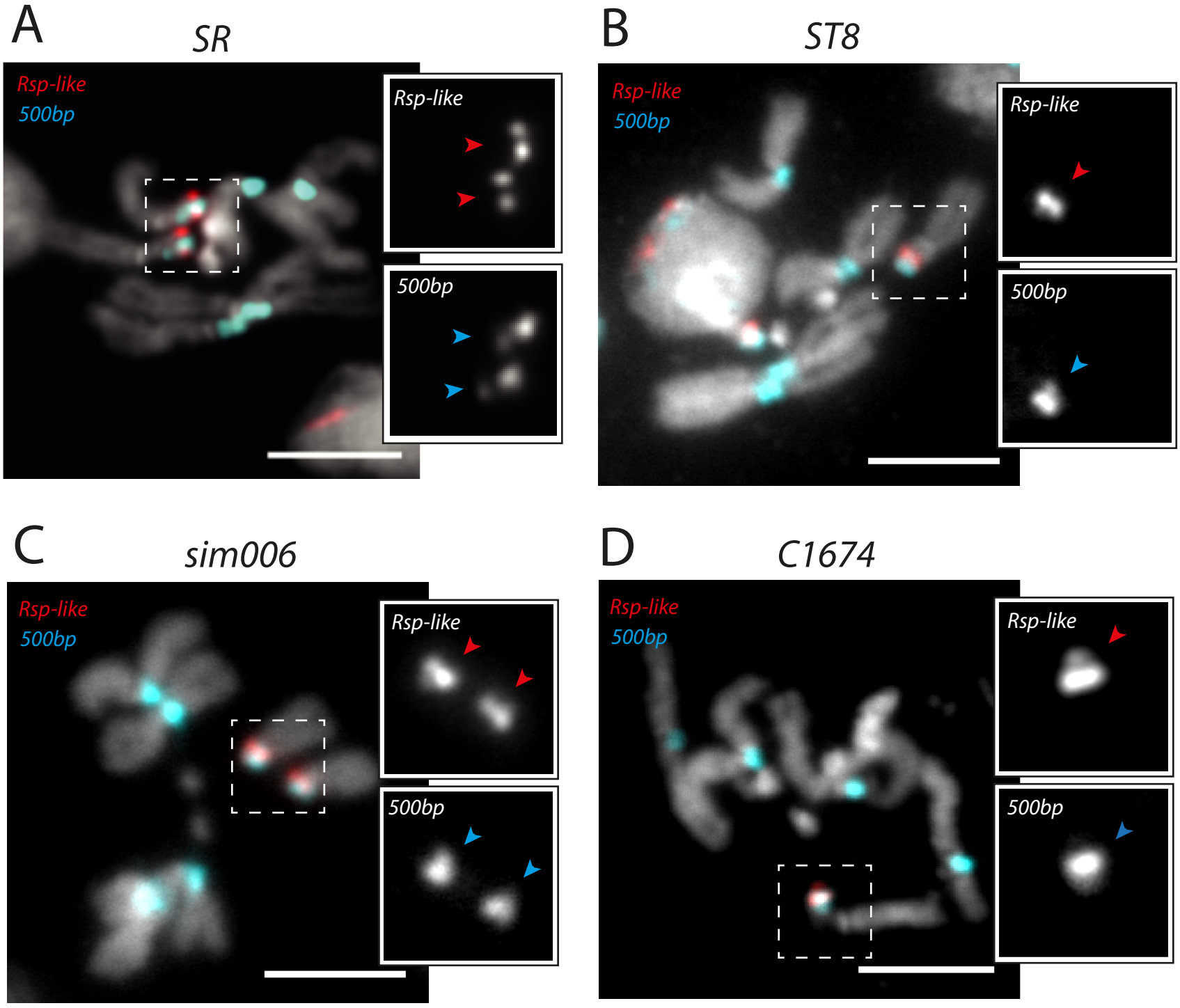
Figure 3: FISH on mitotic chromosomes from larval brains of (A) *SR*, (B) *ST8*, (C) *sim006* and (D) *C167.4* strains. We used oligopaint probes targeting the *Rsp-like* (red) and *500-bp* (blue) satellites. The scale bar represents 5µm. The inset zooms in on the *X* chromosome with a single focus of *Rsp-like* and *500-bp* in each strain. The arrows within the insets point to each focus associated with the *X* chromosome.

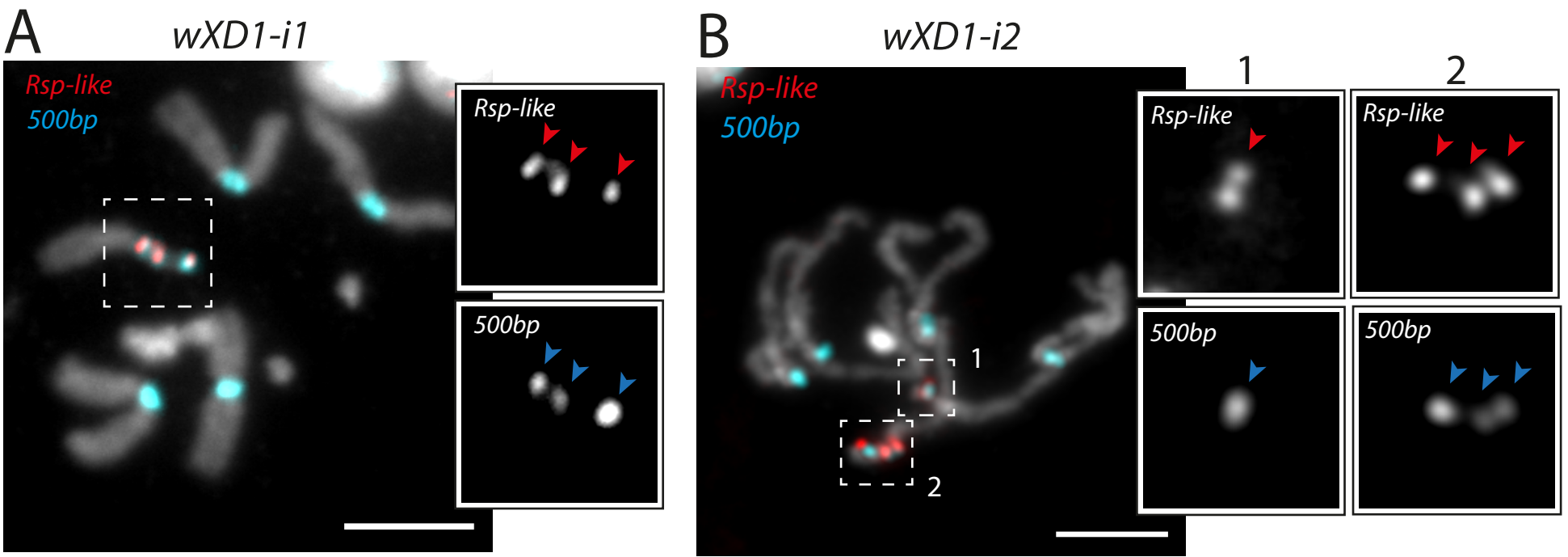
TABLE

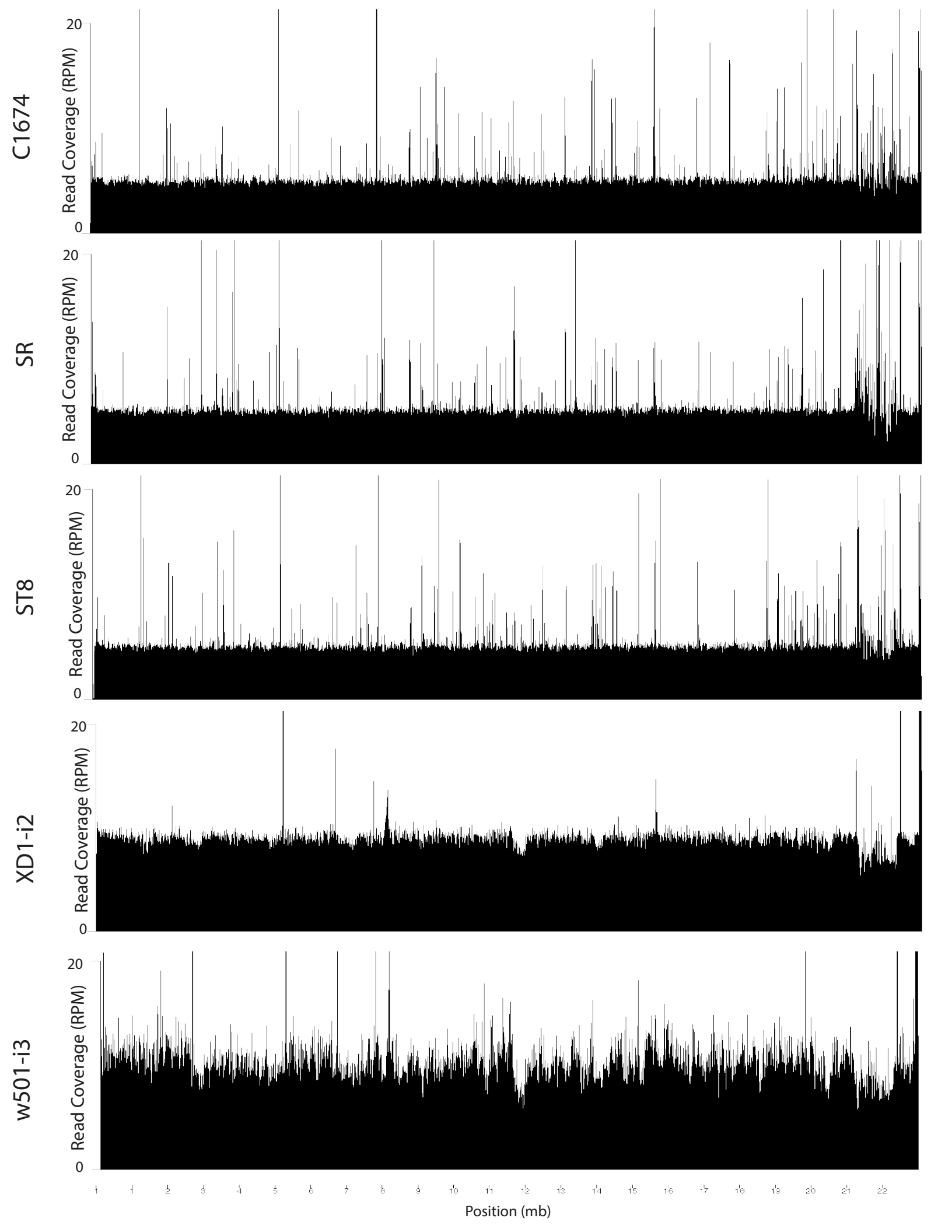
Strain (isolate)	no. of individuals	*X* chromosomes frequencies		
		1-foci	2-foci	3-foci
w^{501} (*w501-i1*)	22	0	0.34	0.66
w^{501} (*w501-i3*)	19	0	0.93	0.07
w^{501} (*w501-i2*)	19	0.21	0.79	0
w^{XD1} (*wXD1-i2*)	40	0.14	0	0.86
w^{XD1} (*wXD1-i1*)	22	0	0	1
SR	18	1	0	0
C167.4	18	1	0	0
ST8	20	1	0	0
Sim006	20	1	0	0

Table 1: A summary of structural variation involving the *500-bp* and *Rsp-like* satellites in the *X* chromosome pericentric heterochromatin within and between isolates of *D. simulans* strains. The isolate identities for w^{501} and w^{XD1} are indicated in parentheses. We report the number of individuals (*i.e.*,brains, which includes both males and females) tested: all spreads examined within an individual brain were consistent (see Materials and Methods). We report the proportion of 1-, 2-, or 3-focus *X* chromosomes among individuals from each isolate. The detailed genotype of each individual tested is presented in SupTable1.


A
w501-i1
Rsp-like
500bp
1
2
Rsp-like
500bp
B
w501-i2
C
w501-i3


A
SR
Rsp-like
500bp
Rsp-like
500bp
B
ST8
Rsp-like
500bp
Rsp-like
500bp
C
sim006
Rsp-like
500bp
Rsp-like
500bp
D
C1674
Rsp-like
500bp
Rsp-like
500bp


A
wXD1-i1
Rsp-like
500bp
Rsp-like
500bp
B
wXD1-i2
Rsp-like
500bp
1
2
1
2
Rsp-like
Rsp-like
500bp
500bp


C1674
SR
ST8
XD1-i2
w501-i3
Read Coverage (RPM)
20
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Position (mb)

SUPPLEMENTAL FIGURES

Supplemental Figure 1. Read coverage across the *X* chromosome assembly is not informative about the presence of structural variation within the isolates. We plotted read coverage (reads per million, RPM) across the *X* chromosomes in five public Illumina datasets for the strains w^{501}, w^{XD1}, *SR*, *ST8*, *C167.8*. The w^{501} and w^{XD1} strains may be polymorphic for multiple *X* pericentromere structural variants but libraries were prepared from pooled females. The breakpoints of possible structural rearrangements that may be present in the w^{501} and w^{XD1} libraries are not obvious from coverage plots on the assembled *X* chromosome. The breakpoint of the structural variant may be beyond the assembled region, but our ability to detect a breakpoint depends on the relative frequency of the different structural variants in the pool of individuals sequenced. We would need multiple biological replicates of each isolate and an assembly that extends through the pericentric heterochromatin to assess whether genomic approaches can detect the structural variation.

w501-i3				
X2Y	**X3Y**	**X2X2**	**X2X3**	**X3X3**
8	1	9	1	0

w501-i2				
X2Y	**X1Y**	**X2X2**	**X2X1**	**X1X1**
6	3	7	3	0

w501-i1				
X2Y	**X3Y**	**X2X2**	**X2X3**	**X3X3**
8	4	1	1	8

wXD1-i1				
X1Y	**X3Y**	**X1X1**	**X1X3**	**X3X3**
9	0	13	0	0

wXD1-i2				
X1Y	**X3Y**	**X1X1**	**X1X3**	**X3X3**
6	11	0	1	19

SR	
X1Y	**X1X1**
9	8

ST8	
X1Y	**X1X1**
13	7

C167.4	
X1Y	**X1X1**
9	9

sim006	
X1Y	**X1X1**
11	9